Introduction to R

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

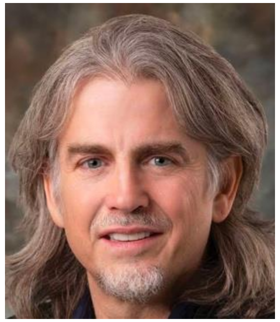
Harvard Chan Bioinformatics Core

Introductions!

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform



John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth Partan

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 Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Alex Bartlett

Elizabeth
Partan

Emma Berdan

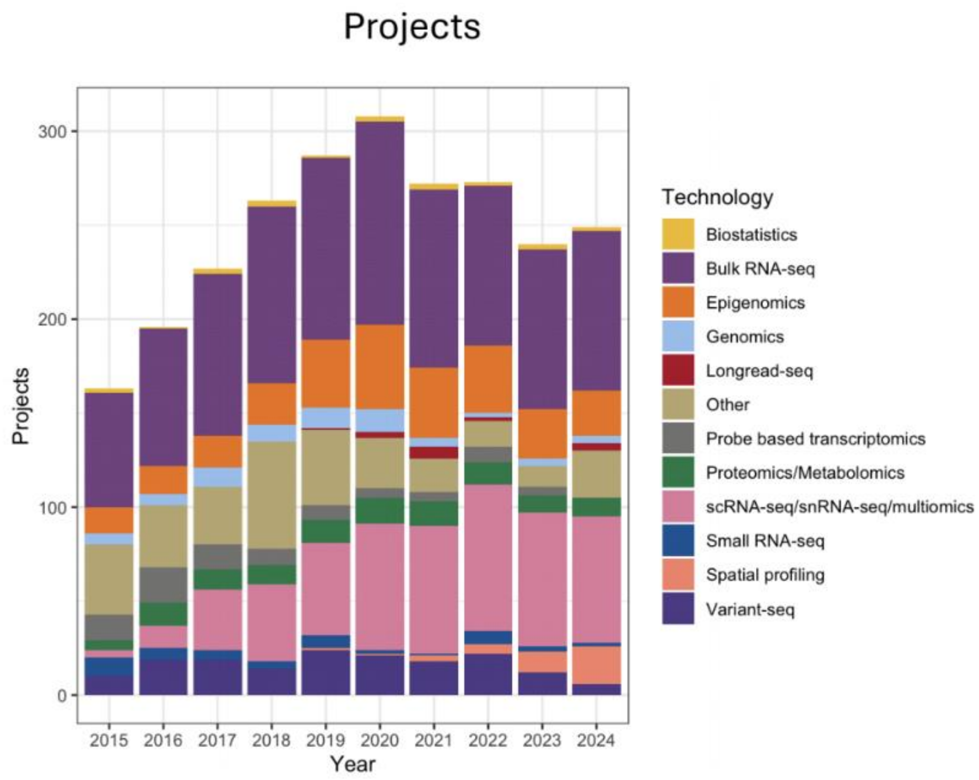
James Billingsley

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



https://bioinformatics.sph.harvard.edu/services

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

HARVARD
CATALYST

THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

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

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

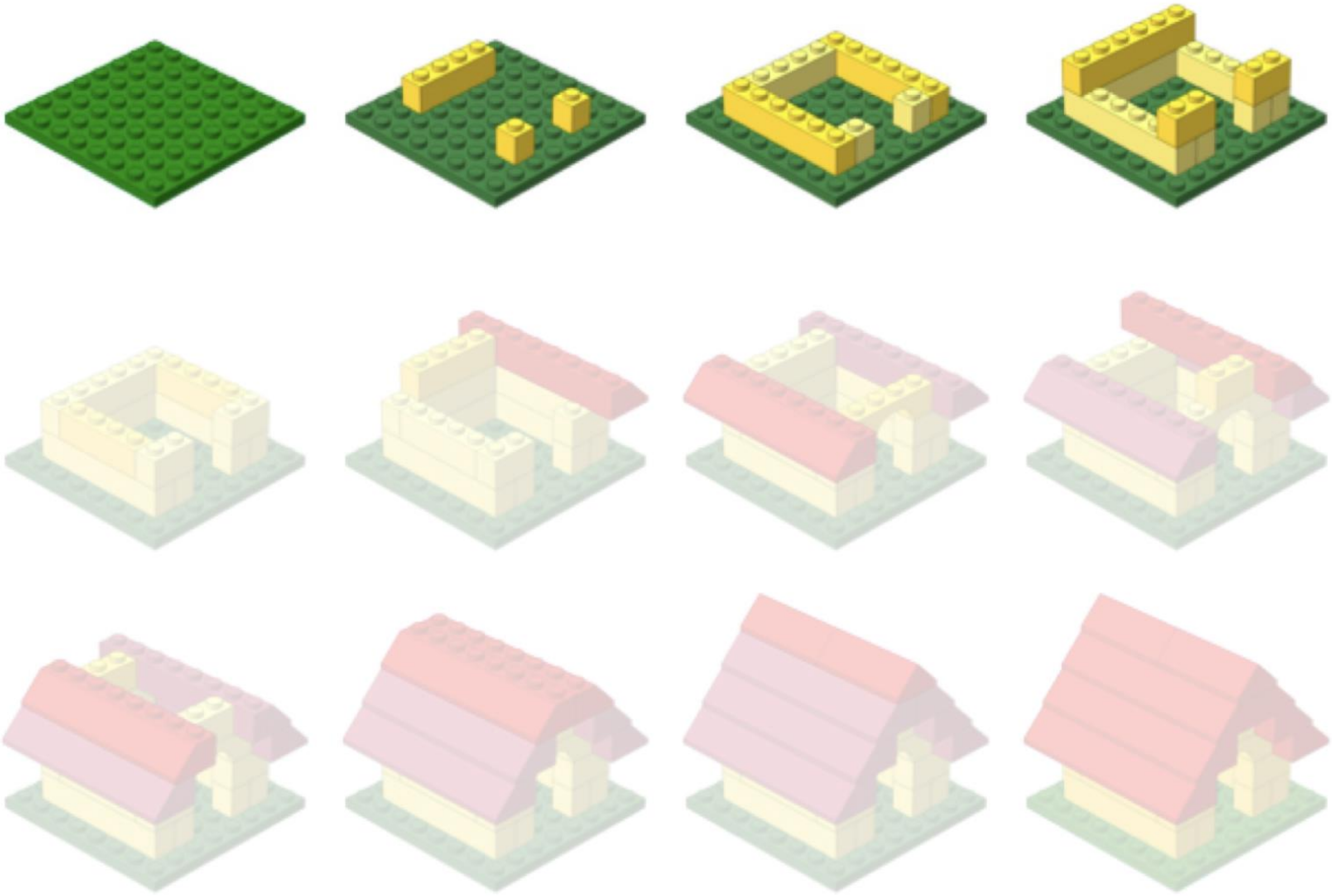
HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD
CATALYST
THE HARVARD CLINICAL AND TRANSLATIONAL SCIENCE CENTER

HARVARD
MEDICAL SCHOOL

Workshop scope

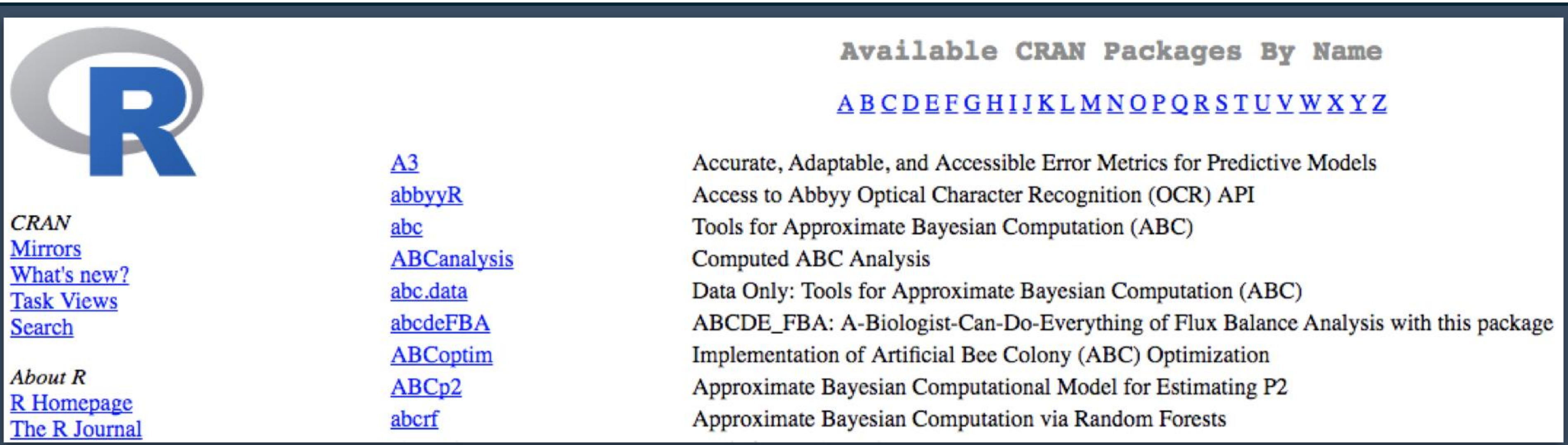


Learning R

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



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

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

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

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

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

Bioconductor

OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

- 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

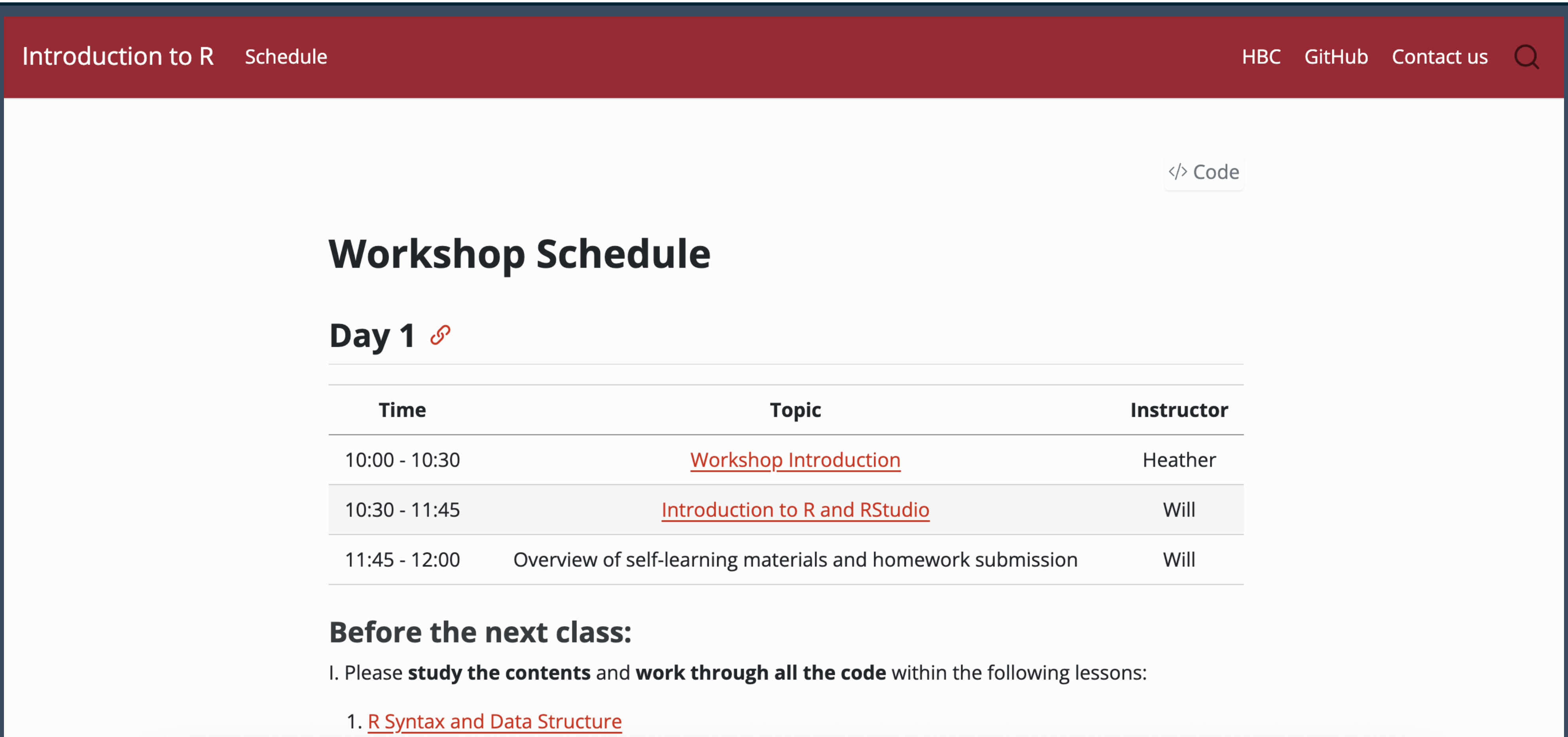
https://www.bioconductor.org

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 schedule



Workshop Schedule

Day 1

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

Before the next class:

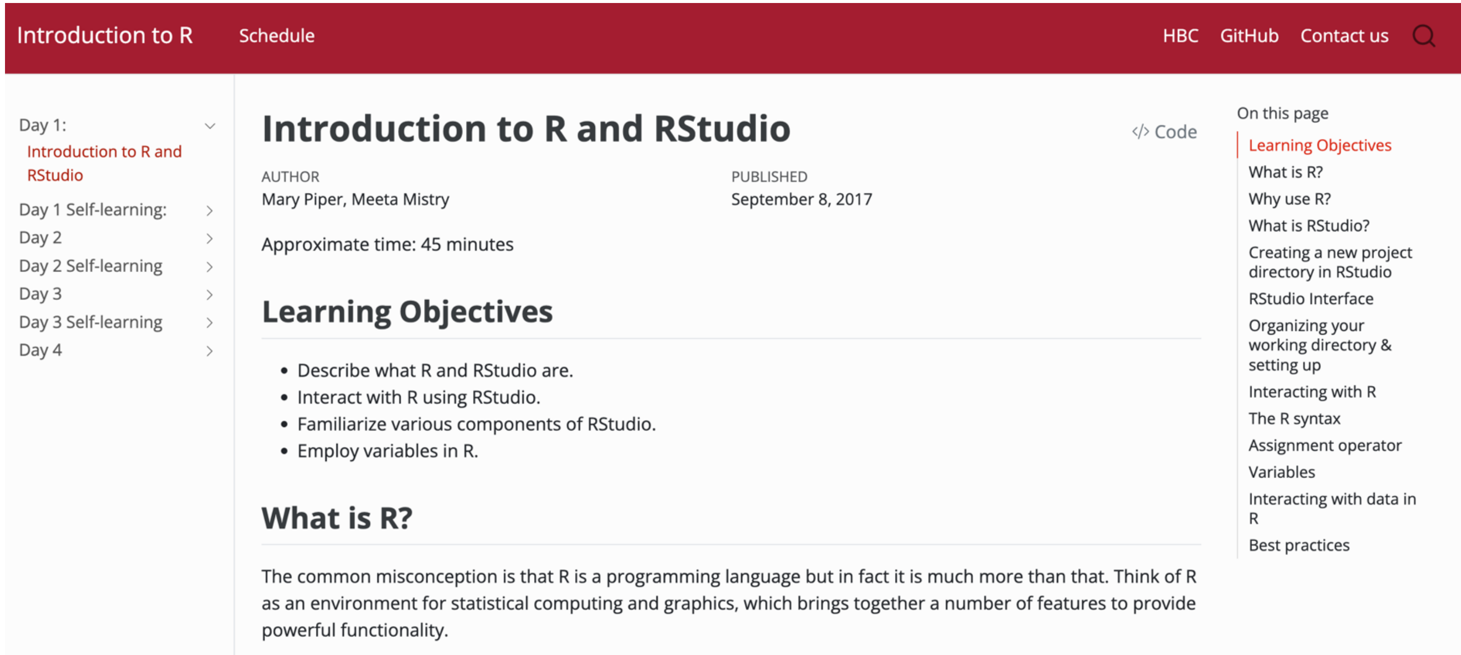
I. Please **study the contents** and **work through all the code** within the following lessons:

1. R Syntax and Data Structure

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

Course materials

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



Introduction to R and RStudio

AUTHOR
Mary Piper, Meeta Mistry

PUBLISHED
September 8, 2017

Approximate time: 45 minutes

Learning Objectives

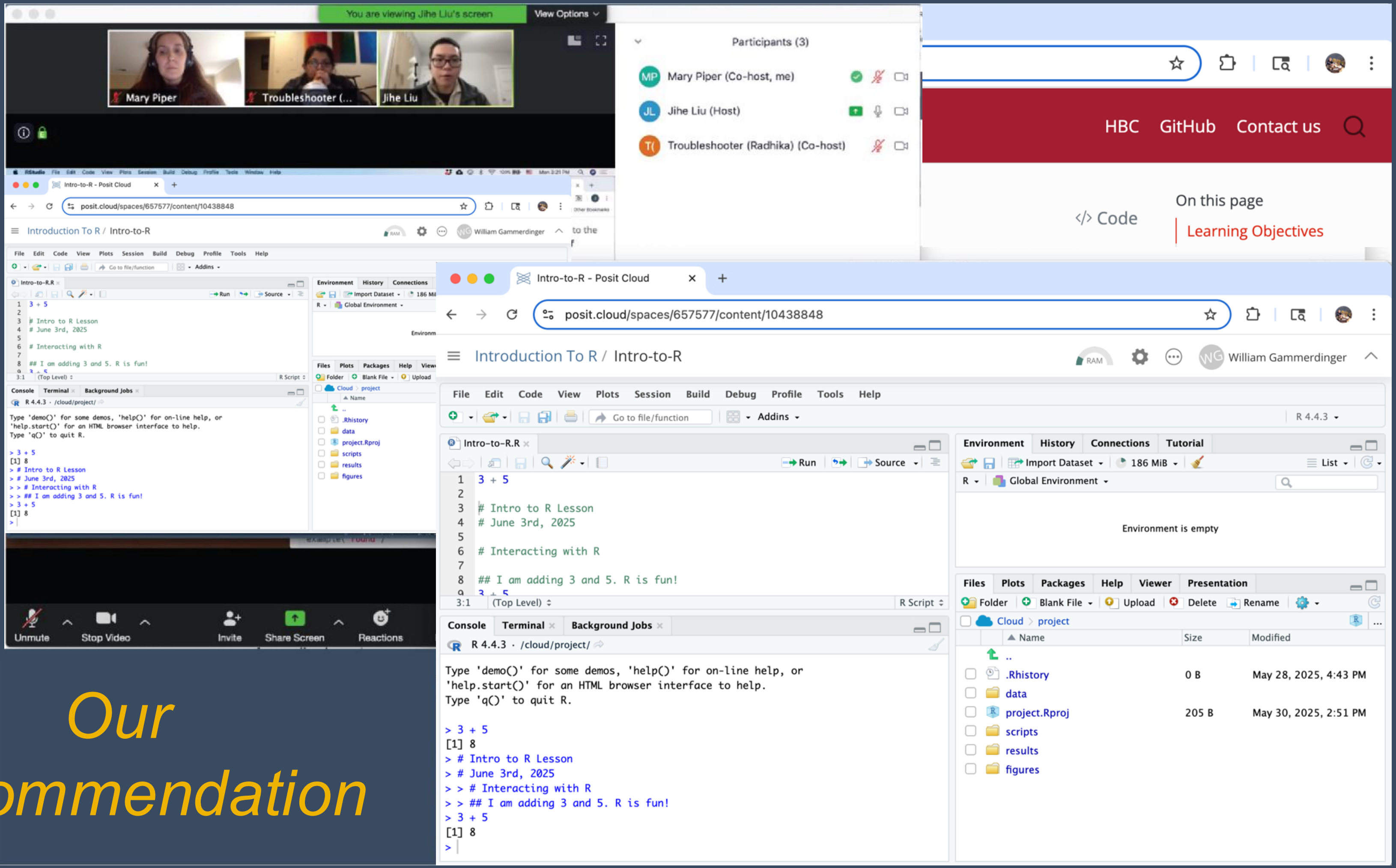
- Describe what R and RStudio are.
- Interact with R using RStudio.
- Familiarize various components of RStudio.
- Employ variables in R.

What is R?

The common misconception is that R is a programming language but in fact it is much more than that. Think of R as an environment for statistical computing and graphics, which brings together a number of features to provide powerful functionality.

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

Single Screen & 3 Windows



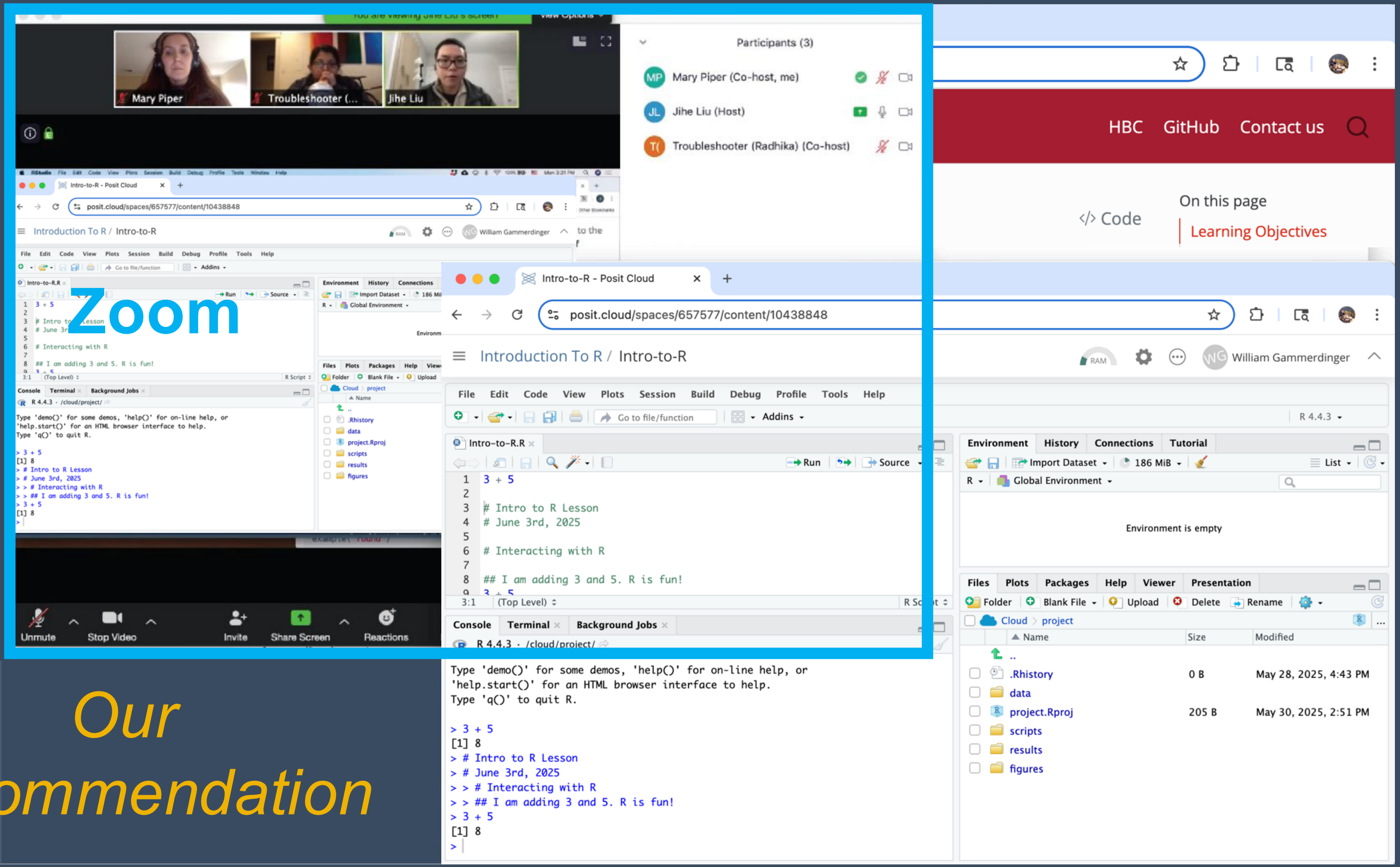
```
3 + 5

# Intro to R Lesson
# June 3rd, 2025

# Interacting with R

## I am adding 3 and 5. R is fun!
```

```
R 4.4.3 · /cloud/project/

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 3 + 5
[1] 8
> # Intro to R Lesson
> # June 3rd, 2025
> > # Interacting with R
> > ## I am adding 3 and 5. R is fun!
> 3 + 5
[1] 8
>
```

Name	Size	Modified
..		
.Rhistory	0 B	May 28, 2025, 4:43 PM
data		
project.Rproj	205 B	May 30, 2025, 2:51 PM
scripts		
results		
figures		

Our Recommendation

Single Screen & 3 Windows



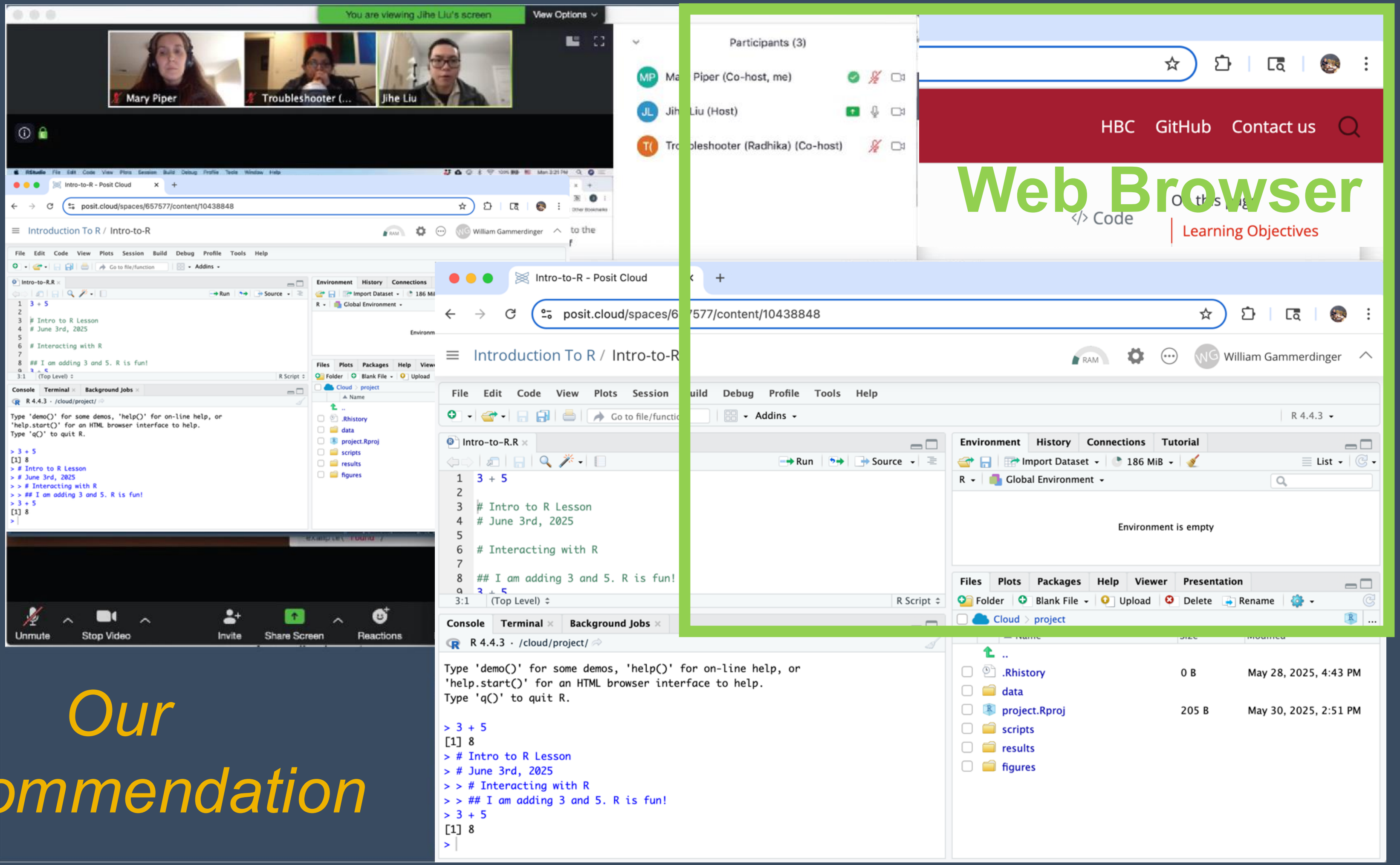
```
3 + 5

# Intro to R Lesson
# June 3rd, 2025

# Interacting with R

## I am adding 3 and 5. R is fun!
3 + 5
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 3 + 5
[1] 8
> # Intro to R Lesson
> # June 3rd, 2025
> > # Interacting with R
> > ## I am adding 3 and 5. R is fun!
> 3 + 5
[1] 8
>
```

Name	Size	Modified
..		
.Rhistory	0 B	May 28, 2025, 4:43 PM
data		
project.Rproj	205 B	May 30, 2025, 2:51 PM
scripts		
results		
figures		

Our Recommendation

Single Screen & 3 Windows



Web Browser

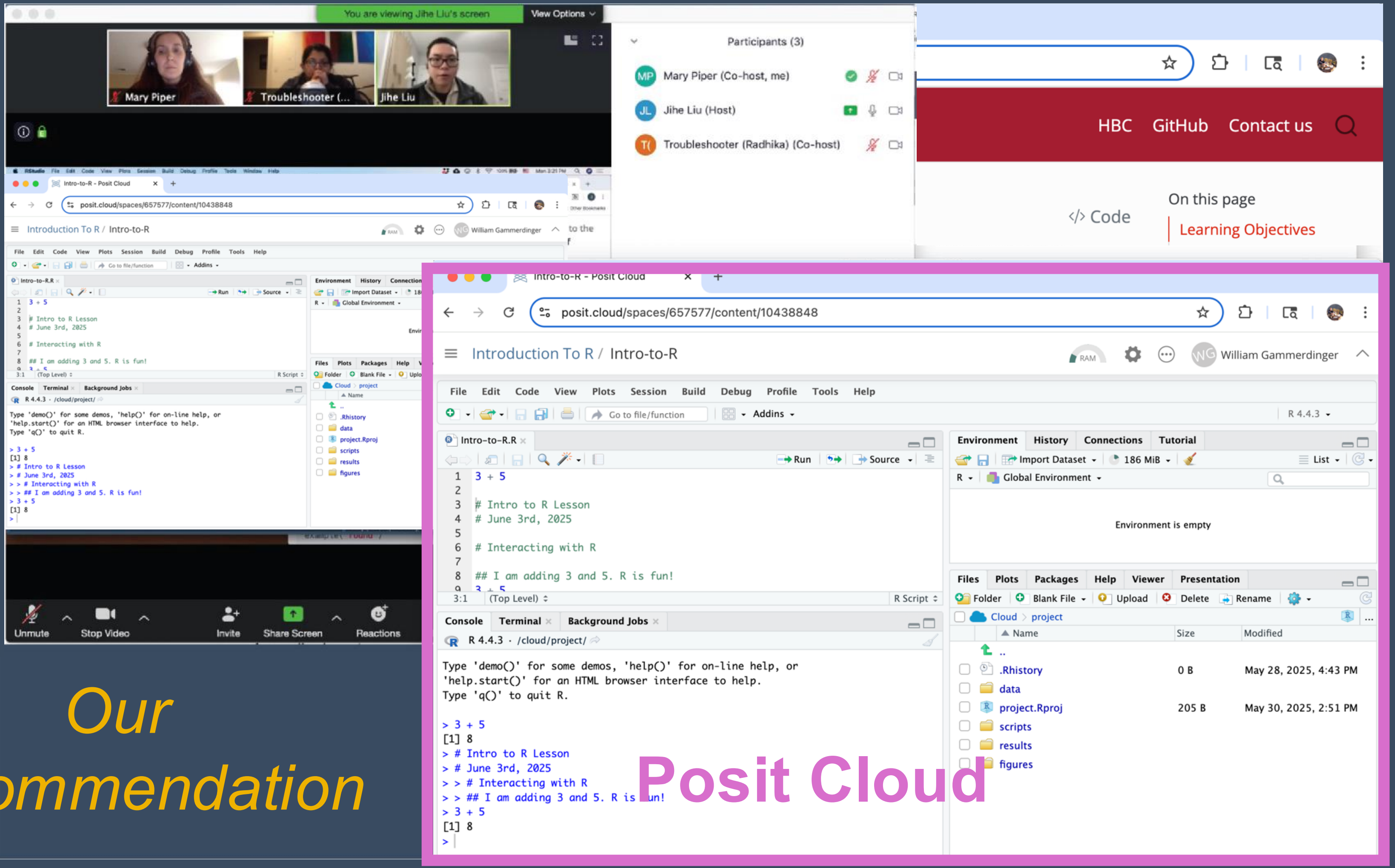
```
3 + 5

# Intro to R Lesson
# June 3rd, 2025

# Interacting with R

## I am adding 3 and 5. R is fun!
3 + 5
```

```
R 4.4.3 · /cloud/project/

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 3 + 5
[1] 8
> # Intro to R Lesson
> # June 3rd, 2025
> > # Interacting with R
> > ## I am adding 3 and 5. R is fun!
> 3 + 5
[1] 8
>
```

Name	Size	Modified
..		
.Rhistory	0 B	May 28, 2025, 4:43 PM
data		
project.Rproj	205 B	May 30, 2025, 2:51 PM
scripts		
results		
figures		

Our Recommendation

Single Screen & 3 Windows



```
3 + 5

# Intro to R Lesson
# June 3rd, 2025

# Interacting with R

## I am adding 3 and 5. R is fun!
3 + 5
```

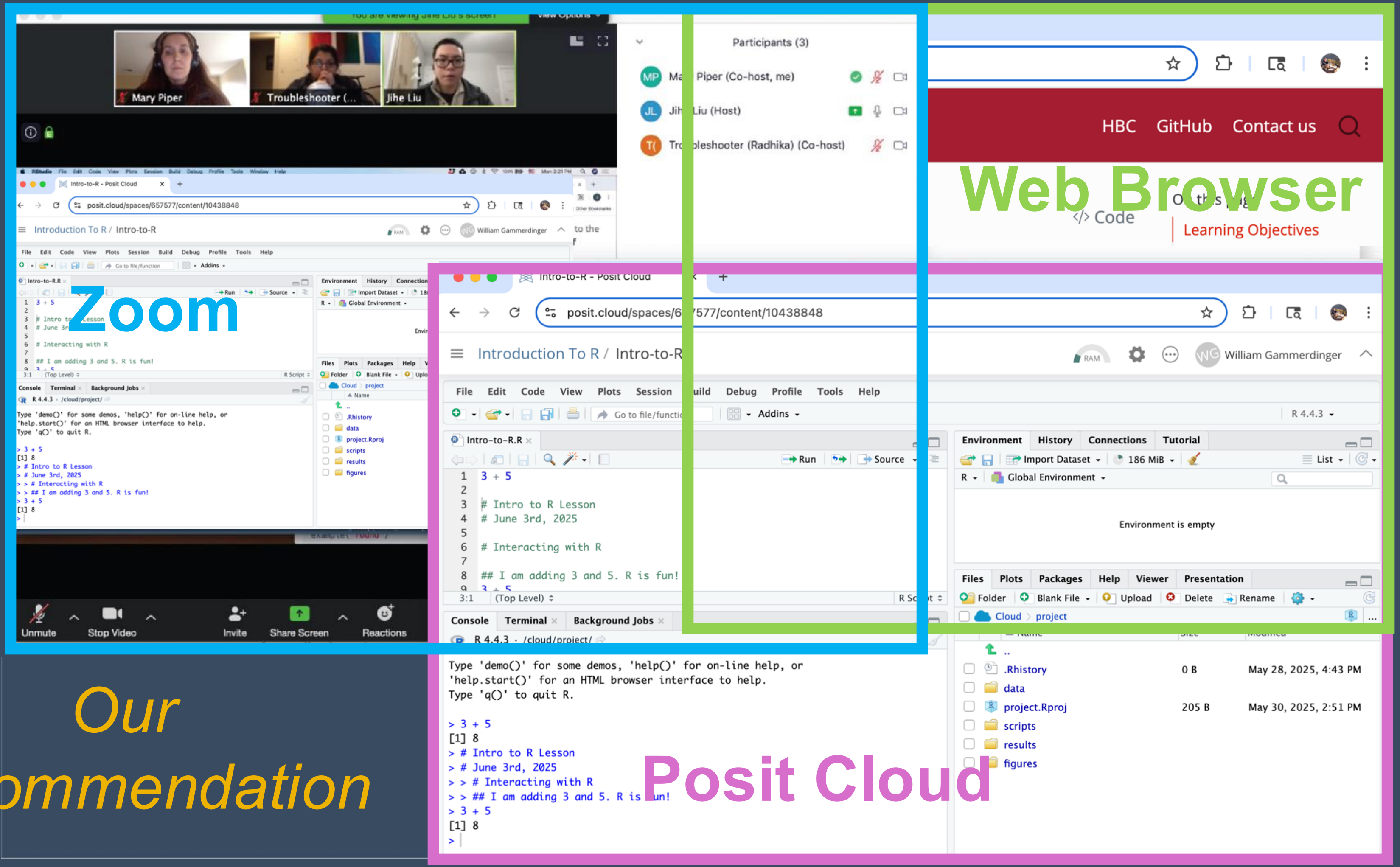
```
R 4.4.3 · /cloud/project/

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 3 + 5
[1] 8
> # Intro to R Lesson
> # June 3rd, 2025
> > # Interacting with R
> > ## I am adding 3 and 5. R is fun!
> 3 + 5
[1] 8
>
```

Name	Size	Modified
.Rhistory	0 B	May 28, 2025, 4:43 PM
data		
project.Rproj	205 B	May 30, 2025, 2:51 PM
scripts		
results		
figures		

Our Recommendation

Posit Cloud

Single Screen & 3 Windows



```
3 + 5

# Intro to R Lesson
# June 3rd, 2025

# Interacting with R

## I am adding 3 and 5. R is fun!
3 + 5
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 3 + 5
[1] 8
> # Intro to R Lesson
> # June 3rd, 2025
> > # Interacting with R
> > ## I am adding 3 and 5. R is fun!
> 3 + 5
[1] 8
>
```

Name	Size	Modified
.Rhistory	0 B	May 28, 2025, 4:43 PM
data		
project.Rproj	205 B	May 30, 2025, 2:51 PM
scripts		
results		
figures		

Our
Recommendation

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

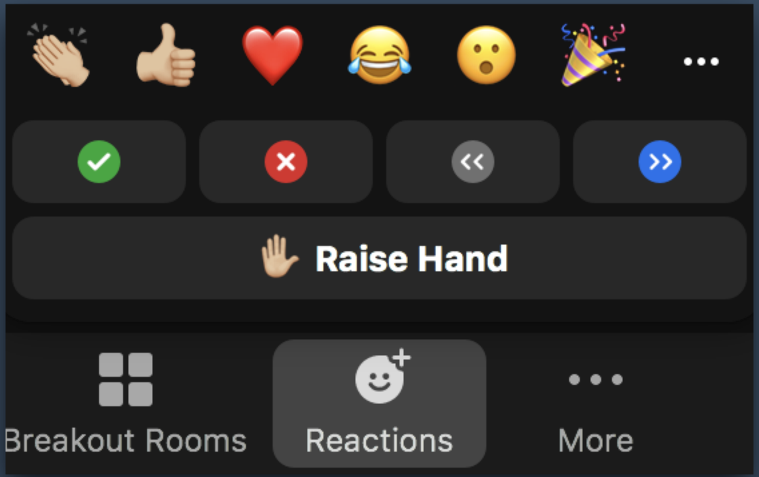
- At-home lessons and exercises after each session
- Cover material not previously discussed
- Provides us feedback to help pace the course appropriately
- 3-5 hours to complete
- Homework load is heavier in the beginning of this workshop series and tapers off

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
  - ✋ Raise Hand when the presenter asks for questions
  - Let the Moderator know

Odds & Ends

- 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 Moderator with a description of the problem

Contact Us

- *HBC training team:* hbctraining@hsph.harvard.edu
- *HBC consulting:* bioinformatics@hsph.harvard.edu