

# Introduction to R

## Invitation to chat:

What was the most enjoyable part of this past weekend?

Harvard Chan Bioinformatics Core

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

Sponsored by DF/HCC, CFAR, and HMS Foundry

# 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/>



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

NIEHS



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER





Shannan Ho Sui  
*Director*



Meeta Mistry  
*Associate Director*



Lorena Pantano  
*Director of Bioinformatics  
Platform*



John Quackenbush  
*Faculty Advisor*



Emma Berdan



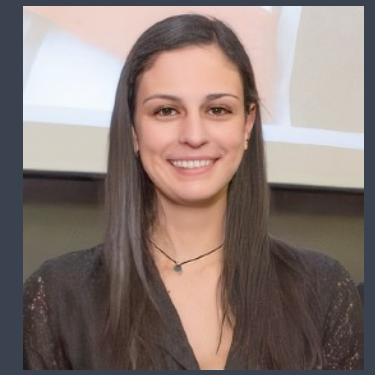
Heather Wick



Will Gammerdinger



Noor Sohail



Alex Bartlett



Upendra Bhattarai



James Billingsley



Zhu Zhuo



Maria Simoneau





Shannan Ho Sui  
*Director*



Meeta Mistry  
*Associate Director*



Lorena Pantano  
*Director of Bioinformatics  
Platform*



John Quackenbush  
*Faculty Advisor*



Emma Berdan



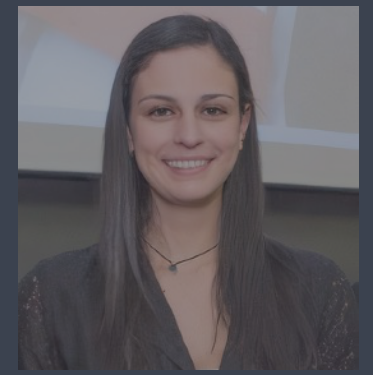
Heather Wick



Will Gammerdinger



Noor Sohail



Alex Bartlett



Upendra Bhattarai



James Billingsley



Zhu Zhuo



Maria Simoneau

# 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 training program is for 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*



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

**DF/HCC**  
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



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

to devote substantial time to material development, training 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\*\*

or the advanced workshops.

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

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



# Training

A key component of the training program is for researchers at Harvard to

[HBC's training team](#) is made up of experts in training and community based research projects to ensure

Our hands-on workshops have an emphasis on **experimentation** for **wet-lab biologists** and their data.

We offer three types of workshops:

1. [Short, 3-hour monthly](#)
2. [Basic Data Skills](#)\*\*
3. [Advanced Topics: Analyzing](#)

\*\*The basic data skills



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

**DF/HCC**  
DANA-FARBER / HARVARD CANCER CENTER



THE HARVARD CLINICAL  
AND TRANSLATIONAL  
SCIENCE CENTER



Our dedicated training team holds workshop to help researchers work with NGS data.

Participants do not have to devote substantial time to material development, but the training team also participate in consultations on best practices in NGS analysis.

Workshops focus on the **analysis of high-throughput sequencing data**, with an emphasis on **reproducibility**. Our workshops are designed to help with planning-based experiments and analysing the resulting

(informatics)

(NGS) data\*\*

for the advanced workshops.

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

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

Introductions!





Shannan Ho Sui  
*Director*



Meeta Mistry  
*Associate Director*



Lorena Pantano  
*Director of Bioinformatics  
Platform*



John Quackenbush  
*Faculty Advisor*



Emma Berdan



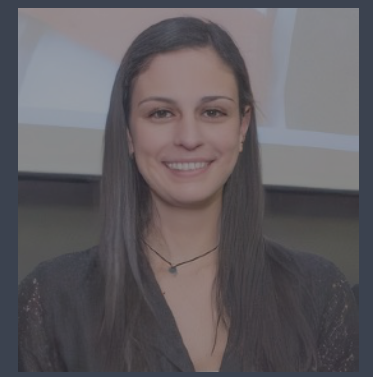
Heather Wick



Will Gammerdinger



Noor Sohail



Alex Bartlett



Upendra Bhattarai



James Billingsley

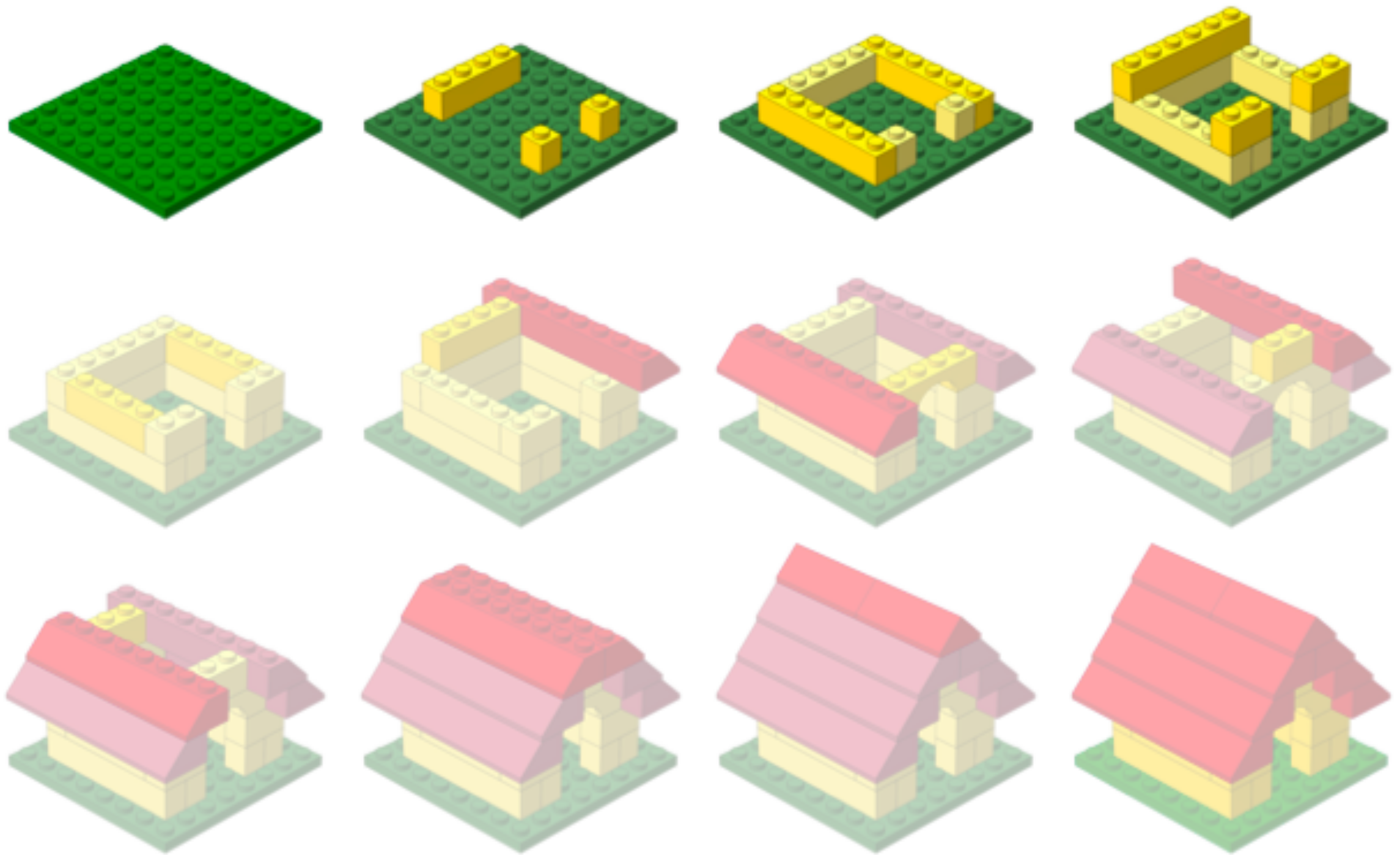


Zhu Zhuo



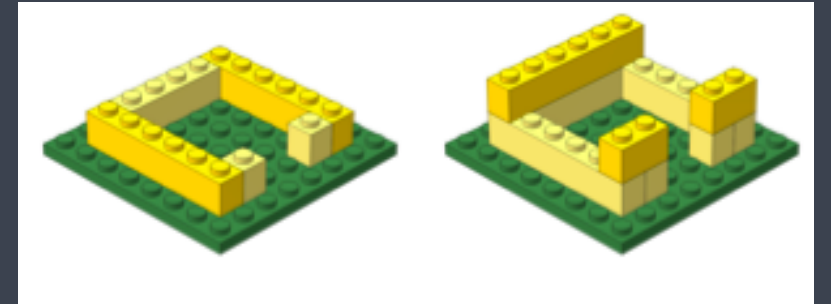
Maria Simoneau

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

(Comprehensive R Archive Network)



[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

- The main repository for R packages
- Easy to install

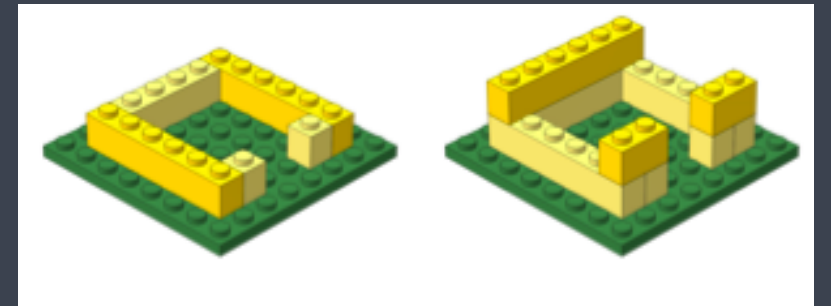
<https://cran.r-project.org/>



- 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](http://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

## Introduction to DGE

[View on GitHub](#)

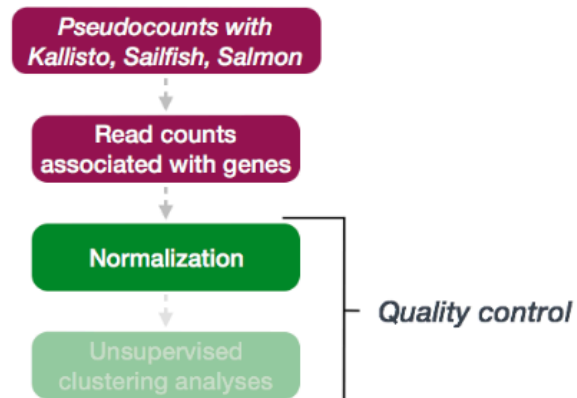
Approximate time: 60 minutes

### 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	<a href="#">Workshop Introduction</a>	Jihe
10:30 - 11:45	<a href="#">Introduction to R and RStudio</a>	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?

The image is a composite of three overlapping windows. At the top is a Zoom meeting window showing three participants: Mary Piper (Co-host, me), Troubleshooter (Radhika) (Co-host), and Jihe Liu (Host). Below the Zoom window is an RStudio window. The RStudio window has a script editor with the following code:

```
1 # Assignment operator
2 x <- 3
3
4 # Functions
5 getwd()
6
7 sqrt(81)
8
9 round(3.14159)
10 ?round
11
```

The RStudio window also shows a console with the following output:

```
> x <- 3
> # Functions
> getwd()
[1] "/Users/marypiper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

At the bottom right is a terminal window showing the execution of the R commands and their output:

```
~/Desktop/R-testing - RStudio
> x <- 3
> # Functions
> getwd()
[1] "/Users/marypiper/Desktop/R-testing"
> sqrt(81)
[1] 9
> round(3.14159)
[1] 3
> ?round
>
```

The RStudio window also shows a documentation pane for the 'rounding of numbers' function, which includes a description of the function and its usage.



# Single screen & 3 windows?

**ZOOM**

*Our recommendation*

```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```

```
1 # Assignment operator  
2 x <- 3  
3  
4 # Functions  
5 getwd()  
6  
7 sqrt(81)  
8  
9 round(3.14159)  
10 ?round  
11
```

```
> x <- 3  
> # Functions  
> getwd()  
[1] "/Users/marypiper/Desktop/R-testing"  
> sqrt(81)  
[1] 9  
> round(3.14159)  
[1] 3  
> ?round  
>
```

**Rounding of Numbers**

**Description**

`ceiling` takes a single numeric argument `x` and returns a numeric vector containing the smallest integers not less than the corresponding elements of `x`.

`floor` takes a single numeric argument `x` and returns a numeric vector containing the largest integers not greater than the corresponding elements of `x`.

`trunc` takes a single numeric argument `x` and returns a numeric vector containing the integers formed by truncating the values in `x` toward 0.

`round` rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

`signif` rounds the values in its first argument to the specified number of significant digits.

**Usage**

```
ceiling(x)  
floor(x)  
trunc(x, ...)
```

# Single screen & 3 windows?

**Web browser**

```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```

```
1 # Assignment operator  
2 x <- 3  
3  
4 # Functions  
5 getwd()  
6  
7 sqrt(81)  
8  
9 round(3.14159)  
10 ?round  
11
```

```
> x <- 3  
> # Functions  
> getwd()  
[1] "/Users/marypiper/Desktop/R-testing"  
> sqrt(81)  
[1] 9  
> round(3.14159)  
[1] 3  
> ?round  
>
```

**R Documentation: Rounding of Numbers**

**Description**

`ceiling` takes a single numeric argument `x` and returns a numeric vector containing the smallest integers not less than the corresponding elements of `x`.

`floor` takes a single numeric argument `x` and returns a numeric vector containing the largest integers not greater than the corresponding elements of `x`.

`trunc` takes a single numeric argument `x` and returns a numeric vector containing the integers formed by truncating the values in `x` toward 0.

`round` rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about "round to even" when rounding off a 5.

`signif` rounds the values in its first argument to the specified number of significant digits.

**Usage**

```
ceiling(x)  
floor(x)  
trunc(x, ...)
```

*Our  
recommendation*

# Single screen & 3 windows?

The image is a composite of three overlapping windows. The top window is a Zoom meeting interface showing three participants: Mary Piper (Co-host, me), Jihe Liu (Host), and Troubleshooter (Radhika) (Co-host). The middle window is an RStudio IDE showing a script with the following code:

```
1 # Assignment operator
2 x <- 3
3
4 # Functions
5 getwd()
6
7 sqrt(81)
8
9 round(3.14159)
10 ?round
11 |
```

The bottom window is an R documentation page for the 'Rounding of Numbers' section. It includes a description of the `ceiling`, `floor`, `trunc`, `round`, and `signif` functions.

**Our recommendation**

**RStudio**

# Single screen & 3 windows?

**Web browser**

```
483  
484  
485 getwd()  
486  
487 # square root function  
488 sqrt(81)  
489  
490 # round function  
491 round(3.14159)  
492 ?round  
493  
494  
495
```

```
1 # Assignment operator  
2 x <- 3  
3  
4 # Functions  
5 getwd()  
6  
7 sqrt(81)  
8  
9 round(3.14159)  
10 ?round  
11 |
```

**ZOOM**

**Our recommendation**

**RStudio**

Participants (3)  
MP Mary Piper (Co-host, me)  
JL Jihe Liu (Host)  
T Troubleshooter (Radhika) (Co-host)

Environment History Connections  
Global Environment  
Values  
number 15  
x 5  
y 10

Environment History Connections  
Global Environment  
Values  
x 3

Rounding of Numbers  
Description  
ceiling takes a single numeric argument x and returns a numeric vector containing the smallest integers not less than the corresponding elements of x.  
floor takes a single numeric argument x and returns a numeric vector containing the largest integers not greater than the corresponding elements of x.  
trunc takes a single numeric argument x and returns a numeric vector containing the integers formed by truncating the values in x toward 0.  
round rounds the values in its first argument to the specified number of decimal places (default 0). See 'Details' about 'round to even' when rounding off a 5.  
signif rounds the values in its first argument to the specified number of significant digits.  
Usage  
ceiling(x)  
floor(x)  
trunc(x, ...)



# 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)

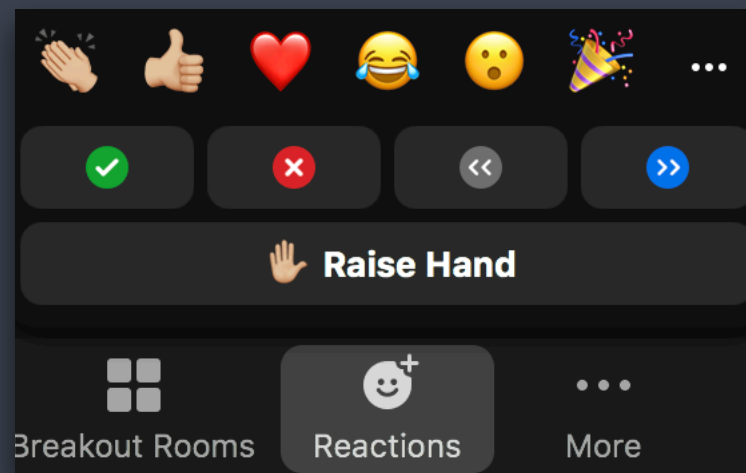
- ❖ Quit/minimize all applications that are not required for class

# 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



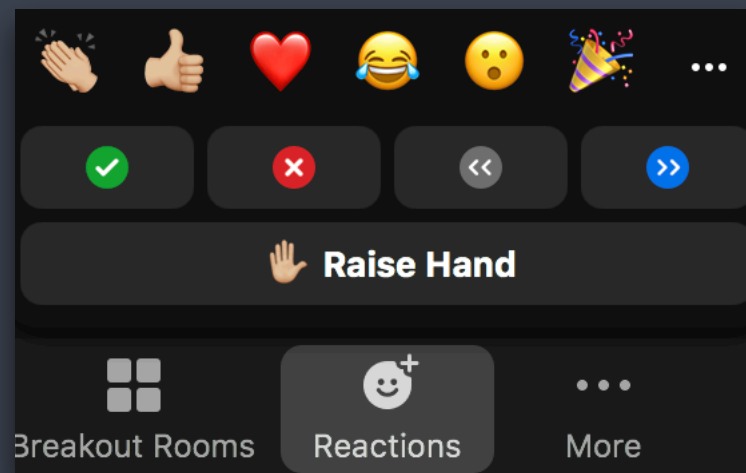
**Raise Hand**

- 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](mailto:hbctraining@hsph.harvard.edu)

*HBC consulting:* [bioinformatics@hsph.harvard.edu](mailto:bioinformatics@hsph.harvard.edu)

Twitter

[@bioinfocore](https://twitter.com/bioinfocore)