

Introduction to R

Harvard Chan Bioinformatics Core

<https://tinyurl.com/intro-to-r-1-5>

Sponsored by HMS (TnT) and HSCI



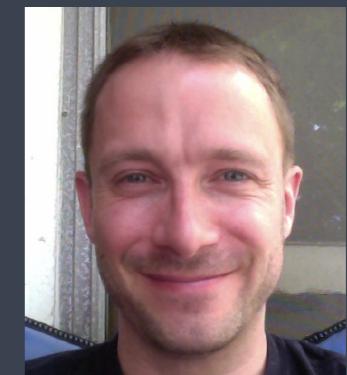
Shannan Ho Sui



John Hutchinson



Brad Chapman



Rory Kirchner



Meeta Mistry



Radhika Khetani



Mary Piper



Victor Barrera



Lorena Pantano

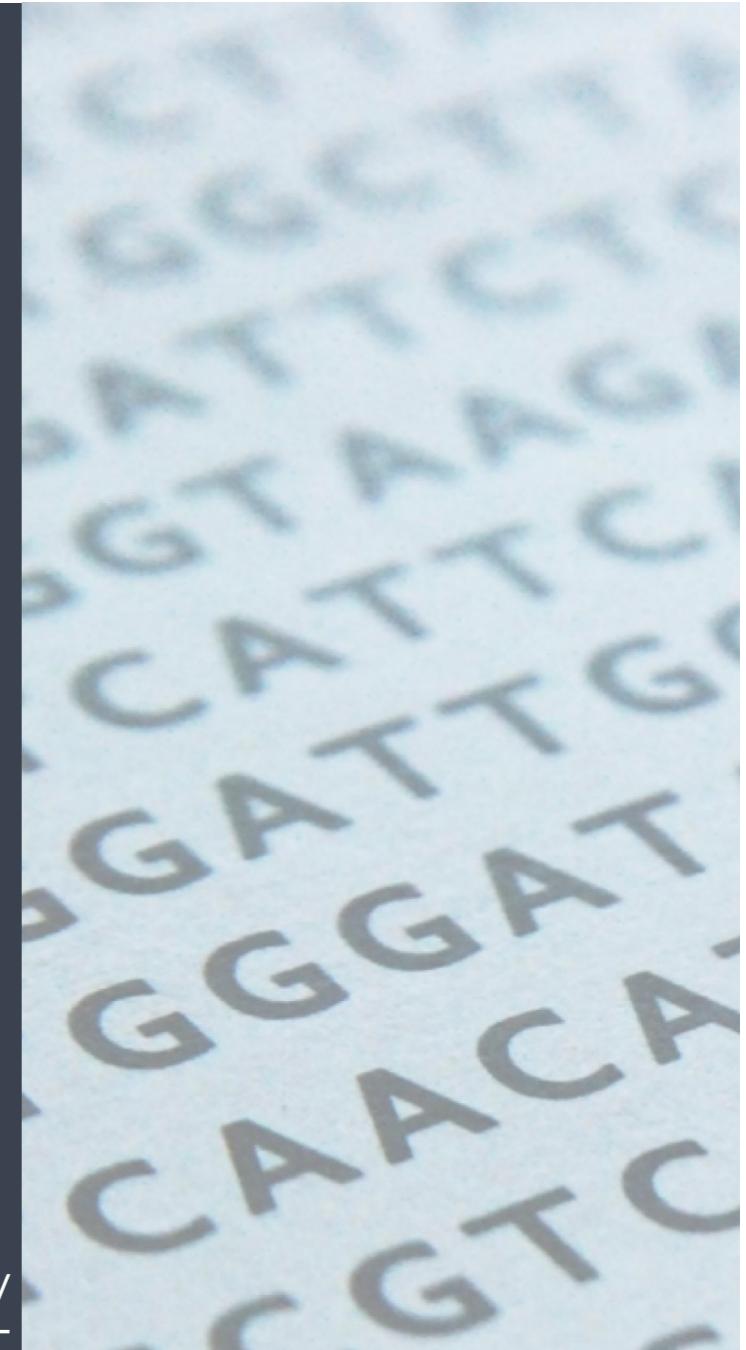


Peter Kraft

Consulting

- RNA-seq analysis: bulk, single cell, small RNA
- ChIP-seq and ATAC-seq analysis
- Genome-wide methylation
- WGS, resequencing, exome-seq and CNV studies
- QC & analysis of gene expression arrays
- Functional enrichment analysis
- Grant support

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

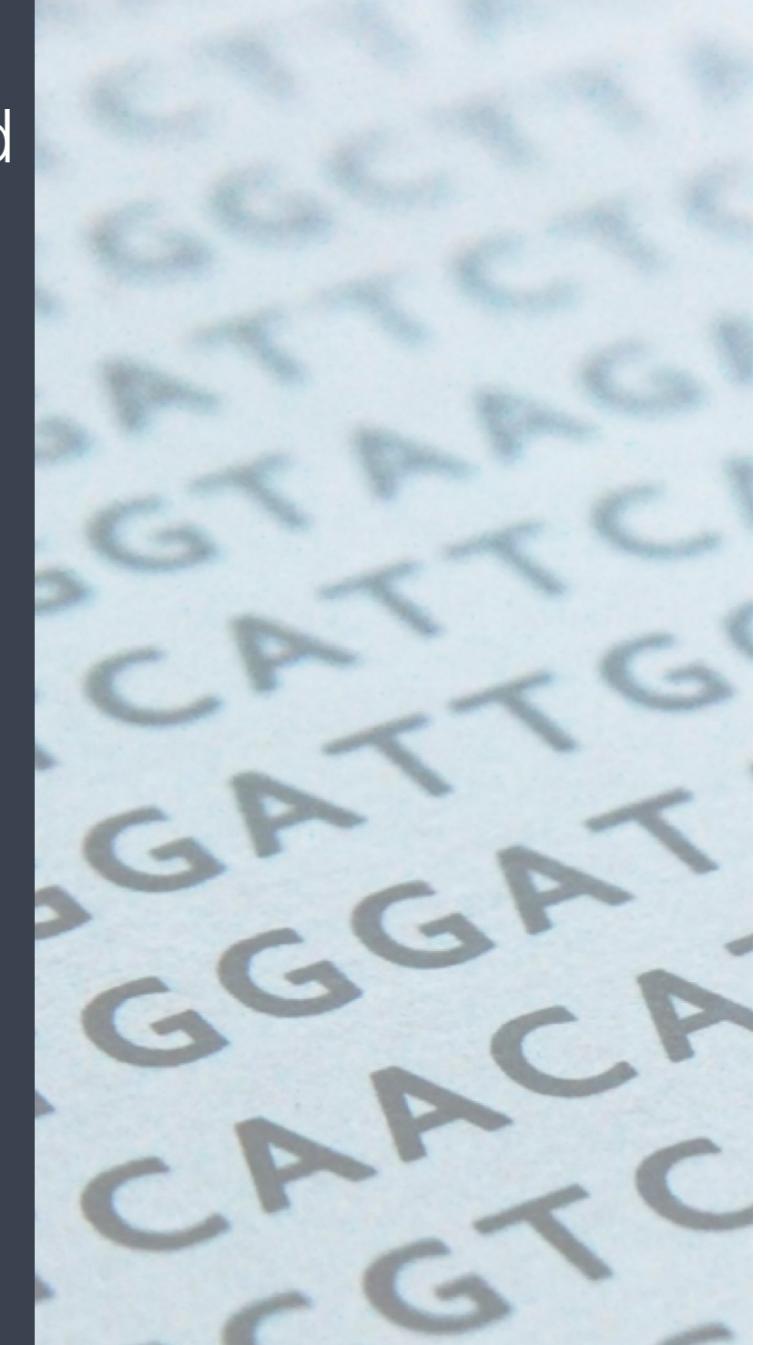


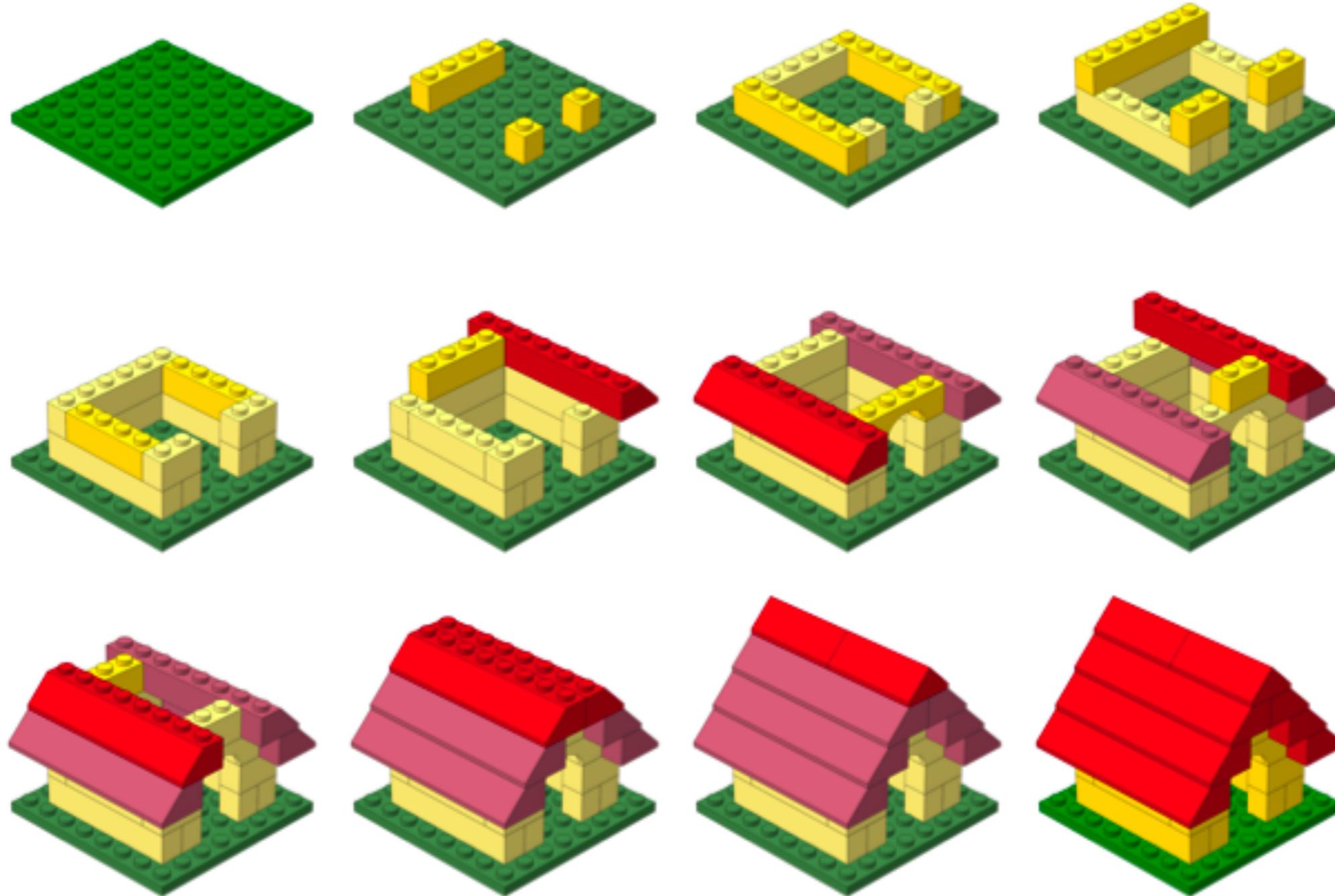
Training

- Short workshops on introductory, intermediate and advanced topics related to NGS data analysis

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

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

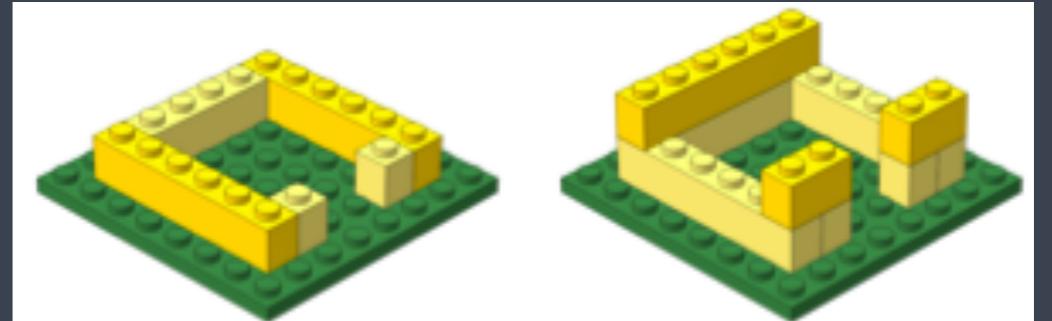




<http://anoved.net/tag/lego/page/3/>

Setting up to perform Bioinformatics analysis

Setting up...



- ✓ Introduction to the command-line interface (shell, Unix, Linux)
 - Dealing with large data files
 - Performing bioinformatics analysis
 - Using tools
 - Accessing and using compute clusters
- ✓ R
 - Parsing and working with smaller results text files
 - Statistical analysis, e.g. differential expression analysis
 - Generating figures from complex data

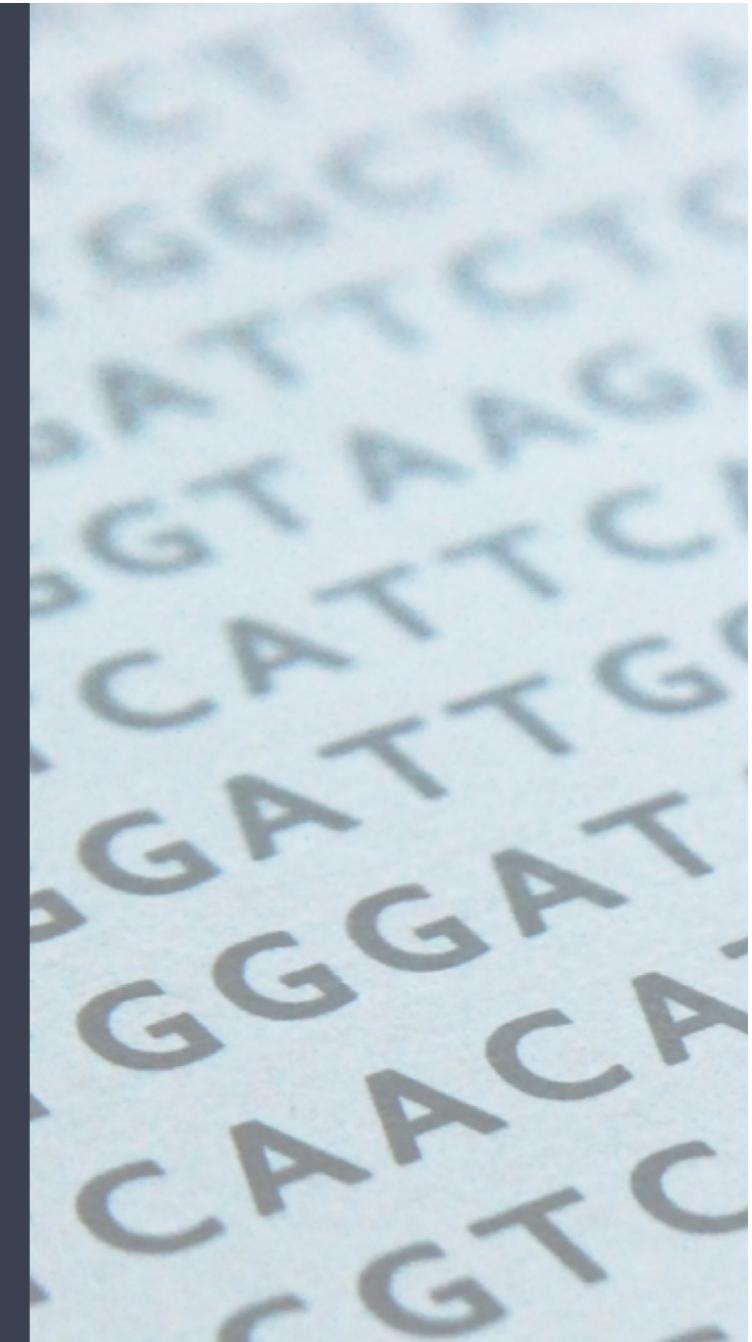
Training

We have divided our short workshops into 2 categories:

1. **Basic Data Skills** - No prior programming knowledge needed (no prerequisites)
2. **Advanced Topics: Analysis of high-throughput sequencing (NGS) data** - Certain “Basic” workshops required as prerequisites.

Any participants wanting to take an advanced workshop will have to have taken the appropriate basic workshop(s) within the past 6 months.

https://hbctraining.github.io/main/training_spring2019.html

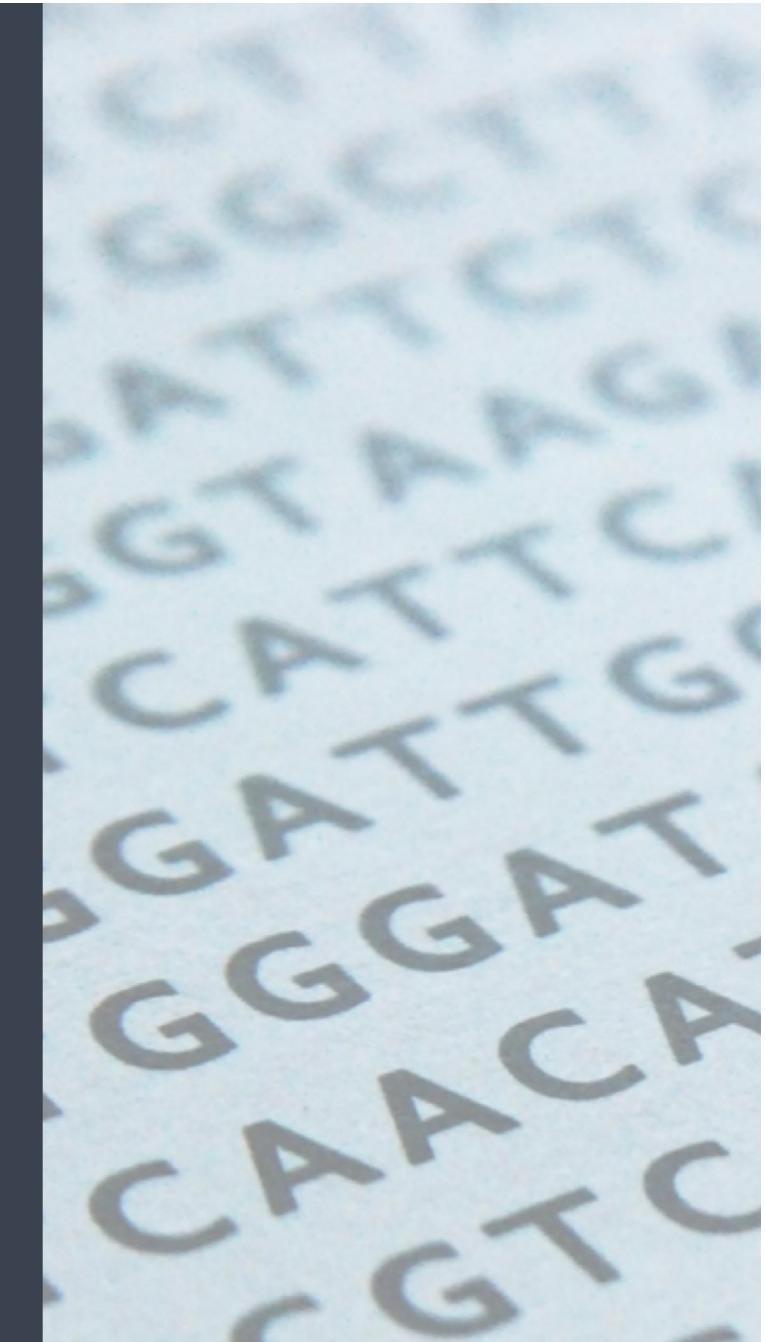


Training

Spring 2019 Schedule

We are structuring the training schedule such that it gives interested researchers several opportunities to take the basic workshops.

Topic	Category	Date	Duration	Prerequisites
Introduction to the command-line interface (shell)	Basic	January 17th	1 day	None
Introduction to R	Basic	January 28th & 29th	1.5 days	None
Introduction to the command-line interface (shell)	Basic	February 13th	1 day	None
Introduction to (bulk) RNA-seq	Advanced	Late February	2 days	Intro to shell
Introduction to R	Basic	Early March	1.5 days	None
Introduction to differential gene expression analysis (bulk RNA-seq)	Advanced	Late March	2 days	Intro to R
Introduction to the command-line interface (shell)	Basic	Early April	1 day	None
Introduction to ChIP-seq analysis	Advanced	Late April	2 days	Intro to R + shell

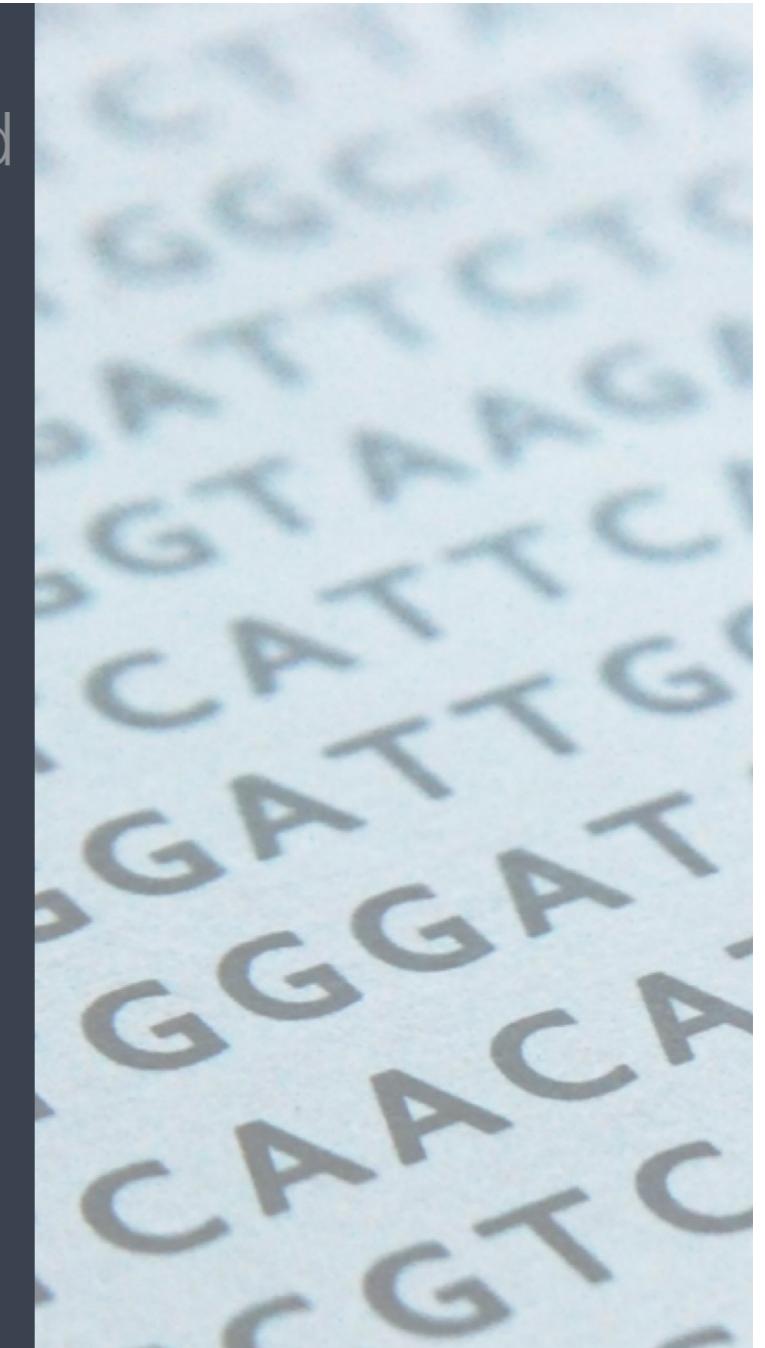


Training

- Short workshops on introductory, intermediate and advanced topics related to NGS data analysis
- Monthly, 2-3 hour, hands-on and free workshops on “Current Topics in Bioinformatics”

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

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

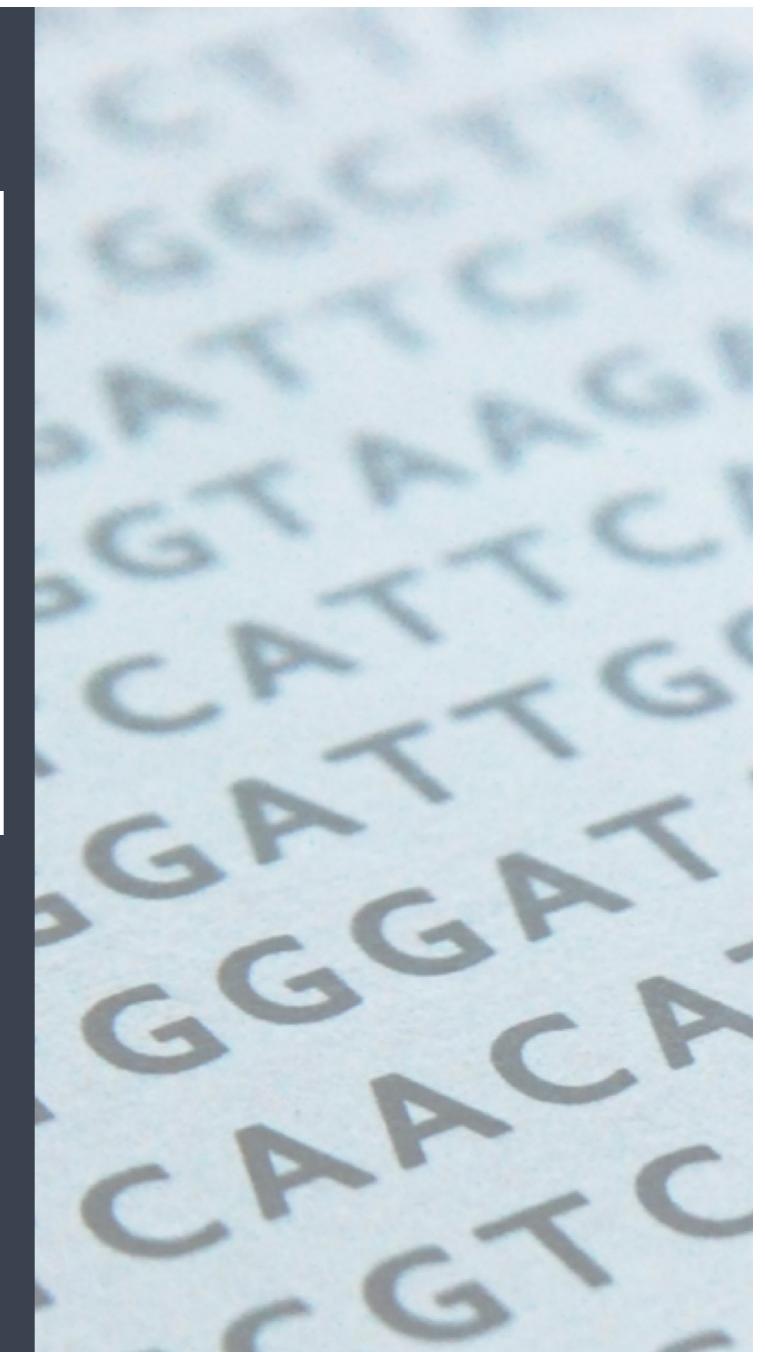


Training

<https://tinyurl.com/spring2019-modules>

Spring 2019 Schedule:

Topic	Date	Time	Location	Pre requisites
Introduction to R	January 10th	1 PM	Kresge (HSPH), room 502	None
Introduction to tidyverse and visualizations with ggplot2	February 6th	1 PM	FXB (HSPH), room G12	Beginner R
Gene annotations and functional analysis of gene lists	March 6th	1 PM	FXB (HSPH), room G12	Beginner R
Generating research analysis reports with RMarkdown	April 3rd	1 PM	FXB (HSPH), room G12	Beginner R

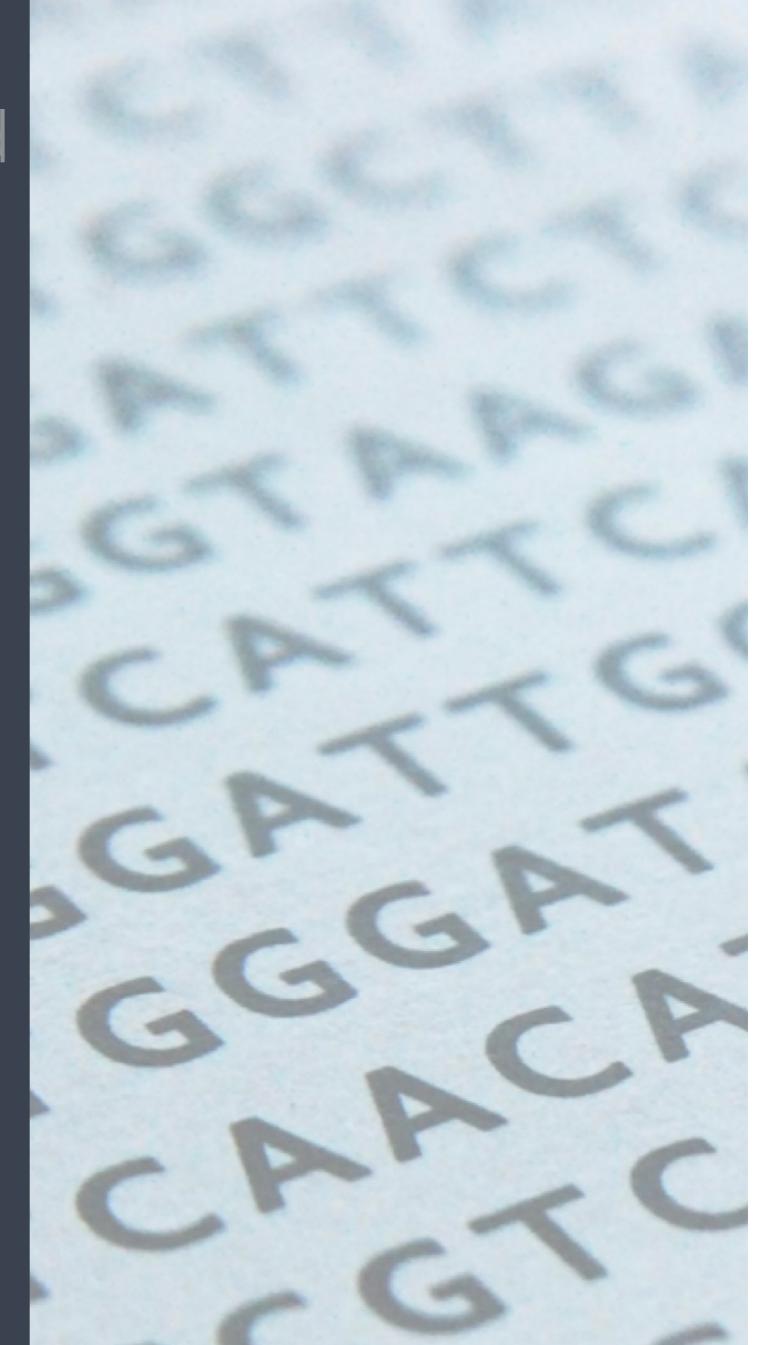


Training

- Short workshops on introductory, intermediate and advanced topics related to NGS data analysis
- Monthly, 2-3 hour, hands-on and free workshops on “Current Topics in Bioinformatics”
- In-depth courses (8- or 12-day formats)

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

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





HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS / CFAR
Bioinformatics
Core

HSCI
HARVARD STEM CELL
INSTITUTE

Center for Stem
Cell
Bioinformatics

 **HARVARD CATALYST**
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

 **HARVARD**
MEDICAL SCHOOL

Harvard
Catalyst
Bioinformatics
Consulting

HMS
Tools &
Technology
Committee



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS / CFAR
Bioinformatics
Core

HSCI
HARVARD STEM CELL
INSTITUTE

Center for Stem
Cell
Bioinformatics

 **HARVARD CATALYST**
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

 **HARVARD**
MEDICAL SCHOOL

Harvard
Catalyst
Bioinformatics
Consulting

HMS
Tools &
Technology
Committee



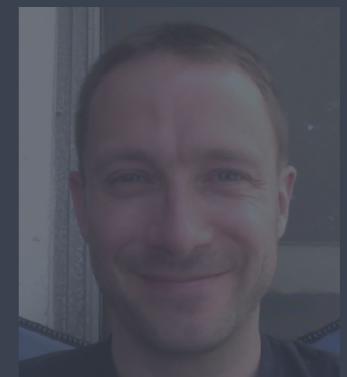
Shannan Ho Sui



John Hutchinson



Brad Chapman



Rory Kirchner



Meeta Mistry



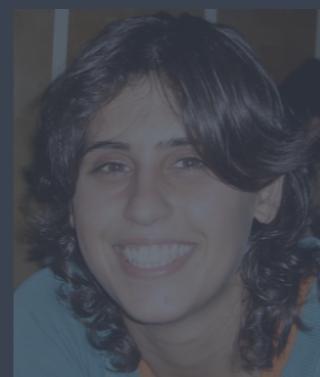
Radhika Khetani



Mary Piper



Victor Barrera



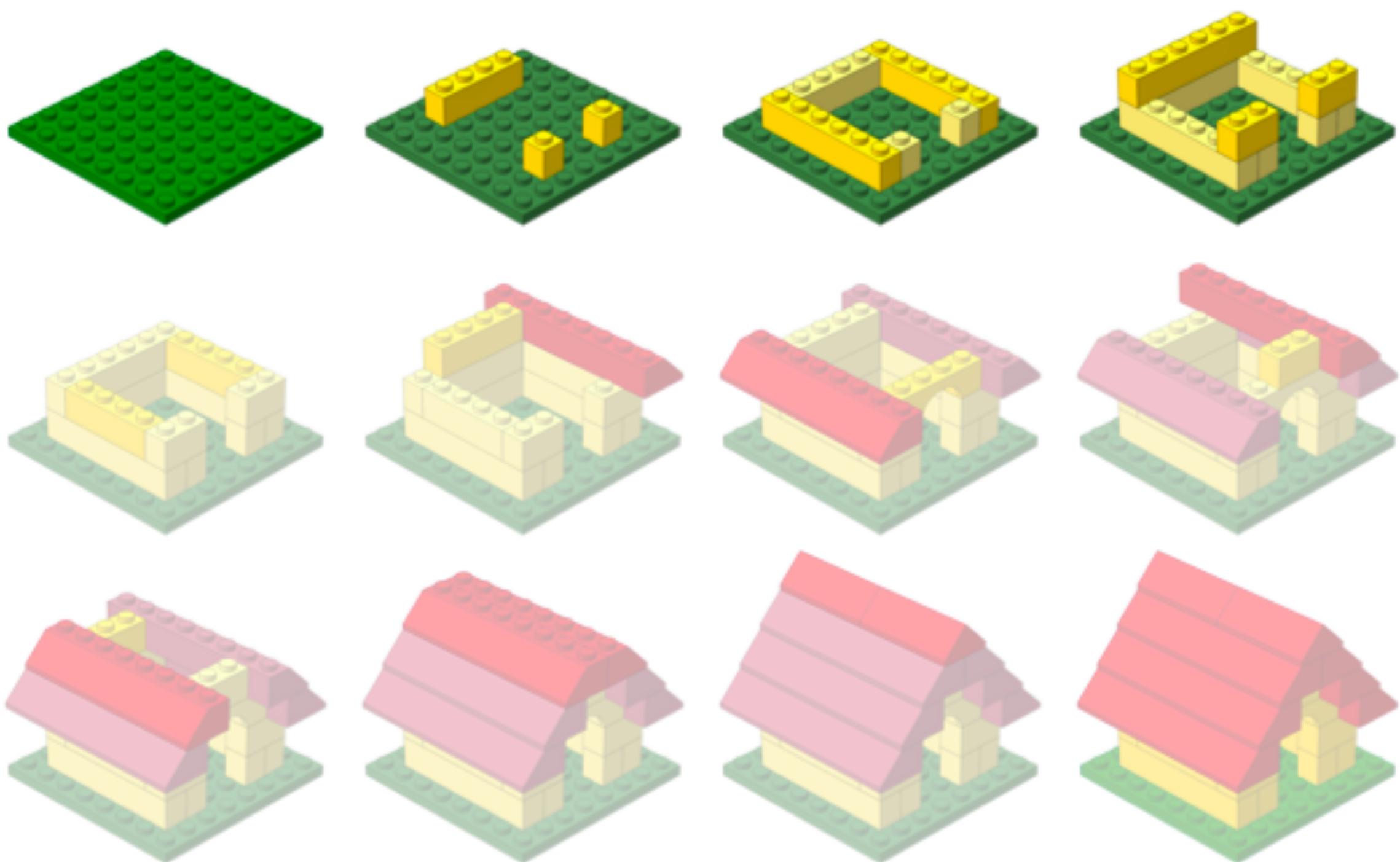
Lorena Pantano



Peter Kraft

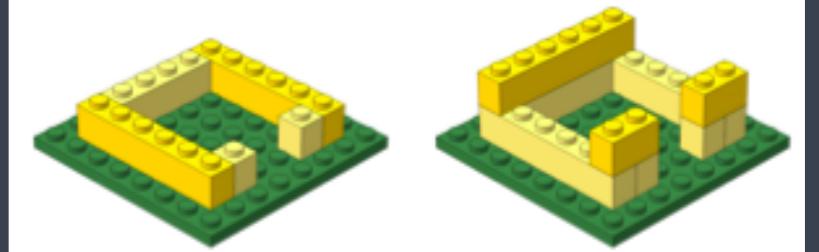
Class Introductions!

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
- ✓ Visualize data using *ggplot2*

Logistics

Course webpage (wiki)

<https://tinyurl.com/intro-to-r-1-5>

Course materials online



Approximate time: 70 min

Learning Objectives

- Employ variables in R.
- Describe the various data types used in R.
- Construct data structures to store data.

The R syntax

Now that we know how to talk with R via the script editor or the console, we want to use R for something more than adding numbers. To do this, we need to know more about the R syntax.

Below is an example script highlighting the many different “parts of speech” for R (syntax):

- the **comments** `#` and how they are used to document function and its content
- **variables** and **functions**
- the **assignment operator** `<-`

The 2 Window problem...

The screenshot shows the RStudio interface with two windows overlapping. The top window is titled "Intro-to-R.R" and contains R code. The bottom window is titled "Console" and shows the R environment.

Code in Intro-to-R.R:

```
351 animals[4, 2] <- "Gray"
352 animals$color <- factor(animals$color)
353 animals$new2 <- c(1,2,3)
354
355 vector1 <- c(6:11)
356 data.frame(animals[, 1:2], vector1, animals[, 3:4])
357
358
359
360
361
362:1 (Top Level) ▾
```

Console Output:

```
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.

[Workspace loaded from ~/Dropbox (HBC)/HBC Team Folder (1)/Teaching/Intro-to-R/.RData]

>
```

Environment Tab (Top Window):

Global Environment	6 obs. of 4 variables
behavior	6 obs. of 3 variables
vector1	6 obs. of 1 variables

File Explorer (Bottom Window):

Name	Size	Modified
..		
.RData	5.8 MB	May 3, 2018, 1:40
.Rhistory	17.4 KB	Nov 15, 2018, 1:2
data		
de_sleuth.R	2.6 KB	Oct 10, 2018, 10:2
figures		
Intro-to-R.R	11.9 KB	May 1, 2018, 3:38

Text in Bottom Window:

```
rownames(metadata)

metadata[c("sample10", "sample12"),]
```

Selecting using indices with logical operators

With dataframes, similar to vectors, we can use logical vectors for specific columns in the dataframe to select only the rows in a dataframe with TRUE values at the same position or index as in the logical vector. We can then use the logical vector to return all of the rows in a dataframe where those values are TRUE.

Odds and Ends

- ❖ Name tags: Tent Cards
- ❖ Post-its
- ❖ Phones on vibrate/silent

Contact us!

HBC training team: hbctraining@hsph.harvard.edu

HBC consulting: bioinformatics@hsph.harvard.edu

Twitter

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