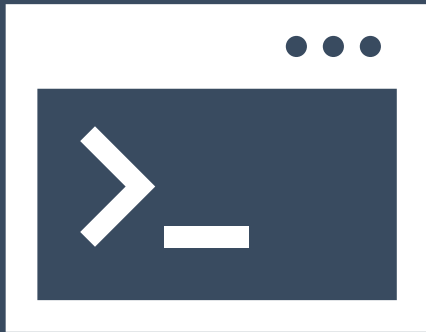


Introduction to R

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



Harvard Chan Bioinformatics Core



Introductions!





Shannan Ho Sui
Director



Lorena Pantano
*Director of Bioinformatics
Platform*



John Quackenbush
Faculty Advisor



James Billingsley



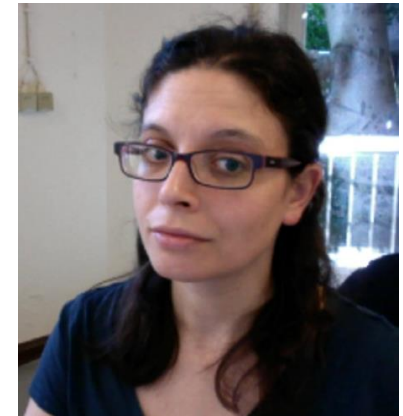
Elizabeth
Partan



Will Gammerdinger



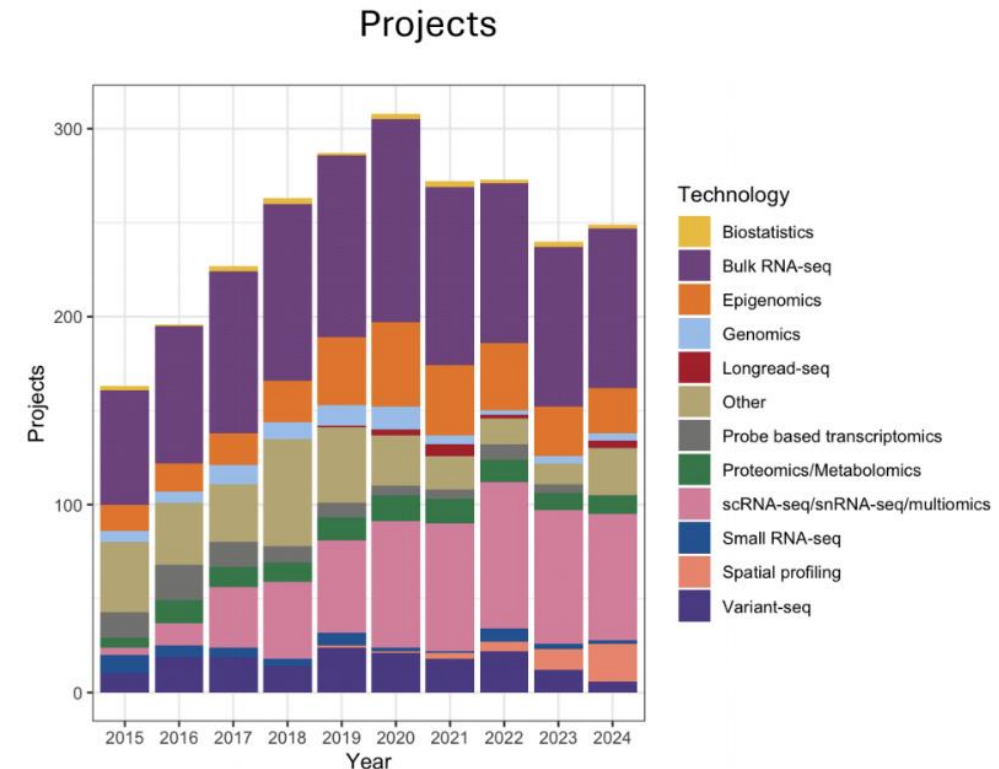
Noor Sohail



Emma Berdan

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



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://hsph.harvard.edu/research/bioinformatics/training/>

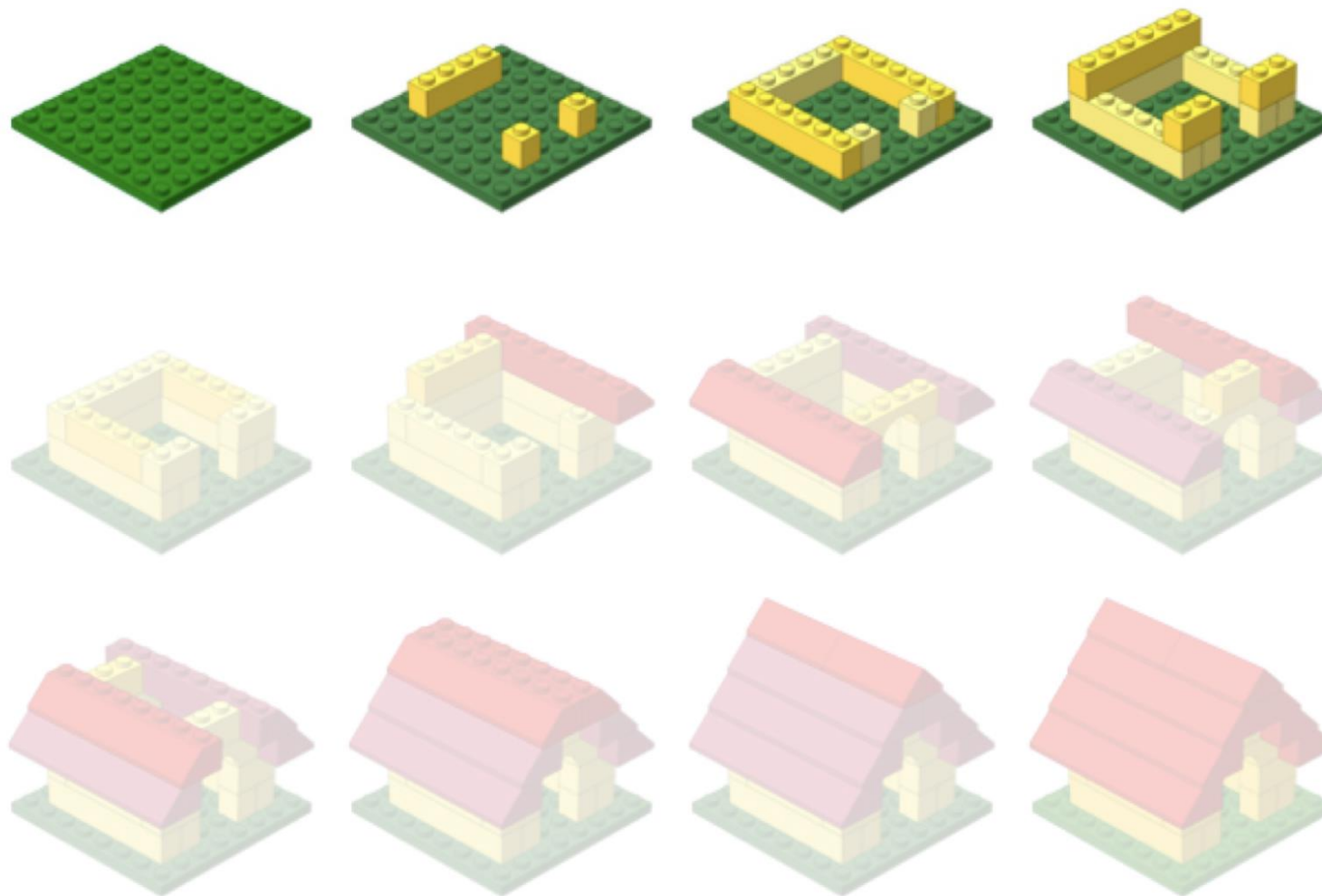
Training

- ❖ 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



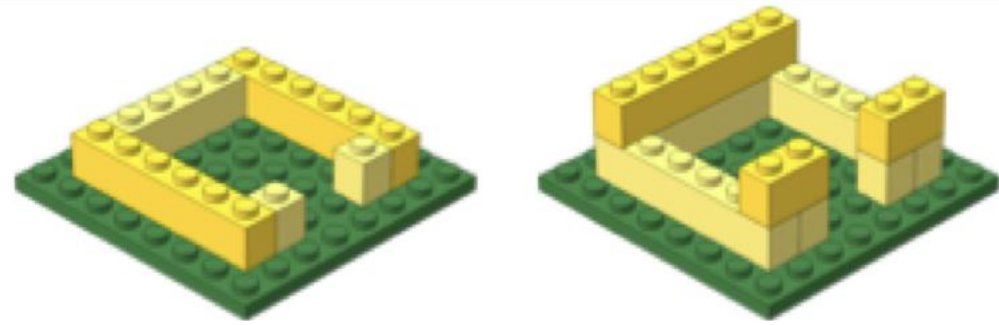
<https://hsph.harvard.edu/research/bioinformatics/training/>

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](#)

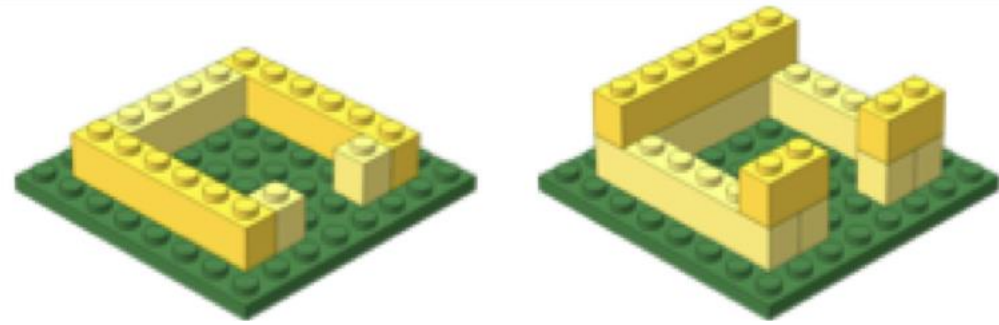
Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
Access to Abbyy Optical Character Recognition (OCR) API
Tools for Approximate Bayesian Computation (ABC)
Computed ABC Analysis
Data Only: Tools for Approximate Bayesian Computation (ABC)
ABCDE_FBA: A-Biologist-Can-Do-Everything of 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

<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

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

[Introduction to R](#) [Schedule](#) [HBC](#) [GitHub](#) [Contact us](#) 

 Code

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](#)

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

Course materials

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

[Introduction to R](#) [Schedule](#) [HBC](#) [GitHub](#) [Contact us](#)

Day 1:
[Introduction to R and RStudio](#)

Day 1 Self-learning: >

Day 2 >

Day 2 Self-learning >

Day 3 >

Day 3 Self-learning >

Day 4 >

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.

</> Code

On this page

[Learning Objectives](#)

[What is R?](#)

[Why use R?](#)

[What is RStudio?](#)

[Creating a new project directory in RStudio](#)

[RStudio Interface](#)

[Organizing your working directory & setting up](#)

[Interacting with R](#)

[The R syntax](#)

[Assignment operator](#)

[Variables](#)

[Interacting with data in R](#)

[Best practices](#)

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

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
- ❖ Name tags
- ❖ Post-its
 - ❖  green - I am all set
 - ❖  red - I need time/help
- ❖ Phones on vibrate/silent
- ❖ Bathrooms

Contact Us

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