

Introduction to R (R Bootcamp)

Mary Piper, PhD

Radhika Khetani, PhD

Harvard Chan Bioinformatics Core

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

What does a bioinformatics core do?

Mission

HBC provides a single point of contact for Harvard researchers interested in bioinformatics support, applying genomic approaches together with established and developing methodologies from epidemiology, environmental health, biostatistics and bioinformatics to improve human health.

Core staff aid researchers within the Harvard community with the management, integration and contextual analysis of biological high-throughput data, focusing on next-generation sequencing support. We provide consulting, training, tools, databases and best practices, foster collaboration and a community of bioinformatics activities, and build a unified infrastructure supporting a diverse set of experimental systems and high-throughput biological data.

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

How does one work with a core?

Working with us

Consults are billed at \$150/hour after an initial meeting to discuss your project. We will provide you with an estimate of the number of hours required, a quote and a start date as part of our [Memorandum of Understanding \(PDF\)](#). During the project you will receive detailed monthly billing updates and ongoing analysis results via our project page.



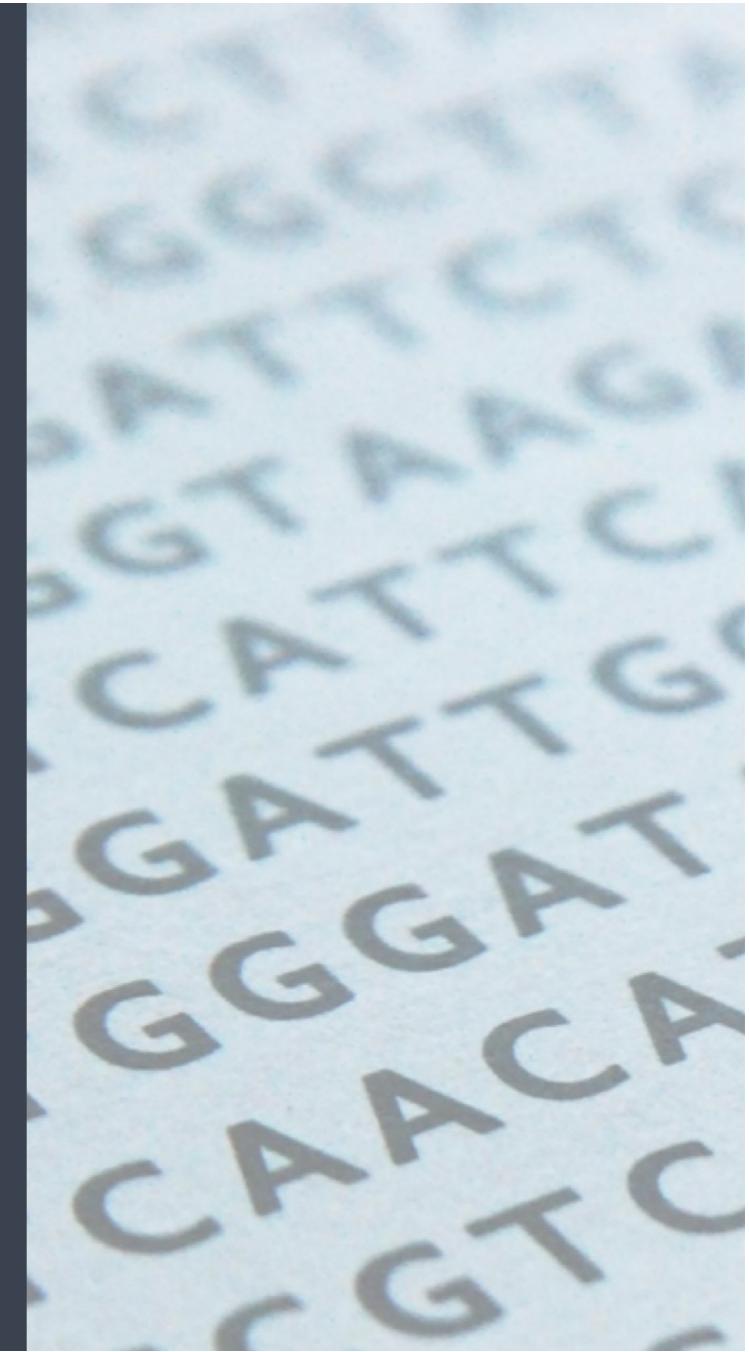
Fees for pilot projects and other short term consults can be subsidized in many cases through support from our funding partners. If in doubt ask us for more information!

Please note that payment of fees for data analysis services and authorship are not mutually exclusive. Depending on their contributions to a project, HBC Data Analysts might be considered collaborators at the same level as other academic colleagues who contribute intellectually and receive funding for work on a project. We follow the general guidelines for authorship laid out by the International Committee of Medical Journal Editors. The recovery of Core expenses through the recharge system does not exclude the possibility for authorship for Core personnel. Similarly, authorship does not substitute for payment of Core expenses for services rendered.

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

Consulting and analysis services

- RNA-seq, small RNA-seq and ChIP-seq analysis
 - Genome-wide methylation
 - WGS, resequencing, exome-seq and CNV studies
 - Quality assurance and analysis of gene expression arrays
 - Functional enrichment analysis
 - Grant support





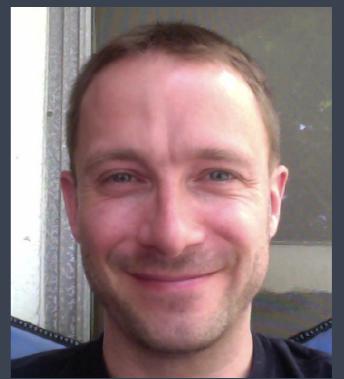
Shannan Ho Sui
Director



John Hutchinson
Associate Director



Victor Barrera



Rory Kirchner



Meeta Mistry



Radhika Khetani
Training Director



Mary Piper



James Billingsley



Zhu Zhuo



Sergey Naumenko



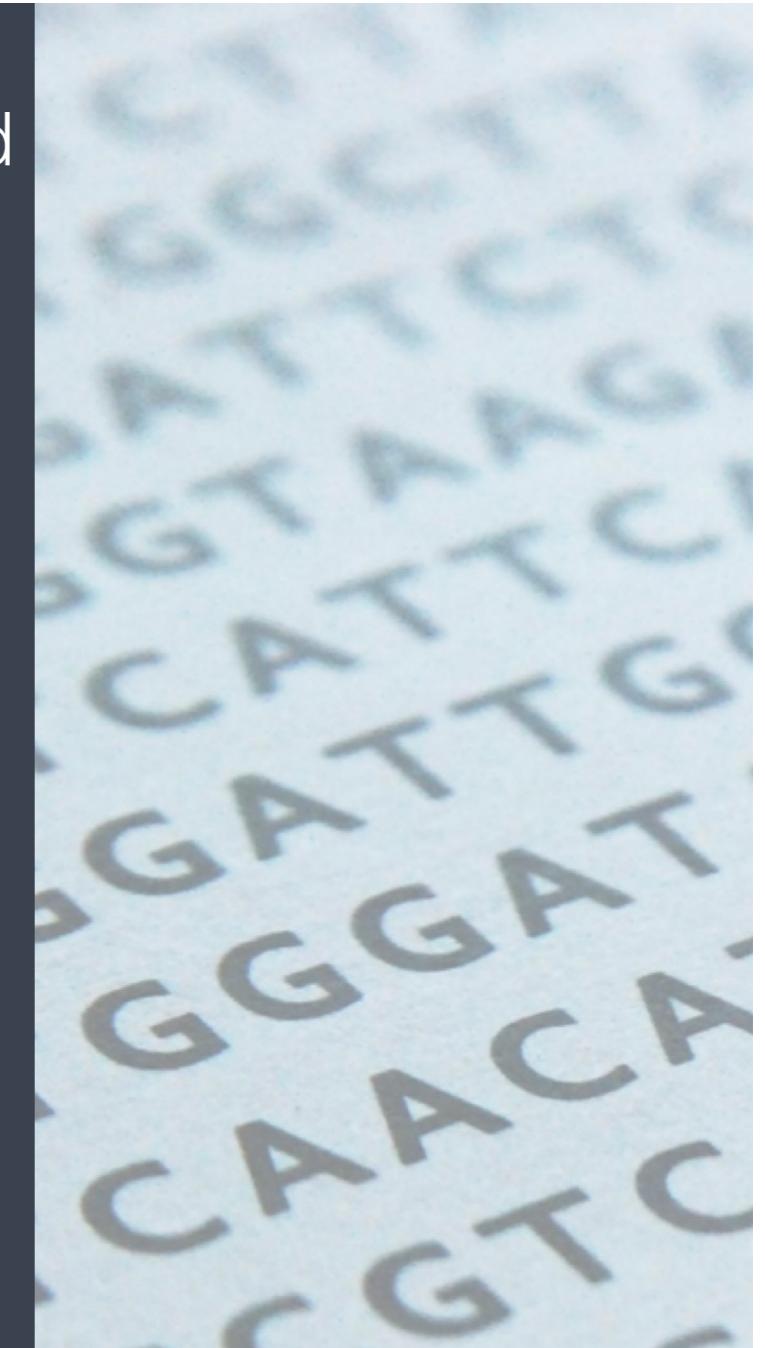
Peter Kraft
Faculty Advisor

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/>





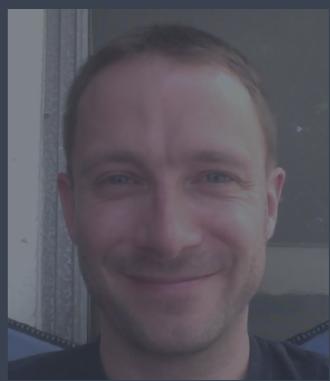
Shannan Ho Sui
Director



John Hutchinson
Associate Director



Victor Barrera



Rory Kirchner



Meeta Mistry



Radhika Khetani
Training Director



Mary Piper



James Billingsley



Zhu Zhuo



Sergey Naumenko



Peter Kraft
Faculty Advisor



HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

NIEHS / CFAR
Bioinformatics
Core

HSCI
HARVARD STEM CELL
INSTITUTE

Center for Stem
Cell
Bioinformatics

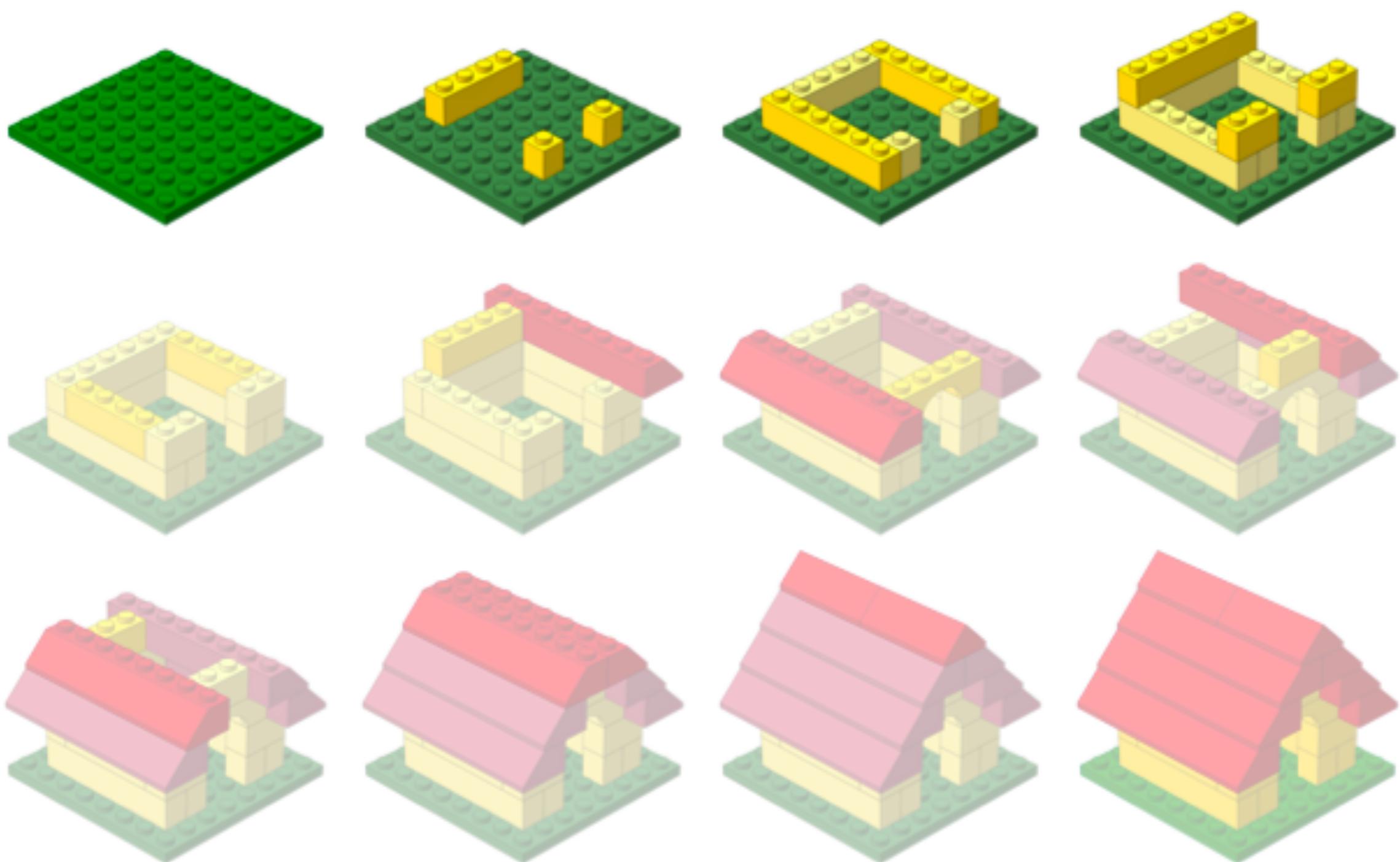
 **HARVARD CATALYST**
THE HARVARD CLINICAL
AND TRANSLATIONAL
SCIENCE CENTER

 **HARVARD**
MEDICAL SCHOOL

Harvard
Catalyst
Bioinformatics
Consulting

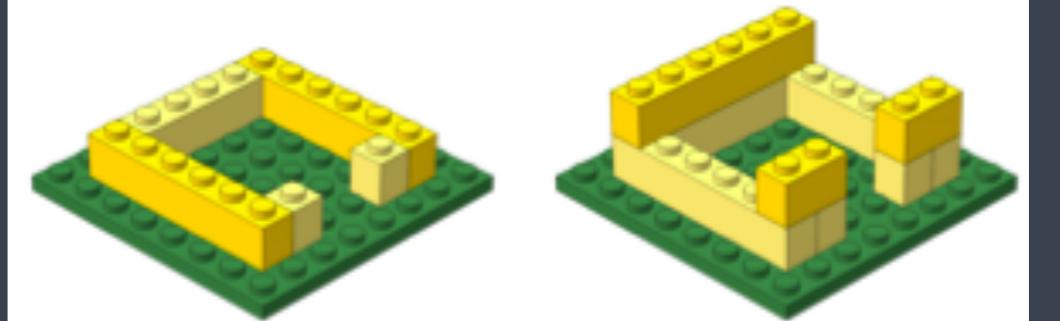
HMS
Tools &
Technology

Workshop scope



Learning R

Learning Objectives



- ✓ Comfortably use RStudio (a graphical interface for R)
- ✓ Fluently interact with R using RStudio
- ✓ Become familiar with R syntax
- ✓ Understand and use data structures in R
- ✓ Install packages and use functions in R
- ✓ Visualize data using *ggplot2*

Visualization

Open source

Data science



Platform
agnostic

Computational
statistics

What is R?

R is an environment for statistical computing and graphics, which brings together a number of features to provide powerful functionality.

- Data handling, wrangling, and storage for big data
- Wide array of statistical methods can be used and added on
- Superior graphical capabilities
- Easy to install on any platform and use (and it's free!)
- Open source with a large and growing community of peers

What is RStudio?

RStudio provides an integrated development environment (IDE) with many features to make using R easier and is a great alternative to working on R in a terminal.

- Graphical user interface, not just a command prompt
- Great learning tool
- Free for academic use
- Program agnostic
- Open source



www.rstudio.com

Logistics

Course webpage (wiki)

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

Course materials online



Introduction to R

[View on GitHub](#)

Approximate time: 70 min

Learning Objectives

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

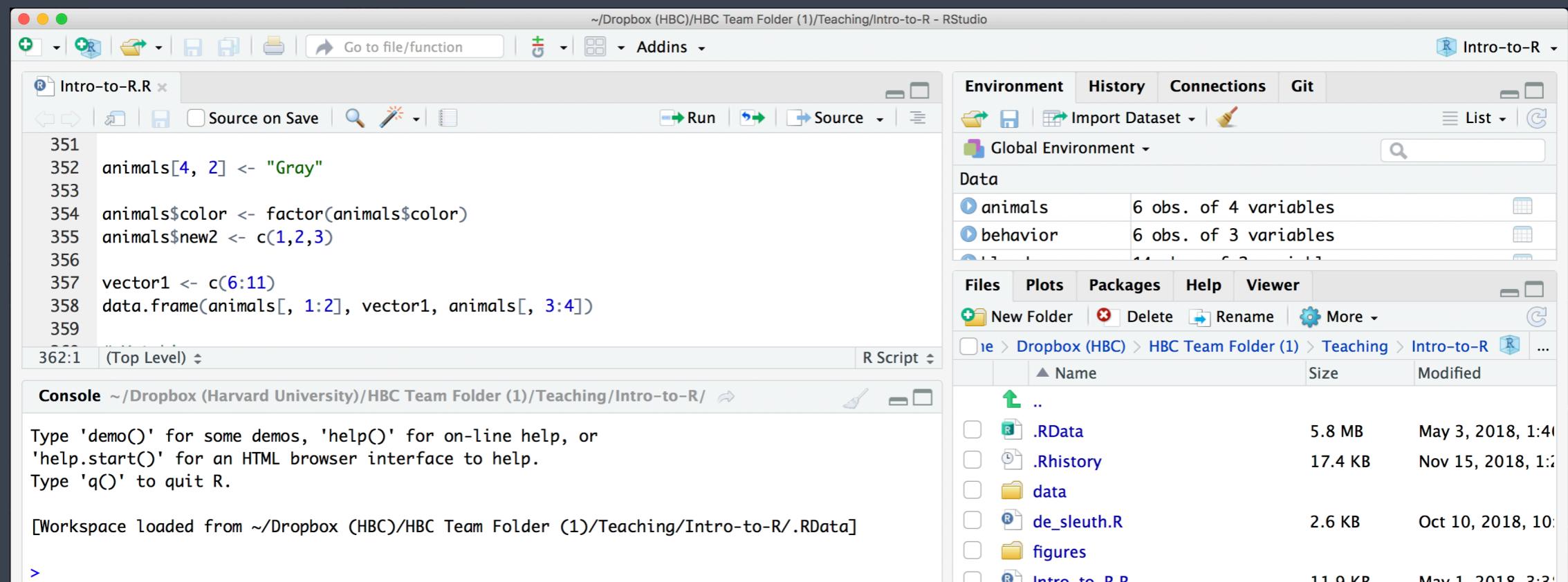
The R syntax

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...



```
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

- ❖ 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)