

# Introduction to R

Introduction to R for Public Health Researchers

# Welcome to class!

1. Introductions
2. Class overview
3. Getting R up and running

# About Us

## **Andrew Jaffe**

Lead Investigator, Lieber Institute for Brain Development

Assistant Professor, Department of Mental Health, JHSPH

PhD in Epidemiology, MHS in Bioinformatics

Email: [ajaffe@jhu.edu](mailto:ajaffe@jhu.edu)

# About Us

## **John Muschelli**

Assistant Scientist, Department of Biostatistics

PhD in Biostatistics, ScM in Biostatistics

Email: [jmusche1@jhu.edu](mailto:jmusche1@jhu.edu)

# What is R?

- ▶ R is a language and environment for statistical computing and graphics
- ▶ R is the open source implementation of the S language, which was developed by Bell laboratories
- ▶ R is both open source and open development

(source: <http://www.r-project.org/>)

# Why R?

- ▶ Powerful and flexible
- ▶ Free (open source)
- ▶ Extensive add-on software (packages)
- ▶ Designed for statistical computing
- ▶ High level language

# Why not R?

- ▶ Fairly steep learning curve
  - ▶ “Programming” oriented
  - ▶ Minimal interface
- ▶ Little centralized support, relies on online community and package developers
- ▶ Annoying to update
- ▶ Slower, and more memory intensive, than the more traditional programming languages (C, Java, Perl, Python)

# Introductions

What do you hope to get out of the class?

Why else to use R?



## Course Website

[http://www.aejaffe.com/winterR\\_2017/](http://www.aejaffe.com/winterR_2017/)

Materials will be uploaded the night before class

# Learning Objectives

- ▶ Reading data into R
- ▶ Recoding and manipulating data
- ▶ Writing R functions and using add-on packages
- ▶ Making exploratory plots
- ▶ Understanding basic programming syntax
- ▶ Performing basic statistical tests

# Course Format

- ▶ 3 modules per class session, each approximately 1 hour
  - ▶ “Interactive” Lecture with RStudio + slides
  - ▶ Lab/Practical experience

# Grading

1. Attendance/Participation: 20%
2. Nightly Homework: 3 × 15%
3. Final “Project”: 35%

# Grading

- ▶ **Homework 1:** Due Tuesday by class
- ▶ **Homework 2:** Due Thursday by class
- ▶ **Homework 3:** Due Friday by class
- ▶ **Project:** Due 2 weeks after class ends

# Installing R

- ▶ Install the latest version from:  
<http://cran.r-project.org/>
- ▶ Install RStudio

## Useful (+Free) Resources

- ▶ Homework will involve working through:  
<http://tryr.codeschool.com/>
- ▶ DataCamp <http://www.datacamp.com>
- ▶ UCLA Institute for Digital Research and Education:  
<http://www.ats.ucla.edu/stat/r/>
- ▶ R reference card: <http://cran.r-project.org/doc/contrib/Short-refcard.pdf>
- ▶ Undergrad Guide to R: <https://sites.google.com/site/undergraduateguidetor/>
- ▶ Quick R: <http://statmethods.net/>
- ▶ Data Camp: <https://www.datacamp.com/>