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**

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)

## Introductions

What do you hope to get out of the class?

Why R?

## Course Website

<http://www.aejaffe.com/summerR_2016/>

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 x 15%
3. Final "Project": 35%

## Grading

* **Homework 1**: Due Tuesday 6/14 by 5pm
* **Homework 2**: Due Thursday 6/15 by class
* **Homework 3**: Due Friday 6/16 by class
* **Project**: Due Friday 7/1 by 5pm

## 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)

## Installing R

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

## R Studio

(Makes R easier)

* Integrated Development Environment (IDE) for R
  + Syntax highlighting, code completion, and smart indentation
  + Execute R code directly from the source editor
  + Easily manage multiple working directories using projects
  + Workspace browser and data viewer
  + Plot history, zooming, and flexible image and PDF export
  + Integrated R help and documentation
  + Searchable command history
* <http://www.rstudio.com/>

## Working with R

* The R Console "interprets" whatever you type
  + Calculator
  + Creating variables
  + Applying functions
* "Analysis" Script + Interactive Exploration
  + Static copy of what you did (reproducability)
  + Try things out interactively, then add to your script
* R revolves around functions
  + Commands that take input, performs computations, and returns results
  + Many come with R, but people write external functions you can download and use

## Useful R Studio Shortcuts

* Ctrl + Enter (Cmd + Enter on OS X) in your script evaluates that line of code
  + It's like copying and pasting the code into the console for it to run.
* Ctrl+1 takes you to the script page
* Ctrl+2 takes you to the console
* <http://www.rstudio.com/ide/docs/using/keyboard_shortcuts>

## 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/>