

Module 1

Introduction to R

Andrew Jaffe
Instructor

Welcome to class!

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

About Me

Investigator at the Lieber Institute for Brain Development

PhD in Epidemiology, MHS in Bioinformatics

Email: andrew.jaffe@libd.org

TAs

Andrew Azman

- PhD Candidate in Epidemiology
- Email: aazman@jhsph.edu

Kate Grabowski

- PhD Candidate in Epidemiology, ScM in Epidemiology
- Email: mgrabows@jhsph.edu

Introductions

What do you hope to get out of the class?

Why R?

Course Website

<http://biostat.jhsph.edu/~ajaffe/rsummer2013.html>

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
- Performing basic statistical tests
- Understanding basic programming syntax

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 Wednesday 6/12 by class Homework 2: Due Thursday 6/13 by class Homework 3:
Due Friday 6/14 by class Project: Due Sunday 6/16 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/>

Note that you must manually update R, often at your own peril...

R Studio

- 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/>

RStudio

File Edit Code View Plots Session Project Build Tools Help

Go to file/function

Project: (None)

Workspace History

Import Dataset

Data

| | |
|------|------------------------|
| cars | 50 obs. of 2 variables |
|------|------------------------|

Files Plots Packages Help

Zoom Export Clear All

```
1 # this is an example of R studio
2
3 data(cars)
4 plot(cars)
```

1:33 (Top Level) R Script

Console ~/Lieber/Classes/WinterInstituteR/Lectures/lecture1/

```
output file:
C:\Users\Andrew\Documents\Lieber\Classes\winterInstituteR\Lectures\lecture1\index.m
[1] "index.html"
> data(cars)
> plot(cars)
>
```

A scatter plot showing the relationship between speed (x-axis) and distance (y-axis) for the 'cars' dataset. The x-axis is labeled 'speed' and ranges from 0 to 25 with major ticks every 5 units. The y-axis is labeled 'dist' and ranges from 0 to 120 with major ticks every 20 units. The plot contains 50 data points represented by open circles, showing a positive correlation between speed and distance.

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 (reproducibility)
 - 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
- `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

- The homework will involve working through: <http://tryr.codeschool.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/>