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: ajaffe@jhu.edu

TAs

John Muschelli

ScM in Biostatistics, 3rd year PhD student in Biostatistics

Email: jmuschel@jhsph.edu

Introductions

What do you hope to get out of the class?

Why R?

Course Website

http://biostat.jhsph.edu/~ajaffe/rsummer2014.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
- 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 Monday 6/30 by 5pm

Homework 2: Due Wednesday 7/2 by class

Homework 3: Due Thursday 7/3 by class

Project: Due Wednesday 7/11 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/

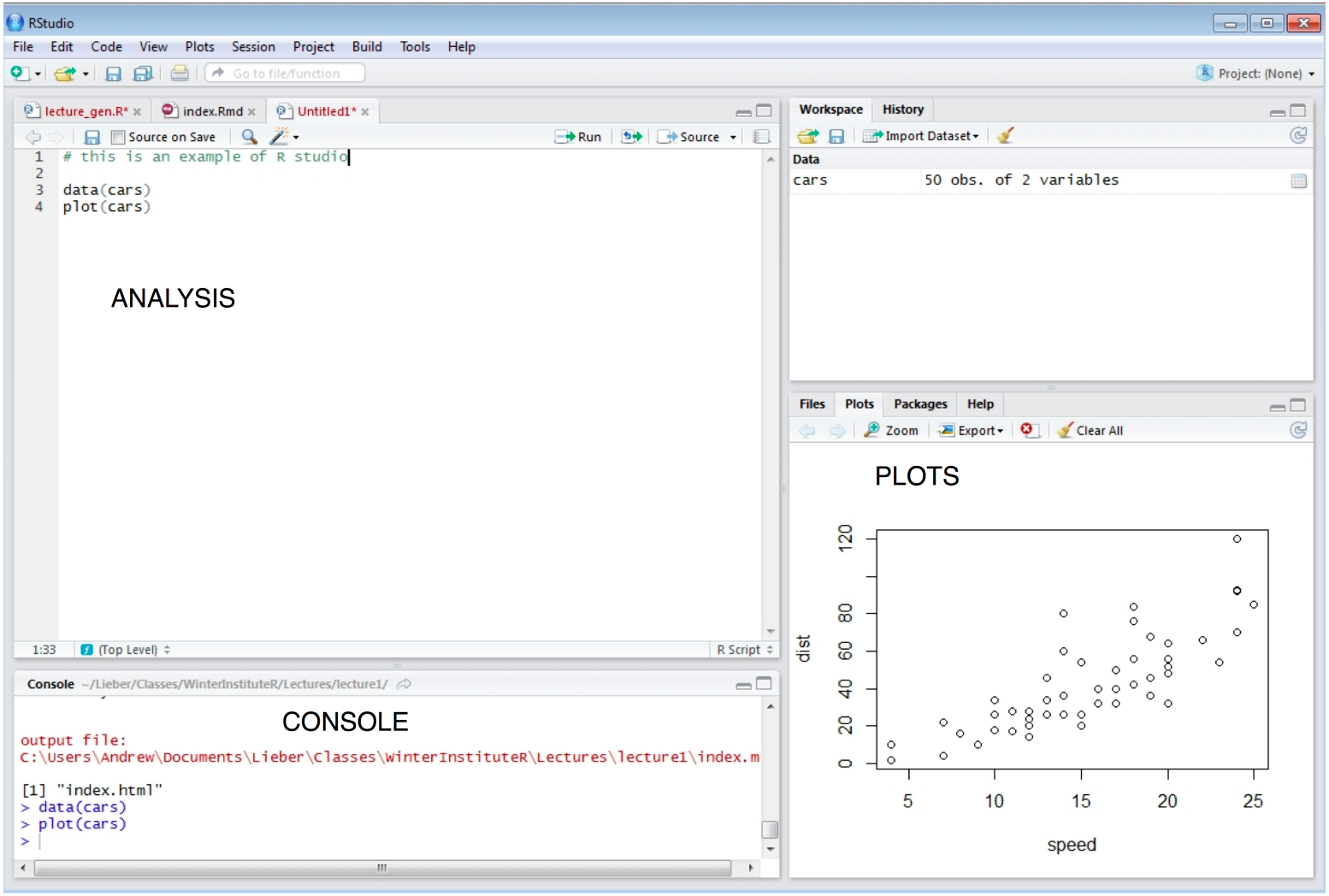
If you have an older version of R, you may not need to update. If you do want to update, re-install and run

```
update.packages(ask=FALSE)
```

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/


RStudio
File Edit Code View Plots Session Project Build Tools Help
Go to file/function
Project: (None)
lecture_gen.R*
index.Rmd
Untitled1*
Source on Save
Run
Source
1 # this is an example of R studio
2
3 data(cars)
4 plot(cars)
ANALYSIS
1:33
(Top Level)
R Script
Workspace
History
Import Dataset
Data
cars
50 obs. of 2 variables
Files
Plots
Packages
Help
Zoom
Export
Clear All
PLOTS
dist
speed
0
20
40
60
80
120
5
10
15
20
25
Console ~/Lieber/Classes/WinterInstituteR/Lectures/lecture1/
CONSOLE
output file:
C:\Users\Andrew\Documents\Lieber\Classes\WinterInstituteR\Lectures\lecture1\index.m
[1] "index.html"
> data(cars)
> plot(cars)
>

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/

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/