Introduction to R for Data Management and Analysis

Session 1

Welcome!

Introductions

- Name
- Program/Concentration
- Familiarity with R
- Expectations
- Fun fact?

Class structure

- Lecture and short exercises: 90 minutes
- Break: 10 minutes
- Exercise: 40 minutes
- Discussion: 10 minutes

Grading

- Attendance: 15%
- Quizzes: 20%
- Exercises: 65%

Course Outline

Date	Topics	Assessment
Tues. June 1	Introductions About this class Features of the R language Getting help and troubleshooting	-
Thurs. June 3	Classes and data types data.frame Import/Export subsets	Exercise

Announcements

- Slack available at https://cunysphcode.slack.com
- Syllabus available on Blackboard
- Zoom link in syllabus

Today's class

- Installation / RStudio Cloud
- What is R, RStudio, Git, GitHub?
- How do I get started?
- Recommendations for RStudio setup
- Features of the R language
- What is a function?
- Getting help within R
- General / R housekeeping tips
- Getting help and troubleshooting
- In-class exercises
- Motivating Examples
- Learning R Trello Board
- Q & A

Installation / RStudio Cloud

- We will use RStudio Cloud first
  - https://www.rstudio.com/products/cloud/
- Install R and RStudio later

What is R?

- R is a programming language and environment for statistical computing and data visualization
- "Base R" refers to the standalone suite of pre-packaged functions that allow R to function as a language
- Extensions of the R language are what are called "packages"
- A **package** is a container of functions that give R additional flexibility

What is RStudio?

- IDE Interactive Development Environment
- Console + Help + Figures + Project Management
- Let's have a look at it!

What is Git / GitHub?

- GitHub is a public repository of user generated code / analyses
- Provides a foundation for reproducible reports
- Versioning is done using software called `git`
- `git` takes care of versioning of all files in a repository (project)

How do I get started?

- First download the latest R version from r-project.org
- Install R with all the default settings
- Download RStudio from RStudio.com
- RStudio allow you to select the R version installed in your system.

Recommendations for RStudio setup

- Tools > Global Options
- Don't restore .RData into workspace
- Never save workspace to .RData on exit

Features of the R Language

- case sensitive!!
- Spaces are ignored (except in names)
- works with functions
- vectorized operations
- objects
- help pages
- `?reserved`

What is a function?

- a series of steps wrapped up into a single command
- a name followed by parentheses `help()`
- arguments (e.g., `functionname(argument1 = "default")`)
- input / output

Getting help within R

Pseudo code	Example code
install.packages(packagename)	install.packages("dplyr")
?functionname	?select
?package::functionname	?dplyr::select
? 'Reserved keyword or symbol' (or backticks)	? '%>%'
??searchforpossiblyexistingfunctionandortopic	??simulate
help(package = "loadedpackage")	help("dplyr")
browseVignettes("packagename")	browseVignettes("dplyr")

General tips

- Learning R will be frustrating
- Learning a language
- Practice promotes familiarity

R Housekeeping tips

- Maintain a clean R "global" environment
- Save your scripts rather than outputs
- Use object names that are descriptive
- Improve readability with clean formatting

Getting help and troubleshooting

- Critically important
- "Debugging" your script
- Step by step, line by line process

In-class Exercises

- R as a calculator exercise

Motivating examples

- http://shiny.rstudio.com/gallery/google-charts.html
- https://shiny.sph.cuny.edu/PowerCalc/

Learning R Trello Board

- Link is also in the syllabus
- https://trello.com/b/7VsveLu5

VIM / Emacs (Extra)

- Tools for 'efficient' typing
- (Optional) Vim game
- Emacs download

Q & A

- Questions?