Introduction to R for Data Management and Analysis

Marcel Ramos, MPH

Session 7

Notes from previous class

- lm and friends
- t-test with a formula vs vector
- object names with a period (.)
- 'apply'ing over a data.frame

Today's outline

- General Tips
- Troubleshooting
- Exploratory functions
- GitHub basics
- Data Analysis Workflow

General R Learning Tips

- Effective troubleshooting is important
  - Get familiar with errors
  - Stack Overflow
  - Helping others on Slack
- Keep inputs and outputs in mind
- Practice, practice, practice
  - Mock examples
  - Answer your own questions in the console
  - Search for answers online
  - Follow online example by running them
- Make use of online resources, help pages
  - RStudio cheatsheets
- Create a catalog...

Catalog

- Create a catalog of functions that are useful for a particular task
- Use your reference websites / textbooks
  - R Cookbook by Winston Chang
  - Quick-R
  - R for Data Science - Hadley Wickham
  - Handbook for Statistical Analyses using R - 3rd Ed.

Troubleshooting

- Getting help within R
  - Documentation pages (examples)
- Getting help outside of R
  - stackoverflow.com
    - Reproducible examples reprex::reprex
  - Quick-R website
- Finding the error in the code
  - Run code line by line to identify location
  - Isolate the error by breaking the process into small pieces

Exploring an object

- head & tail
- class
- length or dim
- dimnames
- attributes
- str
- args on a function

Data Analysis Workflow

- RStudio setup
- Maintain a clean environment
  - No restoring .RData
  - Never save workspace on exit
- Top to bottom RMarkdown
- Import data (usually)
- Clean
  - subset / filter
- Reshape (wide / long)
- summarize / aggregate
- visually explore (EDA)
- graph
  - ggplot2
  - baseR
  - lattice
- Generate the report
  - knitting the document

Example

- https://www.opencasestudies.org/
  - https://www.opencasestudies.org/ocs-healthexpenditure/