Introduction to R and RStudio

Anthony Chau

UCI Center for Statistical Consulting

2021/01/12 (updated: 2021-02-09)

Pre-workshop tasks:

1. Download R version 4.0.3
2. Download RStudio version >= 1.3.1
3. View and download workshop materials at Google Drive

Post-workshop

1. Post-workshop survey will be sent out
2. Recordings will be made available shortly after

Outline

Time	Topic
11:00am - 11:15am	Overview of R & RStudio
11:15am - 11:30am	Objects and functions
11:30am - 12:00pm	Data types and data structures
12:00pm - 12:05pm	Break
12:05pm - 12:30pm	Data Management
12:30pm - 12:55pm	Dataset demo (with plotting)
12:55pm - 1:00pm	Questions

Learning Objectives

1. **Programming**: learn R programming syntax to interact with your data in complex and useful ways

2. **Data Management**: learn how to represent and manipulate data using R data structures

3. **Exploratory Data Analysis**: learn how to run descriptive statistics and generate visualizations of your data

4. **Reproducible research**: learn how to use R and RStudio to manage your projects and analyses in a reproducible manner

Slide structure

- Slides will be a mix of text, images, and code
- For code examples, look for a light-yellow box
- R comments are prefixed by `#`
  - Comments are useful notes about the R code
  - Comments are not R code
- R code is not prefixed by anything
- R output is prefixed by `#>`
  - Output is displayed in the console pane

```
# This is a comment
sum(1, 2, 3, 4, 5)
#> [1] 15
```

Workshop structure

1. Lecture
2. Short demos in RStudio (when applicable)
3. Checkpoint Questions
    - Quick review questions after a topic/section
    - Volunteers to answer and answer question
4. Live coding (last ~ 30 minutes)

Overview of R

What is R?

R is a programming language with strengths in data management, data visualization, and statistics.

Why R?

- Free
- Welcoming community (IMO)
  - Lots of R bloggers, twitter presence, community groups (RLadies, local groups)
- Open source (non-proprietary)
- Active community development to solve problems
  - If you are trying to solve a problem, someone else has probably written a R solution that you could use
- Flexible and extensible
  - Dashboards
  - Automate periodic reports
  - Develop novel statistical tools

Why not R?

- Other software already suits your needs.
- Legacy systems:
  - Highly common to have years or decades of code and programs written in other languages.
  - Cheaper to maintain existing systems than to invest the resources and human hours to convert to R.
- High investment of time and energy to learn and be productive in R
- You collaborate with others who do not use R

Overview of RStudio

What is RStudio?

RStudio is a software tool for R that provides utilities for writing R code, data analysis, and project management

Some features

- Code completion
- Syntax highlighting
- Integrated help and documentation
- Develop and test interactive graphics
- Author reports, books, slides

Navigating RStudio

- Rstudio can be divided into different panes. Below are the most important panes.

Console

- Area to write R code
- Meant for interactive use and quick checks

Script

- Another area to write R code
- Similar to a Word Document but for code
- Analogous to Stata do files, SPSS and SAS syntax files

Navigating RStudio

Environment

- Area to view and inspect all your objects and data

Files

- Area to navigate your files
- Can open up R and other supported files directly

Navigating RStudio

Packages

- Lists all downloaded packages
- Able to download and update packages
- Packages are user-written R code to solve a specific problem
- Packages are analogous to Stata user-written functions, SAS and SPSS macros

Help

- Contains documentation for anything R-related
- Get help by typing in script or console: `help("topic")` or `?topic`