

Lecture 2: Describing Data

STAT 515 — Statistical Methods I

Dr. Yen-Yi Ho

Department of Statistics, University of South Carolina

What is R?

- **Statistical Language:** Free, open-source programming language specialized in data analysis and graphics.
- **Ecosystem:** Backed by a vast community and thousands of specialized packages via CRAN.
- **Core Use Cases:**
 - Data cleaning and manipulation
 - Statistical modeling and testing
 - High-quality data visualization ('ggplot2')

What is RStudio (Posit)?

- **IDE:** An Integrated Development Environment designed specifically for R.
- **Dependency:** R must be installed on your computer before RStudio can work.
- **Why use it?:** Provides syntax highlighting, project management, workspace inspection, and integrated help menus in a single screen.

The 4 Panes of RStudio

- **Top-Left (Source/Editor):** Where you write, edit, and save R scripts.
- **Bottom-Left (Console):** Where code is executed and results print out.
- **Top-Right (Environment/History):** Tracks active variables, data frames, and loaded objects.
- **Bottom-Right (Files/Plots/Help):** File manager, graphical plot viewer, package status, and documentation.

First Steps in R

- Using R as a calculator:

Code

```
> 5 + 3  
> sqrt(16)
```

- Object Assignment ('<-'):

Code

```
> x <- c(1, 2, 3, 4, 5)  
> mean(x)
```

Summary

- **R** does the heavy-duty statistical computation.
- **RStudio** makes writing and organizing R code easier.
- Both are essential, free tools for modern data science.

Questions?