Introduction to R for Data Analysis

Michael Malick

Simon Fraser University

mmalick@sfu.ca

Course Info

- Format: Two-day course
- Website: www.michaelmalick.com/R

Course Overview

The objective of this two-day course is to provide participants with a foundational knowledge of the R programming language and its applications for data analysis. The course uses lecture, examples, and problem sets to help participants learn the necessary skills and concepts to use the R language to analyze independent datasets.

Course Outline

1. R Basics
    - Overview of R
    - Working directory
    - Workspace
    - History
    - Help
    - Packages
    - Scripts
    - Basic math

- Functions
- Errors and warnings

2. Data Structures and Subsetting
    - Vectors and classes
    - Matrices
    - Dataframes and factors
    - Lists
3. Reading and Writing Data
    - File formats
    - Writing data: `write.table()`
    - Reading data: `read.table()`
4. Data Manipulation
    - Order vectors and dataframes
    - Add and remove columns in dataframes
    - Missing values
    - Combining datasets: `rbind()`
    - Merging datasets: `merge()`
    - Reshaping datasets: reshape2 package
5. Base Graphics
    - Graphics overview
    - Plotting functions
    - Graphics devices
    - Graphical parameters: `par()`
    - Low level plotting commands
6. Lattice Graphics
    - Grid graphics overview

- Formulas and conditioning
- Groups, scales, legends
- Panel functions

7. Data Summaries: `apply` family of functions
   - Why `apply`?
   - Summarize by row and columns: `apply()`
   - Summarize by groups: `tapply()` and `by()`
8. Intro to Statistical Models
   - Overview of linear models
   - Fitting a linear regression model
   - Model diagnostics
   - Model comparisons: `AIC()`
9. Automation: `for` loops
   - Why `for` loops?
   - `for` loop syntax
   - Data and graphics examples
10. Automation: writing functions
    - Why write a function?
    - Function syntax
    - Errors and warnings
    - Debugging: `browser()`
    - Pass through: `...`