

R By Example

Dave Armstrong
Associate Professor

University of Western Ontario
Department of Political Science
Department of Statistics and Actuarial Sciences (by courtesy)

e: dave@quantoid.net
w: www.quantoid.net/teachicpsr/rbyexample

1 Overview and Course Objectives

This course is designed to get you over the basic hurdles you will face when beginning to learn **R**. It will cover some of the basic tasks that you face as quantitative researchers and will put you in a position to extend your knowledge of **R** in whatever direction is required to meet your own needs. The course takes an example-based approach, so for much of the workshop, we discuss applications and use cases that would appear in the “real world”. Over the course of the workshop, you will be encouraged to work through examples of common uses of R in applied research. For many of you, R will be a new language and like learning any new language, practice makes perfect.

The goal of the course is not to make you R experts, no short course could do that. However, what we can do is help you understand how R works, what sorts of things it does well and how R might fit into your research work flow. Further, the hope is that participants in the course will become sufficiently proficient with R to use it in their own work during and after the course.

This course is also not technically a “statistics” course. That said, we will certainly be covering plenty of statistical concepts and I am happy to try venture in to some of the fundamentals of those techniques as time and suitability permit.

We will meet from 9AM to 5PM each day with one hour for lunch. Each day will be broken into two 3.5 hour sessions, each with a 15-minute break. Within each 3.5 hour session, there will be approximately 2.5 hours of instruction and then around an hour to try out what you’ve learned on your own data. We can adjust the relative balance of these two items as required.

1.1 Computing

The workshop will be held in a computer lab and all of the necessary technology will be supplied to all participants. However, I would encourage people to bring their own computers so you get a chance to work with R in a computing context that is familiar. Before the course, you will be given instructions on exactly which software you need for the course.

1.2 Course Materials

Each day, you will be given a printed handout that will work through that day's material. The course material will also be available electronically, along with code and data on the course's website and will be available to course participants.

2 Outline

Below is a brief outline of what I intend to cover. I will try as much as possible to get through all the material. In the event that all material cannot be covered in the allotted time, I will make every effort to cover the material of most interest to course participants. Further, if desired material is not on the syllabus we can potentially amend it to include the desired material.

1. Introduction
2. The Basics
 - (a) Getting R
 - (b) Using R
 - (c) Assigning Output to Objects
 - (d) Vectors and Matrices
 - (e) Object Orientation
 - (f) Reading in Data: SPSS, Stata, .csv, text
 - (g) Functions, Syntax and Arguments
 - (h) Merging Datasets
 - (i) Writing/Saving data and objects
3. Dealing with Data
 - (a) Examining Data: Summaries and Descriptives
 - (b) Data Types
 - (c) Recoding and Adding New Variables
 - (d) Missing Data
 - (e) Filtering and Sorting

4. Statistics

- (a) Cross-tabs and Categorical Measures of Association
- (b) Measures of Association
- (c) Linear Models
 - i. Adjusting the Base Category in Categorical Variables
 - ii. Model Diagnostics
 - iii. Linear Hypothesis Tests
 - iv. Factors and Interactions
 - v. Non-linearity: Transformations and Polynomials
 - vi. Model Fit and testing
- (d) Generalized Linear Models and the Like
 - i. Binary DV Models
 - ii. Other GLMs
 - iii. Ordinal DV Models
 - iv. Multinomial DV Models
- (e) Multi-level Models
 - i. Random intercepts models
 - ii. Random coefficients models
 - iii. Multiple levels of nesting

5. Help, Troubleshooting, Finding Functions

- (a) Finding Packages on CRAN
- (b) Troubleshooting: Warnings and Errors
- (c) Help: Books and Web

6. Primer on Good Graphics

- (a) Graphical Perception
- (b) Advice

7. Different Graphical Packages in R.

8. Base Graphics.

- (a) Default Plotting Methods
- (b) Fine Control Over Your Graphs
- (c) Example: Building a Scatterplot.
- (d) Interactivity in Base Graphics

9. ggplots

- (a) Common Graphs in ggplot: Scatterplots, Bar Graphs, Histograms, etc...
 - (b) From Deducer (an R GUI) to Code
 - (c) Faceting and Grouped Data
10. Effects Package: Making Effects Plots for Interactions and Non-Linear Models
- (a) Transformations and Interactions in Linear Models
 - (b) Predicted Probabilities in GLMs and other Choice Models
11. Maps
12. Interactivity
- (a) iplots
 - (b) d3
13. Repeated Calculations
- (a) `*apply`
 - (b) `aggregate`
 - (c) `by`
 - (d) Loops
 - i. Example: Permutation Test of Significance for Φ Coefficient.
 - (e) If-then Statements
14. Basic Function Writing
- (a) Calculating a Mean
 - (b) Summary with Robust Standard Errors
 - (c) Changing Existing Function Defaults
 - (d) `.Rprofile` and the `.First` and `.Last` functions.
15. Clustered Standard Errors and Grouped Data
16. Reproducible Research