Introduction to R: 
How to Use R for Data Management, Data Analysis, and Graphical Display

2016 ICPSR Summer Program
August 15-19, 2016
Ann Arbor, MI

General Information
Location: Angell Hall Computing Classroom “C”
Meeting Time: 5:00 PM-8:00 PM

We will begin promptly at 5:00 PM each day. Please arrive and be ready to go (i.e. have new datasets and materials downloaded). We will have one ten-minute break during each session.

Instructor
Joseph Waddington, Ph.D.
Assistant Professor
Department of Educational Policy Studies and Evaluation (EPE)
College of Education
University of Kentucky

E-mail: rjwaddington@uky.edu

Contact Information & Communication
For any questions, please be sure to e-mail me and I will address your question as soon as possible. I will primarily use e-mail to communicate with you and encourage you to do the same with me and others in our class. Please check your e-mail regularly for announcements, information, and updates. I expect you to be up-to-date on all workshop communications. If you have questions or if something is unclear, please do not hesitate to ask. For file storage, we will be using Google Drive. I will upload and share data files, electronic versions of handouts, R scripts, and any other workshop materials.
Course Description & Objectives

The “R” statistical software package has become widely used to perform data management, conduct statistical analyses, and produce graphical displays of data across the social, behavioral, health, and other sciences. R is an open-source, code-based program that combines the ability to conduct analyses with a convenient facility for programming. R’s comprehensive network (CRAN), provides thousands of “add-on” packages for use with advanced quantitative analyses.

This workshop will introduce users to the R programming environment and its use as a data analysis package. Participants in the workshop will learn to use R for data management; conducting and interpreting descriptive analyses, basic hypothesis tests, and regression analyses; producing graphical displays and visual presentations of data; and other advanced topics as time permits. This workshop is ideal for staff, researchers, analysts, graduate students, and faculty who are seeking a brief and applied introduction to using R for data analysis in their own work, research projects, or course instruction.

The workshop will feature both a lecture and a lab component. During lecture, the instructor will demonstrate basic features and coding in R to manage data, conduct analyses, and produce graphical displays. During lab, the instructor will lead participants through guided examples with real social science data on topics and techniques that mirror the same ones covered in the day’s lecture. All data will be provided by the instructor. At the end of the workshop, we anticipate that participants will have a basic understanding of how to use R to conduct quantitative analyses, as well as prepared to use R’s add-on packages for further advanced techniques.

Workshop Schedule

Day 1 (Mon. Aug. 15): Data Management and Descriptive Analysis in R
Topics: downloading and using add-on packages; reading data into R; variable management using the “dplyr” package (viewing, labeling, recoding, computing new variables); univariate descriptive analysis (means, standard deviations, frequencies, measures of position); graphics (histogram)

Day 2 (Tue. Aug. 16): More Descriptive Analysis and Hypothesis Testing in R
Topics: bivariate descriptive analysis (cross-tabulations, correlations); producing scatterplots and other two-variable displays; basic hypothesis testing (z- and t-tests, one-way ANOVA, chi-square).

Day 4 (Wed. Aug. 17): Regression Analysis in R
Topics: computing and interpreting output of bivariate and multiple regression analyses; interaction terms. Reception with Dinner Provided @ 6:30 PM.

Day 4 (Thu. Aug. 18): More Regression Analysis and Graphical Displays in R
Topics: additional descriptive graphics; regression diagnostics and displays; other displays using the “ggplot2” package; weighting

Day 5 (Fri. Aug. 19): Advanced Data Analysis in R
Topics: logistic and multilevel regression models; post-estimation calculations; exporting results. Post-Workshop Celebration @ 7:30 PM.