The R Statistical Computing Environment
Basics and Beyond
Getting Started with R

John Fox

McMaster University

ICPSR/Berkeley 2016

Outline

To be covered as time permits:

- Getting started with R, RStudio, and R Markdown
- Linear and generalized linear models in R
- Mixed-effects models with the **nlme**, **lme4**, and **car** packages
- Survival analysis with the **survival** package
- Structural-equation modeling with the **sem** package
- Data and data management in R
- R programming
- R graphics
- Building R packages

Getting Started With R
What is R?

- A statistical programming language and computing environment, implementing the S language.
- Two implementations of S:
  - S-PLUS: commercial, for Windows and (some) Unix/Linux, eclipsed by R.
  - R: free, open-source, for Windows, Mac OS X, and (most) Unix/Linux.

Getting Started With R
What is R?

- How does a statistical programming environment differ from a statistical package (such as SPSS)?
  - A package is oriented toward combining instructions and rectangular datasets to produce (voluminous) printouts and graphs. Routine, standard data analysis is easy; innovation or nonstandard analysis is hard or impossible.
  - A programming environment is oriented toward transforming one data structure into another. Programming environments such as R are *extensible*. Standard data analysis is easy, but so are innovation and nonstandard analysis.

Getting Started With R
Why Use R?

- Among statisticians, R has become the de-facto standard language for creating statistical software. Consequently, new statistical methods are often first implemented in R.
- There is a great deal of built-in statistical functionality in R, and many (literally thousands of) add-on packages available that extend the basic functionality.
- R creates fine statistical graphs with relatively little effort.
- The R language is very well designed and finely tuned for writing statistical applications.
- (Much) R software is of very high quality.
- R is easy to use (for a programming language).
- R is free (in both of senses: costless and distributed under the Free Software Foundation's GPL).
- The RStudio IDE makes it easy to use R for programming, package development and, routine data analysis, including writing reports with embedded R code, for reproducible research.

Getting Started With R
This Workshop

- The purpose of this workshop is to get participants started using R.
- The statistical content is largely assumed known.
- Much of the workshop is based on J. Fox and S. Weisberg, *An R Companion to Applied Regression, Second Edition*, Sage (2011), and on-line appendices.
- More advanced participants may prefer to read, or want to read in addition, W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S, Fourth Edition*. New York: Springer, 2002
- Additional materials and links are available on the web site for the book: *http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/index.html* or *tinyurl.com/carbook*
- The book is associated with R packages (called **car**) and **effects** that implement a variety of methods helpful for analyzing data with linear and generalized linear models and for visualizing fitted statistical models.

Getting Started With R
This Workshop

- Other references are given on the lecture series web site.
- Workshop web site: *http://socserv.mcmaster.ca/jfox/Courses/R-course-Berkeley/index.html* or *tinyurl.com/Berkeley-R-course*