Introduction to R

Brad Jones
Department of Political Science

University of California, Davis
What is R?

- A programming environment for data analysis and graphics.
- R is an open-source, freely-available statistical package.
- R implements a dialect of the S language that is the basis for the commercial S-PLUS system.
Where can you get R

- Comprehensive R Archive Network (http://cran.r-project.org)
The R environment

- CLI (command line interface) as opposed to a GUI (graphical user interface).

- A GUI typically allows a user to access many functions through menus and dialog boxes.

- The program looks “old school” but it is not. It’s simplicity is a virtue.
Advantages of R

- A strong emphasis on graphics and visualizing data.
- Has a wide range of standard statistical applications, augmented by freely available add-on libraries.
- The programming language is easy to use and finely tuned to the development of statistical applications.
- Object-oriented design features.
- Easy to write functions and combine related sets of functions and data in R packages.
- Strong connection between theoretical model specifications and R language implementations.
Disadvantages of R

- Probably not the best for data management related tasks.
- STATA, SPSS, Excel etc. are fairly easy to use, with many models and options already implemented.
- However, we can import a variety of data sources into the R environment.
Illustration to Follow

- How to read data from a file into an R data frame.
- How to draw some simple graphs.
- Understand differences between Mac vs. PC platforms [they are minimal]
- How to perform least squares regression.
- How to check the adequacy of the preliminary regression model using a variety of diagnostic methods.
This example is from John Fox’s book *An R and S-Plus Companion to Applied Regression*. The data was originally used by Duncan (1961) and deals with occupation prestige. I have some code that shows how to measure the effect of income and education on prestige.
There are 100s contributed packages available on the CRAN web site.

In this class you may need to access these packages.

Which are freely available!

The only way you can learn is by doing however.

Code given by Verzani should work seamlessly. Try it out!
R Tips

- Create a dedicated directory on your computer for this class
- Write out your code AND SAVE IT in this directory for future use!
- In other words, don’t try and reinvent the wheel each time.
- Do not complain about how hard it is. It is not hard, it is just new.
- Do not wait until the last minute to complete assignments that require use of R
R References (in addition to Verzani)

- *Data Analysis and Graphics Using R*  
  by John Maindonald and John Braun

- *An R and S-PLUS Companion to Applied Regression*  
  by John Fox

- *An Introduction to R*  
  by W.N. Venables, D.M. Smith and the R Development Core Team

- *Introductory Statistics with R*  
  by Peter Dalgaard

- Lots of documentation on the CRAN web site.