WORKING WITH DATA

An Introduction to R

15.071x – The Analytics Edge

What is R?

- A software **environment** for data analysis, statistical computing, and graphics
- A programming **language**
  - Natural to use, complete data analyses in just a few lines

History of R

- Originated from S
  - A statistical programming language developed by John Chambers at Bell Labs in the 1970s
- The first version of R was developed by Robert Gentleman and Ross Ihaka at the University of Auckland in the mid-1990s
  - Wanted a better statistical software in their Macintosh teaching laboratory
  - An open-source alternative

Why Use R?

- There are many choices for data analysis software
  - SAS, Stata, SPSS, Excel, MATLAB, Minitab, pandas
  - So why are we using R?
- Free (open-source project)
- Widely used
  - More than 2 million users around the world
  - New features are being developed all the time
  - A lot of community resources
- Easy to re-run previous work and make adjustments
- Nice graphics and visualizations

Using R

- We will just use the R command line interface
- If you want to try a graphical user interface (GUI), here are some popular choices:
  - RStudio (http://www.rstudio.com)
  - Rattle (http://rattle.togaware.com)

R Resources

- Official page: http://www.r-project.org
- Download page: http://www.cran.r-project.org
- Some helpful websites:
  - http://www.statmethods.net
  - www.rseek.org
  - http://www.ats.ucla.edu/stat/r/
  - http://finzi.psych.upenn.edu/search.html
- Looking for a command? Google it
- Best way to learn R is through trial and error