

## 22S:166 Computing in Statistics

### Introduction to R

Lecture 5  
September 10, 2007

Kate Cowles  
374 SH, 335-0727  
kcowles@stat.uiowa.edu

### What R is

- “an integrated suite of software facilities for data manipulation, calculation, and graphics display” (*An Introduction to R*, Venables, Ripley, and the R Core team)
  - data handling and storage capabilities
  - operators for calculations on arrays and matrices
  - data analysis tools
  - graphical capabilities
  - programming language
  - planned and coherent system

- an implementation of S language
  - S language was developed at AT&T-Bell Labs
    - \* first version 1976
  - S-Plus is a commercial version of S (begin in 1987)
    - \* sold and supported by Insightful Corp.
    - \* GUI
    - \* many formats supported for graphics export and data input/output
    - \* runs on Windows, UNIX, Linux (not Macintosh)

- advantages of S
  - extendible
    - \* users write new functions in S language
      - just as developers do
    - \* excellent documentation for adding functions to system
    - \* users can create their own data types
    - \* huge international community of users constantly contribute new capabilities
    - \* contrast with SAS
      - very hard to write new SAS procedures
      - users write in different language (SAS macro or IML) than developers
  - high-level language
    - \* only a few commands required to do complex things

- language is connected to data while executing  
example (from *Statistical Computing and Graphics* course notes by Frank Harrell)

```
if(is.factor(x) | is.character(x) |
   (is.numeric(x) & length(unique(x)) < 20))
  table(x) else quantile(x)
```

computes quantiles of  $\mathbf{x}$  if  $\mathbf{x}$  is numeric and has at least 20 distinct values, requeryency table otherwise

- object-oriented
  - \* fewer commands to learn because the same command can be applied to different types of objects
- Harrell: “best scientific graphics available”
  - \* Harrell: “SAS graphics are ugly, inflexible, have poor defaults, difficult to program”

## Starting and running R interactively on Linux

- recommendation: use a separate subdirectory for each major project you do with R
- in a terminal window, get into the desired subdirectory and start R by entering

R

- R commands may be issued interactively
- to quit

q()

- follow prompts as to whether you want to save *workspace*
- if you don’t save it, any new objects (data, functions, results) created during the current R session will be lost

## R

- international team of statisticians started developing R in early 1990’s
  - to provide open source alternative to S-Plus
  - to provide S implementation on Linux (not supported by S-Plus then)
- easy to download and install from web sites
- excellent documentation
- user-contributed libraries called packages expand capabilities
- runs on Windows, UNIX, Linux, Macintosh
- no GUI on most platforms
- fewer data import/export capabilities than S-Plus
  - although add-on packages provide more
  - no export specifically to Powerpoint

- Part B of this lecture is from Maindonald, J.H. “Using R for Data Analysis and Graphics: Introduction, Code and Commentary,” available as contributed documentation on [www.cran.r-project.org](http://www.cran.r-project.org)