

All about R - by Example

Mark S. Handcock

CSSS 506

Based on notes by Julian Faraway, University of Michigan

What is R?

- A language and environment for statistical computing and graphics.
- Based on S language created by Chambers et al. at Bell Labs
- A “GNU S” GNU - a recursive acronym for “GNU’s Not Unix”
- Free and open-sourced software distributed under a GNU-style copyleft.

Features

- Data handling and storage facility
- Branching and looping.
- Suite of operators for calculations on arrays, in particular matrices
- Large, coherent, integrated collection of intermediate tools for data analysis
- Graphical facilities for data analysis and display either on-screen or on hardcopy
- Modular programming using functions.
- Most functions in R are written in R.
- A fully developed “functional” programming language
- C, C++ and Fortran code can be linked and called at run time
- Wide variety of user contributed packages

Brief R session

```
> data()  
> data(stackloss)  
> help(stackloss)  
> dim(stackloss)  
> plot(stackloss)  
> summary(stackloss)  
> # This is a comment line  
> # Let's do a regression ...  
> #  
> stacklm <- lm(stack.loss ~ ~ Air.Flow ,data=stackloss)  
> summary(stacklm)  
> plot(stacklm)  
> X <- as.matrix(cbind(1,stackloss[1:3,]))  
> solve(t(X) %*% X,t(X) %*% stackloss$stack.loss)  
> stacklm$coef  
> library()
```

How to get it

- Main site at www.r-project.org
- Windows, Macintosh and Unix versions.
- Windows download:
 - Basic package, documentation and several popular packages (15MB)
 - Just the executables (2MB or 2 floppies)
- Regular updates are nice but requires maintenance.

Books

- Becker, Chambers and Wilks *The New S Language* - “The Blue Book” (Sv2)
- Chambers and Hastie *Statistical Models in S* - “The White Book” (Sv3)
- Chambers *Programming with Data* - “The Green Book” (Sv4)
- Venables and Ripley *Modern Applied Statistics with S-PLUS (3rd Ed)*
Recommended
- Venables and Ripley *S Programming* **Not for Beginners**
- Any book on S-PLUS will be usable.

Free Documentation

- *An introduction to R* - **The Official Guide**
- *Using R for Data Analysis and Graphics* by John Maindonald - **Contributed Guide**
- Online help pages
- See the “Links to Resources” on the CSSS 506 homepage

S-PLUS

- Original S language developed by Chambers et al at Bell Labs in 1984.
- S-PLUS is based on S - company formed in late 80's now owned by Insightful has exclusive license for S code.
 - Corporate Focus - Finance, Pharma, etc
 - Highly developed graphical user interface
 - Web and Server based services
 - Educational licensing conditions vary
 - Now at Version 6

R compared with S-PLUS

- Statistical capabilities now comparable - more new developments in R
- Code largely portable between R and S-PLUS
 - R stores data differently
 - More esoteric code differences e.g. scoping rules
- Internal coding entirely different - speed differences
- Support
 - R has an active mailing list (best for intermediate/advanced users)
 - S-PLUS also has a mailing list but also has more formal and reliable support desk

History

- Developed by **R**oss Ihaka and **R**obert Gentleman in NZ
- Based on commands described in the Blue, White and Green S books.
- Not a clone of S-PLUS - has extra functions.
- Reliable and comprehensive for the last 2-3 years - v1.0 released February 2000.
- Current version is 1.4.0
- Now supported by a volunteer team of developers together with many contributors

Teaching with R

- Not menu-driven or GUI-based - requires more effort to get started
- Enables, even requires, students to implement maths
- Free software, all platforms (and docs) makes life easy
- How to learn R
 - See a brief introductory lecture with handout (this lecture!)
 - Work through examples, answer questions needing only those commands
 - You are encouraged to work in groups
- The coming lectures on R
 - I will provide handout with (most) commands and output
 - You should annotate handouts
 - Suggest alternative analyses, etc.

Using R at UW

- On CSSCR machines ...
- Increasingly on C&C machines ...
- Install on your own machine e.g., PC, Mac or LINUX.

Thesis Research with R

- Customized Graphics - where do most plots in Stats journals come from?
- Rapid prototyping and experimentation with new procedures
- Used to be too slow for simulations - but now R is better and CPUs faster
- Great way to make research available for practical use