The statistical programming language and computing environment S has become the de-facto standard among statisticians and has made substantial inroads in the social sciences. The S language has two implementations: the commercial product S-PLUS, and the free, open-source R. Both are available for Windows and Unix/Linux systems; R, in addition, runs on Macintoshes. This lecture series and associated workshops will use R.

A statistical package, such as SPSS or SAS, is primarily oriented toward combining instructions with rectangular case-by-variable datasets to produce (often voluminous) printouts. Such packages make routine data analysis relatively easy, but they make it relatively difficult to do things that are innovative or non-standard, or to add to the built-in capabilities of the package. In contrast, a good statistical computing environment also makes routine data analysis easy, but it additionally supports convenient programming; this means that users can extend the already impressive facilities of S. Statisticians and others have taken advantage of the extensibility of S to contribute more than 1000 freely available “packages” of R programs. As well, S is especially capable in the area of statistical graphics, reflecting its origin at Bell Labs, a centre of graphical innovation.

The first two (lecture) sessions are meant to provide a basic overview of and introduction to R, including to statistical modeling in R – in effect, using R as a statistical package. The following four to five workshop sessions pick up where the basic lectures leave off, and combine lecture material with hands-on experience. The workshop sessions are intended to provide the background required to use R seriously for data analysis and presentation, including an introduction to S programming and to the design of custom statistical graphs, unlocking the power in the R statistical programming environment. The topics for the workshops in session 3-7 are somewhat flexible, depending upon participants’ interests: the topics given here are suggestions.

An outline of the classes follows (with chapter references to Fox, 2002):

1. Getting started with S (Ch. 1)
2. Statistical models in S (Ch. 4, 5, & appendices)
3. Data in S; using R packages (Ch. 2 & 3)
4-5. Programming in S (Ch. 8)
6. S graphics (Ch. 7)
7 (interest permitting). Building R packages
CD/ROM

I’ll distribute a CD/ROM with the Windows version of R, including all of the contributed packages on the Comprehensive R Archive Network (CRAN) web site, along with other R-related resources.

Course Web Site

Materials for the course will be deposited at

BIBLIOGRAPHY

This has been an explosion of sources on S (particularly on R, with many titles of the form “X with R”). The following list is by no means complete.

Principal Text

The principal text for this short course is J. Fox, *An R and S-PLUS Companion to Applied Regression*, Sage, 2002. The book is now slightly dated, but still suitable as an introduction. Additional materials are available on the web site for the book, <http://socserv.mcmaster.ca/jfox/Books/Companion/index.html>, including several appendices (on structural-equation models, mixed models, survival analysis, etc.); scripts for the examples in all of the chapters and appendices; information on acquiring and installing R; and more. The book is associated with the car package for R. Alternatively (or additionally), more advanced students may wish to use W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S* as their principal source.

Manuals

S-PLUS is distributed with a set of manuals, as is R.

S-PLUS manuals are also available on the Insightful Corporation web site, at <http://www.insightful.com/support/documentation.asp>.

Likewise, R manuals are also available at the CRAN (Comprehensive R Archive Network) web site, <http://cran.r-project.org/manuals.html>.

A manual for S-PLUS Trellis Graphics (also useful for the lattice package in R) is at <http://cm.bell-labs.com/cm/ms/departments/sia/doc/trellis.user.pdf>. 
Programming in S


Selected Statistical Methods Programmed in S


to statistical modeling, with frequent references to Harrell’s Hmisc and Design packages for S-PLUS and R.


C. Loader, *Local Likelihood and Regression*. New York: Springer, 1999. Another text on nonparametric regression and density estimation, using the `locfit` package (in S-PLUS and R). Although the text is less readable than Bowman and Azzalini, the `locfit` software in very capable.


**Other Sources (Some Free)**