

R Workshop, Lund, May 14-16 2007.

This workshop is organized under a STINT [Institutional Grants Programme \(IGP\)](#) involving the [Department of Mathematical Statistics](#) in Lund and several universities in Australia and South Korea. Anyone interested in participating in the workshop should contact [Nader Tajvidi](#).

General

R has an unrivalled ability to handle a huge variety of statistical analyses, including some that take advantage of recent theoretical and computational advances. It is the work of leading experts in statistical computing, drawn from many different countries. It implements the flexible and powerful S language, so that it is highly extensible. It has become the system of choice for professionals who wish to develop new routines.

The R system provides, straightforwardly, well-designed publication-quality plots than can include mathematical symbols and formulae. While the user has full control when required, careful default graphical design choices keep user intervention to a minimum.

R runs on Windows, MacOS and a wide variety of UNIX and similar platforms (including FreeBSD and Linux), See the website <http://www.r-project.org/> for further information on R.

The presenter

John Maindonald (Centre for Mathematics and Its Applications, Australian National University), will present the workshop. John has had wide experience in many different areas of statistical application. He is the senior author of the widely used book *Data Analysis and Graphics Using R: An Example-Based Approach* (Cambridge University Press, 2003). John's interests include: Analysis of Microarray Data, Statistical Computation, Statistical Perspectives on Data Mining, and issues in Research Planning and Practical Data Analysis.

Notes

Notes will be supplied. Additionally, there will be extensive use of material from the text: Maindonald, J.H. and Braun, W.J. 2006. *Data Analysis and Graphics Using R. An Example-Based Approach*. 2nd edition, Cambridge University Press, to appear in January 2007.
[URL: <http://wwwmaths.anu.edu.au/~johnm/r-book.html>]

Use of computers

A limited number of computers will be available for use of participants. Alternatively, participants may bring their own laptops.

Who should attend?

This three-day workshop is intended for individuals with a good level of statistical knowledge, and who wish to learn to use R, or to improve their competence with R. It will demonstrate the use of R for data manipulation, for graphics, and for practical data analysis.

There will be some use of graphical user interfaces, but most input will be from the command line. Participants should be comfortable with the typing of language commands at the command line.

Participants are encouraged to attend on all days. Some may however wish to attend only for the overview of the R system on day one. Those who already have a good working knowledge of R may attend on days 2 and 3 only, or may wish to limit their attendance to the contributed talks on day 3. Offers of talks that demonstrate the use of R for practical data analysis, or that are otherwise R-related, will be welcomed.

Data for Possible Use in the Course

Participants who have data that they think might be suitable for use for illustrative purposes in the course are invited to contact the presenter ([John Maindonald](#)) with details.

Draft Schedule

DAY 1

9 - 9.30: Registration and setup

9.30 – 10.00am: Setup details & the working environment –

Getting help; Installing and using packages

Setting the working directory; Sessions and workspaces;

10.00 – 10.45am: Laboratory – Practice with R setup.

10.45 - 11.15: Morning tea

11.15 – 12.00: Overview and key concepts:

Demonstrations and examples,

Simple functions, Data frames, Data input, Simple graphs

12.00 - 1.15: Lunch

1.15 – 3.00: The R Language

Numeric, character and logical vectors, factors

Missing values, tables

Common sources of difficulty (and an occasional graph)

Functions, anonymous functions, and generic functions

Data manipulation and other functions; `sapply()`

Model, table and graphics formulae

3.00 - 3.30: Afternoon tea

3.30 – 5.00: Laboratory exercises

DAY 2 (May be varied, depending on preferences of participants, and on how much is covered on the first day)

9.00 - 10.45: Base and Trellis Graphics (including image plots)

Exploratory data analysis, graphs in model-based analysis, and presentation graphs

10.45 - 11.15: Morning tea

11.15 - 12.00:

Fitting statistical models

Model formulae, linear models, simple analysis of variance

12.00 - 1.15: Lunch

1.15 - 3.00:

Extensions of simple linear models; diagnostics,
Regression examples, regression splines
Generalized linear models

3.00 - 3.30: Afternoon tea

3.30 – 5.00: Laboratory exercises

DAY 3 Priority will be given to talks that discuss the use of R for practical data analysis, or that are otherwise likely to be of interest to users of R. Anyone who wishes to contribute should contact [Nader Tajvidi](#) or [John Maindonald](#).

Topics that may be canvassed, in any time that remains, include:

Multivariate methods and discriminant methods

Issues and topics left over from days 1 and 2

Overview of important R packages

Analysis of data supplied by participants
(I will need prior access to the data, and will be able to use only 2 or 3 such datasets.)

Preparation of documents that incorporate output from R – the Sweave compendium concept

Uses for advanced R features; computing with formulae and expressions, etc.

Creation of packages.