

Data analysis methods in weather and climate research

Dr. David B. Stephenson
Department of Meteorology
University of Reading
www.met.rdg.ac.uk/cag



Overview of R statistical language

- Brief history
- Why use R?
- The RCLIM project
- Demonstration

(c) 2005 D.B.Stephenson@reading.ac.uk

1

Types of statistical software

- Statistical language-based software
e.g. Splus, R, SAS
- Interactive Spreadsheet-like packages e.g.
Minitab, SPSS, Excel
- Data analysis software with stat routines e.g.
MATLAB, PV-Wave, IDL
- Home made subroutines
e.g. numerical recipes, friend's code, etc.

(c) 2005 D.B.Stephenson@reading.ac.uk

2

Brief history of R language

- 1960s: AT&T develop data handling language (S)
- 1970s: S gets updated and sold as Splus
- 1980s: Freeware version R developed from old S
- 1990s: Splus and R get widely used by statisticians
- 2000-: R starts to get more widely used e.g. by university statistics departments

(c) 2005 D.B.Stephenson@reading.ac.uk

3

Why use R?

- It's one of the most powerful high-level languages for doing statistical analysis
- It has built-in functions and contributed libraries that can do most modern statistical methods
- It allows you to learn up-to-date statistical practices from its comprehensive online help
- It enables you to communicate easily with statisticians
- It runs on most operating systems (UNIX, PC, Mac)
- It is freely available – 26Mb download from:

www.r-project.org

Some useful commands

```
> # The # symbol starts a comment statement
> help.start()      # launches help pages in browser
> q()               # quit R - save workspace saves all objs
> ?command         # gives help page on "command"
> command          # lists the function "command"
> read.table("file.txt") # reads in multicolumn tabbed data
> summary(x)       # summarises data object x
> dims(x)          # gives dimensions of array x
> new<-command     # puts output of command into new
```

(c) 2005 D.B.Stephenson@reading.ac.uk

5

Some useful tips

- Set up a separate work folder for each project
- In R, do **File > Change dir ...** to set this as your default directory
- Do **File > New script** to open up a text file for writing and saving your commands and analysis results
- Use **Ctrl-R** to execute highlighted commands in the script
- Use **Ctrl-S** to save the script (do this frequently)
- Use **Windows > Tile** to stop your windows getting cluttered up
- Use the help commands if you are not sure about anything:
`help.start()`, `?command`, `help.search("keyword")`
- To quit do `q()` and say yes to save workspace.
- Click on the `.RData` file (R icon) in the project directory to open R again

(c) 2005 D.B.Stephenson@reading.ac.uk

6

The RCLIM project

R is often faster and more accurate than many of the commercial packages such as MATLAB and IDL. However, it has a few areas that need improving for it to be able to do climate analysis:

- read/write large gridded fields in netcdf format
- do nice geographical contour maps
- do climate analysis at many grid points

The RCLIM project started in Reading in 2005 has started to write code to address these gaps. An RCLIM webpage will appear soon here:

www.reading.met.ac.uk/cag/rclim

(c) 2005 D.B.Stephenson@reading.ac.uk

7

Now for a live demonstration!

- Sampling the population (what is your height?)
- Exploratory Data Analysis
- Two sample comparison

(c) 2005 D.B.Stephenson@reading.ac.uk

8
