

WHAT I KNOW ABOUT R (Windows version)

(An evolving document)

January 18, 2002

Lance Waller

- **WHAT IS R?**

R is a free statistical analysis package. Its syntax follows the command language of S+ very closely but some of internal implementation is different. The internal implementation details are largely transparent to the user, except that looping tends to work much better in R than in S+. (This is not to say looping is perfect in R, but I do have at least one example of nested loops (20 iterations of the outer loop, 500 of the inner loop) that completely stalls a Unix workstation in S+ but runs quickly in R on a Windows machine).

More information on R appears on the website (see next question).

- **WHERE IS R?**

<http://cran.r-project.org/>

(Note: No `www` in the URL). This is the website of CRAN (the Comprehensive R Archive Network), and contains links to the source code and installation software. There are mirror sites as well, including a connection through StatLib at Carnegie Mellon University at

<http://lib.stat.cmu.edu/>

- **HOW DO I GET R?**

To get the source files for Windows, go to

<http://cran.r-project.org/>

Look under “Frequently used pages”. In the box under “Precompiled Binary Distribution”, click on “Windows (95 and later)”. Next, click on the “base” directory. The `ReadMe` and `ReadMe.rw1040` files contains a list of what each file contains, and some detailed documentation on how the Windows version of R differs from the Unix version). From this directory, you need to download the following file:

`SetupR.exe` (15Mb)

This file contains the set-up wizard and the binary files. It is large and can take a while to download. R is already installed on several of the computers in the Biostatistics computer lab.

There is also more helpful documentation on the CRAN website. On the lefthand side of the webpage (under the R icon) find “Documentation” and click on the link to “Manual”. Here you can download PDF files of *An Introduction to R* which provides an overview of using R in a variety of statistical analysis situations. If you click on “Contributed” instead of “Manuals” you’ll find a few other PDF files for beginners.

- **HOW DO I INSTALL R IN WINDOWS?**

The package `SetupR.exe` is very nice and uses wizard-like windows to install the base system and also lets you add additional libraries to the system without reinstalling the whole thing (see below). Just click on the `SetupR.exe` icon and follow the directions. It will create several subdirectories. Suppose you intalled R in the directory

```
c:\R\
```

Two subdirectories of particular note are

```
c:\R\rw1040\library
```

and

```
c:\R\rw1040\bin
```

as outlined in questions below.

At one point you will have a checklist of files to choose from. This also tells you how much space you need. The HTML help files add a lot of space. You don’t *need* to have the compiled HTML help if you have the text help, BUT the HTML help runs through Netscape/Internet Explorer and has a search engine so you can find help about keywords rather than trying to remember the exact command names. (When in R, you start the HTML help by typing `help.start()`, including the parantheses).

The file `SetupR.exe` also creates a shortcut to `Rgui`. `Rgui` is the GUI (graphical user interface) version of R that runs in Windows. You begin with a command window and R adds graphics windows as needed.

- **HOW DO I ADD EXTRA LIBRARIES TO R?**

Go to

<http://cran.r-project.org/>

Page down to “Frequently used pages”. On the right-hand side (Windows), click on “contributed packages”. This gives a directory of the .zip files with the libraries.

But how do you figure out which ones you want? Go back to the CRAN page (URL above), and click on “Sources of contributed packages” under “All Platforms” (the top box). Here’s a list that describes what is in each package. Read the “Index of Contents” to see what is in each packages, then download the PDF manual file for more detail.

When you have what you want, go back to the main CRAN page, click on “Windows (95 or later)” under “Precompiled Binary Distributions”. Now click on the directory “contrib”, then download the .zip file. Next run SetupR.exe, click on “An add-on package”, give the path name and install it. When it installs, it creates a folder with the name of the package. This folder needs to be placed in directory

```
c:\R\rw1040\library\
```

for R to know where it is. Either specify this directory in SetupR.exe or move the new folders there after installation.

To call the add-on package `akima` within R, enter

```
library(akima)
```

on the command line in Rgui. Then you can access any commands in the add-on package.

NOTE: Each time you start Rgui you’ll need to call the packages you want to include (e.g., you need to say `library(akima)` again the next time you run Rgui).

- **WHAT EXTRA LIBRARIES DOES LANCE USE?**

So far:

- * `akima` (includes the `interp()` function that is standard in S+ but not in R);
- * `KernSmooth` (kernel smoothing routines from Wand and Jones (1995) *Kernel Smoothing*, Chapman and Hall/CRC);
- * `VR` (actually several libraries containing R versions of all the S+ routines from Venables and Ripley’s (1994) *Modern Applied Statistics Using S-plus*, Springer);

- * `sgeostat` (spatial statistics routines by James Majure, NOTE: you also need to load the add-on package `tripack` for `sgeostat` to work).
- * `splancs` (spatial point process routines by Barry Rowlingson and Peter Diggle, Lancaster University).

- **IS THERE ANY INTRODUCTORY DOCUMENTATION FOR R?**

As noted above, there is helpful documentation on the CRAN website. On the lefthand side of the webpage (under the R icon) find “Documentation” and click on the link to “Manuals”. Here you can download PDF files of *An Introduction to R* which provides an overview of using R in a variety of statistical analysis situations. If you click on “Contributed” instead of “Manuals” you’ll find a few other PDF files for beginners.

- **WHAT ARE SOME COOL THINGS YOU CAN DO IN R THAT YOU CAN’T IN S+?**

- * `read.csv()`: This reads in a comma-delimited `.csv` file without specifying all the details in `read.table()` (e.g., `sep=’, ’`). Very handy for inputting data from an Excel worksheet (save as `.csv`, open `Rgui`, use `read.csv()`).
- * `filled.contour()`: makes a contour plot and fills in the spaces between contour lines with color gradients. Makes it very easy to tell if a collection of contours means a peak or a valley.
- * `persp()`: slightly different from the S+ version. R’s `persp()` lets you identify the viewing location by polar coordinates and specify the vanishing point perspective. This works much better than the `eye` part of S+’s `persp`. Also R’s `persp()` lets you add color and shading (you can even control the coordinates of the light source for shading). Very nice.
- * As mentioned above, nested looping works better in R than S+. I have an example, if interested.

- **SOME EXAMPLES OF THINGS LANCE HAS FOUND THAT WORK DIFFERENTLY IN R AND S+**

- * If you are fitting a regression model with polynomial terms, e.g.,

$$E[Y] = \beta_0 + \beta_1 x + \beta_2 x^2$$

in S+ you say

```
polyreg <- lm(y ~ x + x^2)
```

If you try this in R, you get results for the model

$$E[Y] = \beta_0 + \beta_1 x,$$

so apparently you can't define the new covariate x^2 on the fly. Instead you need to say

```
polyreg <- lm(y ~ x + I(x^2))
```

This is only vaguely revealed in the documentation. I suppose the `I()` notation works in S+, but at least `lm()` also works in S+ without it.

• SOME OTHER TIDBITS.

- * *How to get out of Rgui:* Enter `q()` at the command prompt. If you want variables you defined during your session to be available next session, save your workspace when prompted.
- * *Lance's standard operating procedure.* Have either a Notepad or Wordpad file open with my commands. Open Rgui and cut and paste commands from my file to the command line. This lets me save a long sequence of commands, and reduces the amount of retyping commands over and over. I also generally "tile" the windows: after the first plot, click on the command window in Rgui (to make it active), then chose "Tile" from the "Windows" menu.
- * *Making postscript files of figures.* Although you can click on the right mouse button (when the cursor is over the graphics window) and choose "Save as postscript", this doesn't work very well (you get a tiny copy of your plot in the middle of an otherwise empty page. Instead, enter the command

```
postscript("c:\R\myspsfile.ps")
```

then enter the commands to make your plot (nothing will show on your screen since it will be output to the postscript file). When you are done enter

```
dev.off()
```

which closes the .ps file (the current graphics device). If you only give a file name and not the full path name, look for your file in the directory

```
c:\R\rw1040\bin\
```