News

The Newsletter of the R Project

Volume 6/4, October 2006

Editorial

by Paul Murrell

Welcome to the second regular issue of R News for 2006, which follows the release of R version 2.4.0. This issue reflects the continuing growth in R's sphere of influence, with articles covering a wide range of extensions to, connections with, and applications of R. Another sign of R's growing popularity was the success and sheer magnitude of the second useR! conference (in Vienna in June), which included over 180 presentations. Balasubramanian Narasimhan provides a report on useR! 2006 on page 45 and two of the articles in this issue follow on from presentations at useR!.

We begin with an article by Max Kuhn on the **odfWeave** package, which provides literate statistical analysis *à la* **Sweave**, but with ODF documents as the medium, rather than TEX documents. Next up is Jim Lemon with an introduction to his **plotrix** package for producing a variety of plot customisations. From graphical output to graphical user interfaces, Adrian Bowman, Crawford, and Bowman describe the **rpanel** package (introduced at useR!) which provides a friendlier wrapper on the **tcltk** package, with some nice examples of simple interactive graphics.

Matthew Pocernich takes a moment to stop and smell the CO_2 levels and describes how R is involved in some of the reproducible research that informs the climate change debate.

The next article, by Roger Peng, introduces the **filehash** package, which provides a new approach to the problem of working with large data sets in R. Robin Hankin then describes the **gsl** package, which implements an interface to some exotic mathematical functions in the GNU Scientific Library.

The final three main articles have more statistical content. Wolfgang Lederer and Helmut Küchenhoff describe the **simex** package for taking measurement error into account. Roger Koenker (another useR! presenter) discusses some non-traditional link functions for generalised linear models. And Víctor Leiva and co-authors introduce the **bs** package, which implements the Birnbaum-Saunders Distribution.

In other contributions, Susan Holmes provides a book review of Fionn Murtagh's book "Correspondence Analysis and Data Coding with Java and R" and Uwe Ligges provides an R Help Desk article on how to find the source code behind an R function.

The issue ends with our regular updates on changes to R itself, new packages on CRAN, new

Contents of this issue:

members and contributors to the R Foundation, and recent news from the Bioconductor project.

Finally, I would like to remind everyone that the next DSC conference is taking place in Auckland, New Zealand, next February 15 & 16. I hope to see you here!

Paul Murrell
The University of Auckland, New Zealand
paul@stat.auckland.ac.nz

Sweave and the Open Document Format – The odfWeave Package

by Max Kuhn

Introduction

When documenting the results of a data analysis, creating the report can easily take more time than the analysis. There are at least two approaches to this

- Create and format tables and plots, then manually add text around these elements. In this case, the analysis code has no connection to the report.
- Take a combined approach, where the analysis results and the report are created at the same time using the same source code.

In the latter case, the `Sweave` function (Leisch, 2002) is a powerful component of R that can be used to combine R code with LaTeX so that the output is embedded in the processed document.

The user could write a document in LaTeX that contains `Sweave` tags. These tags encapsulate R commands. For example, an in–line `Sweave` tag might look like `\Sexpr{sqrt(2)}`. When the LaTeX document is processed by R's `Sweave` function, the `\Sexpr` tag is a signal to R to insert the results of the command `sqrt(2)` in place of the tag.

When "weaving" with LaTeX, the user has the ability to embed textual R output, as well as more complex types of R output (such as tables). After weaving, standard LaTeX tools can be used to generate a final document file, usually in postscript or PDF format.

For example, R package vignettes are created using LaTeX and `Sweave` (Leisch, 2003). For those readers who are new to `Sweave`, typing `vignette()` at the R prompt will display a list of installed packages that contain manuals that were written using LaTeX and `Sweave`. The package sources contain the underlying `Sweave`/LaTeX files and are a good resource for getting started with `Sweave` and LaTeX.

The capabilities of `Sweave` were later extended to HTML format in the **R2HTML** package. This provided the same functionality, but uses the HTML markup language.

While `Sweave` is highly effective at producing beautiful documents that show the many features of R, there are at least two limitations created by using LaTeX or HTML as the markup language. First, it requires the user to have a moderate knowledge of the markup languages to create and produce the documents.

Secondly, using these particular markup languages limits the main file formats to Postscript, PDF or HTML. If the user is collaborating with scientists or researchers who are not comfortable with editing LaTeX or HTML, it is difficult for them to add their subject-specific insight to the document. For example, at Pfizer, R is used as a computational engine for gene expression and computational chemistry data analysis platforms. Scientists can upload their data, specify models and generate results. We would like to provide them with the results of the analysis in a format that enables them to frame the statistical output in the context of their problem

The Open Document Format and The `odfWeave` Package

The Open Document Format (ODF) is an XML–based markup language. ODF is an open, non–proprietary format that encompasses text documents, presentations and spreadsheets. Version 1.0 of the specification was finalized in May of 2005 (OASIS, 2005). One year later, the format was approved for release as an ISO and IEC International Standard.

ODF has the advantage of being the default format for the OpenOffice (version 2.0 and above). OpenOffice is is a free, open source office suite and can export files to many different formats, including MS Word, rich text format, HTML, PDF and others. Other applications, such as Koffice, use the Open Document Format. Also, Microsoft has recently indicated that it will host an open source project converting documents to ODF from within Microsoft Office.

If OpenOffice documents were to support `Sweave` tags, reports could be edited and formatted in a sophisticated GUI environment. This would allow R users to include `Sweave` tags without directly editing