Changes in R

by the R Core Team

New features in version 1.3.0

- Changes to connections:
  - New function `url()` to read from URLs. `file()` will also accept URL specifications, as will all the functions which use it.
  - File connections can now be opened for both reading and writing.
  - Anonymous file connections (via `file()`) are now supported.
  - New function `gzfile()` to read from / write to compressed files.
  - New function `fifo()` for connections to / from fifos (on Unix).
  - Text input from file, pipe, fifo, gzfile and url connections can be read with a user-specified encoding.
  - New functions `readChar()` to read character strings with known lengths and no terminators, and `writeChar()` to write user-specified lengths from strings.
  - `sink()` now has a stack of output connections, following S4.
  - `sink()` can also be applied to the message stream, to capture error messages to a connection. Use carefully!
  - `seek()` has a new 'origin' argument.
  - New function `truncate()` to truncate a connection open for writing at the current position.
  - New function `socketConnection()` for socket connections.
  - The 'blocking' argument for file, fifo and socket connections is now operational.
- Changes to date/time classes and functions:
  - Date/time objects now all inherit from class `"POSIXt"`.
  - New function `difftime()` and corresponding class for date/time differences, and a `round()` method.
  - Subtraction and logical comparison of objects from different date/time classes is now supported. NB: the format for the difference of two objects of the same date/time class has changed, but only for objects generated by this version, not those generated by earlier ones.
  - Methods for `cut()`, `seq()`, `round()` and `trunc()` for date/time classes.
  - Convenience generic functions `julian()`, `weekdays()`, `months()`, and `quarters()` with methods for `"POSIXt"` objects.
- Coercion from real to integer now gives `NA` for out-of-range values, rather than the most extreme integer of the same sign.
- The Ansari-Bradley, Bartlett, Fligner-Killeen, Friedman, Kruskal-Wallis, Mood, Quade, *t*, and Wilcoxon tests as well as `var.test()` in package **ctest** now have formula interfaces.
- The matrix multiplication functions `%*%` and `crossprod()` now use a level-3 BLAS routine dgemm. When R is linked with the ATLAS or other enhanced BLAS libraries this can be substantially faster than the previous code.
- New functions `La.eigen()` and `La.svd()` for eigenvector and singular value decompositions, based on LAPACK. These are preferred to `eigen()` and `svd()` for new projects and can make use of enhanced BLAS routines such as ATLAS. They are used in `cancor()`, `cmdscale()`, `factanal()` and `princomp()` and this may lead to sign reversals in some of the output of those functions.
- Provided the FORTRAN compiler can handle `COMPLEX*16`, the following routines now handle complex arguments, based on LAPACK code: `qr()`, `qr.coef()`, `qr.solve()`, `qr.qy()`, `qr.qty()`, `solve.default()`, `svd()`, `La.svd()`.
- `aperm()` uses strides in the internal C code and so is substantially faster (by Jonathan Rougier).
- The four `bessel[IJKY](x,nu)` functions are now defined for nu < 0.
- `[dpqr]nbinom()` also accept an alternative parametrization via the mean and the dispersion parameter (thanks to Ben Bolker).
- New generalised "birthday paradox" functions `[pq]birthday()`.
- `boxplot()` and `bxp()` have a new argument 'at'.
- New function `capabilities()` to report optional capabilities such as jpeg, png, tcltk, gzfile and url support.
- New function `checkFF()` for checking foreign function calls.

 ISSN 1609-3631

- New function `col2rgb()` for color conversion of names, hex, or integer.
- `coplot()` has a new argument 'bar.bg' (color of conditioning bars), gives nicer plots when the conditioners are factors, and allows factors for x and y (treated almost as if `unclass()`ed) using new argument 'axlabels'. [Original ideas by Thomas Baummann]
- 'hessian' argument added to `deriv()` and its methods. A new function `deriv3()` provides identical capabilities to `deriv()` except that 'hessian' is `TRUE` by default. `deriv(*, *, func = TRUE)` for convenience.
- New `dev.interactive()` function, useful for setting defaults for `par(ask=*)` in multifigure plots.
- `dist()` in package **mva** can now handle missing values, and zeroes in the Canberra distance.
- The default method for `download.file()` (and functions which use it such as `update.packages()`) is now `"internal"`, and uses code compiled into R.
- `eigen()` tests for symmetry with a numerical tolerance.
- New function `formatDL()` for formatting description lists.
- New argument 'nsmall' to `format.default()`, for S-PLUS compatibility (and used in various packages).
- `?/help()` now advertises `help.search()` if it fails to find a topic.
- `image()` is now a generic function.
- New function `integrate()` with S-compatible call.
- New function `is.unsorted()` the C version of which also speeds up `.Internal(sort())` for sorted input.
- `is.loaded()` accepts an argument 'PACKAGE' to search within a specific DLL/shared library.
- Exact *p*-values are available for the two-sided two-sample Kolmogorov-Smirnov test.
- `lm()` now passes '`...`' to the low level functions for regression fitting.
- Generic functions `logLik()` and `AIC()` moved from packages **nls** and **nlme** to **base**, as well as their lm methods.
- New components in `.Machine` give the sizes of `long`, `long long` and `long double` C types (or 0 if they do not exist).
- `merge.data.frame()` has new arguments, 'all[.xy]' and 'suffixes', for S compatibility.
- `model.frame()` now calls na.action with the terms attribute set on the data frame (needed to distiguish the response, for example).
- New generic functions `naprint()`, `naresid()`, and `napredict()` (formerly in packages **MASS** and **survival5**, also used in package **rpart**). Also `na.exclude()`, a variant on `na.omit()` that is handled differently by `naresid()` and `napredict()`.

  The default, lm and glm methods for `fitted()`, `residuals()`, `predict()` and `weights()` make use of these.
- New function `oneway.test()` in package **ctest** for testing for equal means in a one-way layout, assuming normality but not necessarily equal variances.
- `options(error)` accepts a function, as an alternative to an expression. (The Blue Book only allows a function; current S-PLUS a function or an expression.)
- `outer()` has a speed-up in the default case of a matrix outer product (by Jonathan Rougier).
- `package.skeleton()` helps with creating new packages.
- New `pdf()` graphics driver.
- `persp()` is now a generic function.
- `plot.acf()` makes better use of white space for 'nser > 2', has new optional arguments and uses a much better layout when more than one page of plots is produced.
- `plot.mts()` has a new argument 'panel' providing the same functionality as in `coplot()`.
- `postscript()` allows user-specified encoding, with encoding files supplied for Windows, Mac, Unicode and various others, and with an appropriate platform-specific default.
- `print.htest()` can now handle test names that are longer than one line.
- `prompt()` improved for data sets, particularly non-dataframes.
- `qqnorm()` is now a generic function.
- `read.fwf()` has a new argument 'n' for specifying the number of records (lines) read in.
- `read.table()` now uses a single pass through the dataset.

- `rep()` now handles lists (as generic vectors).
- `scan()` has a new argument 'multi.line' for S compatibility, but the default remains the opposite of S (records can cross line boundaries by default).
- `sort(x)` now produces an error when `x` is not atomic instead of just returning `x`.
- `split()` now allows splitting on a list of factors in which case their interaction defines the grouping.
- `stl()` has more optional arguments for fine tuning, a `summary()` and an improved `plot()` method.
- New function `strwrap()` for formatting character strings into paragraphs.
- New replacement functions `substr<-()` and `substring<-()`.
- Dataset `swiss` now has row names.
- Arguments 'pkg' and 'lib' of `system.file()` have been renamed to 'package' and 'lib.loc', respectively, to be consistent with related functions. The old names are deprecated. Argument 'package' must now specify a single package.
- The Wilcoxon and Ansari-Bradley tests now return point estimators of the location or scale parameter of interest along with confidence intervals for these.
- New function `write.dcf()` for writing data in Debian Control File format. `parse.dcf()` has been replaced by (much faster) internal `read.dcf()`.
- Contingency tables created by `xtabs()` or `table()` now have a `summary()` method.
- Functions `httpclient()`, `read.table.url()`, `scan.url()` and `source.url()` are now deprecated, and hence 'method="socket"' in `download.file()` is. Use url connections instead: in particular URLs can be specified for `read.table()`, `scan()` and `source()`.
- Formerly deprecated function `getenv()` is now defunct.
- Support for package-specific demo scripts (R code). `demo()` now has new arguments to specify the location of demos and to allow for running base demos as part of 'make check'.
- If not explicitly given a library tree to install to or remove from, respectively, `R CMD INSTALL` and `R CMD REMOVE` now operate on the first directory given in `R_LIBS` if this is set and non-null, and the default library otherwise.
- `R CMD INSTALL` and `package.description()` fix some common problems of 'DESCRIPTION' files (blank lines, ...).
- The `INSTALL` command for package installation allows a '`--save`' option. Using it causes a binary image of the package contents to be created at install time and loaded when the package is attached. This saves time, but also uses a more standard way of source-ing the package. Packages that do more than just assign object definitions may need to install with '`--save`'. Putting a file 'install.R' in the package directory makes '`--save`' the default behavior. If that file is not empty, its contents should be R commands executed at the end of creating the image.

  There is also a new command line option '`--configure-vals`' for passing variables to the configure script of a package.
- `R CMD check` now also checks the keyword entries against the list of standard keywords, for code/documentation mismatches (this can be turned off by the command line option '`--no-codoc`'), and for sufficient file permissions (Unix only). There is a new check for the correct usage of `library.dynam()`.

  It also has a new command line option '`--use-gct`' to use '`gctorture(TRUE)`' when running R code.
- `R CMD Rd2dvi` has better support for producing reference manuals for packages and package bundles.
- configure now tests for the versions of jpeg ($\geq$ 6b), libpng ($\geq$ 1.0.5) and zlib ($\geq$ 1.1.3). It no longer checks for the CXML/DXML BLAS libraries on Alphas.
- Perl scripts now use `Cwd::cwd()` in place of `Cwd::getcwd()`, as `cwd()` can be much faster.
- 'R::Dcf.pm' can now also handle files with more than one record and checks (a little bit) for continuation lines without leading whitespace.
- New manual 'R Installation and Administration' with fuller details on the installation process: file 'INSTALL' is now a brief introduction referencing that manual.
- New keyword 'internal' which can be used to hide objects that are not part of the API from indices like the alphabetical lists in the HTML help system.

- Under Unix, shlib modules for add-on packages are now linked against R as a shared library ('libR') if this exists. (This allows for improved embedding of R into other applications.)
- New mechanism for explicitly registering native routines in a DLL/shared library accessible via `.C()`, `.Call()`, `.Fortran()` and `.External()`. This is potentially more robust than the existing dynamic lookup, since it checks the number of arguments, type of the routine.
- New mechanism allowing registration of C routines for converting R objects to C pointers in `.C()` calls. Useful for references to data in other languages and libraries (e.g. C and hdf5).
- The internal ftp/http access code maintains the event loop, so you can download whilst running tcltk or Rggobi, say. It can be hooked into package **XML** too.

New features in version 1.2.3

- Support for configuration and building the Unix version of R under Mac OS X. (The 'classic' Macintosh port is 'Carbonized' and also runs under that OS.)
- `dotchart()` and `stripchart()` become the preferred names for `dotplot()` and `stripplot()`, respectively. The old names are now deprecated.
- Functions in package **ctest** now consistently use `+/-Inf` rather than `NA` for one-sided confidence intervals.

New features in version 1.2.2

- The Macintosh port becomes a full member of the R family and its sources are incorporated as from this release. See 'src/macintosh/INSTALL' for how that port is built.
- The API header files and export files 'R.exp' are released under LGPL rather than GPL to allow dynamically loaded code to be distributed under licences other than GPL.
- `postscript()` and `xfig()` devices now make use of genuine Adobe afm files, and warn if characters are used in string width or height calculations that are not in the afm files.
- Configure now uses a much expanded search list for finding a FORTRAN 77 compiler, and no longer disallows wrapper scripts for this compiler.
- New Rd markup `\method{GENERIC}{CLASS}` for indicating the usage of methods.
- `print.ftable()` and `write.ftable()` now have a 'digits' argument.
- `undoc()` has a new 'lib.loc' argument, and its first argument is now called 'package'.

Changes on CRAN

by Kurt Hornik and Friedrich Leisch

CRAN packages

The following extension packages from 'src/contrib' were added since the last newsletter.

AnalyzeIO Functions for I/O of ANALYZE image format files. By Jonathan L Marchini.

CoCoAn Two functions to compute correspondence analysis and constrained correspondence analysis and to make the associated graphical representation. By Stephane Dray.

GLMMGibbs Generalised Linear Mixed Models are an extension of Generalised Linear Models to include non-independent responses. This package allows them to be fitted by including functions for declaring factors to be random effects, for fitting models and generic functions for examining the fits. By Jonathan Myles and David Clayton.

GeneSOM Clustering Genes using Self-Organizing Map. By Jun Yan.

Oarray Generalise the starting point of the array index, e.g. to allow `x[0, 0, 0]` to be the first element of `x`. By Jonathan Rougier.

PTAk A multiway method to decompose a tensor (array) of any order, as a generalisation of SVD also supporting non-identity metrics and penalisations. 2-way SVD with these extensions is also available. The package includes also some other multiway methods: PCAn (Tuckern) and PARAFAC/CANDECOMP with these extensions. By Didier Leibovici.