

Changes in R

by the R Core Team

New features in version 1.2.1

- Package `mva` has new functions `factanal()`, `varimax()`, `promax()`, and `examples`.
- New functions `readBin()` and `writeBin()` to transfer binary data to and from connections.
- `merge()` is partially moved to C to reduce its memory usage.
- `library(help = PKG)` now displays the contents of the package's 'DESCRIPTION' file in addition to its 'INDEX'.
- `Sd2Rd` can handle S4-style documentation too: see "*Writing R Extensions*".
- `prompt()` now also works with a character argument (useful for producing many '*.Rd' files in a loop).
- The Unix front-end shell script now ignores a value for `R_HOME` found in the environment.
- Connections functions such as `file()` now accept a description of length > 1, with a warning.
- All text-mode connections now accept input with LF, CR or CRLF line endings. This means that `readLines()` can be used on DOS files and `source()` on Mac files, for example.
Also, CRLF-terminated files can be used as `stdin` on Unix, and files with last lines without an EOL mark can be used as `stdin` and `source()`-ed on Unix and Windows.
- 'DESCRIPTION' file has a new recommended 'Maintainer:' field.
- `stars()` now uses a larger 'cex' for the labels, and 'cex' and 'lwd' are now arguments. Further, the argument names (`xlim`, `ylim`, `axes`) are now consistent with other plot functions. The key symbol is not clipped anymore into the plot region by default.
- Date-time quantities are now printed with the `timezone`, if known.
- R CMD `build` now ignores all files specified (via Perl regexps) in file '`Rbuildignore`' in the top-level source directory of a package.
- Horizontal boxplots are possible with '`horizontal = TRUE`'.

- `all.equal()` on lists now compares them as generic vectors, that is they are equal if have identical names attributes and all components are equal.
- Invalid lines in '`.Renviron`' now give warnings when R is started.
- Argument '`na.last`' implemented for `rank()`.

New features in version 1.2.0

- There is a new memory management system using a generational garbage collector. This improves performance, sometimes marginally but sometimes by double or more. The workspace is no longer statically sized and both the vector heap and the number of nodes can grow as needed. (They can shrink again, but never below the initially allocated sizes.) See `?Memory` for a longer description, including the new command-line options to manage the settings.
- values of '`--min-nsize`' up to 50M (2Gb on 64-bit Solaris) are allowed.
- A (preliminary) version of S4-like connections has been added, and most functions which take a 'file' argument can now work with a connection as well as a file name. For more details, see the chapter on Connections in the "*R Data Import/Export*" manual.
- New command-line option '`--no-restore-history`' implied by '`--vanilla`'.
- Command-line option '`--no-restore`' is now '`--no-restore-data`' and '`--no-restore`' implies '`--no-restore-*`' (currently 'data' and 'history').
- The more recent GNU `regex` from `grep-2.4.2` is used. This uses locale-based ordering for ranges on platforms with `strcoll`.
- The print routines now escape '"' (as '\") in a character string only when it is printed as a quoted string. This makes `print()`, `quote=FALSE`) and `cat()` consistent.
- The standard methods for `add1()` and `drop1()` now attempt to cope with missing values by using a subset of the data that is "cleaned" by `na.action` for the maximal model under consideration.

- `anova()` for 3 or more `lm` objects now behaves compatibly with `S` and `anova.glm1ist()`. The old behaviour is still available by calling `anova1ist.lm()` directly.
- `anova()` for multiple `lm` and `glm` objects no longer truncates the formula printed. There is much more extensive documentation for `anova()` methods.
- New method `as.data.frame.table()` for converting the array-based representation of a contingency table to a data frame containing the classifying factors and the corresponding counts.
- New function `assocplot()` for producing Cohen-Friendly association plots.
- `autoload()` accepts 'lib.loc' and other arguments to `library()`.
- `bxp()` has new argument 'frame.plot', as `plot.default()`.
- `contour()` now has 'axes' and 'frame.plot' args.
- `contrasts(, FALSE)` now always returns an identity matrix, to make `model.matrix` compatible with `S`. This affects models such as `lm(y ~ o - 1)` where `o` is an ordered factor.
- 'where' command added to `debug()`.
- `demo(dynload)` (which used the superseded `call_R` interface) has been removed.
- Class "dendrogram" in package `mva` providing general support for tree-like structures (plotting, cutting, ...).
- `dev.copy2eps()` and `dev2bitmap()` preserve the aspect ratio of the copied device if just one of 'width' and 'height' is specified.
- `dump()` has new argument 'append', argument 'fileout' has been renamed to 'file' (for consistency with all other functions).
- `edit.default()` now checks for an unset 'editor' argument, and terminates with an error if the editor cannot be run.
- The 'mode' argument of `exists()` and `get()` is interpreted as `mode(x)` rather than `typeof(x)`, following `S`.
- New functions `file.access()` and `file.info()` for information on files on the user's file systems.
- New convenience function `file.copy()`.
- `file.show()` allows 'pager' argument to be an R function, and consequently, the 'pager' option can be an R function.
- Formatting (and printing) of `data.frames` with complex objects is improved. `toString()` was added as a new function.
- `format()` has a new argument 'justify' controlling the justification of character strings (and factors).
- Formula objects now have an environment and code manipulating them needs to take care to preserve it or set an appropriate environment.
- New function `fourfoldplot()` for producing fourfold displays of 2 by 2 by k contingency tables.
- `gc()` now reports the space allocated, not the space free, since the total space is now variable.
- New primitive `gc.time()` to report on time spent in garbage collection.
- `hclust()` takes new argument 'members' allowing dissimilarity matrices both for singletons (as until now) and clusters.
- `help()` has an additional 'pager' argument which may be passed to `file.show()` (useful for ESS fans).
- There is now an R 'Hershey' list object for Hershey vector font computations and documentation.
- `hist()` now returns an object of class "histogram" and calls the new function `plot.histogram()` for plotting. It now also allows character labels.
- `if(*)` now gives a more intelligible error message when '*' cannot be coerced to logical.
- `inherits()` is now an internal function and compatible with `S`.
- New function `lag.plot()` in package `ts`.
- `legend()` has a new argument 'pt.bg'.
- The commands history can be loaded with `loadhistory()`, saved with `savehistory()` and displayed with `history()`, under Windows and under Unix using the readline or GNOME interfaces.
- `mad()` has new (logical) arguments 'low' and 'high' (the first giving `S` compatibility).
- New function `manova()` and summary method.

- Function `mantelhaen.test()` in package `ctest` now can deal with general $I \times J \times K$ tables. In addition, in the $2 \times 2 \times K$ case, it can also perform an exact conditional test of independence, and gives confidence intervals for the common odds ratio.
- `model.frame()` now uses the environment of its formula argument, rather than the parent environment, to evaluate variables not found in the data argument. See `help(formula)`.
- `mosaicplot()` can now also create extended mosaic plots, which visualize the residuals from a log-linear model using color and outline.
- New utility function `n2mfrow()`.
- `nlm(check.analyticals = TRUE)` now warns if the supplied gradient and/or hessian are of the wrong length.
- New function `object.size()` to give approximate memory allocation.
- `optim()` now checks the length of an analytical gradient at each evaluation.
- The L-BFGS-B method of `optim()` now support tracing, at several levels of detail.
- `options(check.bounds = TRUE)` makes each vector extension by sub-assignment produce a warning.
- `options(width)` now admits to a limit (previously 200, now 10000) and gives a more informative message if out of range (as it does now for digits and expressions).
- Function `path.expand()` to do tilde-expansion on file paths. This provides an interface to `R_ExpandFileName`, which is now a documented entry point.
- `.Platform` has new component `endian`, useful for binary file manipulations.
- `plot.function()` and `curve()` now take `xlim` as default for `(from,to)` if the former is specified.
- `plot.hclust()` allows arguments `'main'`, `'sub'`, etc., and has non-empty defaults for these.
- `plot.ts(x,y)` now allows to suppress labels and lines; it is better documented.
- The `postscript()` driver now allows a user-specified family so, for example, one can use the same fonts in diagrams as in running text.
- The `postscript()` driver allows its prolog to be changed (by an expert) via object `.ps.prolog`.
- `prop.table()` and `margin.table()` now work with an empty `'margin'`.
- Formerly deprecated function `provide()` is now defunct.
- New functions `read.delim()/read.delim2()` to make it easier to read delimited files as Windows programs tend to create (usually TAB separated).
- New `readLines()` function to read a file line-by-line.
- New functions `reshapeLong()` and `reshapeWide()` emulating Stata's `reshape` command. These are still labeled experimental and might be improved (or removed) in later versions.
- `row.names()` and `row.names<-()` are now generic functions which call `rownames()` as their default method and have methods for class `"data.frame"`.
- New function `Rprof()` for profiling R expressions under Unix. Configure with `'--enable-R-profiling'` (on by default) to make this operational.
- `save(, oldstyle=TRUE)` has been withdrawn.
- `scan()` and `read.table()` have a new argument `'fill'` which can be set `TRUE` to allow reading files with unequal number of fields per line. (Programs like Excel have a habit of creating such files when exporting.)
- `scan()` and `read.table()` have a new argument `'blank.lines.skip'` to allow blank lines to be read.
- `scan()` now reads empty character fields as `"` not `"NA"` unless `"` is included in `na.strings`.
- `smooth()` in package `eda` has a better default (`3RS3R` instead of `3RSR`) and more arguments, e.g., `'twiceit'` for some S compatibility and `'kind = "3R"'` for running medians of 3.
- `strsplit()` has a new argument `'extended'` controlling whether to use extended (the default) or basic regular expressions for splitting.
- `Sys.getenv()` becomes the preferred name for `getenv()`, which is now deprecated.
- New functions `Sys.getlocale()` and `Sys.setlocale()` to query and set aspects of the locale of the R process, and `Sys.localeconv()` to find the default decimal point, etc.

- New function `Sys.info()` for platform, host and user information.
- New function `Sys.putenv()` to set environment variables.
- New function `Sys.sleep()` to suspend execution for a while.
- Date-time support functions with classes "POSIXct" and "POSIXlt" to represent dates and times (resolution 1 second) in the POSIX formats. Functions include `Sys.time()`, `as.POSIXct()`, `strptime()`, `strftime()`, and methods for `format`, `plot`, `c`, There are conversion functions for objects from packages **date** and **chron**; unlike those packages these support functions know about time zones (if the OS does).
- **tcltk** package now has `tkpager()` which is designed to be used by `file.show()` and shows help pages etc. in separate text widgets.
- **tcltk** is now more careful about removing the objects representing widgets in the R workspace when the windows are destroyed (e.g., using window manager controls)
- **tcltk** package has had several canvas functions implemented.
- **tcltk** now wraps callbacks to R in a `try()` construct—the nonlocal return from R's error handling could bring the Tk system into a strange state.
- New demos for **tcltk**: `tkfaq`, `tkfilefind`, `tkcanvas`.
- `termplot()` now has an 'ask' argument.
- `terms()` creates objects which now inherit from class "formula", so for example `as.formula(terms.object)` needs to be replaced by `formula(terms.object)`.
- `traceback()` is now printed un-quoted and labelled by the frame number.
- New argument 'recursive' to `unlink()`. The default behaviour on Unix is now that of `rm -f`, not `rm -rf`. `unlink()` is now compatible across platforms.
- New functions `write.ftable()` and `read.ftable()` for writing out and reading in flat contingency tables.
- `write.table()` now quotes factor columns if 'quote=TRUE', and has a new argument 'qmethod' to control the escaping of embedded quotes in character or factor columns.
- New function `xtabs()` providing a formula interface to cross tabulation.
- The "R Data Import/Export" ('R-data.texi') manual has been added.
- The set of valid R names is now described (at last) in R-intro.
- The "R Language Definition" ('R-lang.texi') manual is now included and built in the same way as the other manuals.
- The R manuals (R-intro, R-exts, ...) are converted to HTML format (if the necessary Texinfo tools are available) and linked into the top HTML help page.
- The header file 'R.h' and those included from it are now usable with C++ code.
- New header file 'R_ext/Boolean.h': `Rboolean` type with `TRUE` and `FALSE` enum constants.
- New header file 'Rgraphics.h' to allow addons to use graphics structures.
- Recommended include file 'Rmath.h' replaces 'R_ext/Mathlib.h'.
- Bessel, beta and gamma functions are now documented as part of the API. Undocumented entry points are no longer in the header files, and some are no longer visible.
- `Calloc` and `Realloc` failures now give size information.
- 'DESCRIPTION' file in installed packages has a new 'Built:' field giving build information (R version, platform, date).
- Much improved support for C++ code in addon packages under Unix. New configure/build variables `SHLIB_CXXLD` and `SHLIB_CXXLD_FLAGS` for specifying the command and flags needed for building shared libraries containing objects from a C++ compiler. Configure tries to get these right in typical cases (GNU tools and/or common platforms). C++ source suffixes '.cpp' and '.C' are now recognized in addition to '.cc'.
- Configure/build variables `MAINLD` and `MAINLD_FLAGS` are renamed to `MAIN_LD` and `MAIN_LD_FLAGS` for consistency with other `MAIN_*` variables, similarly for `SHLIBLD` and `SHLIBLD_FLAGS`.
- Configure/build variable `FLIBS` now only contains the FORTRAN 77 intrinsic and run-time libraries needed for linking a FORTRAN 77 program or shared library (as determined by configure). BLAS library detection was extended, with results saved to the Make variable `BLAS_LIBS` which is also available to addon packages.

- R CMD `build` and `check` have been completely re-written in Perl. In addition to running examples, `check` now also checks the directory structure and control files, makes a temporary installation and runs \LaTeX on the help pages. `build` has been reduced to cleaning, rewriting indices and creating tar files.

The same files of Perl code are now also used under Windows.

- Add-ons for utilities like Perl or \LaTeX have now

a central place in `'$R_HOME/share'`. Migration of existing files might take a while, though.

- Preliminary support for building R as a shared library ('libR') under Unix. Use `configure` with option `'--enable-R-shlib'` or do `make libR` in directory `'src/main'` to create the shared library.

There is also a linker front-end R CMD `LINK` which is useful for creating executable programs linked against the R shared library.

Changes on CRAN

by Kurt Hornik and Friedrich Leisch

Introduction

This column, named *Changes on CRAN*, will be one of the regular columns appearing in every volume of the newsletter. We will try to shortly summarize all changes on CRAN and the R web pages, list new or updated extension packages, new manuals, etc.

Split of CRAN and R homepage

During the second half of 2000 we have split up R's web pages into two separate web sites:

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

The first is meant as R's central homepage, giving information on the R project and everything related to it. The second—CRAN—acts as the download area, carrying the software itself, extension packages, PDF manuals; in short everything you may need to download for using R.

The main motivations for this "artificial" split were:

- CRAN is mirrored on a number of sites worldwide, hence we should try to keep it as small as possible to make life for the mirror sites easier. It should carry the material users of R need to download on a regular basis, such that having a mirror nearby pays off.
- We do not want to be (technically) limited in the possibilities how to present material on the homepage of R. However, a site like CRAN that is intended for other sites to mirror is very limited, because the administrators have no control of the mirror web servers. Hence, not even

the simplest CGI scripts are possible, everything has to be hardcoded into physically existing HTML files.

Both sites are closely linked to each other, and we try to avoid duplicating information in as much as possible. In fact, <http://www.r-project.org/> and <http://cran.r-project.org/> are aliases for the same machine.

New CRAN mirrors

Laszlo Tornoci of the Semmelweis University Medical School in Budapest has set up a CRAN mirror for Hungary. Daniele Medri, with the Economics Faculty of University of Bologna, is working on setting up an Italian mirror.

Thus far, CRAN mirrors only exist in Europe and North America. In the interest of preserving bandwidth, we would like to add mirrors in other continents as well. Please contact wwwadmin@cran.r-project.org if you are interested in providing a new CRAN country mirror.

CRAN packages

CRAN contains R extension packages in four locations:

`'src/contrib'`: The main location containing packages that pass R CMD `check` at least on one platform¹.

`'src/contrib/Devel'`: Packages that are under development, incomplete, do not pass R CMD `check` or where the authors think they are not ready for the main section.

¹Debian GNU/Linux, the platform CRAN itself runs on and is used by the CRAN maintainers. The reason is a simple and practical one: before installing a package on CRAN we check it ...