

# Changes in R

by the R Core Team

## User-visible changes

- `plot(<lm object>)` uses a new default 'which = 5' for the fourth panel when 'which' is not specified.
- The SVN revision number will appear after the date in the welcome message. The date shown is now the date of the last change to the sources rather than the date the sources were prepared.
- `is.null(expression())` now returns FALSE. Only NULL gives TRUE in `is.null()`.
- `graphics::xy.coords`, `xyz.coords` and `n2mfrow` have been moved to the `grDevices` name space (to be available for grid as well).  
`graphics::boxplot.stats`, `contourLines`, `nclass.*`, and `chull` have been moved to the `grDevices` name space. The C code underlying `chull()` has been moved to package `grDevices`.
- `split(x, f)`, `split<-()` and `unsplit()` now by default split by all levels of a factor `f`, even when some are empty. Use `split(x, f, drop = TRUE)` if you want the old behavior of dropping empty levels. `split()` and `split<-()` are S3 generic functions with new arguments 'drop' and '...' and all methods now should have 'drop' and '...' arguments as well.
- The default for 'allowEscapes' in both `read.table()` and `scan()` has been changed to FALSE.
- The default for 'gcFirst' in `system.time()` is now TRUE.

## New features

- `.Platform` has a new component 'path.sep', the separator used between paths in environment variables such as `PATH` and `TEXINPUTS`.
- `anova.mlm()` now handles the single-model case.
- Hexadecimal values are now allowed for `as.numeric()` and `as.integer()` on all platforms, and as integer constants in R code.
- `attach()` now prints an information message when objects are masked on the search path by or from a newly attached database.
- `axis()` now returns 'at' positions.

- `axis()` has a new argument 'hadj' to control horizontal adjustment of labels.
- `axis.Date()` and `axis.POSIXct()` now accept a 'labels' argument (contributed by Gavin Simpson).
- `barplot()` now has arguments 'log = ""' and 'add = FALSE' (as in `barplot2()` from package 'gplots').
- `baseenv()` has been added, to return the base environment. This is currently still NULL, but will change in a future release.
- `boxplot()` now responds to supplying 'yaxs' (via `bxp()`). (Wish of PR#8072.)
- `capabilities()` has a new component 'NLS'.
- `cbind()` and `rbind()` now react to 'deparse.level' = 0,1,2 (as in another system not unlike R).
- Experimental versions of `cbind()` and `rbind()` in methods package, based on new generic function `cbind2(x,y)` and `rbind2()`. This will allow the equivalent of S4 methods for `cbind()` and `rbind()` — currently only after an explicit activation call, see `?cbind2`.
- New functions `cdplot()` and `spineplot()` for conditional density plots and spine plots or spinograms. Spine plots are now used instead of bar plots for x-y scatterplots where y is a factor.
- `checkDocFiles()` in package 'tools' now checks for bad \usage lines (syntactically invalid R code).
- The nonparametric variants of `cor.test()` now behave better in the presence of ties. The "spearman" method uses the asymptotic approximation in that case, and the "kendall" method likewise, but adds a correction for ties (this is not necessary in the Spearman case).
- The X11 `dataentry()` now has support for X Input Methods (contributed by Ei-ji Nakama).
- `density()` is now an S3 generic where `density.default()` former `density()` has new argument 'weights' for specifying observation masses different than the default 1/N — based on a suggestion and code from Adrian Baddeley.
- `download.packages()` now carries on if it encounters a download error (e.g. a repository with a corrupt index).

- `dump()` now skips missing objects with a warning rather than throw an error.
- Added "POSIXlt" methods for `duplicated()` and `unique()`.
- Function `encoded_text_to_latex()` in package tools translates Latin 1,2,9 and UTF-8 encoded character vectors to LaTeX escape sequences where appropriate.
- `encodeString()` allows `justify = "none"` for consistency with `format.default()`. Some argument names have been lengthened for clarity.
- `file()`, `fifo()` and `pipe()` now (if possible) report a reason if they fail to open a connection.

- `format.default()` now has a 'width' argument, and 'justify' can now centre character strings.

`format.default()` has new arguments 'na.encode' to control whether NA character strings are encoded (true by default), and 'scientific' to control the use of fixed/scientific notation for real/complex numbers.

How `format()` works on a list is now documented, and uses arguments consistently with their usage on an atomic vector.

- `format.info()` now has a 'digits' argument, and is documented to work for all atomic vectors (it used to work for all but raw vectors.).
- New function `glob2rx()` for translating 'wildcard' aka 'globbing' to regular expressions.
- There is a new function `gregexpr()` which generalizes `regexpr()` to search for all matches in each of the input strings (not just the first match).
- `[g]sub()` now have a 'useBytes' argument like `grep()` and `regexpr()`.
- `[g]sub(perl = TRUE)` support `\L` and `\U` in the replacement.
- `iconv()` has been moved from 'utils' to 'base'.
- `identify()`'s default method has additional arguments 'atpen' and 'tolerance' (following S).
- `KalmanForecast()` and `KalmanLike()` now have an optional argument `fast=FALSE` to prevent their arguments being modified.
- Exact p-values are available in `ks.test()` for the one-sided and two-sided one-sample Kolmogorov-Smirnov tests.

- `labels()` now has a method for "dist" objects (replacing that for `names()` which was withdrawn in 2.1.0).
- `library()` now explicitly checks for the existence of directories in 'lib.loc': this avoids some warning messages.
- `loadNamespace(keep.source=)` now applies only to that namespace and not others it might load to satisfy imports: this is now consistent with `library()`.
- `match.arg()` has a new argument 'several.ok = FALSE'.
- `max.col()` has a new argument for non-random behavior in the case of ties.
- `memory.profile()` now uses the type names returned by `typeof()` and no longer has two unlabelled entries.
- `methods()` now warns if it appears to have been called on a non-generic function.
- The default `mosaicplot()` method by default draws grey boxes.
- `n1minb()`, similar to that in S-PLUS, added to package 'stats'.
- New algorithm "port" (the `n12sol` algorithm available in the Port library on netlib) added to the `nls()` function in the 'stats' package.
- `object.size()` now supports more types, including external pointers and weak references.
- `options()` now returns its result in alphabetical order, and is documented more comprehensively and accurately. (Now all options used in base R are documented, including platform-specific ones.)  
Some options are now set in the package which makes use of them (`grDevices`, `stats` or `utils`) if not already set when the package is loaded.
- New option("OutDec") to set the decimal point for output conversions.
- New option("add.smooth") to add smoothers to a plot, currently only used by `plot.lm()`.
- `pie()` has new optional arguments 'clockwise' and 'init.angle'.
- `plot.lm()` has two new plots (for 'which' = 5 or 6), plotting residuals or cook distances versus (transformed) leverages - unless these are constant. Further, the new argument 'add.smooth' adds a loess smoother to the point plots by default, and 'qqline = TRUE' adds a `qqline()` to the normal plot. The default for 'sub.caption' has been improved for long calls.

- `R.home()` has been expanded to return the paths to components (which can as from this version be installed elsewhere).
- `readbin()` and `writeBin()` now support raw vectors as well as filenames and connections.
- `read.dcf()` can now read gzipped files.
- `read.table()` now passes `'allowEscapes'` to `scan()`.
- `sample(x, size, prob, replace = TRUE)` now uses a faster algorithm if there are many reasonably probable values. (This does mean the results will be different from earlier versions of R.) The speedup is modest unless `'x'` is very large and `'prob'` is very diffuse so that thousands of distinct values will be generated with an appreciable frequency.
- `scatter.smooth()` now works a bit more like other plotting functions (e.g., accepts a data frame for argument `'x'`). Improvements suggested by Kevin Wright.
- `signif()` on complex numbers now rounds jointly to give the requested number of digits in the larger component, not independently for each component.
- New generic function `simulate()` in the `'stats'` package with methods for some classes of fitted models.
- `smooth.spline()` has a new argument `'keep.data'` which allows to provide `residuals()` and `fitted()` methods for smoothing splines.
- Attempting `source(file, chdir=TRUE)` with a URL or connection for `'file'` now gives a warning and ignores `'chdir'`.
- `source()` closes its input file after parsing it rather than after executing the commands, as used to happen prior to 2.1.0. (This is probably only significant on Windows where the file is locked for a much shorter time.)
- `split()`, `split<-()`, `unsplit()` now have a new argument `'drop = FALSE'`, by default not dropping empty levels; this is *\*not\** back compatible.
- `sprintf()` now supports asterisk `'*'` width or precision specification (but not both) as well as `'*1$'` to `'*99$'`. Also the handling of `'%'` as conversion specification terminator is now left to the system and doesn't affect following specifications.
- The plot method for `stl()` now allows the colour of the range bars to be set (default unchanged at "light gray").
- Added `tclServiceMode()` function to the `tcltk` package to allow updating to be suspended.
- `terms.formula()` no longer allows `'.'` in a formula unless there is a (non-empty) `'data'` argument or `'allowDotAsName = TRUE'` is supplied. We have found several cases where `'data'` had not been passed down to `terms()` and so `'.'` was interpreted as a single variable leading to incorrect results.
- New functions `trans3d()`, the 3D -> 2D utility from `persp()`'s example, and `extendrange()`, both in package `'grDevices'`.
- `TukeyHSD()` now returns p-values adjusted for multiple comparisons (based on a patch contributed by Fernando Henrique Ferraz P. da Rosa).
- New functions `URLEncode()` and `URLdecode()`, particularly for use with `file://` URLs. These are used by e.g. `browse.env()`, `download.file()`, `download.packages()` and various `help()` print methods.
- Functions `utf8ToInt()` and `intToUtf8()` to work with UTF-8 encoded character strings (irrespective of locale or OS-level UTF-8 support).
- `[dqp]wilcox` and `wilcox.test` work better with one very large sample size and an extreme first argument.
- `write()` has a new argument `'sep'`.
- `write.csv[2]` now also support `'row.names = FALSE'`.
- The specification of the substitutions done when processing Renviron files is more liberal: see `?Startup`. It now accepts forms like `R_LIBS=${HOME}/Rlibrary:${WORKGRP}/R/lib`.
- Added recommendation that packages have an overview man page `<pkg>-package.Rd`, and the `promptPackage()` function to create a skeleton version.
- Replacement indexing of a data frame by a logical matrix index containing NAs is allowed in a few more cases, in particular always when the replacement value has length one.
- Conversion of `.Rd` files to  $\LaTeX$  now handles encoding more comprehensively, including some support for UTF-8.
- The internal regex code has been upgraded to `glibc-2.3.5`. Apart from a number of bug fixes, this should be somewhat faster, especially in UTF-8 locales.

- PCRE has been updated to version 6.2.
- zlib has been updated to version 1.2.3.
- bzip2 has been updated to version 1.0.3.
- Complex arithmetic is now done by C99 complex types where supported. This is likely to boost performance, but is subject to the accuracy with which it has been implemented.
- The printing of complex numbers has changed, handling numbers as a whole rather than in two parts. So both real and imaginary parts are shown to the same accuracy, with the 'digits' parameter referring to the accuracy of the larger component, and both components are shown in fixed or scientific notation (unless one is entirely zero when it is always shown in fixed notation).
- Error messages from `.C()` and `.Fortran()`, and from parsing errors, are now more informative.
- The date and date-time functions work better with dates more than 5000 years away from 1970-01-01 (by making dubious assumptions about the calendar in use).
- There is now a traditional Chinese translation, and a much more extensive Russian translation.

## Deprecated & defunct

- Capability "IEEE754" is defunct.
- `loadURL()` is defunct: use `load(url())`.
- `delay()` is defunct: use `delayedAssign()` instead.
- The 'CRAN' argument to `update.packages()`, `old.packages()`, `new.packages()`, `download.packages()` and `install.packages()` is defunct in favour of 'repos'.
- `write.table0()` is deprecated in favour of the much faster `write.table()`.
- `format.char()` is deprecated in favour of `format.default()`.
- `R_HOME/etc/Rprofile` is no longer looked for if `R_HOME/etc/Rprofile.site` does not exist. (This has been undocumented since R 1.4.0.)
- `CRAN.packages()` is deprecated in favour of `available.packages()`.
- `Rd.sty` no longer processes pre-2.0.0 conversions containing `\Link`.

- The stubs for the defunct device `GNOME/gnome` have been removed.
- `print.matrix()` (which has been identical to `print.default` since R 1.7.0) has been removed.

## Installation

- `LD_FLAGS` now defaults to `-L/usr/local/lib64` on most Linux 64-bit OSes (but not ia64). The use of `lib/lib64` can be overridden by the new variable `LIBnn`.
- The default installation directory is now `prefix/LIBnn/R`, `/usr/local/lib64/R` on most 64-bit Linux OSes and `/usr/local/lib/R` elsewhere.
- The places where the doc, include and share directory trees are installed can be specified independently: see the R-admin manual.
- We now test for `wctrans_t`, as apparently some broken OSes have `wctrans` but not `wctrans_t` (which is required by the relevant standards).
- Any external BLAS found is now tested to see if the complex routine `zdotu` works correctly: this provides a compatibility test of compiler return conventions.
- Installation without NLS is now cleaner, and does not install any message catalogues.
- `src/modules/lapack/dlanc.f` is now compiled with `-ffloat-store` if `f2c/gcc` are used, as well as if `g77` is used.
- All the Fortran code has been checked to be fully F77 compliant so there are no longer any warnings from F95 compilers such as `gfortran`.
- The (not-recommended) options `-with-system-zlib`, `-with-system-bzlib` and `-with-system-pcre` now have 'system' in the name.
- If a Java runtime environment is detected at configure time its library path is appended to `LD_LIBRARY_PATH` or equivalent. New Java-related variables `JAVA_HOME` (path to JRE/JDK), `JAVA_PROG` (path to Java interpreter), `JAVA_LD_PATH` (Java library path) and `JAVA_LIBS` (flags to link against JNI) are made available in `Makeconf`.
- Ei-ji Nakama was contributed a patch for FPU control with the Intel compilers on ix86 Linux.

## Mac OS X Installation

- `-with-blas="-framework vecLib"` `-with-lapack` and `-with-aqua` are now the default configure options.
- The default framework version name was changed to not contain the patch level (i.e. it is now 2.2 instead of 2.2.0). Also it can be overridden at configure time by setting `FW_VERSION` to the desired name.
- The Rmath stand-alone library is now correctly installed inside the R.framework if R was configured as a framework. In addition, `make install-Rmath-framework` will install a stand-alone Rmath framework in `/Library/Frameworks` (unless overridden by `RMATH_FRAMEWORK_DIR` specifying full framework path and name including the `.framework` extension).

## Package Installation

- The encoding for a packages' `00Index.html` is chosen from the `Encoding:` field (if any) of the `DESCRIPTION` file and from the `\encoding{}` fields of any Rd files with non-ASCII titles. If there are conflicts, first-found wins with a warning.
- `R_HOME/doc/html/packages.html` is now remade by R not Perl code. This may result in small changes in layout and a change in encoding (to UTF-8 where supported).
- The return value of `new.packages()` is now updated for any packages which may be installed.
- `available.packages()` will read a compressed `PACKAGES.gz` file in preference to `PACKAGES` if available on the repository: this will reduce considerably the download time on a dialup connection.  
The downloaded information about a repository is cached for the current R session.
- The information about library trees found by `installed.packages()` is cached for the current session, and updated only if the modification date of the top-level directory has been changed.
- A data index is now installed for a package with a `'data'` dir but no `'man'` dir (even though it will have undocumented data objects).
- `contrib.url` path for `type="mac.binary"` has changed from `bin/macosx/<version>` to `bin/macosx/<arch>/contrib/<version>` where `<arch>` corresponds to `R.version$arch`

## Utilities

- `checkFF()` used by R CMD check has since R 2.0.0 not reported missing `PACKAGE` arguments when testing installed packages with namespaces. It now
  - treats installed and source packages in the same way.
  - reports missing arguments unless they are in a function in the namespace with a `useDynLib` declaration (as the appropriate DLL for such calls can be searched for).
- `Rd2dvi` sets the `encoding(s)` used appropriately. If UTF-8 encoding is used, `latex >= 2003/12/01` is required.
- `codoc()` allows help files named `pkg_name-defunct.Rd` to have undocumented arguments (and not just `base-defunct.Rd`).

## C-level facilities

- C function `massdist()` called from `density()` has new argument `'xmass'` (= weights).
- Raw vectors passed to `.C()` are now passed as `unsigned char *` rather than as `SEXP`s. (Wish of Keith Frost, PR#7853)
- The search for symbols in a `.C/.Call/...` call without a package argument now searches for an enclosing namespace and so finds functions defined within functions in a namespace.
- `R_max_col()` has new (5th) argument `'*ties_meth'` allowing non-random behavior in the case of ties.
- The header files have been rationalized: the BLAS routine `LSAME` is now declared in `BLAS.h` not `Linpack.h`, `Applic.h` no longer duplicates routines from `Linpack.h`, and `Applic.h` is divided into API and non-API sections.
- `memory.c` has been instrumented so that Valgrind can track R's internal memory management. To use this, configure using `-with-valgrind-instrumentation=level` where level is 1 or 2. Both levels will find more bugs with `gctorture(TRUE)`. Level 2 makes Valgrind run extremely slowly.
- Some support for raw vectors has been added to `Rdefines.h`.
- `R_BaseEnv` has been added, to refer to the base environment. This is currently equal to `R_NilValue`, but it will change in a future release.

## Bug fixes

- `%/%` has been adjusted to make `x == (x %/ y) + y * (x %/ y)` more likely in cases when extended-precision registers were interfering.
- Operations on POSIXct objects (such as `seq()`, `max()` and subsetting) try harder to preserve time zones and warn if inconsistent time zones are used.
- `as.function.default()` no longer asks for a bug report when given an invalid body. (PR#1880, PR#7535, PR#7702)
- Hershey fonts and grid output (and therefore lattice output) now rescale correctly in fit-to-window resizing on a Windows graphics device. Line widths also scale now.
- Plotmath has more support for multibyte characters (contributed by Ei-ji Nakama).
- The `X11()` device now hints the window manager so that decorations appear reliably under e.g. the GNOME WM (contributed by Ei-ji Nakama).
- Subsetting a matrix or an array as a vector used to attempt to use the row names to name the result, even though the array might be longer than the row names. Now this is only done for 1D arrays when it is done in all cases, even matrix indexing. (Tidies up after the fix to PR#937.)
- Constants in `mathlib` are declared 'const static double' to avoid performance issues with the Intel Itanium compiler.
- The parser checks the format of numeric constants more thoroughly so for example '123E-' is no longer valid.
- `contourLines()` no longer requires an open device (used to start a device unnecessarily). Fix suggested by Barry Rowlingson.
- `capabilities()` used partial matching but was not documented to: it no longer does so.
- `kernel(1,0)` printed wrongly; `kernel(<name-string>, *)` now returns a named kernel in all cases; `plot(kernel(.),..)` is more flexible.
- `qgamma(1,s)` didn't give +Inf for some s.
- `installed.packages()` and `download.packages()` now always return a matrix as documented, possibly with 0 rows (rather than a 0-length character vector or NULL).
- Arithmetic operations on data frames no longer coerce the names to syntactically valid names.
- Units are now properly recycled in grid layouts when 'widths' or 'heights' are shorter than the number of columns or rows (PR#8014).
- `DF <- data.frame(A=1:2, B=3:4); DF[1, 1:3] <- NULL` gave a wrong error message.
- `spline()/spinefun()`'s C code had a memory access buglet which never lead to incorrect results. (PR#8030)
- `sum()` was promoting logical arguments to double not integer (as `min()` and other members of its group do).
- `loess()` had a bug causing it to occasionally miscalculate standard errors (PR#7956). Reported by Benjamin Tyner, fixed by Berwin Turlach.
- `library(keep.source=)` was ignored if the package had a namespace (the setting of `options("keep.source.pkgs")` was always used).
- `hist.POSIXct()` and `hist.Date()` now respect `par("xaxt")`.
- The 'vfont' argument was not supported correctly in `title()`, `mtext()`, and `axis()`. The 'vfont' argument is superseded by the `par(family=)` approach introduced in 2.0.0. This bug-fix just updates the warning messages and documentation to properly reflect the new order of things.
- The C-level function `PrintGenericVector` could overflow if asked to print a length-1 character vector of several thousand characters. This could happen when printing a list matrix, and was fatal up to 2.1.1 and silently truncated in 2.1.1 patched.
- What happened for `proc.time()` and `system.time()` on (Unix-alike) systems which do not support timing was incorrectly documented. (They both exist but throw an error.) Further, `system.time()` would give an error in its `on.exit` expression.
- `weighted.residuals()` now does sensible things for `glm()` fits: in particular it now agrees with an `lm()` fit for a Gaussian `glm()` fit. (PR#7961).
- The 'lm' and 'glm' methods for `add1()` took the weights and offset from the original fit, and so gave errors in the (dubious) usage where the upper scope resulted in a smaller number of cases to fit (e.g. by omitting missing values in new variables). (PR#8049)

- `demo()` had a 'device' argument that did nothing (although it was documented to): it has been removed.
- Setting new levels on a factor dropped all existing attributes, including class "ordered".
- `format.default(justify="none")` now by default converts NA character strings, as the other values always did.
- `format.info()` often gave a different field width from `format()` for character vectors (e.g. including missing values or non-printable characters).
- `axis()` now ensures that if 'labels' are supplied as character strings or expressions then 'at' is also supplied (since the calculated value for 'at' can change under resizing).
- Defining S4 methods for "[" had resulted in changed behavior of S3 dispatch in a very rare case which no longer happens.
- Fixed segfault when PostScript font loading fails, e.g., when R is unable to find afm files (reported by Ivo Welch).
- R CMD BATCH <file> now also works when <file> does not end in a newline on Unix-alike platforms.
- `terms.formula()` got confused if the 'data' argument was a list with non-syntactic names.
- `prompt()` and hence `package.skeleton()` now produce \*.Rd files that give no errors (but warnings) when not edited, much more often.
- `promptClass()` and `promptMethods()` now also escape "%" e.g. in "%\*%" and the latter gives a message about the file written.
- `wilcox.test()` now warns when `conf.level` is set higher than achievable, preventing errors (PR#3666) and incorrect answers with extremely small sample sizes.
- The default (protection pointer) stack size (the default for '-max-ppsize') has been increased from 10000 to 50000 in order to match the increased default options("expressions") (in R 2.1.0).
- The R front-end was expecting `-gui=tk` not `Tk` as documented, and rejecting `-gui=X11`.
- `Rdconv -t latex` protected only the first « and » in a chunk against conversion to guillemets.
- `callNextMethod()` and `callGeneric()` have fixes related to handling arguments.
- `ls.diag()` now works for fits with missing data. (PR#8139)
- `window.default()` had an incorrect tolerance and so sometimes created too short a series if 'start' or 'end' were zero.
- Some (fairly pointless) cases of reshape left a temporary id variable in the result (PR#8152)
- R CMD build used 'tar xhf' which is invalid on FreeBSD systems (and followed tar chf, so there could be no symbolic links in the tarball).
- Subassignment of length zero vectors to NULL gave garbage answers. (PR#8157)
- Automatic coercion of raw vectors to lists was missing, so for a list (or data frame) `z, z[["a"]] <- raw_vector` did not work and now does. This also affected `DF$a <- raw_vector` for a data frame DF.
- The internal code for `commandArgs()` was missing PROTECTs.
- The width for `strwrap()` was used as one less than specified.
- R CMD INSTALL was not cleaning up after an unsuccessful install of a non-bundle which was not already installed.

## Changes on CRAN

by Kurt Hornik

### New contributed packages

**ArDec** Implements the autoregressive decomposition of a time series based on the constructive approach in West (1997). Particular cases include the extraction of trend and seasonal components from a monthly time series. Un-

certainty on the resulting components can be derived from sampling of the autoregressive model which is written as a linear regression model and handled on a Bayesian framework. By S. M. Barbosa.

**BACCO** Bayesian analysis of computer code software. The bundle contains routines that evaluate the formulae of Kennedy, O'Hagan, and Oakley. By Robin K. S. Hankin.