# NEWS for R version 3.0.2 (2013-09-25)

NEWS R News



## CHANGES IN R 3.0.2

## **NEW FEATURES:**

- The 'NEWS' files have been re-organized.

  This file contains news for R >= 3.0.0: news for the 0.x.y, 1.x.y and 2.x.y releases is in files 'NEWS.0', 'NEWS.1' and 'NEWS.2'. The latter files are now installed when R is installed. An HTML version of news from 2.10.0 to 2.15.3 is available as 'doc/html/NEWS.2.html'.
- sum() for integer arguments now uses an integer accumulator of at least 64 bits and so will be more accurate in the very rare case that a cumulative sum exceeds 2<sup>53</sup> (necessarily summing more than 4 million elements).
- The example() and tools::Rd2ex() functions now have parameters to allow them to ignore \dontrun markup in examples. (Suggested by Peter Solymos.)
- str(x) is considerably faster for very large lists, or factors with 100,000 levels, the latter as in PR#15337.
- col2rgb() now converts factors to character strings not integer codes (suggested by Bryan Hanson).
- tail(warnings()) now works, via the new `[` method.
- There is now support for the LaTeX style file 'zi4.sty' which has in some distributions replaced 'inconsolata.sty'.
- unlist(x) now typically returns all non-list xs unchanged, not just the "vector" ones. Consequently, format(lst) now also works when the list lst has non-vector elements.
- The tools::getVignetteInfo() function has been added to give information about installed vignettes.
- New assertCondition(), etc. utilities in tools, useful for testing.
- Profiling now records non-inlined calls from byte-compiled code to BUILTIN functions.
- Various functions in **stats** and elsewhere that use non-standard evaluation are now more careful to follow the namespace scoping rules. E.g. **stats::lm()** can now find **stats::model.frame()** even if **stats** is not on the search path or if some package defines a function of that name.
- If an invalid/corrupt .Random.seed object is encountered in the workspace it is ignored with a warning rather than giving an error. (This allows R itself to rely on a working RNG, e.g. to choose a random port.)

• seq() and seq.int() give more explicit error messages if called with invalid (e.g. NaN) inputs.

- When parse() finds a syntax error, it now makes partial parse information available up to the location of the error. (Request of Reijo Sund.)
- Methods invoked by NextMethod() had a different dynamic parent to the generic. This was causing trouble where S3 methods invoked via lazy evaluation could lose track of their generic. (PR#15267)
- Code for the negative binomial distribution now treats the case size == 0 as a one-point distribution at zero.
- abbreviate() handles without warning non-ASCII input strings which require no abbreviation.
- read.dcf() no longer has a limit of 8191 bytes per line. (Wish of PR#15250.)
- formatC(x) no longer copies the class of x to the result, to avoid misuse creating invalid objects as in PR#15303. A warning is given if a class is discarded.
- Dataset npk has been copied from MASS to allow more tests to be run without recommended packages being installed.
- The initialization of the regression coefficients for non-degenerate differenced models in arima() has been changed and in some examples avoids a local maximum. (PR#15396)
- termplot() now has an argument transform.x to control the display of individual terms in the plot. (PR#15329)
- format() now supports digits = 0, to display nsmall decimal places.
- There is a new read-only par() parameter called "page", which returns a logical value indicating whether the next plot.new() call will start a new page.
- Processing Sweave and Rd documents to PDF now renders backticks and single quotes better in several instances, including in '\code' and '\samp' expressions.
- utils::modifyList() gets a new argument keep.null allowing NULL components in the replacement to be retained, instead of causing corresponding components to be deleted.
- tools::pkgVignettes() gains argument check; if set to TRUE, it will warn when it appears a vignette requests a non-existent vignette engine.

# **UTILITIES:**

- R CMD check --as-cran checks the line widths in usage and examples sections of the package Rd files.
- R CMD check --as-cran now implies '--timings'.
- R CMD check looks for command gfile if a suitable file is not found. (Although file is not from GNU, OpenCSW on Solaris installs it as gfile.)
- R CMD build (with the internal tar) checks the permissions of 'configure' and 'cleanup' files and adds execute permission to the recorded permissions for these files if needed, with a warning. This is useful on OSes and file systems which do not support execute permissions (notably, on Windows).
- R CMD build now weaves and tangles all vignettes, so suggested packages are not required during package installation if the source tarball was prepared with current R CMD build.
- checkFF() (used by R CMD check) does a better job of detecting calls from other packages, including not reporting those where a function has been copied from another namespace (e.g. as a default method). It now reports calls where .NAME is a symbol registered in another package.

• On Unix-alike systems, R CMD INSTALL now installs packages group writably whenever the library (lib.loc) is group writable. Hence, update.packages() works for other group members (suggested originally and from a patch by Dirk Eddelbuettel).

- R CMD javareconf now supports the use of symbolic links for JAVA\_HOME on platforms which have realpath. So it is now possible to use
  - R CMD javareconf JAVA\_HOME=/usr/lib/jvm/java-1.7.0
  - on a Linux system and record that value rather than the frequently-changing full path such as '/usr/lib/jvm/java-1.7.0-openjdk-1.7.0.25.x86\_64'.
- (Windows only.) Rscript -e requires a non-empty argument for consistency with Unix versions of R. (Also Rterm -e and R -e.)
- $\bullet$  R CMD check does more thorough checking of declared packages and namespaces. It reports
  - packages declared in more than one of the 'Depends', 'Imports', 'Suggests' and 'Enhances' fields of the 'DESCRIPTION' file.
  - namespaces declared in 'Imports' but not imported from, neither in the 'NAMESPACE' file nor using the :: nor ::: operators.
  - packages which are used in library() or requires() calls in the R code but were already put on the search path via 'Depends'.
  - packages declared in 'Depends' not imported via the 'NAMESPACE' file (except the standard packages). Objects used from 'Depends' packages should be imported to avoid conflicts and to allow correct operation when the namespace is loaded but not attached.
  - objects imported *via* ::: calls where :: would do.
  - objects imported by :: which are not exported.
  - objects imported by ::: calls which do not exist.

See 'Writing R Extensions' for good practice.

- R CMD check optionally checks for non-standard top-level files and directories (which are often mistakes): this is enabled for '--as-cran'.
- LaTeX style file upquote.sty is no longer included (the version was several years old): it is no longer used in R. A much later version is commonly included in LaTeX distributions but does not play well with the ae fonts which are the default for Sweave vignettes.
- R CMD build makes more use of the 'build' sub-directory of package sources, for example to record information about the vignettes.

#### INSTALLATION and INCLUDED SOFTWARE:

- The macros used for the texinfo manuals have been changed to work better with the incompatible changes made in texinfo 5.x.
- The minimum version for a system xz library is now 5.0.3 (was 4.999). This is in part to avoid 5.0.2, which can compress in ways other versions cannot decompress.
- The included version of PCRE has been updated to 8.33.
- The included version of zlib has been updated to 1.2.8, a bug-fix release.
- The included version of xz utils's liblzma has been updated to 5.0.5.
- Since javareconf (see above) is used when R is installed, a stable link for JAVA\_HOME can be supplied then.
- Configuring with '--disable-byte-compilation' will override the 'DESCRIPTION' files of recommended packages, which typically require byte-compilation.
- More of the installation and checking process will work even when TMPDIR is set to a path containing spaces, but this is not recommended and external software (such as texi2dvi) may fail.

#### PACKAGE INSTALLATION:

- Installation is aborted immediately if a LinkingTo package is not installed.
- R CMD INSTALL has a new option --no-byte-compile which will override a 'ByteCompile' field in the package's 'DESCRIPTION' file.
- License 'BSD' is deprecated: use 'BSD\_3\_clause' or 'BSD\_2\_clause' instead. License 'X11' is deprecated: use 'MIT' or 'BSD\_2\_clause' instead.
- Version requirements for LinkingTo packages are now recognized: they are checked at installation. (Fields with version requirements were previously silently ignored.)
- The limit of 500 S3method entries in a NAMESPACE file has been removed.
- The default 'version' of Bioconductor for its packages has been changed to the upcoming '2.13', but this can be set by the environment variable R\_BIOC\_VERSION, e.g. in file 'Renviron.site'.

#### **C-LEVEL FACILITIES:**

- 'Rdefines.h' has been tweaked so it can be included in C++ code after 'R\_ext/Boolean.h' (which is included by 'R.h').

  Note that 'Rdefines.h' is not kept up-to-date, and 'Rinternals.h' is preferred for new code.
- eval and applyClosure are now protected against package code supplying an invalid rho.

#### DEPRECATED AND DEFUNCT:

- The unused namespace argument to package.skeleton() is now formally deprecated and will be removed in R 3.1.0.
- plclust() is deprecated: use the plot() method for class "hclust" instead.
- Functions readNEWS() and checkNEWS() in package tools are deprecated (and they have not worked with current 'NEWS' files for a long time).

## DOCUMENTATION:

• 'An Introduction to R' has a new chapter on using R as a scripting language including interacting with the OS.

## **BUG FIXES:**

- help.request() could not determine the current version of R on CRAN. (PR#15241)
- On Windows, file.info() failed on root directories unless the path was terminated with an explicit ".". (PR#15302)
- The regmatches<-() replacement function mishandled results coming from regexpr(). (PR#15311)
- The help for setClass() and representation() still suggested the deprecated argument representation=. (PR#15312)
- R CMD config failed in an installed build of R 3.0.1 (only) when a sub-architecture was used. (Reported by Berwin Turlach.)
- On Windows, the installer modified the 'etc/Rconsole' and 'etc/Rprofile.site' files even when default options were chosen, so the MD5 sums did not refer to the installed versions. (Reported by Tal Galili.)
- plot(hclust(), cex =) respects cex again (and possibly others similarly). (Reported by Peter Langfelder.)
- If multiple packages were checked by R CMD check, and one was written for a different OS, it would set --no-install for all following packages as well as itself.

 qr.coef() and related functions did not properly coerce real vectors to complex when necessary. (PR#15332)

- ftable(a) now fixes up empty dimnames such that the result is printable.
- package.skeleton() was not starting its search for function objects in the correct place if environment was supplied. (Reported by Karl Forner.)
- Parsing code was changing the length field of vectors and confusing the memory manager. (PR#15345)
- The Fortran routine ZHER2K in the reference BLAS had a comment-out bug in two places. This caused trouble with eigen() for Hermitian matrices. (PR#15345 and report from Robin Hankin)
- vignette() and browseVignettes() did not display non-Sweave vignettes properly.
- Two warning/error messages have been corrected: the (optional) warning produced by a partial name match with a pairlist, the error message from a zero-length argument to the : operator. (Found by Radford Neal; PR#15358, PR#15356)
- $\bullet$  svd() returned NULL rather than omitting components as documented. (Found by Radford Neal; PR#15360)
- mclapply() and mcparallel() with silent = TRUE could break a process that uses stdout output unguarded against broken pipes (e.g., zip will fail silently). To work around such issues, they now replace stdout with a descriptor pointed to '/dev/null' instead. For this purpose, internal closeStdout and closeStderr functions have gained the to.null flag.
- log(), signif() and round() now raise an error if a single named argument is not named x. (PR#15361)
- deparse() now deparses raw vectors in a form that is syntactically correct. (PR#15369)
- The jpeg driver in Sweave created a JPEG file, but gave it a '.png' extension. (PR#15370)
- Deparsing of infix operators with named arguments is improved. (PR#15350)
- mget(), seq.int() and numericDeriv() did not duplicate arguments properly. (PR#15352, PR#15353, PR#15354)
- kmeans(algorithm = "Hartigan-Wong") now always stops iterating in the QTran stage. (PR#15364).
- read.dcf() re-allocated incorrectly and so could segfault when called on a file with lines of more than 100 bytes.
- On systems where mktime() does not set errno, the last second before the epoch could not be converted from POSIX1t to POSIXct. (Reported by Bill Dunlap.)
- add1.glm() miscalculated F-statistics when df > 1. (Bill Dunlap, PR#15386).
- stem() now discards infinite inputs rather than hanging. (PR#15376)
- The parser now enforces C99 syntax for floating point hexadecimal constants (e.g. 0x1.1p0), rather than returning unintended values for malformed constants. (PR#15234)
- model.matrix() now works with very long LHS names (more than 500 bytes). (PR#15377)
- integrate() reverts to the pre-2.12.0 behaviour: from 2.12.0 to 3.0.1 it sometimes failed to achieve the requested tolerance and reported error estimates that were exceeded. (PR#15219)
- strptime() now handles '%W' fields with value 0. (PR#15915)
- R is now better protected against people trying to interact with the console in startup code. (PR#15325)

- Subsetting 1D arrays often lost dimnames (PR#15301).
- Unary + on a logical vector did not coerce to integer, although unary did.
- na.omit() and na.exclude() added a row to a zero-row data frame. (PR#15399)
- All the (where necessary cut-down) vignettes are installed if R was configured with '--without-recommended-packages'.
- source() did not display filenames when reporting syntax errors.
- Syntax error reports misplaced the caret pointing out the bad token.
- (Windows only) Starting R with R (instead of Rterm or Rgui) would lose any zero-length strings from the command line arguments. (PR#15406)
- Errors in the encoding specified on the command line via --encoding=foo were not handled properly. (PR#15405)
- If x is a symbol, is.vector(x, "name") now returns TRUE, since "name" and "symbol" should be synonyms. (Reported by Hervé Pagès.)
- R CMD rtags works on platforms (such as OS X) with a XSI-conformant shell command echo. (PR#15231)
- is.unsorted(NA) returns false as documented (rather than NA).
- R CMD LINK did not know about sub-architectures.
- system() and system2() are better protected against users who misguidedly have spaces in the temporary directory path.
- file.show() and edit() are now more likely to work on file paths containing spaces. (Where external utilities are used, not the norm on Windows nor in R.app which should previously have worked.)
- Packages using the **methods** package are more likely to work when they import it but it is not attached. (Several parts of its C code were looking for its R functions on the search path rather than in its namespace.)
- lgamma(-x) is no longer NaN for very small x.
- (Windows) system2() now respects specifying stdout and stderr as files if called from Rgui. (PR#15393)
- Closing an x11() device whilst locator() or identify() is in progress no longer hangs R. (PR#15253)
- list.dirs(full.names = FALSE) was not implemented. (PR#15170)
- format() sometimes added unnecessary spaces. (PR#15411)
- all.equal(check.names = FALSE) would ignore the request to ignore the names and would check them as attributes.
- The symbol set by tools::Rd2txt\_options(itemBullet=) was not respected in some locales. (PR#15435)
- mcMap() was not exported by package parallel. (PR#15439)
- plot() for TukeyHSD objects did not balance dev.hold() and dev.flush() calls on multi-page plots. (PR#15449)



# CHANGES IN R 3.0.1

## **NEW FEATURES:**

- chooseCRANmirror() and chooseBioCmirror() gain an ind argument (like setRepositories()).
- mcparallel has a new argument mc.interactive which can modify the interactive flag in the child process. The new default is FALSE which makes child processes non-interactive by default (this prevents lock-ups due to children waiting for interactive input).

- scan() now warns when end-of-file occurs within a quoted string.
- count.fields() is now consistent with scan() in its handling of newlines in quoted strings. Instead of triggering an error, this results in the current line receiving NA as the field count, with the next line getting the total count of the two lines.
- The default method of image() will plot axes of the class of xlim and ylim (and hence of x and y if there is a suitable range() method). Based on a suggestion of Michael Sumner.
- load() now has a verbose argument for debugging support, to print the names of objects just before loading them.
- When loading a serialized object encounters a reference to a namespace which cannot be loaded, this is replaced by a reference to the global environment, with a warning.
- pairs() gains a line.main option for title placement.
- The remaining instances in which serialization to a raw vector was limited to 2GB have been unlimited on a 64-bit platform, and in most cases serialization to a vector of more than 1GB will be substantially faster.

#### **UTILITIES:**

• R CMD config now make use of personal 'Makevars' files under '~/.R' and a site file 'Makevars.site', in the same way as R CMD SHLIB and R CMD INSTALL. This makes the utility more useful in package configure scripts.

On Windows finding the personal files may require the environment variable HOME set. The old behaviour can be obtained with the new options '--no-user-files' and '--no-site-files'.

### PACKAGE INSTALLATION:

• Alternatives to the site and user customization files 'Makevars.site' and '~/.R/Makevars' can be specified *via* the environment variables R\_MAKEVARS\_SITE and R\_MAKEVARS\_USER respectively. These can be used to suppress the use of the default files by setting an empty value (where possible) or a non-existent path.

## **BUG FIXES:**

- sys.source() did not report error locations when keep.source = TRUE.
- as.POSIXct.numeric was coercing origin using the tz argument and not "GMT" as documented (PR#14973).
- The active binding to assign fields in reference classes has been cleaned up to reduce dependence on the class' package environment, also fixing bug in initializing read-only fields (inspired by a report from Hadley Wickham).
- str(d) no longer gives an error when names(d) contain illegal multibyte strings (PR#15247).
- Profiling of built-in functions with line.profiling= TRUE did not record the line from which they were called.
- citation(pkg) dropped the header and footer specified in the 'CITATION' file (PR#15257).
- Quotes were handled differently when reading the first line and reading the rest, so read.table() misread some files that contained quote characters (PR#15245).
- cat() with sep a character vector of length greater than one and more than one argument was using separators inconsistently (PR#15261).
- On Windows in R 3.0.0, savePlot() failed because of an incorrect check on the argument count.

• unzip(list = TRUE) returned Names as a factor and not a character vector (as documented) for the internal method. (Noticed by Sean O'Riordain.)

- contourLines() now checks more comprehensively for conformance of its x, y and z arguments (it was used incorrectly in package R2G2).
- Saved graphics display lists are R version-specific. Attempting to load workspaces containing them (or some other version-specific objects) aborted the load in R 3.0.0 and earlier; now it does a partial load and generates a warning instead.
- In R 3.0.0, identify() and locator() did not record information correctly, so replaying a graph (e.g. by copying it to another device) would fail. (PR#15271)
- Calling file.copy() or dirname() with the invalid input "" (which was being used in packages, despite not being a file path) could have caused a segfault. dirname("") is now "" rather than "." (unless it segfaulted).
- supsmu() could read/write outside its input vectors for very short inputs (seen in package rms for n = 4).
- as.dendrogram()'s hclust method uses less memory and hence gets considerably faster for large (n ~ 1000) clusterings, thanks to Daniel Müllner. (PR#15174)
- The return value when all workers failed from parallel::mclapply(mc.preschedule = TRUE) was a list of strings and not of error objects. (Spotted by Karl Forner and Bernd Bischl.)
- In R 3.0.0, when help() found multiple pages with the same alias, the HTML display of all the selections was not produced. (PR#15282)
- splinefun(method="monoH.FC") now produces a function with first argument named x and allows deriv=3, as documented. (PR#15273)
- summaryRprof() would only read the first chunksize lines of an Rprof file produced with line.profiling=TRUE. By default, this is the first 100 seconds. (PR#15288)
- lsfit() produced an incorrect error message when argument x had more columns than rows or x had a different number of rows than y. (Spotted by Renaud Gaujoux.)
- Binary operations on equal length vectors copied the class name from the second operand when the first had no class name, but did not set the object bit. (PR#15299)
- The trace() method for reference generator objects failed after those objects became function definitions.
- write.table() did not check that factors were constructed correctly, and so caused a segment fault when writing bad ones. (PR#15300)
- The internal HTTP server no longer chokes on POST requests without body. It will also pass-through other request types for custom handlers (with the method stored in Request-Method header) instead of failing.



## CHANGES IN R 3.0.0

## SIGNIFICANT USER-VISIBLE CHANGES:

- Packages need to be (re-)installed under this version (3.0.0) of R.
- There is a subtle change in behaviour for numeric index values  $2^{31}$  and larger. These never used to be legitimate and so were treated as NA, sometimes with a warning. They are now legal for long vectors so there is no longer a warning, and x[2^31] <- y will now extend the vector on a 64-bit platform and give an error on a 32-bit one.
- It is now possible for 64-bit builds to allocate amounts of memory limited only by the OS. It may be wise to use OS facilities (e.g. ulimit in a bash shell, limit in csh), to set limits on overall memory consumption of an R process, particularly in a

multi-user environment. A number of packages need a limit of at least 4GB of virtual memory to load.

64-bit Windows builds of R are by default limited in memory usage to the amount of RAM installed: this limit can be changed by command-line option '--max-mem-size' or setting environment variable R\_MAX\_MEM\_SIZE.

 Negative numbers for colours are consistently an error: previously they were sometimes taken as transparent, sometimes mapped into the current palette and sometimes an error.

#### **NEW FEATURES:**

- identical() has a new argument, ignore.environment, used when comparing functions (with default FALSE as before).
- There is a new option, options(CBoundsCheck=), which controls how .C() and .Fortran() pass arguments to compiled code. If true (which can be enabled by setting the environment variable R\_C\_BOUNDS\_CHECK to 'yes'), raw, integer, double and complex arguments are always copied, and checked for writing off either end of the array on return from the compiled code (when a second copy is made). This also checks individual elements of character vectors passed to .C().
  - This is not intended for routine use, but can be very helpful in finding segfaults in package code.
- In layout(), the limits on the grid size have been raised (again).
- New simple provideDimnames() utility function.
- Where methods for length() return a double value which is representable as an integer (as often happens for package Matrix), this is converted to an integer.
- Matrix indexing of dataframes by two-column numeric indices is now supported for replacement as well as extraction.
- setNames() now has a default for its object argument, useful for a character result.
- StructTS() has a revised additive constant in the loglik component of the result: the previous definition is returned as the loglik0 component. However, the help page has always warned of a lack of comparability of log-likelihoods for non-stationary models. (Suggested by Jouni Helske.)
- The logic in aggregate.formula() has been revised. It is now possible to use a formula stored in a variable; previously, it had to be given explicitly in the function call.
- install.packages() has a new argument quiet to reduce the amount of output shown.
- Setting an element of the graphics argument lwd to a negative or infinite value is now an error. Lines corresponding to elements with values NA or NaN are silently omitted. Previously the behaviour was device-dependent.
- Setting graphical parameters cex, col, lty, lwd and pch in par() now requires a length-one argument. Previously some silently took the first element of a longer vector, but not always when documented to do so.
- Sys.which() when used with inputs which would be unsafe in a shell (e.g. absolute paths containing spaces) now uses appropriate quoting.
- as.tclObj() has been extended to handle raw vectors. Previously, it only worked in the other direction. (Contributed by Charlie Friedemann, PR#14939.)
- New functions cite() and citeNatbib() have been added, to allow generation of in-text citations from "bibentry" objects. A cite() function may be added to bibstyle() environments.
- A sort() method has been added for "bibentry" objects.

• The bibstyle() function now defaults to setting the default bibliography style. The getBibstyle() function has been added to report the name of the current default style.

- scatter.smooth() now has an argument lpars to pass arguments to lines().
- pairs() has a new log argument, to allow some or all variables to be plotted on logarithmic scale. (In part, wish of PR#14919.)
- split() gains a sep argument.
- termplot() does a better job when given a model with interactions (and no longer attempts to plot interaction terms).
- The parser now incorporates code from Romain Francois' **parser** package, to support more detailed computation on the code, such as syntax highlighting, comment-based documentation, etc. Functions getParseData() and getParseText() access the data.
- There is a new function rep\_len() analogous to rep.int() for when speed is required (and names are not).
- The undocumented use rep(NULL, length.out = n) for n > 0 (which returns NULL) now gives a warning.
- demo() gains an encoding argument for those packages with non-ASCII demos: it defaults to the package encoding where there is one.
- strwrap() converts inputs with a marked encoding to the current locale: previously it made some attempt to pass through as bytes inputs invalid in the current locale.
- Specifying both rate and scale to [dpqr]gamma is a warning (if they are essentially the same value) or an error.
- merge() works in more cases where the data frames include matrices. (Wish of PR#14974.)
- optimize() and uniroot() no longer use a shared parameter object across calls. (nlm(), nlminb() and optim() with numerical derivatives still do, as documented.)
- The all.equal() method for date-times is now documented: times are regarded as equal (by default) if they differ by up to 1 msec.
- duplicated() and unique() gain a nmax argument which can be used to make them much more efficient when it is known that there are only a small number of unique entries. This is done automatically for factors.
- Functions rbinom(), rgeom(), rhyper(), rpois(), rnbinom(), rsignrank() and rwilcox() now return integer (not double) vectors. This halves the storage requirements for large simulations.
- sort(), sort.int() and sort.list() now use radix sorting for factors of less than 100,000 levels when method is not supplied. So does order() if called with a single factor, unless na.last = NA.
- diag() as used to generate a diagonal matrix has been re-written in C for speed and less memory usage. It now forces the result to be numeric in the case diag(x) since it is said to have 'zero off-diagonal entries'.
- backsolve() (and forwardsolve()) are now internal functions, for speed and support for large matrices.
- More matrix algebra functions (e.g. chol() and solve()) accept logical matrices (and coerce to numeric).
- sample.int() has some support for n ≥ 2<sup>31</sup>: see its help for the limitations.
   A different algorithm is used for (n, size, replace = FALSE,prob = NULL) for n > 1e7 and size <= n/2. This is much faster and uses less memory, but does give different results.</li>

• approxfun() and splinefun() now return a wrapper to an internal function in the stats namespace rather than a .C() or .Call() call. This is more likely to work if the function is saved and used in a different session.

- The functions .C(), .Call(), .External() and .Fortran() now give an error (rather than a warning) if called with a named first argument.
- Sweave() by default now reports the locations in the source file(s) of each chunk.
- clearPushBack() is now a documented interface to a long-existing internal call.
- aspell() gains filters for R code, Debian Control Format and message catalog files, and support for R level dictionaries. In addition, package utils now provides functions aspell\_package\_R\_files() and aspell\_package\_C\_files() for spell checking R and C level message strings in packages.
- bibentry() gains some support for "incomplete" entries with a 'crossref' field.
- gray() and gray.colors() finally allow alpha to be specified.
- monthplot() gains parameters to control the look of the reference lines. (Suggestion of Ian McLeod.)
- Added support for new %~% relation ("is distributed as") in plotmath.
- domain = NA is accepted by gettext() and ngettext(), analogously to stop() etc.
- termplot() gains a new argument plot = FALSE which returns information to allow the plots to be modified for use as part of other plots, but does not plot them. (Contributed by Terry Therneau, PR#15076.)
- quartz.save(), formerly an undocumented part of R.app, is now available to copy a device to a quartz() device. dev.copy2pdf() optionally does this for PDF output: quartz.save() defaults to PNG.
- The default method of pairs() now allows text.panel = NULL and the use of <foo>.panel = NULL is now documented.
- setRefClass() and getRefClass() now return class generator functions, similar to setClass(), but still with the reference fields and methods as before (suggestion of Romain Francois).
- New functions bitwNot(), bitwAnd(), bitwOr() and bitwXor(), using the internal interfaces previously used for classes "octmode" and "hexmode".

  Also bitwShiftL() and bitwShiftR() for shifting bits in elements of integer vectors.
- New option "deparse.cutoff" to control the deparsing of language objects such as calls and formulae when printing. (Suggested by a comment of Sarah Goslee.)
- colors() gains an argument distinct.
- New demo(colors) and demo(hclColors), with utility functions.
- list.files() (aka dir()) gains a new optional argument no.. which allows to exclude "." and ".." from listings.
- Multiple time series are also of class "matrix"; consequently, head(), e.g., is more useful.
- encodeString() preserves UTF-8 marked encodings. Thus if factor levels are marked as UTF-8 an attempt is made to print them in UTF-8 in RGui on Windows.
- readLines() and scan() (and hence read.table()) in a UTF-8 locale now discard a UTF-8 byte-order-mark (BOM). Such BOMs are allowed but not recommended by the Unicode Standard: however Microsoft applications can produce them and so they are sometimes found on websites.
  - The encoding name "UTF-8-BOM" for a connection will ensure that a UTF-8 BOM is discarded.
- mapply(FUN, a1, ...) now also works when a1 (or a further such argument) needs a length() method (which the documented arguments never do). (Requested by Hervé Pagès; with a patch.)

• .onDetach() is supported as an alternative to .Last.lib. Unlike .Last.lib, this does not need to be exported from the package's namespace.

- The srcfile argument to parse() may now be a character string, to be used in error messages.
- The format() method for ftable objects gains a method argument, propagated to write.ftable() and print(), allowing more compact output, notably for LaTeX formatting, thanks to Marius Hofert.
- The utils::process.events() function has been added to trigger immediate event handling.
- Sys.which() now returns NA (not "") for NA inputs (related to PR#15147).
- The print() method for class "htest" gives fewer trailing spaces (wish of PR#15124).
  - Also print output from HoltWinters(), nls() and others.
- loadNamespace() allows a version specification to be given, and this is used to check version specifications given in the 'Imports' field when a namespace is loaded.
- setClass() has a new argument, slots, clearer and less ambiguous than representation. It is recommended for future code, but should be back-compatible. At the same time, the allowed slot specification is slightly more general. See the documentation for details.
- mget() now has a default for envir (the frame from which it is called), for consistency with get() and assign().
- close() now returns an integer status where available, invisibly. (Wish of PR#15088.)
- The internal method of tar() can now store paths too long for the 'ustar' format, using the (widely supported) GNU extension. It can also store long link names, but these are much less widely supported. There is support for larger files, up to the 'ustar' limit of 8GB.
- Local reference classes have been added to package **methods**. These are a technique for avoiding unneeded copying of large components of objects while retaining standard R functional behavior. See ?LocalReferenceClasses.
- untar() has a new argument restore\_times which if false (not the default) discards the times in the tarball. This is useful if they are incorrect (some tarballs submitted to CRAN have times in a local timezone or many years in the past even though the standard required them to be in UTC).
- replayplot() cannot (and will not attempt to) replay plots recorded under R < 3.0.0. It may crash the R session if an attempt is made to replay plots created in a different build of R >= 3.0.0.
- Palette changes get recorded on the display list, so replaying plots (including when resizing screen devices and using dev.copy()) will work better when the palette is changed during a plot.
- chol(pivot = TRUE) now defaults to LAPACK, not LINPACK.
- The parse() function has a new parameter keep.source, which defaults to options("keep.source").
- Profiling via Rprof() now optionally records information at the statement level, not just the function level.
- The Rprof() function now quotes function names in in its output file on Windows, to be consistent with the quoting in Unix.
- Profiling via Rprof() now optionally records information about time spent in GC.
- The HTML help page for a package now displays non-vignette documentation files in a more accessible format.

• To support options(stringsAsFactors = FALSE), model.frame(), model.matrix() and replications() now automatically convert character vectors to factors without a warning.

- The print method for objects of class "table" now detects tables with 0-extents and prints the results as, e.g., ''. (Wish of PR#15198.)
- Deparsing involving calls to anonymous functions has been made closer to reversible by the addition of extra parentheses.
- The function utils::packageName() has been added as a lightweight version of methods::getPackageName().
- find.package(lib.loc = NULL) now treats loaded namespaces preferentially in the same way as attached packages have been for a long time.
- In Windows, the Change Directory dialog now defaults to the current working directory, rather than to the last directory chosen in that dialog.
- available.packages() gains a "license/restricts\_use" filter which retains only packages for which installation can proceed solely based on packages which are guaranteed not to restrict use.
- New check\_packages\_in\_dir() function in package tools for conveniently checking source packages along with their reverse dependencies.
- R's completion mechanism has been improved to handle help requests (starting with a question mark). In particular, help prefixes are now supported, as well as quoted help topics. To support this, completion inside quotes are now handled by R by default on all platforms.
- The memory manager now allows the strategy used to balance garbage collection and memory growth to be controlled by setting the environment variable R\_GC\_MEM\_GROW. See ?Memory for more details.
- ('For experts only', as the introductory manual says.) The use of environment variables R\_NSIZE and R\_VSIZE to control the initial (= minimum) garbage collection trigger for number of cons cels and size of heap has been restored: they can be overridden by the command-line options --min-nsize and --min-vsize; see ?Memory.
- On Windows, the device name for bitmap devices as reported by .Device and .Devices no longer includes the file name. This is for consistency with other platforms and was requested by the lattice maintainer.
- win.metafile() still uses the file name: the exact form is used by package tkrplot.
- set.seed(NULL) re-initializes .Random.seed as done at the beginning of the session if not already set. (Suggestion of Bill Dunlap.)
- The breaks argument in hist.default() can now be a function that returns the breakpoints to be used (previously it could only return the suggested number of breakpoints).
- File 'share/licenses/licenses.db' has some clarifications, especially as to which variants of 'BSD' and 'MIT' is intended and how to apply them to packages. The problematic licence 'Artistic-1.0' has been removed.

# LONG VECTORS: This section applies only to 64-bit platforms.

- There is support for vectors longer than  $2^{31}-1$  elements. This applies to raw, logical, integer, double, complex and character vectors, as well as lists. (Elements of character vectors remain limited to  $2^{31}-1$  bytes.)
- Most operations which can sensibly be done with long vectors work: others may return the error 'long vectors not supported yet'. Most of these are because they explicitly work with integer indices (e.g. anyDuplicated() and match()) or because other limits (e.g. of character strings or matrix dimensions) would be exceeded or the operations would be extremely slow.

• length() returns a double for long vectors, and lengths can be set to  $2^{31}$  or more by the replacement function with a double value.

- Most aspects of indexing are available. Generally double-valued indices can be used to access elements beyond  $2^{31} 1$ .
- There is some support for matrices and arrays with each dimension less than 2<sup>31</sup> but total number of elements more than that. Only some aspects of matrix algebra work for such matrices, often taking a very long time. In other cases the underlying Fortran code has an unstated restriction (as was found for complex svd()).
- dist() can produce dissimilarity objects for more than 65536 rows (but for example hclust() cannot process such objects).
- serialize() to a raw vector is unlimited in size (except by resources).
- The C-level function R\_alloc can now allocate 2<sup>35</sup> or more bytes.
- agrep() and grep() will return double vectors of indices for long vector inputs.
- Many calls to .C() have been replaced by .Call() to allow long vectors to be supported (now or in the future). Regrettably several packages had copied the non-API .C() calls and so failed.
- .C() and .Fortran() do not accept long vector inputs. This is a precaution as it is very unlikely that existing code will have been written to handle long vectors (and the R wrappers often assume that length(x) is an integer).
- Most of the methods for sort() work for long vectors.
  rank(), sort.list() and order() support long vectors (slowly except for radix sorting).
- sample() can do uniform sampling from a long vector.

#### PERFORMANCE IMPROVEMENTS:

- More use has been made of R objects representing registered entry points, which is more efficient as the address is provided by the loader once only when the package is loaded.
  - This has been done for packages base, methods, splines and tcltk: it was already in place for the other standard packages.
  - Since these entry points are always accessed by the R entry points they do not need to be in the load table which can be substantially smaller and hence searched faster. This does mean that .C / .Fortran / .Call calls copied from earlier versions of R may no longer work but they were never part of the API.
- Many .Call() calls in package base have been migrated to .Internal() calls.
- solve() makes fewer copies, especially when b is a vector rather than a matrix.
- eigen() makes fewer copies if the input has dimnames.
- Most of the linear algebra functions make fewer copies when the input(s) are not double (e.g. integer or logical).
- A foreign function call (.C() etc) in a package without a PACKAGE argument will only look in the first DLL specified in the 'NAMESPACE' file of the package rather than searching all loaded DLLs. A few packages needed PACKAGE arguments added.
- The **@<-** operator is now implemented as a primitive, which should reduce some copying of objects when used. Note that the operator object must now be in package **base**: do not try to import it explicitly from package **methods**.

## PACKAGE INSTALLATION:

• The transitional support for installing packages without namespaces (required since R 2.14.0) has been removed. R CMD build will still add a namespace, but a .First.lib() function will need to be converted.

R CMD INSTALL no longer adds a namespace (so installation will fail), and a .First.lib() function in a package will be ignored (with an installation warning for now).

As an exception, packages without a 'R' directory and no 'NAMESPACE' file can still be installed.

• Packages can specify in their 'DESCRIPTION file' a line like

#### Biarch: yes

to be installed on Windows with '--force-biarch'.

- Package vignettes can now be processed by other engines besides Sweave; see 'Writing R Extensions' and the tools::vignetteEngine help topic for details.
- The '\*.R' tangled source code for vignettes is now included in tarballs when R CMD build is used to produce them. In R 3.0.0, '\*.R' files not in the sources will be produced at install time, but eventually this will be dropped.
- The package type "mac.binary" now looks in a path in the repository without any Mac subtype (which used to be 'universal' or 'leopard'): it looks in 'bin/macosx/contrib/3.0' rather than 'bin/macosx/leopard/contrib/2.15'). This is the type used for the CRAN binary distribution for OS X as from R 3.0.0.
- File 'etc/Makeconf' makes more use of the macros \$(CC), \$(CXX), \$(F77) and \$(FC), so the compiler in use can be changed by setting just these (and if necessary the corresponding flags and FLIBS) in file '~/.R/Makevars'.

This is convenient for those working with binary distributions of R, e.g. on OS X.

#### **UTILITIES:**

• R CMD check now gives a warning rather than a note if it finds calls to abort, assert or exit in compiled code, and has been able to find the '.o' file in which the calls occur.

Such calls can terminate the R process which loads the package.

- The location of the build and check environment files can now be specified by the environment variables R\_BUILD\_ENVIRON and R\_CHECK\_ENVIRON, respectively.
- R CMD Sweave gains a '--compact' option to control possibly reducing the size of the PDF file it creates when '--pdf' is given.
- R CMD build now omits Eclipse's '.metadata' directories, and R CMD check warns if it finds them.
- R CMD check now does some checks on functions defined within reference classes, including of .Call() etc calls.
- R CMD check --as-cran notes assignments to the global environment, calls to data() which load into the global environment, and calls to attach().
- R CMD build by default uses the internal method of tar() to prepare the tarball. This is more likely to produce a tarball compatible with R CMD INSTALL and R CMD check: an external tar program, including options, can be specified *via* the environment variable R\_BUILD\_TAR.
- tools::massageExamples() is better protected against packages which re-define base functions such as cat() and get() and so can cause R CMD check to fail when checking examples.
- R CMD javareconf has been enhanced to be more similar to the code used by configure.

There is now a test that a JNI program can be compiled (like configure did) and only working settings are used.

It makes use of custom settings from configuration recorded in 'etc/javaconf'.

• The '--no-vignettes' argument of R CMD build has been renamed to the more accurate '--no-build-vignettes': its action has always been to (re)build vignettes and never omitted them.

R CMD check accepts '--no-build-vignettes' as a preferred synonym for '--no-rebuild-vignettes'.

#### DEPRECATED AND DEFUNCT:

- The ENCODING argument to .C() is defunct. Use iconv() instead.
- The .Internal(eval.with.vis) non-API function has been removed.
- Support for the converters for use with .C() has been removed, including the oft misused non-API header 'R\_ext/RConverters.h'.
- The previously deprecated uses of array() with a 0-length dim argument and tapply() with a 0-length INDEX list are now errors.
- 'Translation' packages are defunct.
- Calling rep() or rep.int() on a pairlist or other non-vector object is now an error.
- Several non-API entry points have been transferred to packages (e.g. R\_zeroin2) or replaced by different non-API entry points (e.g. R\_tabulate).
- The 'internal' graphics device invoked by .Call("R\_GD\_nullDevice", package = "grDevices") has been removed: use pdf(file = NULL) instead.
- The .Fortran() entry point "dqrls" which has not been used by R since version 2.15.1 is no longer available.
- Functions traceOn() and traceOff() in package methods are now defunct.
- Function CRAN.packages() is finally defunct.
- Use of col2rgb(0) is defunct: use par("bg") or NA instead.
- The long-defunct functions Rd\_parse(), anovalist.lm(), categpry(), clearNames(), gammaCody(), glm.fit.null(), lm.fit.null(), lm.wfit.null(), manglePackageNames(), mauchley.test(), package.contents(), print.coefmat(), reshapeLong(), reshapeWide(), tkclose(), tkcmd(), tkfile.dir(), tkfile.tail(), tkopen(), tkputs(), tkread(), trySilent() and zip.file.extract() have been removed entirely (but are still documented in the help system).
- The unused dataPath argument to attachNamespace() has been removed.
- grid.prompt() has been removed: use devAskNewPage() instead.
- The long-deprecated intensities component is no longer returned by hist().
- mean() for data frames and sd() for data frames and matrices are defunct.
- chol(pivot = FALSE, LINPACK = TRUE), ch2inv(LINPACK = TRUE), eigen(EISPACK = TRUE), solve(LINPACK = TRUE) and svd(LINPACK = TRUE) are defunct: LAPACK will be used, with a warning.
- The keep.source argument to library() and require() is defunct. This option needs to be set at install time.
- Documentation for real(), as.real() and is.real() has been moved to 'defunct' and the functions removed.
- The maxRasters argument of pdf() (unused since R 2.14.0) has been removed.
- The unused fontsmooth argument has been removed from the quartz() device.
- All the (non-API) EISPACK entry points in R have been removed.
- chol(pivot = TRUE, LINPACK = TRUE) is deprecated.
- The long-deprecated use of \synopsis in the 'Usage' section of '.Rd' files will be removed in R 3.1.0.

• .find.package() and .path.package() are deprecated: only the public versions without the dot have ever been in the API.

• In a package's 'DESCRIPTION' file,

License: X11

is deprecated, since it includes 'Copyright (C) 1996 X Consortium' which cannot be appropriate for a current R package. Use 'MIT' or 'BSD\_2\_clause' instead.

#### **CODE MIGRATION:**

- The C code underlying base graphics has been migrated to the **graphics** package (and hence no longer uses .Internal() calls).
- Most of the .Internal() calls used in the stats package have been migrated to C code in that package.
  - This means that a number of .Internal() calls which have been used by packages no longer exist, including .Internal(cor) .Internal(cov), .Internal(optimhess) and .Internal(update.formula).
- Some .External() calls to the base package (really to the R executable or shared library) have been moved to more appropriate packages. Packages should not have been using such calls, but some did (mainly those used by integrate()).

### PACKAGE parallel:

- There is a new function mcaffinity() which allows getting or setting the CPU affinity mask for the current R process on systems that supports this (currently only Linux has been tested successfully). It has no effect on systems which do not support process affinity. Users are not expected to use this function directly (with the exception of fixing libraries that break affinity settings like OpenBLAS) the function is rather intended to support affinity control in high-level parallel functions. In the future, R may supplement lack of affinity control in the OS by its own bookkeeping via mcaffinity() related to processes and threads it spawns.
- mcparallel() has a new argument mc.affinity which attempts to set the affinity of the child process according to the specification contained therein.
- The port used by socket clusters is chosen randomly: this should help to avoid clashes observed when two users of a multi-user machine try to create a cluster at the same time. To reproduce the previous behaviour set environment variable R\_PARALLEL\_PORT to 10187.

#### **C-LEVEL FACILITIES:**

- There has been some minor re-organization of the non-API header files. In particular, 'Rinternals.h' no longer includes the non-API header 'R\_exts/PrtUtil.h', and that no longer includes 'R\_exts/Print.h'.
- Passing NULL to .C() is now an error.
- .C() and .Fortran() now warn if "single" arguments are used with DUP = FALSE, as changes to such arguments are not returned to the caller.
- C entry points R\_qsort and R\_qsort\_I now have start and end as size\_t to allow them to work with longer vectors on 64-bit platforms. Code using them should be recompiled.
- A few recently added C entry points were missing the remapping to Rf\_, notably [dpq]nbinom\_mu.
- Some of the interface pointers formerly available only to R.app are now available to front-ends on all Unix-alikes: one has been added for the interface to View().
- $\bullet$  PACKAGE = "" is now an error in .C() etc calls: it was always contrary to the documentation.

• Entry point rcont2 has been migrated to package stats and so is no longer available.

- R\_SVN\_REVISION in 'Rversion.h' is now an integer (rather than a string) and hence usable as e.g. #if R\_SVN\_REVISION < 70000.
- The entry points rgb2hsv and hsv2rgb have been migrated to package grDevices and so are no longer available.
- R\_GE\_version has been increased to 10 and name2col removed (use R\_GE\_str2col instead). R internal colour codes are now defined using the typedef rcolor.
- The REPROTECT macro now checks that the protect index is valid.
- Several non-API entry points no longer used by R have been removed, including the Fortran entry points chol, chol2inv, cg, ch and rg, and the C entry points Brent\_fmin, fft\_factor and fft\_work.
- If a .External call is registered with a number of arguments (other than -1), the number of arguments passed is checked for each call (as for other foreign function calls).
- It is now possible to write custom connection implementations outside core R using 'R\_ext/Connections.h'. Please note that the implementation of connections is still considered internal and may change in the future (see the above file for details).

#### INTERNATIONALIZATION:

- The management of translations has been converted to R code: see ?tools::update\_pkg\_po.
- The translations for the R interpreter and RGui.exe are now part of the base package (rather than having sources in directory 'po' and being installed to 'share/locale'). Thus the base package supports three translation domains, R-base, R and RGui.
- The compiled translations which ship with R are all installed to the new package translations for easier updating. The first package of that name found on .libPaths() at the start of the R session will be used. (It is possible messages will be used before .libPaths() is set up in which case the default translations will be used: set environment variable R\_TRANSLATIONS to point to the location of the intended translations package to use this right from the start.)
- The translations form a separate group in the Windows installer, so can be omitted if desired.
- The markup for many messages has been changed to make them easier to translate, incorporating suggestions from Łukasz Daniel.

## INSTALLATION:

- There is again support for building without using the C 'long double' type. This is required by C99, but system implementations can be slow or flawed. Use configure option '--disable-long-double'.
- make pdf and make install-pdf now make and install the full reference index (including all base and recommended packages).
- The 'reference manual' on the Windows GUI menu and included in the installer is now the full reference index, including all base and recommended packages.
- R help pages and manuals have no ISBNs because ISBN rules no longer allow constantly changing content to be assigned an ISBN.
- The Windows installer no longer installs a Start Menu link to the static help pages; as most pages are generated dynamically, this led to a lot of broken links.
- Any custom settings for Java configuration are recorded in file 'etc/javaconf' for subsequent use by R CMD javareconf.

• There is now support for makeinfo version 5.0 (which requires a slightly different '.texi' syntax).

- The minimum versions for '--use-system-zlib' and --use-system-pcre are now tested as 1.2.5 and 8.10 respectively.
- On Windows, the stack size is reduced to 16MB on 32-bit systems: misguided users were launching many threads without controlling the stack size.
- configure no longer looks for file '~/.Rconfig': '~/.R/config' has long been preferred

#### BUG FIXES:

- When R CMD build is run in an encoding other than the one specified in the package's 'DESCRIPTION' file it tries harder to expand the authors@R field in the specified encoding. (PR#14958)
- If R CMD INSTALL is required to expand the authors@R field of the 'DESCRIPTION' file, it tries harder to do so in the encoding specified for the package (rather than using ASCII escapes).
- Fix in package **grid** for pushing a viewport into a layout cell, where the layout is within a viewport that has zero physical width OR where the layout has zero total relative width (likewise for height). The layout column widths (or row heights) in this case were being calculated with non-finite values. (Reported by Winston Chang.)
- solve(A, b) for a vector b gave the answer names from colnames(A) for LINPACK = TRUE but not in the default case.
- La.svd() accepts logical matrices (as documented, and as svd() did).
- legend() now accepts negative pch values, in the same way points() long has.
- Parse errors when installing files now correctly display the name of the file containing the bad code.
- In Windows, tcltk windows were not always properly constructed. (PR#15150)
- The internal functions implementing parse(), tools::parseLatex() and tools::parse\_Rd() were not reentrant, leading to errors in rare circumstances such as a garbage collection triggering a recursive call.
- Field assignments in reference class objects via \$<- were not being checked because the magic incantation to turn methods on for that primitive operator had been inadvertently omitted.
- setHook(hookname, value, action="replace") set the hook to be the value, rather than a list containing the value as documented. (PR#15167)
- If a package used a 'NEWS.Rd' file, the main HTML package index page did not link to it. (Reported by Dirk Eddelbuettel.)
- The primitive implementation of 0<- was not checking the class of the replacement. It now does a check, quicker but less general than slot<-. See the help.
- split(x, f) now recycles classed objects x in the same way as vectors. (Reported by Martin Morgan.)
- pbeta(.28, 1/2, 2200, lower.tail=FALSE, log.p=TRUE) is no longer -Inf; ditto for corresponding pt() and pf() calls, such as pt(45, df=5000, lower.tail=FALSE,log.p=TRUE). (PR#15162)
- The Windows graphics device would crash R if a user attempted to load the graphics history from a variable that was not a saved history. (PR#15230)
- The workspace size for the predict() method for loess() could exceed the maximum integer size. (Reported by Hiroyuki Kawakatsu.)
- ftable(x, row.vars, col.vars) now also works when the \*.vars arguments are (integer or character vectors) of length zero.

 $\bullet$  Calling cat() on a malformed UTF-8 string could cause the Windows GUI to lock up. (PR#15227)

• removeClass(cc) gave "node stack overflow" for some class definitions containing "array" or "matrix".

# CHANGES in previous versions

• Older news can be found in text format in files {'NEWS.0', 'NEWS.1' and 'NEWS.2' in the R home directory. News in HTML format for R versions from 2.10.0 to 2.15.3 is in 'doc/html/NEWS.2.html'.