NEWS for R Under development (unstable) (2014-08-20 r66449)

NEWS

R News



CHANGES IN R-devel

NEW FEATURES:

- anyNA() gains a recursive argument.
- When x is missing and names is not false (including the default value), Sys.getenv(x, names) returns an object of class "Dlist" and hence prints tidily.
- (Windows.) shell() no longer consults the environment variable SHELL: too many systems have been encountered where it was set incorrectly (usually to a path where software was compiled, not where it was installed). R_SHELL, the preferred way to select a non-default shell, can be used instead.
- Some unusual arguments to embedFonts() can now be specified as character vectors, and the defaults have changed accordingly.
- Functions in the Summary group duplicate less. (PR#15798)
- (Unix-alikes.) system(cmd, input =) now uses 'shell-execution-environment' redirection, which will be more natural if cmd is not a single command (but requires a POSIX-compliant shell). (PR#15508)
- read.fwf() and read.DIF() gain a fileEncoding argument, for convenience.
- Graphics devices can add attributes to their description in .Device and .Devices. Several of those included with R use a "filepath" attribute.
- pmatch() uses hashing in more cases and so is faster at the expense of using more memory. (PR#15697)
- pairs() gains new arguments to select sets of variables to be plotted against each other.
- file.info(, extra_cols = FALSE) allows a minimal set of columns to be computed: on some systems this can be significantly faster with large file lists.
- New function dir.exists() in base to test efficiently whether one or more paths exist and are directories.

• dput() and friends gain new controls '"hexNumeric"' and '"digits17"' which output double and complex quantities as, respectively, binary fractions (exactly, see sprintf("%a")) and as decimals with up to 17 significant digits.

- save(), saveRDS() and serialize() now support ascii = NA which writes ASCII files using the sprintf("%a")) for double/complex quantities. This is read-compatible with ascii = TRUE but avoids binary->decimal->binary conversions with potential loss of precision. Unfortunately Windows' lack of C99 compliance means that the format cannot be read correctly there in R before 3.1.2.
- The default for formatC(decimal.mark =) has been changed to be getOption("OutDec"); this makes it more consistent with format() and suitable for use in print methods, e.g. those for classes "density", "ecdf", "stepfun" and "summary.lm".
 - getOption("OutDec") is now consulted by the print method for class "kmeans", by
 cut(), dendrogram(), plot.ts() and quantile() when constructing labels and for
 the report from legend(trace = TRUE).
 (In part, wish of PR#15819.)
- printNum() and hence format() and formatC() give a warning if big.mark and decimal.mark are set to the same value (period and comma are not uncommonly used for each, and this is a check that conventions have not got mixed).
- merge() can create a result which uses long vectors on 64-bit platforms.
- dget() gains a new argument keep.source which defaults to FALSE for speed (dput() and dget() are most often used for data objects where this can make dget() many times faster).
- Packages may now use a file of common macro definitions in their help files, and may import definitions from other packages.
- A number of macros have been added in the new 'share/Rd' directory for use in package overview help pages, and promptPackage() now makes use of them.
- tools::parse_Rd() gains a new permissive argument which converts unrecognized macros into text. This is used by utils:::format.bibentry to allow LaTeX markup to be ignored.
- options(OutDec =) can now specify a multi-byte character, e.g., options(OutDec = "\u00b7") in a UTF-8 locale.
- utils::URLencode() was updated to use unreserved and reserved characters from RFC 3986, http://tools.ietf.org/html/rfc3986, instead of RFC 1738.
- is.recursive(x) is no longer true when x is an external pointer, a weak reference or byte code; the first enables all.equal(x, x) when x <- getClass(.).
- ls() (aka objects()) and as.list.environment() gain a new argument sorted.
- The "source" attribute (which has not been added to functions by R since before R version 2.14.0) is no longer treated as special.
- An experimental function returnValue() has been added to give on.exit() code access to a function's return value for debugging purposes.

INSTALLATION and INCLUDED SOFTWARE:

 The \donttest sections of R's help files can be tested by make check TEST_DONTTEST=TRUE.

UTILITIES:

There is support for testing the \dontrun and \donttest parts of examples in packages.

tools::testInstalledPackage() accepts new arguments commentDontrun = FALSE and commentDonttest = FALSE.

- R CMD check gains '--run-dontrun' and '--run-donttest' options.
- R CMD check now reports empty importFrom declarations in a 'NAMESPACE' file, as these are common errors (writing importFrom(Pkg) where import(Pkg) was intended).
- R CMD check --as-cran now checks code usage directly on the package namespace without loading and attaching the package and its suggests and enhances.
- The HTML generated by tools::Rd2HTML() and tools::toHTML() methods is now XHTML 1.0 Strict.

C-LEVEL FACILITIES:

- New C API utility logspace_sum(logx[], n).
- Entry points rbinom_mu, rnbinom_mu and rmultinom are remapped (by default). This requires packages using them to be re-installed.

BUG FIXES:

- Fixed two obscure bugs in pairlist subassignment, reported by Radford Neal as part of pqR issue 16.
- Fixes for bugs in handling empty arguments and argument matching by name in log().



CHANGES IN R 3.1.1 patched

NEW FEATURES:

- embedFonts() now defaults to format = "ps2write" for '.ps' and '.eps' files. This is available in Ghostscript 9.x (since 2010) whereas the previous default, format = "pswrite", was removed in 9.10.
- For consistency with [dpqr]norm(), [dp]lnorm(sdlog = 0) model a point mass at exp(mulog) rather than return NaN (for an error).
- capabilities() now reports if ICU is compiled in for use for collation (it is only actually used if a suitable locale is set for collation, and never for a C locale).
- (OS X only.) Package **tcltk** checks when loaded if it is linked against the CRAN X11-based Tcl/Tk and if so that the Tcl/Tk component and the X11 libraries are installed. This allows more informative error messages to be given.
 - The X11() device and X11-based versions of the data editor and viewer (invoked by edit() and View() for data frames and matrices from command-line R) check that the X11 libraries are installed.
- icuSetCollate() allows locale = "default", and locale = "none" to use OS services rather than ICU for collation.
 - Environment variable R_ICU_LOCALE can be used to set the default ICU locale, in case the one derived from the OS locale is inappropriate (this is currently necessary on Windows).

• New function icuGetCollate() to report on the ICU collation locale in use (if any).

INSTALLATION and INCLUDED SOFTWARE:

- The configure script reports on the more important capabilities/options which are not compiled in.
- When building R as a shared library, the '-L\${R_HOME}/lib\${R_ARCH}' flag is placed earlier in the link commands used during installation and when packages are installed: this helps ensure that the current build has priority if an R shared library has already been installed by e.g. install-libR in a library mentioned in LDFLAGS (and not in 'your system's library directory' as documented). (Wish of PR#15790.)
- LaTeX package upquote is no longer required for R's use of inconsolata.

BUG FIXES:

- dmultinom() did not handle non-finite probabilities correctly.
- prettyNum(x, zero.print=*) now also works when x contains NAs.
- A longstanding bug in nlminb() on Windows was traced to a compiler bug in gcc 4.6.3; a workaround has been put in place. (PR#15244 and PR#15914).
- Rendering of \command in HTML has been improved: this is particularly evident on the help page for INSTALL.
- as.hexmode(x) and as.octmode(x) now behave correctly for some numeric x, e.g., c(NA, 1) or c(1, pi).
- drop1() failed if the scope argument had no variables to drop. (PR#15935)
- edit() (and hence fix()) failed if an object had a non-character attribute named "source" (an attribute that had been used in R prior to version 2.14.0).
- callGeneric() could fail if the generic had ... as a formal argument. (PR#15937).



CHANGES IN R 3.1.1

NEW FEATURES:

- When attach() reports conflicts, it does so compatibly with library() by using message().
- R CMD Sweave no longer cleans any files by default, compatibly with versions of R prior to 3.1.0. There are new options '--clean', '--clean=default' and '--clean=keepOuts'.
- tools::buildVignette() and tools::buildVignettes() with clean = FALSE no longer remove any created files. buildvignette() gains a keep argument for more cleaning customization.
- The Bioconductor 'version' used by setRepositories() can now be set by environment variable R_BIOC_VERSION at runtime, not just when R is installed. (It has been stated that Bioconductor will switch from 'version' 2.14 to 'version' 3.0 during the lifetime of the R 3.1 series.)
- Error messages from bugs in embedded 'Sexpr' code in Sweave documents now report the source location.

• type.convert(), read.table() and similar read.*() functions get a new numerals argument, specifying how numeric input is converted when its conversion to double precision loses accuracy. The default value, "allow.loss" allows accuracy loss, as in R versions before 3.1.0.

- For some compilers, integer addition could overflow without a warning. R's internal code for both integer addition and subtraction is more robust now. (PR#15774)
- The function determining the default number of knots for smooth.spline() is now exported, as .nknots.smspl().
- dbeta(, a,b), pbeta(), qbeta() and rbeta() are now defined also for a = 0, b = 0, or infinite a and b (where they typically returned NaN before).
- Many package authors report that the RStudio graphics device does not work correctly
 with their package's use of dev.new(). The new option dev.new(noRStudioGD =
 TRUE) replaces the RStudio override by the default device as selected by R itself, still
 respecting environment variables R_INTERACTIVE_DEVICE and R_DEFAULT_DEVICE.
- readRDS() now returns visibly.
- Modifying internal logical scalar constants now results in an error instead of a warning.
- install.packages(repos = NULL) now accepts http:// or ftp:// URLs of package archives as well as file paths, and will download as required. In most cases repos = NULL can be deduced from the extension of the URL.
- The warning when using partial matching with the \$ operator on data frames is now only given when options("warnPartialMatchDollar") is TRUE.
- Package help requests like package?foo now try the package foo whether loaded or not.
- General help requests now default to trying all loaded packages, not just those on the search path.
- Added a new function promptImport(), to generate a help page for a function that
 was imported from another package (and presumably re-exported, or help would not
 be needed).

INSTALLATION and INCLUDED SOFTWARE:

- configure option '--with-internal-tzcode' can now be used with variable rsharedir.
- The included version of PCRE has been updated to 8.35.
- There is a new target make uninstall-libR to remove an installed shared/static 'libR'.
 - make install-libR now works if a sub-architecture is used, although the user will need to specify libdir differently for different sub-architectures.
- There is more extensive advice on which LaTeX packages are required to install R or to make package manuals (as done by R CMD check) in the 'Writing R Extensions' manual.
- Compilers/linkers were handling the visibility control in 'src/extra/xz' inconsistently (and apparently in some cases incorrectly), so it has been simplified. (PR#15327)
- (Windows) There is updated support for the use of ICU for collation: see the 'R Installation and Administration Manual'.

BUG FIXES:

dbinom(x, n), pbinom(), dpois(), etc, are slightly less restrictive in checking if n is integer-valued. (Wish of PR#15734.)

- pchisq(x, df, ncp, log.p = TRUE) is more accurate and no longer underflows for small x and ncp < 80, e.g, for pchisq(1e-5, df = 100, ncp = 1, log = TRUE). (Based on PR#15635 and a suggestion by Roby Joehanes.)
- The s ("step into") command in the debugger would cause R to step into expressions evaluated there, not just into functions being debugged. (PR#15770)
- The C code used by strptime() rejected time-zone offsets of more than +1200 (+1245, +1300 and +1400 can occur). (PR#15768)
- (Windows only.) png(type = "cairo", antialias = "gray") was not accepted. (PR#15760)
- Use of save(..., envir=) with named objects could fail. (PR#15758)
- Sweave() mis-parsed 'Sexpr' expressions that contained backslashes. (PR#15779)
- The return value from options (foo = NULL) was not the previous value of the option. (PR#15781)
- enc2utf8() and enc2native() did not always mark the encoding of the return values when it was known.
- dnbinom(x, size = <large>, mu, log = TRUE) no longer underflows to -Inf for large mu, thanks to a suggestion from Alessandro Mammana (MPI MolGen, Berlin).
- pbeta(x, a, b, log = TRUE) no longer behaves discontinuously (in a small x-region) because of denormalized numbers. Also, pbeta(1-1e-12, 1e30, 1.001, log=TRUE) now terminates "in real time".
- The "CRAN" filter (see available.packages()) no longer removes duplicates other than of packages on CRAN, and does not fail if there is no CRAN repository in getOption("repos").
- The device listing from dev2bitmap() and bitmap() was truncated to 1000 characters: modern versions of GhostScript on most platforms have many more devices.
- (Windows.) Commands such as Sys.which() and pipe() which needed to find the full path to a command could segfault if the 'long' path name was much longer than the 'short' path name (which Sys.which() returns), as the behaviour of the Windows API call had changed.
- R CMD build will fail with an error if one of the packages specified in the 'VignetteBuilder' field is not installed. (Without loading those packages it cannot be ascertained which files are intended to be vignettes. This means that the 'VignetteBuilder' packages have to be installed for package checking too.) (Wish of PR#15775.)
- Misguided attempts to use chull() with non-finite points now give an error (related to PR#15777).
- For a formula with exactly 32 variables the 32nd variable was aliased to the intercept in some C-level computations of terms, so that for example attempting to remove it would remove the intercept instead (and leave a corrupt internal structure). (PR#15735)
- anyDuplicated() silently returned wrong values when the first duplicate was at an index which was too large to be stored in an integer vector (although a lot of RAM and patience would have been needed to encounter this).
- tools::Rd2ex(commentDontrun = FALSE) failed if the block had only one line.

• Hexadecimal constants such as 0x110p-5L which were incorrectly qualified by L were parsed incorrectly since R 3.0.0, with a slightly garbled warning. (PR#15753)

- system() returned success on some platforms even if the system was unable to launch a process. (PR#15796)
- (Windows Rgui console.) Unbuffered output was sometimes not output immediately if the prompt was not on the last line of the console.
- The built-in help server did not declare the encoding for the 'DESCRIPTION' or other text files to be the package encoding, so non-ASCII characters could be displayed incorrectly.
- R is now trying harder to not cleanup child processes that were not spawned by mcparallel() on platforms that provide information about the source process of the SIGCHLD signal. This allows 3rd party libraries to manage the exit status of children that they spawn without R interfering.
- mcmapply() was only parallelizing if the number of jobs was bigger than the number of cores. It now parallelizes if the number of jobs is more than one.
- Auto-printing would re-evaluate its argument when trying to dispatch to a print method. This is now avoided when possible.
- Unserializing (including load() and readRDS()) could silently return incorrect numeric values from ASCII saves if there was a read error.
- getParseData() could return incorrect values for the parents of some elements. (Reported by Andrew Redd.)
- Attempting to use data frames of 2³¹ or more rows with merge() or to create a merged data frame of that size now gives a clearer error message.
- parse() did not check its file argument was a connection if it was not a character string, so e.g. parse(FALSE) attempted to read from stdin.

 Nor did dump() and dput().
- The "help.try.all.packages" option was ignored when the shortcut syntax for help was used, e.g. ?foo.
- A potential segfault in string allocation has been fixed. (Found by Radford Neal.)
- Potential memory protection errors in sort() and D() have been fixed. (Found by Radford Neal.)
- Fixed a lack of error checking in graphics event functions. (Found by Radford Neal; a different patch used here than the one in pqR.)
- numericDeriv() sometimes miscalculated the gradient. (PR#15849, reported originally by Radford Neal)



CHANGES IN R 3.1.0

NEW FEATURES:

• type.convert() (and hence by default read.table()) returns a character vector or factor when representing a numeric input as a double would lose accuracy. Similarly for complex inputs.

If a file contains numeric data with unrepresentable numbers of decimal places that are intended to be read as numeric, specify colClasses in read.table() to be "numeric".

• tools::Rdiff(useDiff = FALSE) is closer to the POSIX definition of diff -b (as distinct from the description in the man pages of most systems).

- New function anyNA(), a version of any(is.na(.)) which is fast for atomic vectors, based on a proposal by Tim Hesterberg. (Wish of PR#15239.)
- arrayInd(*, useNames = TRUE) and, analogously, which(*, arr.ind = TRUE) now make use of names(.dimnames) when available.
- is.unsorted() now also works for raw vectors.
- The "table" method for as.data.frame() (also useful as as.data.frame.table()) now passes sep and base arguments to provideDimnames().
- uniroot() gets new optional arguments, notably extendInt, allowing to auto-extend the search interval when needed. The return value has an extra component, init.it.
- switch(f, ...) now warns when f is a factor, as this typically happens accidentally where the useR meant to pass a character string, but f is treated as integer (as always documented).
- The parser has been modified to use less memory.
- The way the unary operators (+ !) handle attributes is now more consistent. If there is no coercion, all attributes (including class) are copied from the input to the result: otherwise only names, dims and dimnames are.
- colorRamp() and colorRampPalette() now allow non-opaque colours and a ramp in opacity via the new argument alpha = TRUE. (Suggested by Alberto Krone-Martins, but optionally as there are existing uses which expect only RGB values.)
- grid.show.layout() and grid.show.viewport() get an optional vp.ex argument.
- There is a new function find_gs_cmd() in the tools package to locate a GhostScript executable. (This is an enhanced version of a previously internal function there.)
- object.size() gains a format() method.
- There is a new family, "ArialMT", for the pdf() and postscript() devices. This will only be rendered correctly on viewers which have access to Monotype TrueType fonts (which are sometimes requested by journals).
- The text and PDF news files, including 'NEWS' and 'NEWS.2', have been moved to the 'doc' directory.
- combn(x, simplify = TRUE) now gives a factor result for factor input x (previously user error). (Related to PR#15442.)
- Added utils::fileSnapshot() and utils::changedFiles() functions to allow snapshots and comparison of directories of files.
- make.names(names, unique=TRUE) now tries to preserve existing names. (Suggestion of PR#15452.)
- New functions cospi(x), sinpi(x), and tanpi(x), for more accurate computation of cos(pi*x), etc, both in R and the C API. Using these gains accuracy in some cases, e.g., inside lgamma() or bessell(). (Suggested by Morten Welinder in PR#15529.)
- print.table(x, zero.print = ".") now also has an effect when x is not integer-valued.
- There is more support to explore the system's idea of time-zone names. Sys.timezone() tries to give the current system setting by name (and succeeds at least on Linux, OS X, Solaris and Windows), and OlsonNames() lists the names in the system's Olson database. Sys.timezone(location = FALSE) gives the previous behaviour.

• Platforms with a 64-bit time_t type are allowed to handle conversions between the "POSIXct" and "POSIXIt" classes for date-times outside the 32-bit range (before 1902 or after 2037): the existing workarounds are used on other platforms. (Note that time-zone information for post-2037 is speculative at best, and the OS services are tested for known errors and so not used on OS X.)

Currently time t is usually long and hence 64-bit on Universities 64-bit platforms:

- Currently time_t is usually long and hence 64-bit on Unix-alike 64-bit platforms: however in several cases the time-zone database is 32-bit. For R for Windows it is 64-bit (for both architectures as from this version).
- The "save.defaults" option can include a value for compression_level. (Wish of PR#15579.)
- colSums() and friends now have support for arrays and data-frame columns with 2³¹ or more elements.
- as.factor() is faster when f is an unclassed integer vector (for example, when called from tapply()).
- fft() now works with longer inputs, from the 12 million previously supported up to 2 billion. (PR#15593)
- Complex svd() now uses LAPACK subroutine ZGESDD, the complex analogue of the routine used for the real case.
- Sweave now outputs '.tex' files in UTF-8 if the input encoding is declared to be UTF-8, regardless of the local encoding. The UTF-8 encoding may now be declared using a LaTeX comment containing the string %\SweaveUTF8 on a line by itself.
- file.copy() gains a copy.date argument.
- Printing of date-times will make use of the time-zone abbreviation in use at the time, if known. For example, for Paris pre-1940 this could be 'LMT', 'PMT', 'WET' or 'WEST'. To enable this, the "POSIX1t" class has an optional component "zone" recording the abbreviation for each element.
 - For platforms which support it, there is also a component "gmtoff" recording the offset from GMT where known.
- (On Windows, by default on OS X and optionally elsewhere.) The system C function strftime has been replaced by a more comprehensive version with closer conformance to the POSIX 2008 standard.
- dnorm(x, log = FALSE) is more accurate (but somewhat slower) for |x| > 5; as suggested in PR#15620.
- Some versions of the tiff() device have further compression options.
- read.table(), readLines() and scan() have a new argument to influence the treatment of embedded nuls.
- Avoid duplicating the right hand side values in complex assignments when possible.
 This reduces copying of replacement values in expressions such as Z\$a <- a0 and ans[[i]] <- tmp: some package code has relied on there being copies.</p>
 Also, a number of other changes to reduce copying of objects; all contributed by or based on suggestions by Michael Lawrence.
- The fast argument of KalmanLike(), KalmanRun() and KalmanForecast() has been replaced by update, which instead of updating mod in place, optionally returns the updated model in an attribute "mod" of the return value.
- arima() and makeARIMA() get a new optional argument SSinit, allowing the choice
 of a different state space initialization which has been observed to be more reliable
 close to non-stationarity: see PR#14682.

• warning() has a new argument noBreaks., to simplify post-processing of output with options(warn = 1).

- pushBack() gains an argument encoding, to support reading of UTF-8 characters using scan(), read.table() and related functions in a non-UTF-8 locale.
- all.equal.list() gets a new argument use.names which by default labels differing components by names (if they match) rather than by integer index. Saved R output in packages may need to be updated.
- The methods for all.equal() and attr.all.equal() now have argument check.attributes after ... so it cannot be partially nor positionally matched (as it has been, unintentionally).
 - A side effect is that some previously undetected errors of passing empty arguments (no object between commas) to all.equal() are detected and reported.
 - There are explicit checks that check.attributes is logical, tolerance is numeric and scale is NULL or numeric. This catches some unintended positional matching. The message for all.equal.numeric() reports a "scaled difference" only for scale != 1.
- all.equal() now has a "POSIXt" method replacing the "POSIXct" method.
- The "Date" and "POSIXt" methods of seq() allows by = "quarter" for completeness (by = "3 months" always worked).
- file.path() removes any trailing separator on Windows, where they are invalid (although sometimes accepted). This is intended to enhance the portability of code written by those using POSIX file systems (where a trailing / can be used to confine path matching to directories).
- New function agrepl() which like grepl() returns a logical vector.
- fifo() is now supported on Windows. (PR#15600)
- sort.list(method = "radix") now allows negative integers (wish of PR#15644).
- Some functionality of print.ts() is now available in .preformat.ts() for more modularity.
- mcparallel() gains an option detach = TRUE which allows execution of code independently of the current session. It is based on a new estranged = TRUE argument to mcfork() which forks child processes such that they become independent of the parent process.
- The pdf() device omits circles and text at extremely small sizes, since some viewers were failing on such files.
- The rightmost break for the "months", "quarters" and "years" cases of hist.POSIX1t() has been increased by a day. (Inter alia, fixes PR#15717.)
- The handling of DF[i,] <- a where i is of length 0 is improved. (Inter alia, fixes PR#15718.)
- hclust() gains a new method "ward.D2" which implements Ward's method correctly. The previous "ward" method is "ward.D" now, with the old name still working. Thanks to research and proposals by Pierre Legendre.
- The sunspot.month dataset has been amended and updated from the official source, whereas the sunspots and sunspot.year datasets will remain immutable. The documentation and source links have been updated correspondingly.
- The summary() method for "lm" fits warns if the fit is essentially perfect, as most of the summary may be computed inaccurately (and with platform-dependent values).

Programmers who use summary() in order to extract just a component which will be reliable (e.g., \$cov.unscaled) should wrap their calls in suppressWarnings().

INSTALLATION and INCLUDED SOFTWARE:

- The included version of LAPACK has been updated to 3.5.0.
- There is some support for parallel testing of an installation, by setting TEST_MC_CORES to an integer greater than one to indicate the maximum number of cores to be used in parallel. (It is worth specifying at least 8 cores if available.) Most of these require a make program (such as GNU make and dmake) which supports the \$MAKE -j nproc syntax.

Except on Windows: the tests of standard package examples in make check are done in parallel. This also applies to running tools::testInstalledPackages().

The more time-consuming regression tests are done in parallel.

The package checks in make check-devel and make check-recommended are done in parallel.

- More of make check will work if recommended packages are not installed: but recommended packages remain needed for thorough checking of an R build.
- The version of 'tzcode' included in 'src/extra/tzone' has been updated. (Formerly used only on Windows.)
- The included (64-bit) time-zone conversion code and Olson time-zone database can be used instead of the system version: use configure option '--with-internal-tzcode'. This is the default on Windows and OS X. (Note that this does not currently work if a non-default rsharedir configure variable is used.)
 - (It might be necessary to set environment variable TZ on OSes where this is not already set, although the system timezone is deduced correctly on at least Linux, OS X and Windows.)
 - This option also switches to the version of strftime included in directory 'src/extra/tzone'.
- configure now tests for a C++11-compliant compiler by testing some basic features. This by default tries flags for the compiler specified by 'CXX', but an alternative compiler, options and standard can be specified by variables 'CXX1X', 'CXX1XFLAGS' and 'CXX1XSTD' (e.g., '-std=gnu++11').
- R can now optionally be compiled to use reference counting instead of the NAMED mechanism by defining SWITCH_TO_REFCNT in 'Rinternals.h'. This may become the default in the future.
- There is a new option '--use-system-tre' to use a suitable system tre library: at
 present this means a version from their git repository, after corrections. (Wish of
 PR#15660.)

PACKAGE INSTALLATION:

- The CRANextra repository is no longer a default repository on Windows: all the binary versions of packages from CRAN are now on CRAN, although CRANextra contains packages from Omegahat and elsewhere used by CRAN packages.
- Only vignettes sources in directory 'vignettes' are considered to be vignettes and hence indexed as such.
- In the 'DESCRIPTION' file.

License: X11

- is no longer recognized as valid. Use 'MIT' or 'BSD_2_clause' instead, both of which need '+ file LICENSE'.
- For consistency, entries in '.Rinstignore' are now matched case-insensitively on all platforms.
- Help for S4 methods with very long signatures now tries harder to split the description in the 'Usage' field to no more than 80 characters per line (some packages had over 120 characters).
- R CMD INSTALL --build (not Windows) now defaults to the internal tar() unless R_INSTALL_TAR is set.
- There is support for compiling C++11 code in packages on suitable platforms: see 'Writing R Extensions'.
- Fake installs now install the contents of directory 'inst': some packages use this to install e.g. C++ headers for use by other packages that are independent of the package itself. Option '--no-inst' can be used to get the previous behaviour.

DEBUGGING:

- The behaviour of the code browser has been made more consistent, in part following the suggestions in PR#14985.
- Calls to browser() are now consistent with calls to the browser triggered by debug(), in that Enter will default to n rather than c.
- A new browser command **s** has been added, to "step into" function calls.
- A new browser command f has been added, to "finish" the current loop or function.
- Within the browser, the command help will display a short list of available commands.

UTILITIES:

- Only vignettes sources in directory 'vignettes' are considered to be vignettes by R CMD check. That has been the preferred location since R 2.14.0 and is now obligatory.
- For consistency, R CMD build now matches entries in '.Rbuildignore' and 'vignettes/.install_extras' case-insensitively on all platforms (not just on Windows).
- checkFF() (called by R CMD check by default) can optionally check foreign function calls for consistency with the registered type and argument count. This is the default for R CMD check --as-cran or can be enabled by setting environment variable _R_CHECK_FF_CALLS_ to 'registration' (but is in any case suppressed by '--install=no'). Because this checks calls in which .NAME is an R object and not just a literal character string, some other problems are detected for such calls. Functions suppressForeignCheck() and dontCheck() have been added to allow package authors to suppress false positive reports.
- R CMD check --as-cran warns about a false value of the 'DESCRIPTION' field 'BuildVignettes' for Open Source packages, and ignores it. (An Open Source package needs to have complete sources for its vignettes which should be usable on a suitably well-equipped system).
- R CMD check --no-rebuild-vignettes is defunct:
 R CMD check --no-build-vignettes has been preferred since R 3.0.0.
- R CMD build --no-vignettes is defunct:
 - R CMD build --no-build-vignettes has been preferred since R 3.0.0.

• R CMD Sweave and R CMD Stangle now process both Sweave and non-Sweave vignettes. The tools::buildVignette() function has been added to do the same tasks from within R.

- The flags returned by R CMD config --ldflags and (where installed) pkg-config --libs libR are now those needed to link a front-end against the (shared or static) R library.
- 'Sweave.sty' has a new option '[inconsolata]'.
- R CMD check customizations such as _R_CHECK_DEPENDS_ONLY_ make available packages only in 'LinkingTo' only for installation, and not for loading/runtime tests.
- tools::checkFF() reports on .C and .Fortran calls with DUP = FALSE if argument check_DUP is true. This is selected by R CMD check by default.
- R CMD check --use-gct can be tuned to garbage-collect less frequently using gctorture2() via the setting of environment variable _R_CHECK_GCT_N_.
- Where supported, tools::texi2dvi() limits the number of passes tried to 20.

C-LEVEL FACILITIES:

- (Windows only) A function R_WaitEvent() has been added (with declaration in header'R.h') to block execution until the next event is received by R.
- Remapping in the 'Rmath.h' header can be suppressed by defining 'R_NO_REMAP_RMATH'.
- The remapping of rround() in header 'Rmath.h' has been removed: use fround() instead.
- ftrunc() in header 'Rmath.h' is now a wrapper for the C99 function trunc(), which might as well be used in C code: ftrunc() is still needed for portable C++ code.
- The never-documented remapping of prec() to fprec() in header 'Rmath.h' has been removed.
- The included LAPACK subset now contains ZGESDD and ZGELSD.
- The function LENGTH() now checks that it is only applied to vector arguments. However, in packages length() should be used. (In R itself LENGTH() is a macro without the function overhead of length().)
- Calls to SET_VECTOR_ELT() and SET_STRING_ELT() are now checked for indices which are in-range: several packages were writing one element beyond the allocated length.
- allocVector3 has been added which allows custom allocators to be used for individual vector allocations.

DEPRECATED AND DEFUNCT:

- chol(pivot = TRUE, LINPACK = TRUE) is defunct.

 Arguments EISPACK for eigen() and LINPACK for chol(), chol2inv(), solve() and svd() are ignored: LAPACK is always used.
- .find.package() and .path.package() are defunct: only the versions without the initial dot introduced in R 2.13.0 have ever been in the API.
- Partial matching when using the \$ operator on data frames now throws a warning and may become defunct in the future. If partial matching is intended, replace foo\$bar by foo[["bar", exact = FALSE]].
- The long-deprecated use of \synopsis in the 'Usage' section of '.Rd' files has been removed: such sections are now ignored (with a warning).

- package.skeleton()'s deprecated argument namespace has been removed.
- Many methods are no longer exported by package **stats**. They are all registered on their generic, which should be called rather than calling a method directly.
- Functions readNEWS() and checkNEWS() in package tools are defunct.
- download.file(method = "lynx") is deprecated.
- .C(DUP = FALSE) and .Fortran(DUP = FALSE) are now deprecated, and may be disabled in future versions of R. As their help has long said, .Call() is much preferred. R CMD check notes such usages (by default).
- The workaround of setting R_OSX_VALGRIND has been removed: it is not needed in current valgrind.

BUG FIXES:

- Calling lm.wfit() with no non-zero weights gave an array-overrun in the Fortran code and a not very sensible answer. It is now special-cased with a simpler answer (no qr component).
- Error messages involving non-syntactic names (e.g., as produced by `\r` when that
 object does not exist) now encode the control characters. (Reported by Hadley Wickham.)
- getGraphicsEvent() caused 100% usage of one CPU in Windows. (PR#15500)
- nls() with no start argument may now work inside another function (scoping issue).
- pbeta() and similar work better for very large (billions) ncp.
- Where time zones have changed abbreviations over the years, the software tries to more consistently use the abbreviation appropriate to the time or if that is unknown, the current abbreviation. On some platforms where the C function localtime changed the tzname variables the reported abbreviation could have been that of the last time converted.
- all.equal(list(1), identity) now works.
- Bug fix for pushing viewports in grid (reported by JJ Allaire and Kevin Ushey).
 NOTE for anyone poking around within the graphics engine display list (despite the warnings not to) that this changes what is recorded by grid on the graphics engine display list.
- Extra checks have been added for unit resolution and conversion in **grid**, to catch instances of division-by-zero. This may introduce error messages in existing code and/or produce a different result in existing code (but only where a non-finite location or dimension may now become zero).
- Some bugs in TRE have been corrected by updating from the git repository. This
 allows R to be installed on some platforms for which this was a blocker (PR#15087
 suggests Linux on ARM and HP-UX).
- ? applied to a call to an S4 generic failed in several cases. (PR#15680)
- The implicit S4 generics for primitives with . . . in their argument list were incorrect. (PR#15690)
- Bug fixes to methods::callGeneric(). (PR#15691)
- The bug fix to aggregrate() in PR#15004 introduced a new bug in the case of no grouping variables. (PR#15699)
- In rare cases printing deeply nested lists overran a buffer by one byte and on a few platforms segfaulted. (PR#15679)

• The dendrogram method of as.dendrogram() was hidden accidentally, (PR#15703), and order.dendrogram(d) gave too much for a leaf d. (PR#15702)

- R would try to kill processes on exit that have pids ever used by a child process spawned by mcparallel even though the current process with that pid was not actually its child.
- cophenetic() applied to a "dendrogram" object sometimes incorrectly returned a "Labels" attribute with dimensions. (PR#15706)
- printCoefmat() called from quite a few print() methods now obeys small getOption("width") settings, line wrapping the '"signif. codes"' legend appropriately. (PR#15708)
- model.matrix() assumed that the stored dimnames for a matrix was NULL or length 2, but length 1 occurred.
- The clipping region for a device was sometimes used in base graphics before it was set.



CHANGES IN R. 3.0.3

NEW FEATURES:

- On Windows there is support for making '.texi' manuals using texinfo 5.0 or later: the setting is in file 'src/gnuwin32/MkRules.dist'.

 A packaging of the Perl script and modules for texinfo 5.2 has been made available at http://www.stats.ox.ac.uk/pub/Rtools/.
- write.table() now handles matrices of 2³¹ or more elements, for those with large amounts of patience and disc space.
- There is a new function, La_version(), to report the version of LAPACK in use.
- The HTML version of 'An Introduction to R' now has links to PNG versions of the figures.
- There is some support to produce manuals in ebook formats. (See 'doc/manual/Makefile'. Suggested by Mauro Cavalcanti.)
- On a Unix-alike Sys.timezone() returns NA if the environment variable TZ is unset, to distinguish it from an empty string which on some OSes means the 'UTC' time zone.
- The backtick may now be escaped in strings, to allow names containing them to be constructed, e.g. `\``. (PR#15621)
- read.table(), readLines() and scan() now warn when an embedded nul is found in the input. (Related to PR#15625 which was puzzled by the behaviour in this unsupported case.)
- (Windows only.) file.symlink() works around the undocumented restriction of the Windows system call to backslashes. (Wish of PR#15631.)
- KalmanForecast(fast = FALSE) is now the default, and the help contains an example of how fast = TRUE can be used in this version. (The usage will change in 3.1.0.)
- strptime() now checks the locale only when locale-specific formats are used and caches the locale in use: this can halve the time taken on OSes with slow system functions (e.g., OS X).

• strptime() and the format() methods for classes "POSIXct", "POSIX1t" and "Date" recognize strings with marked encodings: this allows, for example, UTF-8 French month names to be read on (French) Windows.

- iconv(to = "utf8") is now accepted on all platforms (some implementations did already, but GNU libiconv did not: however converted strings were not marked as being in UTF-8). The official name, "UTF-8" is still preferred.
- available.packages() is better protected against corrupt metadata files. (A recurring problem with Debian package shogun-r: PR#14713.)
- Finalizers are marked to be run at garbage collection, but run only at a somewhat safer later time (when interrupts are checked). This circumvents some problems with finalizers running arbitrary code during garbage collection (the known instances being running options() and (C-level) path.expand() re-entrantly).

INSTALLATION and INCLUDED SOFTWARE:

• The included version of PCRE has been updated to 8.34. This fixes bugs and makes the behaviour closer to Perl 5.18. In particular, the concept of 'space' includes 'VT' and hence agrees with POSIX's.

PACKAGE INSTALLATION:

- The new field 'SysDataCompression' in the 'DESCRIPTION' file allows user control over the compression used for 'sysdata.rda' objects in the lazy-load database.
- install.packages(dependencies = value) for value = NA (the default) or value = TRUE omits packages only in LinkingTo for binary package installs.

C-LEVEL FACILITIES:

- The long undocumented remapping of rround() to Rf_fround() in header 'Rmath.h' is now formally deprecated: use fround() directly.
- Remapping of prec() and trunc() in the 'Rmath.h' header has been disabled in C++ code (it has caused breakage with libc++ headers).

BUG FIXES:

- getParseData() truncated the imaginary part of complex number constants. (Reported by Yihui Xie.)
- dbeta(x, a, b) with a or b within a factor of 2 of the largest representable number could infinite-loop. (Reported by Ioannis Kosmidis.)
- provideDimnames() failed for arrays with a 0 dimension. (PR#15465)
- rbind() and cbind() did not handle list objects correctly. (PR#15468)
- replayPlot() now checks if it is replaying a plot from the same session.
- rasterImage() and grid.raster() now give error on an empty (zero-length) raster. (Reported by Ben North.)
- plot.lm() would sometimes scramble the labels in plot type 5. (PR#15458 and PR#14837)
- min() did not handle NA_character_ values properly. (Reported by Magnus Thor Torfason.)
- (Windows only.) readRegistry() would duplicate default values for keys. (PR#15455)

• str(..., strict.width = "cut") did not handle it properly when more than one line needed to be cut. (Reported by Gerrit Eichner.)

- Removing subclass back-references when S4 classes were removed or their namespace unloaded had several bugs (e.g., PR#15481).
- aggregate() could fail when there were too many levels present in the by argument. (PR#15004)
- namespaceImportFrom() needed to detect primitive functions when checking for duplicated imports (reported by Karl Forner).
- getGraphicsEvent() did not exit when a user closed the graphics window. (PR#15208)
- Errors in vignettes were not always captured and displayed properly. (PR#15495)
- contour() could fail when dealing with extremely small z values. (PR#15454)
- Several functions did not handle zero-length vectors properly, including browseEnv(), format(), gl(), relist() and summary.data.frame(). (E.g., PR#15499)
- Sweave() did not restore the R output to the console if it was interrupted by a user in the middle of evaluating a code chunk. (Reported by Michael Sumner.)
- Fake installs of packages with vignettes work again.
- Illegal characters in the input caused parse() (and thus source()) to segfault. (PR#15518)
- The nonsensical use of nmax = 1 in duplicated() or unique() is now silently ignored.
- qcauchy(p, *) is now fully accurate even when p is very close to 1. (PR#15521)
- The validmu() and valideta() functions in the standard glm() families now also report non-finite values, rather than failing.
- Saved vignette results (in a '.Rout.save' file) were not being compared to the new ones during R CMD check.
- Double-clicking outside of the list box (e.g., on the scrollbar) of a Tk listbox widget generated by tk_select.list() no longer causes the window to close. (PR#15407)
- Improved handling of edge cases in parallel::splitindices(). (PR#15552)
- HTML display of results from help.search() and ?? sometimes contained badly constructed links.
- c() and related functions such as unlist() converted raw vectors to invalid logical vectors. (PR#15535)
- (Windows only) When a call to system2() specified one of stdin, stdout or stderr to be a file, but the command was not found (e.g., it contained its arguments, or the program was not on the PATH), it left the file open and unusable until R terminated. (Reported by Mathew McLean.)
- The bmp() device was not recording res = NA correctly: it is now recorded as 72 ppi.
- Several potential problems with compiler-specific behaviour have been identified using the 'Undefined Behaviour Sanitizer' in conjunction with the clang compiler.
- hcl() now honours NA inputs (previously they were mapped to black).
- Some translations in base packages were being looked up in the main catalog rather than that for the package.
- As a result of the 3.0.2 change about 'the last second before the epoch', most conversions which should have given NA returned that time. (The platforms affected include Linux and OS X, but not Windows nor Solaris.)

• rowsum() has more support for matrices and dataframes with 2^{31} or more elements. (PR#15587)

- predict(<lm object>, interval = "confidence", scale = <something>) now works. (PR#15564)
- The bug fix in 3.0.2 for PR#15411 was too aggressive, and sometimes removed spaces that should not have been removed. (PR#15583)
- Running R code in a **tcltk** callback failed to set the busy flag, which will be needed to tell OS X not to 'App Nap'.
- The code for date-times before 1902 assumed that the offset from GMT in 1902 was a whole number of minutes: that was not true of Paris (as recorded on some platforms).
- Using Sys.setlocale to set LC_NUMERIC to "C" (to restore the sane behavior) no longer gives a warning.
- deparse() now deparses complex vectors in a way that re-parses to the original values. (PR#15534, patch based on code submitted by Alex Bertram.)
- In some extreme cases (more than 10¹⁵) integer inputs to dpqrxxx() functions might have been rounded up by one (with a warning about being non-integer). (PR#15624)
- Plotting symbol pch = 14 had the triangle upside down on some devices (typically screen devices). The triangle is supposed to be point up. (Reported by Bill Venables.)
- getSrcref() did not work on method definitions if rematchDefinition() had been used.
- KalmanForecast(fast = FALSE) reported a (harmless) stack imbalance.
- The count of observations used by KalmanRun() did not take missing values into account.
- In locales where the abbreviated name of one month is a partial match for the full name of a later one, the %B format in strptime() could fail. An example was French on OS X, where 'juin' is abbreviated to 'jui' and partially matches juillet. Similarly for weekday names.
- pbeta(x, a, b, log.p = TRUE) sometimes underflowed to zero for very small and very differently sized a, b. (PR#15641)
- approx() and approxfun() now handle infinite values with the "constant" method. (PR#15655)
- stripchart() again respects reversed limits in xlim and ylim. (PR#15664)



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 2⁵³ (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.)
- 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.
- R CMD check analyses ::: calls.

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 when R is installed.

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)

- 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.



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 2³¹ 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 ≥ 2³¹: 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.
- 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 time zone 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.
- \bullet 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 2³¹ 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 2³⁵ 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().
- 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 @<- 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.
- 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 'doc' directory. News in HTML format for R versions from 2.10.0 to 2.15.3 is in 'doc/html/NEWS.2.html'.