# NEWS for R version 4.0.0 (2020-04-24)

NEWS

R News



# CHANGES IN 4.0.0

## SIGNIFICANT USER-VISIBLE CHANGES:

- Packages need to be (re-)installed under this version (4.0.0) of R.
- matrix objects now also inherit from class "array", so e.g., class(diag(1)) is c("matrix", "array"). This invalidates code incorrectly assuming that class(matrix\_obj)) has length one.
  - S3 methods for class "array" are now dispatched for matrix objects.
- There is a new syntax for specifying *raw* character constants similar to the one used in C++: r"(...)" with ... any character sequence not containing the sequence ')". This makes it easier to write strings that contain backslashes or both single and double quotes. For more details see ?Quotes.
- R now uses a 'stringsAsFactors = FALSE' default, and hence by default no longer converts strings to factors in calls to data.frame() and read.table().
  A large number of packages relied on the previous behaviour and so have needed/will need updating.
- The plot() S3 generic function is now in package base rather than package graphics, as it is reasonable to have methods that do not use the graphics package. The generic is currently re-exported from the graphics namespace to allow packages importing it from there to continue working, but this may change in future.
  - Packages which define S4 generics for plot() should be re-installed and package code using such generics from other packages needs to ensure that they are imported rather than rely on their being looked for on the search path (as in a namespace, the base namespace has precedence over the search path).

## REFERENCE COUNTING:

- Reference counting is now used instead of the NAMED mechanism for determining when objects can be safely mutated in base C code. This reduces the need for copying in some cases and should allow further optimizations in the future. It should help make the internal code easier to maintain.
  - This change is expected to have almost no impact on packages using supported coding practices in their C/C++ code.

# **MIGRATION TO PCRE2:**

• This version of R is built against the PCRE2 library for Perl-like regular expressions, if available. (On non-Windows platforms PCRE1 can optionally be used if PCRE2 is not available at build time.) The version of PCRE in use can be obtained *via* extSoftVersion(): PCRE1 (formerly known as 'PCRE') has versions <= 8, PCRE2 versions >= 10.

- Making PCRE2 available when building R from source is strongly recommended (preferably version 10.30 or later) as PCRE1 is no longer developed: version 8.44 is 'likely to be the final release'.
- PCRE2 reports errors for some regular expressions that were accepted by PCRE1. A hyphen now has to be escaped in a character class to be interpreted as a literal (unless first or last in the class definition). '\R', '\B' and '\X' are no longer allowed in character classes (PCRE1 treated these as literals).
- Option PCRE\_study is no longer used with PCRE2, and is reported as FALSE when that is in use.

#### **NEW FEATURES:**

- assertError() and assertWarning() (in package tools) can now check for *specific* error or warning classes *via* the new optional second argument classes (which is not back compatible with previous use of an unnamed second argument).
- DF2formula(), the utility for the data frame method of formula(), now works without parsing and explicit evaluation, starting from Suharto Anggono's suggestion in PR#17555.
- approxfun() and approx() gain a new argument na.rm defaulting to true. If set to false, missing y values now propagate into the interpolated values.
- Long vectors are now supported as the seq argument of a for() loop.
- str(x) gets a new deparse.lines option with a default to speed it up when x is a large call object.
- The internal traceback object produced when an error is signalled (.Traceback), now contains the calls rather than the deparse() d calls, deferring the deparsing to the user-level functions .traceback() and traceback(). This fulfils the wish of PR#17580, reported including two patch proposals by Brodie Gaslam.
- data.matrix() now converts character columns to factors and from this to integers.
- package.skeleton() now explicitly lists all exports in the 'NAMESPACE' file.
- New function .S3method() to register S3 methods in R scripts.
- file.path() has some support for file paths not in the session encoding, e.g. with UTF-8 inputs in a non-UTF-8 locale the output is marked as UTF-8.
- Most functions with file-path inputs will give an explicit error if a file-path input in a marked encoding cannot be translated (to the native encoding or in some cases on Windows to UTF-8), rather than translate to a different file path using escapes. Some (such as dir.exists(), file.exists(), file.access(), file.info(), list.files(), normalizePath() and path.expand()) treat this like any other non-existent file, often with a warning.
- There is a new help document accessed by help("file path encoding") detailing how file paths with marked encodings are handled.
- New function list2DF() for creating data frames from lists of variables.
- iconv() has a new option sub = "Unicode" to translate UTF-8 input invalid in the 'to' encoding using '<U+xxxx>' escapes.
- There is a new function infoRDS() providing information about the serialization format of a serialized object.

• S3 method lookup now by default skips the elements of the search path between the global and base environments.

- Added an argument add\_datalist(\*,small.size = 0) to allow the creation of a 'data/datalist' file even when the total size of the data sets is small.
- The backquote function bquote() has a new argument splice to enable splicing a computed list of values into an expression, like ,@ in LISP's backquote.
- The formula interface to t.test() and wilcox.test() has been extended to handle one-sample and paired tests.
- The palette() function has a new default set of colours (which are less saturated and have better accessibility properties). There are also some new built-in palettes, which are listed by the new palette.pals() function. These include the old default palette under the name "R3". Finally, the new palette.colors() function allows a subset of colours to be selected from any of the built-in palettes.
- n2mfrow() gains an option asp = 1 to specify the aspect ratio, fulfilling the wish and extending the proposal of Michael Chirico in PR#17648.
- For head(x,n) and tail() the default and other S3 methods notably for vector n, e.g. to get a "corner" of a matrix, has been extended to array's of higher dimension thanks to the patch proposal by Gabe Becker in PR#17652. Consequently, optional argument addrownums is deprecated and replaced by the (more general) argument keepnums. An invalid second argument n now leads to typically more easily readable error messages.
- New function .class2() provides the full character vector of class names used for S3 method dispatch.
- Printing methods(..) now uses a new format() method.
- sort.list(x) now works for non-atomic objects x and method = "auto" (the default) or "radix" in cases order(x) works.
- Where they are available, writeBin() allows long vectors.
- New function deparse1() produces one string, wrapping deparse(), to be used typically in deparse1(substitute(\*)), e.g., to fix PR#17671.
- wilcox.test() enhancements: In the (non-paired) two-sample case, Inf values are treated as very large for robustness consistency. If exact computations are used, the result now has "exact" in the method element of its return value. New arguments tol.root and digits.rank where the latter may be used for stability to treat very close numbers as ties.
- readBin() and writeBin() now report an error for an invalid endian value. The affected code needs to be fixed with care as the old undocumented behavior was to swap endian-ness in such cases.
- sequence() is now an S3 generic with an internally implemented default method, and gains arguments to generate more complex sequences. Based on code from the S4Vectors Bioconductor package and the advice of Hervé Pagès.
- print()'s default method and many other methods (by calling the default eventually and passing ...) now make use of a new optional width argument, avoiding the need for the user to set and reset options("width").
- memDecompress() supports the RFC 1952 format (e.g. in-memory copies of gzip-compressed files) as well as RFC 1950.
- memCompress() and memDecompress() support long raw vectors for types "gzip" and "zx".
- sweep() and slice.index() can now use names of dimnames for their MARGIN argument (apply has had this for almost a decade).

• New function proportions() and marginSums(). These should replace the unfortunately named prop.table() and margin.table(). They are drop-in replacements, but also add named-margin functionality. The old function names are retained as aliases for back-compatibility.

- Functions rbinom(), rgeom(), rhyper(), rpois(), rnbinom(), rsignrank() and rwilcox() which have returned integer since R 3.0.0 and hence NA when the numbers would have been outside the integer range, now return double vectors (without NAs, typically) in these cases.
- matplot(x,y) (and hence matlines() and matpoints()) now call the corresponding methods of plot() and lines(), e.g, when x is a "Date" or "POSIXct" object; prompted by Spencer Graves' suggestion.
- stopifnot() now allows customizing error messages via argument names, thanks to a patch proposal by Neal Fultz in PR#17688.
- unlink() gains a new argument expand to disable wildcard and tilde expansion. Elements of x of value "~" are now ignored.
- mle() in the stats4 package has had its interface extended so that arguments to the negative log-likelihood function can be one or more vectors, with similar conventions applying to bounds, start values, and parameter values to be kept fixed. This required a minor extension to class "mle", so saved objects from earlier versions may need to be recomputed.
- The default for pdf() is now useDingbats = FALSE.
- The default fill colour for hist() and boxplot() is now col = "lightgray".
- The default order of the levels on the y-axis for spineplot() and cdplot() has been reversed.
- If the R\_ALWAYS\_INSTALL\_TESTS environment variable is set to a true value, R CMD INSTALL behaves as if the '--install-tests' option is always specified. Thanks to Reinhold Koch for the suggestion.
- New function R\_user\_dir() in package tools suggests paths appropriate for storing R-related user-specific data, configuration and cache files.
- capabilities() gains a new logical option Xchk to avoid warnings about X11-related capabilities.
- The internal implementation of **grid** units has changed, but the only visible effects at user-level should be
  - a slightly different print format for some units (especially unit arithmetic),
  - faster performance (for unit operations) and
  - two new functions unitType() and unit.psum().

Based on code contributed by Thomas Lin Pedersen.

- When internal dispatch for rep.int() and rep\_len() fails, there is an attempt to dispatch on the equivalent call to rep().
- Object .Machine now contains new longdouble.\* entries (when R uses long doubles internally).
- news() has been enhanced to cover the news on R 3.x and 2.x.
- For consistency, N <-NULL; N[[1]] <-val now turns N into a list also when val) has length one. This enables dimnames(r1)[[1]] <-"R1" for a 1-row matrix r1, fixing PR#17719 reported by Serguei Sokol.
- deparse(..), dump(..), and dput(x,control = "all") now include control option "digits17" which typically ensures 1:1 invertibility. New option control = "exact" ensures numeric exact invertibility via "hexDigits".
- When loading data sets via read.table(), data() now uses 'LC\_COLLATE=C' to ensure locale-independent results for possible string-to-factor conversions.

• A server socket connection, a new connection type representing a listening server socket, is created via serverSocket() and can accept multiple socket connections via socketAccept().

- New function socketTimeout() changes the connection timeout of a socket connection
- The time needed to start a homogeneous 'PSOCK' cluster on 'localhost' with many nodes has been significantly reduced (package parallel).
- New globalCallingHandlers() function to establish global condition handlers. This allows registering default handlers for specific condition classes. Developed in collaboration with Lionel Henry.
- New function tryInvokeRestart() to invoke a specified restart if one is available and return without signaling an error if no such restart is found. Contributed by Lionel Henry in PR#17598.
- str(x) now shows the length of attributes in some cases for a data frame x.
- Rprof() gains a new argument filter.callframes to request that intervening call frames due to lazy evaluation or explicit eval() calls be omitted from the recorded profile data. Contributed by Lionel Henry in PR#17595.
- The handling of \${F00-bar} and \${F00:-bar} in 'Renviron' files now follows POSIX shells (at least on a Unix-alike), so the first treats empty environment variables as set and the second does not. Previously both ignored empty variables. There are several uses of the first form in 'etc/Renviron'.
- New classes argument for suppressWarnings() and suppressMessages() to selectively suppress only warnings or messages that inherit from particular classes. Based on patch from Lionel Henry submitted with PR#17619.
- New function activeBindingFunction() retrieves the function of an active binding.
- New "cairoFT" and "pango" components in the output of grSoftVersion().
- New argument symbolfamily in cairo-based graphics devices and new function cairoSymbolFont() that can be used to provide the value for that argument.

#### Windows:

- Rterm now works also when invoked from MSYS2 terminals. Line editing is possible when command winpty is installed.
- normalizePath() now resolves symbolic links and normalizes case of long names of path elements in case-insensitive folders (PR#17165).
- md5sum() supports UTF-8 file names with characters that cannot be translated to the native encoding (PR#17633).
- Rterm gains a new option '--workspace' to specify the workspace to be restored. This allows equals to be part of the name when opening *via* Windows file associations (reported by Christian Asseburg).
- Rterm now accepts ALT+xxx sequences also with NumLock on. Tilde can be pasted with an Italian keyboard (PR#17679).
- R falls back to copying when junction creation fails during package checking (patch from Duncan Murdoch).

# DEPRECATED AND DEFUNCT:

- Make macro 'F77\_VISIBILITY' has been removed and replaced by 'F\_VISIBILITY'.
- Make macros 'F77', 'FCPIFCPLAGS' and 'SHLIB\_OPENMP\_FCFLAGS' have been removed and replaced by 'FC', 'FPICFLAGS' and 'SHLIB\_OPENMP\_FFLAGS' respectively. (Most make programs will set 'F77' to the value of 'FC', which is set for package compilation. But portable code should not rely on this.)

• The deprecated support for specifying C++98 for package installation has been removed.

- R CMD config no longer knows about the unused settings 'F77' and 'FCPIFCPLAGS', nor 'CXX98' and similar.
- Either PCRE2 or PCRE1 >= 8.32 (Nov 2012) is required: the deprecated provision for 8.20–8.31 has been removed.
- Defunct functions mem.limits(), .readRDS(), .saveRDS(),..find.package(), and .path.package() from package base and allGenerics(), getAccess(), getAll-Methods(), getClassName(), getClassPackage(), getExtends(), getProperties(), getPrototype(), getSubclasses(), getVirtual(), mlistMetaName(), removeMethodsObject(), seemsS4Object(), traceOff(), and traceOn() from methods have been removed.

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

- installChar is now remapped in 'Rinternals.h' to installTrChar, of which it has been a wrapper since R 3.6.0. Neither are part of the API, but packages using installChar can replace it if they depend on 'R >= 3.6.2'.
- Header 'R\_ext/Print.h' defines 'R\_USE\_C99\_IN\_CXX' and hence exposes Rvprintf and REvprintf if used with a C++11 (or later) compiler.
- There are new Fortran subroutines dblepr1, realpr1 and intpr1 to print a scalar variable (gfortran 10 enforces the distinction between scalars and length-one arrays). Also labelpr to print just a label.
- R\_withCallingErrorHandler is now available for establishing a calling handler in C code for conditions inheriting from class error.

# **INSTALLATION** on a UNIX-ALIKE:

- User-set 'DEFS' (e.g., in 'config.site') is now used for compiling packages (including base packages).
- There is a new variant option '--enable-lto=check' for checking consistency of BLAS/LAPACK/LINPACK calls see 'Writing R Extensions'.
- A C++ compiler default is set only if the C++11 standard is supported: it no longer falls back to C++98.
- PCRE2 is used if available. To make use of PCRE1 if PCRE2 is unavailable, configure with option '--with-pcre1'.
- The minimum required version of libcurl is now 7.28.0 (Oct 2012).
- New make target distcheck checks
  - R can be rebuilt from the tarball created by make dist,
  - the build from the tarball passes make check-all,
  - the build installs and uninstalls,
  - the source files are properly cleaned by make distclean.

## **UTILITIES:**

- R --help now mentions the option --no-echo (renamed from --slave) and its previously undocumented short form -s.
- R CMD check now optionally checks configure and cleanup scripts for non-Bourne-shell code ('bashisms').
- R CMD check --as-cran now runs \donttest examples (which are run by example()) instead of instructing the tester to do so. This can be temporarily circumvented during development by setting environment variable \_R\_CHECK\_DONTTEST\_EXAMPLES\_ to a false value.

#### PACKAGE INSTALLATION:

• There is the beginnings of support for the recently approved C++20 standard, specified analogously to C++14 and C++17. There is currently only limited support for this in compilers, with flags such as '-std=c++20' and '-std=c++2a'. For the time being the configure test is of accepting one of these flags and compiling C++17 code

## **BUG FIXES:**

- formula(x) with length(x) > 1 character vectors, is deprecated now. Such use has been rare, and has 'worked' as expected in some cases only. In other cases, wrong x have silently been truncated, not detecting previous errors.
- Long-standing issue where the X11 device could lose events shortly after startup has been addressed (PR#16702).
- The data.frame method for rbind() no longer drops <NA> levels from factor columns by default (PR#17562).
- available.packages() and hence install.packages() now pass their ... argument to download.file(), fulfilling the wish of PR#17532; subsequently, available.packages() gets new argument quiet, solving PR#17573.
- stopifnot() gets new argument exprObject to allow an R object of class expression (or other 'language') to work more consistently, thanks to suggestions by Suharto Anggono.
- conformMethod() now works correctly in cases containing a "&& logic" bug, reported by Henrik Bengtsson. It now creates methods with "missing" entries in the signature. Consequently, rematchDefinition() is amended to use appropriate .local() calls with named arguments where needed.
- format.default(\*,scientific = FALSE) now corresponds to a practically most extreme options(scipen = n) setting rather than arbitrary n = 100.
- format(as.symbol("foo")) now works (returning "foo").
- postscript(...,title = \*) now signals an error when the title string contains a character which would produce corrupt PostScript, thanks to PR#17607 by Daisuko Ogawa.
- Certain Ops (notably comparison such as ==) now also work for 0-length data frames, after reports by Hilmar Berger.
- methods(class = class(glm(..))) now warns more usefully and only once.
- write.dcf() no longer mangles field names (PR#17589).
- Primitive replacement functions no longer mutate a referenced first argument when used outside of a complex assignment context.
- A better error message for contour(\*,levels = Inf).
- The return value of contourLines() is no longer invisible().
- The Fortran code for calculating the coefficients component in lm.influence() was very inefficient. It has (for now) been replaced with much faster R code (PR#17624).
- cm.colors(n) etc no longer append the code for alpha = 1, "FF", to all colors. Hence all eight \*.colors() functions and rainbow() behave consistently and have the same non-explicit default (PR#17659).
- dnorm had a problematic corner case with sd == -Inf or negative sd which was not flagged as an error in all cases. Thanks to Stephen D. Weigand for reporting and Wang Jiefei for analyzing this; similar change has been made in dlnorm().

• The optional iter.smooth argument of plot.lm(), (the plot() method for lm and glm fits) now defaults to 0 for all glm fits. Especially for binary observations with high or low fitted probabilities, this effectively deleted all observations of 1 or 0. Also, the type of residuals used in the glm case has been switched to "pearson" since deviance residuals do not in general have approximately zero mean.

- In plot.lm, Cook's distance was computed from unweighted residuals, leading to inconsistencies. Replaced with usual weighted version. (PR#16056)
- Time-series ts(\*,start,end,frequency) with fractional frequency are supported more consistently; thanks to a report from Johann Kleinbub and analysis and patch by Duncan Murdoch in PR#17669.
- In case of errors mcmapply() now preserves attributes of returned "try-error" objects and avoids simplification, overriding SIMPLIFY to FALSE. (PR#17653)
- as.difftime() gets new optional tz = "UTC" argument which should fix behaviour during daylight-savings-changeover days, fixing PR#16764, thanks to proposals and analysis by Johannes Ranke and Kirill Müller.
- round() does a better job of rounding "to nearest" by measuring and "to even"; thanks to a careful algorithm originally prompted by the report from Adam Wheeler and then others, in PR#17668.
  - round(x,dig) for negative digits is much more rational now, notably for large |dig|.
- Inheritance information on S4 classes is maintained more consistently, particularly in the case of class unions (in part due to PR#17596 and a report from Ezra Tucker).
- is() behaves more robustly when its argument class2 is a classRepresentation object.
- The warning message when attempting to export an nonexistent class is now more readable; thanks to Thierry Onkelinx for recognizing the problem.
- choose() misbehaved in corner cases where it switched n -k for k and n was only nearly integer (report from Erik Scott Wright).
- mle() in the stats4 package had problems combining use of box constraints and fixed starting values (in particular, confidence intervals were affected).
- Operator ? now has lower precedence than = to work as documented, so = behaves like <- in help expressions (PR#16710).
- smoothEnds(x) now returns integer type in both cases when x is integer, thanks to a report and proposal by Bill Dunlap PR#17693.
- The **methods** package does a better job of tracking inheritance relationships across packages.
- norm(diag(c(1,NA)),"2") now works.
- subset() had problems with 0-col dataframes (reported by Bill Dunlap, PR#17721).
- Several cases of integer overflow detected by the 'undefined behaviour sanitizer' of clang 10 have been circumvented. One in rhyper() may change the generated value for large input values.
- dotchart() now places the y-axis label (ylab) much better, not overplotting labels, thanks to a report and suggestion by Alexey Shipunov.
- A rare C-level array overflow in chull() has been worked around.
- Some invalid specifications of the day-of-the-year (via %j, e.g. day 366 in 2017) or week plus day-of-the-week are now detected by strptime(). They now return NA but give a warning as they may have given random results or corrupted memory in earlier versions of R.
- socketConnection(server = FALSE) now respects the connection timeout also on Linux.

• socketConnection(server = FALSE) no longer leaks a connection that is available right away without waiting (e.g. on 'localhost').

- Socket connections are now robust against spurious readability and spurious availability of an incoming connection.
- blocking = FALSE is now respected also on the server side of a socket connection, allowing non-blocking read operations.
- anova.glm() and anova.glmlist() computed incorrect score (Rao) tests in no-intercept cases. (André Gillibert, PR#17734)
- summaryRprof() now should work correctly for the Rprof(\*,memory.profiling=TRUE) case with small chunk size (and "tseries" or similar) thanks to a patch proposal by Benjamin Tyner, in PR#15886.
- xgettext() ignores strings passed to ngettext(), since the latter is handled by xngettext(). Thanks to Daniele Medri for the report and all the recent work he has done on the Italian translations.
- data(package = "P") for P in base and stats no longer reports the data sets from package datasets (which it did for back compatibility for 16 years), fixing PR#17730.
- x[[Inf]] (returning NULL) no longer leads to undefined behavior, thanks to a report by Kirill Müller in PR#17756. Further, x[[-Inf]] and x[[-n]] now give more helpful error messages.
- Gamma() family sometimes had trouble storing link name PR#15891

# **BUG FIXES (Windows):**

- Sys.glob() now supports all characters from the Unicode Basic Multilingual Plane, no longer corrupting some (less commonly used) characters (PR#17638).
- Rterm now correctly displays multi-byte-coded characters representable in the current native encoding (at least on Windows 10 they were sometimes omitted, PR#17632).
- scan() issues with UTF-8 data when running in a DBCS locale have been resolved (PR#16520, PR#16584).
- RTerm now accepts enhanced/arrow keys also with ConPTY.
- R can can now be started *via* the launcher icon in a user documents directory whose path is not representable in the system encoding.
- socketConnection(server = FALSE) now returns instantly also on Windows when connection failure is signalled.
- Problems with UTF-16 surrogate pairs have been fixed in several functions, including tolower() and toupper() (PR#17645).

# CHANGES in previous versions

• Older news can be found in text format in files 'NEWS.0', 'NEWS.1', 'NEWS.2' and 'NEWS.3' in the 'doc' directory. News in HTML format for R versions 3.x and from 2.10.0 to 2.15.3 is available at 'doc/html/NEWS.3.html' and 'doc/html/NEWS.2.html'.