

Changes in R

by Tomas Kalibera and Sebastian Meyer

Abstract We present selected changes in the development version of R (referred to as R-devel, to become R 4.4) and provide some statistics on bug tracking activities in 2023.

1 Selected changes in R-devel

R 4.4.0 is due to be released around April 2024. The following gives a selection of changes in R-devel, which are likely to appear in the new release. The summaries below include text contributed by authors of some of the changes: Peter Dalgaard, Martyn Plummer, Brian Ripley, Deepayan Sarkar and Luke Tierney.

- The `anova()` function is used for analysis of variance for linear models and analysis of deviance for generalized linear models (GLMs). Previously the `anova()` function behaved differently for GLMs: it would not show test statistics and p-values by default, instead relying on the user to specify the required test statistic. Thanks to changes to "family" objects already included in R 4.3.0, the `anova()` function can now determine an appropriate default test for comparing two GLMs ("LRT" for families with a fixed dispersion parameter and "F" for families with free dispersion) and will show this along with the associated p-value.
- As part of the process of allowing the use of Rao's score test in connection with `glm()`, the `confint()` method for "glm" objects now allows `test = "Rao"`, as does the underlying `profile()` method. To enable this, the code for these functions, and also the corresponding `plot()` and `pairs()` methods, was copied from the **MASS** package to the R sources before modification. The `pairs()` method has also been revised to better handle the case where only a subset of parameters have been profiled.
- R 4.4.0 will include support for producing single-page HTML reference manuals for an entire package, similar to the PDF reference manuals currently hosted on CRAN package pages. It will also include support for a table of contents in HTML help pages, which is controlled by `options("help.htmltoc")`.
- R 4.3.0 added support for experimenting with alternate object systems by providing the `chooseOpsMethod()` generic for resolving method selection for Ops group generics, and the `nameOfClass()` generic to allow more flexible class representations to be used in `inherits()`. In addition, `@` became an internal generic, `@<` already was. R 4.4.0 will add internal support for bare objects by renaming the S4SXP type to OBJSXP and having `typeof()` return "object" for generic bare objects. For now, generic bare S4 objects are distinguished by having a special bit set; it is hoped that this can eventually be dropped.
- R relies on the system `libiconv` for encoding conversions, especially from UTF-8. Apple replaced completely its `libiconv` in macOS 14 with substantial revisions in 14.1 and 14.2: rather than reporting errors when an exact conversion is not possible, it in almost all cases attempts 'transliteration' so for example permille ("‰") is rendered as "o/oo".
musl (as used by Alpine Linux) has long substituted "****", but we now faced converted strings growing in length. Issues were particularly seen when plotting on pdf() devices and it became clear many package authors had never looked at their graphical output. That suggested that transliteration was a safer route, and now R transliterates if the system `libiconv` has not got there first and so (except in rare cases and under musl) R will give the same PDF output on all platforms.
- `Rprof()`, the sampling profiler in R, now supports profiling in "elapsed" time (a.k.a. wall-clock time, real-time) on Unix in addition to "cpu" time. When profiling in elapsed time, the time advances also while R is waiting on I/O, so it may be preferred for some kinds of analysis in I/O intensive applications. Also, elapsed time profiling is the only one currently supported on Windows, so it is good to have a matching option on Unix.
- R gained initial support for 64-bit ARM hardware on Windows (macOS and Linux machines are already supported). It is already possible to build R and recommended packages from source and they pass their automated checks. Testing and porting of other CRAN packages has been started, with a number of patches contributed to package maintainers. This effort uses an experimental LLVM-based toolchain with the new `flang` compiler, which has been added to `Rtools`. In addition to actually supporting 64-bit ARM Windows machines, which are still rare but emerging, this effort also drives portability improvements of R and R packages. Previously, a lot of this code explicitly or implicitly assumed GCC compilers and Intel CPUs on Windows.

- The R CMD check utility for package development performs some additional checks on R documentation (Rd) files. The most prominent addition (in the sense that over 3000 CRAN packages were affected) is a new note about “lost braces”. In (LaTeX-like) Rd syntax, braces are used to mark arguments and otherwise group tokens; they must be escaped as `\{` and `\}` to be included literally in normal text. The new check tries hard to report relevant mistakes, for example:
 - `code{...}`: missing backslash in front of the macro name
 - `{1, 2}`: in-text set notation, where the braces need escaping or the whole expression needs to be put inside a math `\eqn{}`
 - `\itemize{ ... \item{label}{description} ... }`: Rd code meant as a description list with initial labels; this needs `\describe` instead of `\itemize`, otherwise the element becomes “labeldescription” because an `\itemize \item` does not take any arguments.
- A new binary infix operator `%|%` is defined in **base**. This is the so-called *null coalescing operator*: `x %|% y` expresses “use `x` if not `NULL`, otherwise use `y`”.
- `is.atomic(NULL)` now returns `FALSE` and thus behaves according to the R language definition of an atomic vector (`RShowDoc("R-lang")`, Section 2.1.1), which covers the six basic types logical, integer, double, complex, character and raw. For historical reasons (compatibility with S), `is.atomic(NULL)` gave `TRUE` in `R < 4.4.0`, treating `NULL` loosely as “any vector of size 0”. Similarly, `NCOL(NULL)` returned 1 but now gives 0.
- There is a new startup option `--max-connections` to set the maximum number of connections for the R session. It defaults to 128 as before. Values up to 4096 are allowed, but resource limits may in practice restrict to smaller values. This enables advanced users to configure R in environments where a large number of connections (e.g., network) is needed.
- R 4.4.0 on recent Windows will use the new Segment Heap allocator provided by the system. This new allocator has slightly better performance on some applications than the default Low Fragmentation Heap allocator, with the hope that it would be further improved in future versions of Windows.
- R makes use of a system `libdeflate` library if available, in preference to the system `libz` library. This can speed up decompressing R objects in lazy-loading databases and other operations.

See the `NEWS.Rd` file in the R sources for a more complete list; nightly rendered versions are available at <https://CRAN.R-project.org/doc/manuals/r-devel/NEWS.html> with RSS feeds at <https://developer.R-project.org/RSSfeeds.html>.

2 Bug statistics for 2023

Summaries of bug-related activities over the past year were derived from the database underlying R’s [Bugzilla system](#). Overall, 186 new bugs or requests for enhancements were reported, 204 reports were closed, and 942 comments were added by a total of 120 contributors. The numbers of new reports and contributors were comparable to 2022, but comments increased by 8% and closures by 20%. Higher activity in 2023 was driven by a dedicated effort in reviewing and discussing open reports during the R Project Sprint at the University of Warwick, UK, 30 August to 1 September (Turner and Becker 2023).

Figure 1 shows the monthly numbers of new reports, closures and comments in 2023. Comment activity was relatively low in July and peaked in September due to the sprint.

The top 5 components reporters have chosen for their reports were “Low-level”, “Misc”, “Language”, “Documentation”, and “Accuracy”. 9% of the reports were suggestions for enhancements that were submitted either in the “Wishlist” component or in a specific component but with severity level set to “enhancement”.

References

Turner, Heather, and Gabriel Becker. 2023. “R Project Sprint 2023.” *The R Journal* 15: 299–305. <https://journal.R-project.org/news/RJ-2023-3-sprint>.

Tomas Kalibera
R Core Team
Prague, Czechia
ORCID: 0000-0002-7435-734X
Tomas.Kalibera@R-project.org

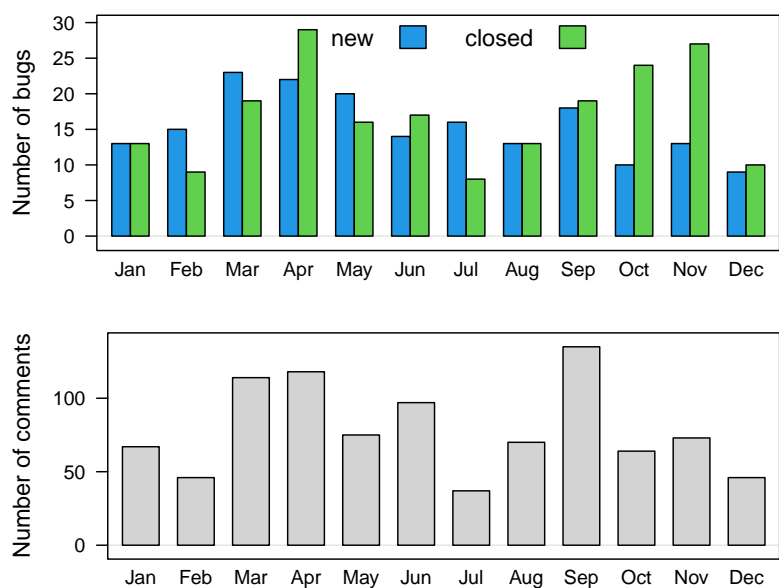


Figure 1: Bug tracking activity by month in 2023.

Sebastian Meyer
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany
ORCID: [0000-0002-1791-9449](https://orcid.org/0000-0002-1791-9449)
Sebastian.Meyer@R-project.org