R Project Sprint 2023

by Heather Turner

Abstract R Project Sprint 2023 was a three-day event at the University of Warwick, UK, that brought together novice and experienced contributors to work alongside members of the R Core Team. 55 members of the R community participated, with external contributors selected to balance technical expertise and provide opportunities for members of historically under-represented groups. Participants worked collaboratively on contributions to base R and on infrastructure supporting contribution. Several small tasks were completed within the duration of the sprint, whilst significant steps were made on larger projects. The event provided a unique opportunity for external contributors to learn about the R development process and to develop their contribution skills.

Introduction

R Project Sprint 2023 was a three-day event hosted at the University of Warwick, UK. The aim of the event was to bring novice and experienced contributors together to work collaboratively with member of the R Core Team - the developers that maintain and develop the code and documentation that forms the base distribution of R ("base R").

Participants

All members of the R Core Team were invited to the event and 11 members were able to participate. Another 13 participants were invited/pre-selected - these included local organizers, representatives from sponsors, and experienced contributors. The remaining 31 participants were selected on the basis of self nomination through an application form (a couple more were selected but ultimately unable to participate). Figure 1 shows group photos taken on Day 2 and and Day 3 of the sprint. Participation was in-person by default, but exceptions were made in a few cases where travel was not possible, e.g., due to visa issues. The number online was higher than anticipated due to travel disruptions; in the end 7 participated online.

Members of demographic groups underrepresented within the contributor community were encouraged to apply for a place, by promoting the event to affinity groups (R-Ladies, MiR, RainbowR, AfricaR, ArabR, AsiaR, and LatinR) and by direct communication with potential participants. Figure 2 shows the geographical distribution of all 55 participants. There were 16 from Europe with 8 from the UK; 13 from North America with 12 from the USA; 7 from Asia with 5 from India; 6 from Latin America with 3 from Argentina; 5 from Africa with 2 from Nigeria; 4 from Oceania - all from New Zealand, and 3 from the Middle East.

We have further information from the nomination form, which was completed by 40 of the 44 invited/selected contributors. Over half (25/40) self-identified as belonging to one or more underrepresented groups. Figure 3 summarises the skills of these contributors as assessed by the selection committee, using data from the nomination forms. A "contributor level" was assigned based on self-ratings of familiarity with relevant concepts and processes, along with answers to free text questions about the applicant's experience and motivation. The committee deliberately selected participants to achieve the balance shown in the first plot of Figure 3: an equal number of advanced and novice contributors, with the remainder have intermediate expertise. The second plot summarises the potential for contribution to translations: 14 of the selected contributors expressed a specific interest in translation; 8 more were surmised to have potential based on country of residence, for the remainder (22) there was no evidence as we did not ask about this explicitly in the form.

Preparation

There were two sides to preparation for the sprint: gathering suitable tasks to work on and helping participants brush up their knowledge and skills.

Anyone with an interest in R development was encouraged to suggest ideas for suitable tasks via the discussion forum on the R Project Sprint 2023 GitHub repository (GitHub Discussions). This provided a space for participating members of R Core to give feedback and for participating contributors to express an interest. Sprint participants could propose a project by adding a page to the Projects section of the sprint website. In the run up to the sprint, ideas and projects were transferred to issues on sprint GitHub repository along with further last-minute ideas from core developers and members of the R Contribution Working Group. This enabled participants to assign themselves to issues and provided a way to track tasks during the sprint.

Figure 1: Photos of sprint participants on Day 2 (top) and Day 3 (bottom), including online participants on screen (not all participants photographed).

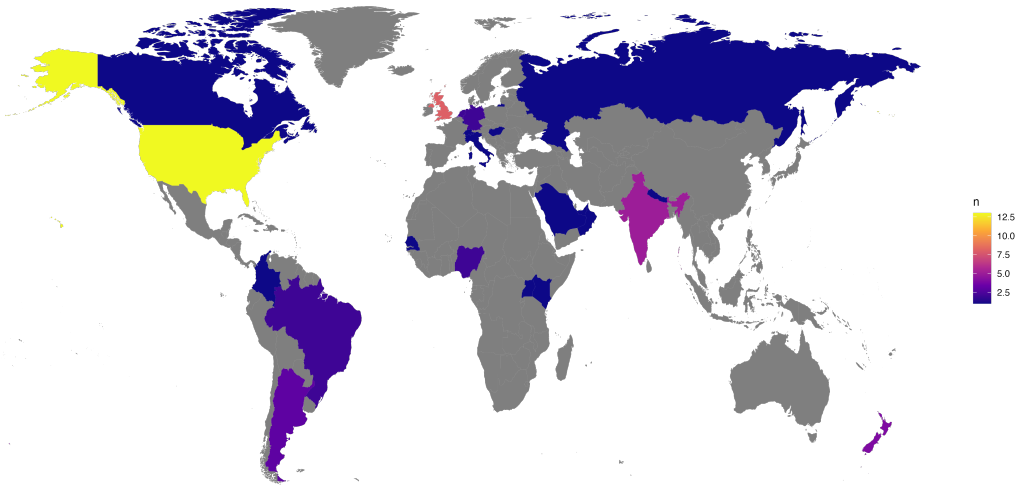

Figure 2: Chloropleth showing the distribution of participants on the world map

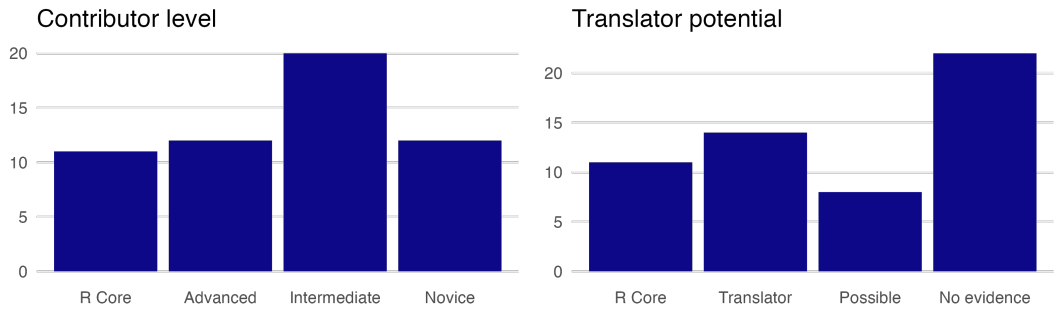

Figure 3: Skills of external contributors as judged by the selection committee. Left: level of expertise in R contribution. Right: potential as a translator of English to other languages.

To help those with less experience prepare for the sprint, participants were pointed to resources created by the R Core Team and the R Contribution Working Group (RCWG), including the R Blog post on reviewing bugs and the useR! 2021 tutorials on analysing bugs/contributing patches and translating messages in R. In addition, participants were encouraged to engage with relevant events, in particular the Debugging in R tutorial run by Shannon Pileggi for R-Ladies Remote, and the C Book Club for R Contributors and R Contributor Office Hours run by the RCWG.

It was expected that participants would be able to build R from source on the laptop they brought to the sprint. People new to this were pointed to the R-admin manual, the R Dev Guide and the prototype GitHub Codespace which provides a virtual environment in which to build R - this was demonstrated in one of the contributor office hours.

Format

The sprint began with a hybrid evening welcome event where Martyn Plummer gave some opening remarks on contributing to the R Project, then participants split into small groups to chat to a member of R Core. This was followed by an informal drinks reception for in-person participants.

Each sprint day started with a kick-off session and ended with a report-back session, both hybrid to include our online participants. On the first day, R Core members gave short talks in these sessions, giving a taste of their work for the R Project. Otherwise the sessions were use to allocate people to tasks and summarise progress.

The remaining day time was spent working in small groups, sometimes arranging hybrid meetings to discussion specific issues.

On the second evening, in-person participants enjoyed a conference dinner, whilst on the final evening the sprint participants joined the Warwick R User Group for a hybrid meetup to share progress at the sprint, this was followed by a buffet dinner for in-person participants.

Translation

A key activity during the sprint was work on translating English strings for localization of R. In 2022, Gergely Daróczi set up a Weblate instance at `https://translate.rx.studio` that provides a user-friendly interface for contributing translations to the R Project. Sprint participants created a new set of guidelines for translators and a new section in the R Dev Guide on How to contribute new translations. Several new features were enabled on the Weblate instance, including translation memory, hyperlinking to the source string location and dedicated reviewers to approve translations. New components were added, so that the instance not only covers base R (messages, warnings, errors and the Windows GUI), but also the Mac GUI and recommended packages.

Figure 4 gives a summary of activity on Weblate during the sprint: around 2000 messages were changed over 14 languages. The vast majority of this activity can be attributed to the sprint directly or indirectly - the Hungarian translations were imported from earlier work in 2011 and the Turkish translations were made by external contributors after the Mac GUI component was added.

Code and Documentation

The remaining activity at the sprint related to code and documentation in base R. Code issues were split into topics to help organize work groups: accessibility, graphics, packages, statistics, translation and low-level. The translation issues here related to infrastructure maintain by the R Core Team, as opposed to Weblate. The low-level topic was a catch-all that covered utility functions and/or issues that required advanced technical expertise, e.g., in C.

Figure 5 shows the progress of issues at the end of the sprint and two months after. An issue is considered closed if a corresponding bug report on R's Bugzilla (`https://bugs.r-project.org`) was closed, if a corresponding patch was committed to base R, or if the issue was closed by an update to a CRAN package. By the end of the sprint, ten issues had been closed. Seven of these were documentation bugs, including one that was closed just before the sprint due to a participant reviewing issues in preparation. However, progress had been made on thirty-four other issues, ranging from discussing the issue, through defining a roadmap, to work in progress or proposing a patch. Two months after the sprint, another twelve issues had been closed and six more had progressed status (e.g., from roadmap to work in progress). These eighteen issues included three that were not started at the sprint, but worked on soon after as follow-up to a partial fix or due to participants reviewing the progress of sprint issues.

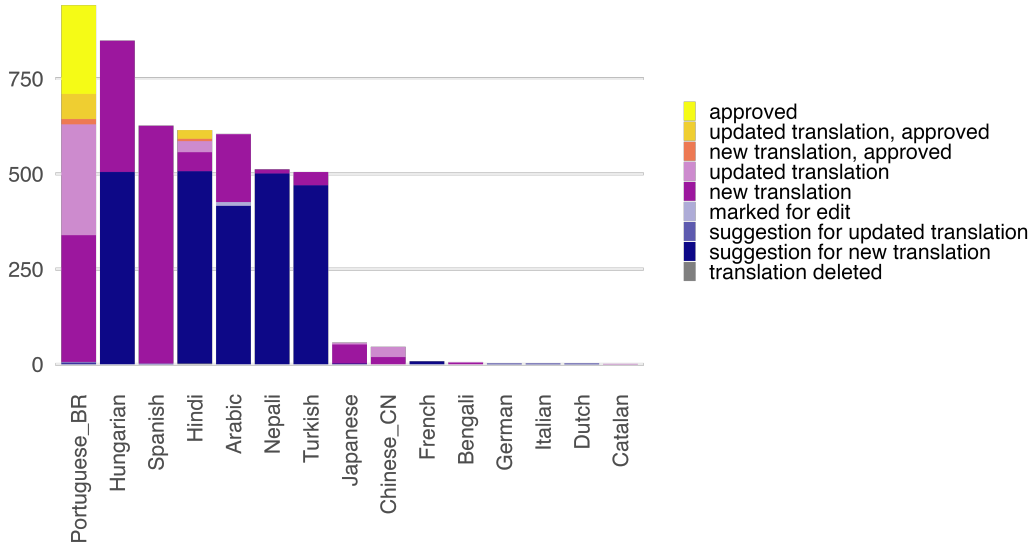

Figure 4: Changes in the R Project components on Weblate during the three days of the sprint

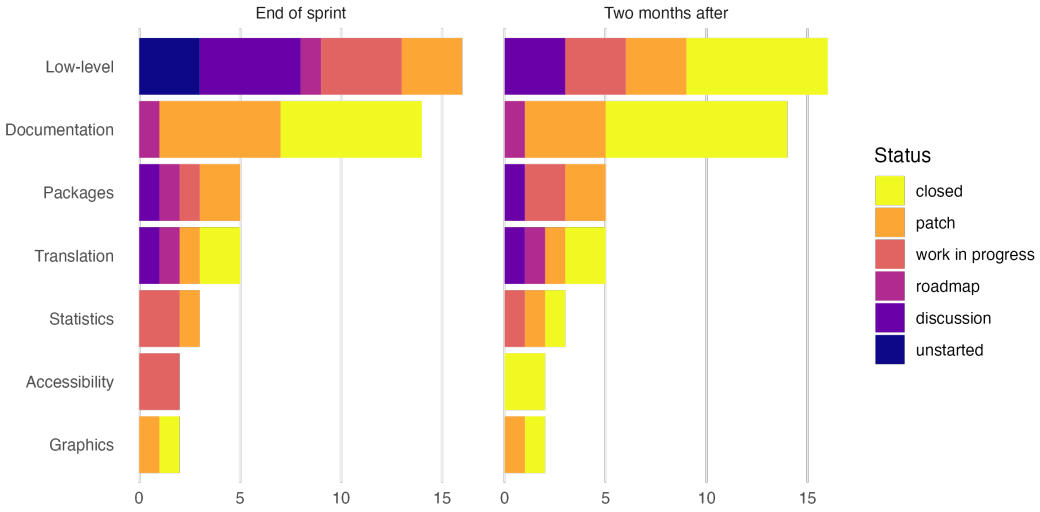

Figure 5: Status of issues at the end of the sprint and two months after

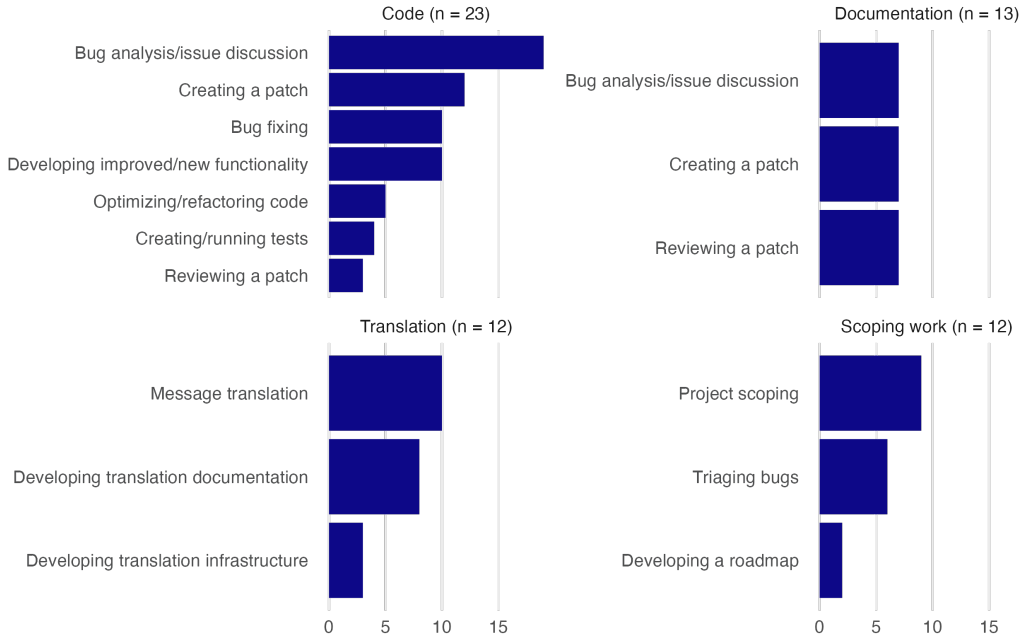

Figure 6: Activities of external contributors, based on 32 responses to participant survey.

The low-level issues included new functionality, e.g., supporting custom parallel backends; refactoring, e.g., improving the speed of scalar random number generation; improving behaviour, e.g., better formating of complex numbers and bug fixes, e.g. managing long names when creating tarballs.

Participants working on documentation began by triaging all open documentation bugs on Bugzilla to identify ones that could be closed without fixing, or ones that appeared straight-forward to patch, hence the high closure rate for these issues. Some closed bugs had been open for several years.

Package-related issues included adding support for defining vignette order, improving messages to CRAN maintainers, and caching installed packages.

Translation-related issues included identifying untranslated strings in the R source files, and creating a roadmap towards internationalization of help pages. The R Consortium are funding a project by Elio Campitelli and Renata Hirota for prototyping work as a first step on this roadmap.

Statistics issues included improving the behaviour of `t.test.formula()` and `wilcox.test.formula()` for paired tests and enhancing `sample.int()` for unequal probability sampling, for which a prototype package was developed after the sprint for testing.

Accessibility focused on two issues faced by screenreader users: logging base graphics and logging R sessions. Functions resulting from this work are now implemented in **BrailleR**.

Finally there were two issues related to graphics, one fixed during the sprint implementing 3-digit hex colors and one larger project on adding alpha masks to the Quartz graphics device.

There were more issues prepared for the sprint than are summarised here, but they were not taken up at the sprint. In some cases there was insufficient support from R Core to pursue the idea, or it was considered out of scope for the sprint, or there were no available participants with relevant skills to take the idea forward. Often participants were interested in multiple issues and were encouraged to favour issues/topics where larger group discussions were taking place, to take advantage of everyone being together.

Participant experience

As well as aiming to make progress on contributions to the R Project, the sprint was intended to develop participants' knowledge and experience in contribution and motivate them to continue contributing after the sprint.

Figure 6 summarises the activities engaged in at the sprint, for 32 out of 44 external contributors that responded to a post-sprint survey. Around two-thirds were involved in working on code issues and around a third worked on documentation and/or translation. Scoping work was also an important activity, that a third of contributors engaged in.

Figure 7 summarises the activities that contributors engaged in for the first time either during the

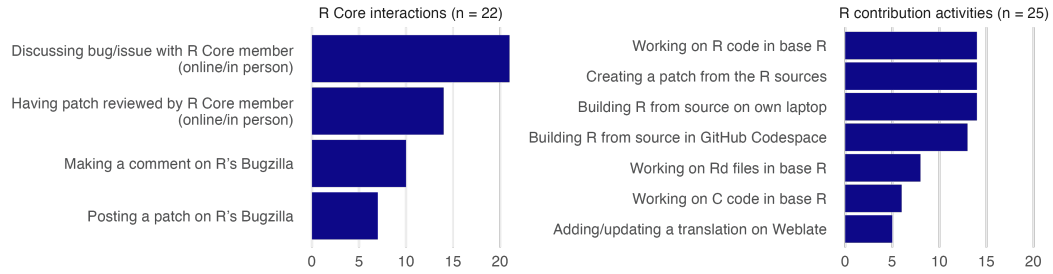

Figure 7: R core interactions and R contribution activities that external contributors engaged in for the first time, based on 32 responses to participant survey.

sprint, or to prepare for or follow up on work done at the sprint. For around two-thirds of contributors it was the first time they had discussed a bug or issue with an R Core member, whether online or in person. About a third commented on Bugzilla for the first time and around a quarter posted their first patch. Around half the contributors built R from source for the first time, either on their own laptop or in the GitHub Codespace (or both) and for about half the contributors it was their first time working on R, C, or Rd files in base R.

Organizers and sponsors

The organization of the sprint was lead by Heather Turner, as part of a research fellowship funded by the UK Engineering and Physical Sciences Research Council. This fellowship provided core funding and was supplemented by sponsorship funding:

- Platinum Sponsor (R Core travel): the R Foundation.
- Gold sponsors (evening events, participant travel): the R Consortium, the Centre for Research in Statistical Methodology Warwick University, posit.
- Silver Sponsors (participant travel): Seminar for Statistics ETH Zurich, Rx Studio, The Prostate Cancer Clincal Trials Consortium, and Google.

Through this sponsorship, travel, accommodation and subsistence was provided for all participants.

Martyn Plummer and Ella Kaye completed the local organizer team. Members of the RCWG helped with the planning, especially Gabe Becker who joined Heather and Ella on the selection committee and helped gather issues in the run up to the sprint.

Summary

R Project Sprint 2023 was a very collaborative event, where external contributors had a unique opportunity to work closely with R Core members. Good progress was made across a broad range of issues with continued impact after the sprint. The feedback from both R Core and external participants was very positive, e.g.,

> Thank you for organizing an incredible sprint and creating space for newcomers
>
> There were many different parts that contributed so well to make it very productive, envigourating, and motivating
>
> From arrival to departure, everything was seamless and I had a great time discovering what it takes to maintain R.
>
> I'm exhausted but also super excited by all the work we did and that I take as homework.

Several participants - as well as R community members that could not attend this time - asked when we would hold a repeat event. Finding funding for ~50 people from around the world to attend a 3-day sprint is quite a challenge. So in the short term we plan to run 1-day events in collaboration with in-person conferences. Whilst this will limit the scope of tasks that can be tackled, we can benefit from people already travelling for the conference, with conference scholarship schemes helping to support inclusion.

Links

- Sprint website: https://contributor.r-project.org/r-project-sprint-2023/
- GitHub repository: https://github.com/r-devel/r-project-sprint-2023

Heather Turner
University of Warwick
Coventry, United Kingdom
https://warwick.ac.uk/heatherturner
ORCiD: 0000-0002-1256-3375
h.turner.1@warwick.ac.uk