The R Project for Statistical Computing

Google Summer of Code Proposal

CI Optimization for R Package Performance Testing

Sagnik Mandal

January 25, 2025 v1.0

Contents

1. Project Info

Project title: Optimizing a performance testing workflow by reusing minified R package versions between CI runs

Short Title: CI Optimization for R Package Performance Testing

Idea Page: Idea Description on the R GSoC Wiki

2. Bio

I am Sagnik Mandal, a sophomore pursuing an Integrated Dual Degree in Materials Science and Technology at IIT (BHU), Varanasi. I have been a self-taught programmer since over 6 years now, and have also contributed to various open source projects, and used to maintain a couple of packages on the Arch Linux User Repository (AUR).

I first learnt about R, when I was following a online course on "Dealing with materials data : collection, analysis and interpretation", this course discussed data analysis using R for materials science. I have also been working under a professor at my institute on a project that involves data analysis and machine learning for material science applications.

I have used Github Actions in some of my personal projects, last year I applied to Joplin for GSoC to work on a similar project where I had planned to use Github Actions to automatically fetch build Joplin Plugins, along with preventing malicious code from being executed. While the project was shortlisted by Joplin, I couldn't make it to the final list of selected students, but this experience has helped me understand the working of Github Actions.

Other information about me can be found on my resume.

3. Contact Information

Name: Sagnik Mandal

Postal Address: 503, Satish Dhawan Hostel, IIT (BHU), Varanasi, Uttar Pradesh, India, 221005 (Timezone: UTC+5.5°)

Telephone(s): +91-7470989815°

Email(s): sagnik.mandal.mst23@iitbhu.ac.in°, acriticalcynic@outlook.com°

Other communications channels: Google Meet, Zoom°, Discord°, WhatsApp°

4. Affiliation

Institution: Indian Institute of Technology (Banaras Hindu University), Varanasi, India

Program: B.Tech. & M.Tech. (Integrated Dual Degree) in Materials Science and Technology

Stage of Completion: Sophomore, expected graduation in 2028

Contact to Verify: Dr. Chandan Upadhyay° (*email: cupadhyay.mst@iitbhu.ac.in°*)

5. Schedule Conflicts

I have my End Semester Examinations from Apr 24 – May 09, 2025, and will most probably be travelling back home within a week after that. This will clash with the first week of the Community Bonding period, but I will be able to work on the project during the rest of the summer without any interruptions. I have already setup my build environment and through the tests I have also tackled parts of the project.

My next semester starts on July 11, 2025, and considering a 175 hour coding period, I will ensure the project is completed before that.

6. Mentors

Evaluating Mentor: Anirban Chetia (anirban166) (*email: ac4743@nau.edu°*)

Co-Mentor(s): Toby Dylan Hocking (tdhock) (*email: toby.hocking@r-project.org°*)

Contact with Mentors: I have been in touch with both Anirban and Toby over email.

7. Coding Plan and Methods

TODO

8. Timeline

TODO

9. Management of Coding Project

TODO

10. Test Submissions

EASY: Script to Minify R Package°

- Wrote a script that is agnostic to the package name and version, and can be used to minify any R package tarball. It also installs the minified package using `R CMD INSTALL`, to ensure that the package is installable.

MEDIUM: Github Action to Minify R Packages and Upload as Artifact°

- Created a GitHub Action that reads the package name and version from the issue description, checks if the package is already minified, and if not, minifies the package and uploads it as an artifact. The action also installs the minified package using `R CMD INSTALL`.
- It then comments on the issue with package size details, and time taken to minify the package.

HARD: Supporting PRs from Forks in Autocomment-atime-results°

- Modified the `Autocomment-atime-results` workflow to support PRs from forks. The workflow is divided into two parts: one that runs the tests and uploads the results as artifacts, and another that downloads the results and comments on the PR with the results. The workflow has been tested to work with PRs from forks and non-forks.
- Then cloned the data.table repository, with all its historical branches required for the atime results, and tested the workflow to work with PRs from forks and non-forks.