

Report Reproducibility in CS Survey

D Gawehns

2024-12-01

Summary of Results

The 57 respondents know about and use tools for reproducible research. The impact, quality and re-usability of research are mentioned by some respondents, which hints at underlying motivators for making research reproducible. Those topics can motivate researchers to work more reproducible and could be explored to design more targeted and sustainable interventions to foster reproducibility in the CS.

Several CS specific tools (e.g., testing proofs or sharing seeds) came up less often than expected. This might be sub-discipline dependent or influenced by how the questions were phrased. As many respondents already share code and data (or consider this important), sharing of computational environments is a logical next step. This seems to not be common practice and a long-hanging fruit for workshops and other interventions.

Limitations: This is a quick summary of the results, sub-discipline specific analyses are following. The questionnaire was filled in by 57 respondents and has a high self-selection bias as respondents are likely to have had some time to think about reproducible research before. The results and conclusions cannot be generalized, but give hints towards interesting topics and themes for tools, workshops and interventions tailored to people working in computer science research.

Background

This report summarizes the answers of respondents on a questionnaire about reproducible research in computer science. The research question was : What do computer scientists understand as reproducible computer science? This small research project was conducted as part of a fellowship at the Netherlands eScience Center. The ethics committee of the science faculty at Leiden University approved this research. There were 57 respondents in total. Respondents were recruited via email lists, linkedin posts and word of mouth. Any researcher affiliated with a research organisation in the Netherlands and considering themselves as working in computer science was invited to participate. The original target of 200 respondents could not be met. Most respondents indicated that they were affiliated with Leiden University.

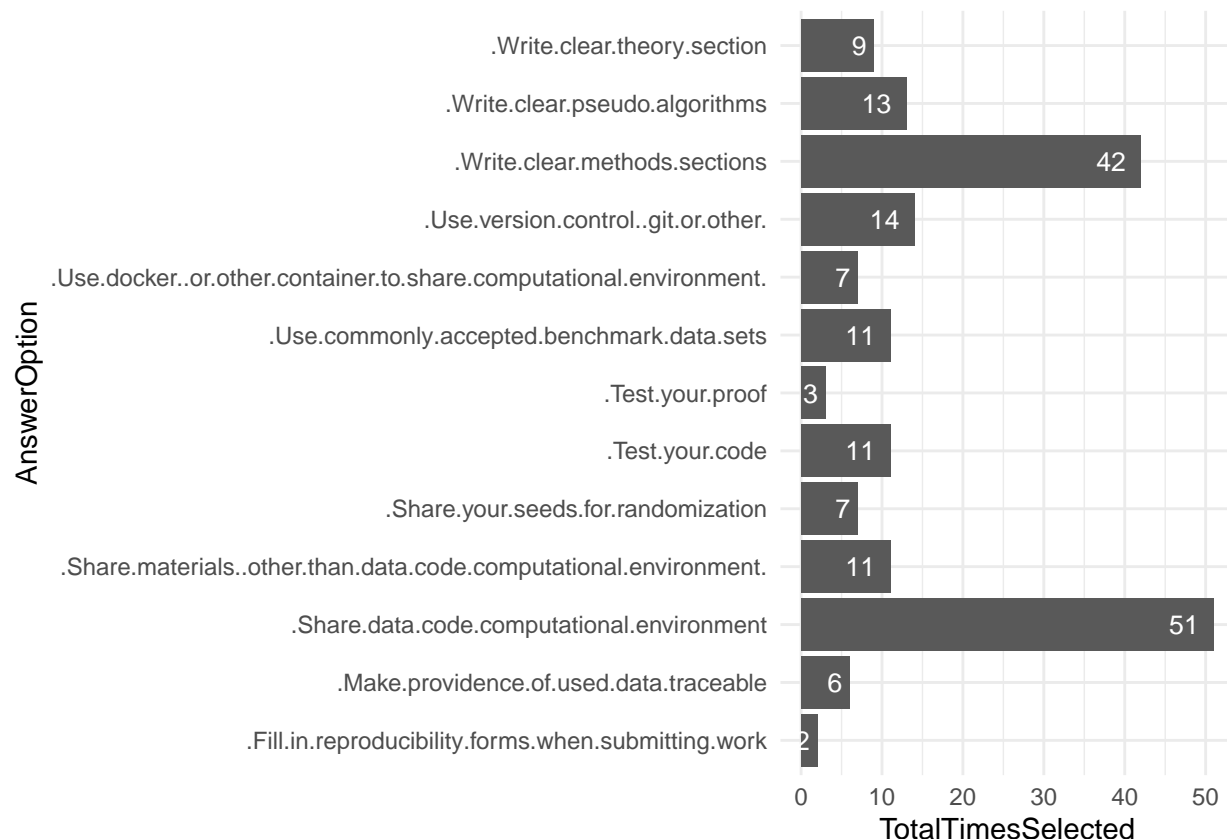
Introduction: What is your first association when you hear the term “Reproducible Research”? (Open Text)

The answers can be categorized into three types: Tools for and, Requirements and Impact of Reproducible Research. Respondents mentioned tools for Reproducible Research such as docker containers or github repositories. Most respondents answered with a requirement for reproducible research. Sharing of materials was mentioned by most respondents, some note that documentation of methods and study setups is a requirement to make a research outputs reproducible. Some respondents mentioned the impact of reproducible research: “Better Science” , ” Robust Science” , ” quality and reputation” and making it possible for others to rerun experiments or studies.

First Question: Which are the most important tools for reproducible research?

From the given options, almost all respondents (51) chose “Sharing of data, code, computational environment”. “Writing a clear method section” was chosen by 42 respondents. Least respondents chose the option “Fill

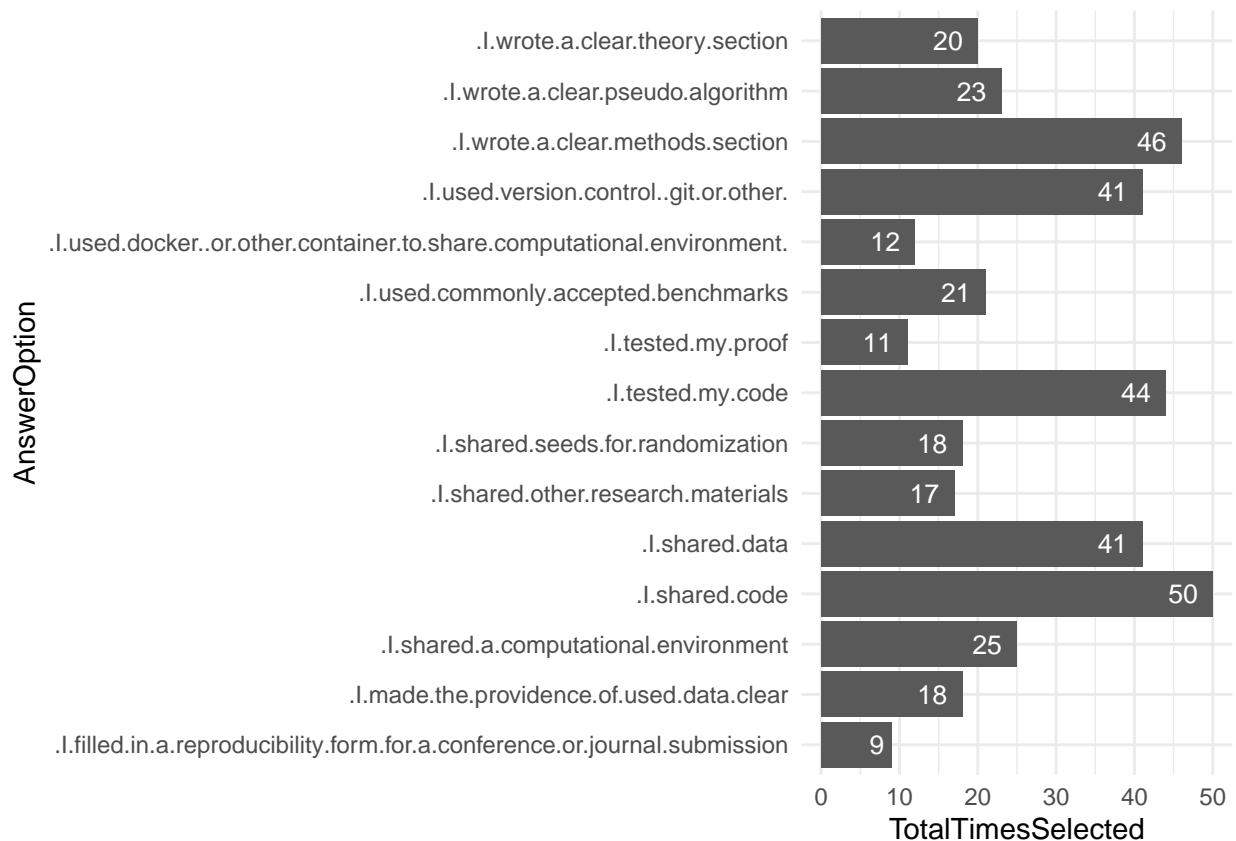
in reproducibility forms”, “Test your proof”, “make data traceable” and ” use docker or others to share computational environment”.



Saving 6.5 x 4.5 in image

Second question: Which tools did you implement?

Almost all respondents shared code and data in the past. Testing code and writing clear methods sections is another common practice as well as using git or other version control tools. Few respondents filled in a reproducibility form for a conference. This might be sub-field dependent. Only 11 respondents tested proofs and 12 shared computational environments. It would be interesting to find out what respondents would need to share their computational environment, given that many already use version control and are open to sharing code and data.



Saving 6.5 x 4.5 in image

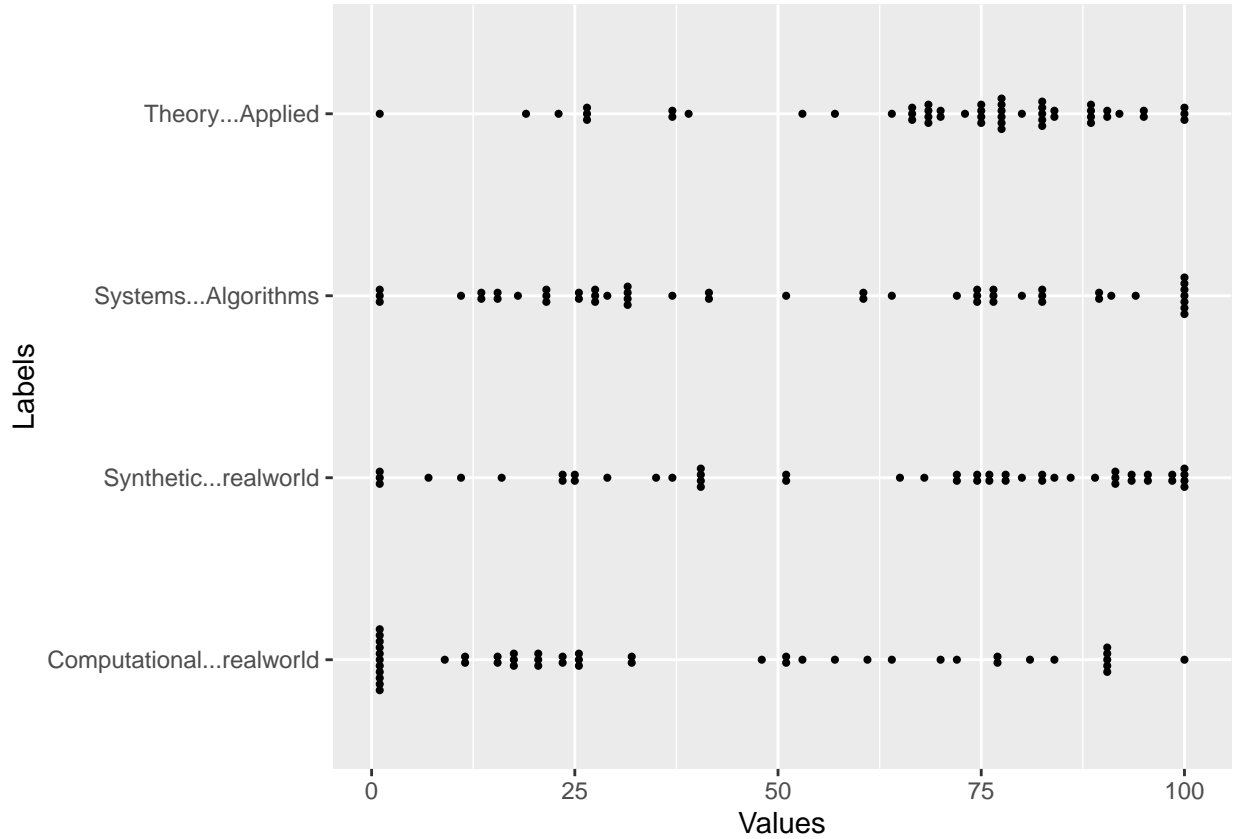
Third question: How could you make your work as least reproducible as possible?

The answer possibilities concerning sharing and accessibility of materials were ticked by most respondents. Not sharing is the easiest way to make reproducibility close to impossible or at least very hard. Similarly, many respondents ticked the option “cherry pick your benchmark - only report successful experiments”, which will obscure and hide relevant information. Writing overly complicated theory sections and referencing obscure sources were selected by 21 and 13 respondents, which underlines that these sections are relevant to computer scientists to not only follow an author’s reasoning but also to build on and reproduce the results of the paper. Tinkering with seeds is something only few respondents chose as effective way to make reproducibility hard, also using many dependencies in code is something respondents didn’t feel as being effective.



Saving 6.5 x 4.5 in image

The last question was about the subfield of CS research the respondents worked in. To find out how much respondents differed in terms of the data they worked with and the methodologies they used, they were asked to use sliders.



The slider questions are based on the assumption that the type of methods, data and work focus influences which reproducibility skills and tools researchers would employ. In broad terms, a theoretical computer scientists is much more like to check a proof than someone working on an application. As a rough summary, respondents work mainly applied and run computational experiments (as opposed to testing hypotheses in the real world or lab). There is a rather even spread when it comes to the use of synthetic versus real world data and an even spread along the continuum of system vs algorithm design.

The slider questions are not validated, nor where they piloted beforehand. The respondents had to fall back to their own understanding of the question without possibility to ask for clarification.

Nevertheless, one can state that many of the respondents work with real world data and some test their hypotheses in the real world. Computer science is not limited to synthetic data and computational experiments. We can also state that a majority of respondents works with an application in mind (to varying degrees) and run computational experiments (here, some respondents chose the extreme slider position indicating a very clear position on the continuum).