

# Track 1 Paper: Surveying the U.S. National Postdoctoral Association Regarding Software Use and Training in Research

Udit Nangia  
National Center for Supercomputing Applications (NCSA)  
University of Illinois Urbana-Champaign  
Urbana, USA  
udit.nangia2@gmail.com

Daniel S. Katz  
NCSA & CS & ECE & iSchool  
University of Illinois Urbana-Champaign  
Urbana, USA  
d.katz@ieee.org

**Abstract**—This paper reports on the results of a 2017 survey conducted by email and web of members of the U.S. National Postdoctoral Association regarding their use of software in research and their training regarding software development. The responses show that 95% of respondents use research software. Of all the respondents, 63% state they could not do their research without research software, 31% could do it but with more effort, and 6% would not find a significant difference in their research without research software. In addition, 54% of respondents have not received any training in software development, though all respondents who develop software for researchers have received either self-taught or formal software development training.

**Keywords**—software use, software training, survey

## I. INTRODUCTION

Research software is essential and/or central to a large fraction of modern research projects [1]. Individual researchers choose to use existing software or to develop new software to best carry out their projects. In recent years, computing researchers have explored the relationship between software development and research in academia. For example, a survey of academic faculty and staff at British universities found that 92% of academics use research software, with 69% saying that their research would not be practical without it. The report also indicated that 56% of respondents develop their own software, and that 21% of them have no training in software development [2, 3]. It is alarming that while a large percentage of research depends on research software and development of this software by academic faculty and staff, a large percentage of software users and a substantial fraction of developers don't have software development training. In addition, a 2008 survey showed that the knowledge required to develop and use scientific software is primarily acquired from peers and through self-study, rather than from formal education and training [4].

This paper expands on the British data, focusing instead on U.S. universities. It is based on a survey conducted through the membership list of the National Postdoctoral Association, which is an organization that advocates and represents postdoctoral scholars in the U.S. research community. Based on the results of this survey, the paper examines and discusses how role in research, gender, and years of experience influence how researchers utilize research software and the training that they have received in developing research software.

## II. METHODOLOGY

The motivation for the work described in this paper is to assess the role of software in conducting research at U.S. universities. This includes fundamental questions of how prevalent software development training is among researcher and how widespread software use is among researchers at different levels and fields.

Questions for the survey were based on the questions asked in a similar survey of British universities [2]. The questions were further modified based on feedback received from conducting a preliminary survey of U.S. universities' faculty and staff. The survey consisted of 14 questions that related to particular aspects of interest in the participants. These included demographic questions, questions assessing the participants' use of research software, and questions evaluating the training that the participants have received in software development.

Respondents who participated in the survey were members of the U.S. National Postdoctoral Association who were present on the organization's mailing list. The survey was distributed through e-mail to the organization's mailing list in March 2017. Members of the organization could input their responses until the end of April 2017. Members participated purely voluntarily, with no external incentives, and their identities were kept anonymous in collecting the data.

Data shown in Table I shows the figures associated with the distribution of the survey. In total, 6281 members of the organization were emailed. Of these, 4473 individuals viewed the email and 209 participants responded to the survey and were included in the analysis. While this is a low response rate (3.33%), it is similar to that of the UK survey [2, 3]. Gizmodo says that response rates can fall below 2% when the respondent population is less-targeted, when contact information is unreliable, or where there is less incentive or little motivation to respond [5]. Because we were unable to provide incentives to respond (other than helping the community) nor were we able to send more than single email request for responses, we are satisfied with this rate, though we recognize there are likely selection effects in our data.

We used descriptive statistical methods to analyze the data obtained from the survey participants. Table II shows the demographic information collected on the participants. Females (55.3%) are more represented than males (42.3%). And individuals conducting research in STEM fields are more

represented in the survey (89.5%) than individuals in non-STEM fields (10.5%).

TABLE I. METRICS ASSOCIATED WITH SURVEY DISTRIBUTION

Total Population Emailed	6281
Total Views	4473
Total Click-Backs	432
Total Responses	209
Response Rate	3.33%

TABLE II. DESCRIPTION OF RESPONDENTS

Gender	Male	Female	Other/Prefer not to say
	42.3%	55.3%	2.4%
Research Discipline	STEM <sup>a</sup>	Non-STEM	Other
	89.5%	10.5%	0%

<sup>a</sup> Science, Technology, Engineering, and Mathematics

### III. RESULTS AND DISCUSSION

We are particularly interested in understanding participating researchers' roles, use of research software, and training in software development.

The initial topic in the survey was concerned with assessing the role the respondents play in conducting research. As shown in Fig. 1, 81% of respondents conduct their own research projects; 50% of respondents support someone else's research; 18% manage researchers or research projects, and only 5% of respondents report developing research software for other

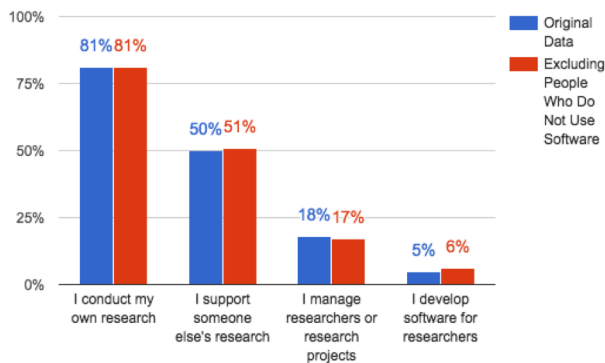


Fig. 1. Assessment of researchers' role between all respondents and respondents excluding those who do not use research software.

researchers. This trend of respondents indicating that they conduct their own research having the largest representation and respondents indicating they develop research software having the lowest response is consistent when the results are stratified by different variables, such as gender and years of experience, as shown in Fig. 2 and Fig. 3 respectively. One distinct feature of the data shown in Fig. 2 is that a greater percentage of men report that they develop software for other researchers than women. Fig. 3 shows that a greater percentage of researchers surveyed who had over 10 years of experience develop software for other researchers. This distinction is particularly pronounced

because no researchers with 5 or less years of experience reported that they develop software for other researchers.

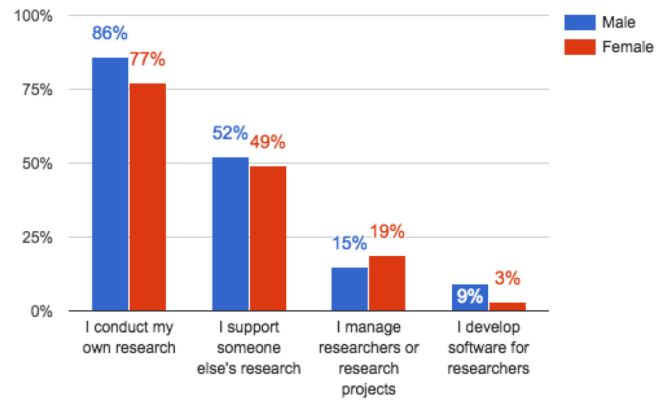


Fig. 2. Assessment of researchers' role between male and female respondents.

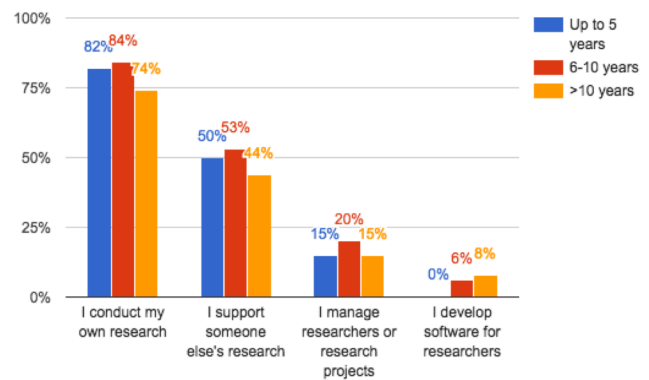


Fig. 3. Assessment of researchers' role between respondents with varying years of experience.

The next topic concerned use of research software. As shown in Fig. 4, 95% of respondents say that they use research software. The use of research software remains as a large majority, even when the data is stratified by gender and years of experience, as shown in Fig. 5 and Fig. 6, in which both genders and all levels of experiences report that at least 85% of respondents use research software. One distinct feature of this data, shown in Fig. 6, is that respondents with only up to 5 years of experience reported that they do not use research software at the highest percentage of 15%. It is also worth mentioning that male and female respondents reported roughly equal percentages of software use as shown in Fig. 5.

The third topic concerns what would happen if the responses the participants could no longer use research software. This was intended to assess the need for research software among the respondents. As shown in Fig. 7, a majority of respondents said that it would not be practical for them to conduct their work without software. This trend remained consistent when the data was stratified by gender and years of experience, as shown in Fig. 8 and Fig. 9, in which all categories demonstrated a majority of respondents claiming that it would not be practical to conduct research without software. One distinct feature can be seen in Fig. 9; respondents with only up to 5 years of experience reported that not using software would make no significant difference to their work at a much greater percentage than any

of the other groupings, corresponding with the fact that members of this group were the most likely not to use research software. Not surprisingly, Fig. 6 and Fig. 9 show a correspondence between those who could do their research without software and those who don't use software in their research, across all levels of experience and gender, but not an equivalence, which is somewhat puzzling and possibly deserves more detailed follow-up. For example, if 5% of male respondents do not use research software, why would only 2% of male respondents say that not having would make no difference to their work, rather than 5%?

Finally, in order to assess the level of training that the respondents have received in software development, the survey

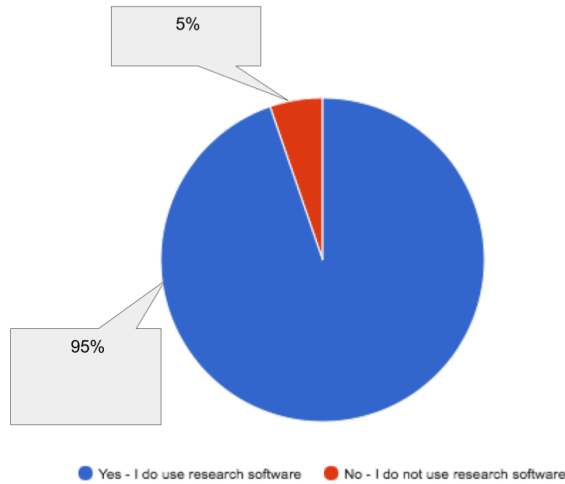


Fig. 4. Use of research software among respondents.

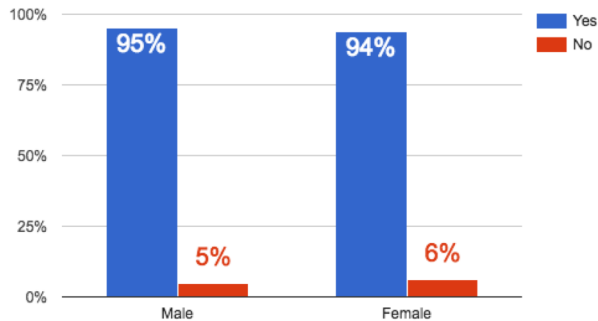


Fig. 5. Use of research software between male and female respondents.

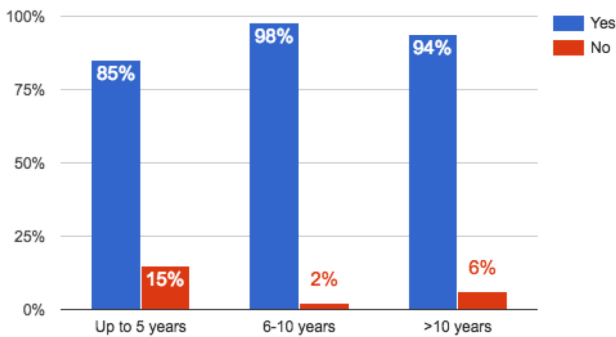


Fig. 6. Use of research software between respondents of varying experience.

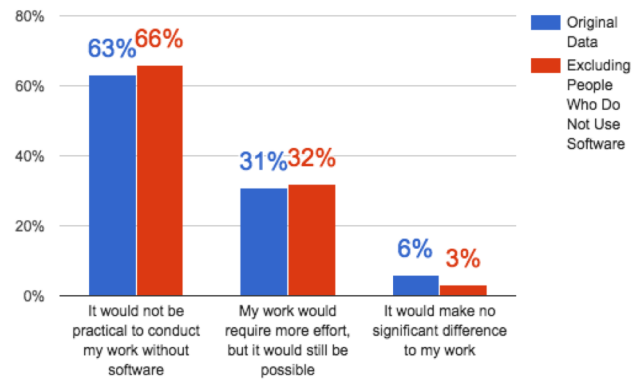


Fig. 7. Assessment of need for research software among all respondents and respondents excluding those who do not use research software.

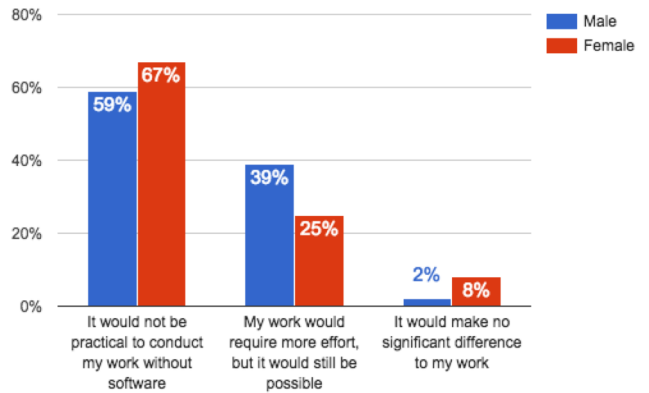


Fig. 8. Assessment of need for research software among male and female respondents.

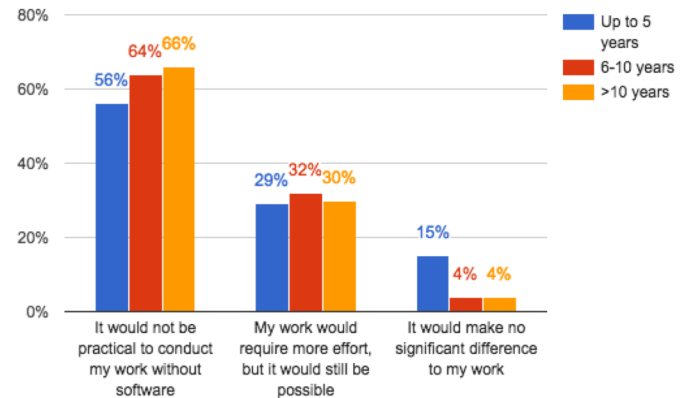


Fig. 9. Assessment of need for research software between respondents of varying experience.

asked "Have you received any training in software development?" The data presented in Fig. 10 summarizes the overall results of this question. The majority of respondents (54%) reported that they had received no training in software development (Fig. 10). Some deviation is seen with regards to gender, as shown in Fig. 11, where female respondents had a large majority (68%) who had not received any training in software development, while male respondents had a majority (64%) reporting that they had received software development training in some form. Fig. 12 shows that respondents with all

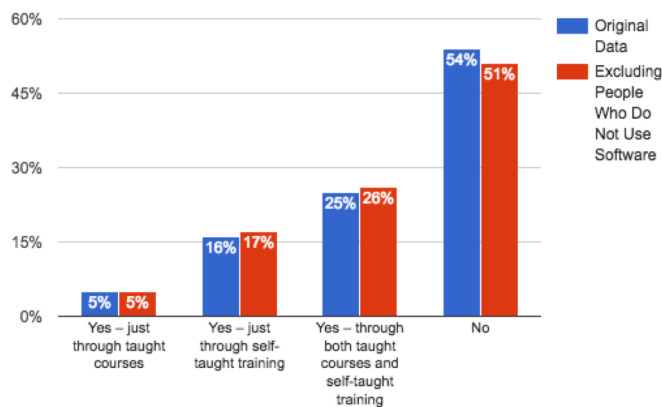


Fig. 10. Training in software development between respondents and respondents excluding those who do not use research software.

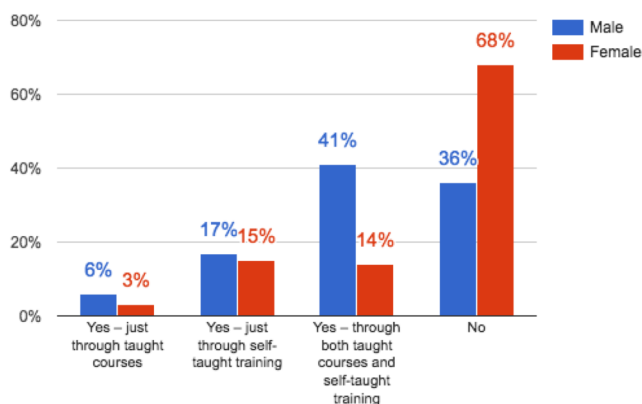


Fig. 11. Training in software development between male and female respondents.

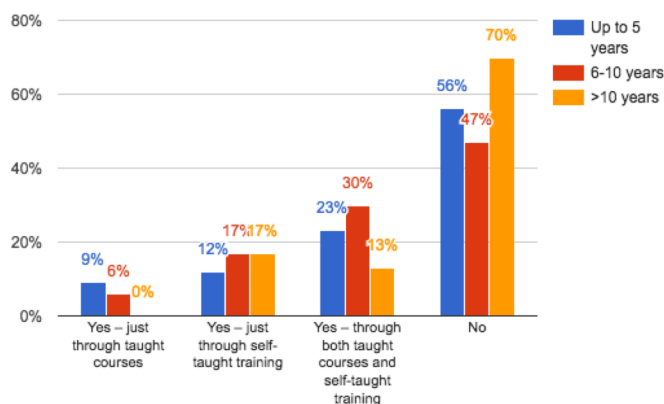


Fig. 12. Training in software development between respondents of varying experience.

levels of experiences reported “No” at the greatest percentage. Respondents with more than 10 years of experience not only reported the greatest percentage of individuals that had not received any training in software development (70%) out of all of the groups, but also showed that 0% of the respondents had received training in software development through taught courses.

More positively, training results for the subset of respondents who produce software for other researchers are shown in Fig. 13. All of them have been trained in software development, through self-taught training, instructed courses, or both. And the vast majority (82%) have received training in software development through taught courses.

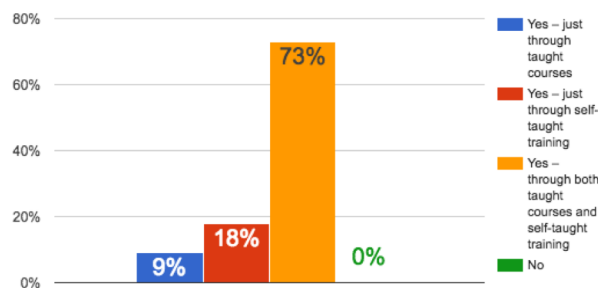


Fig. 13. Training in software development for respondents who develop software for other researchers

#### IV. CONCLUSION

This paper reports on a survey-based study that was conducted to assess the prevalence and need of software use in research across genders and varying years of experience by administering a questionnaire to members of the U. S. National Postdoctoral Association in 2017.

Although the study was conducted at a limited scale, the results of the survey highlight several key issues concerning the use of software in academic research. Research software is in widespread use among respondents, where it plays a vital role in conducting their research. And the small fraction of respondents who develop software for others have had software development training, and most have had formal training.

We recommend that future studies expand on this research by conducting a more comprehensive survey of researchers (beyond postdocs) in the U.S. We also recommend that future studies work to assess the quality of software produced in research settings.

#### REFERENCES

- [1] J. Howison and J. Bullard, “Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature,” *J Assn Inf Sci Tec*, 67: 2137–2155. 2016. doi: 10.1002/asi.23538
- [2] S. Hettrick, “It’s impossible to conduct research without software, say 7 out of 10 UK researchers,” Software Sustainability Institute, 2014. Available at: <https://www.software.ac.uk/blog/2016-09-12-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers>
- [3] S.J. Hettrick, M. Antonioletti, L. Carr, N. Chue Hong, S. Crouch, D. De Roure, et al, “UK Research Software Survey 2014”, Zenodo, 2014. doi: 10.5281/zenodo.14809.
- [4] J. E. Hannay, C. MacLeod, J. Singer, H. P. Langtangen, D. Pfahl and G. Wilson, "How do scientists develop and use scientific software?," 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, Vancouver, BC, 2009, pp. 1-8. doi: 10.1109/SECSE.2009.5069155
- [5] Gizmodo, “Survey Response Rates,” 2015. Available from: <https://www.surveygizmo.com/survey-blog/survey-response-rates/>