



NUMPY COMMUNITY SURVEY 2020 RESULTS



OVERVIEW

In mid-2020, the NumPy Survey Team in partnership with students and faculty from a Master's course in Survey Methodology jointly hosted by the University of Michigan and the University of Maryland conducted the first official NumPy Community Survey. We invited all NumPy stakeholders to share about their experiences with NumPy and voice their thoughts about the future of the project. Over 1,200 users from 75 countries participated to help us map out a landscape of the NumPy community.

Key
Performance
Indicators

TABLE OF CONTENTS

Overview	3
Glossary	6
Community	7
Community – Contribution to open source	8
Community – Mentorship	11
Community – Demographics	13
Usage and development	17
Future directions	20
Acknowledgements	21
Methodology	23

GLOSSARY

Throughout this report, you'll see a few key terms and phrases come up. These provide a consistent framework for how we collect and describe our data within the report.

COMMUNITY For the purpose of this report, the terms "NumPy community" or "community" comprise everyone who uses NumPy, regardless of their activity.

LOCATION Country information for users is based on the responses submitted by the survey participants.

NUMPY USERS Anyone who uses NumPy, regardless of their activity, and participated in this survey.

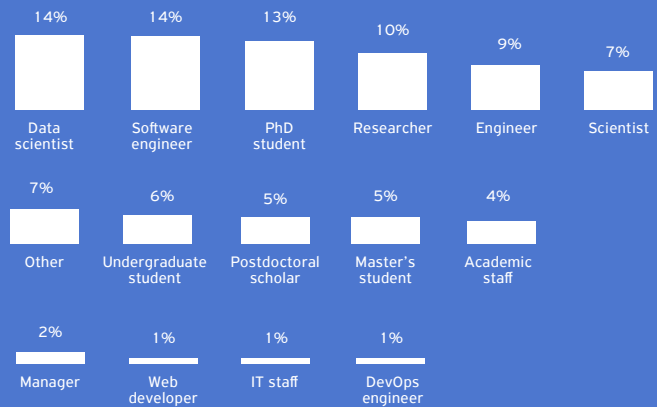
OPEN SOURCE PROJECTS are public repositories with an open source license.

COMMUNITY

Job Roles

41% of NumPy users identify as data scientists, software engineers, or PhD students

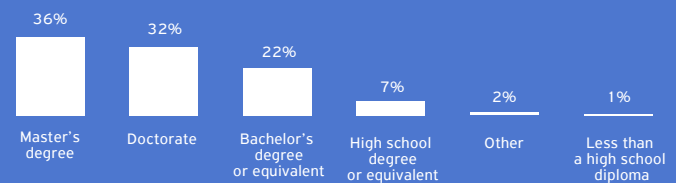
What is your current role in your organization?



Level of Education

1/3 Nine out of ten have at least a Bachelor's degree and one in three holds a PhD

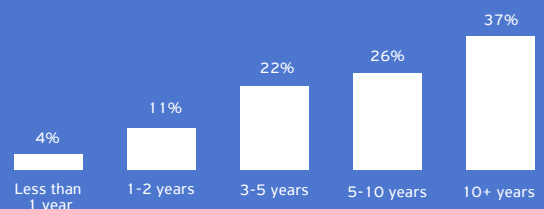
What is the highest degree or level of education you have completed?



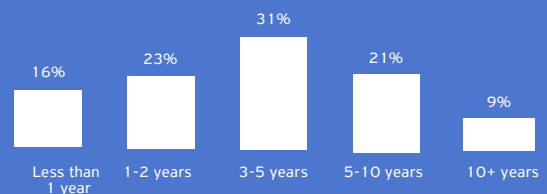
Experience in Programming

63% of survey participants have significant experience in programming, with veterans (10+ years) taking the lead. Interestingly, when it comes to using NumPy, more of our respondents identify as beginners than experienced users.

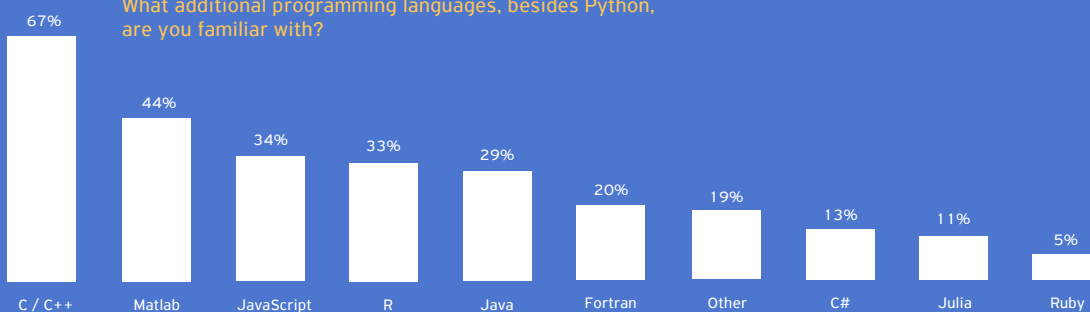
How many years of programming experience do you have?



How many years of experience do you have using NumPy?



What additional programming languages, besides Python, are you familiar with?



67% of NumPy users are familiar with C / C++; 44%, with Matlab.

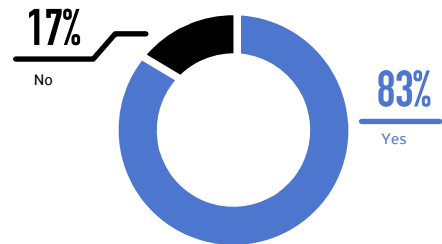
COMMUNITY – CONTRIBUTION TO OPEN SOURCE

Like many open source projects, NumPy largely relies on community contributions for maintenance and improvements. One of the objectives of this survey was to learn more about NumPy users' experiences and motivations as contributors to NumPy and other open source projects.

8/10

Eight out of ten survey participants are interested in contributing to open source software projects. Yet, confusion about where to start and lack of experience and time are often preventing them from making the leap.

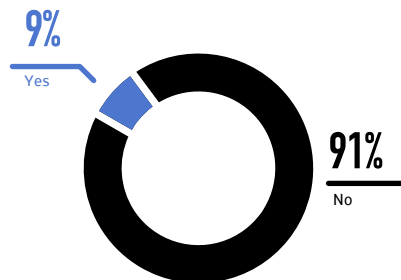
Do you have any interest in contributing to open source projects?



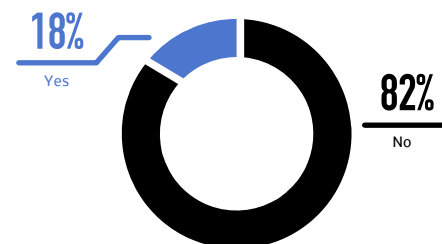
9%

9% of survey participants have contributed to NumPy and one in five of those contributes regularly.

Have you ever contributed to NumPy?



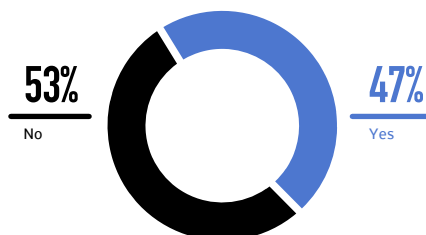
Are you a regular contributor to NumPy?



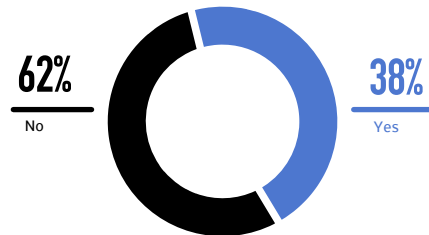
47%

When it comes to other open source projects, 47% of survey participants have contributed at least once, mainly to the projects within the Python ecosystem, and one in three of them contributes to these projects on a regular basis.

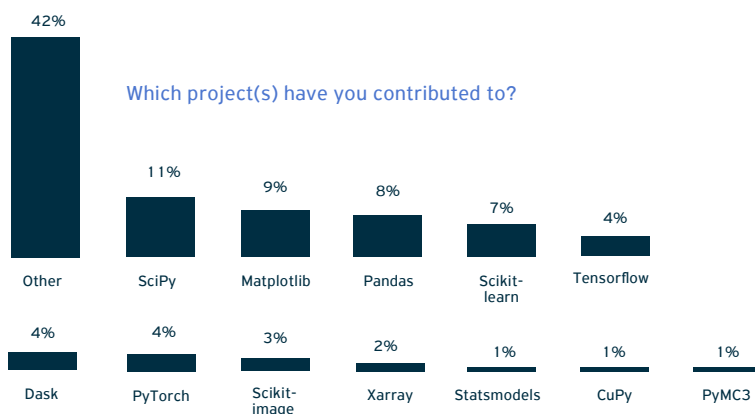
Have you ever contributed to any open source project before, excluding NumPy?



Are you a regular contributor to these open source projects?



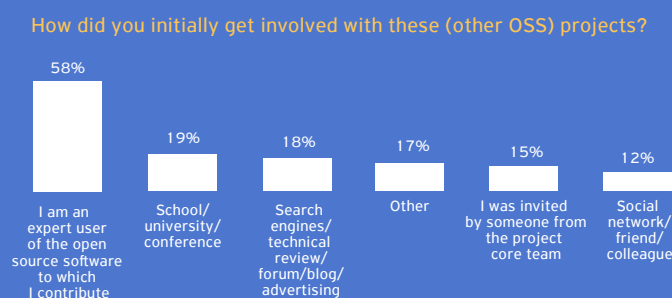
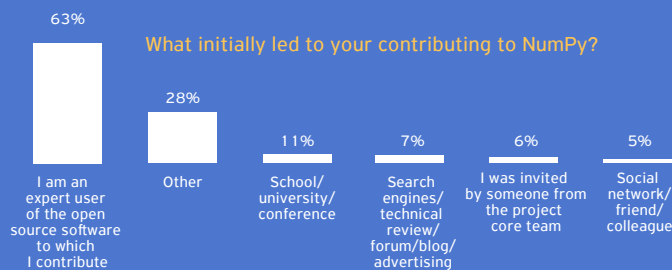
Which project(s) have you contributed to?



Curious what initially led these NumPy users to contributing to open source? Nearly two thirds of them identify as experts in NumPy. A similar pattern is found in the wider open source community.

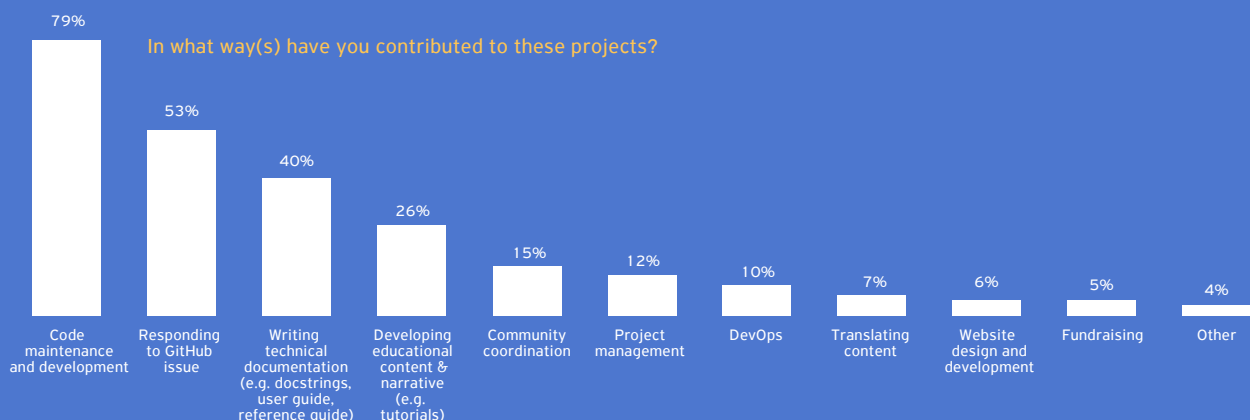
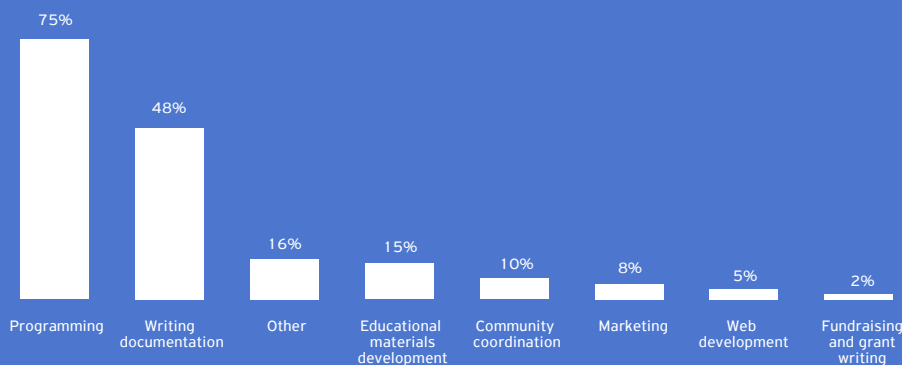
TAKE A NOTE: If you are looking for more contributors to your open source project, reach out to your most experienced users first.

Almost all (97%) of the NumPy contributors also contribute to other OSS projects reflecting NumPy's central place in the scientific Python ecosystem.



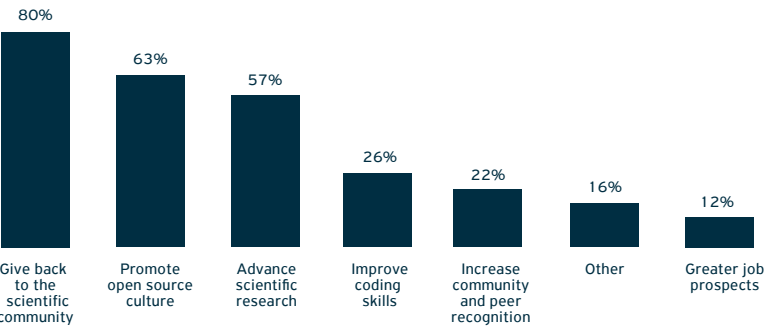
75% of NumPy contributors who participated in this survey have contributed to the project source code; nearly half, to the technical documentation. While 79% of contributors to other open source projects contributed code, 40% pitched in to help with documentation and one in four, with educational content.

In what capacity have you contributed to NumPy?



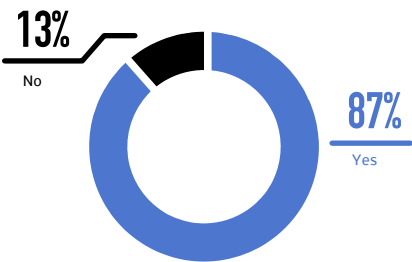
Giving back to the scientific community, promoting open source culture, and advancing scientific research are the top three reasons for respondents to contribute to NumPy.

What are your motivations for contributing to NumPy?



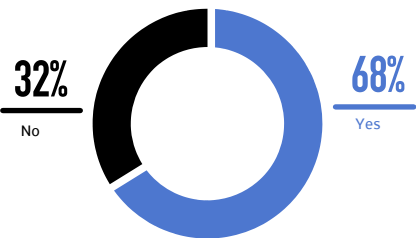
87% 87% of casual NumPy contributors and 100% of those who identify themselves as regulars plan to continue making contributions to the project.

Do you plan to continue making contributions to NumPy?

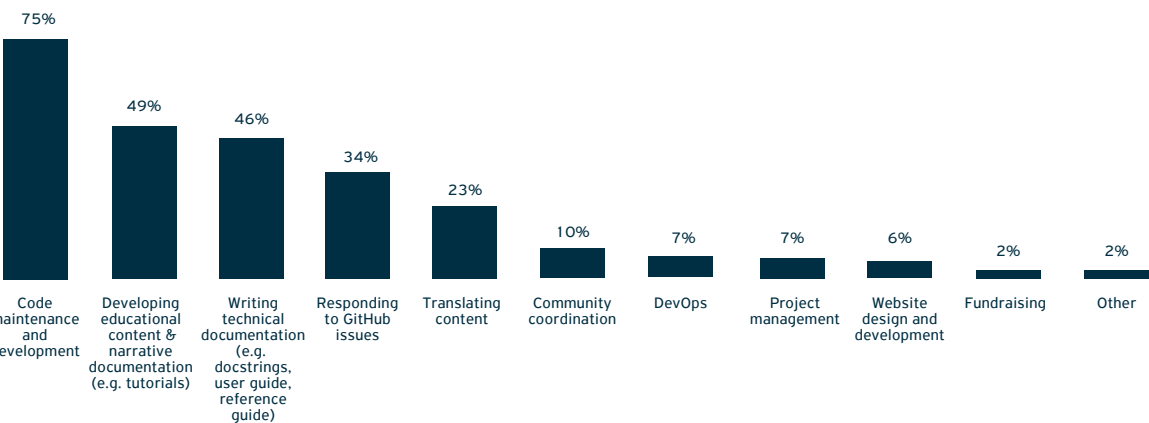


Out of those who have never contributed to NumPy, 68% like the idea of giving back to the NumPy community. 75% of potential contributors expressed their interest in contributing to the source code and nearly half would like to help with educational content or technical documentation.

Are you interested in contributing to NumPy?



In what ways would you be interested in contributing to NumPy?

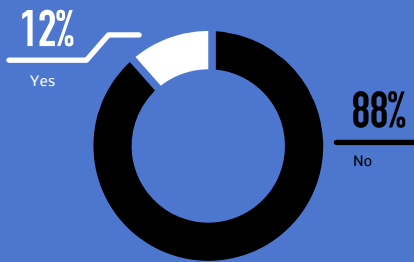


COMMUNITY – MENTORSHIP

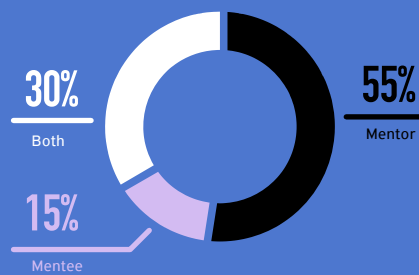
Mentorship is critical to open source community sustainability and growth. In this survey, we attempted to find out more about NumPy users' experiences as mentors and mentees.

12% of respondents took part in at least one mentorship program offered by an open source community. 55% of them served as mentors, 15% joined as mentees, and 30% participated as both mentors and mentees.

Have you ever participated in any mentorship program in any open source community?

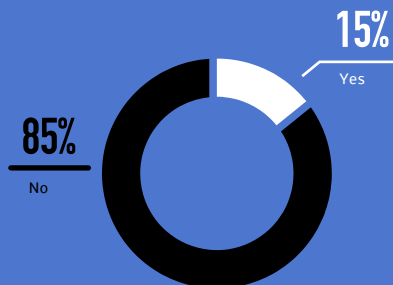


What is/was your role in the program(s)?

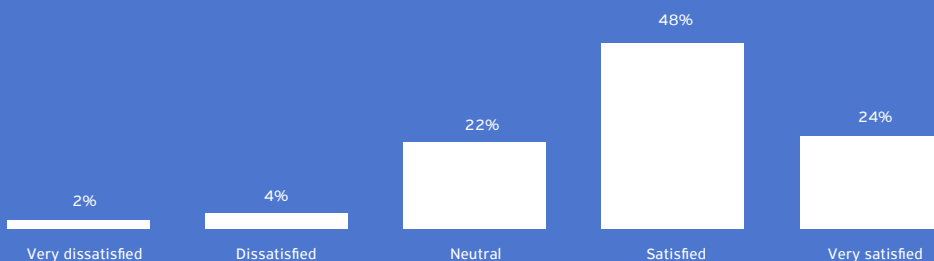


85% of respondents participated in mentorship programs organized by open source communities for free. Nearly three quarters of these respondents feel satisfied with their experiences.

Were you ever charged any fees to participate in the mentorship program?



How satisfied were/are you with your mentorship experience?

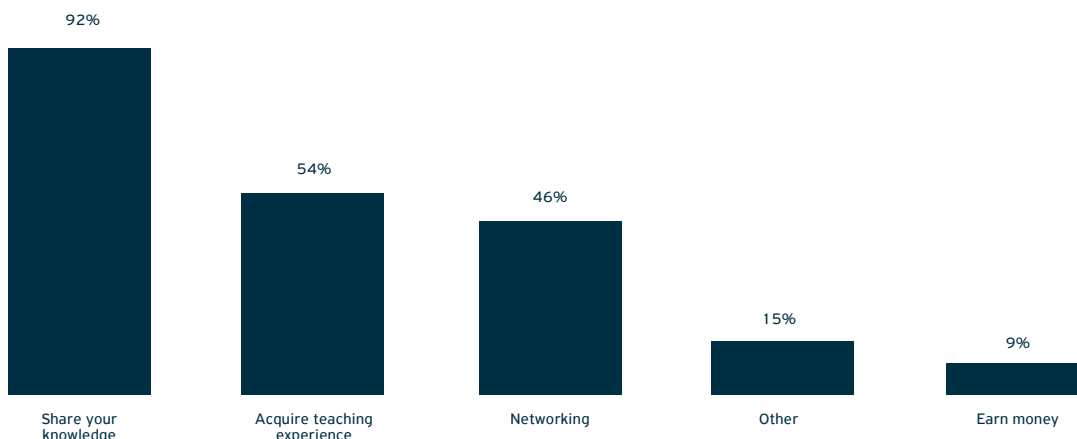


70%

70% of survey participants connected with their mentees or mentors via academic/alumni organizations; about a half, via professional networks or social media.

92% of respondents who served as mentors in these programs enjoyed sharing their knowledge with newer members of the open source community, slightly over a half (54%) viewed these programs as an opportunity to acquire additional teaching experience, and 46% joined to connect with more peers.

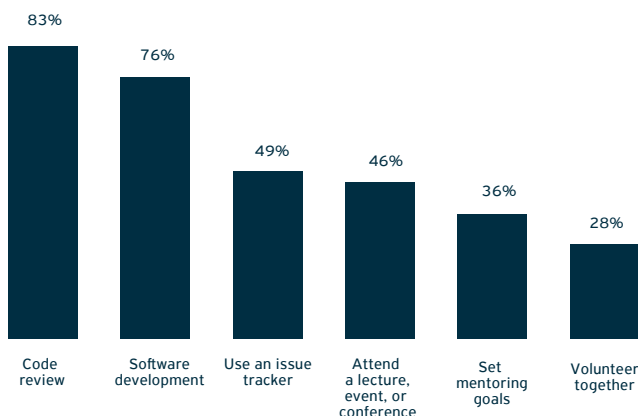
What motivates you to be a mentor?



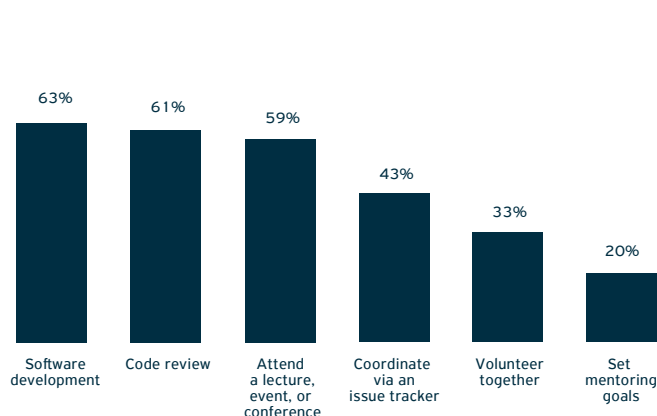
86% of survey participants served as mentors without any financial compensation. As part of mentoring, 83% of respondents reviewed code with their mentees, three quarters collaborated on building software, and about a half used an issue tracker or attended together an industry event.

About 60% of respondents who participated in mentorship programs as mentees reviewed code, collaborated on building software, or attended an industry event together with their mentors.

Have you ever engaged in the following activities with your mentees?

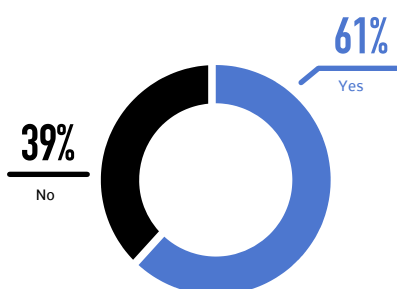


Have you ever engaged in the following activities with your mentors?



61% of NumPy users are interested in participating in a mentorship program offered by NumPy if it were to exist.

If NumPy had a mentorship program, would you be interested in participating?

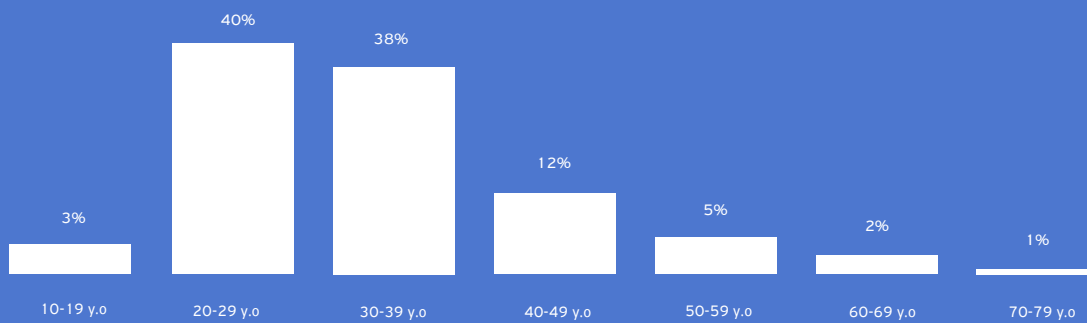


COMMUNITY – DEMOGRAPHICS

Age Range

The majority of respondents (nearly 80%) are in their twenties or thirties.

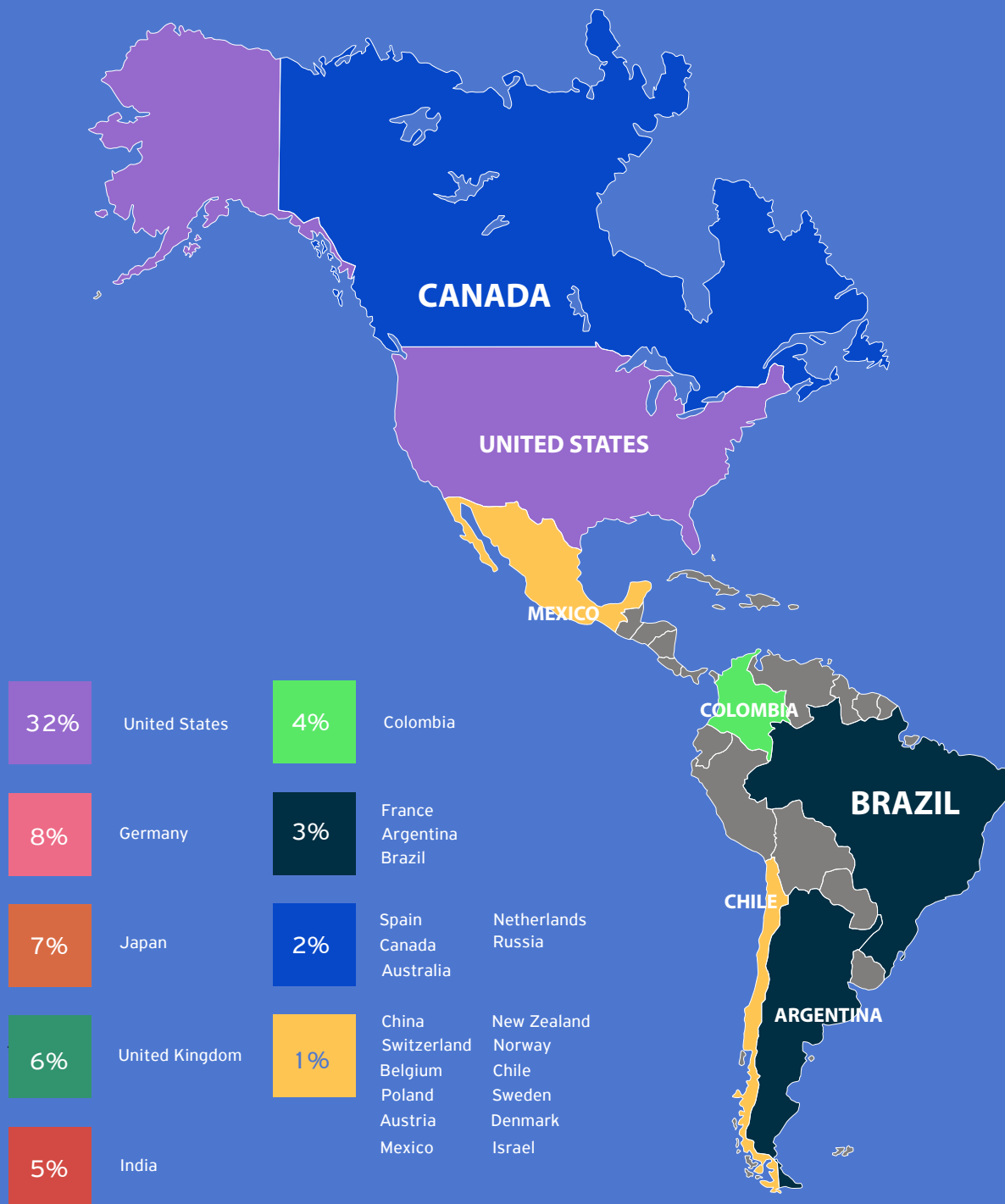
What is your age?

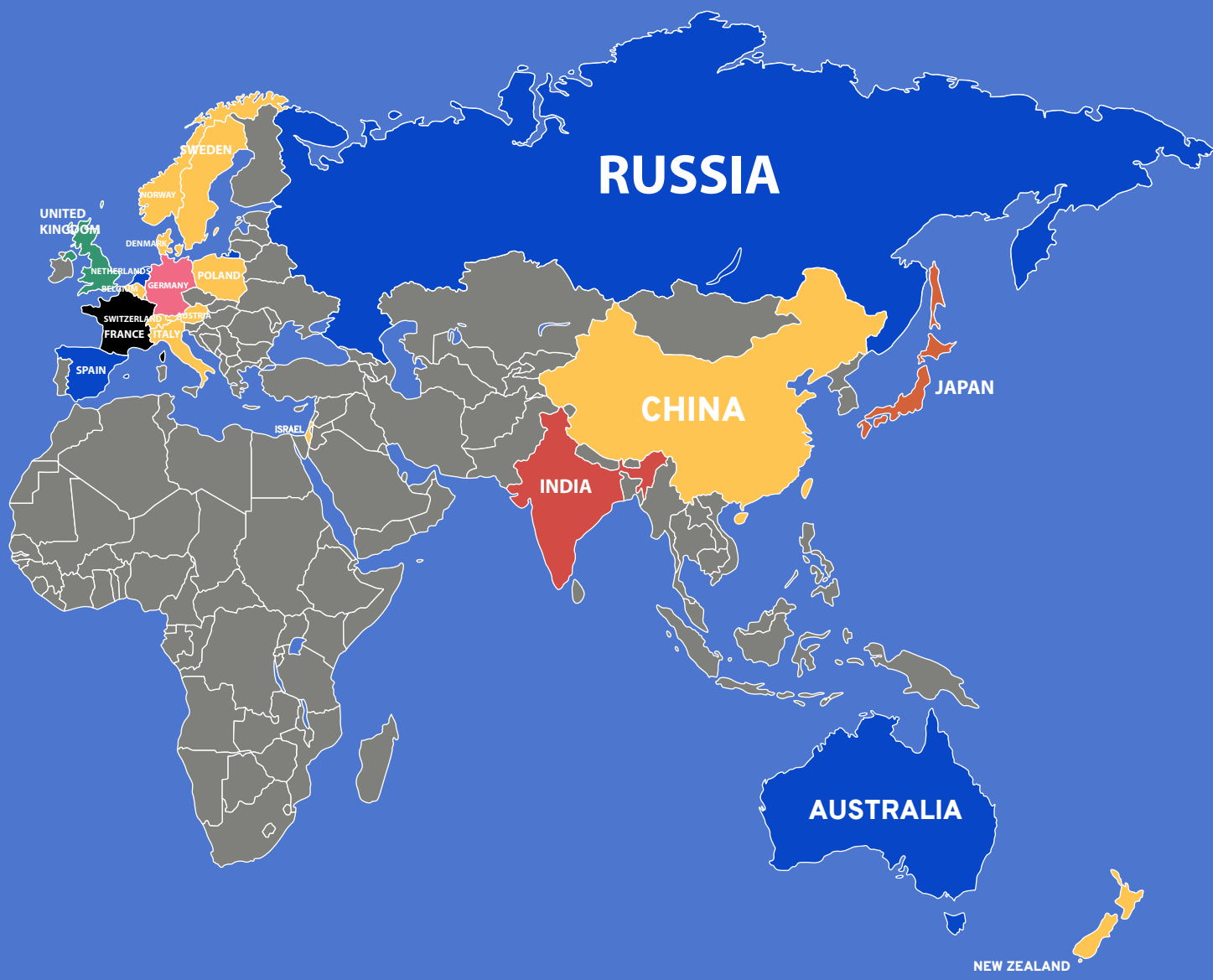


Location

NumPy is a true global open source project. In fact, almost three quarters of survey participants come from outside of the United States. In total, we reached NumPy users living in 75 different countries.

What is your country of residence?

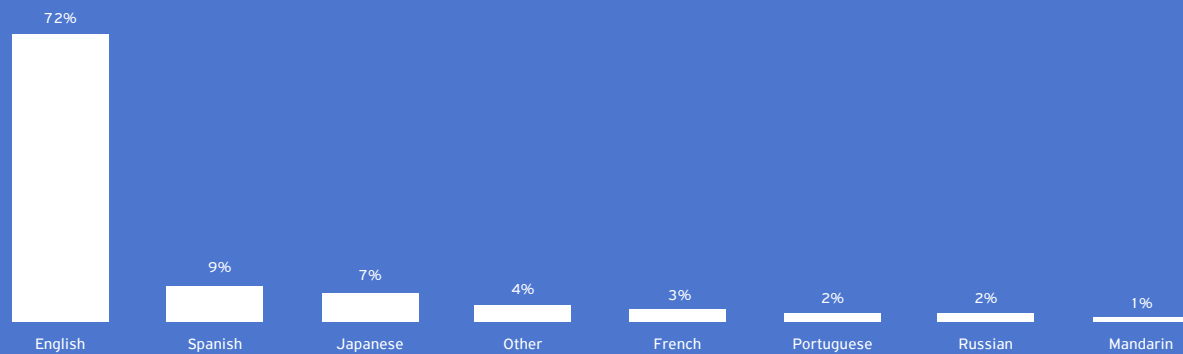




Languages

Over 70% of NumPy users choose English as the preferred language of communication. Still, we received several requests to have the project content (website, documentation, tutorials, etc.) translated into Japanese, Portuguese, and Spanish.

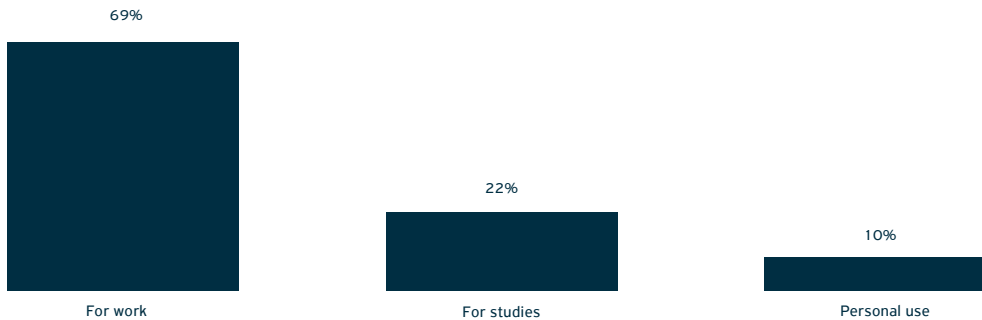
Please specify your preferred language.



USAGE AND DEVELOPMENT

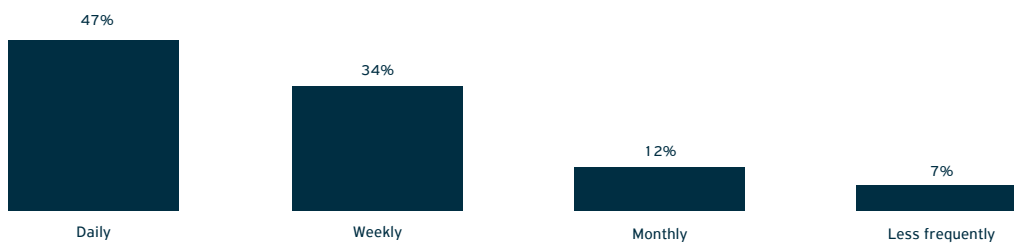
The vast majority of survey participants use NumPy for work; one in five, for studies.

How do you primarily use NumPy?



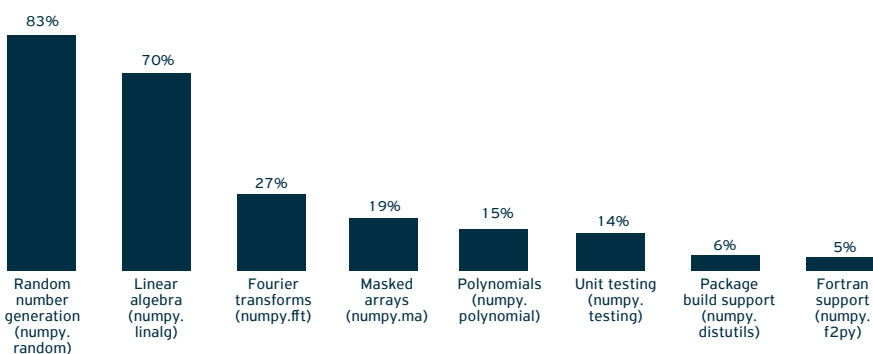
Half of respondents use it on a daily basis; 80%, at least once a week.

How often do you use NumPy?



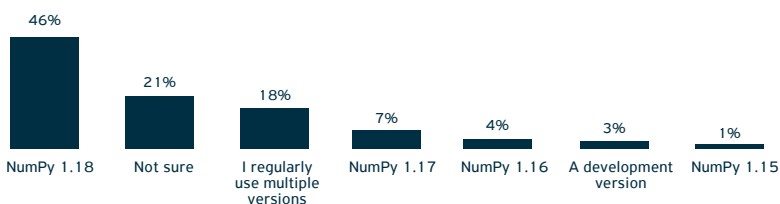
Random number generator, linear algebra, and Fourier transforms are the most often used NumPy sub-packages.

Which components of NumPy do you use regularly?



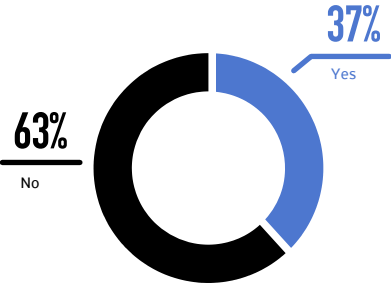
Nearly half of survey participants use NumPy version 1.18.

What version of NumPy are you currently using?



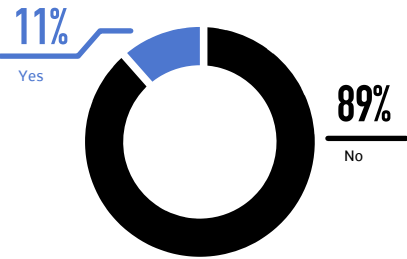
One in three respondents uses the new random API.

NumPy version 1.17 introduced a new API to NumPy's random package. Are you currently using the new random API?



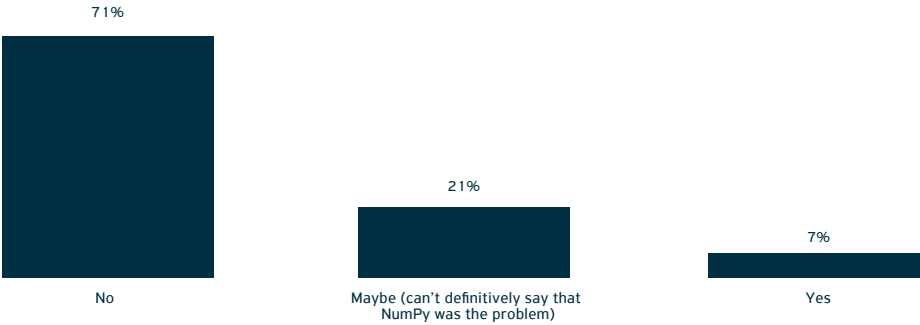
Only 11% have custom C extensions making use of the NumPy C-API.

Do you or your organization have custom C extensions making use of the NumPy C-API (aside from Cython)?

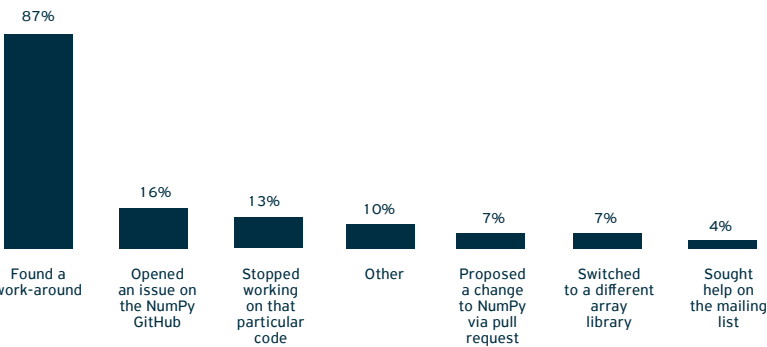


In the last year, 7% of NumPy users experienced problems in code they've written stemming from a problem in NumPy. What did they do about it? 87% found a workaround for it, 16% opened an issue on NumPy GitHub, and 7% proposed a change to NumPy by submitting a pull request.

In the last year, have you experienced problems in code you've written stemming from a problem in NumPy?

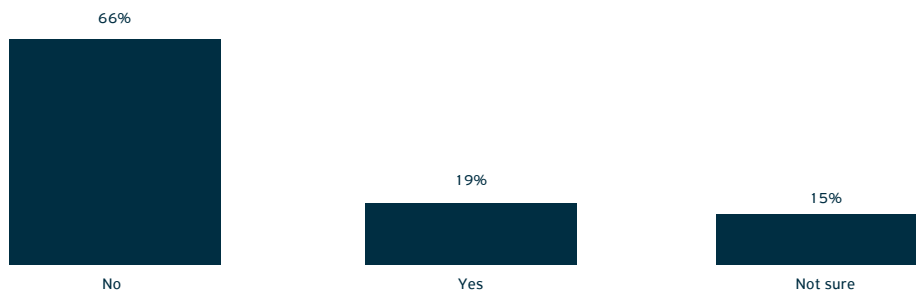


What action(s) did you take in response to this issue?

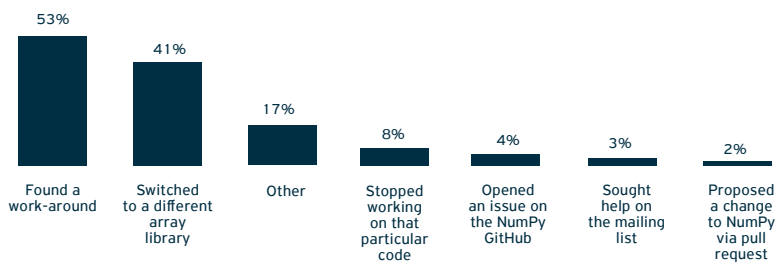


During the same period of time, 19% of respondents encountered a problem involving numerical data that they were unable to solve using NumPy. 53% found a workaround for it, 41% switched to a different library, 5% opened an issue on NumPy GitHub, and 2% proposed a change to NumPy by submitting a pull request.

In the last year, have you encountered a problem involving numerical data that you were unable to solve using NumPy?

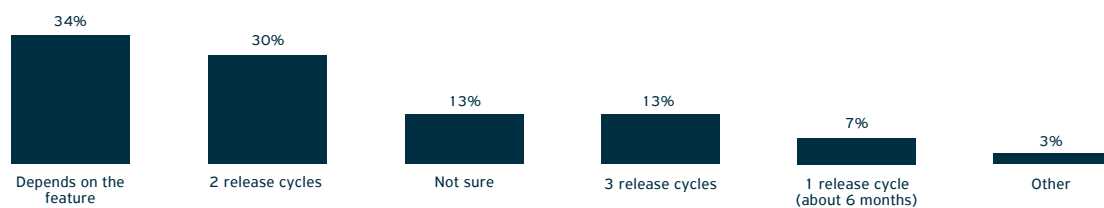


What action(s) did you take in response to this issue?



The NumPy maintainers were eager to know what NumPy users consider as a good deprecation time frame. About one third of survey participants believes it should depend on the feature, while another third suggests 2 release cycles.

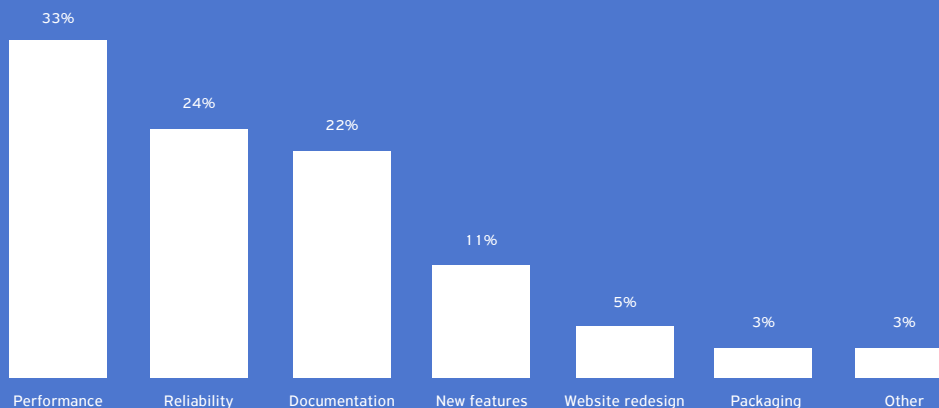
What do you consider as a good deprecation time frame?



FUTURE DIRECTIONS

PROJECT PRIORITIES

NumPy users rank performance as the number one priority for the project. It is followed by reliability and documentation.



Performance

The overall sentiment in the comments is that NumPy performance is already good, but further improvements are possible and should be put high on the project's priority list.

Parallel execution, GPU support, and improved interoperability with other packages, both via the NumPy array protocols and directly with CuPy, Dask, JAX, Numba, Cython, Pythran, PyPy, and deep learning frameworks were mentioned most often. Also, quite a few respondents suggested making better use of modern hardware capabilities like SIMD instructions and simplifying the back-end switches for, in particular, BLAS.

Reliability

NumPy users are almost unanimous that NumPy is very reliable already and it's essential that it remains this way, including while introducing new features and if the speed of development were to increase in the future. Minor points identified for improvement include the clarity of error messages and dealing with inconsistencies and quirks in the parts of NumPy that are used less often.

Documentation

The majority of respondents are happy with the quality of the NumPy documentation. Still, a number of suggestions were submitted for further improvement. "Examples, more examples, and more detailed examples" is the recurrent theme in about half the comments. More tutorials and narrative documentation around best practices and low-level parts of NumPy are also high on NumPy users' wish lists. A common concern about the NumPy documentation is that NumPy is very large, and it's often not easy to navigate the documentation and find the functionality that best fits a user's need.

THE BIG PICTURE

In an open-ended question, we asked: "What single immediate change to NumPy would bring the most value to you as a NumPy user?" Many creative ideas and current pain points were brought up, from community development to API consistency and support for specific scientific use cases. The following topics received by far the most attention: documentation improvements, parallel execution, and GPU support.

Documentation

The comments about the NumPy documentation in response to this question largely overlap with the responses about priorities for NumPy (refer to the summary regarding the documentation in the section above).

GPU support

With nearly 20% of all responses focused on the topic, it's clear that GPU support is high on users' wish lists. Very few respondents brought up CuPy though, and their suggestions were either creating even better compatibility between CuPy and NumPy or implementing native GPU support in NumPy instead.

Parallel Execution

Making better use of modern CPUs was mentioned nearly as often as GPU support. Some respondents suggested guiding users to tools like Dask or Numba to execute NumPy-based code in parallel. Yet, the vast majority called for NumPy itself to provide good multithreading support.

ACKNOWLEDGEMENTS

The inaugural NumPy community survey is an effort of an international group of volunteers who showed a remarkable dedication to the project amidst the extraordinary challenges of the year 2020. Each is listed alphabetically by type of contribution.

Survey Design and Analysis

Ross Barnowski
Sebastian Berg
Xiaoyi Deng
Ralf Gommers
Stephanie Mendoza
Inessa Pawson
Deji Suolang
Mame Fatou Thiam

Translations

Eric Depagne
Xiaoyi Deng
Yuki Dunn
Kiko Correoso Garcia
Jose Guzman
Andrei Ivanov
Paul Ivanov
Siddhartha Kapoor
Tetsuo Koyama
Guilherme Leobas
Dayane Machado
Mahfuza Humayra Mohona
Aerik Pawson
Inessa Pawson
Zijie Poh
Sumera Priyadarsini
Fabrice Silva
Shaloo Shalini
Kriti Singh
Alexandre de Siqueira

Special thanks

Frederick Conrad (Survey Research Center, Institute for Social Research,
University of Michigan)

Danuta Dzierżanowska

Jim Lepkowski (Survey Research Center, Institute for Social Research,
University of Michigan)

Claire McKay Bowen (Urban Institute)

Michael Traugott (Survey Research Center, Institute for Social Research,
University of Michigan)

METHODOLOGY

1,236 NumPy users from 75 countries participated in this online survey, which was fielded from July 2nd to September 2nd, 2020. Respondents were recruited via the NumPy mailing list, the banner display on the numpy.org homepage, announcements at the SciPy'20 conference, and via Twitter.

To engage non-English speaking stakeholders, we conducted the survey in 8 additional languages: Bangla, French, Hindi, Japanese, Mandarin, Portuguese, Russian, and Spanish.

Some totals don't equal 100% due to multiple-choice answers or rounding.