

Table Of Contents

Letter From Our Directors	చ
About Us	4
Our Mission	5
Our Model	6
Our Purpose	7
We Are Labs	8
Sustainable Ecosystem	9
OSS Projects We Contribute To	10
OSS Contributions by the Numbers	11
Project Highlights	12
Community Highlights	13
Leadership Roles in OSS Projects	14
Labs Focus Areas for 2024	15
Stories: Free-threaded CPython	16
Stories: Interactivity And Internationalization of	
Documentation	17
Stories: Packaging	18
Stories: Accelerating Sparse Array	
Computation	19
Stories: NASA ROSES	21
Stories: Accessibility	22
Stories: Narwhals	23

Labs Open Source Fund	24	
Conferences And Events	26	
Labs Open Source Internship Program	27	
Interns And Projects	28	
Running the Labs Open Source Internship	30	
How the Labs Internship Works	31	
Sustaining Labs	32	
Our Community-First Working Model	33	
Co-Ownership Governance Model	35	
Financials	36	
Noteworthy Funding Sources	37	
Diversity Within Labs	38	
Building a Sense of Belonging	41	
How Far We've Come And Looking Ahead		
Acknowledgements	43	
See You Next Year		

Letter From Our Directors

While our path was initially uncertain during the early days of Quansight Labs, we had big dreams driven by a clear mission and bold vision. Over the past six years, Quansight Labs has evolved from a small experimental organization with ambitious aspirations of reimagining and redefining open source sustainability to a still small but mighty organization with incredibly talented staff, a deliberate approach, and strategic direction.

Our clear vision has enabled us to build on it, explore and define the problems we will tackle, and continuously refine our initiatives and goals. It has also shaped how we build long-lasting and meaningful relationships with other stakeholders in the open source and scientific computing ecosystems and beyond (including corporate, academic, and public sectors)—ultimately driving positive impact in the communities we serve and participate in.

From the outset, we have been intentional in building an organization and a diverse and talented team rooted in our values and vision for the future of open source communities and projects. This year was no exception as we strengthened existing initiatives and collaborations and initiated new ones that align with our values of inclusiveness, equity, resiliency, co-creation, and community sovereignty.

Reflecting on another year, we recognize that the path threaded has not been free of uncertainty as changes and challenges within and outside the tech industry have directly and indirectly impacted us as an organization and our staff. However, the need for and reliance on open source and open infrastructure has never been greater. Open source drives innovation and advancements across businesses, Al development, scientific breakthroughs, and critical areas like health, climate, and the public sector.

Nearly every aspect of our modern society is shaped by open source, open infrastructure, and the people building and sustaining these. Yet, despite such staggering success, protecting and sustaining open source in a decentralized, equitable, and community-centric way has never felt more vital, especially as the road ahead seems rather windy and changes in policies and priorities add to the uncertainty of the future of open source.

In such a landscape, it is impossible to look at what our team has achieved and continues to achieve, as well as their resilience and unwavering commitment to our mission and our communities, and not be filled with awe and pride. So, once again, as every year, we are immensely proud to share some highlights from the Quansight Lab's team's work. We hope you enjoy learning about their journey as much as we have enjoyed being part of it.

Tania Allard & Ralf Gommers

Directors, Quansight Labs

About Us



Our Mission

The mission of Quansight Labs is to sustain and grow community-driven open source projects and ecosystems, focusing on the core of the PyData stack and tools and digital infrastructure for data science, Machine Learning (ML) and Artificial Intelligence (AI), and scientific computing.

Open source has become an integral part of our daily lives. Millions of people worldwide use open source software and infrastructure day in and day out.

We fulfill our mission through these activities:

2. Quansight Labs empowers the people who make up the open source community

and fosters growth and inclusion.

1. Quansight Labs facilitates advances and innovations in science, research, education, government, and industry.	We run an annual three-month-long open source internship program for individuals from historically underrepresented groups interested in contributing to and participating in open source.
Our team comprises individuals with diverse backgrounds and skills, including developers, community leaders, designers, accessibility experts, and technical writers. Together, we support the scientific computing ecosystem and its evergrowing community.	Quansight Labs members are actively involved in community building through organizing community calls, contribution sprints, mentoring individuals, organizing community events, and helping coordinate resources and initiatives within the open source ecosystem.
We develop and maintain open source software for ML/AI, data science, scientific research and technical communication. Through this, we foster the creation and dissemination of knowledge.	Quansight Labs has built an access-centered practice to help make scientific computing tools accessible to a broader range of individuals, particularly those with disabilities.
☑ We contribute to many community-driven open source projects to support the long-term sustainability of the ecosystem.	We recently launched a new Labs OSS fund to support the maintenance and development of open source projects that are critical to our ecosystem and are not maintained by Labs members.

Our Model

A core focus of Quansight Labs is to develop innovative yet sustainable funding mechanisms to live up to our mission - sustaining community-driven open source projects and their communities. As such, we have spent the last six years developing and improving our funding and operational models.

Through these, we ensure open source maintainers and community leaders have a significant amount of time dedicated to working on community-driven open source projects while also spending part of their time on consulting projects with Quansight's commercial clients. Such a model includes benefits such as learning and upskilling opportunities for our team, bringing back concrete user feedback and needs, and ideas for open source projects. It also gives us the flexibility to quickly scale up our contributions to the open source projects we support when we receive direct financial support for them.



Our Purpose

The majority of our efforts are aimed at maintaining, strengthening, and evolving existing open source projects that are critical to the scientific computing ecosystem and the PyData stack, including NumPy, Jupyter, SciPy, pandas, conda-forge, and CPython among many others. We provide talented maintainers with structural time for essential day-to-day maintenance and tackling complex problems that require weeks or months of dedicated time. Additionally, we complement this maintenance work with innovation activities to identify and fill critical gaps in the PyData and broader scientific computing ecosystems.

In doing so, we acknowledge that for open source projects and communities to be successful and sustainable, many skills and perspectives are required beyond software development. Expertise in community building, technical writing, graphic design, UI/UX design, marketing, accessibility, fundraising, translation and localization, and project management, among others, are essential to open source. Yet, many of these are often underrepresented or nonexistent in community-driven open source projects. At Quansight Labs, we hire designers, community managers, developer experience engineers, and many others interested in open source to grow the diversity of skills in open source communities and support the projects' growth, sustainability, and maturity. At the same time, we spearhead and support grassroots and advocacy initiatives to formalize and recognize these activities' fundamental role in our ecosystem.

"What I love most about working at Quansight is being able to work on projects which have a long lasting impact on the rest of the open source ecosystem. For pretty much any project that I think of, chances are there's a colleague who's either contributed to it or maintains it, or in some cases even created it."

Marco Gorelli, Senior Software Engineer

"Quansight stands out to me because of its commitment to fostering collaboration between different fields like design, data science, and engineering. I feel empowered by the fact that I can directly contribute to the future of open source while working with people who are experts in their fields."

Smera Goel, UI/UX Designer



We Are Labs



Sustainable Ecosystem



OSS Projects We Contribute To

Our Quansight Labs team continues to actively sustain and contributes to many foundational projects in the PyData and scientific computing ecosystems and beyond, including the Python programming language and stdlib.js.

Last year, we sustained our contributions to the libraries at the core of the PyData ecosystem, including NumPy, pandas, SciPy, Matplotlib, scikit-learn, PyTorch, and JupyterLab to name but a few. We also extended beyond the core of PyData by contributing to high-performance computing, Machine learning, visualization projects, and infrastructure projects like Numba, Ray, Napari, JAX, and Nebari.

We also continued our already substantial support of the Python packaging ecosystem, from continuing our contributions to projects like conda, conda-forge, meson-python, and conda-store.

The main goal of these continued efforts remained improving the end-user and maintainer's experience and workflows by focusing on the tooling for packaging and distributing scientific software.

As part of our mission of building inclusive and sustainable projects and communities, Labs members continued participating in projects focused on Diversity, Equity, Inclusion, and Accessibility within our communities, from participating in the NumFOCUS DISC committee to continue to spearhead accessibility efforts through the Jupyter Accessibility project and directly working on accessibility improvements in projects such as the PyData Sphinx Theme, Bokeh, and more.



































































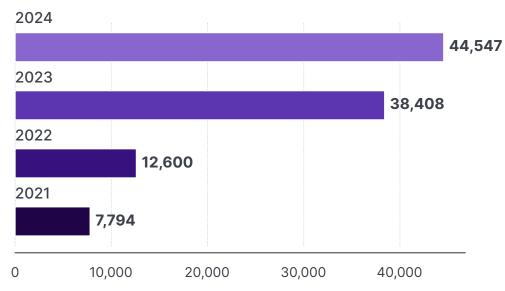




OSS Contributions by the Numbers

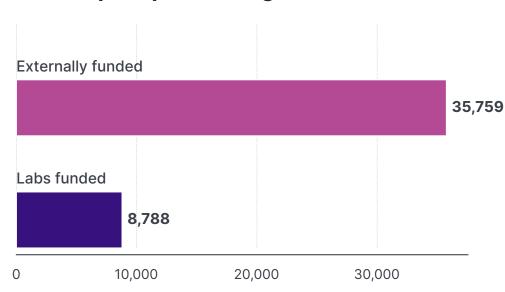
In 2024, we contributed to 40+ open source projects, open source projects, ranging from developer, packaging, and documentation tooling to foundational numerical libraries and ML frameworks and beyond.

Hours spent in OSS projects



We spent 44,547 hours working on these projects. which is a 16% increase compared to last year. Out of those, **8,788** hours were directly funded by Quansight Labs.

Hours spent per funding source



"At Quansight Labs, we work at the foundation of modern scientific computing, shaping the future of open source tools that power research and industry. The scale and impact of our contributions—sustaining critical projects and driving innovation in the PyData ecosystem—are hard to overstate. I feel incredibly fortunate to be part of a team that solves fundamental challenges and helps define what's next for scientific computing."

Project Highlights

conda metadata browser: a Streamlit application to query metadata from conda

packages.

Much of our work's impact comes from the breadth and depth of our contributions to community-driven open source projects. However, we would like to take a minute to celebrate some of the individual project highlights that Labs team members were responsible for or made significant contributions to this year:

\square	NumPy 2.0 was released in June 2024 after a huge amount of development work as well as rollout preparation.		PyData/Sparse added a new Finch.jl backend for most operations, which is significantly faster.
<u> </u>	Free-threaded Python support grew throughout the year, with NumPy 2.1 including support and fixing several longstanding thread safety issues along the	\square	Bokeh completed the first-ever accessibility audit.
	way, and PyO3 0.23.0 also adding support, unlocking the ability to write safe multi-threaded native Python extensions in Rust.	S	<u>Pyodide</u> had 0.26.0 and 0.27.0 releases with a significant number of package upgrades.
\(Narwhals went from an initial release in mid-February 2024 to becoming a required dependency of Altair, HierarchicalForecast, Marimo, Shiny, Plotly, Rio,		Scientific Python projects can now easily add interactive documentation, building on Pyodide, JupyterLite, and jupyterlite-sphinx.
▽	Vegafusion, and more. The Array API standard released v2023 and gained adoption in JAX, MLX, and	\(SciPy supports macOS Accelerate again after a number of years, resulting in improved performance and smaller wheels.
	ndonnx	\sqrt	PyData Sphinx theme had several accessibility improvements and published a
\square'	There were new releases of array-api-compat and array-api-strict, and two new		comprehensive design system.
/	libraries, array-api-extra and array-api-typing, were created.		Napari had <u>several packaging-related enhancements</u> . For instance, with several upstream contributions to the conda/constructor stack (including conda/
(Stdlib.js <u>v0.3.0</u> release included various REPL enhancements, including preview completions, custom keybindings, syntax highlighting, fuzzy auto-complete, and enhanced TAB completion.		menuinst, conda/conda-standalone, and conda/conda), napari moved to using only official sources to build napari installers.
\\ \	, The conda-forge team shipped a <u>new GPU CI</u> for conda-forge, and launched the		

Community Highlights

- Conda-forge launched a new website and refreshed documentation at condaforge.org, which includes a status dashbaord.
- The conda-forge team formalized a security policy and migrated their community forum to Zulip.
- Stdlib established weekly public developer meetings and almost tripled their contributor community.
- OpenBLAS now has a new GitHub organization and polished HTML documentation.

- Two Scientific Python Accessibility events were conducted to raise awareness: Accessibility in your digital day-to-day and Data visualization and accessibility.
- The Labs accessibility team advised the Python Software Foundation (PSF) on improvements to various outreach materials. It also was instrumental in establishing the new PSF User Success Working Group.
- Py-free-threading.github.io was created as a central resource for developers interested in experimenting with free-threaded Python.
- The Data APIs Consortium started having open community workgroup meetings.



Leadership Roles in OSS Projects



Agriya Khetarpal Joined the Pyodide maintainers team and the NumFOCUS Security Committee.



Pavithra Eswaramoorthy Chaired the first NumFOCUS election committee.



Jaime Rodríguez-Guerra Gained commit rights to napari.



Lucy Liu Became a scikit-learn maintainer and a member of the napari core team.



Melissa Weber Mendonça Became a member of the napari core team.



Nathan Goldbaum Gained commit rights on the PyO3 repository.



Pierre Simonard Gained commit rights to Panel.



Ralf Gommers Gained commit rights to OpenBLAS.



Tania Allard Joined the conda steering council and was re-elected to the PSF Board.

Labs Focus Areas for 2024

Our team at Labs brings diverse skill sets, deep domain expertise, and a strong understanding of the open-source scientific Python ecosystem. Each year, we set strategic goals to tackle ecosystem-wide challenges. In 2024, we continued our focus on Python packaging, multiple array and tensor library support, and access-centered practices, while also introducing a new priority: **enhancing scientific Python documentation**. The following stories summarize key projects and initiatives over the past year.



Stories: Free-threaded CPython

Until recently, most code written in the most popular implementation of the Python programming language—CPython—could not efficiently exploit CPUs with multiple cores. With increasing core counts in modern CPUs, the caveats around multithreaded scaling of Python programs present a fundamental challenge to the future popularity of Python and lead to inefficient use of computing resources to work around these scaling limitations.

The Python Steering Council decided in October 2023 to approve PEP 703, making way for a new experimental "free-threaded" build of CPython that can efficiently exploit multicore CPUs by dropping the restrictions on concurrent code execution imposed by CPython's global interpreter lock (commonly referred to as the GIL). Many packages, particularly in the scientific Python space, rely heavily on extensions for Python-written languages like C or C++ that are susceptible to multithreaded data races and other forms of memory unsafety. While PEP 703 was approved, the freethreaded build will remain experimental until community packages are runnable and the single-threaded performance issues present in the free-threaded build are fixed.

The Quansight Labs team and the Python runtime team at Meta worked throughout 2024 to improve the performance of free-threaded Python and allow the community to experiment with the free-threaded build on production workflows. We enabled the goal of improving performance by improving the specialized adaptive interpreter recently introduced to CPython to improve the performance of several patterns found in many Python codebases.

Since production Python workflows, particularly in the Al and Machine Learning spaces, often depend on the scientific Python stack, we initially focused on foundational community projects like Cython, NumPy, SciPy, and pandas. Before our team started, these packages were utterly broken on the free-threaded build. We identified and fixed the issues most commonly hit on the free-threaded build and systematically fixed issues to unblock work that is critically important on projects higher up the stack, like Matplotlib and scikit-learn. We also fixed issues in various packaging tools and libraries. We also wrote documentation (available at https://pyfree-threading, github.io) that describes best practices for porting projects to run on free-threaded Python and to enable efficient scaling of Python programs to utilize modern CPUs efficiently.

While we are continuing this effort in 2025, we are hopeful that our work will help build critical mass and provide community know-how to enable the broader community to port packages they need.

Stories: Interactivity And Internationalization of Documentation

During 2024, the Quansight Labs team dedicated a substantial effort to internationalization and interactivity for documentation in Scientific Python projects. These two initiatives aimed to enhance the accessibility and usability of these tools to serve our global and diverse communities better.

Internationalization

For this project, our goal was to internationalize several projects' websites – not documentation, as this would require a much larger team of translators and considerable review and maintenance efforts. As this effort involves several projects in the Scientific Python ecosystem, we worked to create automation that helped us manage and synchronize content between the projects' websites' sources on GitHub and Crowdin, our chosen internationalization platform, while keeping the translators' workflows as streamlined as possible.

We also started a community of translators working across several languages. We use the Scientific Python Discord server to communicate with, provide feedback, and relay announcements to the volunteer translators. We also plan to run a few translation sprints to help onboard new translators and ensure consistency and quality over different projects and languages.

Currently, translations are available for six different project websites, and target languages include Spanish, Chinese (Simplified), Japanese, Portuguese (Brazilian), Arabic, and Hindi.

Interactive Documentation

The Scientific Python Interactive Documentation project has made significant progress in making documentation interactive - meaning code examples in the documentation can be run in the browser, either inline or by launching a notebook in a separate browser tab - using JupyterLite and the jupyterlite-sphinx extension. Interactive documentation has been successfully deployed for projects such as PyWavelets, NumPy, and SciPy, with plans for other projects to follow in the future.

As part of this work, the jupyterlite-sphinx extension has been refined to improve usability and ensure sensible default behaviors, in collaboration with the extension maintainers and the community. We have also engaged with the Pyodide community to update package recipes in-tree in Pyodide, build WASM wheels in CI jobs of upstream packages to detect potential issues that may affect Pyodide early, and resolve issues related to Pyodide's ability to install and manage wheels for the projects, which is crucial for deploying up-to-date interactive documentation.

Stories: Packaging

The conda ecosystem is more alive than ever. What used to be a family of products developed by a single company is now a multi-stakeholder ecosystem of tools, specifications, and sub-communities.

The conda tooling keeps receiving updates. After shipping a new solver in 2023, we have started several plugins to keep improving it. conda-pypi will help you mix PyPl and conda dependencies more safely, while conda-checkpoints keep timestamped lockfiles of the different operations you performed in each environment. cph list can print the contents of a given conda package, local or remote, without needing to extract anything.

Thanks to the work started for the napari project, constructor and menuinst received many updates that will impact how users can build application installers using conda packages. From signing to customization, these contributions will enable installers like Miniforge to be built more robustly, moving away from the base environment and Windows-only desktop shortcuts.

conda-forge is the biggest community build-farm for conda packages. In 2024, it surpassed the two-million mark for the number of artifacts published! It currently serves an average of one billion downloads a month. Thanks to the funding provided by the CZI EOSS program, we were able to revamp the conda-forge.org website in collaboration with QuantStack. It is now built on top of Docusaurus and features a React-based status page with detailed views of the bot migrations. The redesign received special attention on the accessibility front, ensuring we passed the applicable WCAG criteria.

The new website is accompanied by a <u>Streamlit app</u> that allows you to browse conda artifact metadata without downloading them. This year, we added a new feature to locate artifacts that contain a given file path. This is backed by the public <u>condaforge-paths</u> project.

The build infrastructure also received some updates. In collaboration with MetroStar, we launched a <u>new GPU CI service</u> that the community can apply to use. This leverages a total of six GPUs housed in a 48-CPU server. So far, we have served 650K+ minutes for packages like PyTorch or TensorFlow. We also added compatibility with micromamba and Pixi-based provisioning in conda-smithy, which allowed us to drop the time it takes to run conda-build from 5 minutes to under one minute across all platforms! We are also leveraging cph list so you can inspect the built artifacts directly in the logs.

Both conda and conda-forge moved their primary chat platforms to Zulip. This was decided by their corresponding governing bodies in <u>CEP 18</u> and <u>CFEP 23</u>,respectively. We hope to see you there: <u>conda.zulipchat.com</u> and <u>conda-forge.zulipchat.com</u>.

Stories: Accelerating Sparse Array Computation

Much of the world's information is inherently sparse, with the data underpinning that information largely consisting of zeros or redundant values. While Python has become the dominant programming language for scientific computing, it has long struggled with sparse data. Scientists and engineers have been forced to treat sparse data as dense, leading to unnecessary inefficiencies, or rely on cumbersome workarounds that add complexity and slow computation. To address this gap, the Quansight Labs team has spent the past year collaborating with Saman Amarasinghe's lab at MIT as part of a DARPA SBIR grant to make sparse array computation in Python as seamless, performant and accessible as working with dense arrays.

At the core of our approach is integrating cutting-edge compiler techniques into PyData/Sparse, a widely used open source library for multi-dimensional sparse arrays. These techniques take inspiration from the Tensor Algebra Compiler (TACO), a project pioneered by our collaborators at MIT, which established a universal computational model for sparse array operations. By leveraging TACO's flexible code generation strategy, we can unlock optimizations across sparse array formats and ensure efficient execution on CPUs, GPUs, and other emerging hardware accelerators. At the same time, we have prioritized seamless integration with Python's scientific computing ecosystem by ensuring compliance with the Array API Standard—another foundational project spearheaded by Quansight Labs as part of the Consortium for Python Data API Standards. This alignment allows libraries such as SciPy and scikitlearn to work with sparse data with minimal code modifications and significantly lowers the barriers to adoption.

Over the past year, we have made significant progress. We successfully demonstrated a 100x speedup on key benchmarks compared to existing sparse array libraries while requiring less than a 5% modification to existing code.

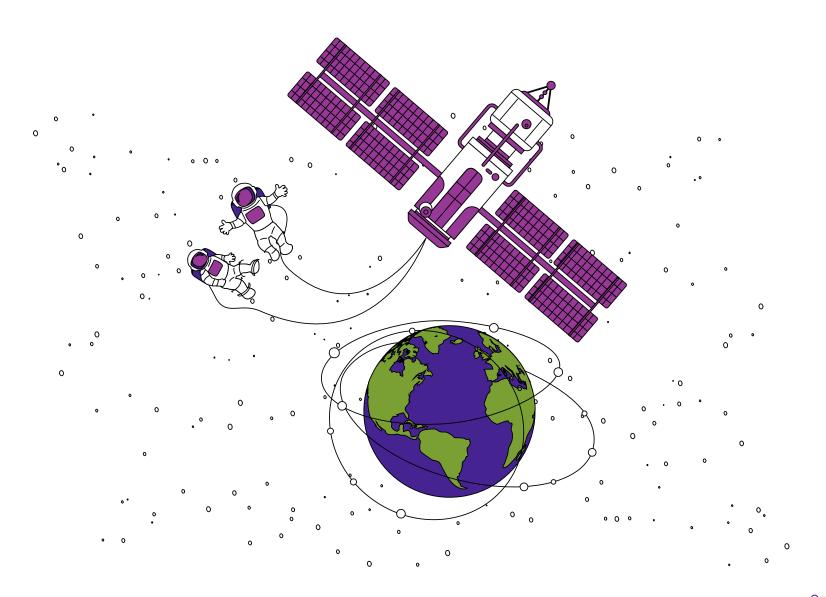
Our contributions to PyData/Sparse have made it a viable drop-in replacement for NumPy, allowing developers to take advantage of sparse computation without sacrificing ease of use. In parallel, our continued participation in the Consortium and our work on the Array API Standard have facilitated new standardization efforts for sparse array interchange, strengthening library interoperability, and reducing ecosystem fragmentation.

Despite these successes, three key challenges remain:

- First, supporting various sparse formats while maintaining high performance across different hardware architectures is a complex undertaking
- Second, ensuring compatibility with evolving Python standards requires ongoing collaboration with the broader scientific computing community.
- Third, transitioning from cutting-edge research to production demands rigorous performance testing and continuous refinement, particularly when optimizing for general-purpose CPUs and specialized accelerators.

¹ This collaboration is based upon work supported by DARPA under Contract No. HR001123C0139. Any opinions, findings and conclusions or recommendations expressed in this report are those of Quansight Labs and do not necessarily reflect the views of DARPA.

As we move into year two, our focus is shifting toward productionization, expanding support for GPUs, and refining a new MLIR-based compilation framework. We aim to bring these advancements to real-world applications, making sparse computation not just an academic exercise but a practical tool for Machine Learning, physics simulations, bioinformatics, and beyond. By pushing the boundaries of what's possible with sparse data in Python, Quansight Labs is laying the foundation for more efficient, scalable, and accessible scientific computing, thus ushering in the next era of accelerated array computation.



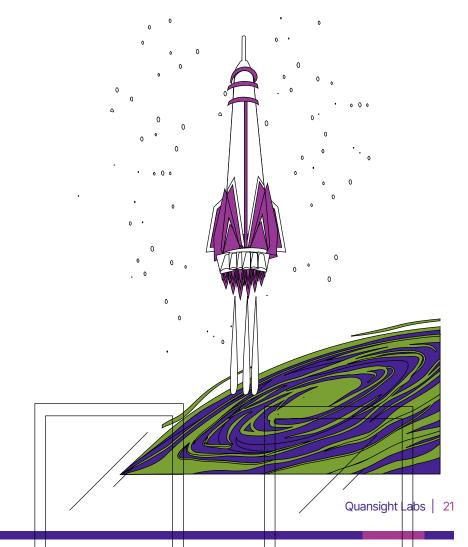
Stories: NASA ROSES

This year, we wrapped up work on a three-year grant from NASA, under its ROSES 2020 program, which supported significant feature and maintenance work across NumPy, pandas, SciPy, and scikit-learn. Close collaboration between maintainers from these four projects has yielded benefits like transferring knowledge of useful tools and techniques, sharing plans for new releases and deprecations, and joint work on cross-project new feature work. A good example of the latter was the variable-length string dtype project-most of the development work happened in NumPy, while a key intended user was pandas—hence NumPy and pandas maintainers worked jointly through the design and implementation to ensure the full pandas test suite passed made the design more robust. The story of this new dtype was told in this SciPy conference talk.

In 2024, a key focus of this project was to support <u>the release of NumPy 2.0</u> and its community-wide rollout. The <u>StringDtype</u> implementation, the Python API and documentation improvements, and the community support were in significant part made possible by this NASA grant.

Another key focus area was array API support in SciPy and scikit-learn, which added support for CuPy (GPU execution) and Dask (lazy/distributed execution) in at least one SciPy and one scikit-learn module. The team learned a lot in the process, and the effort to implement support for array libraries other than NumPy in scikit-learn and SciPy has gradually picked up steam throughout this grant. It has turned out to be a lot of work; however, the benefits–potentially significant performance improvements in particular–promise to be commensurate with that effort. There were also unexpected hurdles, such as the work to make scipy.special code run on GPU had to find a way to deal with shared libraries in Python packages.

The team was a nice mix of long-time maintainers and new contributors and yielded two new maintainers: Meekail Zain became a scikit-learn maintainer in 2022, and Nathan Goldbaum, became a NumPy maintainer in 2023. We hope to repeat this recipe in future grants - contributing renewed energy to the community is as important as the concrete code and documentation improvements.



Stories: Accessibility

This year, access-centered practices (which includes accessibility) continued to be a focus area for us within Labs. Whether as part of improving documentation tooling and themes, auditing data visualization libraries such as Bokeh, contributing to other projects through pull requests, design advice, or participating in community and project level discussions, developers and designers at Labs spent a substantial portion of their time focused on making our ecosystem more accessible to disabled users, and by extension, all of our community.

This year, the Labs team made significant progress towards making the PyDataSphinx Theme (PST), one of the most popular themes for library documentation, more accessible and improving its conformance with WCAG 2.1.

PyData Sphinx Theme Accessibility Audit

The Labs team conducted a comprehensive audit of the PyData Sphinx Theme in 2023, and in the past year, we have remedied the issues identified in the audit. This work included improving keyboard-only use and compatibility with several assistive technologies, improving color contrast across various components, revising all interactive components and their states, enhancing mobile support, enhancing navigation, and making structural adjustments to ensure conformance with WCAG accessibility standards. To our knowledge, PST is currently the first community-driven Sphinx theme to provide such a level of support for assistive technology and with a high level of conformance with WCAG standards.

PyData Sphinx Theme Design System

Continuing the iterative design work on the theme, we consolidated all design components into a <u>comprehensive design system</u>, <u>which was then published on Figma Community</u>.

The idea behind publishing the design system was to make it more accessible, transparent, and open while allowing other members of our ecosystem to build on and learn from all the accessible design work carried out over the last couple of years. As part of this effort, we added detailed design and accessibility documentation to the PyData Sphinx Theme. These resources support the community in implementing the theme effectively and serve as a guide for integrating best practices in accessibility and design within community-driven open source projects.

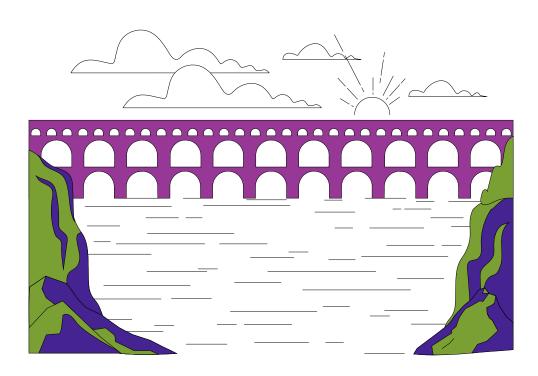
Stories: Narwhals

Narwhals is a lightweight, dependency-free compatibility layer between dataframe libraries. Its journey started one year ago, in February 2024, and it has since grown surprisingly rapidly, given how young it is. It is currently a required dependency of major libraries such as Altair, Formulaic, HierarchicalForecast, Marimo, Plotly, PyShiny, Vegafusion, and more.

The project has eager and lazy APIs to recognize people's different use cases for dataframes. The former assumes that row order is defined and supports operations such as arg_max and extracting Python scalars from Series. The latter makes no assumptions about row order and can, therefore, be used as an SQL frontend—we have recently added DuckDB support, and it's turning out much better than we expected.

The community aspect of the project is surprisingly strong. We have had contributions from nearly 100 unique people, and at least five-six people regularly attend our community calls. We have already done an in-person sprint at PyData Amsterdam, and most of us will meet at PyCon Italy this year.

Quansight, and its partner company, OpenTeams' support has been wonderful, initially via the POSSEE program and later directly funded by Labs. Our future priorities include improving integrations in the ecosystem, getting solid support for DuckDB and PySpark, making the library easier to extend, and defining the project's governance. We're having a real impact by addressing a gap in the ecosystem, and it's truly remarkable to be able to do this whilst at Quansight!

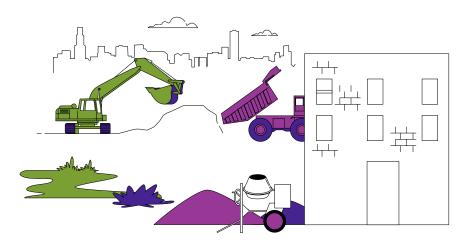


Labs Open Source Fund

Since the creation of Labs, we have continuously strived to refine our working model and drive new initiatives to support our efforts of building and sustaining open source projects and communities. This year was no different, as we launched our Quansight Labs OSS Fund. This annual fund will grant small sponsorships to open source projects that our ecosystem and communities rely on, all of which are nominated and selected by Labs staff members.

Through this fund, Labs provides a direct way for employees and interns to participate in the nomination and selection process to help communities and projects they care about that align with the Labs' mission and need extra support or funding.

The funds will provide direct financial support to maintainers to continue to sustain their projects. They can be used for all aspects of OSS maintenance, from fixing security vulnerabilities, adding new features, improving documentation, mentoring contributors, etc.



Projects must meet the following criteria to be eligible for the OSS Fund:

- The project must have an OSI-approved open source license.
- A Labs employee cannot own or maintain the project.
- The project must have a Code of Conduct.
- This cannot be a project for which Labs already has an active grant or funding source (directly or indirectly).
- The project has little to no significant structural income (relative to its size), i.e., it is mostly or entirely volunteer-driven.
- The project aligns with Labs' mission and values and/or is an upstream dependency of projects we support and sustain.
- The project is actively maintained.

The three awardees of our inaugural Labs OSS Fund are:



CodeMirror

CodeMirror is a code editor component for the web. It can be used in websites to implement a text input field that supports many editing features and has a rich programming interface to allow further extension. It is used by many IDEs, including JupyterLab and Jupyter Notebook.



Mkdocstrings

MKdocstrings is a versatile and language-agnostic MKDocs plugin that enables the automatic generation of documentation from source code.



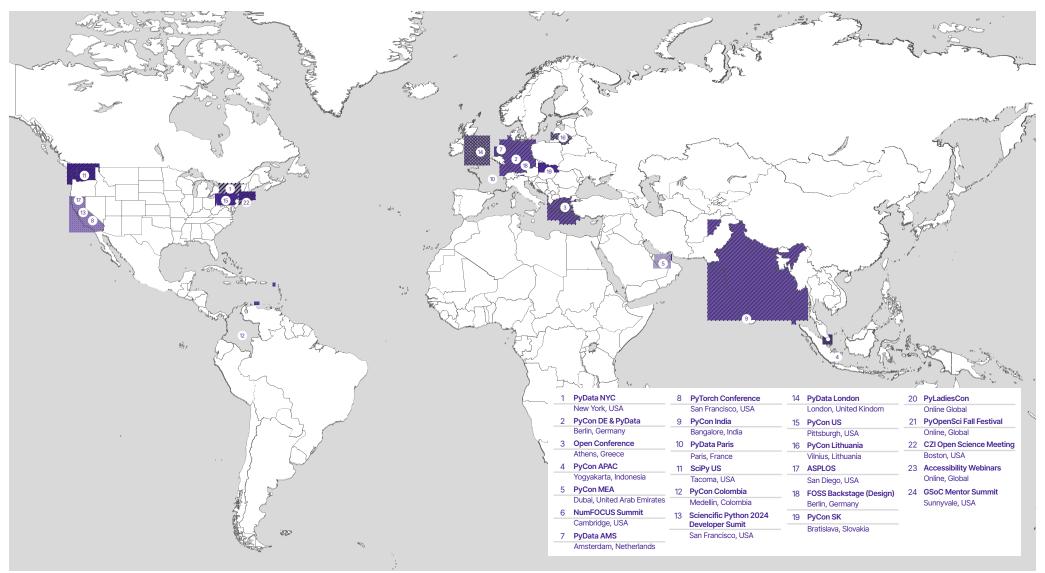
Maturin

Maturin is a tool to build and publish crates with pyo3, cffi, and uniffi bindings and Rust binaries as Python packages with minimal configuration. It supports building wheels for Python 3.8+ on Windows, Linux, macOS, and FreeBSD, can upload them to PyPI, and has basic PyPy and GraalPy support.



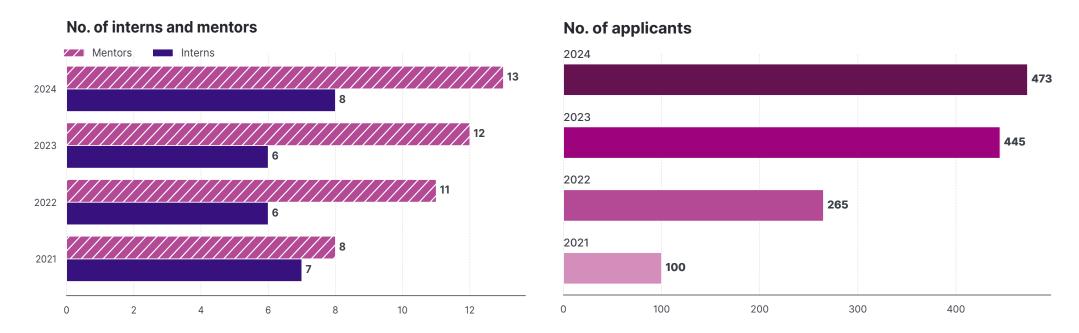
Conferences And Events

Labs members participated in over 20 community conferences and events worldwide, including various PyCon and PyData events, SciPy US, Scientific Python Developer Summit, PyLadiesCon, PyTorch Conference, and many more!



Labs Open Source Internship Program

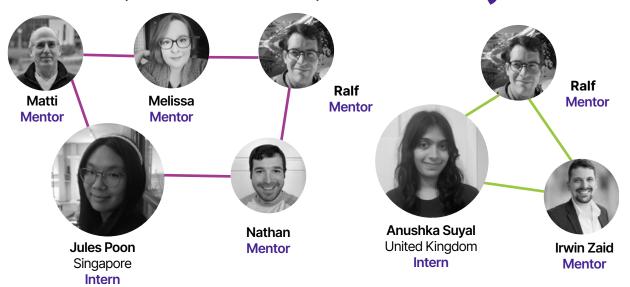
The Labs open source internship has become a staple of our organization and one of our favorite initiatives. This year, we ran the fourth iteration of the program, the most successful to date. We received 473 applications, and 13 mentors supported eight new open source sustainers.



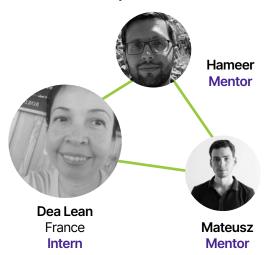
"I worked on the PyData/sparse library. I was given very specific objectives prepared by the core developer, Hameer Abbasi. He was also my mentor. I overhauled the library's website and introduced automated benchmarking in the CI, among other things. Every day was a learning day, not only because of my goals, but because Labs has an environment of learning and sharing. The organization and management of the mentorship was perfect."

Dea María Léon, Quansight Labs Intern 2024

Interns, Mentors, and Projects



Project: Algorithms for data approximations with splines for SciPy.



Project: Sparse documentation and maintenance tasks.

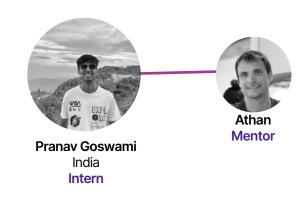
Project: Support for sparse tensors in SciPy.



Project: Implementing a new extended precision data type for NumPy.



Project: Multidimensional cubature integration methods in SciPy.



Project: Extending support for BLAS and LAPACK in web environments by way of stdlib.js



Project: Bokeh accessibility testing.



Project: Polars plugins super-extra-accessible.



Running the Labs Open Source Internship

The Labs Internship Program is a key annual initiative at Labs, offering interns a comprehensive and immersive experience. Over three months, interns actively participated in all Labs' workspaces and meetings. With mentorship and training, they contribute to high-impact open source libraries. We ensure that each internship project is well-defined and strategically aligned with the roadmap of the respective open source projects.

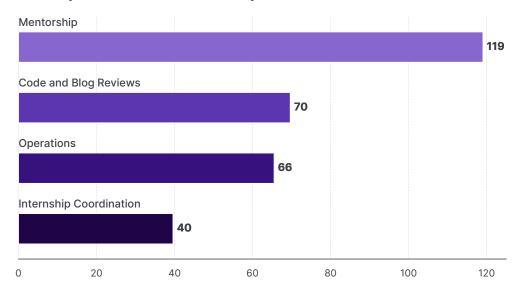
The success of this program requires significant planning, coordination, and execution, as well as an intentional approach to project/intern matching. Labs members dedicated **119 hours** to mentoring interns this year and **40 hours** to program coordination. It is also important to note that functions like evaluating applications, interviewing candidates, onboarding interns to our systems, and other back-office operations account for a substantial amount of effort (a total of 66 hours), so running a mentorship program like this one certainly takes a village with a diverse set of skills.

Our goal is to cultivate supportive and collaborative open source contributors. To achieve this, we encourage interns to engage with open source project communities and participate in ongoing initiatives beyond their individual projects. Throughout the program, interns join weekly check-ins with their peers to share progress and receive feedback. We also provide training sessions based on interns' feedback. At the conclusion of the internship, they present their final results internally to the Quansight team and share their achievements with the broader community through a

blog post. You can read about past internship projects on the Labs Blog.

Seeing our interns thrive within the community is incredibly rewarding, and this initiative continues to be a meaningful and impactful experience for everyone involved.

Hours per task - OSS Internship



How the Labs Internship Works

- OSS sustainers at Labs propose well-scoped internship projects that align with their projects' community roadmap. The projects can span all aspects of open source, including but not limited to documentation, design, and community management.
- Interns are selected through a resume review and interview and are evaluated for project-skill-fit and interest fit. We have noticed this pairing to be a critical part of our internship success.
- During the internship, interns are encouraged to join the OSS project community as supportive and collaborative members.
- As part of the Labs team, interns participate in company and team meetings and have access to the Quansight spaces and resources and some paid time off (PTO) days.
- There are weekly cohort meetings and synchronous and asynchronous checkins between interns and their mentors to ensure interns have sufficient support throughout their internship.
- We do a mid-internship check-in with the mentors to ensure they have all the support they need and check on the project's progress.

- We provide training sessions for interns, including focused sessions on OSS Contribution, CV clinics, code debugging, and more.
- Towards the end of the internship, interns present their work at a Quansight All Hands meeting and write blog posts to share their projects and learning with the broader community.
- Interns often also stay in touch after their internship as Quansight alumni.



Sustaining Labs



Our Community-First Working Model

Our goal at Labs is to facilitate the long-term sustainability of our ecosystem. As a collective, Labs members have vast experience as users, contributors, sustainers, and leaders within critical open source projects and their communities—sometimes spanning a couple of decades.

The following are our core principles guiding how we think about and approach open source sustainability, contributions, collaborations, and engagement with the open source community and our many stakeholders.

The OSS Community Is the Primary Stakeholder

We collaborate with other community members and leaders when working on community-driven projects. That means we favor decision-making processes such as consent or consensus-seeking instead of making unilateral decisions. Additionally, as individuals employed to work on community-driven open source projects, we must be aware of work/load/labor distribution and ensure we are not overloading volunteers in the project. Instead, our work should make their open source participation easier or more enjoyable.

Some ways in which we act on these principles are:

Open issues and shepherd discussions to solicit feedback and input on any signifi-cant efforts or pieces of work.

- Actively seek code reviews or alternative feedback mechanisms from contributors outside Quansight Labs.
- Follow a project's decision-making processes and enhancement proposal processes.
- Invite active participation and collaboration from community members. Our work is easily accessible by the community by leveraging community spaces, open tools and processes for project management, and regularly presenting updates to community stakeholders.
- Routinely assess our progress and allow for and invite feedback from community members.
- Prioritizing work and approaches that help build a healthy, inclusive, and equitable culture, such as seeking collaborations with community members outside Quansight Labs.
- Dedicate substantial time and effort to day-to-day maintenance tasks such as triaging, reviewing contributions, welcoming new contributors, fixing bugs, and updating documentation.
- Being generous on sharing attribution and acknowledging others' efforts and contributions. This includes recognizing and acknowledging all types of contributions, not only those related to code.

- ☐ Encourage community-driven projects to develop and maintain roadmaps and support them in this effort.
- Prioritize work that builds a healthy, inclusive, and sustainability-focused culture in our communities and projects.

Funding for Open Source Projects

Funding opportunities should align clearly with the project and its community's needs and goals. To ensure all our funding-seeking and execution tasks align with this principle, we have adopted the following guidelines:

- Engage with the appropriate leadership groups before seeking funding. Such an approach ensures alignment on the scope, deliverables, and ownership. Also, it allows for the inclusion of other community members—outside of Labs—to collaborate on funding proposals and execution of tasks.
- ☐ Treat collaborators as equal partners.
- Our default way of working is transparent and open; this helps build trust with other interested parties and makes it easier to hold ourselves accountable. We meet this principle in many ways, from making submitted grant proposals public to ensuring our work is easily accessible by the community by leveraging community spaces, open tools and processes for project management, and regularly presenting updates to community stakeholders.

- Routinely assess our progress and allow for and invite feedback from community members.
- Prioritize knowledge sharing and long-term maintainability.
- Prioritize resource sharing with individuals and organizations serving historically marginalized or disenfranchised groups.
- When receiving dedicated funding for an open source project, Labs is a vehicle for the community's interests.



Co-Ownership Governance Model

Quansight Labs adopted a feminist co-leadership model of governance founded on principles of shared ownership, inclusiveness, collaboration, participation, and empowerment. We chose this model as it can help decentralize power dynamics further and support collective leadership at all organizational levels. This has allowed us to apply this shared leadership model across multiple levels within Labs, including Labs' strategic areas. Some of the benefits we've experienced through adopting models of co-leadership for strategic areas are:

- Robust, decentralized, and effective decision-making
- Diversification of ideas and perspectives at strategic levels
- Increased transparency and enhanced delivery
- Continuous assessment of and iteration of strategic priorities and goals
- Increased awareness of our communities' needs and enhanced participation

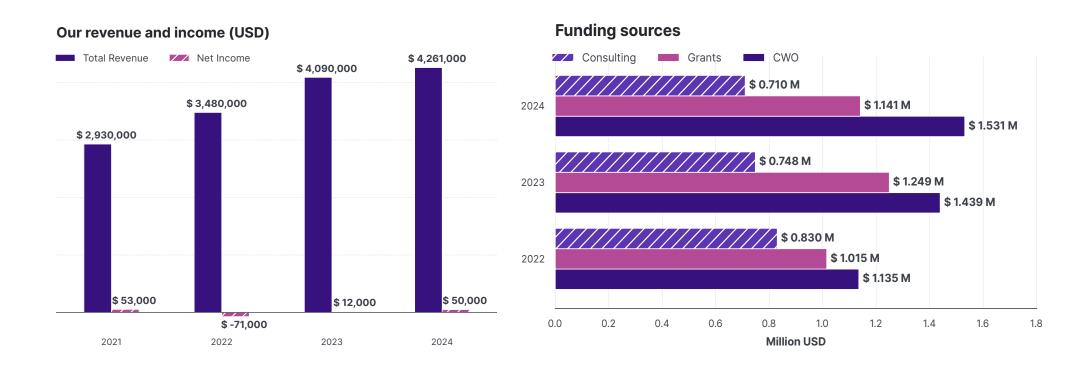
We have now extended our adoption of the feminist co-leadership model for all of our strategic areas within Labs.



Financials

We run Quansight Labs as a public benefit organization. We don't intend to make a profit but rather work at cost on community-driven projects and topics from which we think the community will benefit.

We aim to support our work through a diverse set of income streams. In 2024, our revenue grew very slightly to just over \$4,200,000 USD, with a net income of \$50,000 USD. This income came from three types of sources: community work orders (CWO), grants, and consulting activities.



Noteworthy Funding Sources

This year, our CWO funding remained constant compared to the year before, with a significant amount of support from the Chan Zuckerberg Institute for work on napari, from Anaconda in support of work on conda and the wider conda community, and from the Sovereign Tech Agency to work on OpenBLAS through their Sovereign Tech Fund and conda-store through their Contribute Back Challenges program.

Most of our grant funding came from the Chan Zuckerberg Institute, through its <u>Essential Open Source for Science program</u>, with new two-year grants for Data APIs, Bokeh, and scikit-learn starting mid-2024, and a grant for Scientific Python, and NASA, through the software component of its <u>Research Opportunities in Space and Earth Sciences (ROSES)</u> program.

Consulting income comes from projects done with a variety of industry clients. This income is used to support key unfunded work on community projects. The consulting activities allow open source maintainers in Labs to gain experience with how their projects are used in the industry.



Diversity Within Labs

Quansight Labs is committed to building a diverse and inclusive organization that reflects our world. This principle guides our work at Quansight Labs, from informing existing and new internal policies to supporting or spearheading community initiatives.

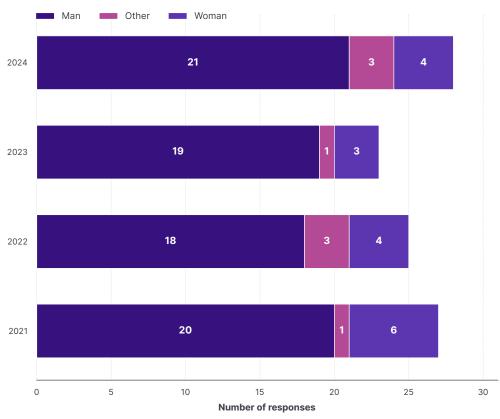
Every year for the last four years we have run an internal voluntary demographics survey. Though voluntary, the participation rates for this survey have been between 90% and 100%, an unusually high participation rate for surveys. Collecting and analyzing data through our hiring processes and through our annual demographics survey has provided incredible insights on Labs' members and has enabled us to ensure everything we do within Labs aligns with the goal of building an equitable and supportive workplace. These data have become invaluable tools for all of our decisionmaking processes.

After four years of data collection, we decided to examine the composition of the Labs' teams over time.

Note that the plots presented here show the number of responses collected throughout the survey, not the number of Labs members. Unless explicitly added and labeled, responses marked as 'Prefer not to say' were removed from the plots. Every question asked through our survey had the option to answer 'Prefer not to say', as well as the option to self-describe, or to skip the question entirely.

While the majority of respondents identify themselves as men, between 17 and 28% of our respondents between 2021 and 2024 identify themselves as women, non-binary, genderqueer, genderfluid, and other gender minorities (grouped under 'Other' to avoid potential identification during analysis).

Gender identity distribution of respondents 2021-2024



55% of our employees have been part of Quansight Labs for 2 or more years. Between 2021 and 2024, 40% to 48% of our respondents self-identified as belonging to a marginalized group in tech and open source based on characteristics such as age, disability status, ethnic/racial background, neurodiversity, socio-economic status, sexual orientation, and gender identity.

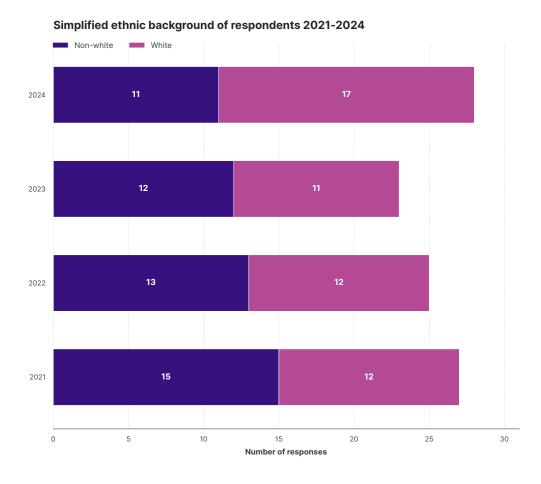
Between 2021 and 2024, 44 to 61% of the Lab members identified as white, while the rest of the respondents identified themselves as belonging to other ethnic backgrounds (including mixed backgrounds). We decided to present these as simplified numbers/categories to avoid potential individual identification.

Self identification as belonging to a marginalized group 2021-2024 Prefer not to say 15 2024 12 2023 15 2022 16 2021

15

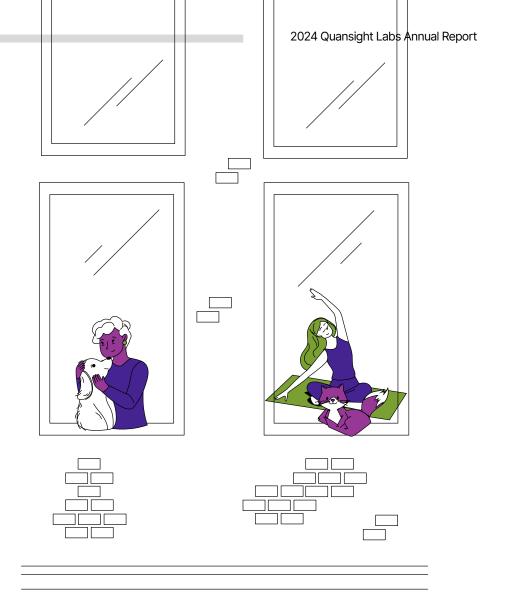
Number of responses

20



Finally, we have seen small shifts in the age ranges of Labs' members, which are directly related to the increasing length of tenure. We continue to have very low attrition numbers.

Age distribution of respondents 2021-2024 26-35 36-45 46-55 56-65 3 2024 2023 2022 2021 10 20 25 30 15 Number of respondents



Building a Sense of Belonging

Quansight Labs is a remote-first organization with a globally distributed team. To maintain strong connections and foster collaboration, we hold regular virtual and inperson gatherings. This includes quarterly virtual all-hands meetings to align on our strategic vision and share project updates, as well as monthly informal company-wide knowledge-sharing sessions. Additionally, we host an annual in-person retreat (Camp Quansight) to connect on a personal level. Our most recent retreat in Monte Isola, Italy, combined work sessions and social activities—it's a cherished and highly anticipated yearly event.







How Far We've Come And Looking Ahead

2024 has been another year full of highlights, surprises, and challenges. From the NumPy 2.0 release, Narwhals taking off, and another amazingly impactful internship program to two Labs members gaining co-authorship on the PyTorch 2 paper and seeing free-threaded Python coming to life, it was an eventful ride.

Challenges often lie in the personal sphere, outside the work environment at Labs. We're fortunate to have a very mutually supportive environment and team. When the unexpected happens, whether related to physical or mental health, family or relationship, or anything else, we do what we can. We're intentional about improving our support structures with this in mind. Hence, we were pleased to be able to implement a sabbatical policy this year to better support the Labs team.

Looking ahead, we see significant changes in the funding landscape coming. Grant funding from institutional funders and private foundations-for open source software and scientific software—will likely decline in 2025 for the first time since we started Quansight Labs. On the other hand, the reliance of the corporate sector on open source—and the PyData ecosystem in particular-continues to grow. We do expect industry funding for community-driven open source to continue to increase. We don't expect to grow in 2025, for the first time since Labs was founded. And we don't need to grow - we have a healthy size, and we'd like to focus on the impact of what we do with our current team.

In the coming year, we intend to continue prioritizing efforts that make the open source projects we support more accessible to groups of users who are not always well served by "business as usual." This includes our work on accessibility (for users with disabilities), interactive documentation (for beginners), and internationalization (for users whose English is not their primary language or a language they speak at all).

In the numerical computing space, a lot is happening - goals include stabilizing GPU support in SciPy and scikit-learn, and helping free-threaded CPython move to its next adoption phase. Also, more Rust! Between Polars, PyO3, Pixi, and Rattler, our team members have gotten significantly more involved in Rust projects, and we expect that to continue. Finally, packaging is always on the agenda: further improving PyTorch support in conda-forge (an unexpected end-of-year sprint!), supporting conda-forge in its transition to rattler-build, making progress towards accelerator support on PyPI, and probably much more!

We continuously balance driving new topics and sustaining existing projects while keeping them healthy. Both matter. We need to do the former because we need to deliver improve-ments and changes that add value—maintaining the status quo and keeping up with external changes isn't enough. We need to do the latter to better serve our communities and because, as people fortunate to be paid to work on projects that are still largely volunteer-driven, we need to do more than our fair share of the boring but necessary maintenance work.

See you on the other side!

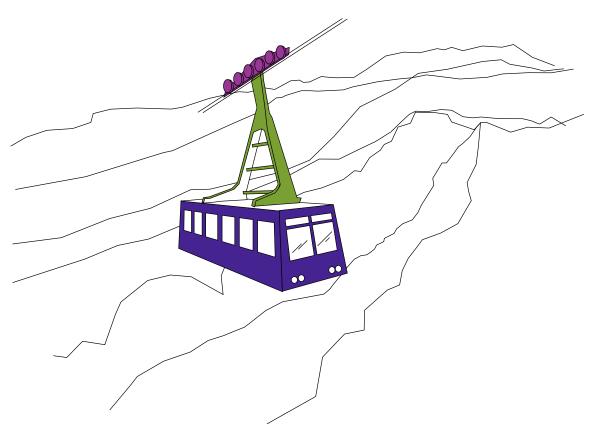
Acknowledgements

On behalf of Quansight Labs, this year's report is presented to you by Pavithra Eswaramoorthy, Tania Allard, and Ralf Gommers.

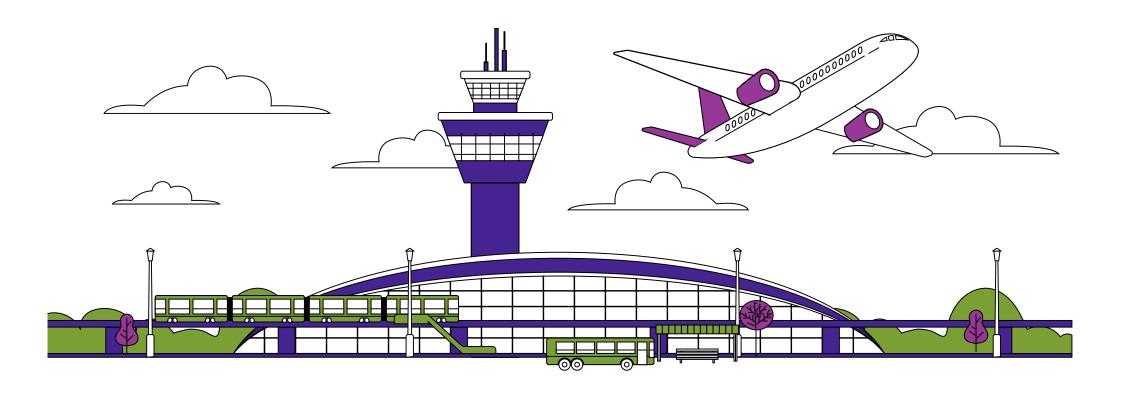
The report is a team-wide effort. Athan Reines, Jaime Rodríguez-Guerra, Marco Gorelli, Melissa Weber Mendonça, Nathan Goldbaum, and Smera Goel contributed the detailed stories. Agriya Khetarpal, Daniel Althviz, and Hameer Abbasi shared project-specific highlights, and Pierre-Olivier Simonard helped us gather data for time-spent calculations.

All the beautiful illustrations and designs are by Irina Fumarel and Paul Otieno, with support from Kate Brack.

A special thanks to Ashley Baal and Athan Reines, who facilitate regular internal communication at Quansight. Their efforts in gathering and sharing updates year-round were essential in curating this report.



See you next year!



Visit labs.quansight.org to learn more about us, or quansight.com to learn about consulting opportunities. Stay in the loop by subscribing to Quansight Quarterly or reach us at connect@quansight.com.









