NUMFOCUS

OPEN CODE = BETTER SCIENCE

2018 ANNUAL REPORT

TABLE OF CONTENTS

Cover photo by Ana Ruvalcaba

LETTER FROM THE EXECUTIVE DIRECTOR AND PRESIDENT

Our 2018 Annual Report highlights the year's achievements and presents an overview of the work your generosity and support has made possible.

Since its founding in 2012, NumFOCUS has served the contributors and users of open source tools responsible for escalating scientific research and discovery. In 2018, we kicked off new initiatives focused on creating a more robust infrastructure to support our efforts over the long term. Procedures were implemented to increase the level of fiscal, operational and legal services provided to our projects, allowing them to focus on the maintenance and management of these critical tools. This year also saw growth in our community-centered programming directed toward supporting project users through education, networking, and inclusivity initiatives.

The success of NumFOCUS is embodied in the accomplishments of our community. We encourage you to take a closer look at the following pages.

LEAH SILEN
NumFOCUS
Executive Director

ANDY TERREL
NumFOCUS
President

PROJECTS...

5 new fiscally sponsored projects were brought on board in 2018, bringing our total number to 25. These included **Bokeh**, an interactive visualization library targeting modern web browsers for presentation; **Cantera**, a software suite solving problems in thermodynamics, chemical reaction rates, and fluid transport processes; **Conda-forge**, a tool for building and distributing unique software packages; **JuMP**, a modeling interface and collection of supporting packages for mathematical optimization embedded in Julia; and **Xarray**, a library providing data structures and analysis tools for working with multidimensional labeled datasets and arrays in Python.

2018 saw numerous project accomplishments: AstroPy released v3.0 and v3.1; **Julia**, after nearly a decade of work released 1.0; **Jupyter** introduced a series of new community workshops; **Matplotlib** brought on three new core developers; **PyMC3** began collaboration with TensorFlow Probability on the design of PyMC4, and **Shogun** began collaboration with the Alan Turing Institute in London.

We were thrilled to award **22 Small Development Grants** to 16 Sponsored and Affiliated projects. The $60,000 in funding was used to modernize documentation, hold sprints and hackathons, fund tutorials, update websites, redesign UX, and complete stalled pull requests.

Two undergraduate students were given the opportunity to work for 3 months on Matplotlib as part of the **John Hunter Matplotlib Summer Fellowship**. Working under the mentorship of a senior contributor, students gained the needed experience and knowledge to become active contributors and potential core maintainers of Matplotlib.

ON THE COMMUNITY FRONT...

One of the year's most exciting initiatives was the **Pandas Worldwide Documentation Sprint** organized by pandas core developer, Marc Garcia. 500 community members in 30 locations participated in the one-day sprint to improve pandas documentation and experience contributing to an open source project, many for the first time.

We're thrilled to feature the profiles of four **outstanding community members**: Reshama Shaikh, organizer of NYC Women in Machine Learning & Data Science and NYC PyLadies; Avik Sengupta, Julia Contributor and JuliaCon Organizer; Noa Shinitski, PyData Berlin Organizer; and Marc Garcia, pandas Core Developer.

The **3rd Annual NumFOCUS Summit** was hosted by Microsoft in NYC. Project leaders from 23 projects as well as other stakeholders were brought together to address project sustainability. Workshop topics included Community Roadmaps, Governance, Fundraising, and Community Culture/Conduct.

The positive impact of events either hosted or supported by NumFOCUS was experienced by more than **6,400 attendees**. The combination of knowledge sharing and networking that takes place at NumFOCUS events is irreplaceable when it comes to strengthening both the project user base and the NumFOCUS ecosystem as a whole.

PyData events spanned 5 continents and 13 countries. As a major fundraiser for NumFOCUS, we thank all those who attended and sponsored for your donations.

DONORS...

Corporate Sponsors play a vital role in our ability to serve and sustain our projects. We applaud your leadership in support of, and participation in, open science research as well as your recognition of the need to ensure the sustainability of the tools that are changing the trajectory of science and industry. 2018 Corporate Sponsorship increased over $100,000! We encourage everyone to personally thank these corporations for their impact on our community.

We're also proud to report that the number of both **individual donations and sustaining memberships more than doubled** in 2018. These NumFOCUS members, many of whom pledged a recurring donation, have stepped up to not only support the projects but to ensure our work in support of these tools continues.

INTERNAL STAKEHOLDERS...

We welcomed **5 new board members** this year and look forward to their continuing leadership and vision. Board members volunteer their time as stewards of our mission and we are grateful for their efforts on behalf of the community.

It is no small task to serve a dynamic, fast growing community but the **staff** of NumFOCUS does an incredible job. There is an abundance of detail that comes with managing the financial, legal, and logistical administration of the 25 Sponsored and 25 Affiliated projects we serve. Our staff works each day to provide the operational foundation for our projects to succeed.

We would love to have you involved. Please visit **numfocus.org** to learn how you can participate or contact us at info@numfocus.org.

PROJECTS

NEW SPONSORED PROJECTS

NumFOCUS welcomed five new sponsored projects in 2018:

Bokeh

Bokeh is an interactive visualization library that targets modern web browsers for presentation. It can help anyone who would like to quickly and easily connect powerful PyData tools to interactive plots, dashboards, and data applications.

Bokeh helps people create rich explorations of data and models on the web, from wherever they are already comfortable and productive (i.e. Python or R). Its goal is to provide elegant, concise construction of versatile graphics and to extend this capability with high-performance interactivity over large or streaming datasets.

Bokeh is used for finance, management, big data mining, graphing business patterns, web applications such as weather data and renewable energy forecasts, viewing energy production data, disaster resource management, displaying agricultural sensor data online, analysis of data from industrial machinery (wind turbines, trains, power plants), and mapping clinical trials using AERO plots, among many other uses.

Cantera

Cantera is an open source software suite which helps users solve problems involving thermodynamics, chemical reaction rates, and fluid transport processes. The software is written to be flexible and efficient, handling these calculations in a way that lets the user shift their attention to other elements of their calculations. Because it is open source, the software can be easily extended to support the needs of any particular researcher.

Cantera has been adopted by a wide range of users across diverse research fields. Its largest user base is in the combustion field, but Cantera has also seen applications in energy storage (such as batteries and fuel cells), geochemistry, chemical processing, thin film deposition, plasma science, and atmospheric/astronomical chemistry.

Conda-forge

Conda-forge builds and distributes software packages, specializing in the hard-to-build or unique packages that often arise in a scientific computing context. Conda-forge is community-driven and community-curated. This means that no package is too domain-specific and all packages undergo review to ensure quality and interoperability.

Conda-forge has a federated structure, where each package has its own list of maintainers, which anyone can join. This helps ensure that people who care about a package are the ones responsible for maintaining it. The conda-forge ecosystem is fully structured around automation. This lets maintainers build packages for systems they do not own (like Windows or Mac), keep their packages up-to-date automatically, and ensure their packages maintain the highest level of quality.

Conda-forge is the preferred way to install large parts of the PyData and R ecosystem. It is used throughout a disparate set of scientific computing domains, from physics to biology to financial technology and everything in between.

JuMP

JuMP is a modeling interface and a collection of supporting packages for mathematical optimization that is embedded in Julia (also a NumFOCUS sponsored project). With JuMP, users formulate various classes of optimization problems with easy-to-read code and then solve these problems using state-of-the-art open-source and commercial solvers. JuMP also makes advanced optimization techniques easily accessible from a high-level language.

JuMP's design leverages advanced features of the Julia programming language to allow users to express complex mathematical optimization problems with a natural notation that mirrors what a user might write on paper. Through this convenient and expressive syntax, JuMP lets users access advanced tools that were previously restricted to low-level proprietary interfaces.

JuMP has been used for routing school buses in Boston, scheduling trains in Canada, and simulating power system operations in North and South America. It is used for teaching and research at numerous universities worldwide. Features and extensions have been developed in collaboration with researchers at universities and national laboratories. A wide range of research papers have cited JuMP. In addition, it is used in many of MIT Sloan's business analytics courses in the undergraduate, MBAn, and MBA programs.

Xarray

Xarray is an open source library providing high-level, easy-to-use data structures and analysis tools for working with multidimensional labeled datasets and arrays in Python. Xarray enables users to perform operations on complex datasets. It interoperates with many of the core libraries in the scientific Python ecosystem, making xarray a powerful high-level tool for data analysis.

Xarray has been used in a wide variety of academic and industry contexts for applications as varied as weather/climate, computational physics, astronomy, biology, econometrics, machine learning and finance. It is a core component of Pangeo, a community platform for Big Data geoscience. Examples of results enabled by xarray include modeling the environmental and socioeconomic impacts of climate change; understanding the life cycle of viruses from single-cell RNA sequencing data; and measuring the speed of galaxies in a telescope survey.

2018 PROJECT HIGHLIGHTS AND ACCOMPLISHMENTS

Astropy:

- ❯ Released v3.0 and v3.1 of the Astropy core package, which are the first Python 3-only versions
- ❯ We published a paper about Astropy in the Astronomical Journal: "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package" (DOI: 10.3847/1538-3881/aabc4f)
- ❯ We passed the 300-contributor mark for the Astropy core package

Bokeh:

- ❯ Joined NumFOCUS!
- ❯ After about 5 years of active development, Bokeh 1.0 was released in October 2018

Cantera:

- ❯ Joined NumFOCUS!
- ❯ Celebrated the release of version 2.4.0 that merged over 70 pull requests and closed over 60 issues
- ❯ Received two NumFOCUS Small Development Grants, leading to a redesign of our website and allowing the core developers and Steering Committee to meet in-person
- ❯ Presented two workshops at regional meetings of the Combustion Institute for nearly 75 Cantera users

Conda-forge:

> Joined NumFOCUS!
> Migrated entire ecosystem from old compiler toolchains to new compilers
> Developed a new auto-tick bot for automatically keeping packages up-to-date
> Became a NumFOCUS fiscally sponsored project

Econ-ARK:

> First pip/conda installable version of Econ-ARK
> Workshops introducing the toolkit at the European Central Bank, the Bank of England, the Bundesbank, the Computing in Economics and Finance conference, and a number of other places
> Hired Patrick Mogensen as a full time mid-doc employee to build some crucial new tools
> Paper describing the Econ-ARK published as part of the proceedings for SciPy conference

FEniCS:

> We released FEniCS Project version 2018.1.0, completing our transition to Python 3
> Nate Sime and Michal Habera (former Google Summer of Code student) joined the Steering Council
> The 2018 edition of the FEniCS Conference was held at the Mathematical Institute, University of Oxford

Julia:

> Julia 1.0 was released in August after nearly a decade of work

JuMP:

> Joined NumFOCUS!
> Hosted over 40 attendees at the JuMP-dev workshop in Bordeaux
> Released JuMP 0.19-beta, which includes a major rewrite of JuMP's infrastructure

Jupyter/IPython:

> Introduced a new community workshops series which gained traction and significant interest from contributors
> JupyterCon was hosted in NYC and featured over 100 speakers, 17 sponsors, and hundreds of attendees
> Various publications: paper on nbgrader, a paper on Binder 2.0, a paper on Reproducible Research Environments with Repo2Docker
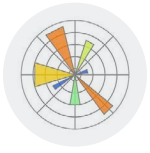
Matplotlib:

> Thomas Caswell officially became lead developer, brought on 3 new core developers (Tim Hoffman, Ernest, Dietmar Schwertberger)
> 200 unique code contributors for the year (average ~26 unique contributors per month)
> One major release (3.0), one minor release (2.2), 8 total patch releases

Nteract:

> nteract released version 0.17 of papermill, a tool for parameterizing, scheduling, and executing Jupyter notebooks that was featured at this year's JupyterCon
> nteract released new documentation to improve contributor and user engagement including documentation for our React components and JavaScript packages, content guides for our documentation writers, and user guides for our users and contributors
> nteract migrated our main code repository to the popular TypeScript language and released alpha versions of new packages that will be used in our implementation of real-time collaboration in notebooks
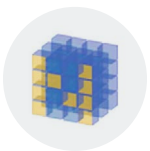
NumPy:

> Started work on grant from Sloan and Moore foundations in the first half of the year. NumPy now has two funded full-time developers for the first time in its history
> Largely as a consequence, we reinvigorated the NumPy Enhancement Proposals (NEPs) as a process to propose and discuss new features, which has resulted in a big step forward towards making the NumPy functions a general API for numerical processing of any types of arrays
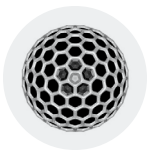
Open Journals:

> Partnered with AAS publishing
> Issued a call for editors and grew our editorial team
> Blew through 500 submissions! (currently at ~550)

pandas:

> World wide documentation sprint in March 2018; 150 pull requests in 24 hours with participation from 30 countries
> 0.23.x series of releases
> First ever in-person development sprint in July in Austin, brought together 12 core devs for discussions on pandas future

PyMC3:

> Won ODSC Open Source Data Science Project award
> Began collaboration with TensorFlow Probability on the design of PyMC4
> Removed Python2 support from PyMC3

PyTables:

> Developed an interface for PyTables and Datasette thanks to support from NumFOCUS Corporate Sponsor ActiveState. Thanks to Javier Sancho for this cool contribution
> Received NumFOCUS Small Development Grant for better support for native HDF5 files, especially in the context of Tables (compound datatypes in HDF5 parlance)

QuantEcon:

> Launched QuantEcon Notes Platform
> New build system that executes code to generate output in QuantEcon Lectures
> Partnership with University of British Columbia (Collaborate on Julia Lectures)

rOpenSci:

> Expanded our open software peer review process: doubled size of editorial team (now 8); quintupled number of peer reviewed packages (43) compared with 2016
> Increased impact of our community: more than half of our R packages (78%) and blog posts (58%) were contributed by community members; gained over 300 new contributors; our software was cited in 241 use cases
> Brought on new Research Software Engineer, Maëlle Salmon

Shogun:

> Started collaboration with the Alan Turing Institute in London, welcoming paid developers Gil Hoben and Lefteris Parasyris
> Had a terrific Google Summer of Code under the NumFOCUS umbrella, bringing in new Shogun contributors Wuwei Lin and Shubham Shukla
> Organized the Machine Learning Open-Source Software (MLOSS) workshop at NeurIPS '18

Stan:

> Created a Stan governing body
> Developed several new major features, including within-chain parallelization
> Held two Stan conferences, one in California and one in Finland

SunPy:

> Moved to using Astropy Time throughout the code base with the help of our GSOC student Vishnunarayan K I
> SunPy got involved in the Python in Heliophysics organization to help the growth of Python and Open Development projects in the wider Heliophysics community
> The SunPy project released a new package "NDCube" which provides tools for working with multi-dimensional arrays with functional coordinate systems provided by astropy WCS

SymPy:

> Two major releases (SymPy 1.2 and SymPy 1.3)
> 7 successful GSoC projects

Xarray:

> Joined NumFOCUS!
> Published a development road map
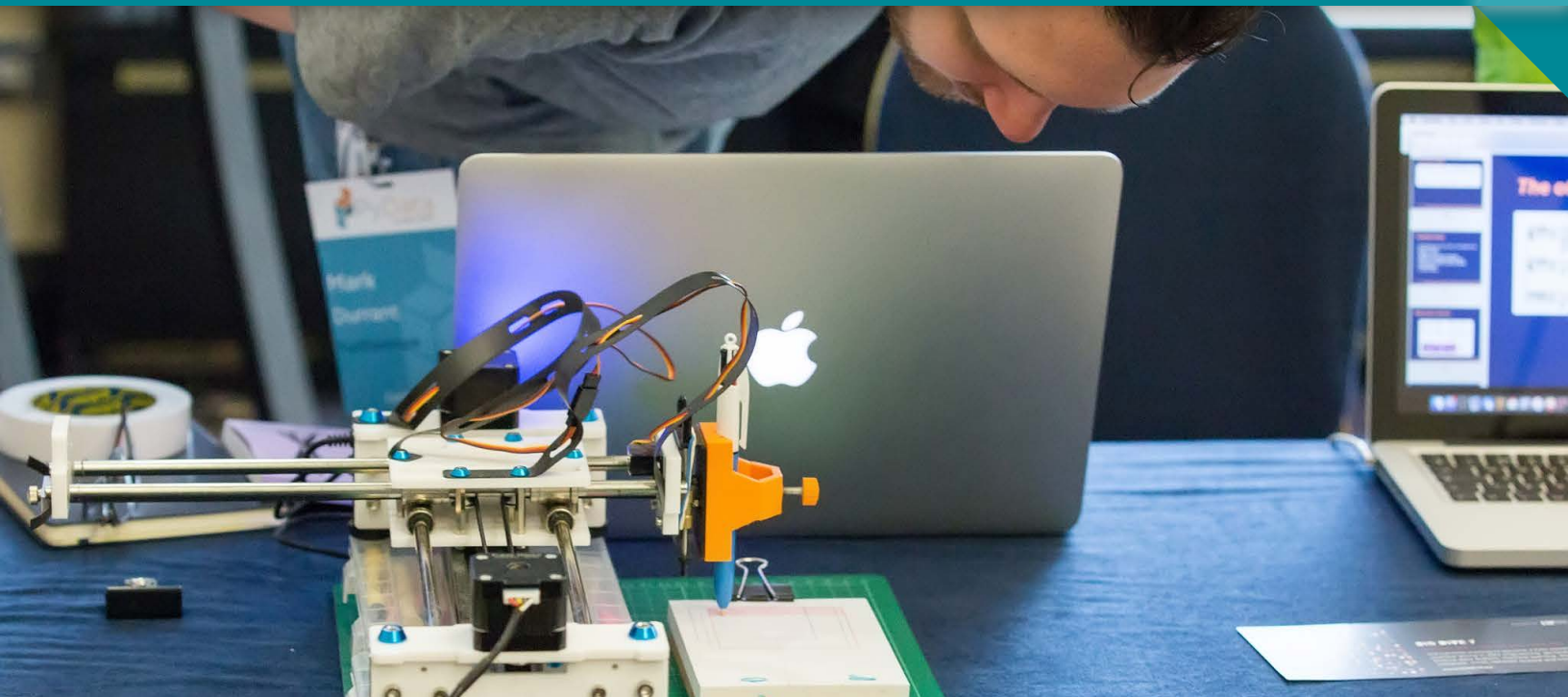> Brought on two new core developers, Deepak Cherian and Spencer Clark

yt

> Released yt 3.5
> Completed two successful Google Summer of Code student projects

GRANTS
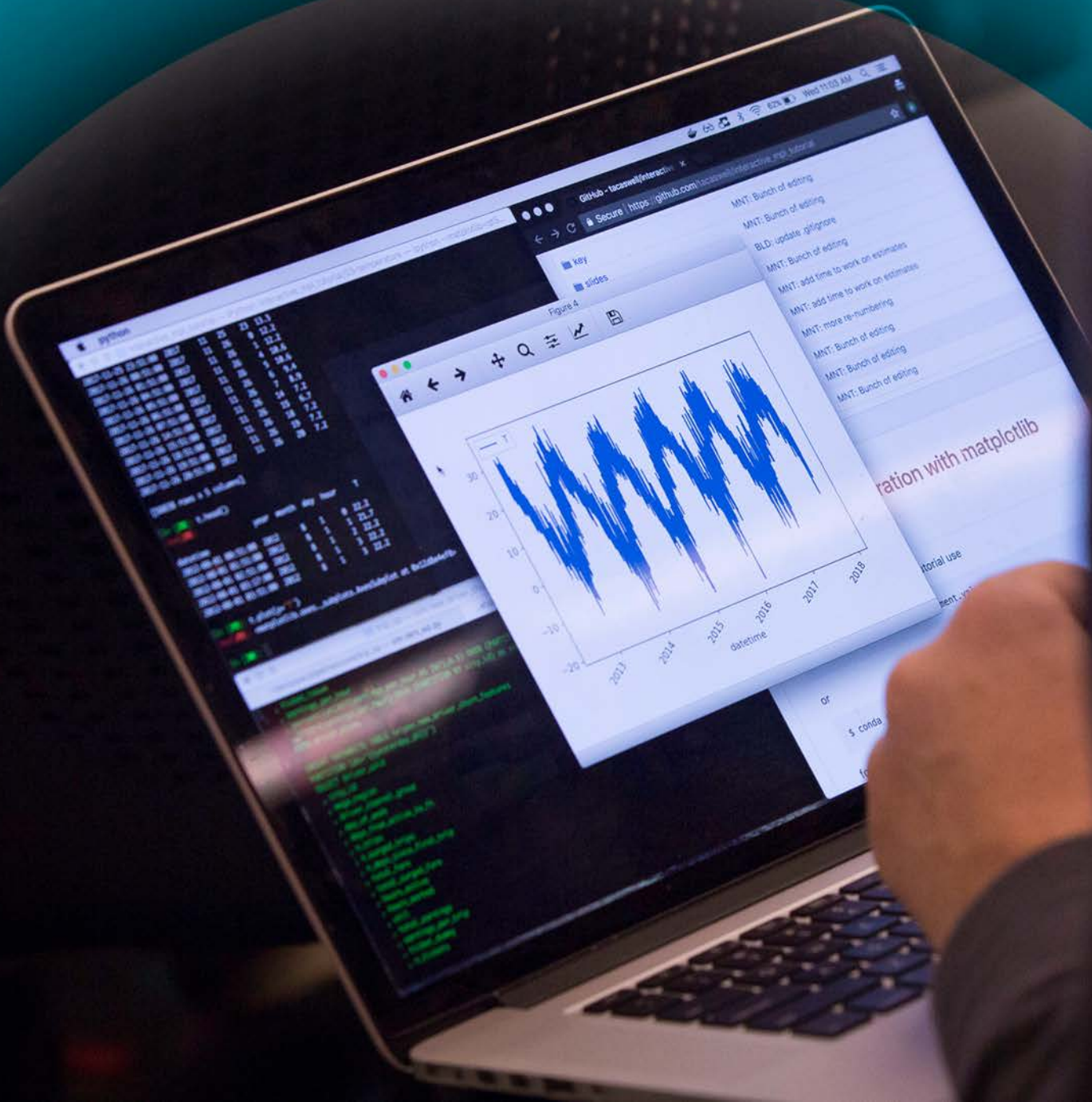
SMALL DEVELOPMENT GRANTS

In 2018, NumFOCUS distributed $60,000 in small developments grants to help our projects improve usability, grow their communities, and speed up the time to major releases. The proposals funded were:

PROJECT	PROPOSAL
Bokeh	• Bokeh Docs Modernization
Cantera	• Modernize, Reorganize, and Update Cantera's Documentation • The 3rd Annual Kinetics Code Conference: Charting near- and long-term directions for Cantera software development
Conda-forge	• conda-forge sprint at SciPy 2019
Gensim	• Modern user-friendly documentation • FastText tutorials

Julia	• BlockBandedMatrices.jl: add support for general array backends (GPU) • Multi-Dimensional Bisection Method for finding the roots of non-linear implicit equation systems	
MDAnalysis	• MDAnalysis tutorial and hackathon	
Open Journals	• Open Journals website update	
Orange Data Mining	• Girls go Data Mining	
Pomegranate	• Improving Documentation, Examples, and Tutorials • Adding compatibility with user-defined Python models	
PyTables	• Better support for native HDF5 files	
SciPy	• Maturing a sparse array implementation for SciPy • An Efficient, High-Level Implementation of Linear Programming	
Shogun	• Fully integrate new parameter framework, unify API/interfaces, and release Shogun 7.0 • Shogun website and UX redesign	
Spyder	• Spyder 4: Making the Scientific Python Development Environment even better	
Statsmodels	• Probability Plots and Generalized Additive Models (finish stalled pull requests)	
SunPy	• Improving the Usability of SunPy's Data Downloader	
SymPy	• MatchPy C++ code generator for SymPy/symengine	

JOHN HUNTER MATPLOTLIB SUMMER FELLOWSHIP

Kimberly Orr and Nabarun Pal were selected as the 2018 John Hunter Matplotlib Summer Fellows.

ABOUT THE FELLOWSHIP

The John Hunter Matplotlib Summer Fellowship, named in memory of Matplotlib creator John Hunter, sponsors one to two students to work full-time for 3 months on Matplotlib during the summer (in the northern hemisphere), supervised and mentored by a senior contributor from the project. The fellowship is designed to help prepare recipients to become active contributors and core maintainers of Matplotlib.

KIMBERLY ORR

Kimberly grew up in Cedar Park, Texas, where she initially discovered her love of programming in her high school computer science classes. She is a computer science and statistics double major at Valparaiso University. Kimberly is actively involved in Valpo's student chapter of the Association for Computing Machinery (ACM), where she enjoys promoting women's involvement in the field and sharing her passions for software engineering and data science. When not doing schoolwork or working as a Teaching Assistant for introductory CS classes, Kimberly enjoys playing oboe in the university's symphony orchestra and playing handbells in the university's handbell choir. After college, she hopes to pursue a career in software engineering and/or data science.

NABARUN PAL

Nabarun Pal is a final year undergraduate student at Indian Institute of Technology Roorkee. He hails from Agartala, the capital of Tripura in Northeast India. He is enthusiastic about open source software development. In his free time, he loves to code, read books or tinker with embedded hardware. Nabarun can also discuss about Internet of Things, Electronics, Robotics with equal spirit. His journey with the field of software and robotics started in his schooling days. He has been representing the college in competitions such as Inter IIT Tech Meet, ABU Robocon, Mercari Hackathon etc. and was involved in projects related to the above domains. He actively participates in conducting open lectures for students in the domains of Introductory Robotics and Control through a curated community of around 2000 members. Besides all the technical stuff, he is an avid foodie, a music buff and loves to read books on his Kindle.
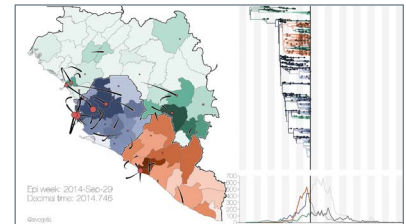
JOHN HUNTER EXCELLENCE IN PLOTTING CONTEST

**NumFOCUS was pleased to contribute prize money for the John Hunter Excellence in Plotting Contest 2018.**

1st The winning entry was "History of Ebola virus epidemic in West Africa 2013-2015," by Gytis Dudas, Luiz Max Carvalho, Trevor Bedford, Andrew J Tatem, Marc A Suchard, Philippe Lemey, & Andrew Rambaut.
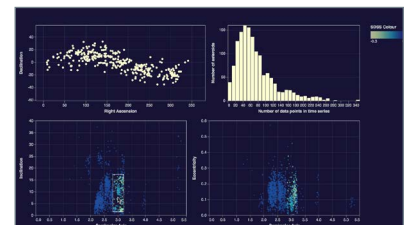

2nd Second prize went to Enrico Garaldi's "Bubbles in the young Universe."
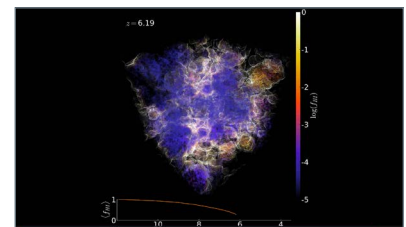
3rd Third to Daniela Huppenkothen's "Making Sense of Asteroids observed with the Gaia Space Telescope."

COMMUNITY

NumFOCUS COMMUNITY PROFILES

HOW ARE YOU INVOLVED WITH NumFOCUS?

I've been representing NumFOCUS in some conferences for the last couple of years. Providing information (and of course stickers) to people interested. Before that, I created and organized PyData Mallorca, before I relocated to London.

I'm also a pandas maintainer, so NumFOCUS helps us with whatever we need. In my particular case I regularly organize development sprints and related events. And I often ask NumFOCUS for advice and other things that can be needed.

WHAT HAS YOUR EXPERIENCE WITH THE NumFOCUS COMMUNITY BEEN LIKE?

I think the first time I was in contact with the NumFOCUS community was at PyData London. I was new to London and I attended many IT events (there is probably one every day of the week). I think PyData was the only one that weren't just talks, but had a sense of community. I still attend it when I can, as I have many friends there including most of the organisers.

Besides that, and the great PyData community at Mallorca, I've been to many PyData conferences, including London, Madrid, Barcelona, Amsterdam and Warsaw. I spoke at some of them, too. I have to admit that with the years it's becoming more difficult to find talks about new things where I can learn like when I started. But I still go to as many conferences as I can, because I realized that meeting old friends and new people is even better than attending the talks.

MARC GARCIA
Pandas Core Developer

"I think the NumFOCUS mission of making sure there are high quality and open tools for science and data is important for everyone."

WHY IS NumFOCUS IMPORTANT TO YOU?

I see NumFOCUS as "just" the community around its mission. And I think the NumFOCUS mission of making sure there are high quality and open tools for science and data is important for everyone.

For big companies this means better tools—and free of use—which has a big impact in their efficiency and productivity. For scientists, students, startups, it means having access to the same tools as the richest corporations. Even for humanity there is a clear impact, if scientists are able to be more efficient and do better and faster research.

I think we really want to make sure that the open source scientific and data tools keep being the best. And as a maintainer of one of the tools, I can tell you that's far from granted at the moment; that's why NumFOCUS is so important.

WHAT WOULD YOU SAY TO SOMEONE WHO WAS WONDERING HOW TO GET INVOLVED WITH NumFOCUS?

It's totally worth it. You can attend your local PyData events, or start a chapter if there isn't one available in your area. You can get involved in one of the open source projects. Or whatever you do, you'll learn a lot, and meet lots of wonderful people.

NumFOCUS COMMUNITY PROFILES

HOW ARE YOU INVOLVED WITH NumFOCUS?

For the last 4 years I had the great opportunity to be a member of the Berlin PyData community, first as an attendee and now as a co-chair of the organizing committee. Already in the first meeting I found the opportunity to contribute to the community by answering the call for volunteers for the upcoming conference. Volunteering in the PyData Berlin 2015 conference was a great experience following which I became more involved in the PyData community and co-organized the 2016 conference.

WHAT HAS YOUR EXPERIENCE WITH THE NumFOCUS COMMUNITY BEEN LIKE?

Being part of the NumFOCUS community has been a wonderful adventure for me. I got to meet so many interesting people with whom I can share ideas and build a community around Python and Data Science.

WHY IS NumFOCUS IMPORTANT TO YOU?

PyData Berlin and NumFOCUS gave me the opportunity to be involved in community building through which I learned a lot about myself and acquired new skills. Through organizing and running conferences with over 500 attendees I got to bring forward my more operational side. I had to hunt for venues, set up CFPs, handle service providers—which are all skills I never knew I had! Most importantly, being involved with NumFOCUS, with every PyData event I get the fulfilment of seeing my plans and ideas actualized on a monthly basis.

NOA SHINITSKI
PyData Berlin Organizer

"PyData Berlin and NumFOCUS gave me the opportunity to be involved in community building through which I learned a lot about myself and acquired new skills."

WHAT WOULD YOU SAY TO SOMEONE WHO WAS WONDERING HOW TO GET INVOLVED WITH NumFOCUS?

Being an organizer of a PyData chapter, I may be a bit biased when saying "join the closest PyData meetup you can find!" There are of course many more ways to be involved: talk at conferences, contribute to open source libraries, answer questions on Stack Overflow and more! If you feel more on the beginner side: attend a meetup or a conference and start a conversation with someone you never met before; improve the documentation of your favourite library; and don't hesitate to reach out to members of the community and express your will to be involved.

NumFOCUS COMMUNITY PROFILES

HOW ARE YOU INVOLVED WITH NumFOCUS?

As a JuliaLang committer and finance chair of JuliaCon, my involvement with NumFOCUS is really to ensure that we support our community in financing our events efficiently, and effectively.

WHAT HAS YOUR EXPERIENCE WITH THE NumFOCUS COMMUNITY BEEN LIKE?

It's been a pleasure working with NumFOCUS, with the staff and the wider community.

WHY IS NumFOCUS IMPORTANT TO YOU?

NumFOCUS provides us with a structure and organizational support that ensures that our community can come together every year with events that are both well organized and extremely productive. It allows us to focus on our communities, leaving some of the necessary administration to NumFOCUS.

WHAT WOULD YOU SAY TO SOMEONE WHO WAS WONDERING HOW TO GET INVOLVED WITH NumFOCUS?

When it comes to numerical computing, the community gathered around NumFOCUS includes some of the most exciting innovations happening in this field. So if you are interested in this subject, there is no better place to be.

AVIK SENGUPTA
Julia Contributor and JuliaCon Organizer

"When it comes to numerical computing, the community gathered around NumFOCUS includes some of the most exciting innovations happening in this field."

NumFOCUS COMMUNITY PROFILES

HOW ARE YOU INVOLVED WITH NumFOCUS?

I attended my first PyData on a scholarship from NumFOCUS in 2014, and that's when I first learned of the organization. I read their newsletters for regular updates. In December 2016, I noticed a call for volunteers for their upcoming diversity committee, and I was excited to participate. I joined the NumFOCUS Diversity in Scientific Computing Committee (DISC) in March 2017. I've worked on a number of different team projects with DISC including the DISCOVER Cookbook (which is a guide to make events more inclusive), drafting the new comprehensive Code of Conduct and organizing the inaugural DISC Unconference which was held in November 2017. I am an organizer for two meetup groups in New York City: Women in Machine Learning & Data Science and PyLadies, and so I connect these communities to NumFOCUS as well.

WHAT HAS YOUR EXPERIENCE WITH THE NumFOCUS COMMUNITY BEEN LIKE?

My experience with NumFOCUS has been wonderfully informative, educational, supportive, welcoming, professional and network expanding. The open source spirit of software development trickles to form an open and welcoming community culture. One of best parts of the NumFOCUS community is how accessible, amiable and inviting it is to people at all levels of scientific work.

WHY IS NumFOCUS IMPORTANT TO YOU?

The most important reasons to me personally are open source, education and diversity & inclusion. There is the the fiscal sponsorship it provides to critical Python programming libraries and tools which I use. And, its diversity and inclusion initiatives are critical and ground-breaking. All of these grass-roots movements need an incubator to thrive, which NumFOCUS provides. Whatever I learn from NumFOCUS, I bring to my other communities, and it has a positive ripple effect.

RESHAMA SHAIKH
Organizer, NYC Women in Machine Learning & Data Science and NYC PyLadies

"One of best parts of the NumFOCUS community is how accessible, amiable and inviting it is to people at all levels of scientific work."

WHAT WOULD YOU SAY TO SOMEONE WHO WAS WONDERING HOW TO GET INVOLVED WITH NumFOCUS?

There are so many ways. Join the DISC Committee. Participate in a PyData conference by organizing, volunteering, attending, speaking, tweeting at the event. And all the same for PyData local meetups. Donate to NumFOCUS. Learn about the projects and work on the issues that are open. For newcomers to the project, participate in an open source sprint. For developers, contribute to open source. Star the GitHub repo for an open source project you use. Follow the project leads on Twitter. Spread the word, tell others about NumFOCUS and ask them to sign up for their newsletter for more information.

THE 2018 NumFOCUS AWARDS

The first annual NumFOCUS Awards recognized individuals who have made substantial contributions to our projects, to our ecosystem, and to the open source scientific computing movement. The awards are meant to highlight exceptional technical, community, and organizational efforts supporting our projects.

NumFOCUS congratulates and thanks all honorees for their outstanding contributions to our community.

COMMUNITY LEADERSHIP AWARD
Ian Ozsvald

Ian was instrumental in launching and growing the PyData London conference and meetup chapter — the largest and most active in the world (over 8k members!). His tireless efforts and good humour (spelled the British way) helped build an incredibly robust community of PyData enthusiasts in the UK. Many regard Ian as a mentor and role model, and he continues to successfully champion new growth for PyData in Europe and elsewhere.

PROJECT SUSTAINABILITY AWARD
Kelle Cruz

Kelle has been an incredibly effective advocate for Astropy within the astronomy community. Her work to help organize and formalize the processes and governance of the project have created a strong structure for future growth of the project. Kelle has also been a leader in creating and maintaining a culture of inclusion within Astropy, which has resulted in successful recruitment and retention of a diverse contributor base. Her efforts have meaningfully advanced Astropy along the path to long-term sustainability for the project.

CORPORATE STEWARDSHIP AWARD
Shahrokh Mortazavi

Shahrokh has been a champion of NumFOCUS projects and programs within Microsoft for a number of years. Thanks to his advocacy, PyData Seattle was held at Microsoft for the first time in 2015 and then again in 2017; Microsoft has also played host to PyData NYC as well as, most recently, the NumFOCUS Summit. Shahrokh is an evangelist and active participant for corporate support of open source scientific computing.

AWARDS WON BY NumFOCUS PROJECTS

ODSC Open Source Data Science Project award — PyMC3

The Open Source Data Science Project award is given in recognition of the significant contributions made by an open source project to the field of data science. At ODSC London 2018, Dr. Thomas Wiecki accepted the award on behalf of the PyMC3 project team.

2017 ACM Software System Award — Project Jupyter

The ACM Software System Award is presented to an institution or individual(s) recognized for developing a software system that has had a lasting influence, reflected in contributions to concepts, in commercial acceptance, or both.

In May of 2018, fifteen members of the Project Jupyter steering council had the honor of accepting this prestigious award on behalf of Project Jupyter. The recipients were: Fernando Pérez, Brian E. Granger, Min Ragan-Kelley, Paul Ivanov, Thomas Kluyver, Jason Grout, Matthias Bussonnier, Damián Avila, Steven Silvester, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Carol Willing, Sylvain Corlay and Peter Parente.

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2018 — Paul M. Romer

Paul Romer, former chief economist of the World Bank and noted champion of open source scientific computing tools, was jointly awarded the Nobel Prize in economics (along with William Nordhaus) for integrating technological innovations into long-run macroeconomic analysis.

Romer is a vocal champion of Jupyter notebooks for transparent analysis and has specifically expressed appreciation for the powerful capabilities and open nature of the NumFOCUS ecosystem of software tools:

"Python libraries let me replicate everything I wanted to do with Mathematica: **Matplotlib** for graphics, **SymPy** for symbolic math, **NumPy** and **SciPy** for numerical calculations, **Pandas** for data, and NLTK for natural language processing. **Jupyter** makes it easy to use Latex to display typeset math. With Matplotlib, Latex works even in the label text for graphs. [...] I'm more productive. I'm having fun."

EVENTS

PYDATA

In 2018, the PyData global network of meetups approximately doubled in size compared to 2017.

over 100,000 members

across 49 countries

in 27 chapters

PyData global network of meetups
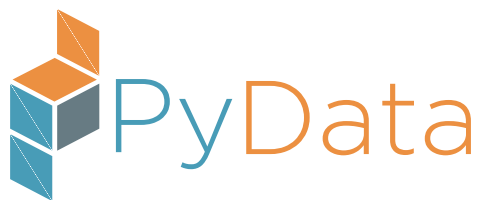
> "My favorite annual conference! Learned a ton, great networking, and very applicable learnings."
>
> —*James Beveridge, Data Scientist*

We also organized PyData conferences and PyData event tracks* across 5 continents, attracting approximately 6,000 participants.

2018 PyData Event Locations

- Florence, Italy*
- Kaunas, Lithuania*
- Edinburgh, UK*
- Budapest, Hungary*
- Tokyo, Japan*

- Karlsruhe, Germany*
- Johannesburg, South Africa*
- London, UK
- Amsterdam, Netherlands
- Berlin, Germany

- Delhi, India
- Córdoba, Argentina
- New York, NY, USA
- Washington, DC, USA
- Los Angeles, CA, USA

SPONSORED PROJECT EVENTS

NumFOCUS projects organized events to further educate the community of users and to expand the community of contributors to the projects. These were:

> StanCon

> FEniCS Con

> Python in Astronomy

> rOpenSci Unconference

> Cascadia R Conference

> JuliaCon

> JupyterCon

> PyMC4 Developers Summit

PROGRAMS

DIVERSITY & INCLUSION

NumFOCUS believes that diverse contributors and community members produce better science and better projects.

The DISC (Diversity & Inclusion in Scientific Computing) Program strives to help create a more diverse community through initiatives and programming devoted to increasing participation by and inclusion of underrepresented people.

2018 DISC Committee Members:

> Samuel Brice

> Gina Helfrich

> Julie Hollek

> Jennifer Klay

> Julia Meinwald

> Leonie Mueck

> Madicken Munk

> Reshama Shaikh

2018 DISC Initiatives

In 2017, the Moore Foundation awarded the DISC Program a generous 2-year grant. The goals of the grant are to produce a scalable kit that can be disseminated to events to help promote best diversity and inclusion practices, assess current diversity programs at NumFOCUS, identify best diversity practices and challenges, and create new initiatives including programs to increase the diversity of project contributors.

In support of these goals, in 2018 the DISC Committee:

• Updated the DISCOVER Cookbook (Diverse & Inclusive Spaces and Conferences: Overall Vision and Essential Resources) to provide a more user-friendly website interface

• Organized an Open Data Science Inclusion Sprint, NYC, October 2018

• Revised the NumFOCUS Code of Conduct to provide greater clarity and specificity
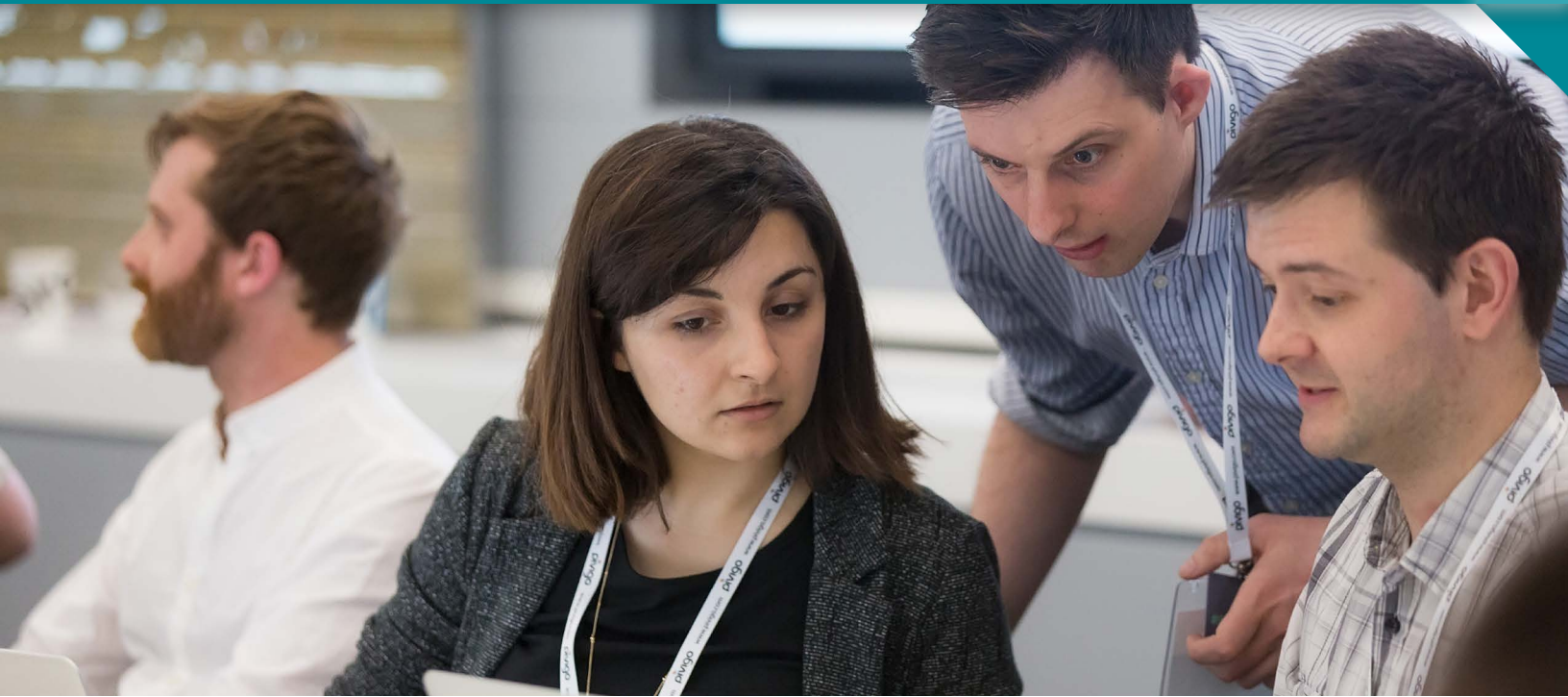
SUSTAINABILITY

NumFOCUS key stakeholders came together September 22–25, 2018 for the annual NumFOCUS Summit, 4 days of intensive work around questions of sustainability.

Attendees included representatives from our sponsored projects, members of the Board, Advisory Council, Staff, representatives from our Corporate Sponsors, as well as other community and corporate leaders.

> Community Roadmaps
> Governance

> Fundraising and Grant Writing
> Culture and Conduct

The diversity of our projects was reflected in the valuable discussions and networking taking place at the Summit. We asked what the most valuable takeaway was from the Sustainability Workshop. Answers included:

"The ability to talk directly with maintainers of other projects who had already gone through several of the stages that I see upcoming for my project, and get their perspective on how to handle them."

"Spur to write down/ document more information."

"Dedicated time to thinking about bigger issues."

"Meeting all the leaders and different projects, hear their thoughts on governance models and funding challenges."

"Presentations and discussions about what road maps are, why we should write them, and what they should look like."

2018 GOOGLE SUMMER OF CODE

NumFOCUS has participated as an umbrella organization for Google Summer of Code (GSoC) since 2015.

Google Summer of Code sponsors students to work full-time for 3 months on open source projects during the summer (in the northern hemisphere), supervised by a senior contributor from the mentoring organization.

In 2018, **47** students participated in GSoC under the NumFOCUS umbrella, working on the following projects:

- Cantera
- Conda Forge
- Data Retriever
- FEniCS
- Gensim

- Julia
- MDAnalysis
- PyMC3
- Shogun
- yt

Astropy, SunPy, and SymPy also participated in GSoC, either directly or as part of another umbrella organization.

NumFOCUS extends a special thanks to our GSoC coordinators, Max Linke and Mridul Seth.
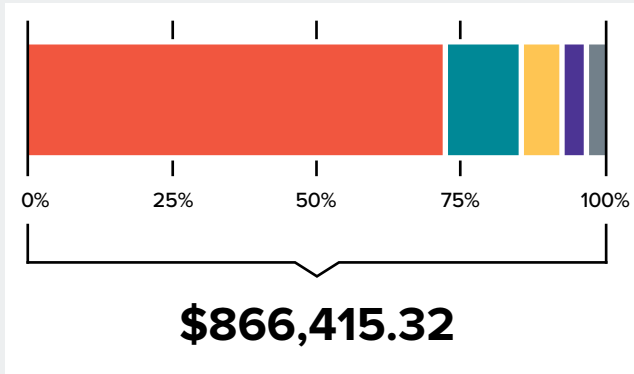
2018 FINANCIALS

INCOME & EXPENSES
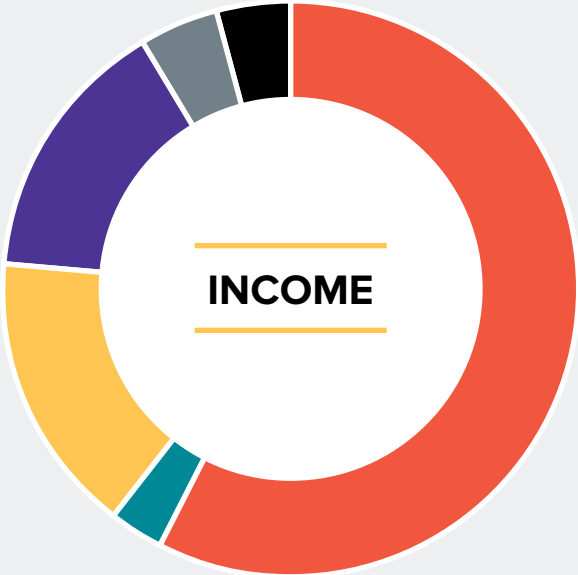
PyData Events
57.5% | $1,021,703.53

Restricted for Programs
3.3% | $58,000.00

Restricted for Projects
15.8% | $281,208.00

Corporate Sponsors
14.8% | $262,186.85

Admin Fees
4.7% | $83,920.50

Individual Donations
3.9% | $68,604.70

Projects
28.8% | $602,927.06

Programs
41.4% | $866,415.32
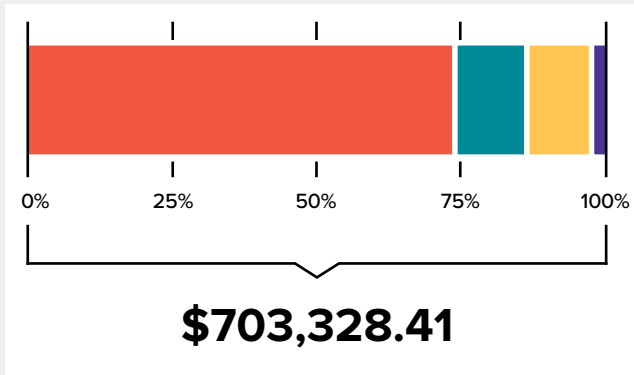
Operations & Fiscal Services
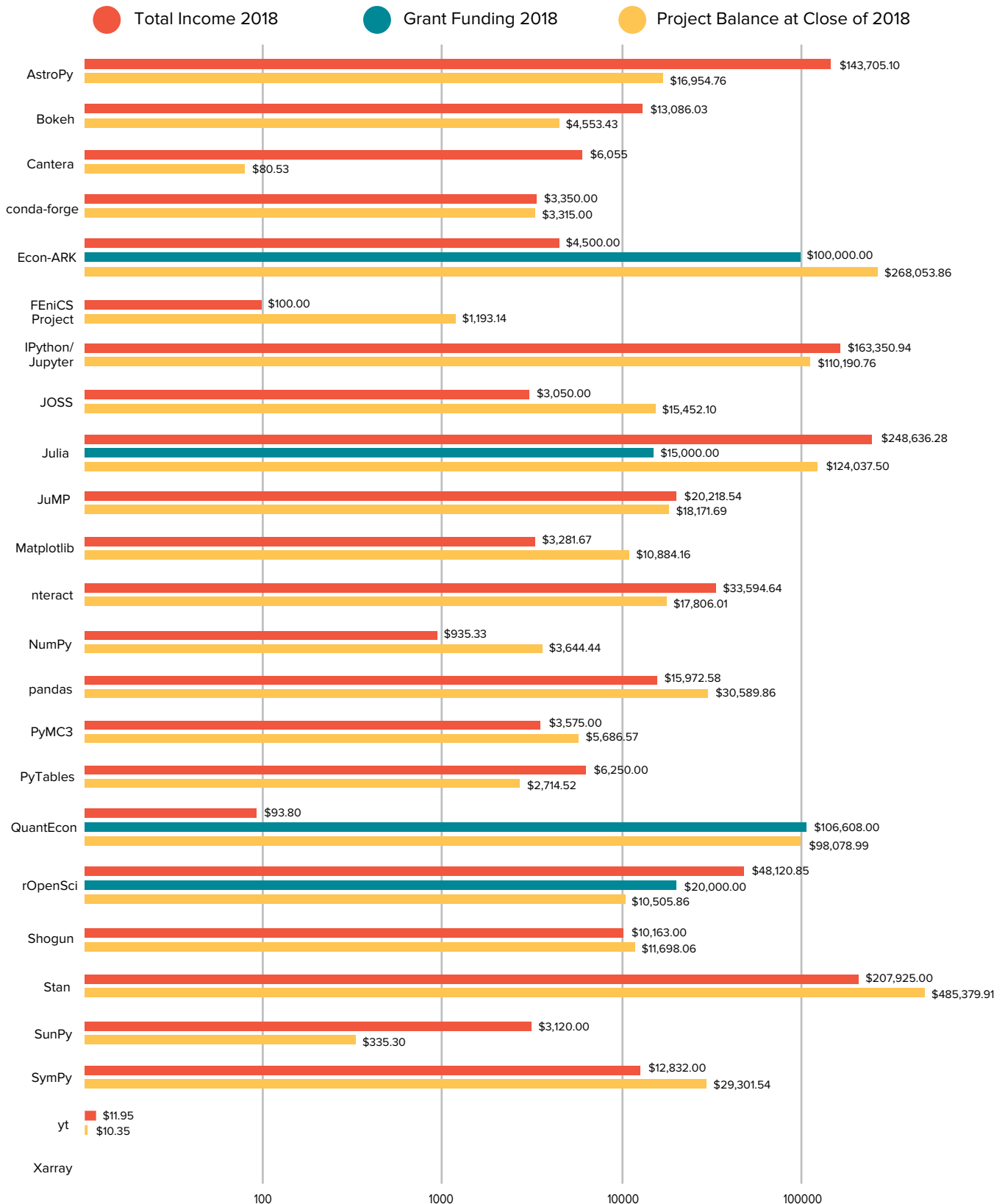29.8% | $703,328.41

EXPENSE DETAILS

PROGRAMS

$866,415.32

- ● PyData │ 72.2% │ $625,425.01
- ● Sustainability │ 13.3% │ $115,645.87
- ● Small Dev Grants │ 7.2% │ $62,300.00
- ● Fellowships & Awards │ 4.3% │ $36,844.44
- ● Google Summer of Code │ 3% │ $26,200.00

OPERATIONS & FISCAL SERVICES

$703,328.41

- ● Staff Salaries │ 73.5% │ $517,260.01
- ● Legal & Professional │ 12.4% │ $87,018.91
- ● General Office Expense │ 11.2% │ $78,940.38
- ● Travel Expenses │ 2.9% │ $20,109.11

PROJECTS DETAIL

Total Income 2018   **Grant Funding 2018**   **Project Balance at Close of 2018**

Project	Total Income 2018	Grant Funding 2018	Project Balance at Close of 2018
AstroPy	$143,705.10		$16,954.76
Bokeh	$13,086.03		$4,553.43
Cantera	$6,055		$80.53
conda-forge	$3,350.00		$3,315.00
Econ-ARK	$4,500.00	$100,000.00	$268,053.86
FEniCS Project	$100.00		$1,193.14
IPython/Jupyter	$163,350.94		$110,190.76
JOSS	$3,050.00		$15,452.10
Julia	$248,636.28	$15,000.00	$124,037.50
JuMP	$20,218.54		$18,171.69
Matplotlib	$3,281.67		$10,884.16
nteract	$33,594.64		$17,806.01
NumPy	$935.33		$3,644.44
pandas	$15,972.58		$30,589.86
PyMC3	$3,575.00		$5,686.57
PyTables	$6,250.00		$2,714.52
QuantEcon	$93.80	$106,608.00	$98,078.99
rOpenSci	$48,120.85	$20,000.00	$10,505.86
Shogun	$10,163.00		$11,698.06
Stan	$207,925.00		$485,379.91
SunPy	$3,120.00		$335.30
SymPy	$12,832.00		$29,301.54
yt	$11.95		$10.35
Xarray			

2018 NumFOCUS CORPORATE SPONSORS

Our Corporate Sponsors invest in the development and sustainability of the open source scientific data stack through their financial support of NumFOCUS.

Many NumFOCUS Corporate Sponsors also allocate employee paid time to development work on NumFOCUS open source projects.

PLATINUM	Bloomberg · Microsoft · IBM
GOLD	Capital One · 2σ TWO SIGMA
SILVER	IMC · CIVIS ANALYTICS · ANACONDA
BRONZE	GEOTAB management by measurement · Google Open Source · Optiver · ActiveState THE OPEN SOURCE LANGUAGES COMPANY · zappi
EMERGING LEADER	QuanSight · Ravelin · plotly · Q Quantopian
WITH SUPPORT FROM	Alfred P. Sloan FOUNDATION · GORDON AND BETTY MOORE FOUNDATION

DONOR LIST

2018 Supporting Donors

Safia Abdalla
Thomas Abernathy
Joshua Adelman
Fatih Akici
Guillaume Alain
Waleed Alhaider
Arnob Almazee
Dominic Antonacci
Anthony Brian Araujo Garcia
Rajen Athreya
Emmanuel Augustin
Hazen Babcock
Christopher Barker
Kevin Bates
James Bednar
Mohammad Ben Salamah
Dom Bennett
Kenneth Benoit
Azamat Berdyshev
Mathieu Besancon
Simon Bettscheider
Chris Binz
Linda Blair
Nitin Borwankar
Jeff Bradberry
The Bresler Foundation, Inc.
     (Alex & Liz Bresler)
phillip brix
Anthony Brown
Peter Bull
Christopher Burns
Stefan Burtscher
Albert Byun
Brendon Caligari
John Callery
Clay Campaigne
Juan Luis Cano
Nandan Padar Chandrashekar
Graham Chester
Jerry Chi
Alin Chis
Youngchan Cho
Daniel Chudnov
Jennifer Clarke
Jafet Cnepa Maceda
Mark Coumans
Adam Cushner
Muhammad Damanhuri
Eric Daniels
Joseph Dasenbrock
Michael Defferrard
Brian Dennis

John Dodson
Jorge Douglas
Allen Downey
Garrett Drapala
Evelyn Drummond
Eric Durham
Nick Dyar
Lukman Edwindra
Phil Elson
Stephen Elston
Ahmet Erdemir
EK Esawi
Derek Evans
Elea Feit
Henry Ferguson
Eric Firing
Donald Fischer
Gordon Fleetwood
Jonathan Frederic - Gift
     Matched by Google
Krissy Freeman
Dave Fulker
Dillon Gardner
Bjarni Gautason
OtoJig GmbH
Nathan Goldbaum
Guillermo Gonzalez Sanchez
Dorian Gray
Bjarne Grimstad
Matthew Gwynne
Raheleh Hadian
John Hagarty
Robert Hamilton
Jessica Hamrick
Harry Hao Xu - Gift Matched
     by Morgan Stanley
Mark Harfouche
Patrick Harrison
Matthew Hawthorn
Brian Hayes
Randy Heiland
Magnus Lie Hetland
Adam Hill
Tim Hoffmann
Robert Hogg
Yaroslav Hovorunov
Tsuntsun Huang
 Hudson River Trading
Thomas Hughes
Thomas Hunger
Taihyun Hwang
Akiyuki Ishikawa

Payoj Jain
Graham Jeffries
Gustavo Jeuken
Malte Joppien
Russell Jurney
Anonymous Katz
Yuji Kawamatsu
Christopher Keller
Kyle Kelley - Gift Matched
     by Netflix
Daniel Kim
Sami Kiviharju
Thomas Kluyver
Ravin Kumar
J.F.J. Laros
The Lawrence Livermore
     National Laboratory
     Helping Others More
     Effectively (HOME)
Sara Lawrence-Sucato
Charles Le Losq
Young Lee
Bob Leiweke
Francesco Lo Conti
Ivan Lobov
Miles Lubin - Gift Matched by
     Google
The Lucey Family Charitable
     Fund
Chatdanai Lumdee
Andy Lustig
Jane Lynn
Jean Maccario
Jay Mangi
Atma Mani
Hemanth Manjunatha
Lessell Manning Jr
Frits Mastik
Joao Matos
Lyndon Maydwell
Timothy Mayes
Stephen McAllister
Jonathan McKenzie
Byron McLendon
Jason Mcmillian
Daniel McNichol
Roy Mendelssohn
Liang Mi
The Michael R. Douglas
     Charitable Fund
Srinivasa Mohan
Guillermo Monge del Olmo

Nick Moore
Ian Moran
Ahamed Mustafa
Paul Nakroshis
Gautham Narayan
Thomas Nesbit
Thomas Niederberger
William O'Shea
Tatsuro Oda
Kristof Ostir
Axel Pahl
David Palmer
Serge-Etienne Parent
Peter Parente
Drew Parsons
Gustavo Patino
Marc Pavese
Christian Peel
Vasily Povalyaev
Roger Powell
Trey Pruitt
Thejaswi Puthraya
Karthik Ram
Florian Rathgeber - Gift
     Matched by Google
Rosen Raychov
Daniel Reeves
Lorenzo Riches
Lindsay Richman
Rob Roy
Alejandro Saez Mollejo
Ritvik Sahajpal
Pranav Sarvaiya
Rajnikant Sarvaiya
Angelika Schneider - Gift
     Matched by Apple
Michael Schroder
Ridgway Scott
Michael Seidel
Mit Shah
Dinesh Shenoy
Petr Simecek - Gift Matched
     by Google
Dimitrios Simopoulos
Pratham Singh
Benjamin Singleton
Indranil Sinharoy
TJ Slezak
Christian Staudt
Monica Stewart - Gift Matched
     by Microsoft
David Strozzi

Michael Stumpf
Timothy Swast - Gift Matched
    by Google
Renee Teate
Jared Thompson
Jeffrey Tratner
Heather Tripp
Shingo Tsuji
Wes Turner

Luca Urpi
Maurits van der Meer
Jed Verity
Divakar Viswanath
Aaron Wald
John Mark Walker
Haoran Wang
Nehemiah Washington
Justin Watson

Robert Weiss
Daniel Weitzenfeld
Debbie Williams
Diana Williams
Denise Williams
Jared Wilson
Lucas Wiman
Matthew Wittmann
Francis Wolinski

Jianhong Xin
David York
Enpeng Yuan
Benjamin Zaitlen
Tauhid Zaman
Michael Zingale

2018 Sustaining Donors

Charles Ahern
Guillermo Ambrosio
RK Aranas
Mohammad Athar
Pauline Barmby
Benjamin Batorsky
Bob Baxley
Michael Becker
Anne Beeman
Monica Bobra
Tim Bock
Vicki Boykis
Erik Bray
Jed Brown
Brandon Burroughs
Gonzalo Bustos
Ken Cantwell
Colin Carroll
Quentin Caudron
Daniel Chen
Pramit Choudhary
Steven Christe
Johnny Cochrane
Scott Collis
Paul Constantine
Joshua Cook
Matthew Craig
Steven Crawford
Zaisheng Dai
Abhijit Dasgupta
Salomon Davila
Adam Deconinck
Eric Dill
Robin Doumerc
Vladislavs Dovgalecs
Paul Egbert
Niclas Ekman Roos
Scott Ellett
Timothy Elser
Joao Fonseca
Mark Fuller
Yankuic Galvan
Francisco Giordano Silva
Greg Goddard

Kevin Goebbert
Joe Greener
Jeffrey Griffin
Matt Hall
Alexander Hasha
Edward Haymore
Nima Hejazi
Jody Hey
Steve Holden
Jefferson Huang
Alexandre Huat
Rick Hubbard
Axel Huebl
David Humeau
Alex Ioannides
Scott Irwin
Zunbeltz Izaola
David Jaluvka
Matthew Jensen
Han Jiadong
Keith Johansen
Andrew Johnson
Britt Johnson
Peter Jones
Ari Kamlani
Daniel Kapitan
Aaron Karper
Gaurav Kaushik
Brian Keegan
Matthew Kelcey
Craig Kelly
Taehun Kim
ivan kissiov
Tim Knight
W. Koehrsen
Leonid Kovalev
Joachim Krois
Tom Ku
Marek Kwiatkowski
Daniel Lakeland
Flavien Lambert
Drew Lancaster
Peter Landwehr
Irene Langkilde-Geary

Jeremy Langley
Scott Lasley
Gregory Lee
Sam Levin
Mathew Lodge
Eric Londo
Remi Louf
Dan Lussier
Timothy Man
Yotam Manor
Brian Mapes
Matthew Martz
Friedhelm Matten
Markus Mayer
Matthew McCormick
Gregory Medlock
Andrey Melentyev
Nancy Melucci
Lorena Mesa
Fred Monroe
Mike Moran
Mike Morgan
Nicholas Murphy
Christopher Musselle
Benjamin Naecker
John Nelson
Kyle Niemeyer
Erfan Noury Qarajalar
Nathaniel Odell
Travis Oliphant
Bernardo Oliveira
Victor Pal
Hilary Parker
Matt Parker
Joshua Patterson
John Pellman
John Peloquin
Cameron Prybol
Monica Puerto
Bruno Quinart
Neil P. Quinn
Ramanathan R
David Rasch
Ara Razo

Greg Reda
Justin Reid
Marjorie Roswell
Kenta Sato
Daniel Saxton
Skipper Seabold
Mohar Sen
David Severski
Will Shanks
Leah Silen
Rachel Slaybaugh
Jonathan Smith
Patrick Sodre
Alexios Spanos
Watee Srinin
Jeff Sternberg
Aidis Stukas
Roland Tanglao
Ravelin Technology
Larry Tooley
James Tsai
Kevin Tyle
Vlad Tyshkevich
Nicholas Ursa
Shankar Vaidyaraman
Bryan Van de Ven
Jake VanderPlas
Angelo Varlotta
Thomas Vetterli
Victor Vicente Palacios
Alejandro Vidal
Liang-Bo Wang
Mauro Werder
Richard West
Matthew Widjaja
Chad Williamson
Benjamin Winkel
David Zihala

NumFOCUS GOVERNANCE

NumFOCUS WELCOMES
NEW BOARD OF DIRECTORS

In 2018, NumFOCUS installed new members of the Board of Directors for the first time in a number of years. Our newly-elected board members are:

SYLVAIN CORLAY — Sylvain has been leading the efforts for NumFOCUS EU incorporation for the past year. He has been involved in the open source scientific computing stack since 2012, maintained the Jupyter interactive widgets framework and co-authored several data visualization packages.

JANE HERRIMAN — After spending a year guiding the Julia Language, a NumFOCUS fiscally sponsored project, as the Director of Diversity and Outreach at Julia Computing, Jane is currently finishing off her PhD at Caltech. She is also a JuliaCon organizer and was active in the DISC working group in its first year.

JAMES POWELL — James has played a vital role in organizing the PyData community while serving as a liaison during his years as a NumFOCUS Vice President. He has attended 30+ PyData events throughout the world helping to recruit the best talent and ideas to better the community. James has also worked with core developers behind numerous NumFOCUS projects to advance their communities.

KATRINA RIEHL — Katrina is currently the Director of Global Data Science at HomeAway, where she manages a large data science program. As a passionate Pythonista for the last 20 years, Katrina has worked on multiple scientific software systems built on open source technologies and furthering Data Science as a discipline at well-known institutions such as Apple, Expedia, and DARPA.

STÉFAN VAN DER WALT — Stéfan has been involved in the open source scientific Python ecosystem for more than a decade, arguing for the adoption of open tooling in making computational science transparent. Most recently, he has been working with the NumPy community and managing two full-time developers at UC Berkeley to explore how funded development can coordinate with and complement community efforts.

They join existing directors **ANDY TERREL** and **LORENA BARBA**, who will continue to serve on the Board.

NumFOCUS expresses our sincere gratitude to the following Board of Directors members whose terms concluded in 2018:

- Ralf Gommers (Secretary)
- Jennifer Klay
- Didrik Pinte (Treasurer)
- Matthew Turk

ADVISORY COUNCIL & STAFF

Advisory Council

Matt Greenwood — Two Sigma

Shahrokh Mortazavi — Microsoft

Brian Granger — Cal Poly

Fernando Pérez — Berkeley

Stefan Karpinski — Julia Computing

Travis Oliphant — Quansight

Jason Grout — Bloomberg

Staff

Leah Silen, Executive Director

Andy Terrel, President

Gina Helfrich, Communications Director and Program Manager for Diversity & Inclusion

Lynn Brubaker, Projects Finance Manager

Jim Weiss, Events Manager

NUMF⬚CUS

OPEN CODE = BETTER SCIENCE

The mission of NumFOCUS is to promote sustainable high-level programming languages, open code development, and reproducible scientific research. We accomplish this mission through our educational programs and events as well as through fiscal sponsorship of open source scientific computing projects. We aim to increase collaboration and communication within the data science and scientific computing community.

NumFOCUS is a 501(c)3 public charity in the United States.

P.O. Box 90596 • Austin, TX 78709
info@numfocus.org
+1 (512) 831-2870

LEARN MORE