

NUMFEDCUS OPEN CODE = BETTER SCIENCE

2017 ANNUAL REPORT

Contents

Letter from the President	2
Achievements in Open Code Development	3
Sponsored Project Achievements	9
NumFOCUS Community Achievements	18
Donors List	28
Corporate Sponsors	32
Financials	33
The NumFOCUS Team	37



OPEN CODE = BETTER SCIENCE

By far my favorite statistic in the 2016 NumFOCUS Annual report was the 37,000 stickers we gave out that year.

Each sticker represented a connection with a community member wanting to proudly show our projects. In this report, that dispassionate statistic has been replaced with real stories of our members' lives. Each has been bettered by the work of the NumFOCUS community. As one reads this report, it is impossible not to see the real impact our small organization is having on the world of science.

In these pages, you will read accounts of people from new community members to a Nobel prize winning economist—each embracing our mission to promote sustainable high-level programming languages, open code development, and reproducible scientific research. Many of our projects are finding more funding, success in publications, and staff to devote to their projects. So many of our accomplishments are due to bringing a diverse community together. You will see that in every aspect of our work, from our new DISCOVER cookbook bringing very real recommendations on diversity at events to our corporate programs bringing in new funding and stakeholders to promote our projects in places we haven't previously been represented. As NumFOCUS grows and matures, we seek to become a better organization that can help open scientific codes and communities grow and be sustained. At our annual summit, it was apparent that as NumFOCUS has grown, we as an organization should have a stronger voice in advocacy of open science principles. We strive to open doors for our projects to raise funds and execute their visions—a reminder that NumFOCUS is not the work of a single project or person, but a collective tens of thousands strong. To this end, our sustainability program is looking at ways to bring our community together and tackle new initiatives. Stay tuned for announcements and programs that are aimed at addressing that conversation.

This report cannot capture all the action in our community and will undoubtedly leave some numbers or activities underrepresented. For example, my statistic of sticker count is no longer reported; instead, we report nearly 2 million dollars in income restricted for projects. We invite you to continue the discussion and support the efforts of our community to use NumFOCUS tools to make the world a better place.

> Andy Terrel, President, NumFOCUS Board of Directors

Achievements in Open Code Development

Julia Joins Petaflop Supercomputer Club

Julia has joined the rarefied ranks of computing languages that have achieved peak performance exceeding one petaflop per second the so-called 'Petaflop Club.'



The Julia application that achieved this milestone is called <u>Celeste</u>. It was developed by a team of astronomers, physicists, computer engineers and statisticians from UC Berkeley, Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center (NERSC), Intel, Julia Computing and the Julia Lab at MIT.

Celeste uses the Sloan Digital Sky Survey (SDSS), a dataset of astronomical images from the Apache Point Observatory in New Mexico that includes every visible object from over 35% of the sky—hundreds of millions of stars and galaxies. Light from the most distant of these galaxies has been traveling for billions of years and lets us see how the universe appeared in the distant past.

Since SDSS data collection began in 1998, the process of cataloging these stars and galaxies was painstaking and laborious. So the Celeste team developed a new parallel computing method to process the entire SDSS dataset. Celeste is written entirely in Julia, and the Celeste team loaded an aggregate of 178 terabytes of image data to produce the most accurate catalog of 188 million astronomical objects in just 14.6 minutes with state-of-the-art point and uncertainty estimates.

Celeste achieved peak performance of 1.54 petaflops using 1.3 million threads on 9,300 Knights Landing (KNL) nodes of the Cori supercomputer at NERSC — a performance improvement of 1,000x in single-threaded execution.

(Thanks to Julia Computing, Inc. for providing this detailed information.)

Data Scientists Make Frequent Use of NumFOCUS Projects at Work

According to Kaggle's 2017 report on <u>The State of Data Science & Machine Learning</u>, NumFOCUS projects are in the Top 40 tools used by data scientists in their work.



OPEN CODE = BETTER SCIENCE



Facebook Makes Sophisticated Forecasting Techniques Available to Non-Experts Thanks to Stan In February 2017, Facebook announced that they had open sourced their forecasting tool, Prophet, which is built on top of Stan, a NumFOCUS sponsored project. According to Facebook, the Prophet forecasting tool is designed to "make it easier for experts and non-experts to make high quality forecasts that keep up with demand."

Stan is a platform for Bayesian modeling, inference, and visualization on the frontiers of applied statistics. The project iTs developed by a team of more than two dozen statisticians, computer scientists, mathematicians, and other researchers, many of whom work in academia rather than industry.

While NumFOCUS fiscally sponsored projects are all primarily used in scientific computing, many of our tools have applications in the world of data science and business. The advances in forecasting made possible through Facebook's reliance on Stan serve as a wonderful example of the productive relationship between academia and industry so often enacted through NumFOCUS projects.

We applaud Facebook's decision to open source the Prophet tool, thereby making its sophisticated forecasting capabilities accessible to a broad audience, and look forward to seeing what new discoveries and innovations will be possible as a result.

NumFOCUS Awards Over \$13,000 in **Grants to Support Our Projects**

This spring the NumFOCUS Board of Directors awarded targeted small development grants to applicants from our sponsored and affiliated projects.

In the wake of a successful 2016 end-of-year fundraising drive, NumFOCUS wanted to direct the donated funds to our projects in a way that would have impact and visibility to donors and the wider community.



NumFOCUS Awarded Grants To The Following Projects:

- () SymPy 1.1 Release Support
- NumExpr-3.0 Beta
- h5py backend for PyTables
- Full Python 3 Support for MDAnalysis

- **Orange Text Analytics Introductory Course for Social Scientists**
- American Meteorological Society Short Course on Open Source Radar Software



Thomas Sargent

Professor of Economics at NYU

Senior Fellow at The Hoover Institution, Stanford University.

Steering Committee for QuantEcon.

How are you involved with NumFOCUS?

I started as an enthusiastic "consumer" benefiting from several NumFOCUS projects by watching Youtube videos and downloading and using Python and Julia code. I am a regular user of Jupyter notebooks (and now Jupyter Lab), SymPy, Matplotlib, IPython, and of course I depend on NumPy. (I even occasionally look at Astropy, because I sort of study some physics as a hobby.) I am also engaged in a couple of big data collection, visualization, and analysis projects that rely on pandas. I have used PyMC3 with success and am now starting to use Stan. It is from PyMC3 and Stan that I learned about things like Hamiltonian MCMC.

So my short answer to the question should have been, "for me, NumFOCUS is a place where I go to school and acquire great tools."

It was a dream come true when John Stachurski and I were lucky enough to convince NumFOCUS to allow our QuantEcon project to become a NumFOCUS sponsored project.

I have also been lucky enough to attend and even give talks at PyData and JuliaCon conferences. I regularly "attend" these almost every day by watching Youtube videos (some over and over again) when I am on the elliptical training machine at the gym.

What has your experience with the NumFOCUS community been like?

My interactions with the NumFOCUS professional staff have all been excellent in every way. In addition to administering our grant efficiently and fairly, NumFOCUS staff have kindly put me in touch with other NumFOCUS workers (e.g., at pandas and Julia) who have answered some of my questions.

Why is NumFOCUS important to you?

I really like being a member of this group. It is a privilege to have QuantEcon's name on the same page with the other members of the community.

What would you say to someone who was wondering how to get involved with NumFOCUS?

For users, I would recommend Youtube videos from NumFOCUS sponsored conferences. And I could tell them how empowering the software constructed by community members is.

Sponsored Project Achievements

UIII

itEcon

29:

NUMF^{OCUS}



nteract

During 2017, the nteract project focused on building an ecosystem of web tools around the desktop application. The nteract team released play.nteract.io and app.nteract.io, two web-based interfaces based on the nteract open source ecosystem and connected to the Binder environment. The nteract team also held the first development sprints in California with attendance from both US-based and international collaborators and contributors. Finally, the nteract team continued to improve the performance and development experience of the different open source tools supported by the team.

Jupyter



2017 was a year of exciting growth within the Jupyter community. One of the highlights was JupyterCon, our first international user conference which was co-hosted by our partners at O'Reilly Media. On the software side, we developed a new JavaScript frontend (JupyterLab) that can be easily extended by third parties and whose components can be reused by others who are building complex web applications that require interactive computing. We also deployed new features JupyterHub and binder in the areas of deployment and more advanced multi-user sharing and publishing capabilities. Binder, notably, has been used to share the research of last year's Nobel Prize in Physics, the LIGO/VIRGO Collaboration. Additionally, we have further developed our key libraries IPython, Jupyter Notebook, nbconvert, and Jupyter Widgets.

SymPy

In January of 2017, we published the <u>SymPy paper</u> in PeerJ CS. **Our paper was the <u>most cited</u> paper in that journal for all of 2017.** Over the summer, we had 9 Google Summer of Code students. You can read more about their projects <u>here</u>. In July, SymPy version 1.1 was released. The changes are listed <u>here</u>.



NumPy

For NumPy, 2017 was a year of solid releases and continuing growth in number of contributors (we crossed the 10,000 pull requests barrier!). In total we made two minor and 4 bugfix releases: 1.12.0-1.12-1 and 1.13.0-1.13.3. The two most significant improvements in those release were much improved support for PyPy, and the introduction of the long awaited "__array_ufunc__" attribute which provides improved ability for classes to override the default ufunc behavior. The most significant event in 2017 was that NumPy received two grants, from the Moore and Sloan foundations — the first direct grant funding ever for NumPy! We're looking forward to the acceleration of development in 2018 that those grants will enable.

Shogun



2017 was one of the most busy years for Shogun ever. Shogun joined NumFOCUS and participated in GSoC with 5 students. We had 2 major and 3 minor releases, a developer hackathon at ETH Zürich and another one (sponsored by NumFOCUS) in Budapest. The team got larger and more diverse: Pan joined as a core developer, she also introduced Shogun at CppCon in Seattle; Lea (and Heiko) represented Shogun at the GSoC and NF summits; Viktor is working on Shogun as part of his job, establishing a native Windows port and easy conda-forge based installation.

PyMC3



This time last year, PyMC3 released its first major version since replacing PyMC2. Version 3.0 was released January 9, 2017, and versions 3.1 and 3.2 were released in June and October, respectively. The PeerJ publication that accompanies PyMC3 has now been cited over 80 times and viewed over 20,000 times since its appearance in 2016, highlighting PyMC3's success in academia and industry. The past year included several notable contributions to the project, including work from two of our Google Summer of Code students, Maxim Kochurov and Bill Engels. Maxim extended and refactored the variational inference module, while Bill greatly expanded and automated our support for Gaussian process models. There were several talks, presentations and tutorials about PyMC3 at international scientific meetings, including PyCon (by Eric Ma, Nicole Carlson, and Chris Fonnesbeck), PyData NYC (Nicole Carlson, and Colin Carroll), PyData San Luis (Chris Fonnesbeck), PyData Berlin (Adrian Seyboldt) and ODSC Europe (Thomas Wiecki). We also set up our discourse forum to facilitate our communication with the users and grow the community. It's been visited more than 129k times (page views) since it was set up in June.



Stan

Stan had a highly productive 2017. The latest release in 2017 was Stan 2.17; we added a suite of new features and performance enhancements, largely around matrix derivatives and differential equation solving, released a new interface in Mathematica, and added new language platforms on top of Stan in Scala, F#, and R. Stan also added a half dozen new developers, received grant funding from a range of sources at Columbia and elsewhere, produced multiple new case studies in core features of Hamiltonian Monte Carlo, did a slew of meetups, talks, and peer reviewed papers, saw several books written about Stan, including one in Japanese, had our first Stan conference with 180 attendees in New York, and much much more. All the code and developer detail is on our GitHub organization, stan-dev.



SunPy

SunPy had ten releases in 2017, including a major release that significantly improved SunPy's website, and data download and astronomical spatial co-ordinate capabilities. Improvements were incorporated from sixteen new SunPy contributors and three Google Summer of Code students. 2017 was the year of the <u>Great American Eclipse</u> and SunPy provided a <u>tutorial</u> demonstrating how to overlay eclipse photographs with NASA solar data. In addition, SunPy is being used by the <u>Eclipse Mega Movie Project</u> to create a movie from eclipse photographs submitted by volunteers as it passed over the United States. SunPy was presented at the 2017 <u>PyAstro, American Astronomical Society Solar Physics</u> <u>Division</u> and <u>American Geophysical Union</u> fall meetings.



Open Journals

The Journal of Open Source Software had 1 major release in 2017, as well as 1 major release for Whedon-API (our bot). The JOSS and Whedon-API codebases are now generalized so that we can create new journals more easily. We also had a nice presence at SciPy 2017. **JOSS had strong growth in 2017; we recently crossed the 200-paper mark.**

NUMF^{OCUS}



Astropy

Astropy had a major release (v2.0) this year, and is feature-complete for v3.0, which will be released at the start of 2018 alongside submission of a paper detailing the v2.0 release in the Astrophysical Journal. The Project also held a successful coordination meeting in September in New York city, along with several workshops for developing educational materials and documentation. Also notable in 2017 was crossing the 200-contributors mark. Finally, we highlight that the notable "kilonova" neutron star merger event (in media reports often called the "biggest science result of 2017") was enabled by Astropy — several of the major papers of that event acknowledged Astropy as critical for their work.

yt



In 2017 yt had two minor releases and one major release, yt 3.4. This release included a number of new features, including support for making plots using matplotlib's interactive GUI backends, a routine for making plots of line queries of fields through a dataset, added support for data objects defined as boolean operations on other data objects, added new APIs for accessing fields and creating data objects in a compact syntax useful for interactive work, and added support for a number of application codes and frameworks, including Athena++, Enzo-p, OpenPMD, and AMRex. We have begun work to extract astrophysics-specific functionality and tools into a separate yt_astro_analysis package with the goal of making the codebase less astrophysics-specific and easier to use for people who do not have an astrophysics background. Several yt developers participated in the SciPy conference, including a talk given by Meagan Lang and Nathan Goldbaum on work to make yt support scalable analysis of extremely large particle datasets. According to Google Scholar, the yt method paper was cited more than 100 times in 2017 and is increasingly a standard tool for visualizing, analyzing, and processing 3D numerical simulations in astrophysics.



FEniCS

The FEniCS Project released version 2017.2.0 on December 5, 2017. New features include support for quadrilateral and hexahedral meshes and XDMF output for time series data, implemented by our <u>Google Summer of Code/NumFOCUS students</u>. 2017.2.0 is the last release that will support Python 2. <u>The FEniCS Conference 2018</u> will take place at the Mathematical Institute, University of Oxford, 21-23 March 2018.

NUMF^{OCUS}

julia

Julia

2017 was a year of booming growth for Julia. Last year, there were 5 Julia versions released, including 0.6.0, and since January of 2017, we've seen a 101% increase in the number of downloads of Julia (904k to 1.815m), a 94% increase in the number of publications citing either of the two flagship Julia papers, a 60% increase in GitHub stars for Julia and its registered packages, a 42% increase in the number of Julia questions on Stack Overflow and the Julia Discourse forum. For the first time, JuliaCon was held on the west coast of the US, in Berkeley, CA. Overall, one of the most exciting breakthroughs came with the Celeste project, where Julia achieved peak performance of over one petaflop per second; it's the first high-level dynamic language to do so. We can't wait to see what 2018 has in store for Julia!

QuantEcon

In 2017, QuantEcon re-hired Matthew McKay as lead developer and hired Natasha Watkins as a pre-doctoral researcher. In September, we put together a series of workshops across the US for incoming PhD students that was attended by approximately 400 people. New content continued to be added to the lectures site, and automatic testing was introduced to monitor lectures for code errors. We became more active on social media by establishing a Twitter and Medium blog for QuantEcon news and discussion. Significant work was put into new projects, including a tool to convert text to Jupyter notebooks called Jupinx, and a platform for sharing code called Bookshelf.

PyTables

In 2017, PyTables had one major and two minor releases by Tom Kooij with support for HDF5 1.10.x and other fixes and improvements. We presented a SciPy 2017 tutorial about "HDF5 take 2: h5py & PyTables" by Tom Kooij. Thanks to the HDF Group who sponsored it. PyTables executed the NumFOCUS grant for continuing the "PyTables and h5py integration project" started during our hackfest in Perth. The grant receiver was Alberto Sabater and he did a great work in porting several of the main objects in PyTables (namely: Array, CArray, EArray and VLArray) on top of h5py. Unfortunately, there was not time left for continuing the work with what is possibly the most important object: the Table. We hope to tackle this in a new grant or somebody's PR. We sent a representative (Alberto Sabater) to the NumFOCUS Sustainability Summit held in Austin, Texas in October. In particular, we participated in the group defining roles inside a project (user, participant, contributor, maintainer...) and how to encourage people to get involved. Thanks to NumFOCUS for sponsoring Alberto's travel costs.







pandas

2017 was a year of growth for pandas. There were 3 major and 3 minor releases, offering significant enhancements, bug fixes and performance improvements, while continuing to streamline the user facing API, in preparation for a 2018 1.0 release. We have seen a massive increase in pandas-tagged posts on StackOverflow, where pandas is the fastest growing Python package. We have seen growth in the number of issues posted and number of pull-requests on the pandas GitHub site. Pandas is the focus of talks and tutorials at many Python conferences, and took home the ODSC West Outstanding Data Science Project award. 2017 saw the release of the revised and updated Python for Data Analysis, by Wes McKinney. Pandas looks forward to an even more exciting 2018 with more uptake from many different fields, users, and developers.

Matplotlib

In 2017 Matplotlib had one major version release with updates to the <u>default styles</u> (with 2 bug-fix releases) and one minor version release (with 1 bug-fix release) adding new features. Feature highlights include: basic support for string categoricals, improved polar plotting and reproducible svg/pdf/ps output. **Significant effort** was put into improving the documentation including overhauling and organizing the <u>examples</u> and <u>tutorials</u>. We had approximately 215 people author comments in 2017, 160 of those for the first time. Matplotlib was used in the Nobel prize winning LIGO work.

rOpenSci



In 2017, rOpenSci released 52 new R packages in areas such as geospatial analysis, taxonomy, database backends, climate data repositories, and tools to automate reproducible research across domains. Many of these came from our community of 400 code contributors through our open peer review system. **Downloads of our software (3.7 million this year) have more than doubled over 2016.** A new series launched in our weekly blog gave an opportunity for 16 authors of rOpenSci packages to share their work with a wider audience. Two exciting developments for us include 1) a <u>new software review collaboration</u> with the journal Methods in Ecology and Evolution, and 2) our new postdoctoral researcher began work to understand barriers to reproducible, open science across research domains that rely on open source software.



Econ-ARK

The Econ-ARK project closed 2017 with the exciting news that Jackie Kazil, tech fellow at Capital One, has joined project in an advisory role. Econ-ARK team members have been investing time in improving web presence and proposing talks for Python developer conferences as a way of attracting potential contributors to the project, including PyCon and SciPy. The team of contributors has expanded to over a dozen, including students from University of Copenhagen and University College London. A main focus in our first 6 months has been outreach to the community of potential users to gauge priorities and needs. In pursuit of those goals, project lead Christopher Carroll has visited the IMF, central banks in Denmark, New Zealand, Australia, several universities, and the home base of the QuantEcon project in Canberra, Australia.



The Carpentries

During 2017, Data and Software Carpentry voted to merge their two projects into a new umbrella project called The Carpentries. Work is still ongoing on that merger. Supporting member numbers doubled during 2017, which made it possible for new staff positions in community development, workshop administration and memberships to be created and filled. We taught 402 workshops last year (up from 389 in 2016) to around 7,940 attendees. We just ran our 100th instructor training event (in Africa, a growing community for us). We badged 510 certified Instructors this year (overall total is 1,619) and we currently have 44 Instructor Trainers. Planning got underway on CarpentryCon, which will be our signature event in 2018.

Cecilia Liao

Data Science Manager at dunnhumby



How are you involved with NumFOCUS?

I am one of the co-organisers of PyData London, involved since the very first conference back in 2014, and I am really excited to be co-chairing this year's conference!

What has your experience with the NumFOCUS community been like?

I have always felt welcomed by the NumFOCUS community. I got involved in the first place because of an innocent tweet; at the time, I was learning to use Python for analysing data, and I remarked on Twitter how much I would love to attend a PyData conference if it were held in London. Before I knew it, I was part of the organising community to make it happen. It didn't matter that I had very little experience with Python or scientific computing at the time; the people in the community are more than ready to accept me and help me learn.

Why is NumFOCUS important to you?

NumFOCUS ensures so many important open source projects are made available to the world. Without the organisation's continuing support in these projects, I wouldn't be able to do what I love doing, working on amazing data science projects in retail and consumer finance space.

What would you say to someone who was wondering how to get involved with NumFOCUS?

Talk to someone in the community! Attend a Meetup, tweet a contributor, or just smile and say hello to anyone involved with NumFOCUS and ask how you can help. There are so many ways to give back to the community — help out with the logistics of an event; write documents, blog or a tutorial; contribute to an open source library; or speak at a Meetup or conference. It feels great to give back to the community that makes our day job that much easier!

NumFOCUS Community Achievements

NUMF^{OCUS}



GSoC under NumFOCUS

Matplotlib (2 students) FEniCS (2 students) PyMC3 (3 students) MDAnalysis (1 student) Data Retriever (1 student) Gensim (3 students)

GSoC under Open Astronomy

AstroPy (4 students) SunPy (4 students

GSoC Direct Participants

Shogun (5 students) SymPy (9 students) Julia (18 students) Over 50 students participated in Google Summer of Code 2017, working on 11 NumFOCUS projects



OPEN CODE = BETTER SCIENCE



Max Linke

Ph.D. Student at the Max Planck Institute of Biophysics

How are you involved with NumFOCUS?

I'm organizing the participation of NumFOCUS in Google Summer of Code (GSoC) for the second year, together with Mridul Seth. Besides this, I'm a maintainer for (NumFOCUS Affiliated Project) MDAnalysis since 2015.

What has your experience with the NumFOCUS community been like?

Because I organize the participation in GSoC, I interact with a lot of people in the NumFOCUS community in many different projects. Every person I had contact with so far has been very friendly and helpful. It's especially nice to see that there are a lot of people involved in GSoC with the aim to get people excited about open source science software and into maintaining/improving the code.

Why is NumFOCUS important to you?

There are several reasons why NumFOCUS is important to me. As a scientist I'm incredibly grateful that the software my research relies on is well funded, continuously improved and kept available. As a maintainer of an affiliated project the Small Development Grants that have been handed out last year are a game changer. Having some money to allow contributors to work full time on a project for a week or month or to meet with others can have a huge impact. With the grant MDAnalysis received last year we managed to add support for Python 3, start working on benchmarks and explore how to use dask to make efficient use of modern CPUs.

What would you say to someone who was wondering how to get involved with NumFOCUS?

The community in my experience is very welcoming and appreciates every contribution, from answering questions on stack overflow / mailing list to documentation improvements and code contributions.

My general recommendation if you want to get involved with any open source project is to pick one that you use regularly and care about. I for example started contributing to MDAnalysis because I used it everyday and wanted to run it on python 3. It finally took 2 years to reach my initial goal but learned a lot in that time and helped to introduce others to the project. Because I use MDAnalysis daily I had fun working on it and saw how this work helped me get stuff done quicker. So for me having fun contributing to a project is a very important part.

If you are just starting remember that everyone is doing this in their spare time. Let maintainers answer your questions, or pull requests on their own time. They are grateful for every contribution that you make! Even if they don't answer right away.

NumFOCUS Project Conferences

StanCon 21 January 2017 in New York City

Python in Astronomy 8-12 May 2017 in Leiden rOpenSci Unconference 25-26 May 2017 in Los Angeles

FEniCS'17 12-14 June 2017 in Luxembourg

JuliaCon 20-24 June 2017 in Berkeley JupyterCon 22-25 August 2017 in New York City

QuantEcon PhD Workshops September 2017 at 8

September 2017 at 8 universities across the U.S.



PyData

PyData provides a forum for the international community of users and developers of data analysis tools to share ideas and learn from one another



PyData is more than a conference series — it is also a global network of local meetups. The PyData network promotes discussion of best practices, new approaches, and emerging technologies for data management, processing, analytics, and visualization. PyData communities approach data science using many languages, including (but not limited to) Python, Julia, and R.

NumFOCUS Community Achievements



Stats



Conferences 3500 Attendees 330 Speakers



Meetups 50k members 73 chapters **33** countries



Twitter 22k followers



2017 PyData Event Locations

Amsterdam, Netherlands	New Delhi, India
Barcelona, Spain	New York, USA
Berlin, Germany	Paris, France
Budapest, Hungary	Rimini, Italy
Cardiff, UK	San Luis, Argentina
Florence, Italy	Seattle, USA
Karlsruhe, Germany	Warsaw, Poland
London. UK	Washington, D.C., USA

"My coworker and I both attended **PyData Seattle**, agreed that **it was the best conference we had ever been to,** and made it a goal to get to another in 2018."

- Phil Anderson, Senior Data Scientist at 84.51





DISC Diversity and Inclusion in Scientific Computing

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

The DISC Program strives to help create a more diverse community through initiatives and programming devoted to increasing participation by and inclusion of underrepresented people.



2017 DISC Committee Members:

- Tobi Bosede Alex Companioni Gina Helfrich Julie Hollek Kari Jordan
- Jennifer Klay Julia Meinwald Erica Moszkowski Manuel Rivas Reshama Shaikh

2017 DISC Initiatives

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.

Made possible by a generous grant from



In support of these goals, in 2017 the DISC Committee produced:

- A survey to collect data on the attendee experience at NumFOCUS-sponsored conferences
- The DISCOVER Cookbook (Diverse & Inclusive Spaces and Conferences: Overall Vision and Essential Resources)
- Women's Intro to Data Science workshop held at the University of San Francisco, with funding made possible by ODSC
- The inaugural <u>DISC Unconference</u> (held at PyData NYC)

Sustainability Program

Made possible by a generous grant from



The purpose of the **Sustainability Workshop** was to gather representatives from the NumFOCUS Sponsored Projects as well as stakeholders from the Board of Directors, Advisory Council, Sustainability Advisory Board, and NumFOCUS staff, in order to create **a shared vision** and strategic action plan for **addressing sustainability across NumFOCUS and its member projects.**



The 2017 NumFOCUS Summit featured a Sustainability Workshop held in Austin, TX at the offices of NumFOCUS' Corporate Sponsor, IBM .

Stakeholder Groups

- Community, Project, and Ops Managers – 8
- Core Developers, Funded – 6
- Core Developers, Unfunded – 6
- Facilitators 3

- Implementers 9
- Project Leads, Funded – 7
- Project Leads, Unfunded – 6
- Sustainers 7
- Attendees 51



Paige Bailey

Sr. Cloud Developer Advocate – Machine Learning/ AI at Microsoft

How are you involved with NumFOCUS?

If you're a Python developer or a data scientist, even if you don't think you're involved with NumFOCUS, you probably are! So many of our discipline's tools are supported financially by the organization: pandas, Matplotlib, Jupyter notebooks, and NumPy, among others. Other than being a delighted user and minor contributor to NumFOCUSsponsored products, I've spoken at PyData Dallas, PyData at Strata + Hadoop, and PyData NYC. Last year, I was also fortunate to attend the Diversity and Inclusion in Scientific Computing (DISC) Unconference at PyData NYC.

What has your experience with the NumFOCUS community been like?

The NumFOCUS community is tirelessly devoted to serving its users and to resourcing open source products that have impact (either financially or through developer contributions). And even though this is very much an organization of engineers, accessibility, inclusion, and diversity initiatives are included for every single event. I've forwarded NumFOCUS's D&I best practices to several conferences and organizations to use as a template for their own efforts.

Why is NumFOCUS important to you?

As a machine learning practitioner who frequently creates predictive models that impact real humans and real lives, ethics are at the forefront of everything I do. Every data set used must be interrogated, and every project should be tirelessly questioned through each step in the data science process. Is this unintentionally marginalizing a demographic? Am I negatively impacting a population? Could this have negative implications on lives?

During the Diversity & Inclusion Unconference, my team was able to build a sort of "Ethical Checklist" for machine learning engineers — a guide for data scientists as they create projects and a series of questions that they can ask to help prevent unintentional bias. If you have energy on this topic, please take a look at the <u>repo on GitHub</u> and contribute!

What would you say to someone who was wondering how to get involved with NumFOCUS?

There are <u>so many</u> meaningful ways to become involved with open source projects, other than just pushing code: documentation improvements, tutorial and sample reviews, triaging issues, tagging issues in GitHub repos as they are posted, adding unit tests, doing design work. Let your personal passion and talents drive the contributions, and ask the current maintainers which adds would be most helpful. You are offering your time — the biggest gift a human can give! — and we appreciate it. So much.

2017 Donors List

1.7



David Adamson Joshua Adelman Denis Akhiyarov Majid alDosari Waleed Alhaider Alexandre Almosni Ahmed Alrasheed Bruno Amorim **RK** Aranas Steven Armour Laksh Arora Mohammad Athar Rajen Athreya **Ricardo Azevedo** Pete Bachant Toni Bagur Cantallops John Bampton **Bob Baxley Michael Becker** Neal Becker David Bellavia Todd Benanzer Iddo Berger Mathias Berger Pravee Bhamidipati Vineet Bhatawadekar Chris Binz Christopher Birster Gerard Blais Paul Blelloch Neraj Bobra Tim Bock Valentina Borghesani Vicki Boykis Rüdiger Braun Jay Bredenberg

Samuel Brice Anthony Brown Mark Brown Nathaniel Brown Peter Bull Marvin Bürkner **Christopher Burns Gonzalo Bustos** Christopher Calloway Andrea Canella Juan Luis Cano Marcos Carreira Colin Carroll Dan Case Blandon Casenave Mizue Castella Maxim Chaplianka Danning Chen Alin Chis Steven Christe Daniel Chudnov Andrey Chufyrev Winston Churchill-Joell Justin Clark Crockett Cobb Elior Cohen Scott Collis Julian Colomina Ken Colton Alex Companioni Jordi Contestí Riccardo Conti **Dolcera** Corporation Francisco Correoso Pierre Costini Matthew Craig

Sahil Dahiya Zaisheng Dai Abhijit Dasgupta Jonathan Davies aartt de beer Institut de I ENS Leonardo De Marchi Rodolfo De Nadai Adam DeConinck Stefania Delprete **Brian Dennis** Jack Devine Karl Diether Eric Dill Dan Dixey Martey Dodoo Paul Donovan Alaina Dooley Vincent Dorie Vladislavs Dovgalecs Joe Downs Pavel Dytrych Jonathan Eckel Sinziana Eckner Blerim Emruli Ahmet Erdemir Adam Eury Henry Ferguson David Fioretti **Donald Fischer** Frederic Foiry Jared Freeman **Brian Fuller** Simon Funke Gauden Galea Amy Galick

Pradeep Ganesan Dennis Gannon Olivier Garet Guido Gay James Ge A. Gel Ajish George Marc Gi Noah Gift Marco Goldin David Goldsmith Aaron Gonzales Jeffrey Gostick Jeffrey Gould Al Graham Joe Greener Julia Greenwald Jeffrey Griffin Capital Group Grover Group, Inc. Prasenjit Gupta Julio Guzman **Bob Haffner** Jack Hale Matt Hall **Robert Hamilton** Chris Hamm **Jiedong Hao** Páll Haraldsson Antti Härkönen **Tim Hawes Brian Hayes** Edward Haymore Peter Heinrich Nima Hejazi Gina Helfrich

OPEN CODE = BETTER SCIENCE

Jody Hey Steve Holden Jeffrey Hollister Morgan Hough **Daniel Hrisca** Chloe Hsu Jefferson Huang Kenneth Hudson Max Humber Tom Hunger Andrew Inglis Kota Ishihara S.M. Islam Zunbeltz Izaola Payoj Jain David Jaluvka Robert Jensen Ewa Jodlowska Cory Jog Keith Johansen Jeremiah Johnson Al Johri Liam Jongsu Kim Max Joseph Moideen Kalladi Daniel Kapitan Sam Kaplan Aaron Karper Brian Keegan Craig Kelly YongHee Kim Brian King Tim Knight Ilia Ko Muhammed Kocabas Chakravarti Kodavanti Lancine Konate Karuna Koppula Chris Kortge Jozef Kosik Leonid Kovalev Sebastian Kretzig Tom Ku Marek Kwiatkowski Hyukgyu Kwon **Flavien Lambert Drew Lancaster** Anton Landberg **Tormod Landet** Peter Landwehr Junpeng Lao John Lashlee Scott Lasley Gregory Lee Kang Lee Young Lee Othmar Lehner Gregory Lett Randall LeVeque Max Linke Anton Logunov Akhil Lohia **Christian Long** Chalmer Lowe Kiatikun Luangkesorn Jose Machicao Don MacMillen Andrey Malakhov **Miguel Manese** Surrendra Markandaya Osvaldo Martin Sergey Maslakov

Frits Mastik Markus Mayer Pamela McA'Nulty Byron McLendon **Edkent Media** Andres Medina Hans Olav Melberg Nicco Mele Andrey Melentyev Nancy Melucci Reda Merzouki Mark Mikofski Scott Miles Claudia Millán **Clayton Miller** Dana Miller Antonio Giulio Minichiello Thomas Moerman Srinivasa Mohan Ehsan Mohyedinkermani Agustin Montero Jose David Montova John Moody Mike Morgan James Morrison Samantha Mowrer Stuart Mumford Nicholas Murphy Christopher Musselle Benjamin Naecker Masaaki Nakano John Nelson Gia Nguyen **David Nicholson** John O'Meara William O'Shea

Tatsuro Oda Nathaniel Odell Johan Öfverstedt Takao Okuda **Travis Oliphant** Bernardo Oliveira Sergi Ortiz Rodriguez Joel Ostblom Axel Pahl Serge-Étienne Parent **Hilary Parker** Matt Parker Paul Patchem Gustavo Patino Grant Paton-Simpson Inti Pedroso Jason Pell John Pellman João Marcello Pereira Jesse Perla Ted Petrou Peter Petto David T. Pitts, Jr. Eraldo Pomponi Vasily Povalyaev Darafei Praliaskouski Ian Price Yonghua Qu Bruno Quinart Neil P. Quinn johnny Radin Varun Rajan Natarajan Ramachandran Karthik Ramasamy Karina Ramos Musalem **Rick Randall**

OPEN CODE = BETTER SCIENCE

David Ranzolin David Rasch **Gunnar Ratsch Rosen Raychov** Greg Reda Justin Reid Sebastien Renard **Christopher Roach** Matthew Roberts David Roher Bernardo Roschke **Kyle Ross** Marjorie Roswell Juanjo Rovira George Rowe **Brett Rudder** Patrick Ryan **Goldman Sachs** Alejandro Sáez Adam Sah Matthew Sanders **Nathan Sanders** TailorDev SAS Kenta Sato **Tanmay Satpathy** Nathanael Schilling **Michael Schroder** Zachary Schubert Laurent Schüpbach **Robert Schwarz Ridgway Scott** Skipper Seabold Mark Servilla Will Shanks Joseph Sheehan Hiroyuki Shigeta

David Shinn Raphael Silva Joseph Slater **Rachel Slaybaugh** Todd Small Adam Smith John Smith Jonathan Smith **Robert Smith** Yaser Souri Sandro Sousa **Emily Spahn Christian Staudt Ryan Steckel** Liliya Stefanova William Stein **Biel Stela** Jan Stembera Jeff Sternberg Heiko Strathmann David Strozzi Susan Sun Tré Sweeney Mark Switajski Tomohiro Takahashi Mrunmay Talegaonkar **Roland Tanglao** Brian Taylor Matt Taylor **Renee Teate** Fernando Teixeira Boon Jin Tek **Bartosz Telenczuk** Pietro Terna Ravi Thambireddy Adam Thomas

Jared Thompson Simon Thordal Paolo Toccaceli Larry Tooley Jeffrey Tratner Shingo Tsuji Vlad Tyshkevich Vladislav Tyshkevich Sergio Uribe Nicholas Ursa Just User Ranjini Vaidyanathan Anne Van Dyk Joe Van Steen Lucas Verney Victor Vicente Alejandro Vidal Vidal Jorge Viram Divakar Viswanath Baskaran Viswanathan Bharath Vullanki Sophie Wagner Jarrett Walker + Associates LLC Aaron Walters Liang-Bo Wang Dean Ware **Tiffany Washburn** Jerry Watkins Stewart Watts Mauro Werder Greg Werner Matthew Widjaja **Tiffany Williams Benjamin Winkel**

Lukas Wittenbecher

Francis Wolinski Mathew Woodyard Jinsong Wu Minwei Xu Kimberly Yang YourCauses Terri Yu Hui Zhang David Žihala Michael Zingale



2017 Corporate Sponsors



With Support From





OPEN CODE = BETTER SCIENCE

2017 Financials















Project Fund Details



Board of Directors



Advisory Council

