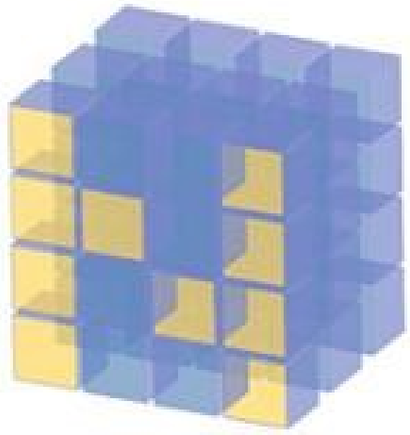


NumPy

Inside NumPy: preparing for the next decade

Matti Picus, Ralf Gommers, Sebastian Berg,
Tyler Reddy, Stéfan van der Walt, Charles Harris

NUMFOCUS
[FISCALLY SPONSORED PROJECT]

Adapted from a talk at SciPyUSA 2019 https://www.youtube.com/watch?v=dBTJD_FDVjU

A very brief history of NumPy

2005: NumPy created

2006: NumPy 1.0

.....: 1-2 releases every year, gradual progress

2015: governance and NumFOCUS relationship formalized

2018: first-ever paid developers hired

The Sloan & Moore grants to BIDS for NumPy

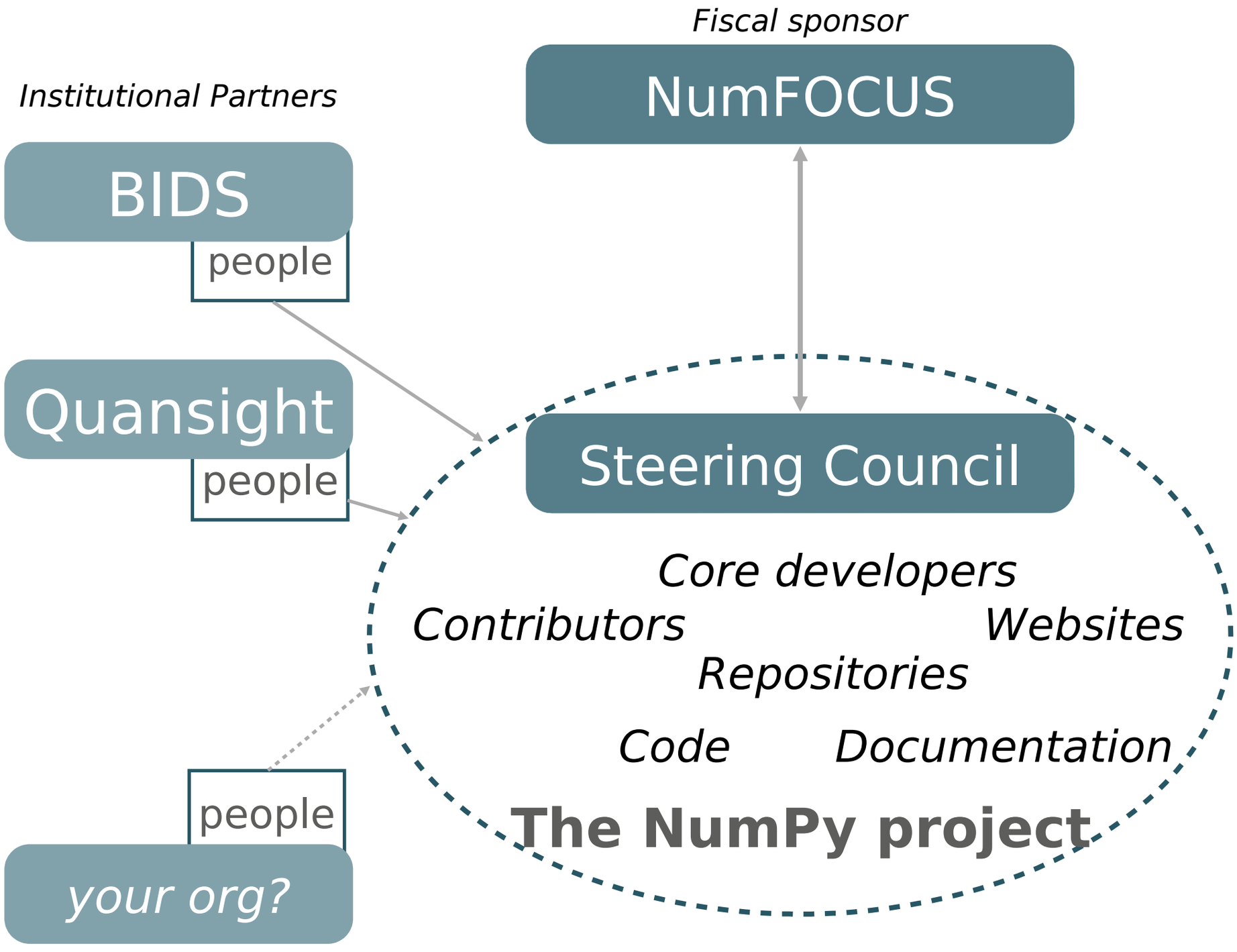
- Two grants for in total $1.3M, Apr 2018 — Oct 2020
- Current co-PIs: Stéfan van der Walt and Fernando Perez
- Social aims: Improve community engagement Grow core team, Diversify contributors
- Technical aims: More flexible & sustainable code, Frequent & consistent releases, Improve data type system, New array protocol

GORDON AND BETTY MOORE FOUNDATION

Berkeley UNIVERSITY OF CALIFORNIA

ALFRED P. SLOAN FOUNDATION

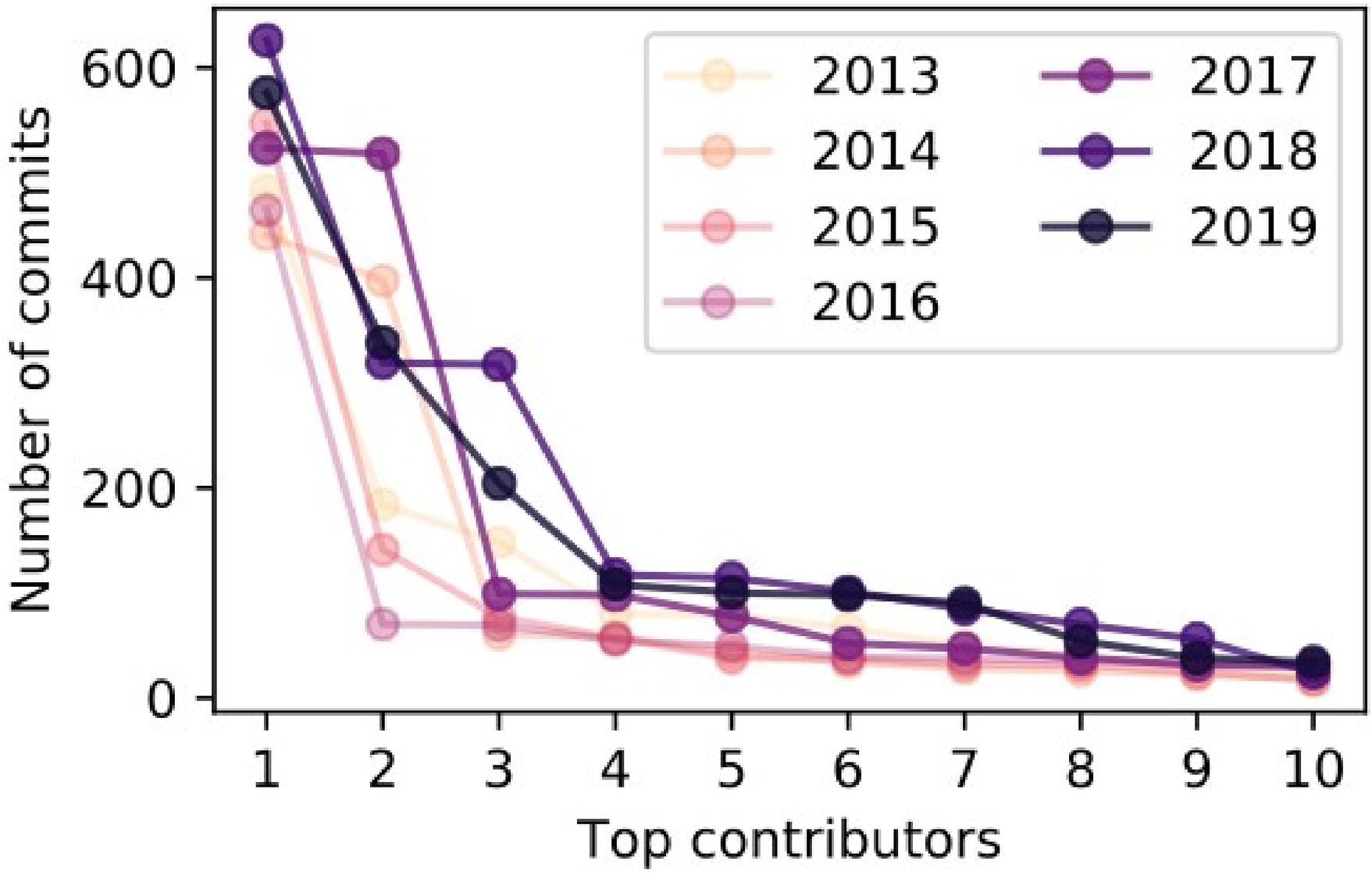
An organizational view of the project



Key questions we'll try to answer in this talk

- Has grant funding over the last year invigorated the NumPy project?
- How healthy/sustainable is NumPy today?
- What does NumPy need in order to thrive?
- What's our vision for NumPy, and what is our plan to achieve that vision?

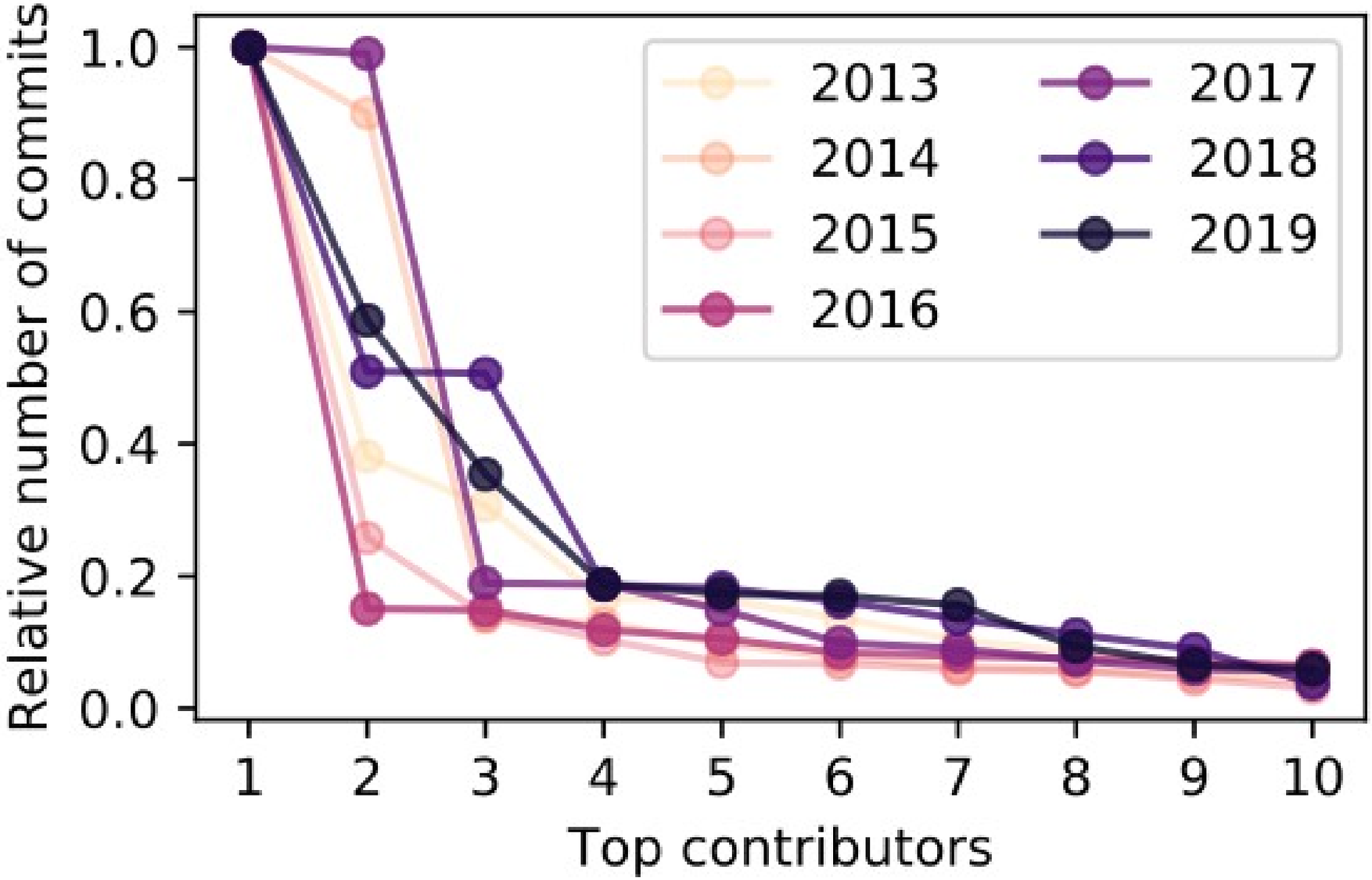
Impact of grant funding — an attempt to quantify



Heavier-tailed distributions in 2018-19: bus factor increased.

Fernando Perez, [Ten years of (interactive) scientific computing](#)

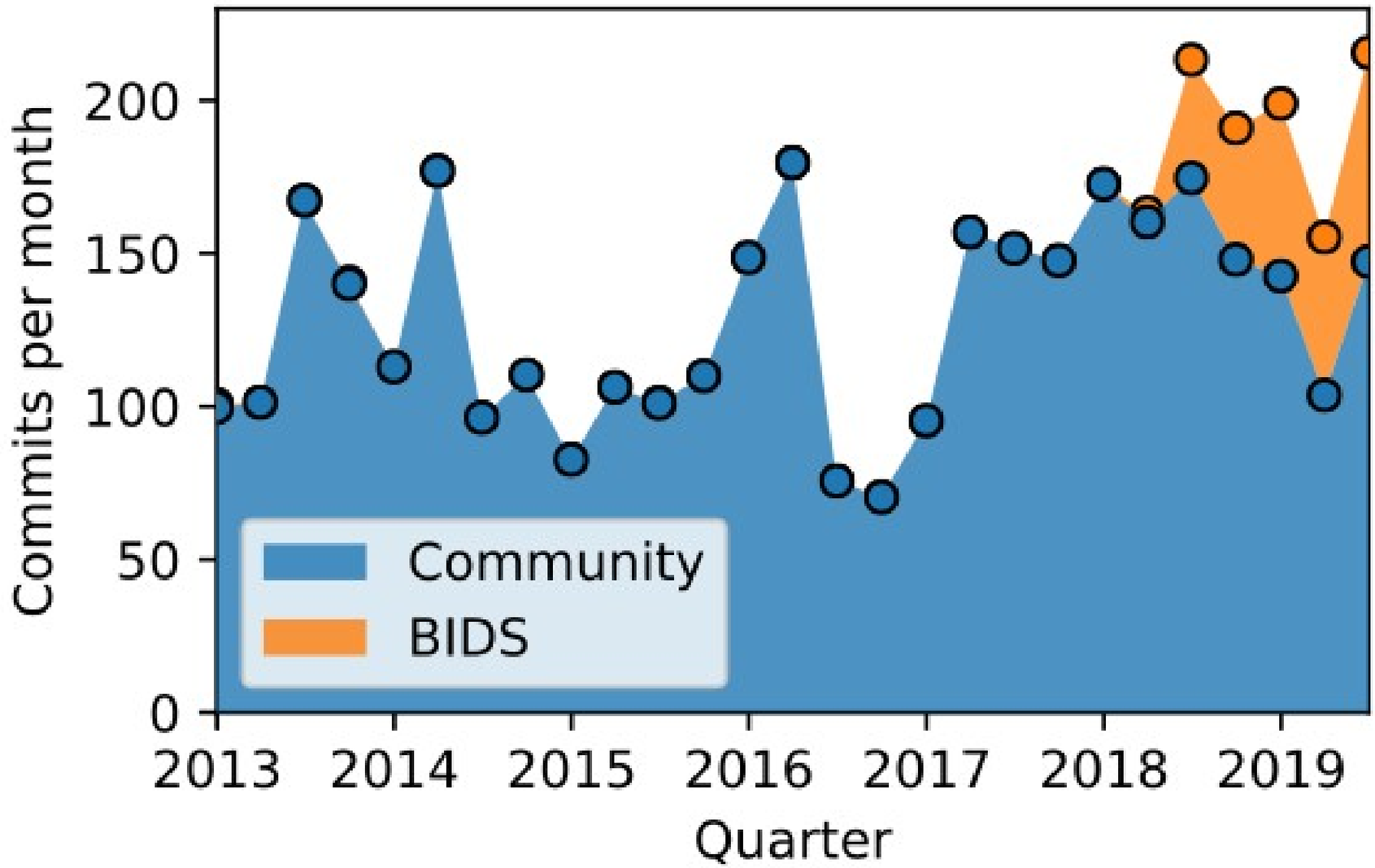
Impact of grant funding — an attempt to quantify



Heavier-tailed distributions in 2018-19: bus factor increased.

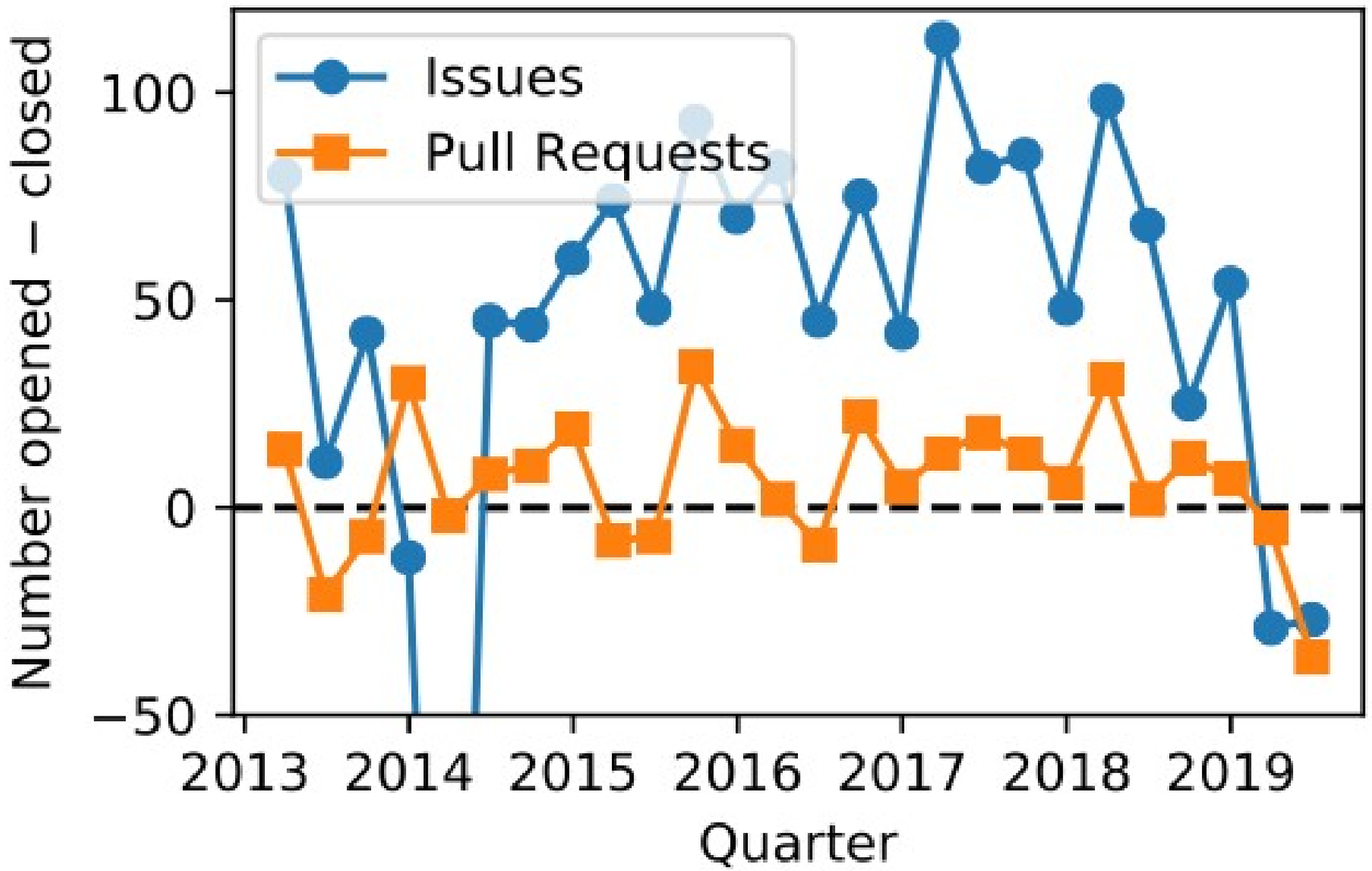
Fernando Perez, Ten years of (interactive) scientific computing

Impact of grant funding — an attempt to quantify



Volunteer contributions relatively stable; total activity up.

Impact of grant funding — an attempt to quantify



Issue/PR count rate of change is key indicator of project health.

Nadia Eghbal, [Methodologies for measuring project health]

Impact of grant funding — qualitative

- We now have a **roadmap**
- We've been able to organize **core team sprints** about every two months with around 5-10 participants
- **NumPy Enhancement Proposal** process revived – including implementing matmul, the new random number module, __array_function__ protocol, and more
- We paid down a decent amount of **technical debt**
- The 1.17.0 release is the **largest release** since 1.7.0 in 2013.
- Planning the ***Array Developer Summit*** for next March for all tensor/array-like projects

Has grant funding over the last year invigorated the NumPy project?

Yes, we think so.

Faster progress and more hands for maintenance also makes it *more fun,* more bandwidth for outreach and education.

How healthy/sustainable is NumPy today?

More healthy than a year ago.

Still a worry though. Bus factor estimate: ~5

Hiring

So who did we hire?

Two full-time engineers, ongoing (till Oct 2020 at least):

- Mar 2018 - ? : Matti Picus (PyPy team member)
- Jun 2018 - Jun 2019: Tyler Reddy (SciPy team member)
- May 2019 - ? : Sebastian Berg (NumPy Steering Council member)
- August 2019 - ? : Warren Weckesser (SciPy team member)

Additionally:

- Supporting Kriti Singh through Outreachy (docs work)
- Participating in GSOD and growing a web/doc team

Funding

Funding — what to pay for?

- Pay for things that are important *and* otherwise won't get done.
- Think broadly. There's more to a project than code!
- Balance maintenance & innovation: keep people motivated!
- Plan for sustainability. Don't start depending on funding unless you’re fairly certain it’s stable.
- Make the life of volunteer maintainers **easier**, not harder!

Funding — who to pay?

- All else being equal, give preference to existing maintainers.
- Pay attention to communication and self-management skills, in addition to technical skills and motivation.
- Consider this an opportunity to make your project more diverse.

Challenges

Sustainability — maintainer bandwidth

Of the 11 Steering Council members:

- 3 are very active (Chuck, Eric, Stephan)
- 3 are paid to work on NumPy
  (Sebastian full-time,
   Stéfan & Ralf a small part of their time (~1 day/wk))
- 5 are in low-activity mode (infrequent emails/commits)

This is still a major challenge! NumPy depends both on a handful of people, and probably on continued funding.

The project beyond code

It’s still hard getting work done on:

- High-level documentation
- Website
- Community building
- Governance and project management
- Long-term planning (both technical and organizational)

Improving NumPy’s culture

- We still struggle with a lack of diversity: *all* maintainers are white and male.
- Our GitHub and mailing list culture is infrequently not as friendly and welcoming as we’d all like it to be.

Vision

Scope & Vision

The key thing NumPy offers is:

an array object (*N-dimensional, in-memory, on CPUs*)
and array computing APIs

NumPy lives at the heart of the numerical Python ecosystem. We want to:

evolve while remaining a stable base,
address bottlenecks that limit how the
wider ecosystem can grow,
and grow and diversify our team and community.

What does NumPy need in order to thrive?

- Sustained funding at a higher level than today.
- More bandwidth from key people for long-term planning, managing the project
- Attracting people in roles that primarily focus on activities other than coding
- Also fund other key components in OSS data science: SciPy, Matplotlib, pandas, Scikit-learn, ...

Parts of a plan to get there

- In the next year, work with tech writers and web developers to start building a docs/web team.
  Efforts already ongoing.

- Build a diverse and robust funding quilt:
  Express our **needs** as a community (not just NumPy), and **ask** funders and major users for contributions:
  Recently started (e.g. pitched to NASA), want to make this much more concrete.

Maintenance		
Key Improvements		