

# A Solid Foundation for Statistics in Python with SciPy

Warren Weckesser  
Berkeley Institute for Data Science

Matt Haberland  
Assistant Professor, Cal Poly



# SciPy – The Library



- Core library in the Python scientific computing ecosystem.
- Started in 2001, currently downloaded millions of times each year.
- Over 6000 other packages depend on SciPy, including scikit-image, scikit-learn, Biopython, PsychoPy, ...
- The web page has over 3000 journal citations.

Just published in *Nature Methods*:

“SciPy 1.0: fundamental algorithms for scientific computing in Python”

<https://www.nature.com/articles/s41592-019-0686-2>

- SciPy includes algorithms for:

linear algebra	sparse matrices	differential equations
optimization	interpolation	special functions
<b>statistics</b>	... and much more.	



## A Solid Foundation for Statistics in Python with SciPy

*What are your goals and expectations for the coming year?*

- Improve the statistics library in SciPy.
  - Implement more tests, and improve existing tests.
  - Expand and improve the library of probability distributions.
  - Tackle older maintenance tasks.
- Improve diversity of SciPy contributors.
  - Engage students in open source development by hosting a coding sprint at Cal Poly in collaboration with clubs that support underrepresented groups.



## A Solid Foundation for Statistics in Python with SciPy

*What do you hope to achieve by learning from or collaborating with other grantees?*

- We would love to learn more about SciPy users.
  - What do you use?
  - What do you need?
- What else can we do to improve diversity of contributors?

*What expertise do you have to share with other grantees?*

- Warren Weckesser and Matt Haberland are long time SciPy developers, with experience in numerical computing, optimization, statistics and signal processing.