# Dr. Kyle Barbary

Data Scientist and Scientific Software Engineer
E-MAIL: kylebarbary@gmail.com
PHONE: 510.684.7308
GITHUB: http://github.com/kbarbary
WEBSITE: http://kylebarbary.com

I'm a data scientist and quantitative researcher with extensive experience in scientific software engineering, large datasets, statistics and modeling. I'm passionate about (1) building better predictive models that have real effects on a product and (2) designing intuitive tools to accelerate research, improve reproducibility and enhance understanding.

## Education

**University of California, Berkeley – Ph.D. in Physics**
Thesis: *High-Redshift Type Ia supernova rates in galaxy cluster and field environments*
Advisor: Saul Perlmutter (2011 Nobel Laureate in Physics)
**Harvey Mudd College – B.S. in Physics with Distinction**
Thesis: *Measurement of the spatial coherence of second harmonic light produced at a gold surface*

## Skills

**Data:** 13+ years experience dealing with large imaging datasets, image and signal processing, modeling, numerical optimization, MCMC methods, Bayesian statistics, hierarchical models.

**Software:** 15+ years experience: Python (numpy, scipy, matplotlib, pandas, cython), experience in C, Julia, SQL, UNIX. Release and maintenance of open-source libraries for astronomy, signal processing, and statistics in Python/C and Julia. Performance optimization of numerical software. 10+ years experience with testing, documentation tools, build systems, packaging, continuous integration, git, and collaborative development.

**Communication:** Extensive experience public speaking, writing journal articles, proposals, blog posts, communicating results.

**Teaching:** Experience planning and teaching workshops and tutorials, mentoring students and junior data scientists and engineers.

**Design:** Experience with website and documentation design (HTML, CSS, markdown, ReST, LaTeX), presentations, logo design for several software projects.

## Selected Software

I wrote and maintain 4 packages on the Python package index and several more on the Jula Package index. See http://kylebarbary.com/software or http://github.com/kbarbary for more details. A few packages where I am the primary author:

**SNCosmo** Python library for supernova cosmology. (★53)

**SEP** Python and C library for Source Extraction and Photometry on astronomical images. (★125)

**Nestle** Nested sampling algorithms for Bayesian evidence in Python. (★65)

**FITSIO.jl** FITS file format I/O library for Julia. (★41)

**Dierckx.jl** 1-d and 2-d basis splines in Julia; Fortran wrapper. (★108)

## Experience

2019-2021 **Senior Data Scientist, Zymergen**
Upon joining Zymergen, the Data Science team was using hierarchical Bayesian models for estimating and removing unwanted process effects from experimental strain data. However, the models were vastly overparameterized and engineered in a way that prevented experimentation and development. Worked to spread knowledge within the team about appropriate parameterization of hierarchical Bayesian models and choice of priors. Designed a new Stan-based Python library and Airflow pipeline that allowed simplification of existing models, rapid development of new models, and applying models to new use cases. New pipeline improved reliability and runtime, and reduced lines of code by a factor of three. Guided the design and implementation of a new outlier detection algorithm and accompanying microservice. Worked on fixing data model limitations that were holding the data science team back, and guided the team's tooling through a transition from MySQL to Snowflake.

2018–2019 **Senior Quantitative Researcher, The Climate Corporation**
Developed and improved predictive models for corn diseases, based on a combination of agronomic studies, disease data and machine learning approaches. Worked with teams across the company to understand and access datasets. Improved runtime of disease risk prediction code by a factor of 10, eliminating a research bottleneck. Developed detailed model of noise in field trial experiments, demonstrating the need for improved datasets. Discovered a cross validation failure with significant impact on product direction.

2017 **Scientific Software Consultant**
Improved and extended the Celeste.jl codebase to run on multiple data sources. Celeste is the first program written in Julia to achieve over 1 petaflop performance.

2014–2017 **Project Scientist, UC Berkeley & Berkeley Institute for Data Science**
**Cosmology Data Science Postdoc**
Rewrote a key piece of the data analysis pipeline for the Nearby Supernova Factory. After rewrite, code returned more accurate answers, ran faster and more reliably, with a tenth the number of lines of code. Released on GitHub. Released open-source Python libraries for supernova cosmology, analysis of astronomical images, and nested sampling (a MCMC-like statistical sampling technique). Developed several core astronomy libraries for Julia and managed the "Julia Astro" organization. Taught tutorials on Python packaging, licensing, testing, documentation, Julia, git and GitHub. Organized several workshops for statistical and computing methods in astronomy.

2012–2014 **Director's Postdoctoral Fellow, Argonne National Laboratory**
Alongside research program, developed modern, extensible Python library for supernova cosmology, enabling studies of systematic uncertainty that were formerly burdensome or prohibitively difficult.

2011–2012 **Postdoctoral Fellow, Lawrence Berkeley National Lab**

## Talks and Publications

Author of over 30 peer-reviewed articles in Astronomy. Referee for several astronomical journals, Supercomputing conference.

See kylebarbary.com and kylebarbary.com/research for selected publications and more details on my work in astronomy and astronomical software.

## Fun facts

I raced road bikes in college and grad school at a national level; now I ride for fun and I'm also getting into mountain biking. I'm an avid skier in the Winter and enjoy backpacking in the Summer.