

JAKE VANDERPLAS

Senior Data Science Fellow

eScience Institute, University of Washington

WRF Data Science Studio
University of Washington
Box 351570
Seattle WA 98195-1570
jakevdp@uw.edu

- I am part of the University of Washington's eScience institute, an interdisciplinary program designed to support data-driven discovery in a wide range of scientific fields. Alongside my institute duties, I work on my own research in astronomy, astrophysics, machine learning, and scalable computation.
- Previously I was an NSF post-doctoral fellow, working jointly between the Astronomy and Computer Science departments at the University of Washington. During that fellowship, I began work on pushing the limits of large, distributed array-based computation with *SciDB*, and develop techniques for data mining and machine learning in large astronomical surveys like *LSST*.
- My PhD research focused on *weak lensing*, a technique which utilizes small gravitational perturbations of light paths to learn about the cosmological distribution and properties of matter in the universe.
- I am interested in encouraging reproducible and open research practices across scientific disciplines. To this end I spend significant time developing and presenting tutorials on scientific computing in the Python programming language, and investing time developing a variety of open-source scientific computing tools.

EDUCATION

- PhD** Department of Astronomy, University of Washington, Seattle, WA
2006-2012 advised by Andrew Connolly and Bhuvnesh Jain (U. Penn)
Thesis: [Karhunen-Loeve Analysis for Weak Gravitational Lensing](#)
- MS** Department of Astronomy, University of Washington, Seattle, WA
2006-2007 advised by Craig Hogan and Andrew Becker
- BS** Calvin College, Grand Rapids MI
1999-2003 Major: Physics; Minors: Mathematics & Japanese
4-year letterman & 1-year captain of the swimming & diving team

EXPERIENCE

Employment

- UW eScience** Senior Data Science Fellow & Director of Research for Physical Sciences
2014-Present University of Washington eScience Institute.
- UW Computer Science** NSF post-doctoral fellowship, CI-TraCS program.
2013-2014 Department of Computer Science, University of Washington. Supervised by Magda Balazinska
- UW Astronomy** Postdoctoral Researcher, LSST Image Simulation group.
2012-2013 Department of Astronomy, University of Washington. Supervised by Andrew Connolly
- UW Planetarium** WorldWide Telescope Planetarium Project Coordinator
2010-2012 University of Washington Planetarium, Seattle WA & Microsoft Research, Redmond WA
- UW Planetarium** K-12 and Community Outreach Coordinator
2008-2010 University of Washington Planetarium, Seattle WA
- Mount Hermon** Experiential Science Educator (4th-8th grade students)
2004-2006 Mount Hermon Outdoor Science School, Santa Cruz CA
- Summit Adventure** Wilderness Instructor: Backpacking, Rock Climbing, and Mountaineering
2004-2005 Summit Adventure, Bass Lake CA
- Japan ESL** Teacher and Tutor of English as a second language
2003-2004 Sendai, Japan

Volunteering

- Neighborhood Advocacy** 2013–2015 As co-chair of the North Delridge Neighborhood Council, I facilitated community gatherings, service work, and other advocacy in our mixed-income neighborhood in southwest Seattle.
- Safe Streets Advocacy** 2010–2014 As a founder of West Seattle Greenways and transportation chair of the Delridge Neighborhood Council, I led the effort to secure grants and city funding totaling over \$2 million for pedestrian and bicycle safety improvements in the neighborhood.
- Pacific Science Center** 2009–2013 As a Science Communication Fellow, I facilitated activities for museum visitors and gave occasional community talks on astronomy and astrophysics.
- Sierra Club** 2007–2012 As a program leader for the Sierra Club's *Inner City Outings* program, I led 3-4 hiking & camping trips each year with Seattle youth.
- UW Planetarium** 2006–2012 Through my graduate career, I participated in the University of Washington Planetarium's K-12 outreach program, facilitating planetarium shows several times each quarter for visitors aged 4 to adult.

Formal Teaching

- CSE 599/ChemE 599** Winter & Fall 2016 Software Engineering for Data Scientists
University of Washington
- Astr 599 / Applied Math 500** Fall 2014 Python for Scientific Computing
University of Washington
- Astr 599** Fall 2013 Scientific computing for Astronomy
University of Washington

Service

- Journal of Open Source Software** 2016-present Editor, Astronomy and Machine Learning
- Python in Astronomy** 2016 SOC member & LOC chair for this 60-person, week-long workshop focused on development of Python tools for astronomers.
- PyData Seattle** 2015 Member of the local organizing committee.
- Computing the Universe** 2015 SOC member for this week-long workshop at the University of California, Berkeley.
- Journal of Statistical Software** 2014-present Associate Editor, primarily focusing on submissions involving Python-language software.
- Astro Hack Week** 2014 Lead organizer for this week-long, 50-person workshop at University of Washington.
- SciPy Conference** 2014 Tutorial co-chair for 2014 SciPy conference.
- PyCon** 2013-2016 Member of the tutorial review committee for PyCon 2013-2016.
- PyData** 2012-2016 Member of the organizing committee for several PyData conferences.

Mentoring

- David Fleming** 2015-Present Graduate Student, UW eScience Institute & Astronomy Department
- Grace Telford** 2015-Present Graduate Student, UW eScience Institute & Astronomy Department
- Jes Ford** 2015-2016 Postdoctoral Fellow, UW eScience Institute & Astronomy Department
- Xiaofeng Meng** 2015-Present Postdoctoral Fellow, UW eScience Institute & Earth and Space Sciences
- SungWon Kwak** 2012–2013 Undergraduate, University of Washington Astronomy
Superimposed High Redshift Spectra

Andy Barr & Undergraduates, University of Washington Pre-MAP program
Devon McMinn *Astronomical Data Processing with LLE*
2008–2009

AWARDS & HONORS

International Astrostatistics Association Winner of the 2016 *Outstanding Publication in Astrostatistics* award for *Statistics, Data Mining, and Machine Learning in Astronomy* (With Z. Ivezić, A. Connolly, and A. Gray)
July 2016

Data Analysis Contest Winner of “Most Insightful” analysis in the 2015 Pronto Data Challenge
November 2015

Data Science Named *Senior Data Science Fellow* at the UW eScience Institute
2014–Present

Data Visualization Runner-up in the 2013 *John Hunter Excellence in Plotting Competition*
July 2013

CIDU Best Paper Recipient of the Best Paper Award, 2012 Conference on Intelligent Data Understanding (CIDU).
October 2012

NSF Fellowship Recipient of a 3-year NSF prize fellowship through the office of CyberInfrastructure CI-TraCS program. NSF Award #1226371.
2012

Calvin College 4-year recipient of the Calvin College Presidential Scholarship; 4 years on MIAA Honor Roll; recipient of the Roger D. Griffioen Physics Scholarship.
1999–2003

COMPUTING

I am an active developer, maintainer, and contributor to several well-known scientific computing packages in the Python community. See my github profile (<http://github.com/jakevdp>) for details.

Skills

- Experienced open source developer, with a specialization in scientific computing, including visualization, data mining and machine learning.
- Expert in the Python Language and extensions such as Cython; very good knowledge of C, C++, and interfacing to legacy Fortran code.
- Experience with a variety of tools and languages, including bash, csh, L^AT_EX, HTML, Javascript, D3js, Git, various database query languages, web templating engines such as Jinja, etc.
- Author of *Pythonic Perambulations*, a popular Python blog covering scientific computing, visualization, statistics, and occasionally whimsical distractions: <http://jakevdp.github.io>

Software

Scikit-Learn I am a member of the core team of [scikit-learn](#), a popular package for performing machine learning in Python. I have contributed in many areas, but most notably routines for efficient 2-point (e.g. nearest neighbors) queries, and algorithms based on these such as k -neighbor classification, kernel density estimation, and manifold learning. I have also presented tutorials on the subject on many occasions, including at the PyCon, SciPy, and PyData conferences.
2010–Present

SciPy I am a maintainer of [SciPy](#), the definitive repository for many scientific computing tools available in Python. My contributions are primarily in the sparse matrix package, including code for efficient solutions of large sparse eigenvalue problems, and for efficient traversal and analysis of large sparse graphs.
2011–Present

Altair I am co-creator of the [Altair](#) project, a declarative statistical visualization library for Python built on the Vega-Lite visualization grammar.
2016–Present

AstroPy I have contributed several components of the suite of statistics tools for the [AstroPy](#) project, a Python package aimed at astronomers. In particular, I wrote the modules for [Bayesian Blocks](#) and the [Lomb-Scargle Periodogram](#).
2014–Present

Others I have created and contributed to many other Python projects, including Matplotlib, IPython, NumPy, Pandas, AstroML, SciDB-Py, Pelican, mpld3, and others. I have also open-sourced much of my research code and teaching materials. More information is available in my [GitHub profile](#).

SELECTED TALKS

**= invited talk

Computing

- **May 2017 **Keynote: The Unexpected Effectiveness of Python in Science**
Python's Visualization Landscape (Submitted Talk)
PyCon 2017, Portland OR
- **April 2017 *The Python Visualization Landscape*
Puget Sound Python Meetup, Seattle WA
- **November 2016 *Declarative Visualization in Python*
Puget Sound Python Meetup, Seattle WA
- **October 2016 *Python and Machine Learning in Astronomy*
TalkPython podcast #81
- **July 2016 *The Python Data Science Stack*
JSM 2016, Chicago IL
- May 2016 *Statistics for Hackers*
PyCon 2016, Portland OR
- **April 2016 *Driving Reproducibility at UW*
NYU Reproducibility Symposium, Brooklyn NY
- **March 2016 *Introduction to Scikit-Learn*
PyData @ Strata, San Jose CA
- **March 2016 *Data Analysis with Pandas*
Fred Hutchinson Cancer Research Institute
- **December 2015 *Statistics for Hackers*
CSSS Seminar, University of Washington
- **December 2015 *An Overview of the Python Scientific Stack*
Python in Geosciences Seminar, University of Washington
- **September 2015 *Statistics for Hackers*
Multithreaded Data Meetup, San Francisco, CA
- July 2015 *Machine Learning with Scikit-Learn* (2 hour tutorial)
Data Science at University of Washington (30 minute talk)
PyData Seattle 2015, Redmond WA
- **July 2015 **Keynote: The State of the Stack**
SciPy 2015, Austin TX
- **June 2015 *Astronomy, Python, and Data Science at UW*
Seattle DAML (Data/Analytics/Machine Learning) meetup
- April 2015 *Losing Your Loops: Fast Numerical Computing with NumPy* (30 minute talk)
Introductory Machine Learning with Scikit-Learn (3 hour tutorial)
PyCon 2015, Montreal, QC
- **December 2014 *Data Science with Python.* (Invited Instructor for 3-day course)
Institute of Health & Society, University of Oslo, Norway
- **October 2014 *Machine Learning with Python*
Practice of Machine Learning Conference, Redmond WA
- **October 2014 *Introduction to NumPy and Matplotlib*
PyData Strata 2014, New York NY
- July 2014 *Frequentism and Bayesianism: What's the Big Deal?*
SciPy 2014, Austin, TX
- **July 2014 *Parallel Computing in Python*
Invited guest lecture, CodeFellows course
- **April 2014 *Python in the Browser Age: Data Exploration in the IPython Notebook*
OpenVisConf 2014, Boston MA

- April 2014 *Exploring Machine Learning with Scikit-learn* (3-hour tutorial)
Diving deeper into Machine Learning with Scikit-learn (3-hour tutorial)
with Olivier Grisel
PyCon 2014, Montreal, QC
- November 2013 *Financial Time-series Data in SciDB*
with Bryan Lewis
PyData NYC 2013
- November 2013 *Efficient Computing with NumPy* (1.5-hour tutorial)
PyData NYC 2013
- November 2013 *Machine Learning with Scikit-Learn* (1.5-hour tutorial)
PyData NYC 2013
- August 2013 *Big Analytics for Python Users Without the Hassles*
with Bryan Lewis, & Travis Oliphant
Webinar presented by Paradigm4
- **July 2013 *Interactive Computing with IPython and ASCOT*
Clawpack Workshop, University of Washington
- July 2013 *An Introduction to Scikit-Learn* (2-part, 8-hour tutorial)
with Gael Varoquaux, & Olivier Grisel
Scipy 2013, Ausin TX
- July 2013 *Introduction to Python* (3-hour tutorial)
Software Carpentry Course, Seattle WA
- April 2013 *Interactive Applications with Matplotlib* (2-hour tutorial)
PyData Silicon Valley, Santa Clara CA
- April 2013 *An Introduction to Scikit-Learn* (3-hour tutorial)
PyCon 2013, Santa Clara CA
- **October 2012 *Scientific Machine Learning with Scikit-learn* (1-hour tutorial)
Interactive Visualization with Matplotlib (1-hour tutorial)
PyData NYC, New York NY
- July 2012 *Machine Learning in Python* (4-hour tutorial)
Scipy 2012, Austin TX
- **March 2012 *Scikit-Learn Tutorial* (1-hour tutorial)
PyData Workshop, Google Campus, Mountain View CA

Astronomy

- **August 2016 *Fundamentals of Frequentist & Bayesian Statistics*
2016 AstroHackWeek, UC Berkeley
- **August 2016 *Git and Github for Driving your Research*
Software Engineering for Astronomers
Introduction to Bayesian Inference
LSSTC Data Science Fellowship Program Summer School, Northwestern University
- **June 2016 *Exploring Galaxy Evolution through Manifold Learning*
Modern Massive Datasets, UC Berkeley
- **April 2016 *Astrostatistics: Opening the Black Box*
Fred Hutchinson Cancer Research Center Data Science Seminar
- **January 2016 *Bayesian Methods in Astronomy*
half-day workshop, 227th AAS meeting
- **December 2015 *Astrostatistics: Opening the Black Box*
Webinar, Inter-institutional Laboratory for e-Astronomy
- **November 2015 *Astrostatistics: Opening the Black Box*
Colloquium, Herzberg Institute of Astrophysics, Victoria, BC
- **November 2015 *Astrostatistics: Opening the Black Box*
Colloquium, Mitchell Institute for Fundamental Physics and Astronomy, College Station, TX

- **June 2015 *Periodograms for Multiband Timeseries* (Invited Talk)
Machine Learning Workshop (Invited instructor)
Local Group Astrostatistics conference, Ann Arbor, MI
- **January 2015 *Code Licensing for Astronomy.* (Invited Short Talk)
225th meeting of the American Astronomical Society, Seattle, WA
- **October 2014 *ESAC Data Analysis and Statistics Workshop.* (Invited Instructor)
European Space Astronomy Center, Madrid, Spain
- January 2014 *AstroML: Python-powered Machine Learning for Astronomy*
Jake VanderPlas, Andrew Connolly, & Zeljko Ivezic
AAS # 223, 2014, Seattle, WA
- **November 2013 *Information Theory and Survey Design*
LBL Cosmology Seminar, Berkeley CA
- October 2013 *LSST and the Time-domain Universe*
Calvin College Physics Seminar, Grand Rapids, MI
- **October 2013 *Unlocking the Universe with Python and LSST*
RuPy conference, Budapest, Hungary
- **August 2013 *Reproducible Astronomy in the LSST Era*
Data Science Seminar, Los Alamos National Labs
- July 2013 *Opening Up Astronomy with Python and AstroML*
Jake VanderPlas, Andrew Connolly, & Zeljko Ivezic
Scipy 2013, Austin TX
- **May 2013 *Information Theory and Survey Design*
UC Davis Cosmology Seminar, Davis CA
- **April 2013 *Observational Tracers of Modified Gravity: Dwarf Disk Galaxies*
Novel Probes of Gravity Workshop, University of Pennsylvania
- October 2012 *AstroML: Machine Learning for Astronomy*
Conference on Intelligent Data Understanding, Boulder CO
- July 2012 *AstroML: Machine Learning for Astronomy*
SciPy Conference, Austin TX
- December 2011 *Processing Shear Maps with Karhunen-Loeve Analysis* (poster)
Jake VanderPlas, Bhuvnesh Jain, & Andrew Connolly
Neuro-Imaging Processing Symposium (NIPS), Granada Spain
- **October 2011 *Alternatives to 2-Point Statistics in Weak Lensing*
DES Collaboration meeting, Philadelphia PA
- **June 2011 *Digital Planetariums for the Masses*
AstroVis, University of Washington
- May 2011 *KL Interpolation of Weak Lensing Shear*
INPA Cosmology Seminar, Lawrence Berkeley National Laboratory, CA
- May 2011 *KL Interpolation of Weak Lensing Shear*
UC Davis Cosmology Seminar, Davis CA
- May 2011 *KL Interpolation of Weak Lensing Shear*
KIPAC Cosmology Seminar, SLAC National Laboratory, CA
- January 2011 *Finding the Odd One Out in Spectroscopic Surveys* (poster)
A. Connolly, S. Daniel, L. Xiong, J. VanderPlas, & J. Schneider
217th AAS meeting, Seattle WA
- January 2011 *3D Reconstruction of the Density Field* (poster)
Jake VanderPlas & Andrew Connolly
217th AAS meeting, Seattle WA
- July 2010 *A New Approach to Tomographic Mapping*
Ten Years of Cosmic Shear, Edinburgh, UK

**November 2007 *SALT-2 Light-curve Fitting for SDSS Supernovae*
SDSS Collaboration Meeting, Fermi National Accelerator Laboratory

Academia & Policy

**December 2016 *Promoting Open Science in the University*
Invited Panel, Open Science and Big Data session at IEEE 2016, Washington DC

**October 2014 *Stemming the Data Science Brain Drain*
Invited talk, TTI/Vanguard *Reprogramming Programming*, Washington DC

August 2014 *Hacking Academia from Inside and Out*
Breakout session, O'Reilly SciFOO, Mountain View CA

General Interest

**June 2013 *The Science of Time Travel*
at the event *Short Films, Big Ideas: The Science of Science Fiction*
Seattle International Film Festival, Seattle WA

**March 2012 *Dark Matter, Dark Energy, and the Fate of the Universe*
Calvin College Physics Colloquium, Grand Rapids MI

**November 2011 *Kinect/WorldWide Telescope Demonstration*
Supercomputing 2011, Seattle WA

**November 2011 *WorldWide Telescope Demonstration*
Partners in Learning Global Forum, Washington DC

**November 2011 *Gravity: A Lens to the Universe*
KCTS9 Queen Anne Science Cafe, Seattle WA

**October 2011 *WorldWide Telescope Demonstration*
Popular Mechanics Breakthrough Awards, New York NY

**March 2011 *Understanding the Dark Side of the Universe*
Pacific Science Center's "Science with a Twist", Seattle WA

**February 2011 *Interconnection in Art and Cosmology*
at the *Traces of the Universe* Art show,
University of Washington, Seattle WA

May 2009 *Dark Matter, Gravitational Lensing, and Cosmology*
Battle Point Astronomical Society, Bainbridge Island, WA

PUBLICATIONS

Books

- [1] J. VanderPlas. *The Python Data Science Handbook*. O'Reilly Media, 2016
- [2] J. VanderPlas. *A Whirlwind Tour of Python*. O'Reilly Media, 2016
- [3] Z. Ivezić, A. Connolly, J. VanderPlas & A. Gray.
Statistics, Data Mining and Machine Learning in Astronomy. Princeton University Press, 2014
(Winner of the IAA's *Outstanding Publication in Astrostatistics* award)

Articles

- [4] F. Rodríguez-Sánchez, B. Marwick, E. Lazowska, J. VanderPlas *Academia's failure to retain data scientists* Science 355.6323:357, 2017
- [5] Parmita Mehta *et al.* *Comparative Evaluation of Big-Data Systems on Scientific Image Analytics Workloads* arXiv:1612.02485, 2016
- [6] J. Ford & J. VanderPlas *Cluster-lensing: A Python Package for Galaxy Clusters & Miscentering* Astronomical Journal, 152:228, 2016
- [7] J. McQueen, M. Meila, J. VanderPlas, Z. Zhang *megaman: Manifold Learning with Millions of points* Journal of Machine Learning Research 17(148):1-5, 2016
- [8] Ariel Rokem *et al.* *Building an Urban Data Science Summer Program at the University of Washington eScience Institute*. Proceedings of the Data For Good Exchange 2015

-
- [9] J. VanderPlas & Z. Ivezić. *Periodograms for Multiband Astronomical Time Series*. ApJ 812:1, 2015
- [10] J. VanderPlas. *Frequentism and Bayesianism: A Python-driven Primer*. Proceedings of the 13th Python in Science (SciPy) Conference, 2014
- [11] J. VanderPlas. *Visualizing Four-Dimensional Asteroids*. Scientific American Magazine, Oct. 2014
- [12] A.J. Connolly *et al.* *An end-to-end simulation framework for the Large Synoptic Survey Telescope*. SPIE 9150:14, 2014
- [13] J. VanderPlas *et al.* *Squeezing a Big Orange into Little Boxes: The AscotDB System for Parallel Processing of Data on a Sphere* IEEE Data Engineering Bulletin 36(4): 11-20 (2013)
- [14] Emad Soroush *et al.* *A Demonstration of Iterative Parallel Array Processing in Support of Telescope Image Analysis*. PVLDB 6(12): 1322-1325 (2013)
- [15] L. Buitinck *et al.* *API design for machine learning software: experiences from the scikit-learn project* European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases (2013)
- [16] L. Palaversa *et al.* *Exploring the Variable Sky with LINEAR. III. Classification of Periodic Light Curves* AJ 146:101, 2013
- [17] V. Vikram, A. Cabré, B. Jain, & J. VanderPlas. *Astrophysical tests of modified gravity: the morphology and kinematics of dwarf galaxies* JCAP 08:20, 2013
- [18] J. VanderPlas, A. Connolly, Z. Ivezić, & A. Gray. *Introduction to AstroML: Machine Learning for Astrophysics*. Proc. of the CIDU, 2012 (**Recipient of the CIDU 2012 Best Paper award**)
- [19] J. VanderPlas, A. Connolly, B. Jain, & M. Jarvis. *Interpolating Masked Weak Lensing Signals with Karhunen-Loeve Analysis*. ApJ 744:180, 2012
- [20] S. Daniel, A. Connolly, A.J. J. Schneider, J. VanderPlas, & L. Xiong *Classification of Stellar Spectra with LLE*. AJ 142:203, 2011
- [21] F. Pedregosa *et al.* *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12:2825, 2011
- [22] B. Jain & J. VanderPlas. *Tests of Modified Gravity with Dwarf Galaxies*. JCAP 10:32, 2011
- [23] J. VanderPlas, A. Connolly, B. Jain, & M. Jarvis. *3D Reconstruction of the Density Field: An SVD Approach to Weak Lensing Tomography*. ApJ 727:118, 2011
- [24] L. Xiong, B. Poczos, J. Schneider, A. Connolly, & J. VanderPlas. *Hierarchical Probabilistic Models for Group Anomaly Detection*. Artificial Intelligence and Statistics (AISTATS), 2011
- [25] H. Lampeitl *et al.* *First-year Sloan Digital Sky Survey-II supernova results: consistency and constraints with other intermediate-redshift data sets*. MNRAS 401:2331, 2010
- [26] LSST Science Collaboration *LSST Science Book, Version 2.0*, arXiv:0912.0201, 2010
- [27] R. Kessler *et al.* *First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters*. ApJS 185:32, 2009
- [28] J. VanderPlas & A. Connolly. *Reducing the Dimensionality of Data: Locally Linear Embedding of Sloan Galaxy Spectra*. AJ 138:1365, 2009
- [29] J. Sollerman *et al.* *First-Year Sloan Digital Sky Survey-II (SDSS-II) Supernova Results: Constraints on Nonstandard Cosmological Models*. ApJ 703:1374, 2009
- [30] R. Kessler *et al.* *SNANA: A Public Software Package for Supernova Analysis*. PASP 121:1028, 2009