Jake VanderPlas

Director of Research – Physical Sciences

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 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 2006-2012
University of Washington, Seattle, WA, advised by Andrew Connolly and Bhuvnesh Jain
Thesis: Karhunen-Loeve Analysis for Weak Gravitational Lensing

MS 2006-2007
University of Washington, Seattle, WA, advised by Craig Hogan and Andrew Becker

BS 1999-2003
Calvin College, Grand Rapids MI
Major: Physics; Minors: Mathematics & Japanese
4-year letterman & 1-year captain of the varsity swim team

Experience

Employment

UW eScience 2014–Present
Senior Data Science Fellow & Director of Research for Physical Sciences
eScience Institute, University of Washington.

UW Computer Science 2013–2014
NSF post-doctoral fellowship, CI-TraCS program.
Department of Computer Science, University of Washington. Supervised by Magda Balazinska

UW Astronomy 2012–2013
Postdoctoral Researcher, LSST Image Simulation group.
Department of Astronomy, University of Washington. Supervised by Andrew Connolly

UW Planetarium 2010–2012
WorldWide Telescope Planetarium Project Coordinator
University of Washington Planetarium, Seattle WA & Microsoft Research, Redmond WA

UW Planetarium 2008–2010
K-12 and Community Outreach Coordinator
University of Washington Planetarium, Seattle WA

Mount Hermon 2004–2006
Experiential Science Educator (4th-8th grade students)
Mount Hermon Outdoor Science School, Santa Cruz CA

Summit Adventure 2004–2005
Wilderness Instructor: Backpacking, Rock Climbing, and Mountaineering
Summit Adventure, Bass Lake CA

Japan ESL 2003–2004
Teacher and Tutor of English as a second language
Sendai, Japan

Volunteering

Neighborhood Advocacy (2013–Present)
As co-chair of the North Delridge Neighborhood Council, I facilitate 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 inner-city 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

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

Astr 102 (Fall 2008)
Introductory Astronomy for Science Majors
head teaching assistant – University of Washington

Astr 150 (Winter 2008)
Planetary Astronomy
teaching assistant – University of Washington

Astr 101 (Fall 2007)
Introductory Astronomy
teaching assistant – University of Washington

Service

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.

AstroData 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 (2013-2014)
Member of the tutorial review committee for PyCon 2013 & 2014.

PyData (2012-2014)
Member of the talk & tutorial review committee for several PyData conferences.

Students Mentored

SungWon Kwak (2012–2013)
Undergraduate, University of Washington Astronomy
Superimposed High Redshift Spectra

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

AWARDS & HONORS

Data Science 2014-Present — *Senior Data Science Fellow* at the UW eScience Institute

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

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

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

Calvin College 1999-2003 — 4-year recipient of the Calvin College Presidential Scholarship.

Calvin College 2000-2001 — Recipient of the Roger D. Griffioen Physics Scholarship.

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, LaTeX, 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 2010–Present — I 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.

SciPy 2011–Present — 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.

AstroML 2012–Present — I am the primary author of AstroML, a Python package devoted to Machine Learning in Astronomy and Astrophysics. Drawing from tools available in SciPy, Scikit-Learn, Matplotlib, and other packages, it provides additional astronomy-specific data analysis routines, loaders for open astronomical datasets, and over 200 examples of data mining, machine learning, and visualization in Astronomy.

SciDB-Py 2013–Present — I created SciDB-py, a Python wrapper of the SciDB database system aimed at efficient distributed array-based computation. This project was in conjunction with engineers at Paradigm4 and at ContinuumIO.

mpld3 2013–Present — I am the author of the mpld3 package, a Python module which converts maplotlib images into interactive D3js visualizations suitable for web publication.

Others — I have made contributions to many other Python projects, including Matplotlib, IPython, NumPy, Pandas, Pelican, and others. I have also open-sourced much of my research code and teaching materials.

SELECTED TALKS

***= invited talk*

Astronomy

**June 2015 *Periodograms for Multiple 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

July 2014 *Frequentism and Bayesianism: What's the Big Deal?*
SciPy 2014, Austin, TX

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

Computing

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 *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

Academia & Policy

**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 Audience

**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] Z. Ivezic, A. Connolly, J. VanderPlas & A. Gray. *Statistics, Data Mining and Machine Learning in Astronomy.* Princeton Univertiy Press, 2014

Articles

[2] J. VanderPlas & Z. Ivezic. *Periodograms for Multiband Astronomical Time Series.* ApJ in prep, 2015

[3] J. VanderPlas. *Frequentism and Bayesianism: A Python-driven Primer.* Proceedings of the 13th Python in Science (SciPy) Conference, 2014

[4] J. VanderPlas. *Visualizing Four-Dimensional Asteroids.* Scientific American Magazine, Oct. 2014

[5] A.J. Connolly *et al. An end-to-end simulation framework for the Large Synoptic Survey Telescope.* SPIE 9150:14, 2014

[6] 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)

[7] Emad Soroush *et al. A Demonstration of Iterative Parallel Array Processing in Support of Telescope Image Analysis.* PVLDB 6(12): 1322-1325 (2013)

[8] 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)

[9] L. Palaversa *et al. Exploring the Variable Sky with LINEAR. III. Classification of Periodic Light Curves* AJ 146:101, 2013

[10] V. Vikram, A. Cabré, B. Jain, & J. VanderPlas. *Astrophysical tests of modified gravity: the morphology and kinematics of dwarf galaxies* JCAP 08:20, 2013

[11] J. VanderPlas, A. Connolly, Z. Ivezic, & A. Gray. *Introduction to AstroML: Machine Learning for Astrophysics.* Proc. of the CIDU, 2012 **(Recipient of the CIDU 2012 Best Paper award)**

[12] J. VanderPlas, A. Connolly, B. Jain, & M. Jarvis. *Interpolating Masked Weak Lensing Signals with Karhunen-Loeve Analysis.* ApJ 744:180, 2012

[13] S. Daniel, A. Connolly, A.J. J. Schneider, J. VanderPlas, & L. Xiong *Classification of Stellar Spectra with LLE.* AJ 142:203, 2011

[14] F. Pedregosa *et al. Scikit-learn: Machine learning in Python.* Journal of Machine Learning Research, 12:2825, 2011

[15] B. Jain & J. VanderPlas. *Tests of Modified Gravity with Dwarf Galaxies.* JCAP 10:32, 2011

[16] 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

[17] L. Xiong, B. Poczos, J. Schneider, A. Connolly, & J. VanderPlas. *Hierarchical Probabilistic Models for Group Anomaly Detection.* Artificial Intelligence and Statistics (AISTATS), 2011

[18] 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

[19] LSST Science Collaboration *LSST Science Book, Version 2.0*, arXiv:0912.0201, 2010

[20] R. Kessler *et al. First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters.* ApJS 185:32, 2009

[21] J. VanderPlas & A. Connolly. *Reducing the Dimensionality of Data: Locally Linear Embedding of Sloan Galaxy Spectra.* AJ 138:1365, 2009

[22] J. Sollerman *et al. First-Year Sloan Digital Sky Survey-II (SDSS-II) Supernova Results: Constraints on Nonstandard Cosmological Models.* ApJ 703:1374, 2009

[23] R. Kessler *et al. SNANA: A Public Software Package for Supernova Analysis.* PASP 121:1028, 2009