

Data Science in Manufacturing

James A. Bednar



Education: UT Austin



TEXAS
The University of Texas at Austin

1993 BS Electrical Engineering

1994 BA Philosophy

2002 PhD Computer Science

2004 Postdoc Computational Neuroscience

Seems crazy, but always focused on building systems as cool as the brain.

(Also partly hedging my bets, in case science didn't work out!)

Faculty: U. Edinburgh Informatics



THE UNIVERSITY
of EDINBURGH

2004 Lecturer

2012 Senior Lecturer

2014 Reader

Academia cares about grants, papers, students, and impact:

- Director, Doctoral Training Centre in Neuroinformatics (\$8M)
- Published 50 papers, 1 book, 2 edited books, hundreds of posters
- Supervised 10 PhD, 19 MSc, 15 UG theses
- Also built open-source Python tools for computational neuroscience (*Param*, *Topographica*, *ImaGen*, *HoloViews*)
- Only 20% of PhDs stay in science.
Regularly asked to talk about industry jobs, but had no idea!



Industry: Anaconda, Inc

2015 Solutions Architect

2019 Manager of Technical Services

2021 Director of Custom Services

Industry cares about money, solutions, and value:

- \$10M or so in funding so far from government and corporate sources
- All our work is based on open source software (OSS)
- Money goes to salaries for engineers improving OSS and applying OSS to solve problems
- Hired three of my ten PhD students (and still have two)
- Together we help improve tools for data science, while working on interesting problems that keep my team motivated

Transition to industry

- My wife wanted to see the sun again and have a back yard
 - And have a plausible retirement plan
 - Couldn't find a faculty position that fit the bill
 - Knew about Anaconda and Enthought from years of SciPy Austin
 - Interviewed at both, got the job at Anaconda, never looked back!
-
- All that time writing [low-impact papers](#) was fairly useless
 - Hundreds of thousands of people download our software every month
 - Becoming critical infrastructure for climate science
(due to large, remote datasets involved)
 - Working closely with clients to apply data science to manufacturing
 - Now also coming back to brain science, funded by Zuckerberg



HoloViz.org provides a set of compatible tools to make it easier to see, understand, and communicate your data, at every stage from discovery to publication.



Panel



hvPlot



HoloViews



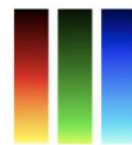
GeoViews



Datashader



Param



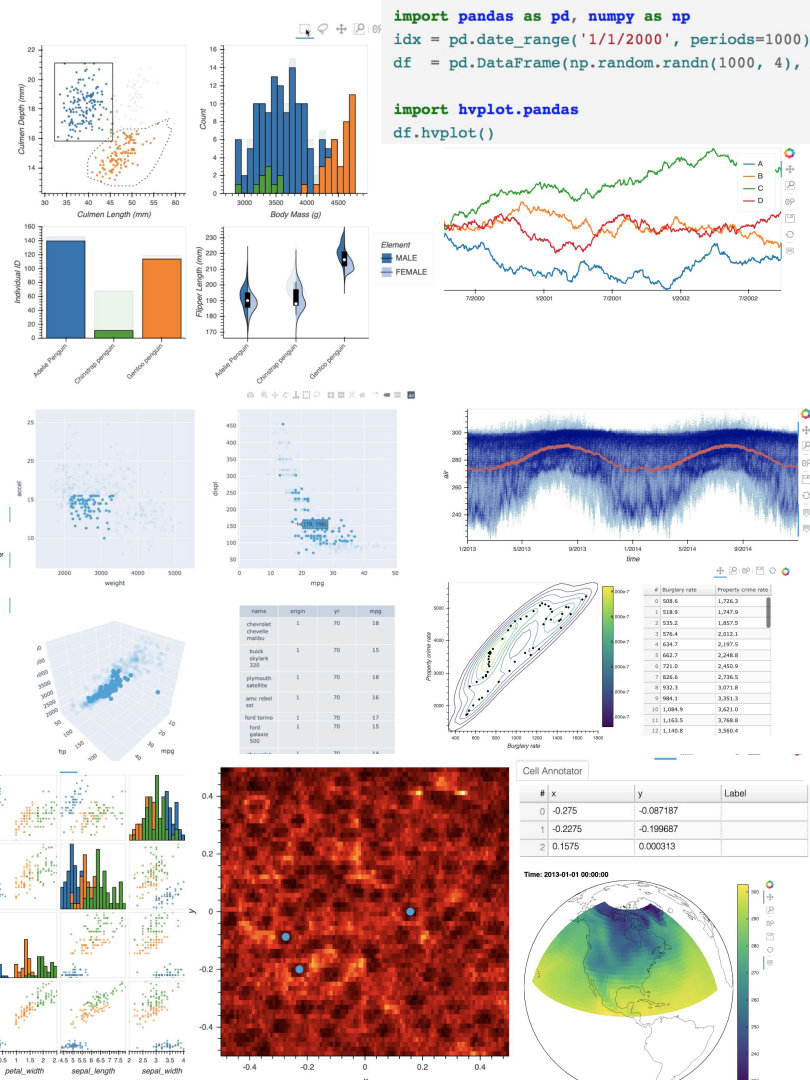
Colorcet



High-level viz/analysis for Bokeh/MPL/ Plotly

hvPlot: Interactive plots from Pandas
.plot()

- Describe your data once, then get plotting for free
- Developed for our academic work
- Now used to build dashboards for corporate clients

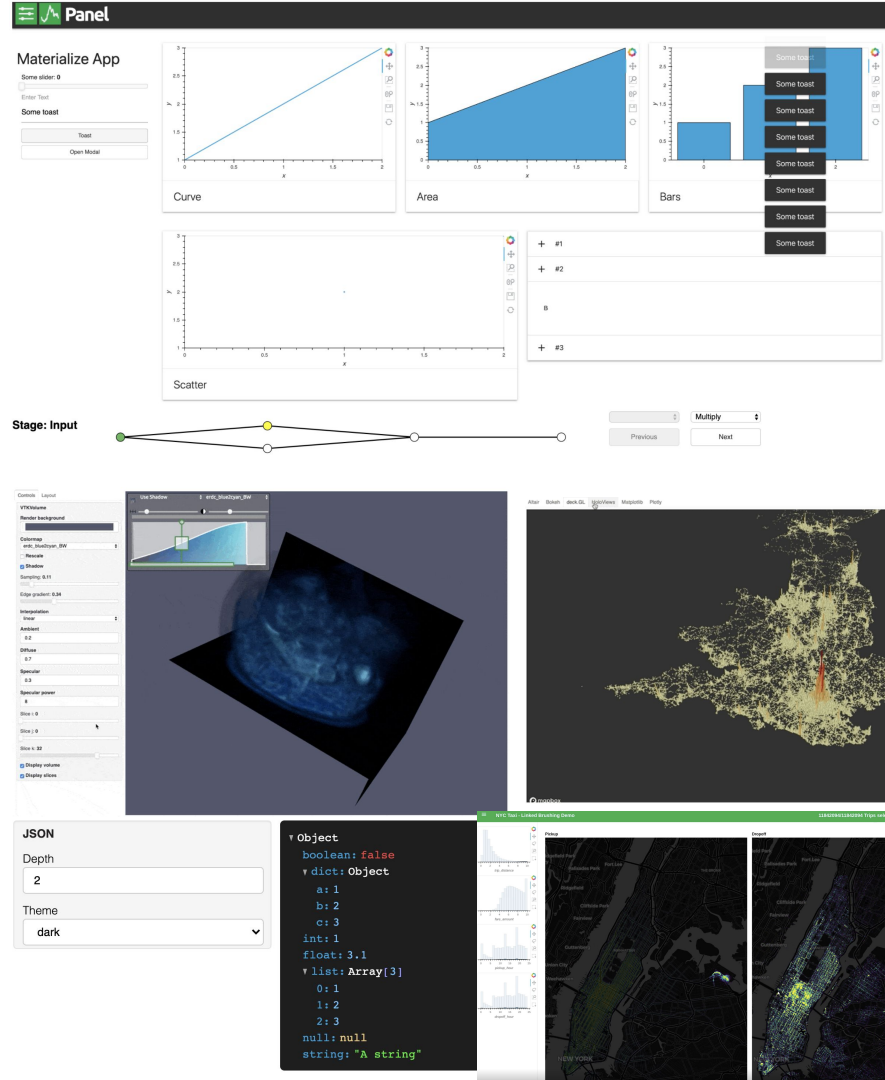




Panel

Easy apps and dashboards for any viz library.

- Fully reactive, Pythonic API for Jupyter or standalone use
- Makes it simple to add widgets to control any plot or table
- Supports nearly all plotting libraries
- Provides fully responsive multi-page apps

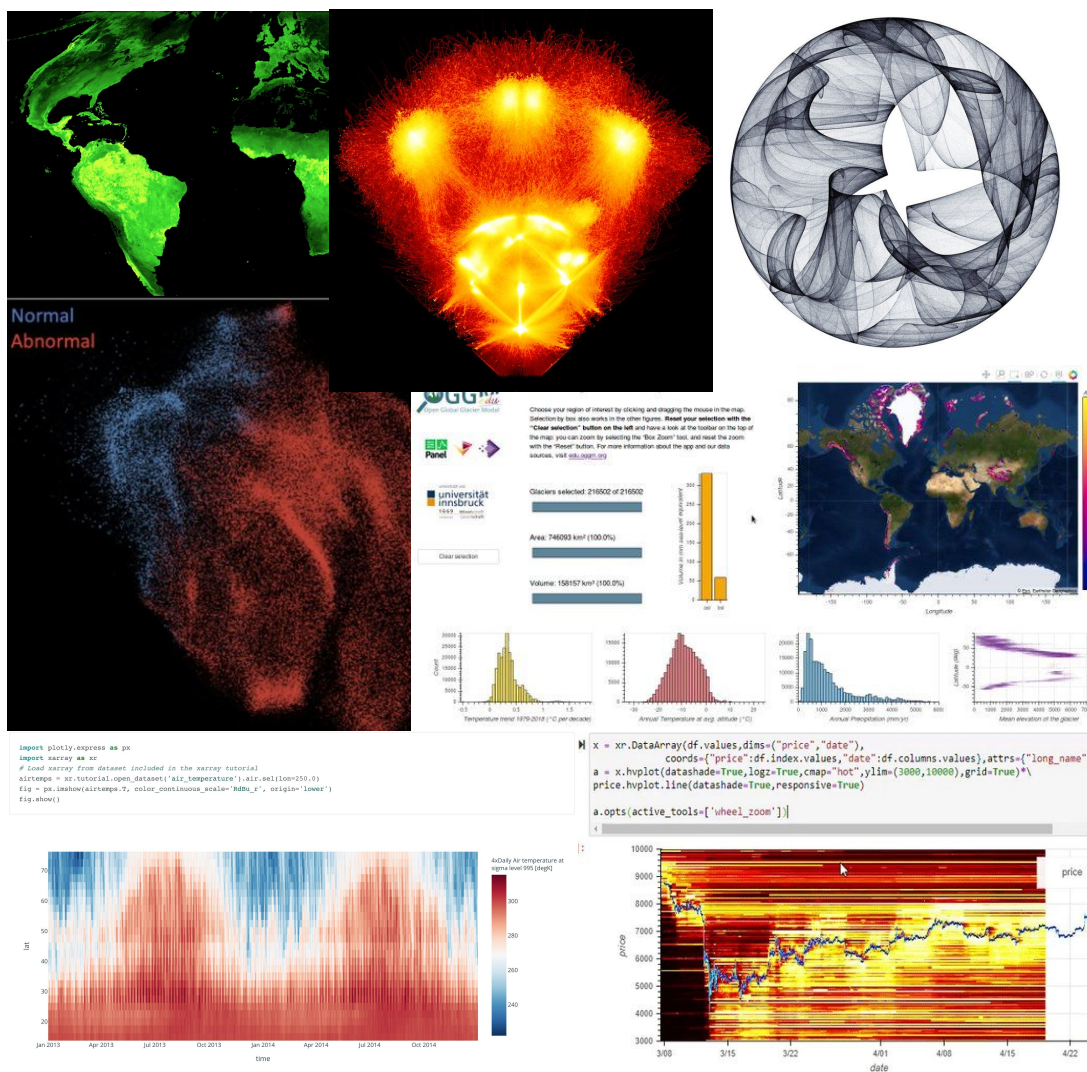




Datashader

*Accurate server-side rendering
of the largest datasets*

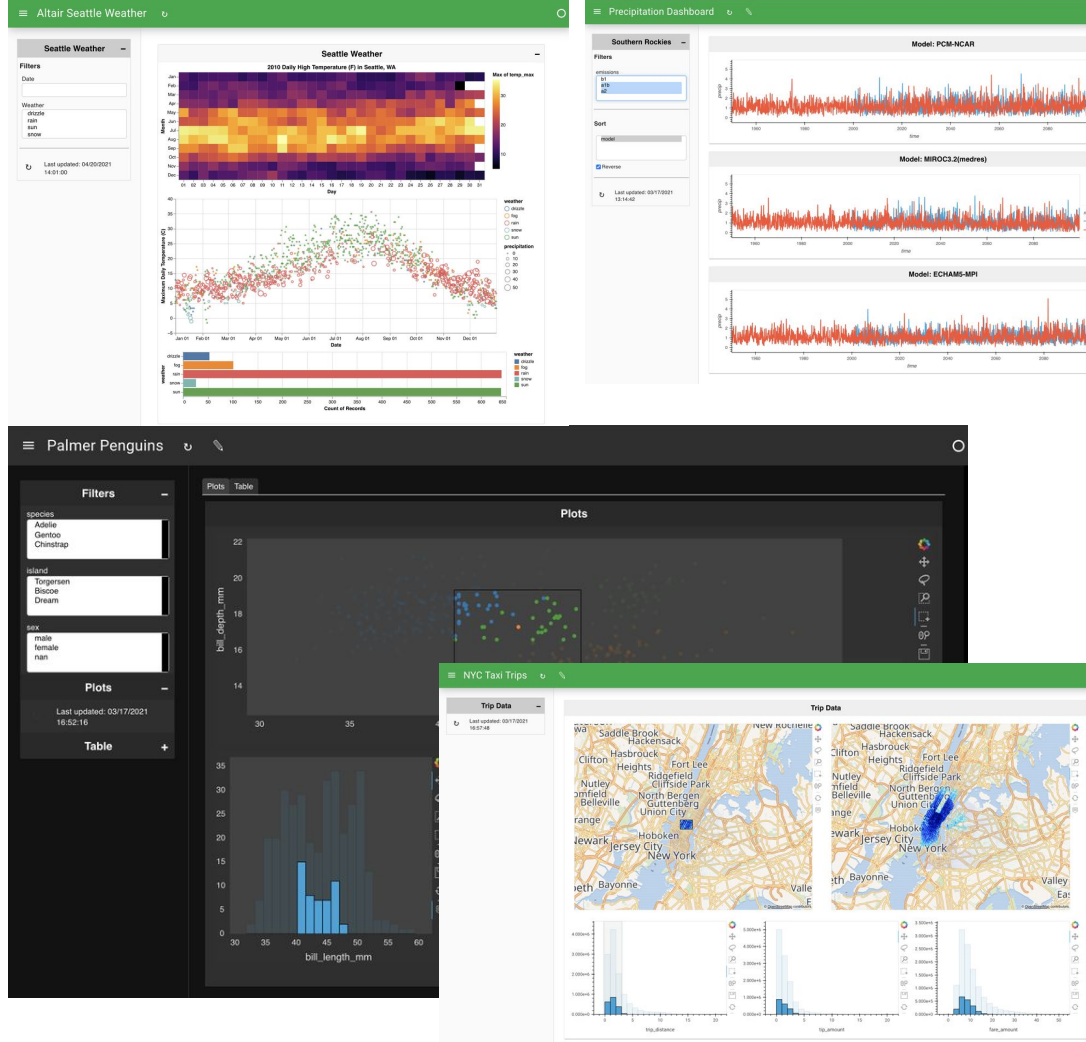
- Renders data as arrays or images
- Makes working with “big data” practical on a laptop
- Scales up to large and/or remote datasets



Lumen

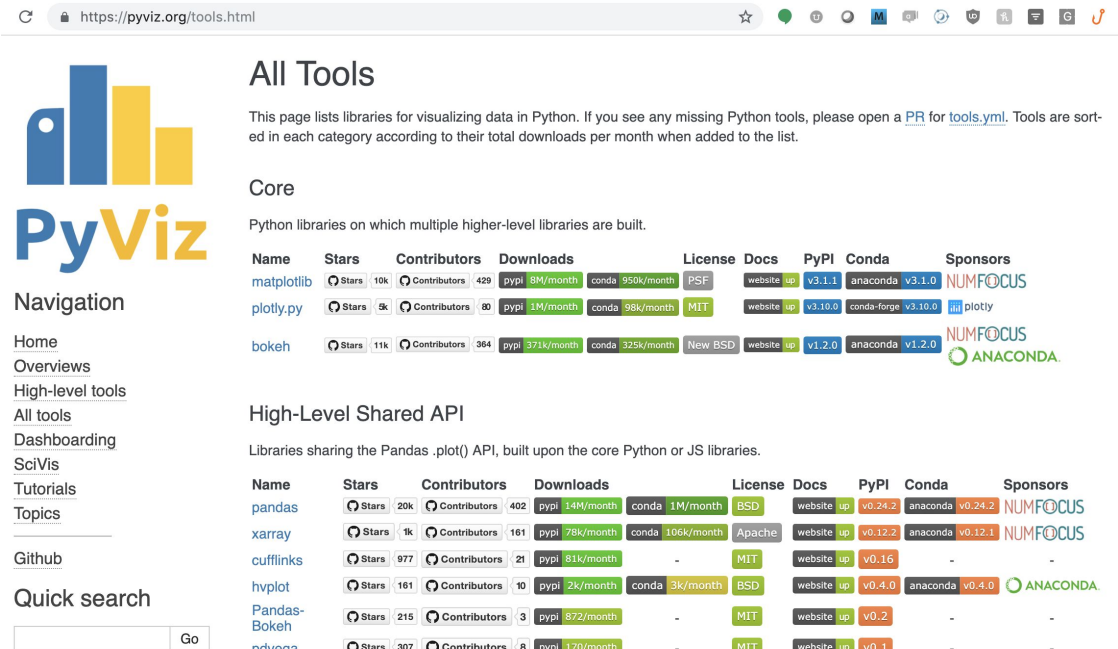
*Low-code or no-code
dashboarding and plotting*

- Declarative .yaml specification for data sources, transforms, filters, and views
- GUI Builder tool
- Extensible in Python
- The power of HoloViz, for everyone!



Also see *PyViz.org*

- An open, non-partisan guide to Python Viz libraries.
- Live status of 90+ tools
- Overviews, comparisons, tutorials



https://pyviz.org/tools.html

All Tools

This page lists libraries for visualizing data in Python. If you see any missing Python tools, please open a [PR](#) for [tools.yml](#). Tools are sorted in each category according to their total downloads per month when added to the list.

Core

Python libraries on which multiple higher-level libraries are built.

Name	Stars	Contributors	Downloads	License	Docs	PyPI	Conda	Sponsors
matplotlib	Stars 10k	Contributors 429	pypi 8M/month conda 950k/month	PSF	website up	v3.1.1	anaconda v3.1.0	NUMFOCUS
plotly.py	Stars 5k	Contributors 80	pypi 1M/month conda 98k/month	MIT	website up	v3.10.0	conda-forge v3.10.0	plotly
bokeh	Stars 11k	Contributors 364	pypi 371k/month conda 325k/month	New BSD	website up	v1.2.0	anaconda v1.2.0	NUMFOCUS ANACONDA

High-Level Shared API

Libraries sharing the Pandas .plot() API, built upon the core Python or JS libraries.

Name	Stars	Contributors	Downloads	License	Docs	PyPI	Conda	Sponsors
pandas	Stars 20k	Contributors 402	pypi 14M/month conda 1M/month	BSD	website up	v0.24.2	anaconda v0.24.2	NUMFOCUS
xarray	Stars 1k	Contributors 161	pypi 78k/month conda 106k/month	Apache	website up	v0.12.2	anaconda v0.12.1	NUMFOCUS
cufflinks	Stars 977	Contributors 21	pypi 81k/month	-	MIT	website up	v0.16	-
hvplot	Stars 161	Contributors 10	pypi 2k/month conda 3k/month	BSD	website up	v0.4.0	anaconda v0.4.0	ANACONDA
Pandas-Bokeh	Stars 215	Contributors 3	pypi 872/month	-	MIT	website up	v0.2	-
pdevega	Stars 307	Contributors 8	pypi 170/month	-	MIT	website up	v0.1	-

Demos

- [Introduction](#): How to make plots
- [Interactive](#): How to make apps and dashboards
- examples.pyviz.org: In-depth examples
- [HoloViz.org](https://holoviz.org): In-depth tutorial
- Also see [hvplot.HoloViz.org](https://hvplot.holoviz.org), or contact me at:

jbednar@anaconda.com



@JamesABednar



jbednar