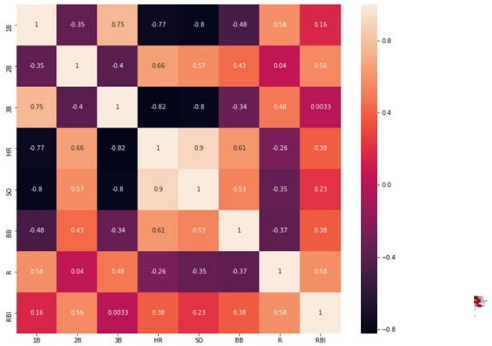
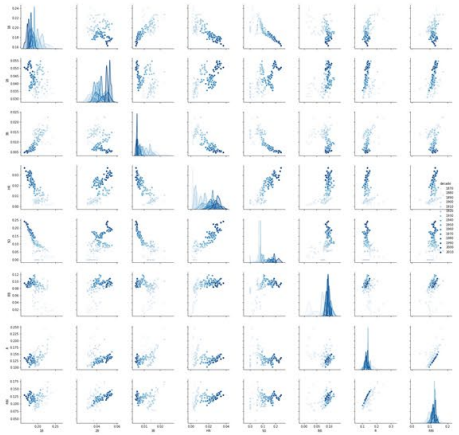
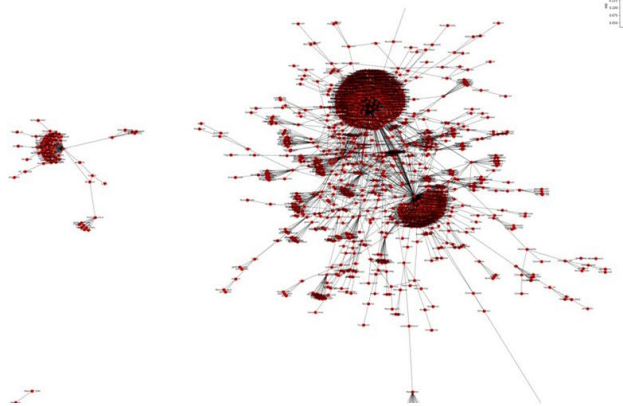
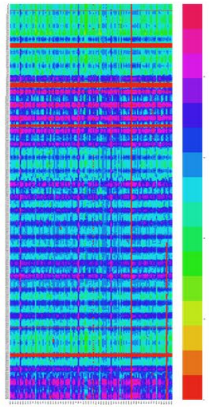
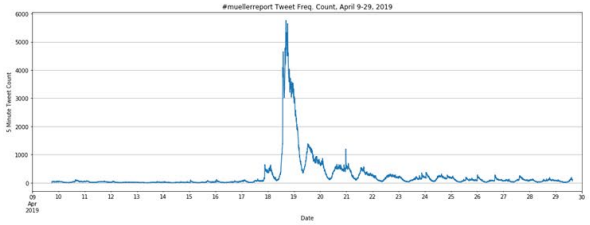
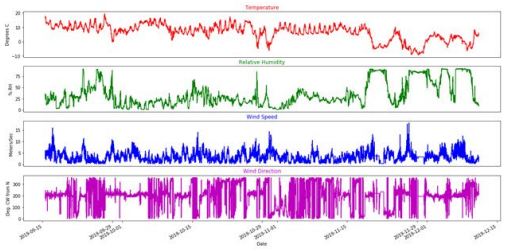












Data Visualization With Python Using Jupyter Notebooks

Jeff Sale

SDSC Learning Design Technologist
XSEDE ECSS Visualization Consultant
XSEDE Workforce Development

https://github.com/sdsc-hpc-training-org/data_vis_with_python

Acknowledgements

- Chris Myers*, Senior Research Associate, Center for Advanced Computing, Adjunct Professor, Dept. of Physics, Cornell University
- Susan Mehringer*, Associate Director, Consulting, Cornell University Center for Advanced Computing, & XSEDE L3 for Training, WBS 2.1.2
- Kate Starbird, Associate Professor, Human Centered Design & Eng., U. of Washington
- Jen Zeimke, Associate Professor, Political Science, John Carroll Univ.
- Mary Thomas and Bob Sinkovits, SDSC User Training

***Data Science With Python Tutorial**

Cornell Virtual Workshop, https://cvw.cac.cornell.edu/

XSEDE Tutorial

Python for Data Science

Cornell Virtual Workshop

Part 1: https://cvw.cac.cornell.edu/PyDataSci1/

Part 2: https://cvw.cac.cornell.edu/PyDataSci2/

Commands to Get Started

Install Bokeh and Holoviews

conda install -c pyviz holoviews bokeh

Install Datashader

conda install datashader

Clone the Holoviews and DataShader Repos

git clone https://github.com/holoviz/holoviews.git

git clone https://github.com/holoviz/datashader.git

Commands to Get Started

Download the Data for this Webinar

wget https://education.sdsc.edu/hpc_training_series_2021/python_vis_webinar_data.zip

Run the shell script to reserve an Expanse compute node and launch a Jupyter Notebook

/cm/shared/apps/sdsc/galyleo/galyleo.sh launch -j notebook -A abc123 -p compute -n 1 -M 8 -t 00:30:00 --conda-env base

The Jupyter Project

“Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.”

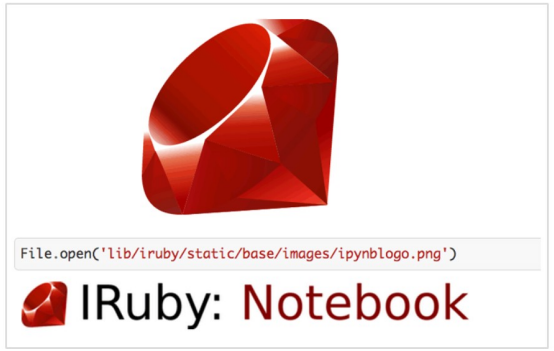
https://jupyter.org/

The Jupyter Menagerie

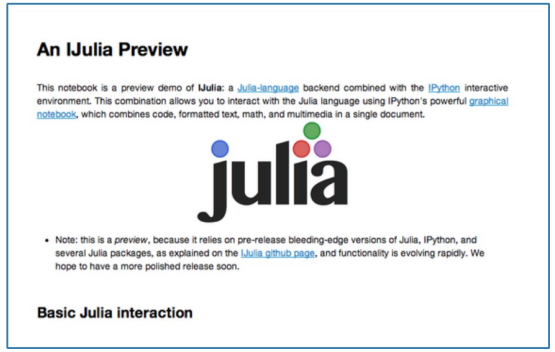
Notebook Widgets

IPython

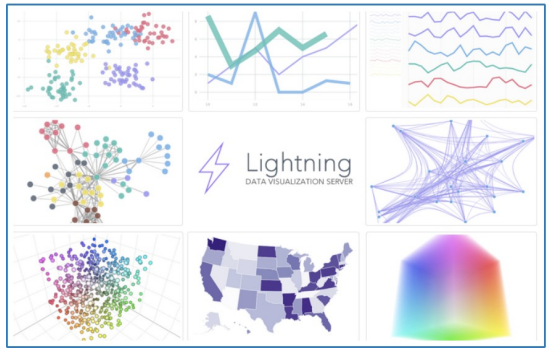
IRuby



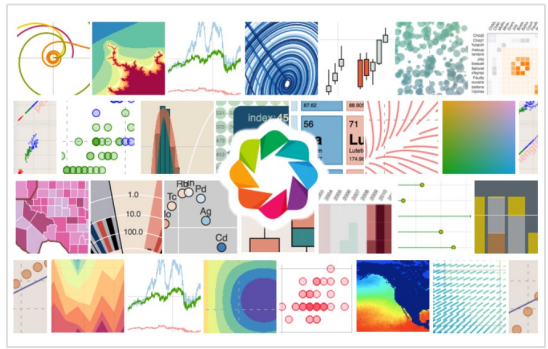
IJulia



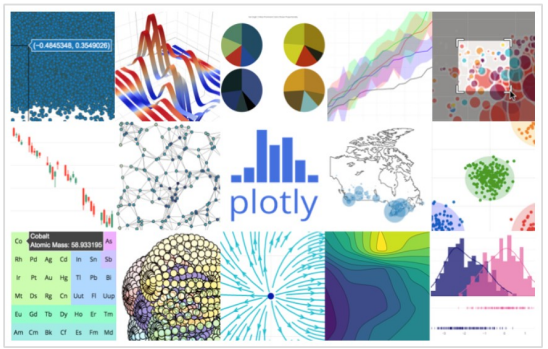
Data Visualization with Lightning

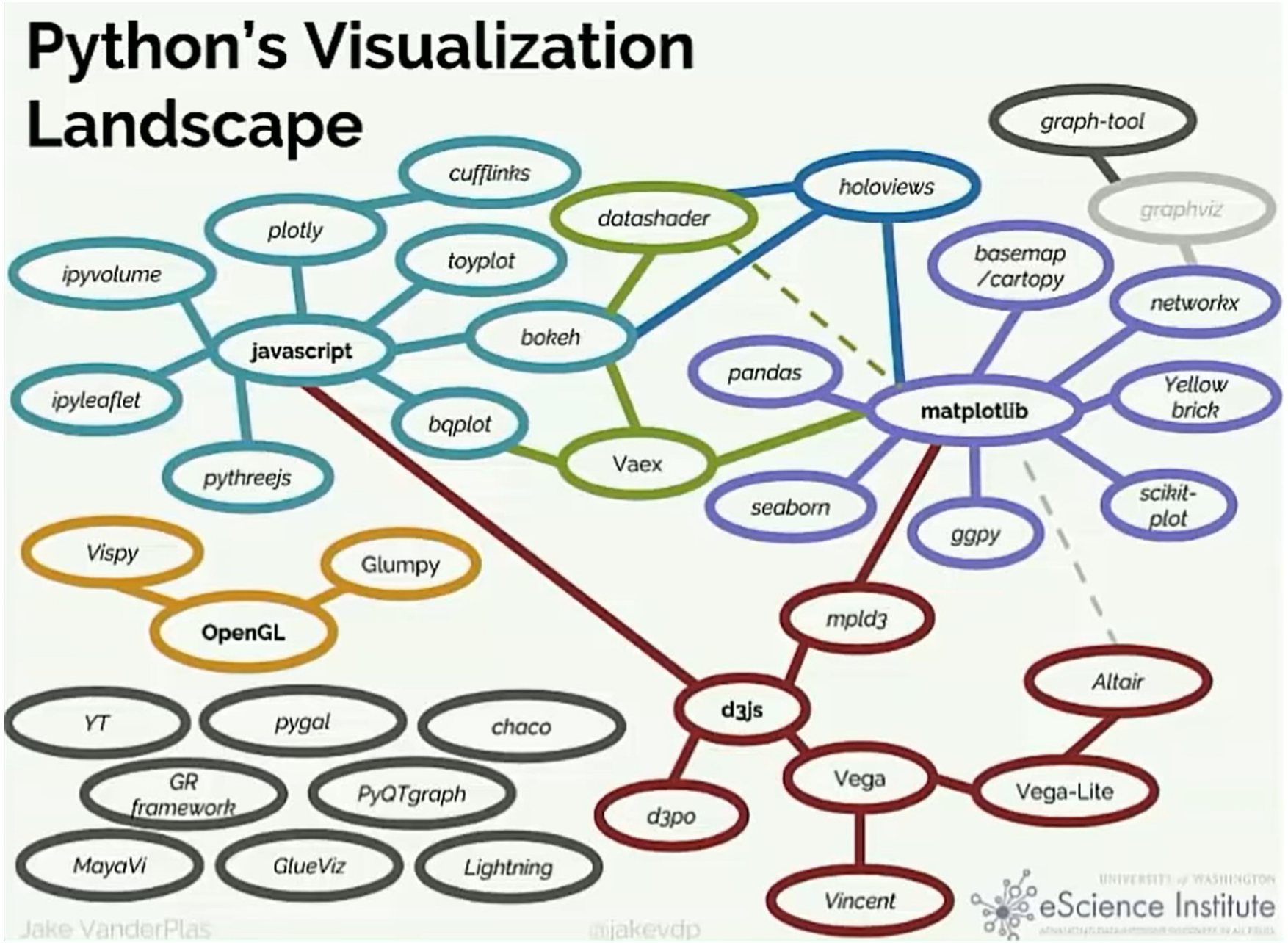


Interactive data visualization with Bokeh



Interactive plots with Plotly





Courtesy: Jake VanderPlas The Python Visualization Landscape PyCon 2017, https://www.youtube.com/watch?time_continue=134&v=eVET9IYgbao&feature=emb_logo

Bokeh

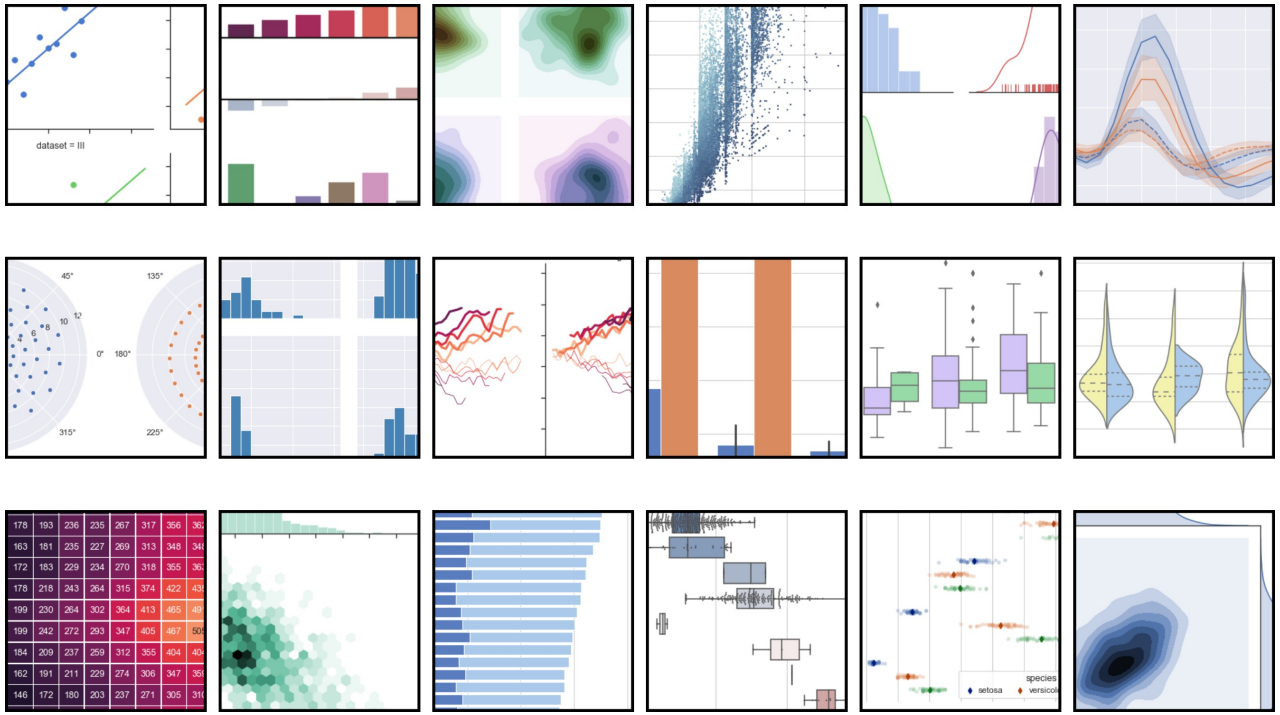
“Bokeh is an interactive visualization library for modern web browsers. It serves as a web-based front end to matplotlib using JavaScript behind the scenes to run in a browser”.

https://docs.bokeh.org/en/latest/index.html

Seaborn

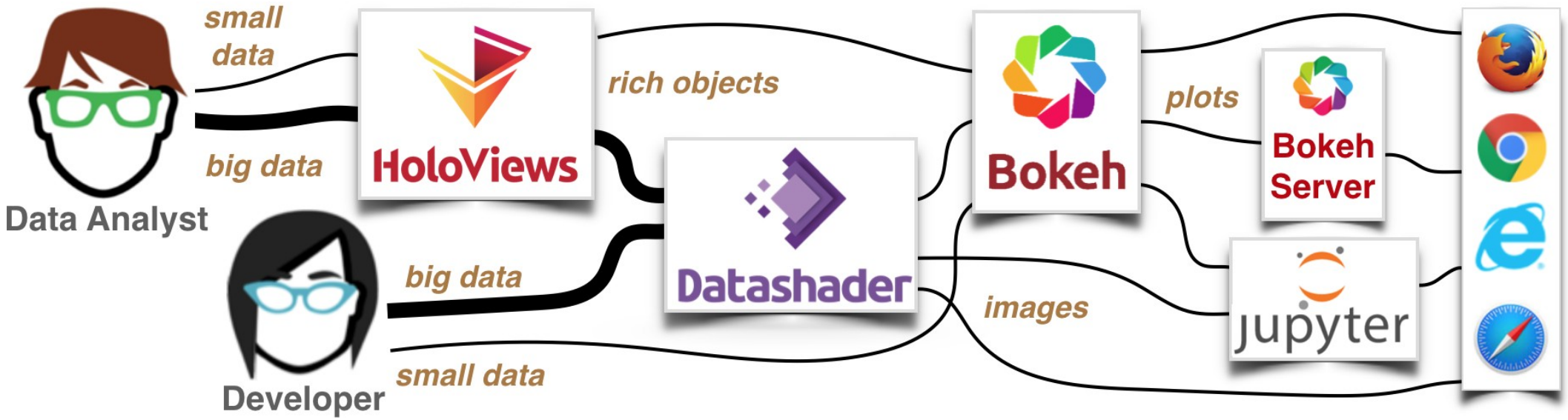
Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

https://seaborn.pydata.org/



HoloViz

Panel hvPlot HoloViews GeoViews Datashader Param Colorcet



Ways to Access Jupyter on XSEDE Systems

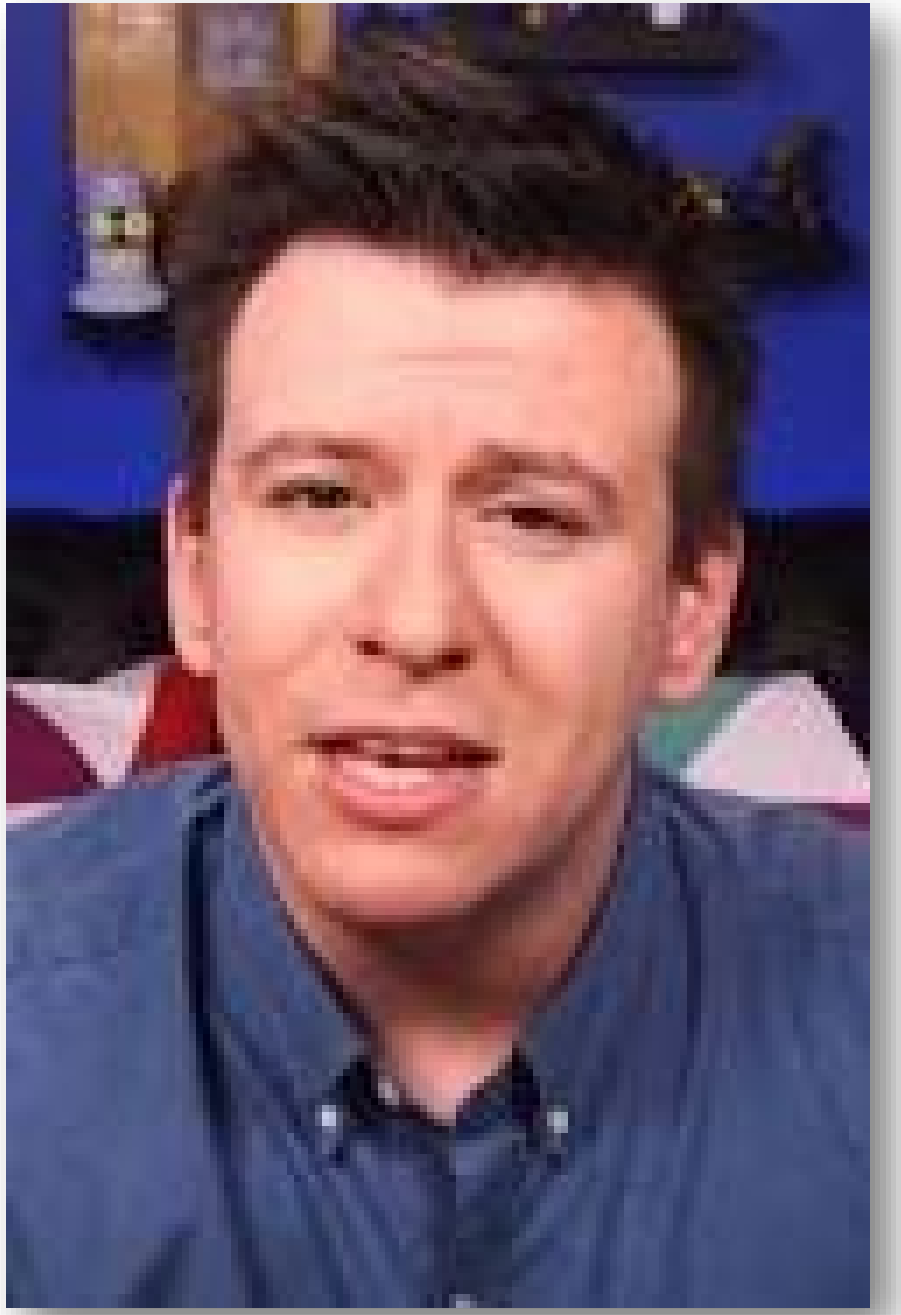
- TACC Visualization Portal
  - https://vis.tacc.utexas.edu/
- SDSC Expanse*
  - https://education.sdsc.edu/training/interactive/202012_running_jupyter_notebooks_on_expanse/index.html
  - https://hpc-training.sdsc.edu/notebooks-101/notebook-101.html
- PSC Bridges 2
  - https://www.psc.edu/user-resources/software/jupyter

*Run new script, galyleo.sh, using this command:

/cm/shared/apps/sdsc/galyleo/galyleo.sh launch -j notebook -A abc123 -p compute -n 1 -M 8 -t 00:30:00 --conda-env base

Visualization Tutorials

- Scientific Visualization with VisIt, Amit Chourasia, SDSC Director of Visualization Services
  - http://users.sdsc.edu/~amit/scivis-tutorial/

- Hyperglyph Visualization, Jeff Sale
  - https://www.iluvdata.org/antz/toroids/tutorials/intro_lessons/index.html

Let's just jump into it!