

If you want to follow along with the Python code,
please go here and try the “Launch into JupyterHub” link:
<https://github.com/benjum/oarc-python-data-viz-1>

Data Visualization with Python Part 1

Ben Winjum

OARC

April 18, 2024



Overview

- Brief intro to some libraries
- Quick dive into matplotlib, pandas, and seaborn
- Comparing plotting processes
 - A few simple plots made with each in turn
- Looking at a few showy examples



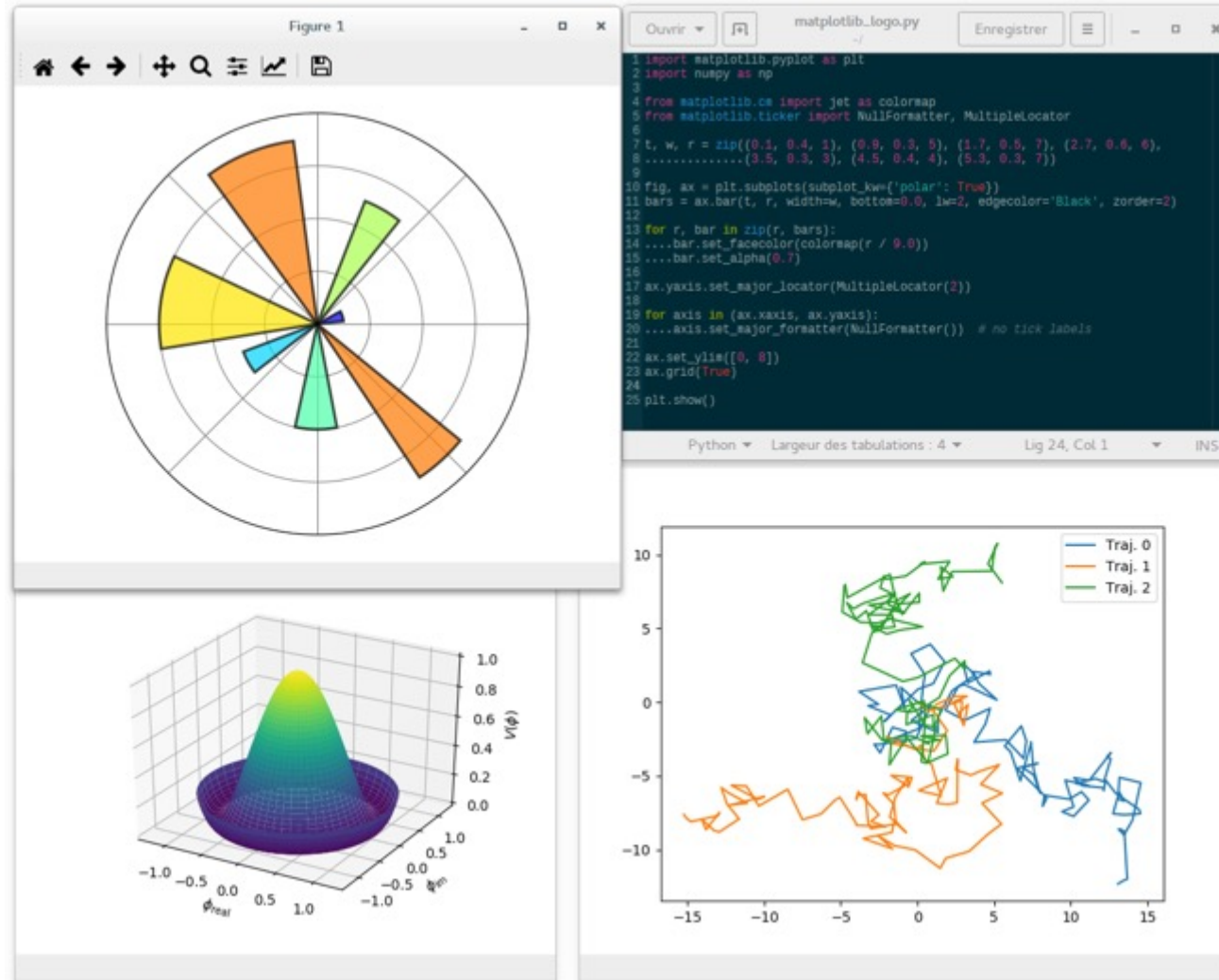
Disclaimer

- Aim: To get your feet wet
 - This is not the whole story, not even close
 - There's so so so *SO* much out there
 - And it's developing very quickly
- I won't cover installing libraries or Python fundamentals
 - Though if you have questions, please reach out to me afterwards
- If this is your first exposure, don't worry:
 - Deeper understanding comes with practice
 - Lots of practice

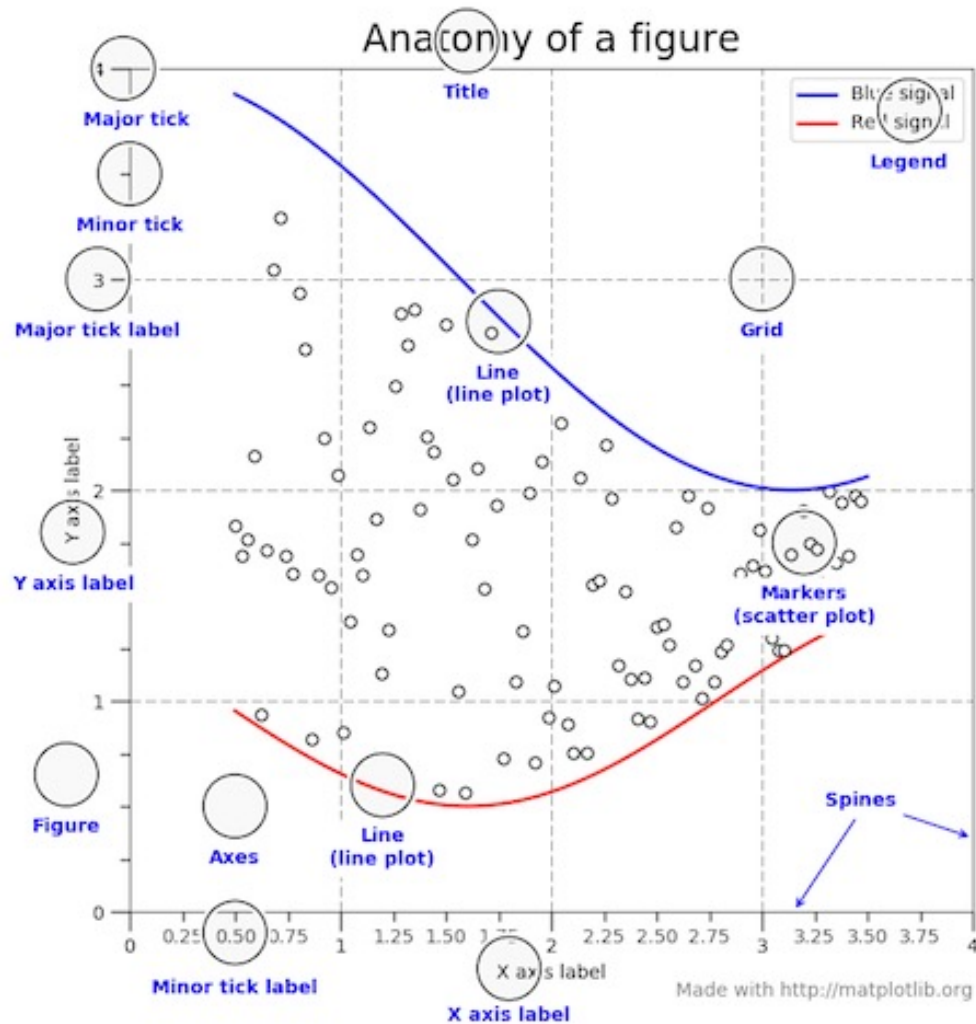


Visualization in Python: matplotlib is the workhorse lib

“Matplotlib tries to make easy things easy and hard things possible.”



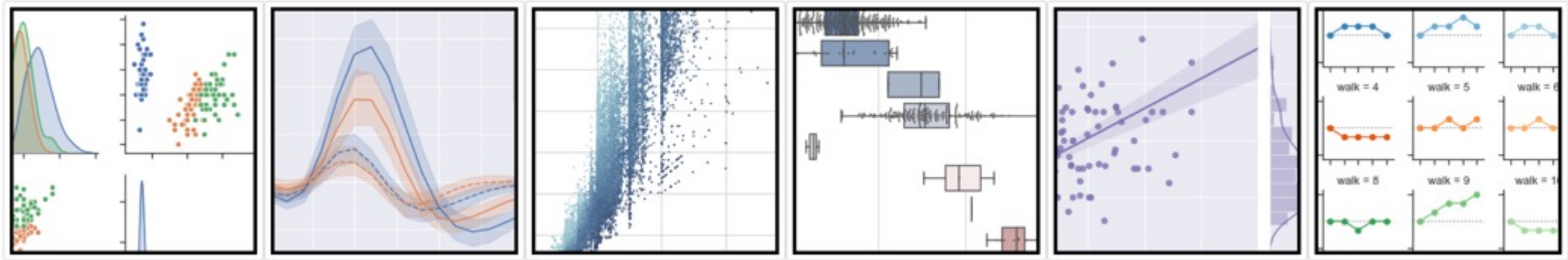
Matplotlib



- Built on the NumPy and SciPy frameworks and initially made to enable interactive Matlab-like plotting via gnuplot from iPython
- Gained early traction with support from Space Telescope Science Institute and JPL
- Easily one of the go-to libraries for academic publishing needs
 - Create publication-ready graphics in a range of formats
 - Powerful options to customize all aspects of a figure
- Matplotlib underlies the plotting capabilities of Pandas and Seaborn

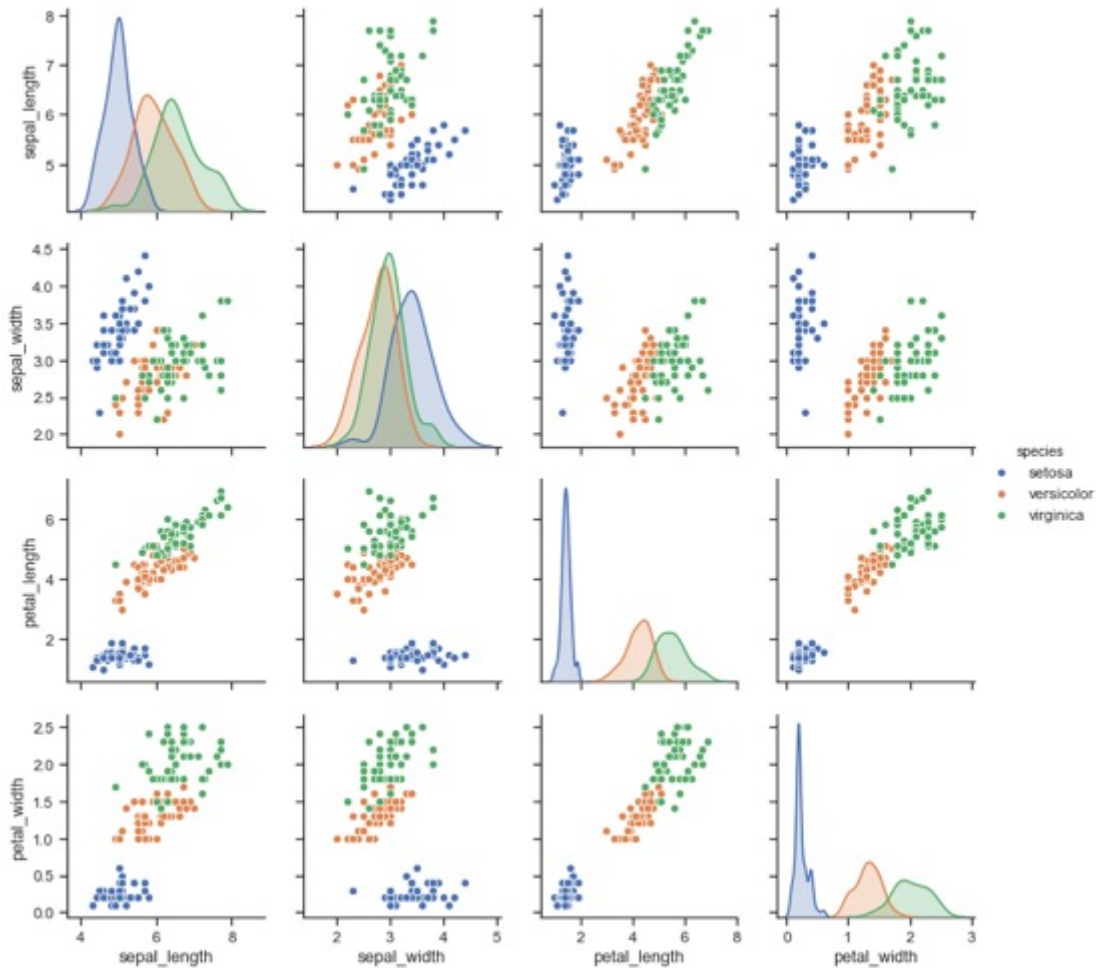
Seaborn

If Matplotlib “tries to make easy things easy and hard things possible,”
Seaborn tries to make a well-defined set of hard things easy too.

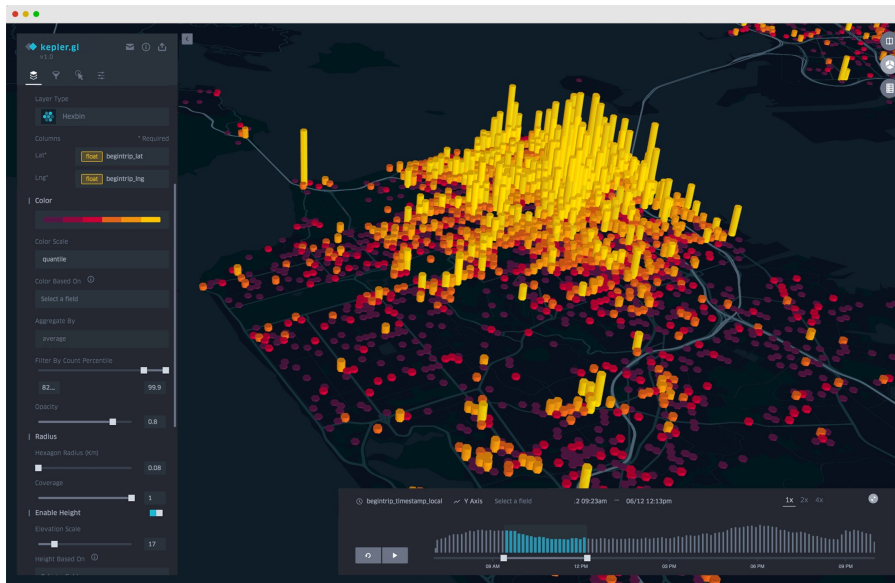
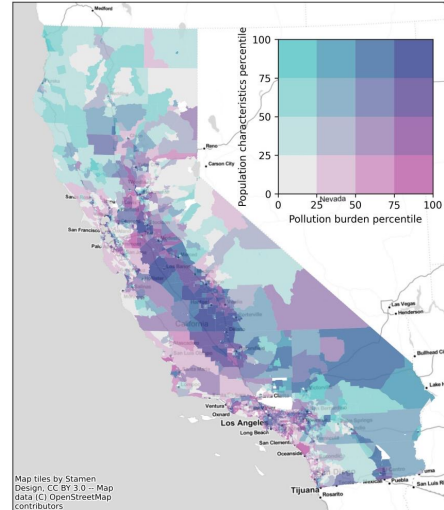
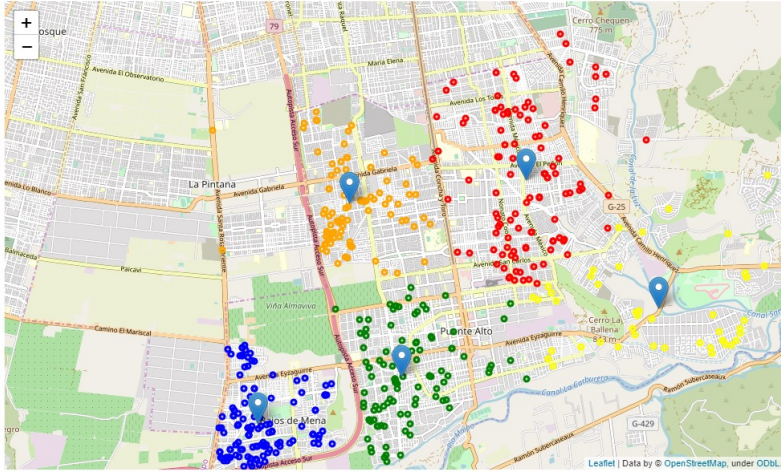


Seaborn

- Built on top of matplotlib and closely integrated with Pandas data structures
- Used for making statistical graphics and using visualization to quickly and easily explore and understand data
- The style settings can also affect Matplotlib plots, even if you don't make them with Seaborn



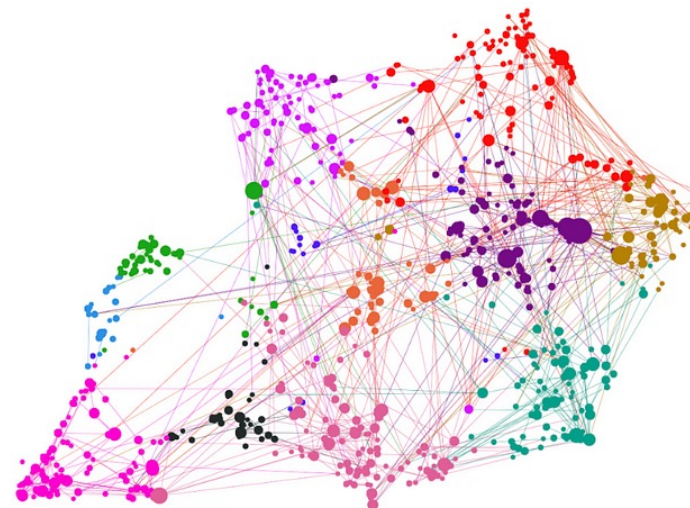
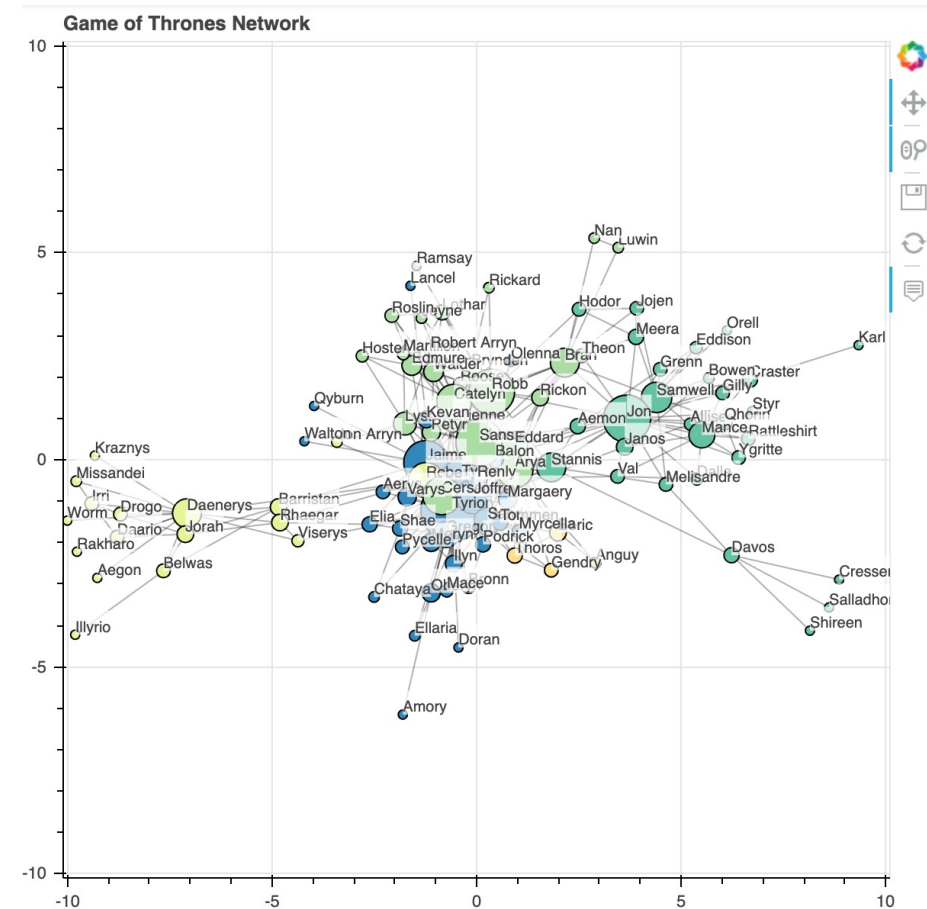
Specialized libraries -> for mapping



- **GeoPandas** extends Pandas datatypes for spatial operations on geometric types; depends on matplotlib for plotting.
 - <https://geopandas.org/en/stable/>
- **Folium** combines Python data wrangling abilities with the leaflet.js library.
 - <https://python-visualization.github.io/folium/>
- **Kepler.gl** is designed for high-performance web-based visualization of large-scale geolocation data.
 - <https://docs.kepler.gl/>

Specialized libraries -> for networks

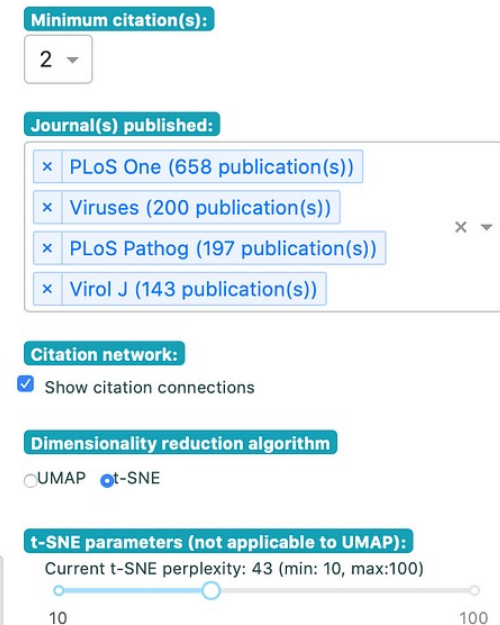
- **Networkx** – Python library for network analysis; provides basic functionality for graph visualization.
- **Bokeh** and **Dash Cytoscape** (from Plotly) – combine Python with Javascript elements to produced very interactive data visualizations



Title: Ifitm3 Polymorphism Rs12252-C Restricts Influenza A Viruses

Journal: Plos One, Published: 2014-10-14

Author(s): Williams, David Evan Joseph; Wu, Wan-Lin; Grotefend, Christopher Robert; Radic, Vladimir; Chung, Changik; Chung, Young-Hwa; Farzan, Michael; Huang, I-Chueh, Citations: 3

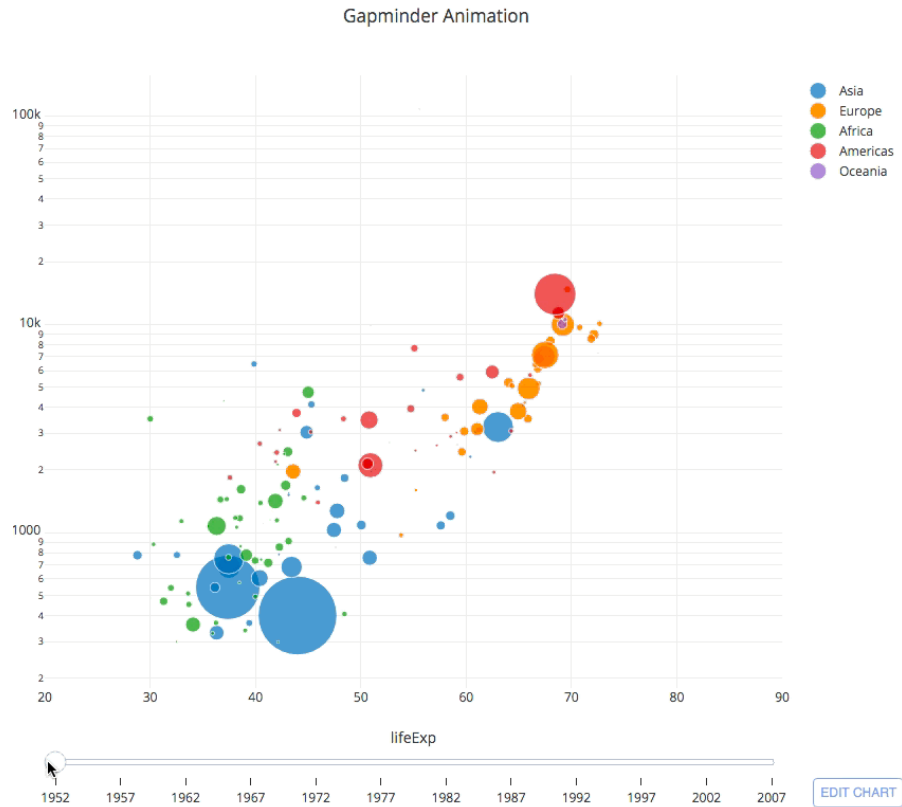


Plotly

The plotly Python library (plotly.py) is an interactive, open-source, and browser-based graphing library



Plotly



- An open-source product of Plotly, Inc., that is built on top of Javascript (plotly.js).
- Enables Python users to create beautiful interactive web-based visualizations that can be displayed in Jupyter notebooks, saved to standalone HTML files, or served as part of pure Python-built web applications using Dash.
- Also has a version for R, as well as other web visualization products

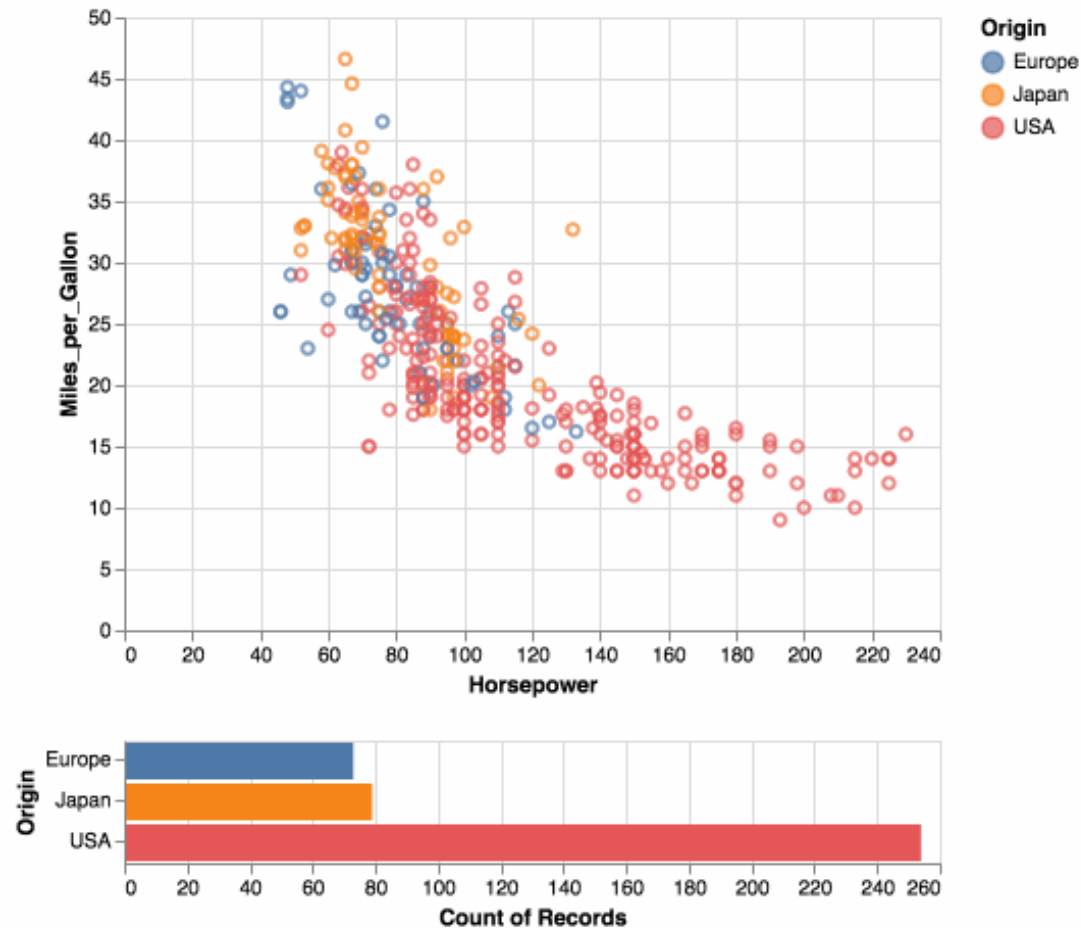
Altair

Altair is a declarative statistical visualization library for Python,
based on Vega and Vega-Lite (high-level grammar of interactive graphics)



Altair

- Based on Vega and Vega-Lite (high-level grammar of interactive graphics)
 - Vega-Lite provides a concise JSON syntax for rapidly generating visualizations to support analysis
 - Its specifications describe visualizations as mappings from data to properties of graphical marks
- Aims for elegant simplicity so focus can be on understanding data

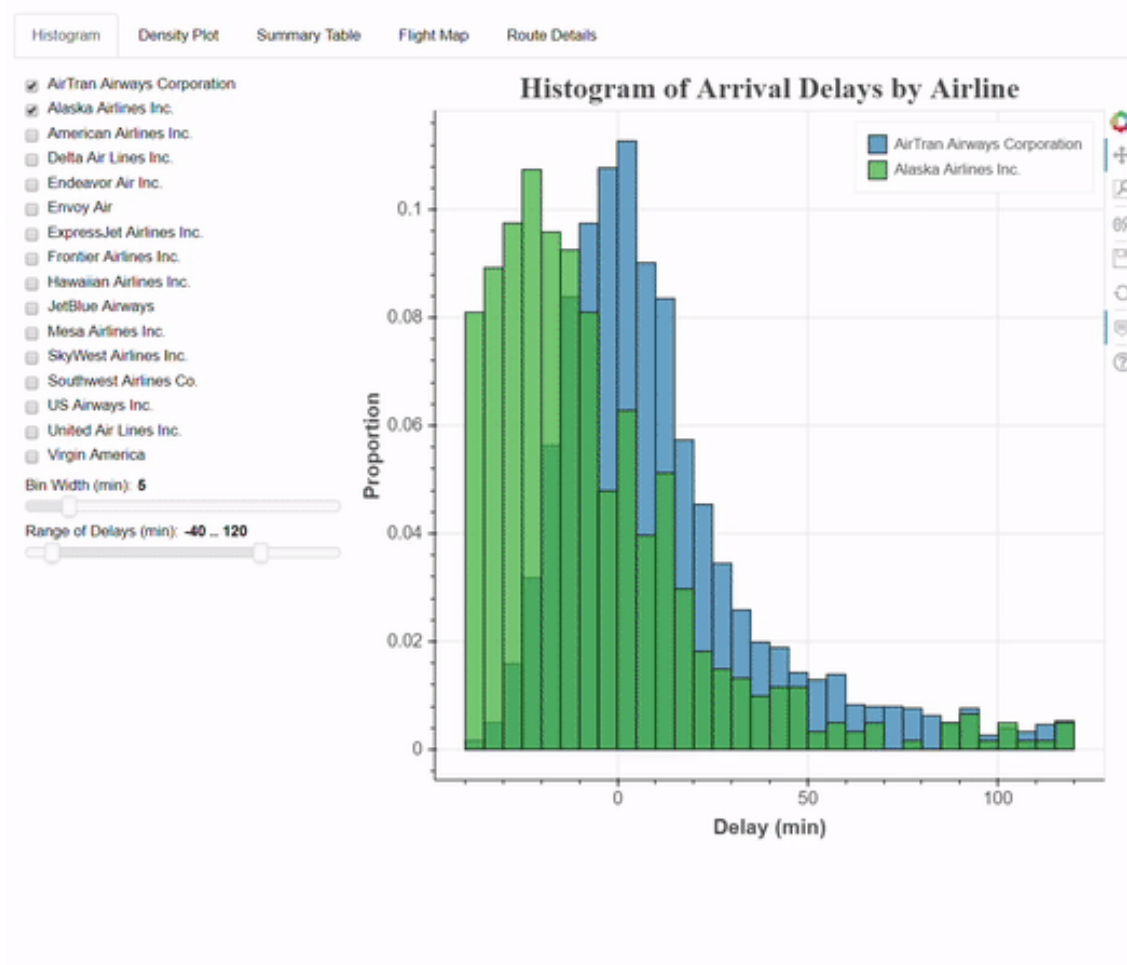


Bokeh

Bokeh creates shareable, interactive data applications for modern browsers ...
all without having to delve into JavaScript or “web tech”.



Bokeh



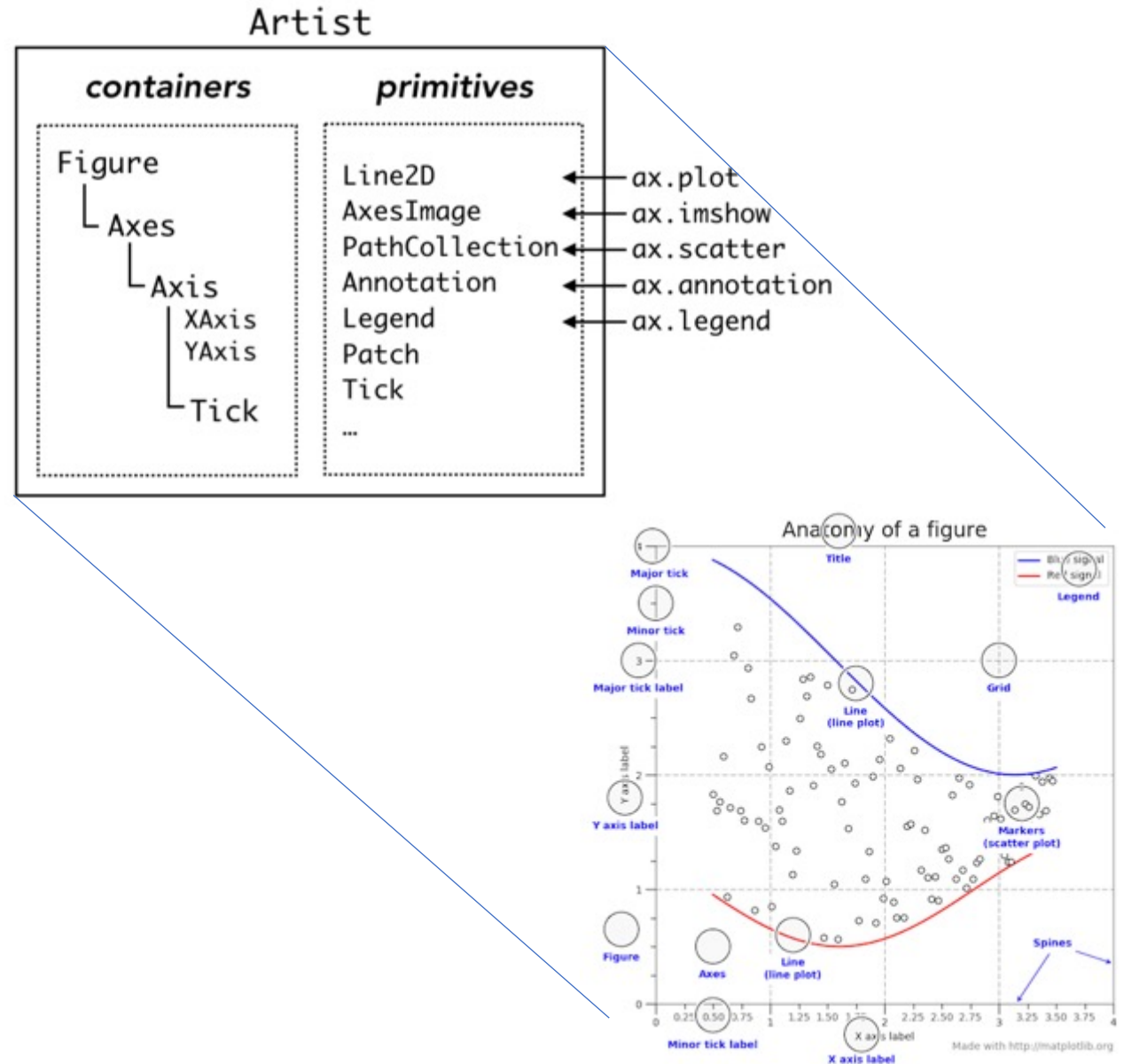
- Originally funded by DARPA
- Produces JSON files which work as input for Javascript, which in turn are used to present data to a web browser
- Aims to help anyone who would like to quickly and easily connect powerful PyData tools to interactive plots, dashboards, and data applications.
- High-performance interactivity over very large or streaming datasets

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

gif obtained from <https://towardsdatascience.com/data-visualization-with-bokeh-in-python-part-iii-a-complete-dashboard-dc6a86aa6e23>

Quick Dive: matplotlib

- Matplotlib is organized in a hierarchy
- At the top: `matplotlib.pyplot`
 - This is a module that provides high-level functions to add elements to the current axes in the current figure
- Lower levels can be accessed by figure and axes objects
 - “Figure”: an object that keeps track of child “Axes” objects (and other things like titles, legends, and the canvas)
 - “Axes”: an object that can be thought of as the plot
 - “Axis” is a different object than “Axes”



Coding time:

<https://github.com/benjum/oarc-python-data-viz-1>