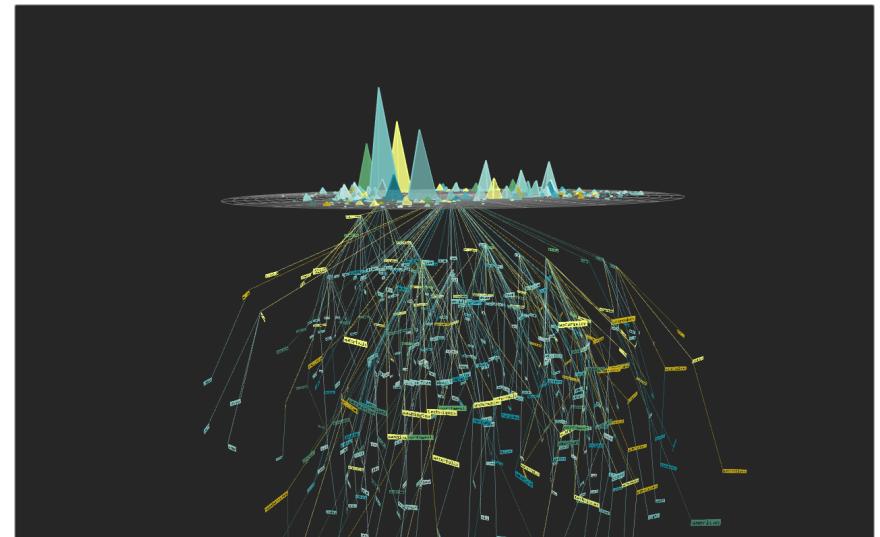
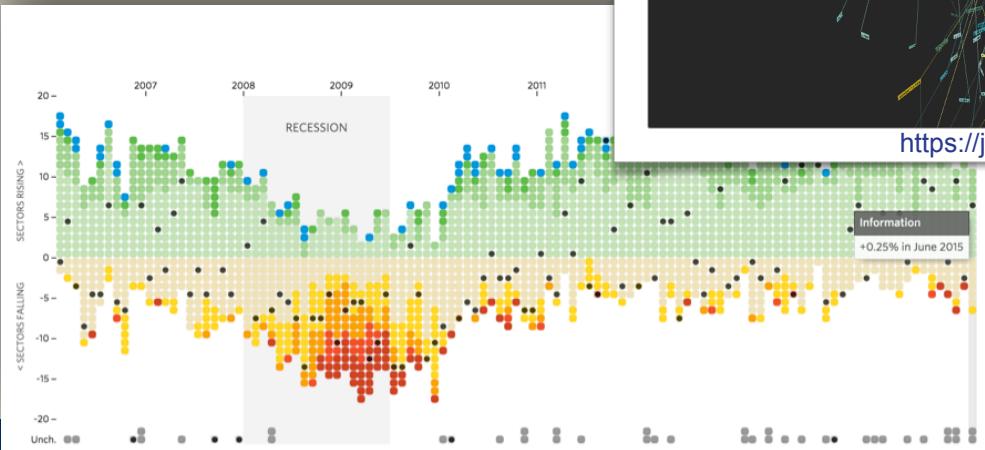


Topic 12: Plotting & Data Visualization



<https://jovianlin.io/data-visualization-seaborn-part-1/>

Plotting & Data Visualization

DATA VISUALIZATION



ILLUSTRATES RAW VALUES



DELIVERS INFORMATION



OFFERS OBJECTIVITY

DataVis refers primarily to specific techniques to represent *data* in a visual language - graphs, plots, charts ... and more.

Successful visualizations (data + design) convey an overview of the data, help you to explain the raw data, explore the data to discover new events or association, and do so in aesthetic ways that do not distract the viewers and so confuse the message.

If you get lost in creating a visual representation of your data, ask yourself what are the implications (or the “so what?” factor) and focus on that message.

Discussion

- What's your experience with **data visualization/plotting**? [Open Discussion]
- Any **specific examples** that you found useful? How/why?
- **Why and when** would we want to visualize data?
- **Beyond data vis/plotting** what is there?
- Recognize any **relationships between DataSci/BigData/DataMining/Stats** and vis? [See SAS, SPSS, IBM, Oracle/Java's white papers on visualization.]

Libraries to help you display data

- [matplotlib](#)
- [seaborn](#) [on top of matplotlib; for statistics and cool visualizations]
- [ggplot](#) [library for R and python]
- [bokeh](#) [interactive vis]
- [pygal](#)
- [plotly](#) [interactive vis]
- [geoplotlib](#) [maps; chloropleth]
- [gleam](#) Other projects
- [missingno](#)
- [leather](#)
- [basemap \(for geographic data\)](#) [requires installing base map library; pillow (but not PIL) <https://pillow.readthedocs.io/en/stable/installation.html>]

Poll:
Who has used
what?

Another great tool of particular value is the JavaScript-oriented d3.js library. Check out Mike Bostock's [Data-Driven Documents](#) homepage and on GitHub. By saving your analysed and processed data, it's easy to stream them into a file to be read later by a website calling the d3.js library.

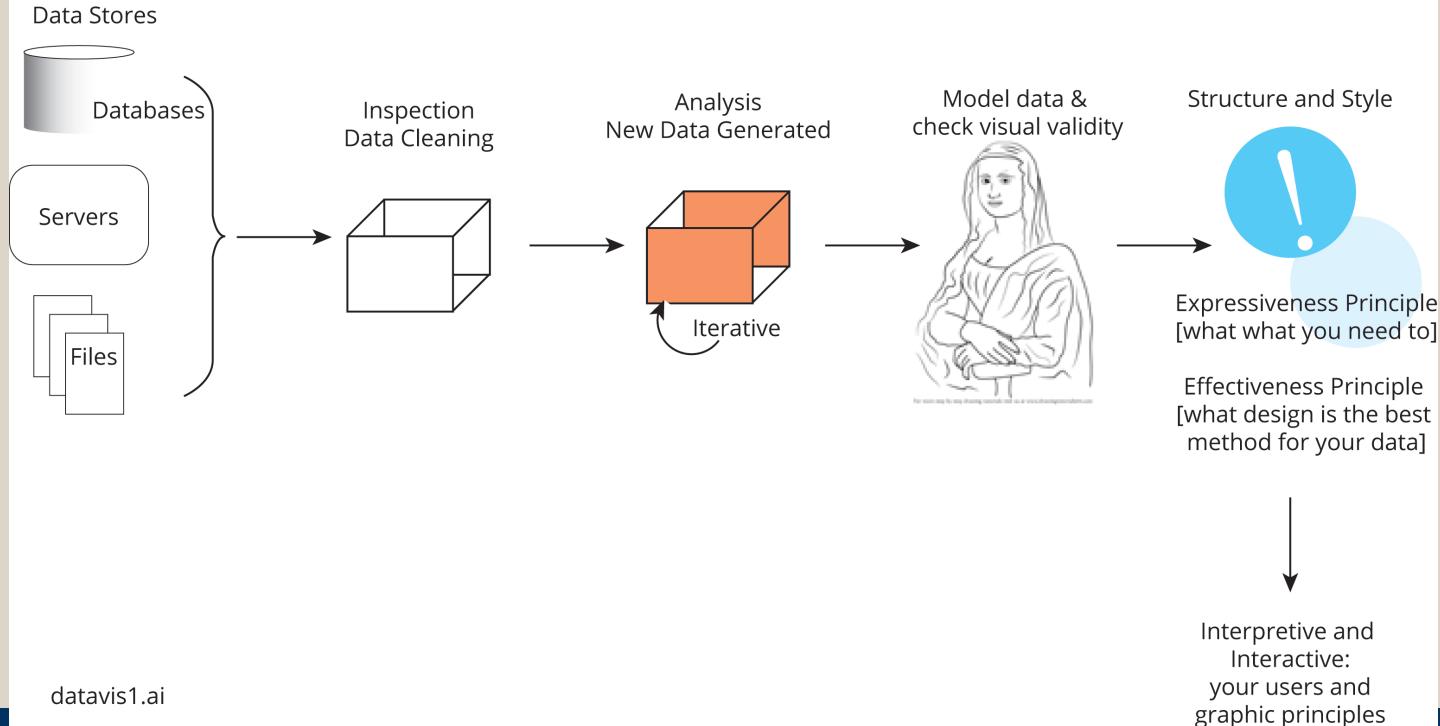
Commercial products - Tableau is very popular.

Matplotlib + Seaborn

- Read about a comparison between Seaborn and Matplotlib
<https://jakevdp.github.io/PythonDataScienceHandbook/04.14-visualization-with-seaborn.html>.
- And this great seaborn tutorial <https://jovianlin.io/data-visualization-seaborn-part-1/>

DataVis | Process

Know your sponsors' needs:
the subject domain, computing possibilities
and limitations, perform a real "needs or
requirements assessment"



Hints about DataVis

Using matplotlib as a the first example, let's say your data are stored in a .csv file ... and that you've already checked for missing values, NaN, and values that are within range and domain, and that you know the data types of each of the columns & rows (whew!).

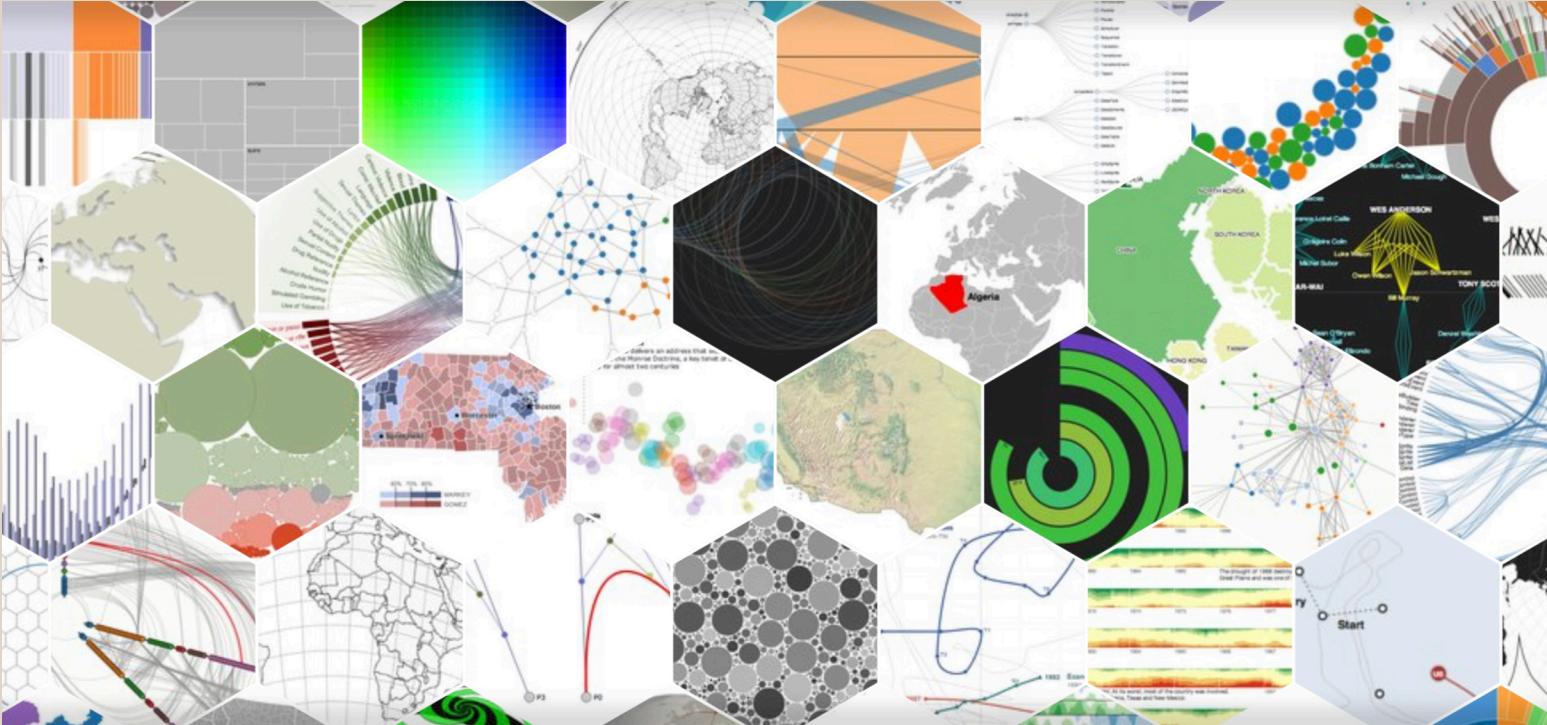
What data structure will hold your data? List? dictionary, Numpy array, other?

Linear? Box plot?
Heatmap? There are *lots* of visualization ideas.
Combine this with graphic design issues.

Color scheme (complementary, tritonic, analogous, etc.); visual elements (line, circle, spacing, size, placement, etc. Perhaps create a dictionary or list to hold these data.

Steps ... present, explain, explore

1. Import libraries
2. Use % to display inline [also other options such as “retina”]
3. Read or access your data file
4. Create the container to hold the visualizations (“figure”); compare this to HTML5’s Container and an SVG container (scalable vector graphics; d3)
5. Framework: axes [placement, ticks], title and labels, subplots
6. Design model and aesthetics [symbols, “visual primitives” [line, dot and other symbols, placement, distance, size, color scheme, alpha]]
7. Shortcuts in the code? [in the commands to create figures, subplots; color schemes stored in a dictionary]
8. User needs & interpretation; output the data plot.



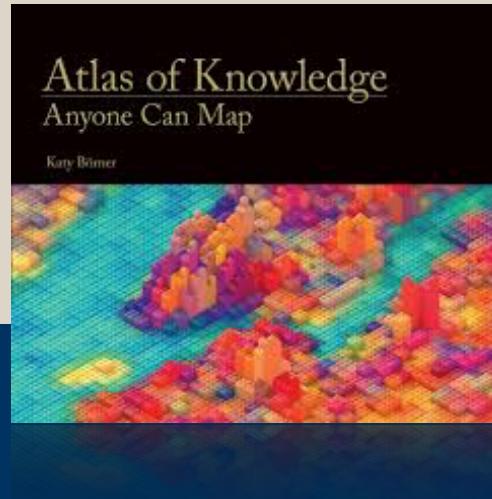
Check out d3js.org for a sample of the many types available.

Dot Graph
Stacked Bar Graph
Proportional Symbol Map
Choropleth
Force-Directed Layout
Radar Graph
Stream Graph
Parallel Coordinate
Stacked Line Graph
Box-and-Whiskers
Radial Tree
Dasymetric
Circular Graph
Tag Cloud
Line Graph
Tree View
Cartogram
Force-Directed Layout
Hive Graph

Pie Chart
Stepped Relief Map
Balloon Tree
Stripe Graph
Arc
Scatter
Bimodal
Crossmap
Steam and Leaf
Bubble Chart
Doughnut
Dot Density
Isoline
Chord
Sankey
Proportional Symbol
Treemap
Arc Graph
History Flow

Elevation
Histogram
Science
Isarithmic
Strip Map
Isochrome Map
Population Pyramid
Enclosure Tree
Dendrogram

Börner, K. *Atlas of knowledge*. MIT Press



W200-Python

Samples (of the many) demo'd

- Line plot - 1 + lines
- Line plots with insert line plot (plot in plot)
- Bar charts [usually with box+whiskers for MCT]; histograms; 2D; 3D
- Regression
- Donuts, radar, polar [these are all circular; popular in dashboards]
- Choropleth [maps - often use svg for visual units (countries, cities, etc) with data superimposed; monochromatic or analogous; map size misrepresents true data values, tho.]
- Heatmaps, Kohonen feature maps
- 3D plots

Identify the code sections for plotting

```
import matplotlib  
import matplotlib.pyplot as plt  
%config InlineBackend.figure_formats = ["retina"]  
import numpy as np
```

```
x = np.linspace(0, 5, 10)  
y = x ** 2
```

```
figure()  
plot(x, y, 'r')  
xlabel('x')  
ylabel('y')  
title('title')  
show()
```

```
fig = plt.figure()
```

```
axes = fig.add_axes([0.1, 0.1, 0.8, 0.8]) # left,  
bottom, width, height (range 0 to 1)
```

```
axes.plot(x, y, 'r')
```

```
axes.set_xlabel('x')  
axes.set_ylabel('y')  
axes.set_title('title');
```

Time for
an Activity
Breakout

Resources

Prof Benoît's Folder

- DataVis_1_Matplotlib.ipynb
- DataVis-extra-fyi.ipynb
- Python for Data Analysis

Prof Huntsinger's Folder

- “Cheet-Sheets”

Resources

- Benoît (2019) *Introduction to information visualization*
- Ware *Information visualization and Design for visual thinking*
- Tufte *Visual display of quantified data, Visual explanation, Beautiful evidence, etc.*
- Steele & Illinsky *Beautiful information* [series of articles]
- Börner, K. *Atlas of knowledge* and *Atlas of science* [various demos of one-off examples]
- Munzner *Visual analysis and design*
- IEEE *Visualization* [professional organization]
- ACM SIGVIS [professional organization]
- *Information visualization* (journal)

White Papers

- Gartner “[Not all BI platforms are created equal](#)”
- Bobriakov (2018) [Comparative analysis of ...](#)
- Tableau ([homepage](#); download and try for free)