

Today's goals:

- Plotting discussion
- Dataviz Lab & Q&A

# Discussion 3

---

COGS 108 SS1\_23

# Plotting

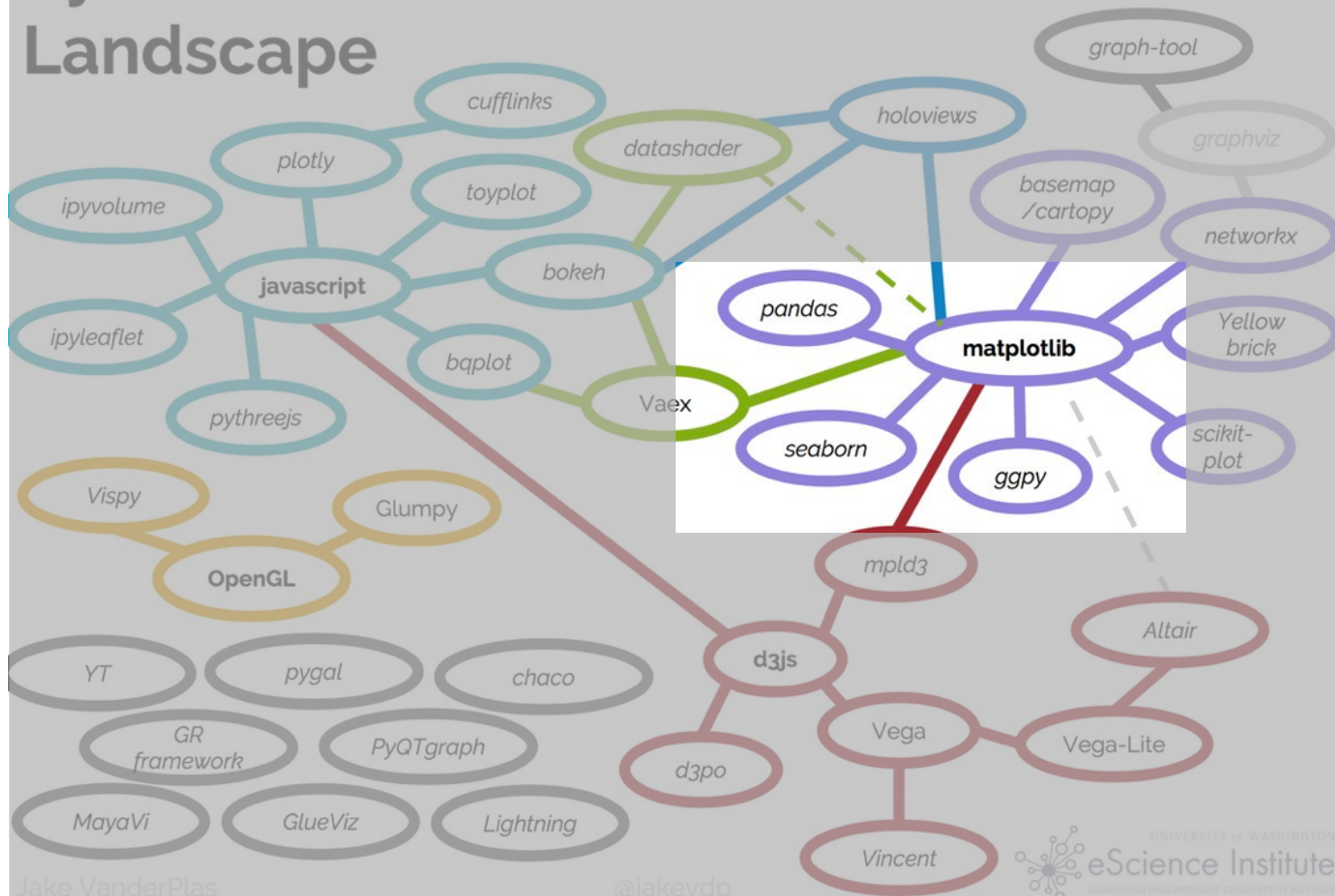
# Why are there so many ways to make the same plot?

- All of these do the same thing:

```
plt.hist(df['income10'], 25)  
df['income10'].hist(bins=25)  
df.hist('income10', bins=25)
```

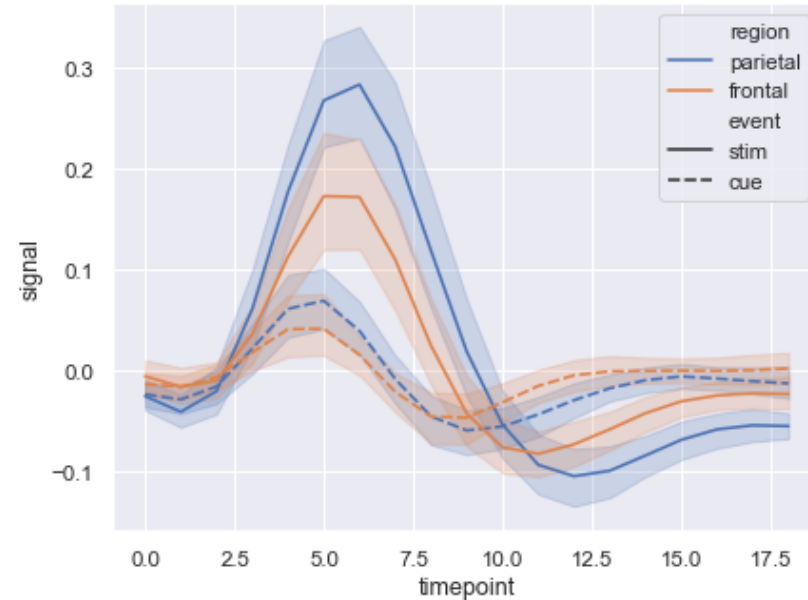
- In Python, most image-based plots created using Matplotlib (plt)
  - **plt.hist**   **plt.bar**   **plt.plot**   etc.
- Pandas gives shortcuts for matplotlib plots. Lines 2 and 3 are shortcuts for line 1.

# Python's Visualization Landscape



# Seaborn (sns)

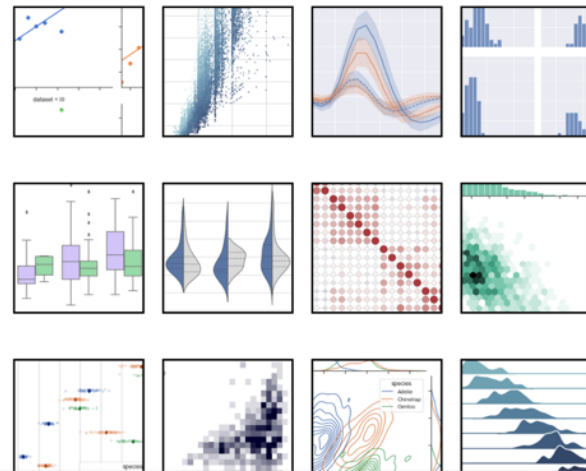
- My personal favorite is the seaborn library.
- Makes common statistical charts easy to create, like bar plots with confidence intervals.
- Again, seaborn is really just a bunch of shortcuts for matplotlib.



# Tips for data visualization

- Nothing wrong with googling the answer
  - E.g. "stacked histogram python", "scatter plot sns", "remove legend seaborn"
- Documentation is your friend: <https://seaborn.pydata.org/api.html>
  - Explore library demos for plot ideas =>
- It takes a long time to make the right plot
  - Exploration before perfection
- Half the time is getting the data in the right format- keep it tidy

Example gallery



<https://seaborn.pydata.org/examples/index.html>

# For more details

- Making good plots is a key skill! This just scratches surface.
- You can get many great jobs just by being able to make informative data visualizations.
  - In academia, publications require good figures, lectures require expressive figures to teach effectively
  - In industry, figures are used in reports, presentations, and product development (and more)
- For more, see Ch 6 of [textbook.ds100.org](https://textbook.ds100.org).

Your Turn: Work on

D3 (data visualization)