Plotting with Pandas

Talking Points: This lesson introduces the Pandas library and the beginnings of Exploratory Data Analysis. The majority of the lesson should be spent going through code – whether that is via Jupyter Slides or a Jupyter Notebook demonstration.

To present this content, begin with `04-plotting-with-pandas.ipynb` to introduce Pandas as a library and data integrity. Transition to the Jupyter Notebook to introduce reading in data, column manipulation, filtering and sorting; conclude with exercises.

Teaching Tips: - There are **Class Questions** littered throughout the notebook. Use as much/little time on these as you see fit relative to how your class is pacing - There is no **Independent Exercise** at the end of this lesson. It is aspirational to have time to let students work entirely independently on this time-wise, so consider doing a guided code-along or paired programming. Use this time to have students set their own challenges. - Pause after learning objectives and level-set for what students will get out of the lesson

---

A Note on Delivery

- This unit's lessons will occur in jupyter notebooks
  - Slides will be an introduction to the lesson (no code, just overview)
  - Then, we'll open a notebook and start coding!

Teaching Tips: - We could have made this into a speaker note, but it's helpful to get it out there so everybody's on the same page - No repl.it for this unit as we'll be in notebooks

---

Plotting with Pandas

- Pandas `.plot()` functionality is effectively a wrapper for matplotlib
- Matplotlib is a charting library for python and scientific computing
- It's considered the de-facto standard for charting locally
  - It's best for scientific papers, EDA, and general introspection of data
  - It's not so great for production level charts that are embedded in applications (check out d3.js

Talking Points:

- Talk briefly about where charts are interpreted, and why different tools may be advantageous

---

So, Pandas and Matplotlib

Whats a wrapper?

- A program that *abstracts* another program to modify its interface

???

- Pandas `.plot()` functionality references matplotlib behind the scenes
- Matplotlib has a reputation for being fairly complex
  - Even for fairly simple charts, you will frequently write loops
  - A fairly plain chart can be 20-30 lines of code
- Pandas helps us here and most charts can be produced with 1-2 lines of code
  - Some functionality is reduced, but *effort is minimized in most cases*