Plotting with Pandas

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!

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

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*

Talk Data to Me

We'll be using three data sets for this lesson:

- Football Records: International football results from 1872 to 2018
- Avocado Prices: Historical data on avocado prices and sales volume in multiple US markets
- Chocolate Bar Ratings: Expert ratings of over 1,700 chocolate bars

All datasets have been graciously downloaded from Kaggle.com, and we'll discover that the right visualization can often replace a bit of fancy machine learning, if done properly.

Chart Types

We'll be covering the following chart types during this lesson:

- Time series line charts
- Categorical bar charts
- Histograms of single columns
- Histograms of entire data frames
- Scatter plots (continuous vs continuous)
- Scatter matricies (multiple scatter plots in a grid)
- Scatter plots with class colors for data points

Let’s Go!

- Open up your dataset!