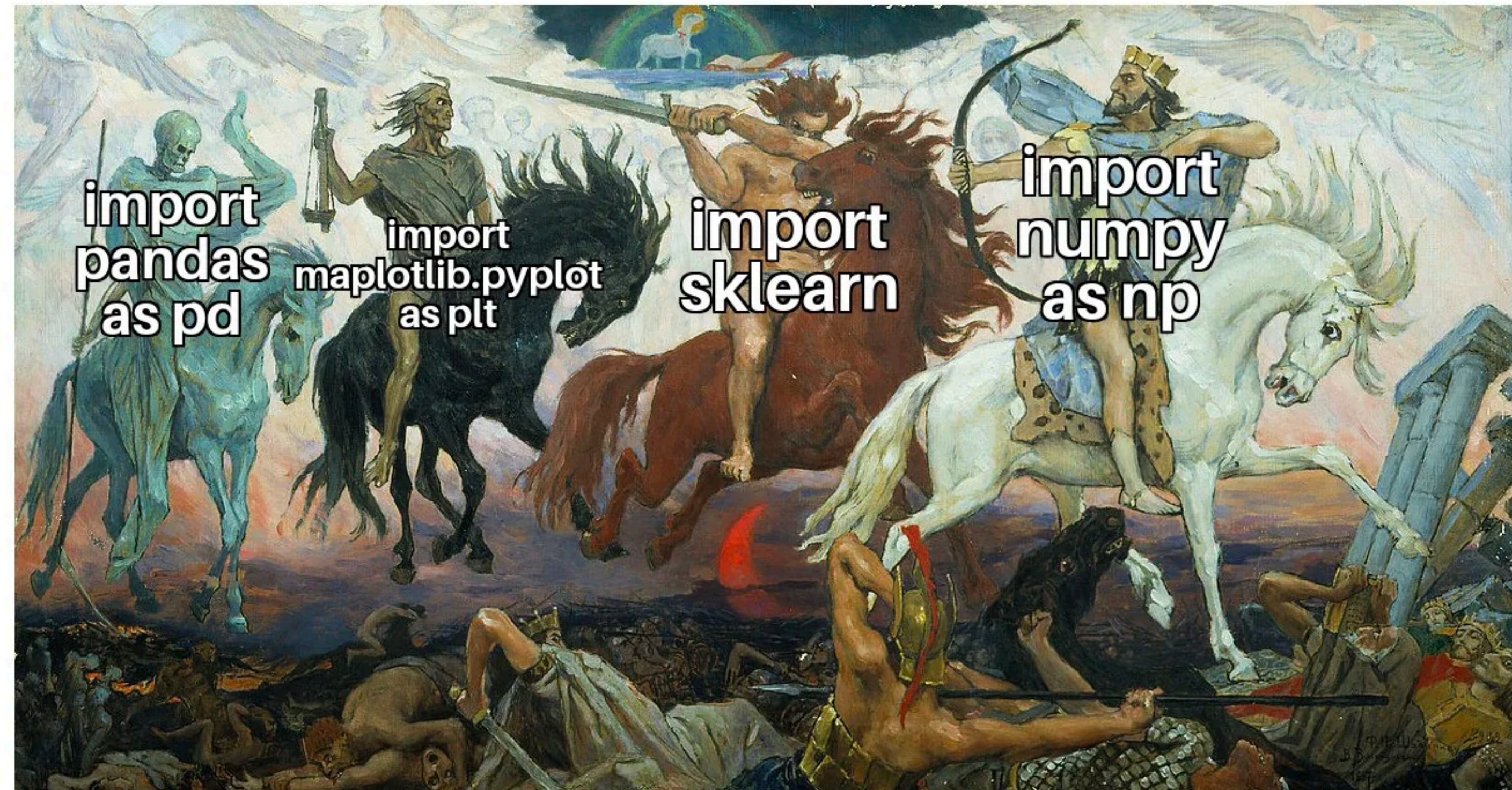


Recitation 9

CS 210

Anamika Lochab

The Four Horsemen of basic machine learning in Python



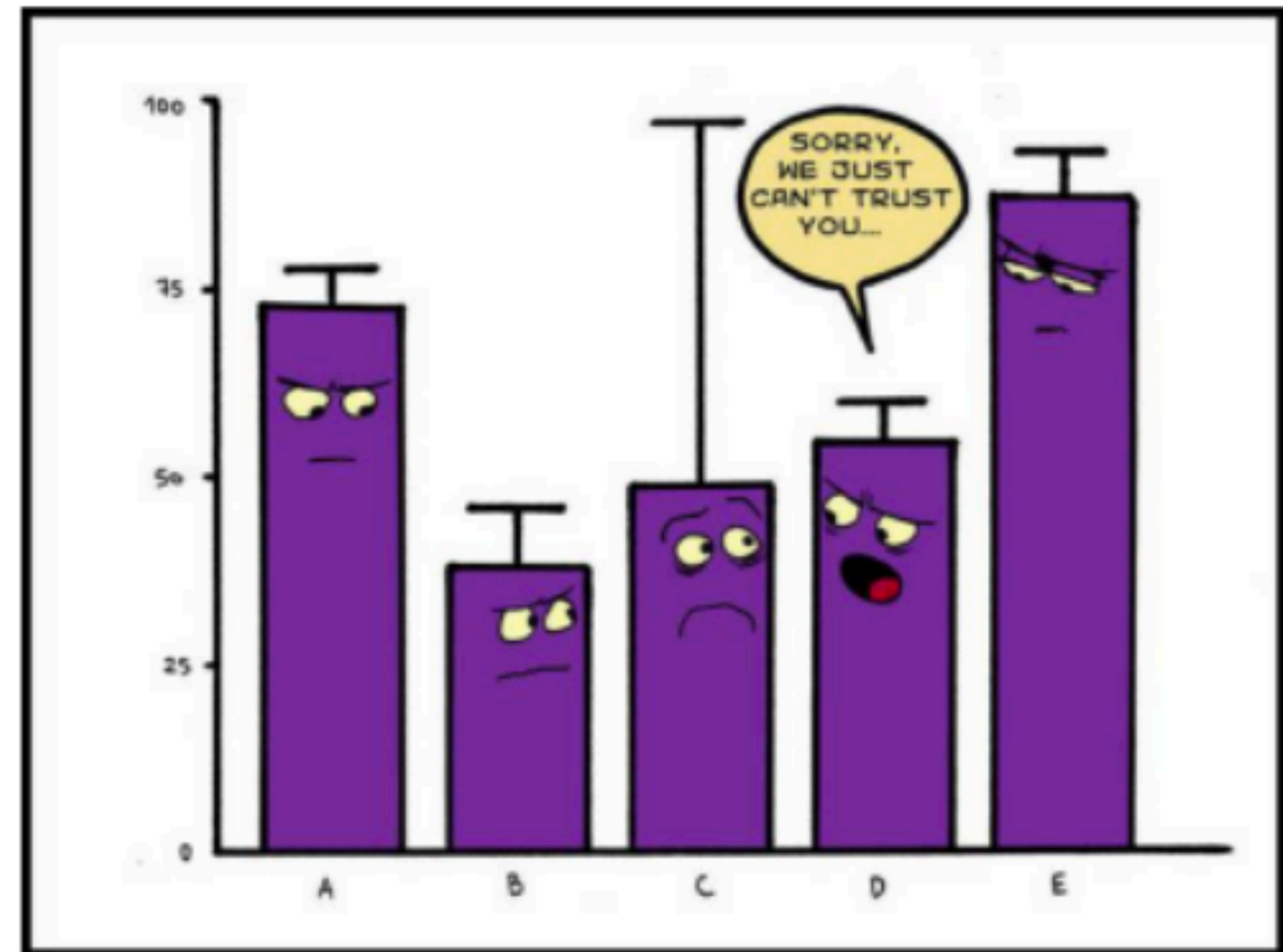
Visualization Objectives

- Record Information
- Analyze data to support reasoning
- Confirm hypotheses
- Communicate ideas to others

Why Visualize?

- To record information
- To point out interesting things
- To communicate information
- To analyze data

A data science meme



Numpy

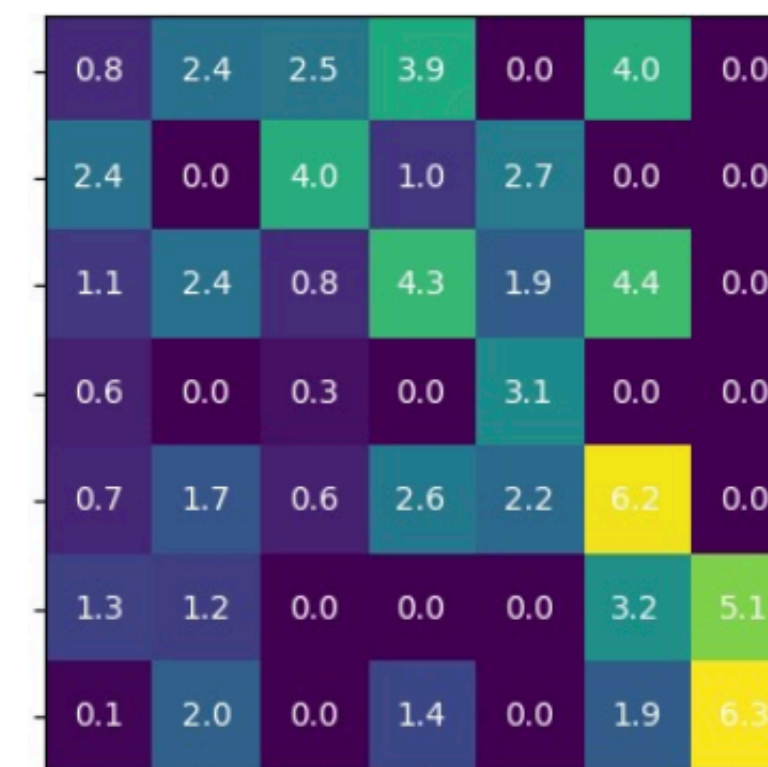
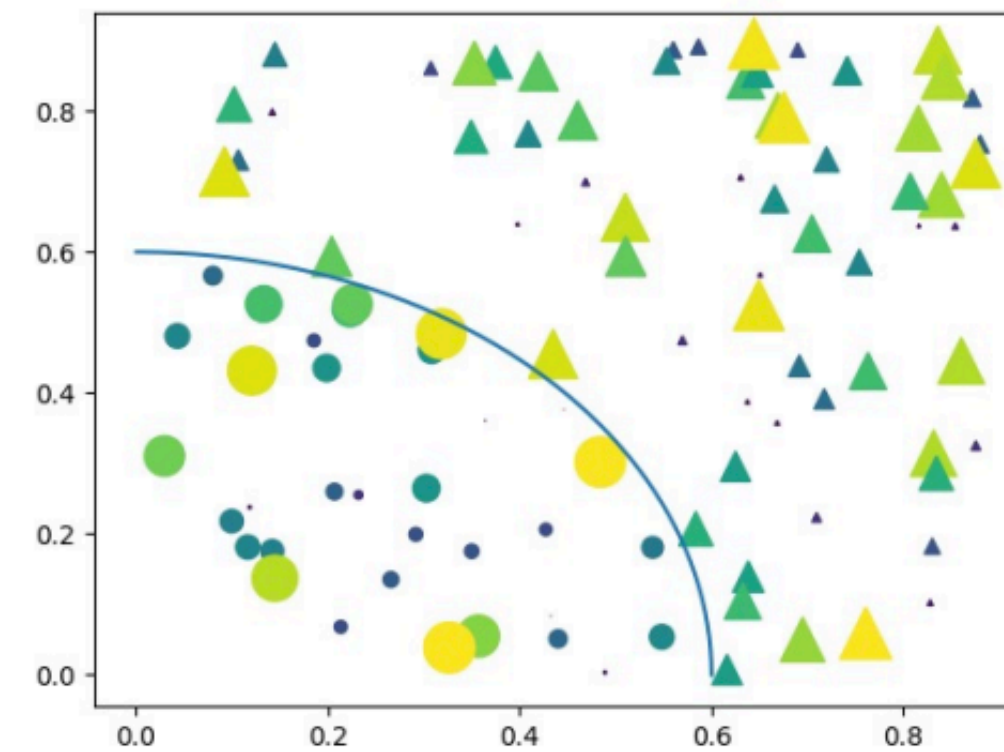
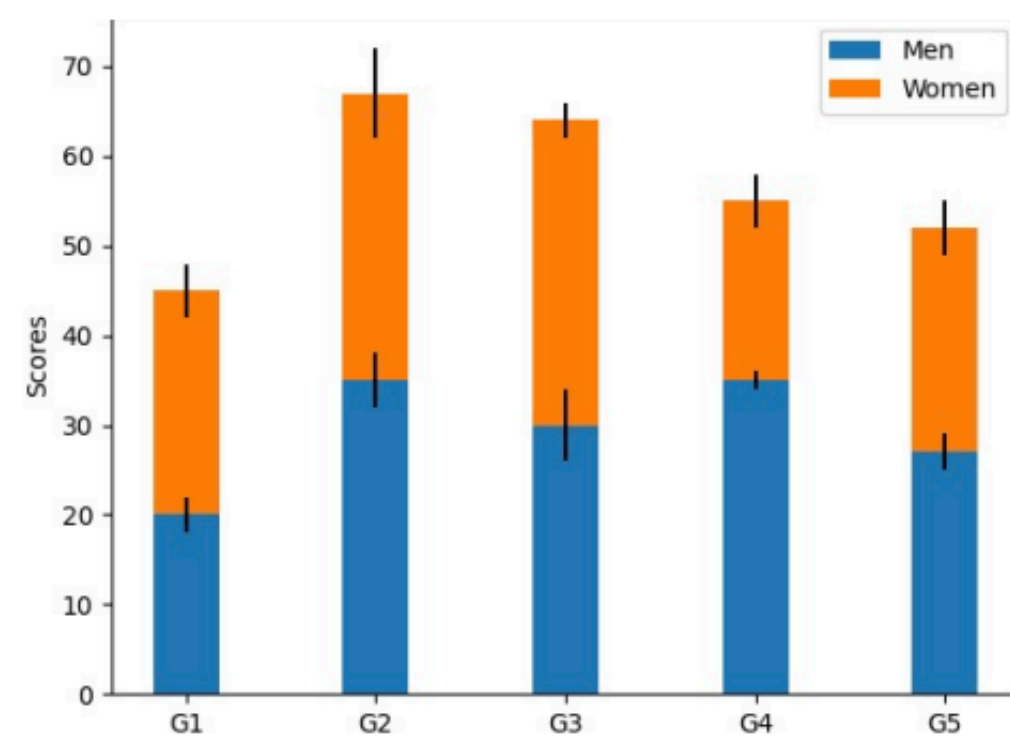
- Fundamental package for scientific computing
- Exceptionally fast – written in C
- Main data structure: • ndarray : n-dimensional arrays of homogeneous data types
- Data manipulation \approx NumPy array manipulation
- Used in other libraries - Matplotlib, pandas, scikit-learn

Pandas

- Fundamental tool for handling and analyzing input data
- Particularly suited for tabular data
- Implements powerful data operations
- Main data structures:
 - DataFrame: A table with rows and columns
 - Series: A single column

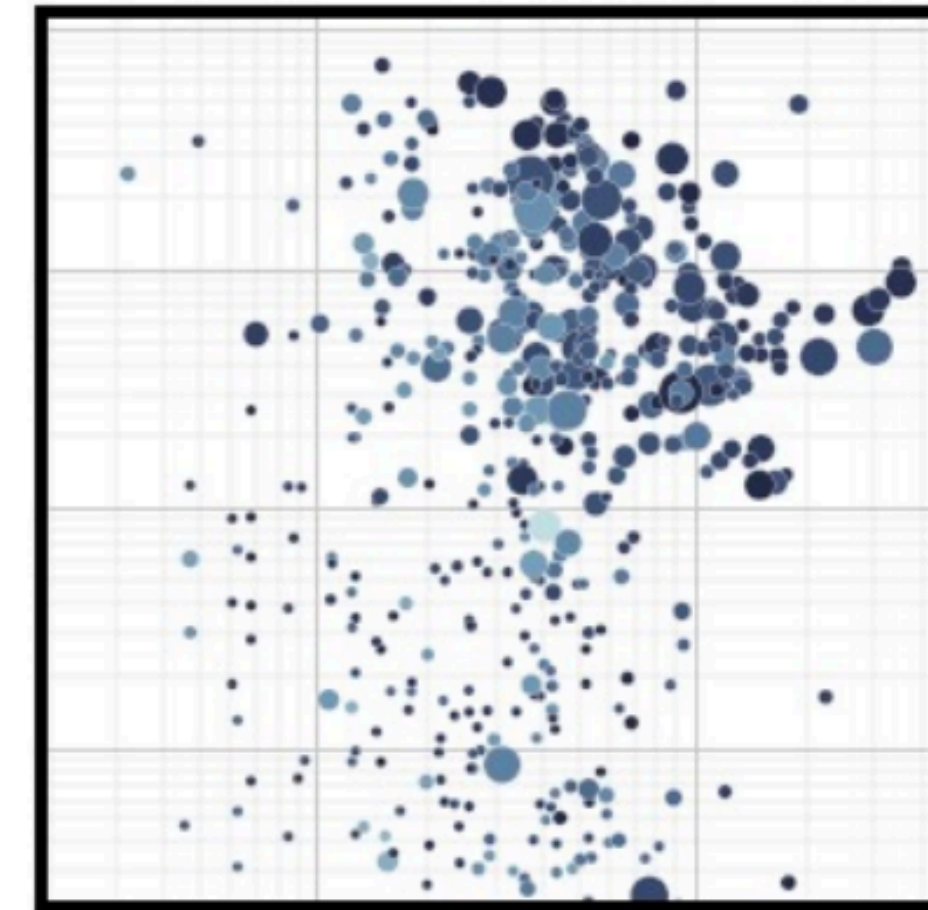
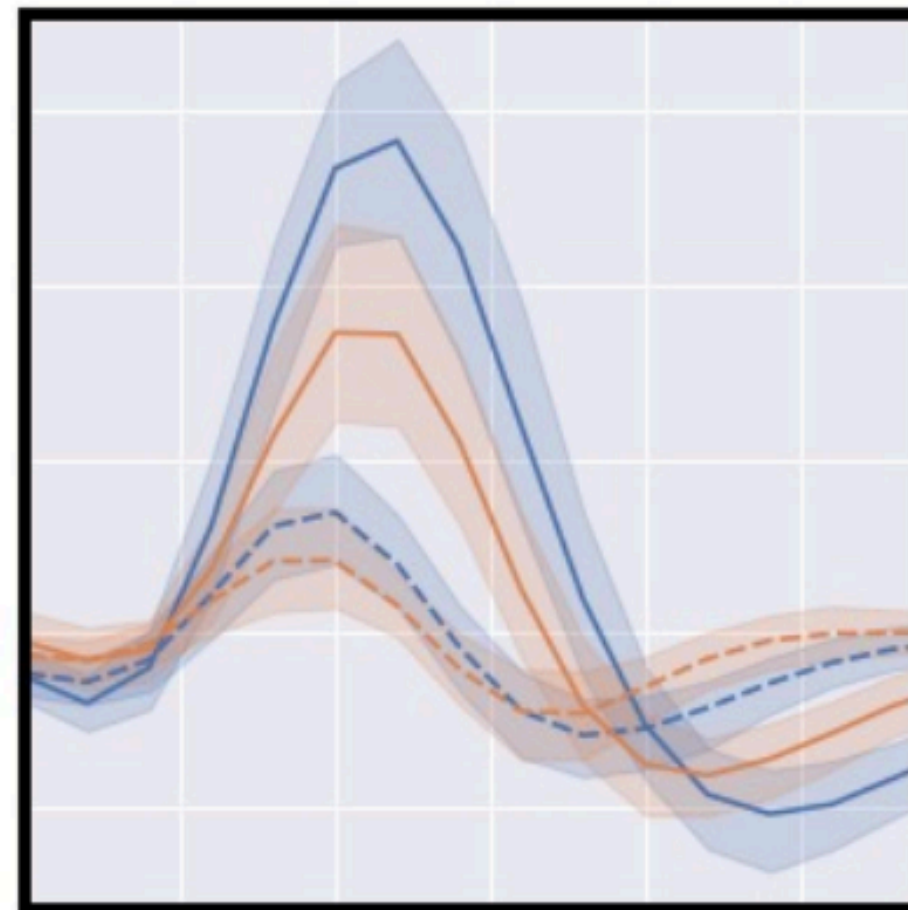
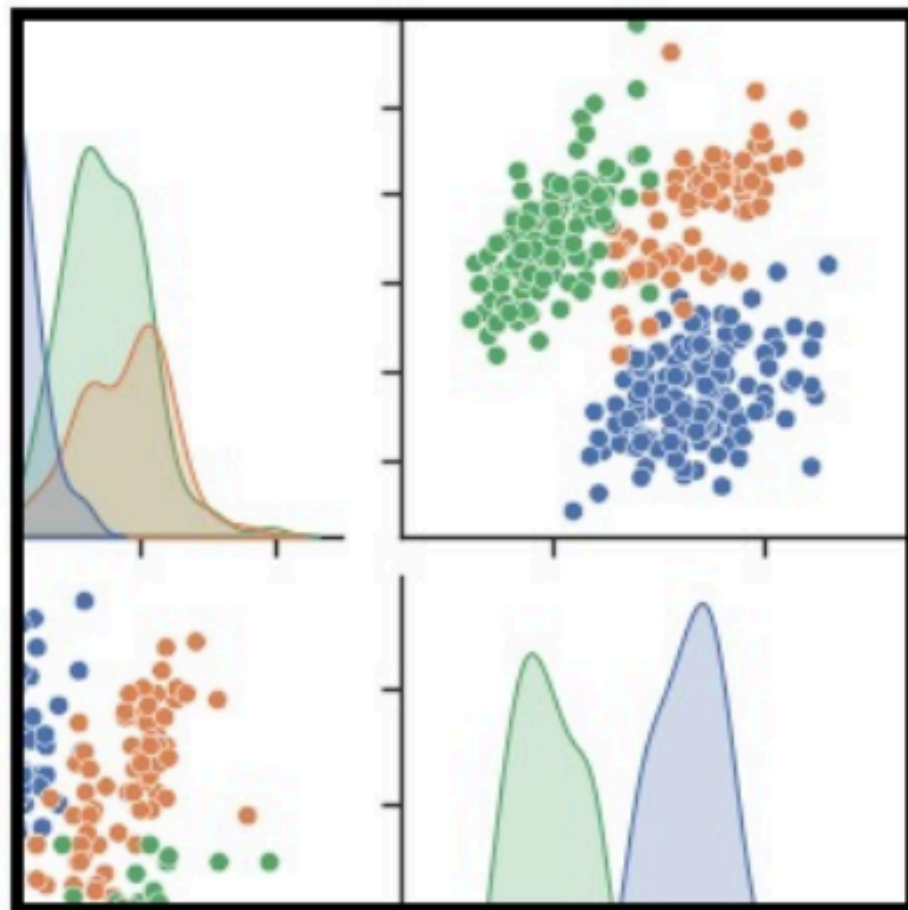
Matplotlib

- Used for basic plotting
- Highly customizable
- Works with NumPy and pandas



Seaborn

- Used for statistical data visualization
- Uses fewer syntax with good default themes
- Integrated to work great with pandas data-frame
- Uses Matplotlib under the hood



Colab File