INTRODUCTION TO PYTHON: DATA VISUALIZATION

Katharine Teykl, M.L.S.
Digital Scholarship Librarian
October 19th, 2022

Learning Objectives

After this workshop, participants will be able to:

- Describe why Python is useful for data visualization

- Identify Python libraries utilized in data visualization

- Create and visualize datasets using the Matplotlib and Pandas libraries

WHAT IS PYTHON?

From Python.org: "...an interpreted, **object-oriented**, high-level programming language with dynamic semantics."

- Object-oriented: focused around data (objects) rather than functions or logic
- Strong focus on syntax and readability
  - PEP 8 Style Guide

WHY IS PYTHON USEFUL FOR DATA VISUALIZATION?

- Adaptable and flexible for a variety of visualizations
  - Scatter plots, line graphs, heatmaps, etc.
  - 3-D visualizations can now be created in Python
- Ability to both read data and visualize data
  - Upload .csv files with pandas to use within visualizations
  - Ability to create small datasets using arrays
  - Simple to manipulate data and create groupings
- Low barrier to entry and beginner-friendly

Libraries for Data Visualization

Matplotlib

- Most popular Python library used to create visualizations of numerical data
    - Line graphs
    - Scatter plots
    - Bar graphs
- Compatible with other libraries used in data analysis
    - Numpy, Pandas
- Not useful for some kinds of data
    - Time series data, geographic/spatial data

Seaborn

- "Extension" of matplotlib
- Able to create more complex visualizations
  - Heat maps, histograms
- More style customizations
  - Ability to create themes
- Check that seaborn is installed in your Python environment!
  - pip install seaborn

Pandas

- Data analysis and manipulation
  - Used in conjuction with visualization libraries in order to read and manipulate data
    - Compatible with Matplotlib
  - Data groupings and dataframes
  - Reads .csv files

Additional Visualization Libraries

- Plotly
  - Creates 40+ web-based visualzations
- GGplot
  - Based on ggplot2 visualization package from R
- Geoplotlib
  - Creates 2-D visualizations of geographic data
- Other libraries to be aware of:
  - Altair, Bokeh, Pygal

Exercise Time!