DATA X

Data Visualization: Matplotlib and Seaborn

Created By:
Joshua Rafael Sanchez
joshuarafael@berkeley.edu
Part 1
Basic Visuals | Matplotlib, Seaborn

Part 2
Interactive Visuals | Plotly, Bokeh, Tableau, etc.
Deeper insights into more interactive and fun data visualization functions. Introduction to Plotly, Bokeh and Tableau.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Matplotlib</strong></td>
<td><strong>Seaborn</strong></td>
<td><strong>Plotly</strong></td>
<td><strong>Tableau</strong></td>
<td><strong>References</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>About Seaborn</td>
<td>About Plotly</td>
<td>About Tableau</td>
<td>Links to Notebooks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Installing Seaborn</td>
<td>Installing Plotly</td>
<td>Tableau Desktop</td>
<td>References Cited</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theme Adjustments (w/ ex)</td>
<td>Using Plotly Offline or Online</td>
<td>No-Code</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plotly Examples</td>
<td>Visualization Tools</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Plotly Alternatives:</td>
<td>Visualization</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>○ Bokeh (w/ ex)</td>
<td>Comparison</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>○ D3.js</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Visualization

What is data visualization?
Data visualization is the graphical representation of information and data.

What makes for effective data visualization?
Visualization transforms data into images effectively and accurately represent information about the data.

What are the advantages of data visualization?
Makes for easier interpretation of patterns and trends as opposed to looking at data in a tabular/spreadsheet format.
Examples of Data Visualizations

Left to Right: John Snow’s 1854 Cholera Outbreak Map, Demographic Gender Breakdown, Government Budget Treemap of Benin
About Data Visualization

**Painting a Picture of Data Visualization:**
- Oxford English Dictionary Definition, 1989: To form a mental image, picture of (something not present or visible to the sight, or of an abstraction); to make visible to the mind or imagination
- There are 3 goals: To **explore** data, to **analyze** data, and/or to **present** data.

**Question: What Would You Like to Show?**
- **Relationships** between variables
- **Composition** of the data over time
- **Distribution** of variable(s) in data
- **Comparison** of data with relation to time, variables, categories, etc.
About Data Visualization
Matplotlib - About

**About Matplotlib:**
- Matplotlib is a comprehensive library for creating static, animated and interactive visualizations in Python.
- Usage: Matplotlib/Pandas is mostly used for quick plotting of Pandas DataFrames and time series analysis.

**Pros and Cons of Matplotlib:**
- Pro: Easy to setup and use.
- Pro: Very customizable.
- Con: Visual presentation tends to be simple compared to other tools.
Matplotlib - Installation

Installing Matplotlib should be straightforward. Sample code for installing packages:

```
# Standard way to import matplotlib from scratch: import matplotlib.pyplot as plt
# you might also see import pylab, however the procedure below is preferred

import numpy as np
import pandas as pd
import matplotlib
import IPython
import matplotlib as mpl
import matplotlib.pyplot as plt  # always import pyplot module as plt (standard)

%matplotlib inline
```
Matplotlib - Object Hierarchy

- **Figure**: Outermost container for a Matplotlib graphic. Can contain multiple Axes objects.
- **Axes**: Actual plots. Contain smaller objects (tick marks, individual lines, etc.)
- **Artist**: Everything that is seen on the figure is an artist.

https://realpython.com/python-matplotlib-guide/
Matplotlib - 2 Approaches to Plotting

1. **Functional/MATLAB Approach (Non-Pythonic)**
   - Most common way of Matplotlib.
   - Pro: Easy approach for interactive use.
   - Con- Not pythonic: Relies on global functions (where variables are declared outside of functions) and displays global figures.

2. **Object-Oriented Approach (Pythonic)**
   - **Recommended** way to use Matplotlib.
   - Pro: Pythonic is object-oriented (you can build plots explicitly using methods of the figure and the classes it contains.)
Matplotlib - Non-Pythonic Example

Example: Combining Line & Scatter Plots From Categorical Variables

```
In [179]:

names = ['grade_a', 'grade_b', 'grade_c',
         'grade_d', 'grade_f']
values = [10, 30, 20, 8, 3]

plt.plot(names, values, 'm-')
plt.scatter(names, values);
```
Matplotlib - Pythonic Example

Example: Simple Line Plot & Bar Plot

```
In [185]:

# f is the canvas object, can contain several plots
# i.e. axes objects (p)
f, ax = plt.subplots()  # returns tuple:

    group_mean = np.mean(values)/4

    # Add a horizontal line denoting average
    ax.axhline(group_mean, ls='--', color='r')

    # data as parameters
    ax.plot([1,2,3,4],[5,2,8,7]);
    ax.hist(np.random.randint(1,4,10));
```
Seaborn - About

About Seaborn:
- Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- Usage: Those who want to create amplified data visuals, especially in color.

Seaborn’s Pros and Cons:
- Pro: Includes higher level interfaces and settings than does Matplotlib.
- Pro: Relatively simple to use, just like Matplotlib.
- Pro: Easier to use when working with Dataframes.
- Con: Like Matplotlib, data visualization seems to be simpler than other tools.
Seaborn - Installation

Installing Seaborn should also be straightforward. Sample code:

```python
from mpl_toolkits.mplot3d import Axes3D
from sklearn.decomposition import PCA
from sklearn import datasets
import seaborn as sns
import pandas as pd
```
Seaborn - Theme Adjustments

Theme Design - Setting Style:
- Use the five built-in themes to style the figure/background of plots:
  - Grids: darkgrid, whitegrid
  - Colors: dark, white, ticks.

Setting Scale:
- Use the four scaling plot presets to customize the size of the plot:
  - In order of relative size: paper, notebook, talk, poster.

Setting Fonts and Line Widths:
- How to change the size of the text:
  - Change the font_scale parameter for sns.set_context().
- How to change the line width of the text:
  - Change the rc parameter for sns.set_context().
Seaborn - Theme Adjustments w/ Examples

Let’s look at the 5 built-in themes to style the figure (background of plots):

- **Grids:** darkgrid, whitegrid
- **Colors:** dark, white, and ticks.

Consider examples using famous [Iris Flower Data Set](https://scikit-learn.org/stable/datasets/index.html#iris-dataset).

- Left graph uses vertical bar plot w/ whitegrid, right graph uses swarm plot with dark.
Seaborn - Theme Adjustments: Color

Option 1- Default & Built-In Color Palettes:

- About: Seaborn has six variations of its default color palette: deep, muted, pastel, bright, dark and colorblind.
- How to use: Use
  sns.color_palette() or
  sns.set_palette() for individual plots.
  - To set a color palette for all plots, use sns.set(*args).
Seaborn - Theme Adjustments: Color

**Option 2- Color Brewer Palettes:**
- About: Created from the research of cartographer Cindy Brewer, these color palettes are specifically chosen as to be easy to interpret ordered categories.
- How to use: Use `sns.color_palette()` or `sns.set_palette()` for individual plots.
  - To set a color palette for all plots, use `sns.set(*args)`.
Seaborn - Theme Adjustments: Color Examples

Left image: Code and resulting plot using default & built-in color palettes.
Right image: Code and resulting plot using a Color Brewer palette.
Matplotlib vs. Seaborn

- Visuals
- Options
- Interactivity
Plotly - About

About Plotly:
- From website: Plotly is an interactive, open-source plotting library that supports over 40 unique chart types.
- Usage: Plotly is advantageous for those who want an interactive environment which many use cases, ranging from statistics to finance to geography and more.

Pros and Cons of Plotly:
- Pro: Make beautiful, interactive, exportable figures in just a few lines of code.
- Pro: Much more interactive & visually flexible than Matplotlib or Seaborn.
- Con: Confusing initial setup to use Plotly without an online account, and lots of code to write.
- Con: Out-of-date documentation and the large range of Plotly tools (Chart Studio, Express, etc.) make it hard to keep up.
Plotly - Installing

**Installing Plotly Offline: (if you want to host locally on your own computer)**

- Steps: You need to import packages and use commands:
  - Resource: Keep checking current version: Initialization for Online Plotting
  - Command to create standalone HTML: `plotly.offline.plot()`
  - Command to create plot in Jupyter Notebook: `plotly.offline.iplot()`

**Installing Plotly Online: (use if you want to host graphs in plotly account)**

- How to: You must create an account to run:
  1. Set up an account at plot.ly
  2. Get a User ID and API keys
  3. Sign keys into the account.
Plotly - Alternatives (Bokeh, D3.js)

**Bokeh:**
- Bokeh is an interactive visualization Python library.
- Provides elegant and concise construction of versatile graphics.
- Usage: Can be used in Jupyter Notebooks and can provide high-performance interactive charts and plots.

**D3.js:**
- D3.js (used with Flask) is a framework used with HTML, CSS, and Javascript together to create visualizations.
- Usage: Use D3.js build-in data-driven transitions for extra customization and elevated visualization for your data.
- Pro: Helps build type of framework you want (Plotly uses D3.js library, here you can use the D3.js library itself; open-source)
- Con: High learning curve; you need to learn HTML, CSS, Javascript
Bokeh - Example

Example of using Bokeh from [article](#). Screenshots of interactive features that Bokeh offers:
Tableau

https://www.tableau.com/
Tableau: Intro & Setup

What Are Dashboards:
- Dashboards act as a data visualization tool where users can easily analyze trends and statistics. It can be a powerful way of communicating results of a Data Science project.
- Examples: Dash by Plotly, Bokeh Dashboards, Google Data Studio, Tableau

About Tableau (Tableau Desktop):
- Pros: Makes the charts and interface almost seamlessly.
- Con: Getting used to the interface and functions.
- Con: Data cleaning/pre-processing easier in Python.

Setting up:
- 1-year free trial of Tableau Desktop for Students. (Paid differs by individual vs organization.)
- Tableau Public (create separate account); share data visualizations with global community.
- Introductory videos are a great resource; robust and go through examples in detail.
Tableau - Tableau Desktop (for Students)

Go to this link to try out a trial: https://www.tableau.com/academic/students

We offer free one-year Tableau licenses to students at accredited academic institutions through our Tableau for Students program. Receive access to our entire eLearning suite once verified.
Tableau - Tableau Desktop (for Students)

When you download the Tableau Desktop Application (MacBook Pro):
Explore: No-Code Visualization Tools

**Infogram:** | [https://infogram.com/app/](https://infogram.com/app/)
- Web-based visualization environment; *infographic environment*.
- Multiple PDF/PNG or HTML-based templates; interactivity built-in.
- Paid version offers: Engagement analytics, team collaboration, consistent product branding.

**Flourish:** | [https://flourish.studio/examples/](https://flourish.studio/examples/)
- Another web-based visualization environment.
- Interest: Interface is pretty straightforward, and visualizations can be really interactive.
- Note: Best for *spreadsheet junkies!*

**Datawrapper:** | [https://www.datawrapper.de/](https://www.datawrapper.de/)
- Web-based visualization and map creation environment.
- **Niche service**, offers some powerful capabilities.
- Fact: Interesting workflow.
## Visualization Tools Comparison

<table>
<thead>
<tr>
<th></th>
<th>Data import &amp; usage</th>
<th>Viz options &amp; customization</th>
<th>Free/paid features</th>
<th>More or less technical?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tableau</strong></td>
<td>- Can import from many data types.</td>
<td>- Many graph options.</td>
<td>- Tableau Public</td>
<td>- More technical due to interface and multitude of options.</td>
</tr>
<tr>
<td></td>
<td>- Robust manipulation.</td>
<td>- Experienced users understand benefit.</td>
<td>- Tableau Desktop (1-Year free trial student)</td>
<td></td>
</tr>
<tr>
<td><strong>Infogr.am</strong></td>
<td>- Can import from some data types.</td>
<td>- Many <strong>infographic</strong> visual options.</td>
<td>- Free w/ account;</td>
<td>- Less technical</td>
</tr>
<tr>
<td></td>
<td>- Some manipulation.</td>
<td>- Drag &amp; drop interface.</td>
<td>- Make publicly available PDF, PNG or HTML</td>
<td>- No code; interface accessible to all.</td>
</tr>
<tr>
<td><strong>Flourish</strong></td>
<td>- Import from Microsoft Excel, CSV, JSON.</td>
<td>- <strong>Graph</strong>, <strong>infographic and slide</strong> options.</td>
<td>- Free w/ account;</td>
<td>- Less technical</td>
</tr>
<tr>
<td></td>
<td>- Some manipulation.</td>
<td>- Straightforward editing interface.</td>
<td>- Embed, PDF, PNG, or HTML</td>
<td>- No code; interface accessible to all.</td>
</tr>
<tr>
<td><strong>Datawrapper</strong></td>
<td>- Import from multiple sources.</td>
<td>- <strong>Static graph</strong> options.</td>
<td>- Free (no account need)</td>
<td>- Less technical.</td>
</tr>
<tr>
<td></td>
<td>- Minimal manipulation.</td>
<td>- <strong>Streamlined process</strong> of creating visualizations</td>
<td>- PDF, PNG, or HTML</td>
<td>- Frequently used by journalists/newspapers.</td>
</tr>
</tbody>
</table>

---

**Berkeley SCET**

<table>
<thead>
<tr>
<th>Matplotlib</th>
<th>Seaborn</th>
<th>Plotly</th>
<th>Tableau</th>
<th>Resources</th>
</tr>
</thead>
</table>

---
References

Data Visualization - References