Data Visualization in Python

It is expected that you already know and use both pandas and matplotlib in your work as a data scientist. For those of you who are comfortable with plotting—meaning that you can readily produce any of the several dozen types of common plots used in data science—then this unit will serve as a refresher on best practices. This module is not a comprehensive overview of the data visualization landscape. We have touched on essential tools and best practices. If necessary, use the resources below to improve your data visualization skills and gain more context.

If you would like additional context a few links are available below:

- Anaconda's article on 'moving toward convergence
- Anaconda's article on the future of Python visualization libraries
- Tutorials for matplotlib

Best practices

Best practices as a data scientist generally require that all work be saved as text files:

1. Simple scripts
2. Modules
3. Python package

Remember to save a maximum amount of code within files, even when using Jupyter. Version control is a key component to effective collaboration and reproducible research.

- Nature article on Git and scientific reproducibility

Essentials of matplotlib

Matplotlib has a "functional" interface similar to MATLAB® that works via the pyplot module for simple interactive use, as well as an object-oriented interface that is more pythonic and better for plots that require some level of customization or modification. The latter is called the artist interface. There is also built in functionality from within pandas for rapid access to plotting capabilities.

- matplotlib tutorial on lifecycle of a plot
- pandas visualization
- matplotlib pyplot interface
- matplotlib artist interface

Other visualization tools

Bokeh - Bokeh supports streaming, real-time data and is used to create interactive, web-ready plots, which can output as JSON objects, HTML documents, or interactive web applications.

plotly - While plotly is widely known as an online platform for data visualization, it can be accessed from a Python notebook. Like Bokeh, Plotly's strength lies in making interactive plots, and it offers some charts not found in most libraries, like contour plots.

dash - Dash is Python framework for building web applications. It built on top of Flask, Plotly.js, React and React Js. It enables you to build dashboards using pure Python.

ggplot - ggplot is a Python visualization library based on R's ggplot2 and the Grammar of Graphics. It lets you construct plots using high-level grammar without thinking about the implementation details.

folium - Based on leaflet.js, it is a great tool for working with geographic data for example choropleth visualizations.

Analytics dashboard in IBM Watson Studio - You can build sophisticated visualizations of your analytics results, communicate the insights that you've discovered in your data on the dashboard and then share the dashboard with others.