04 Data Visualization

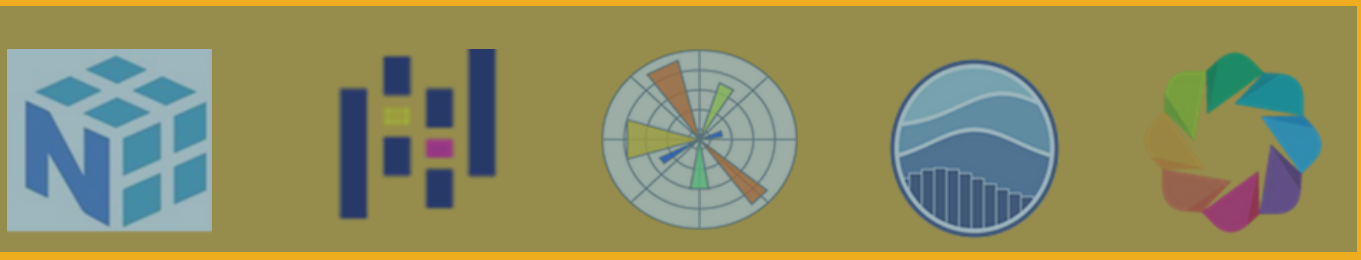
Visualization Objectives

Record information, analyze data to support reasoning, confirm hypotheses, communicate ideas to others

Why Visualization

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

Useful Python Libraries



Numpy Pandas Matplotlib Seaborn Bokeh

Matplotlib

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

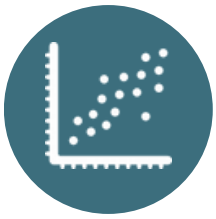
Seaborn

- Used for statistical data visualization
- Uses fewer syntax with good default themes
- Integerated to work great with pandas dataframe
- Uses matplotlib under the hood

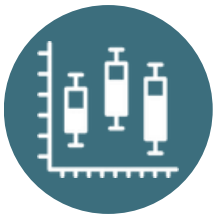
Types of Plots

Line Plots

Bar Plots



Scatter Plots



Box Plots

Histogram

Line Plots

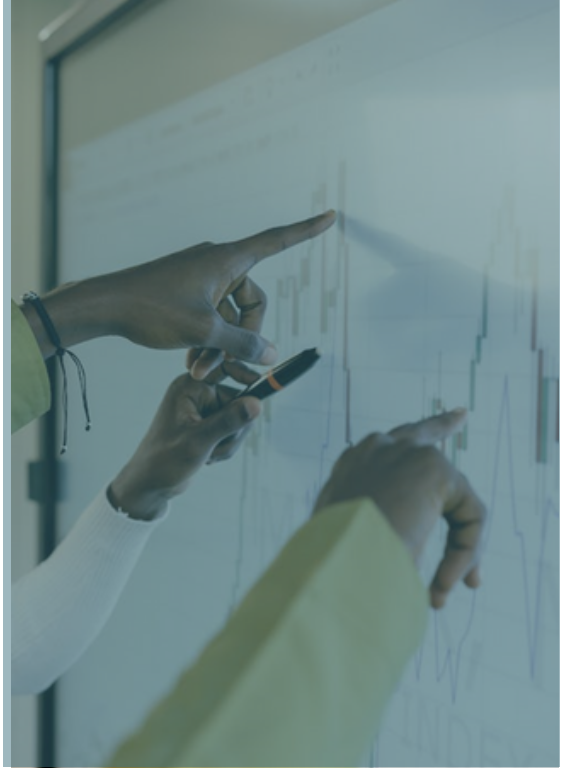
- Used for numeric data,
- Used to show trends
- Compare two or more different variables over time
- Could be used to make predictions

Bar Plots

- Used for nominal or ordinal categories
- Compare data amongst different categories
- Ideal for more than 3 categories
- Can show large data changes over time

Scatter Plots

- Used to visualize relation between two numeric variables
- Used to visualize correlation in a large data set
- Predict behavior of dependent variable based on the measure of the independent variable
- A box-plot is a graph that gives you a good indication of how the values in date are spread out.
- Statistical graph used on sets of numerical data.
- Shows the range, spread, and center.



Histogram

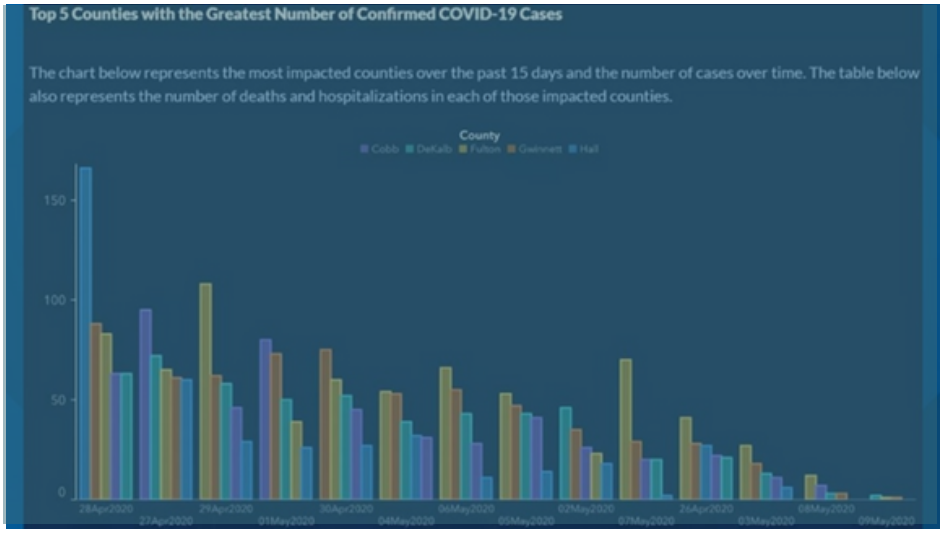
- Used for continous data
- Displays the frequency distribution (shape), summarize large data sets graphically
- Compare multiple distributions.

Choosing the Right Chart

After you learn some kind of the plot types, you see that you should choose the right chart to ensure the information can be delivered clearly.

Avoid Deception

Even if a data scientist is careful to choose the right chart for the right data, there are plenty of ways that data can be displayed in a way to prove a point, often at the cost of undermining the data itself. There are many examples of deceptive charts and infographics!



As the eye is drawn to the right to conclude that, over time, COVID cases have declined in the various counties. In fact, if you look closely at the dates, you find that they have been rearranged to give that deceptive downward trend.

Picking Color

You saw in the 'Florida gun violence' chart how color can provide an additional layer of meaning to charts. While color meaning might be different in different parts of the world, and tend to change in meaning according to their shade.

Generally speaking, color meanings include:

Color	Meaning
Red	Power
Blue	Trust, loyalty
Yellow	Happiness, caution
Green	Ecology, luck, envy
Orange	Vibrance

Styling Your Charts For Readability

Charts are not meaningfoll if they are not read-able! Tahe a moment to consider styling the width and eight of your chart to scale well with your data. If one variable (such as all of states) need to be displayed, Show them vertically on the Y axis if possible so as to avoid a horizontally-scrolling chart.

Label your axes, provide a legend if necessary and offer toollips for better comprehension of data.

If your data is textual and verbose on the X axis, you can angle the text for better readability.

Reference

- https://www.tutorialspoint.com/python/python_dictionary.htm
- https://www.w3schools.com/python/python_dictionaries.asp

Picture Source

- unsplash.com
- pexels.com
- pixabay.com