Data Visualization

Penyusun Modul: Chairul Aulia
Editor: Rina Fitriyani, Silfa Rahma Aulia

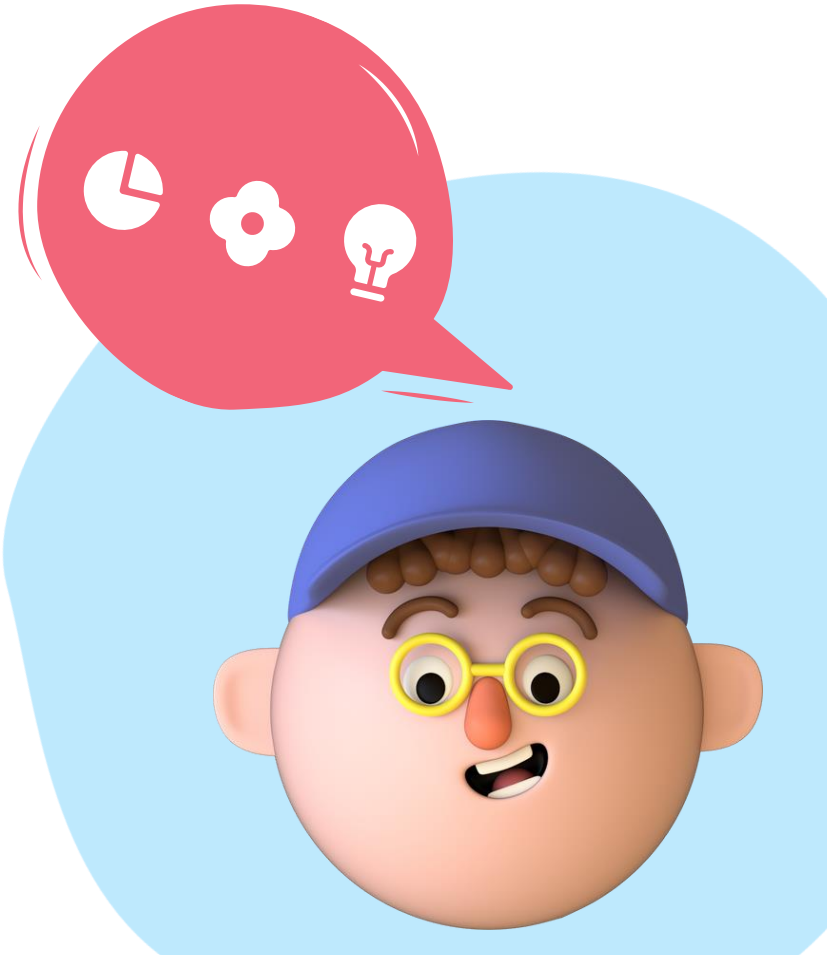
Hold On!

Remember our last topic about **visualization**?

Let’s continue our talk!

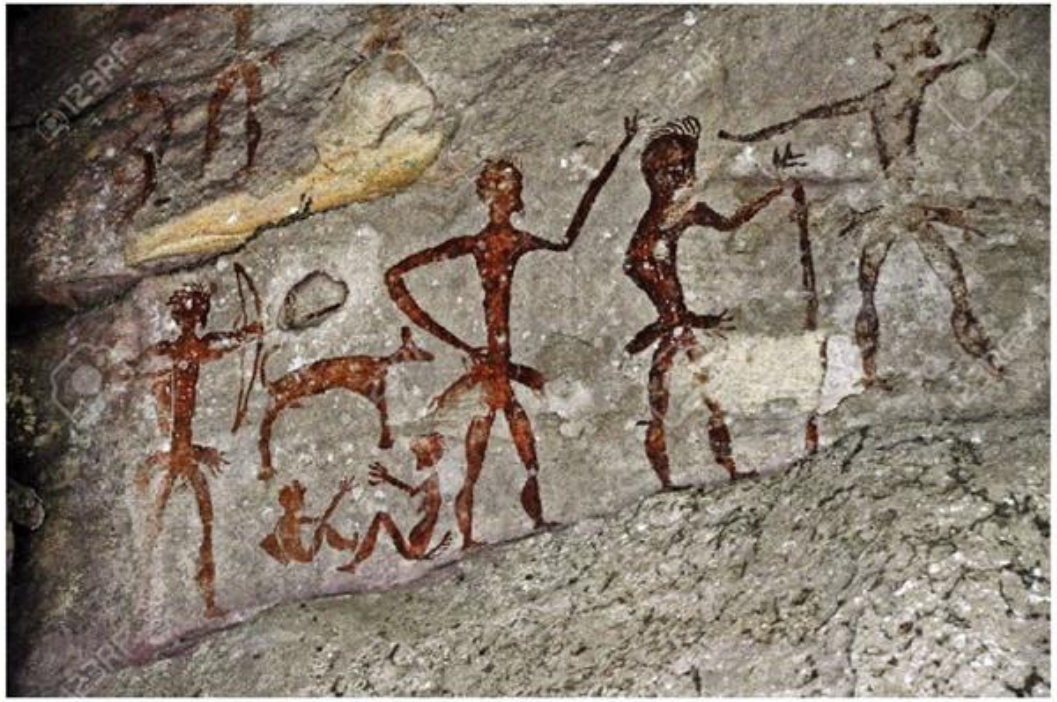
Visualization Objectives

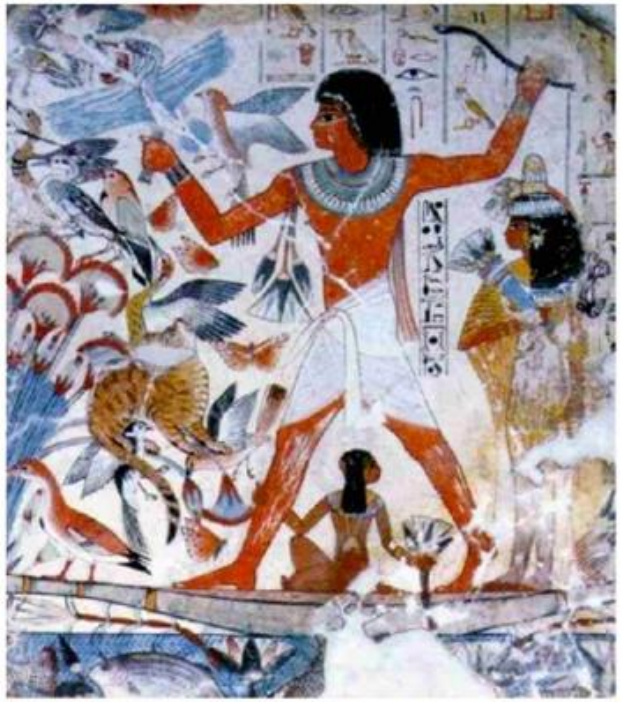
- Record information
- Analyze data to support reasoning
- Confirm hypotheses
- Communicate ideas to others



Why Visualization?

To record information





Why Visualization?

To point out interesting things

MTHIVLWYADCEQGHKILKMTWYN

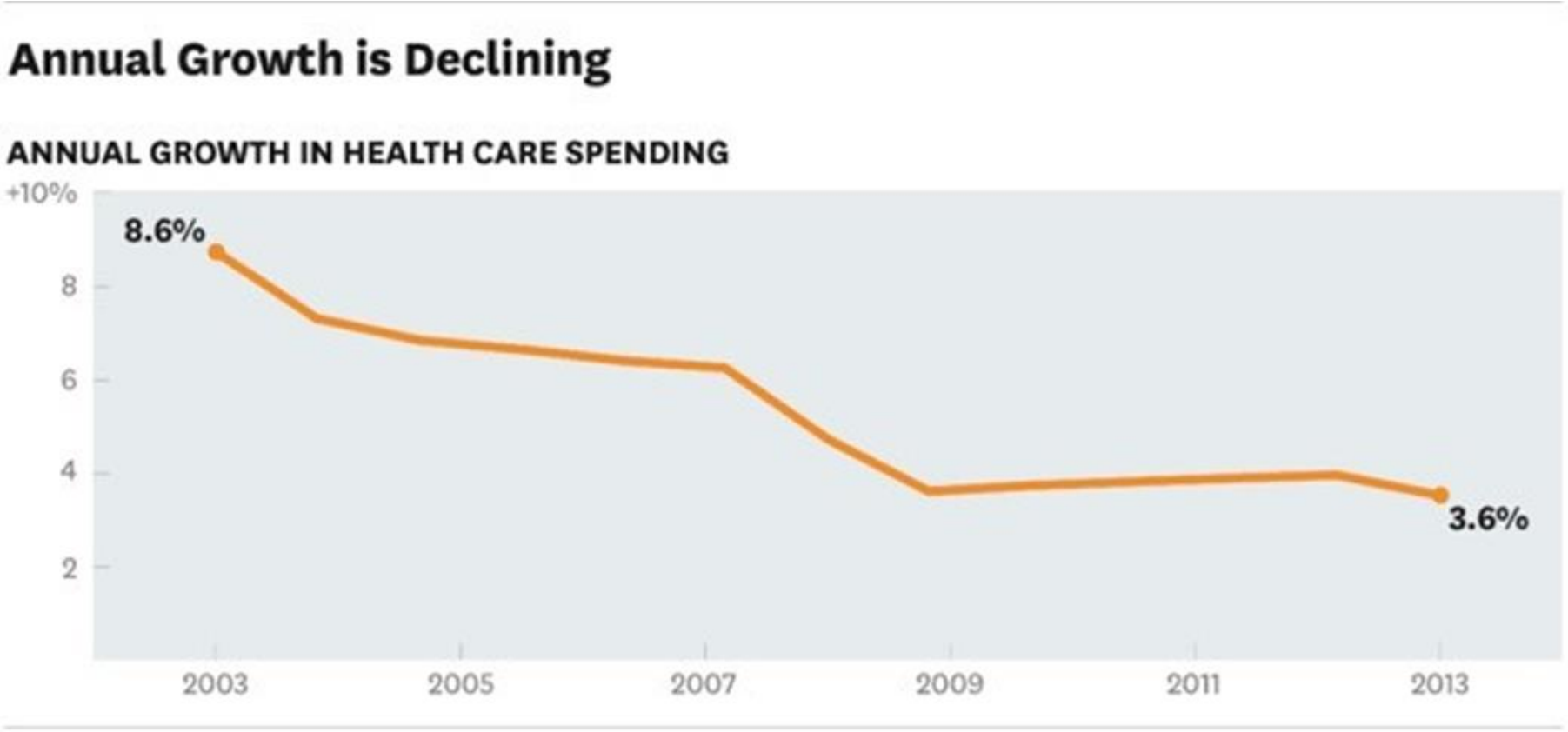
ARDCAIREQGHLVKMFPSTWYARN

GFPSVCEILQGKMFPSNDRCEQDIFP

SGHLMFHKMVPSTWYACEQTWRN

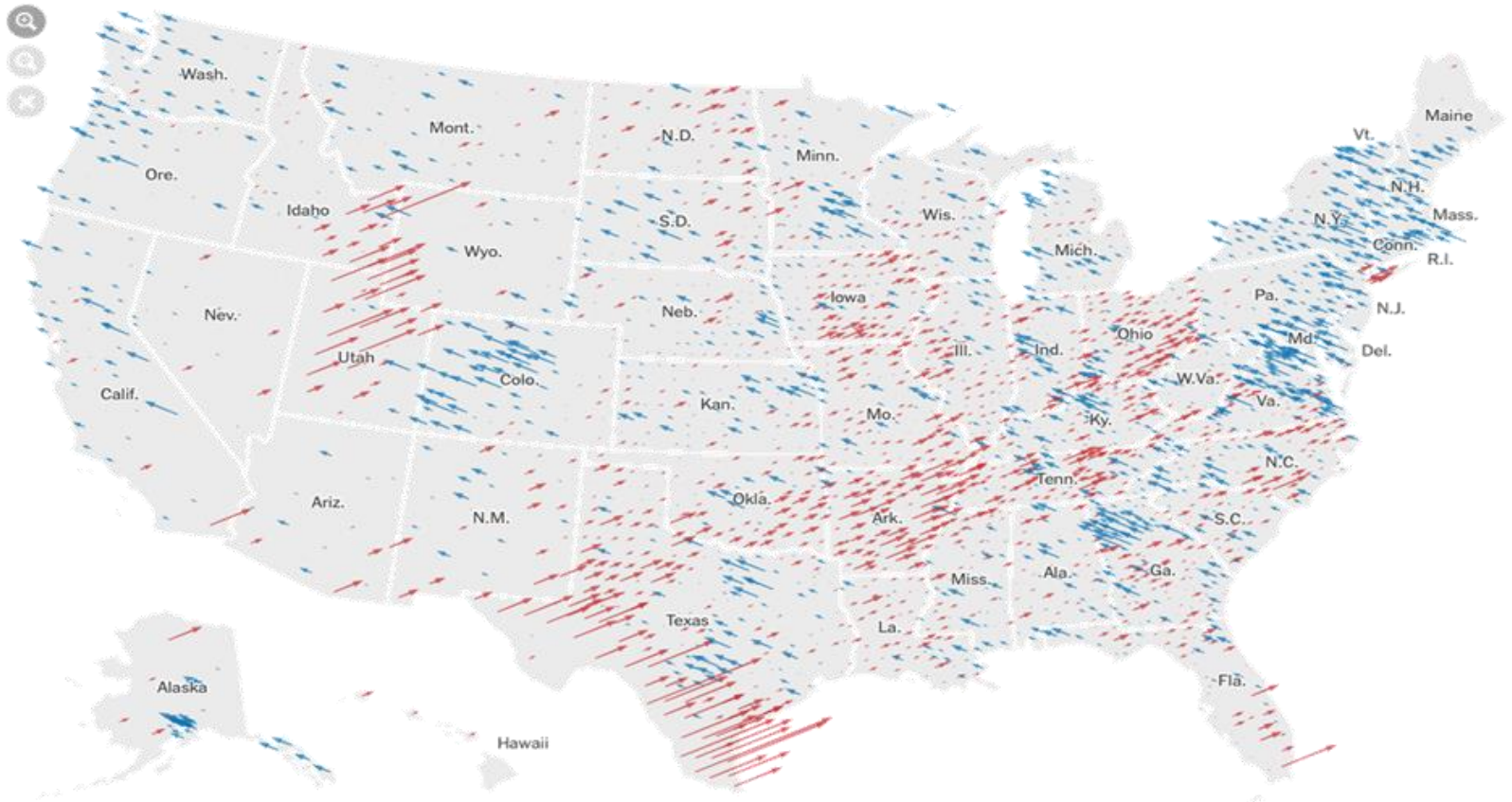
Why Visualization?

To communicate information



Why Visualization?

To analyze data



Useful Python Libraries

Numpy

Pandas

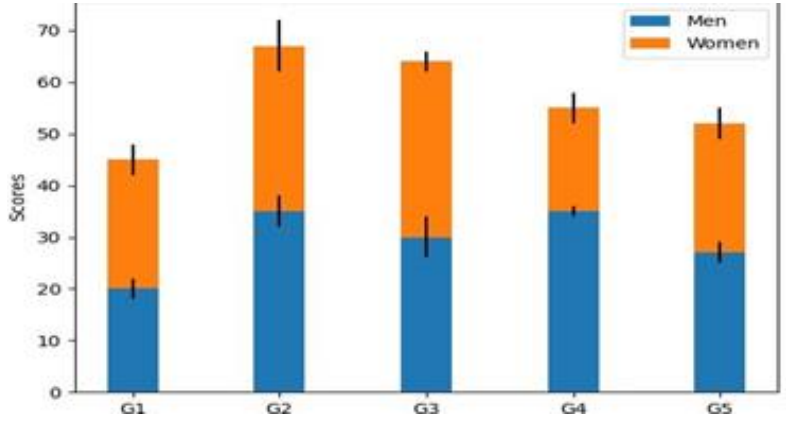
Matplotlib

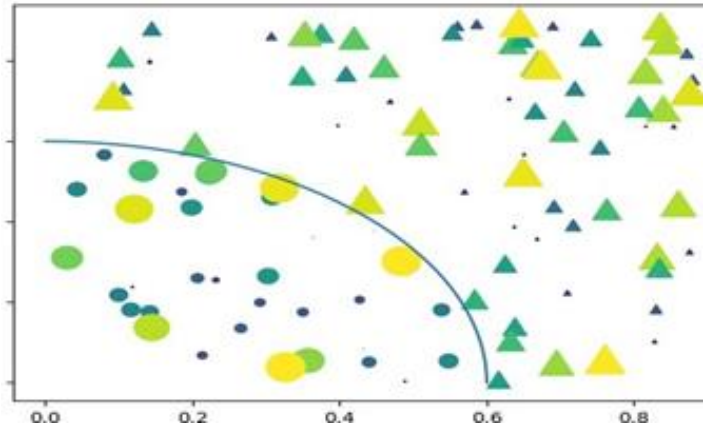
Seaborn

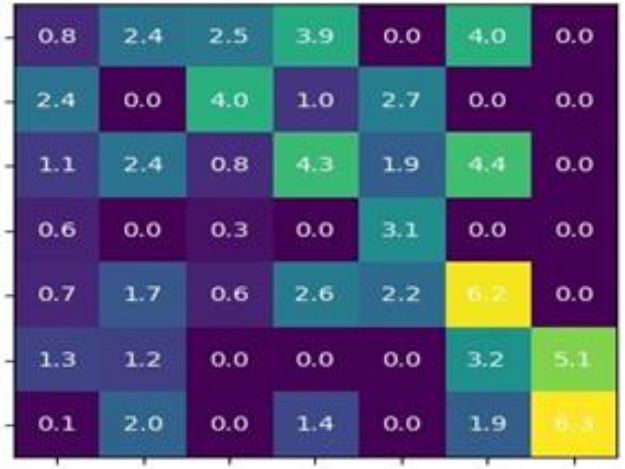
Bokeh

Matplotlib

- Used for basic plotting
- Highly customizable
- Works with Numpy and Pandas

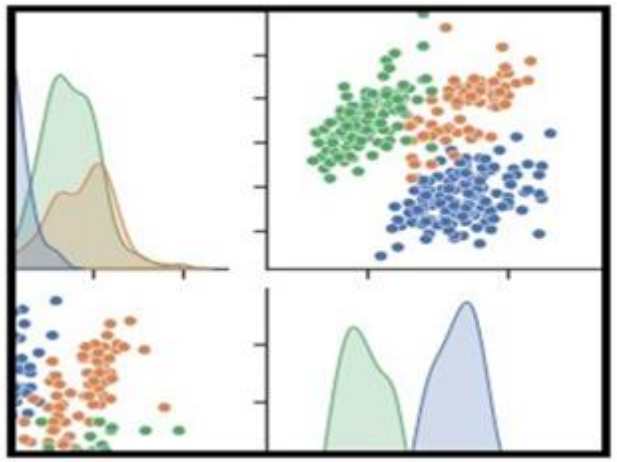


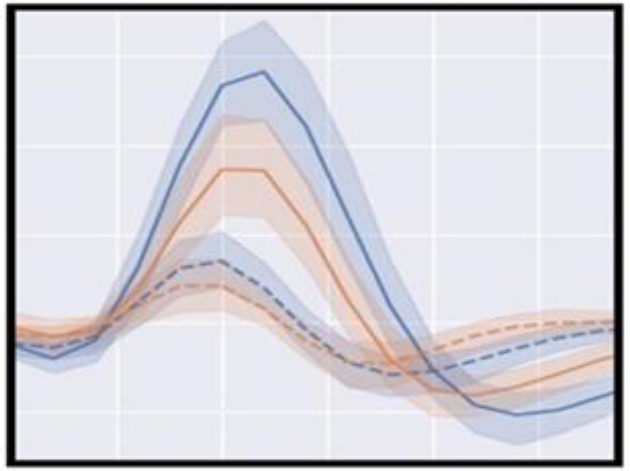


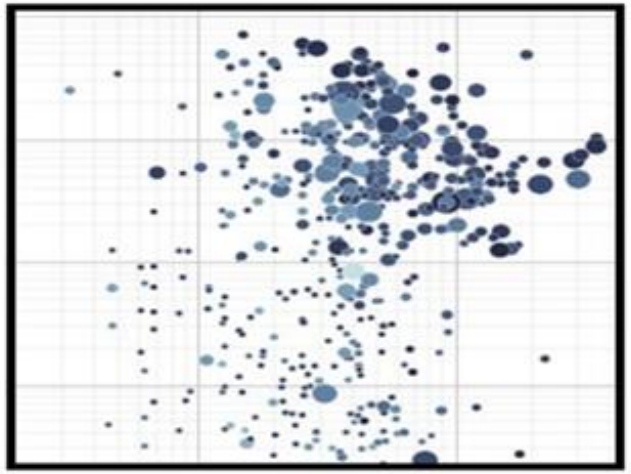


Seaborn

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



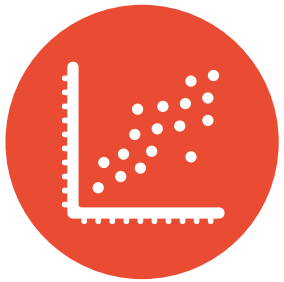




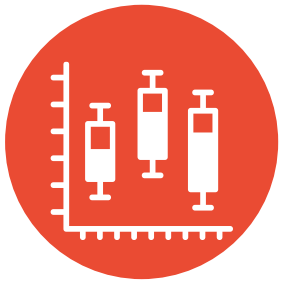
Types of Plots

Line Plots

Bar Plots



Scatter Plots

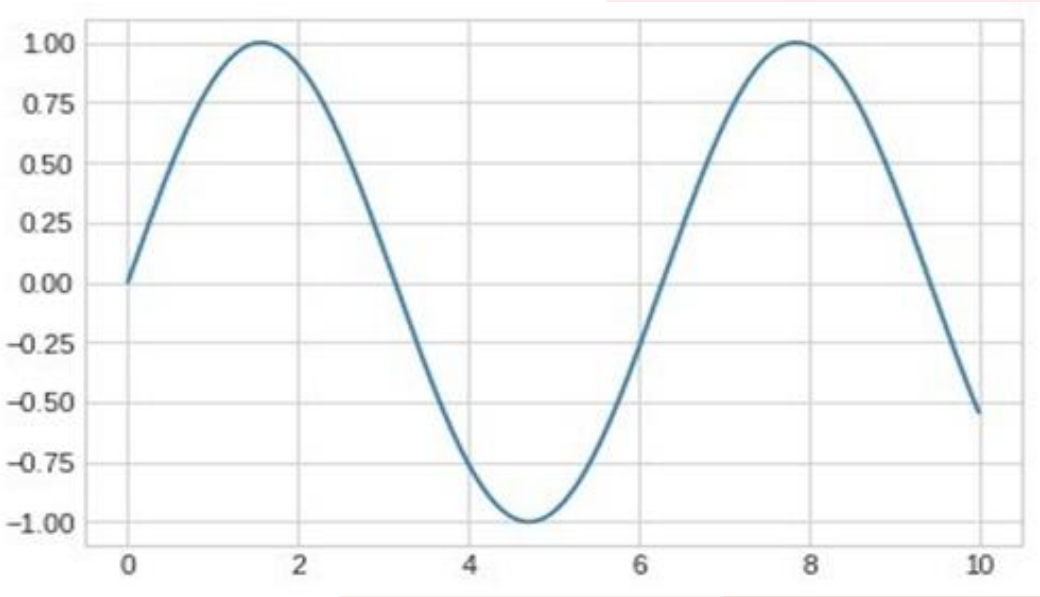


Box Plots

Histogram

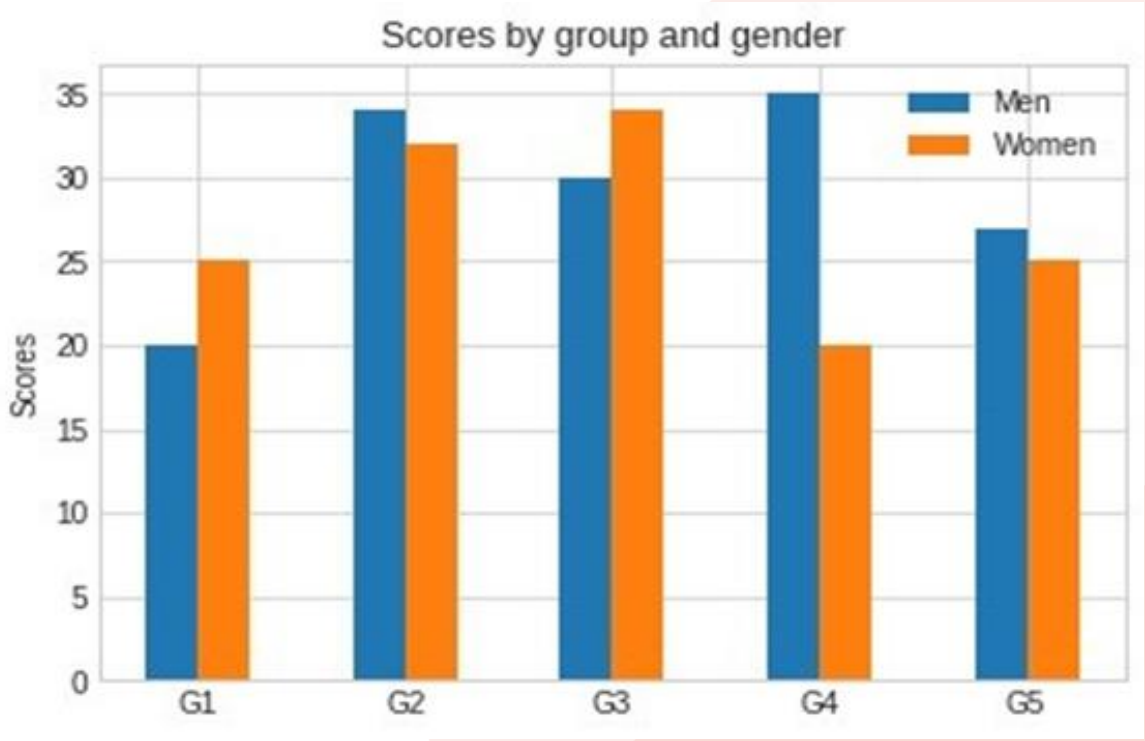
Line Plots

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



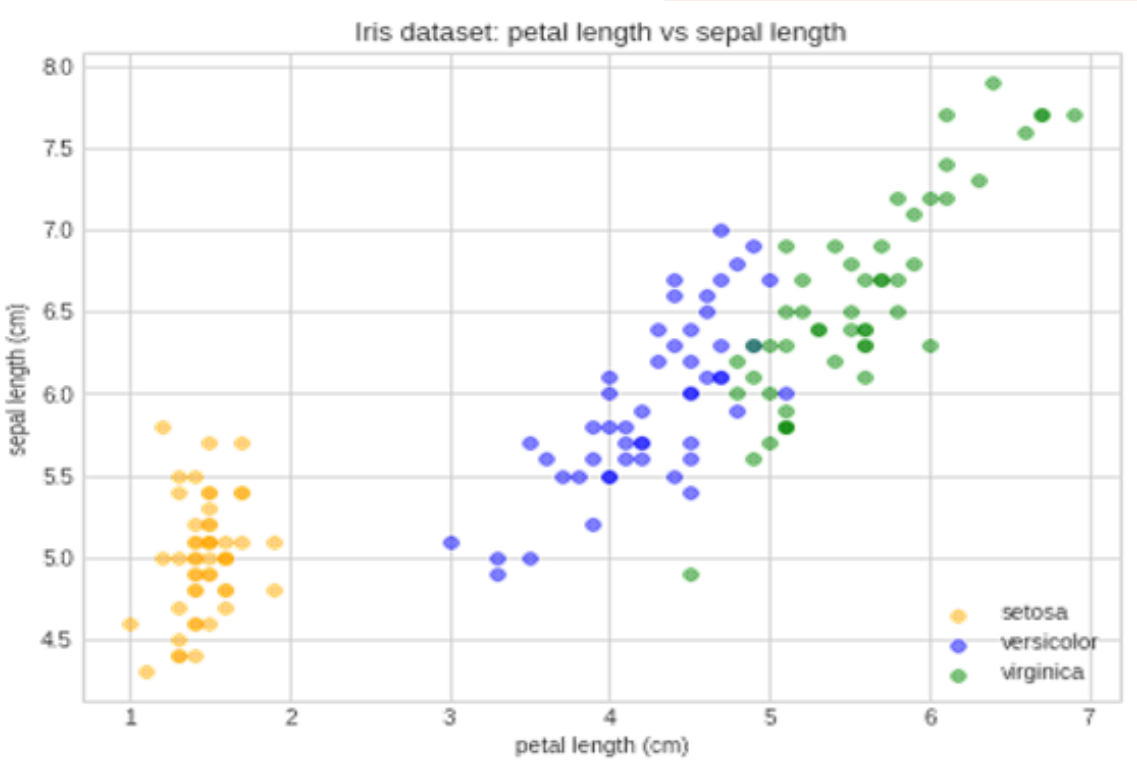
Bar Plots

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



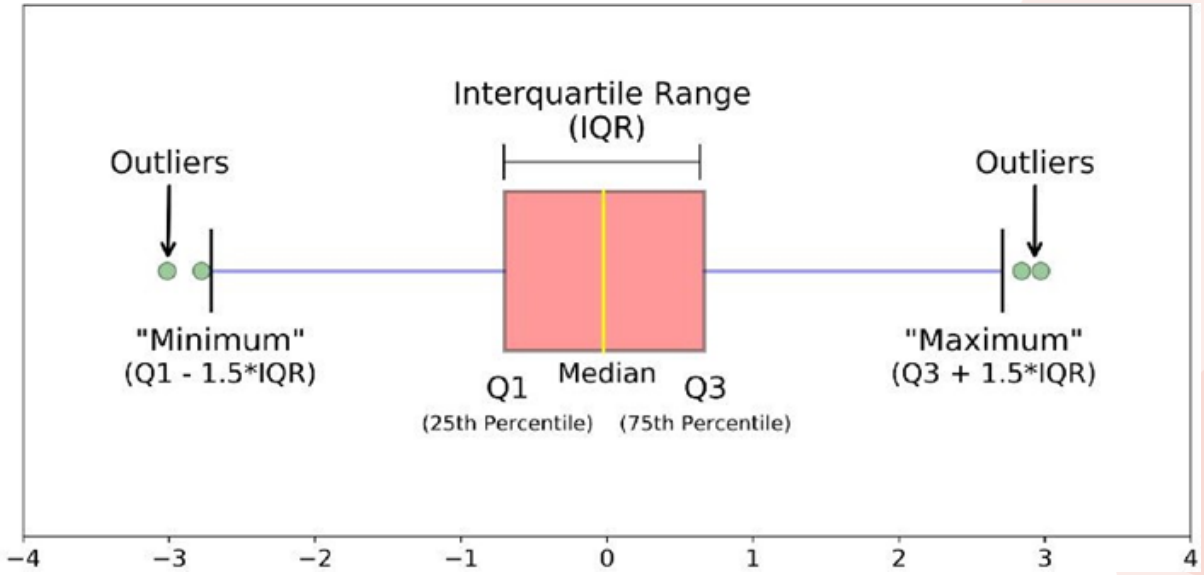
Scatter Plots

- 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.



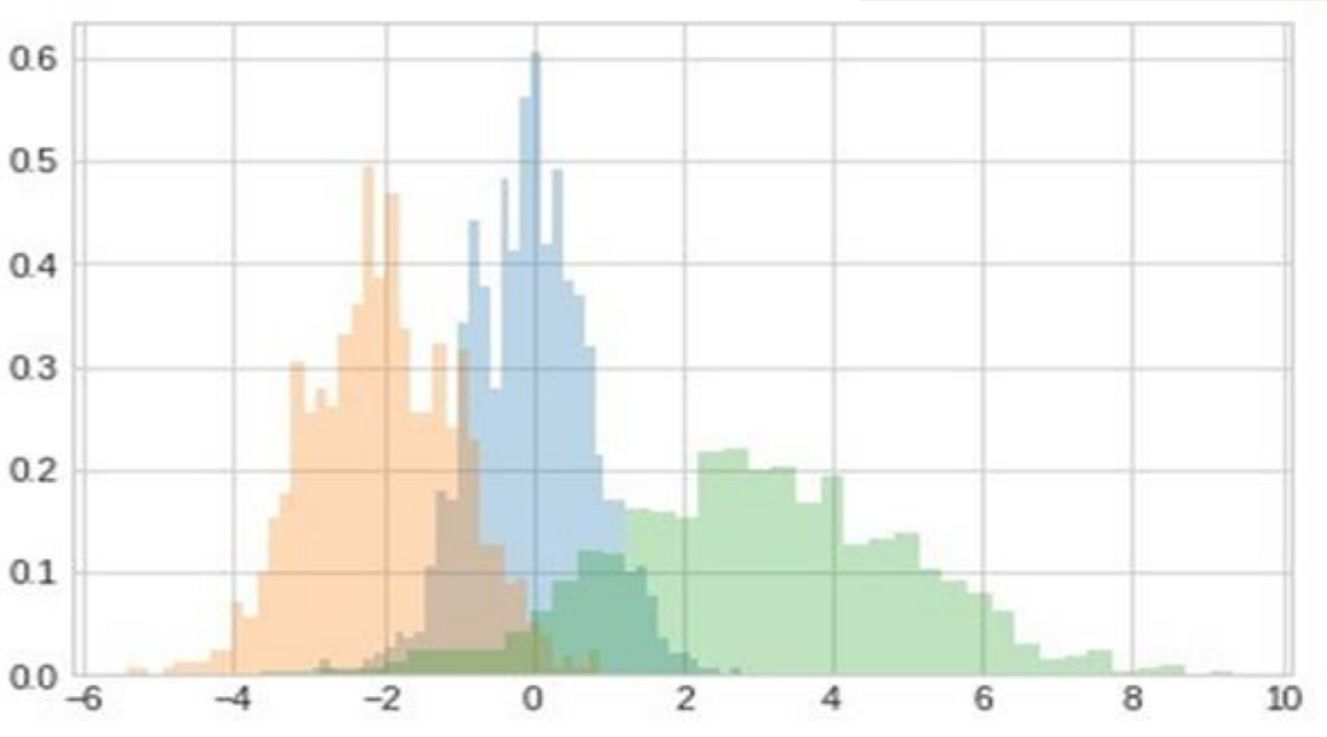
Scatter Plots

- A box-plot is a graph that gives you a good indication of how the values in the data are spread out.
- aka whisker plot
- Statistical graph used on sets of numerical data
- Shows the range, spread, and center



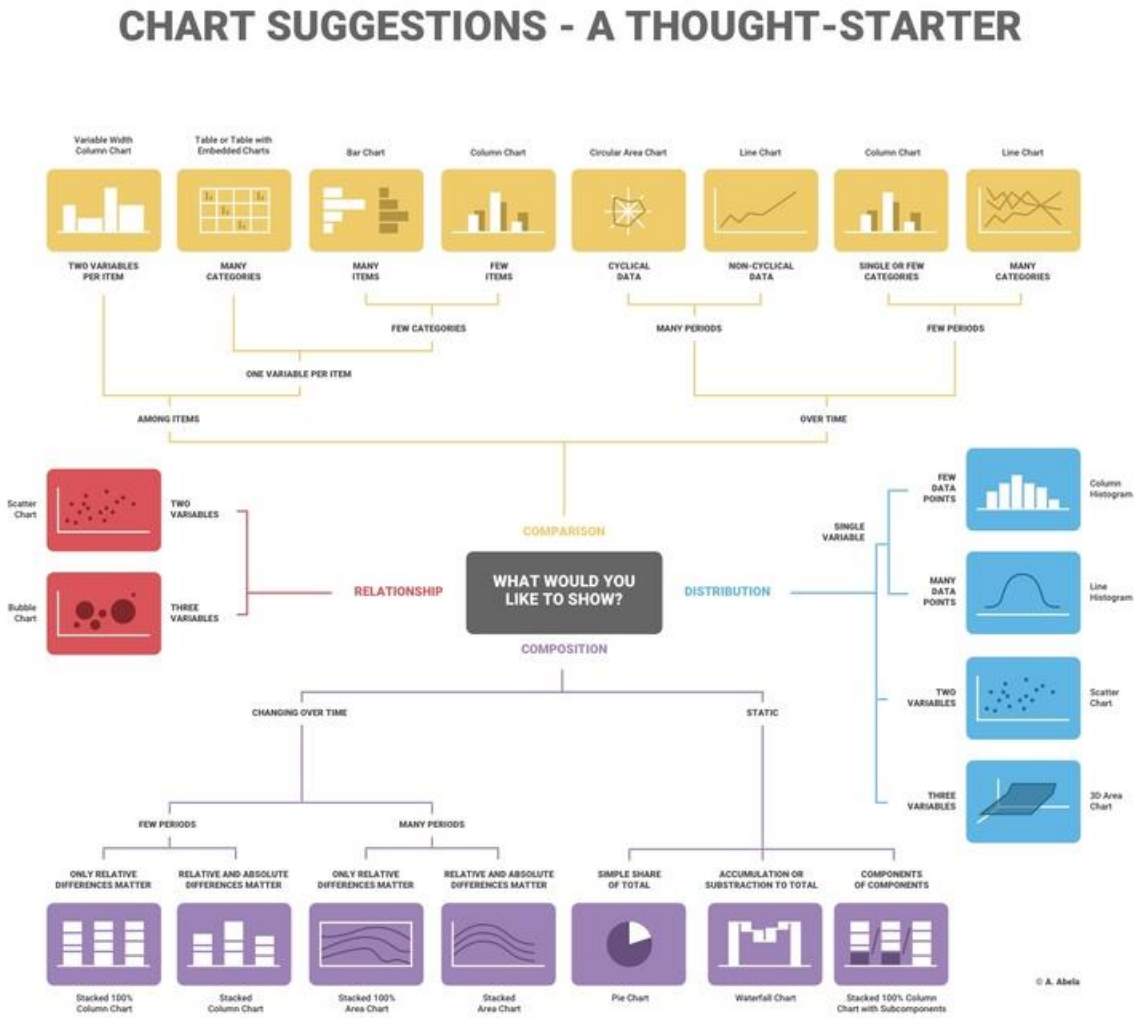
Histogram

- Used for continuous data
- Displays the frequency distribution (shape)
- Summarize large data sets graphically
- Compare multiple distributions



Choosing the Right Chart

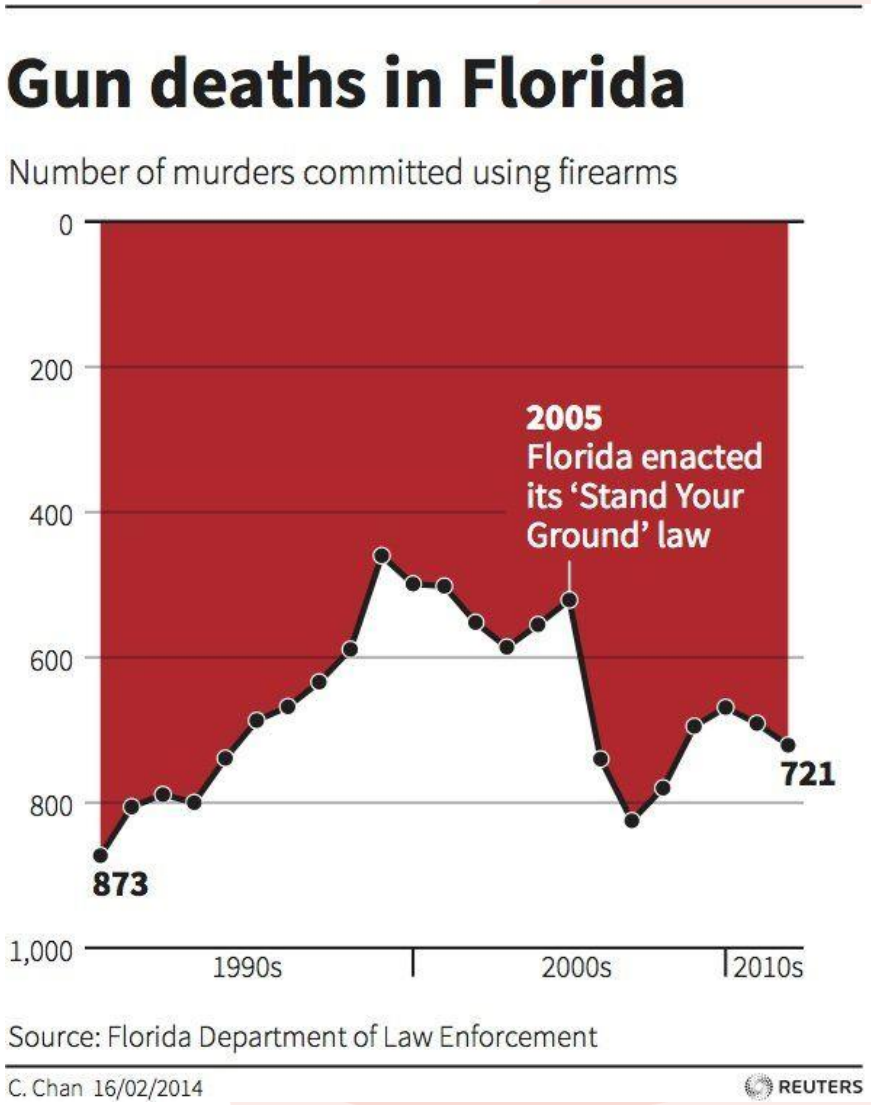
After you learn some kind of the plot types, you see that you should choose the right chart to ensure the information can be delivered clearly.



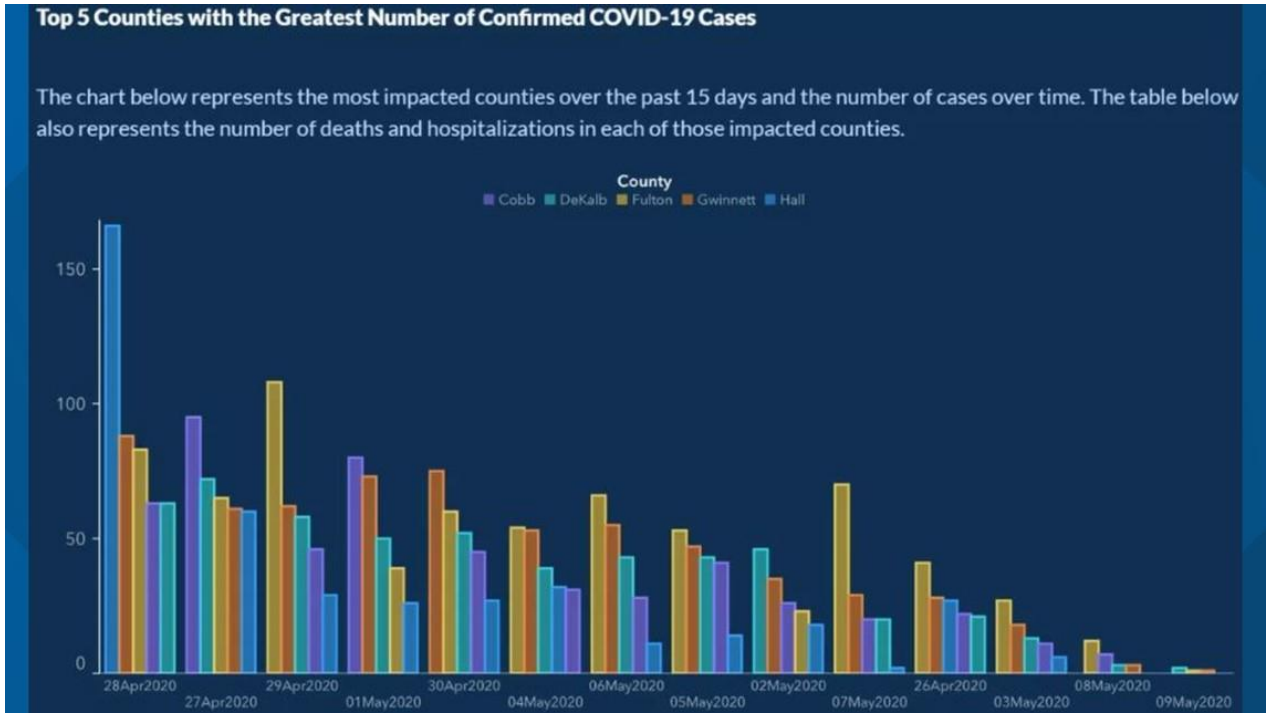
https://infogram.com/page/choose-the-right-chart-data-visualization

Avoid Deception

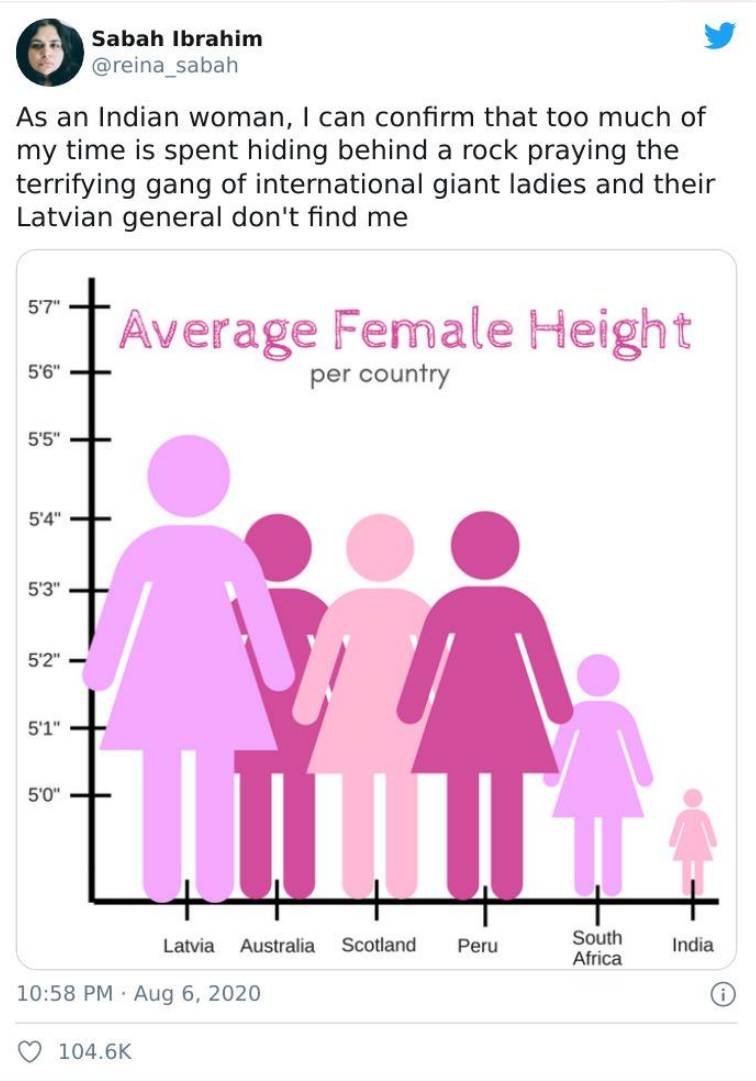
Even if a data scientist is careful to choose the right chart for the right data, there are plenty of ways that data can be displayed in a way to prove a point, often at the cost of undermining the data itself. There are many examples of deceptive charts and infographics!



Avoid Deception



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.



This strange chart shows how proportion can be manipulated, to hilarious effect:

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:

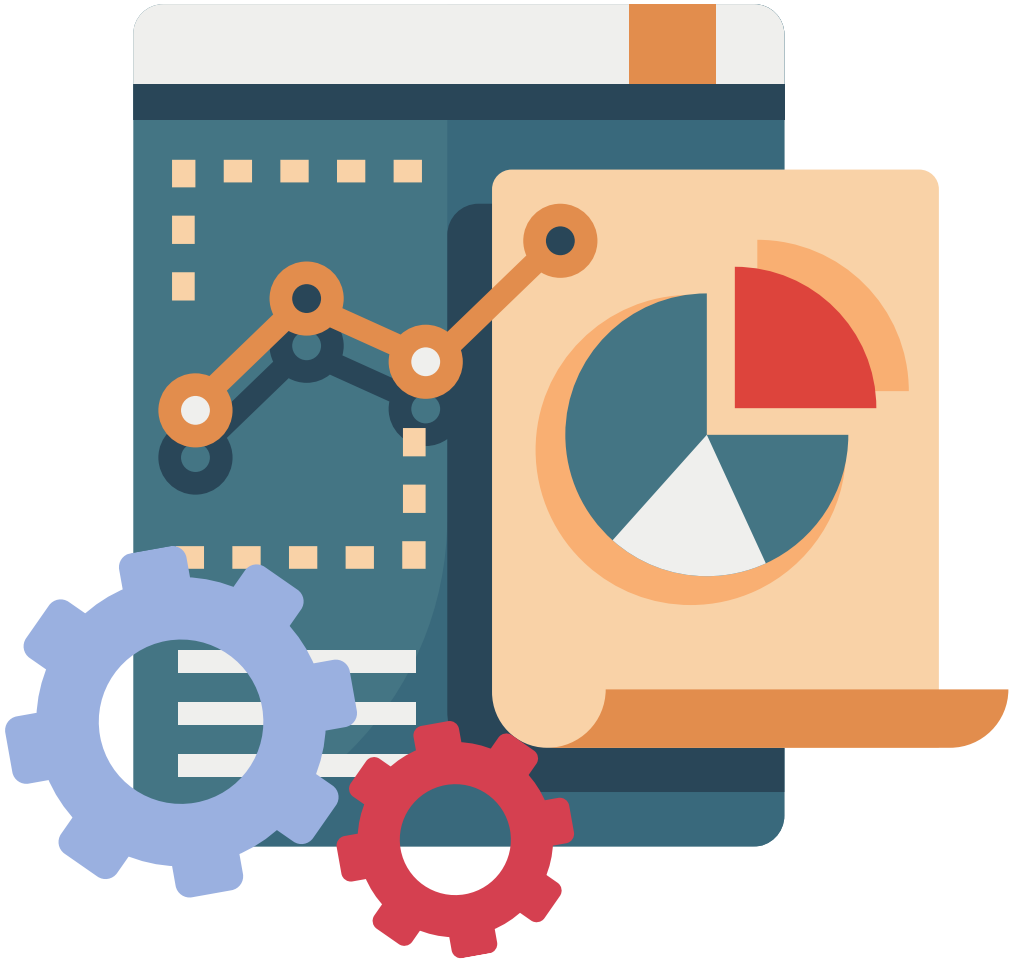
Color	Meaning
Red	Power
Blue	trust, loyalty
Yellow	happiness, caution
Green	ecology, luck, envy
orange	vibrance

Styling your charts for readability

Charts are not meaningful if they are not readable! Take a moment to consider styling the width and height of your chart to scale well with your data. If one variable (such as all 50 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 tooltips for better comprehension of data.

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



As usual,
Let's try it out!

Open the Data_Visualization notebook file that you've got on JupyterLab and start exploring

Exercise

Now, try doing the same thing with the winequality-white dataset

Of course, you are free to explore the plots and variables