DA1: DESIGNING APPLICATIONS IN PYTHON
LECTURE #4
DATA VISUALIZATION

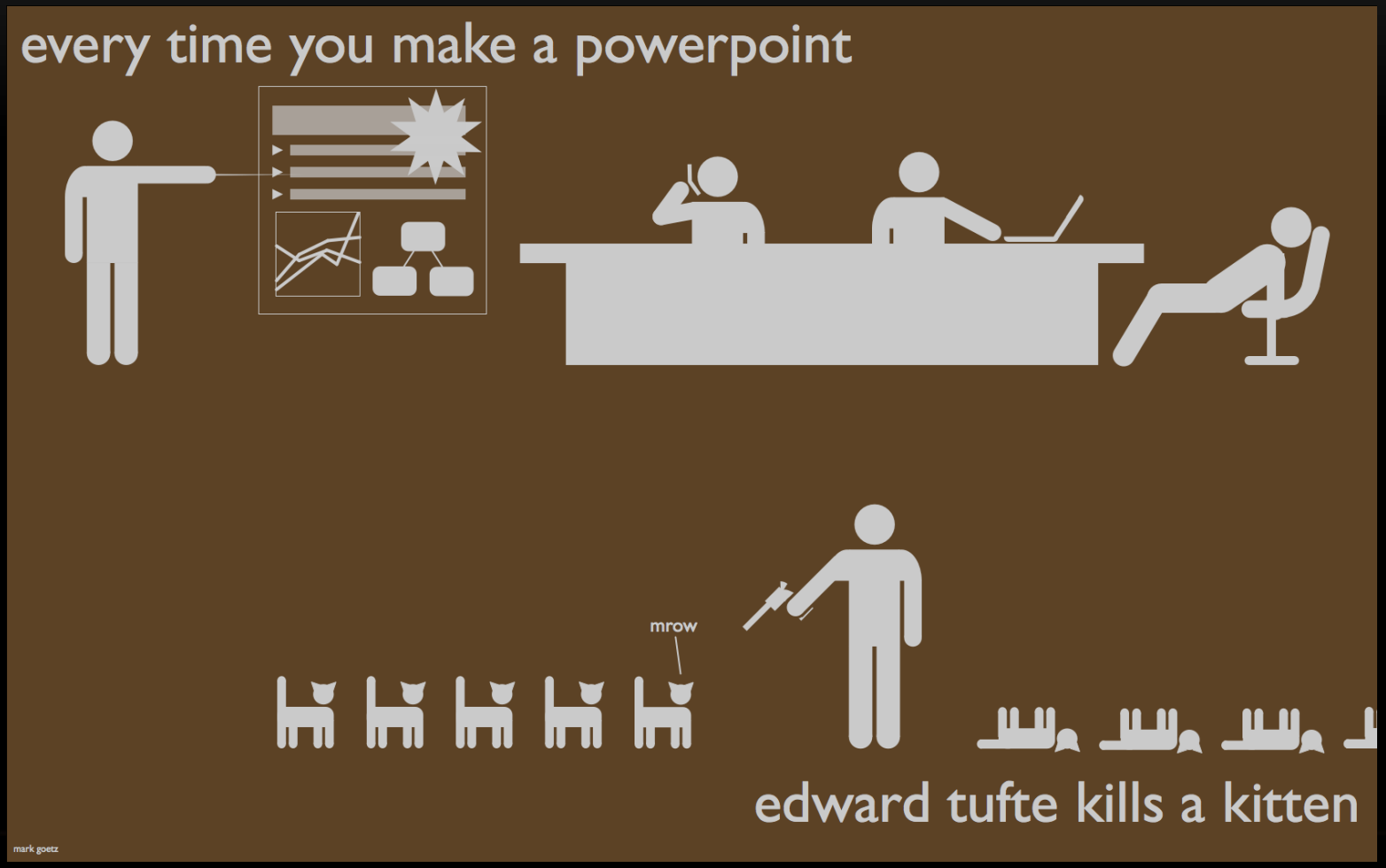
Aidos Sarsembayev

CONTENT

- Charts types
- How to pick the right one?
- Visualization libraries in Python

Making sense of facts, numbers, and measurements is a form of art – the art of **data visualization**. There is a load of data in the sea of noise. To turn your numbers into knowledge, your job is not only to separate noise from the data, but also to present it the right way.

POWER POINT


every time you make a powerpoint
mrow
edward tufte kills a kitten
mark goetz

DATA VISUALIZATION. BEST PRACTICES

There are four **basic presentation types** that you can use to present your data:

- Comparison
- Composition
- Distribution
- Relationship

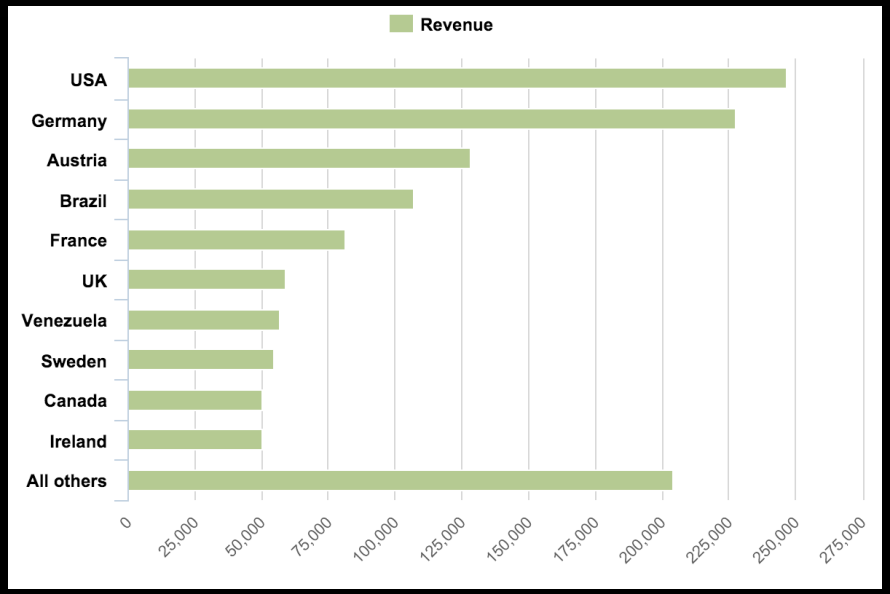
DATA VISUALIZATION. BEST PRACTICES

In order to choose, you have to answer few questions:

- How many variables do you want to show in a single chart? One, two, three, many?
- How many items (data points) will you display for each variable? Only a few or many?
- Will you display values over a period of time, or among items or groups?

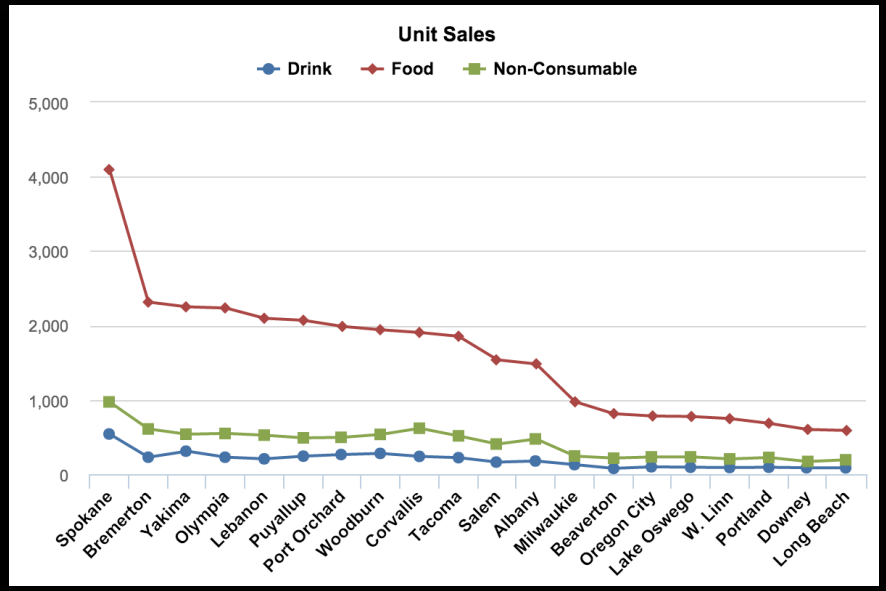
DATA VISUALIZATION. BEST PRACTICES

- Bar charts are good for comparisons,



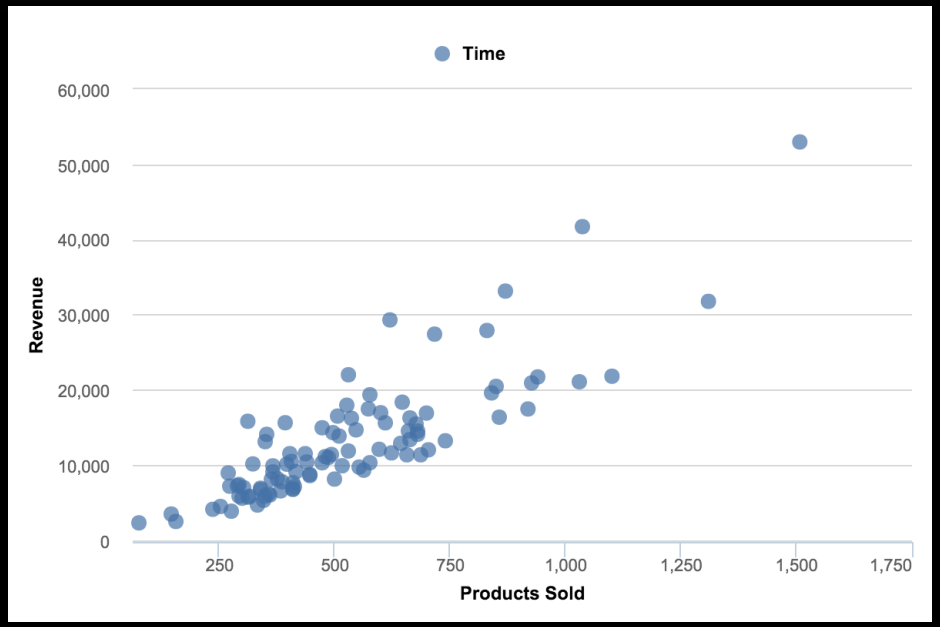
DATA VISUALIZATION. BEST PRACTICES

- Line charts work better for trends.



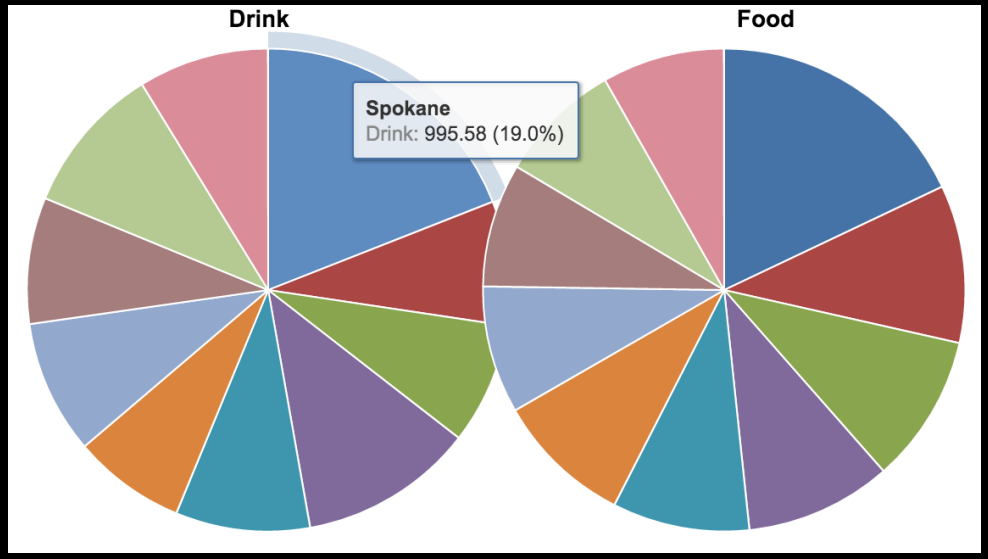
DATA VISUALIZATION. BEST PRACTICES

- Scatter plot charts are good for relationships and distributions

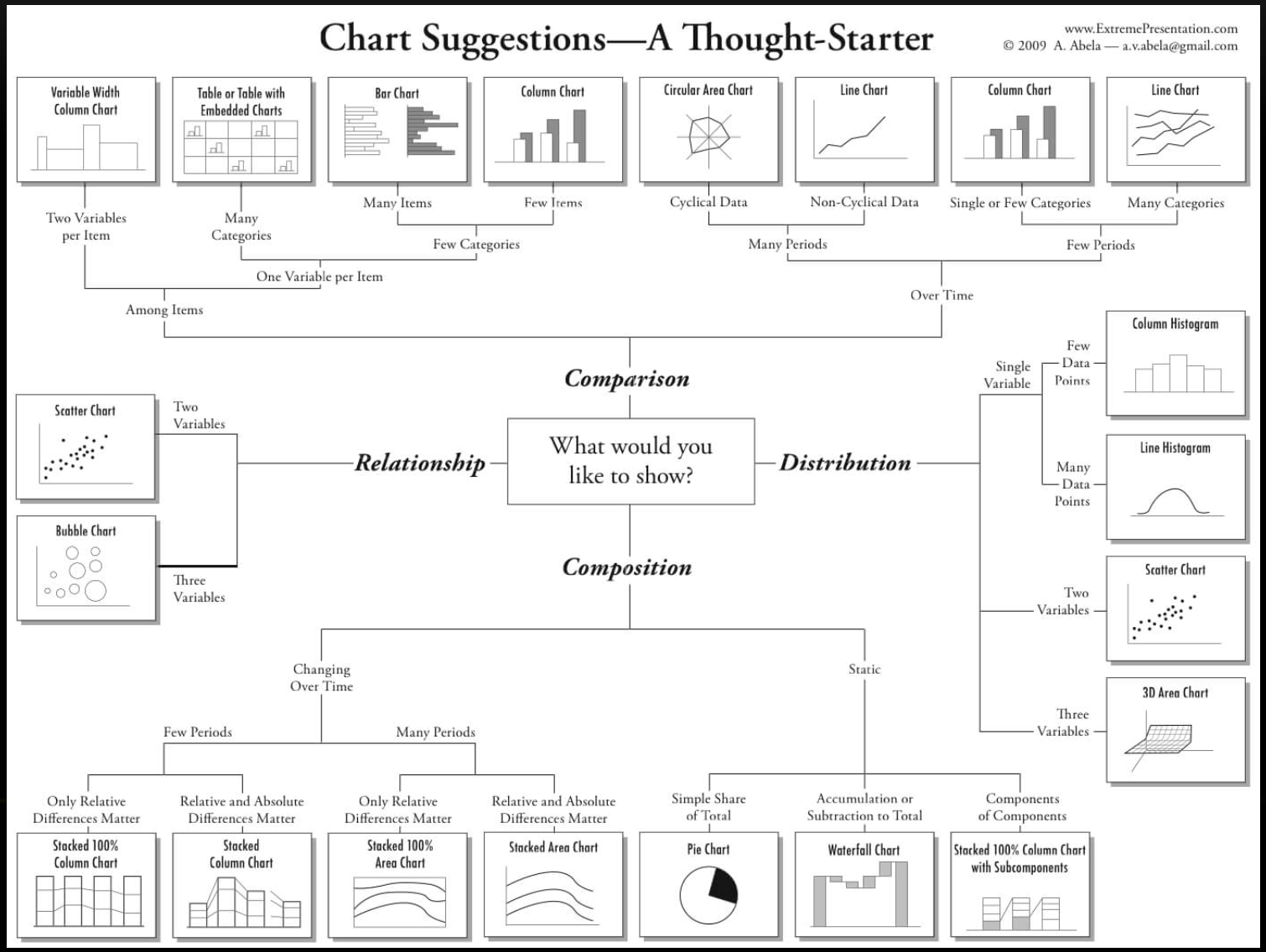


DATA VISUALIZATION. BEST PRACTICES

- Pie charts should be used only for simple compositions — never for comparisons or distributions.

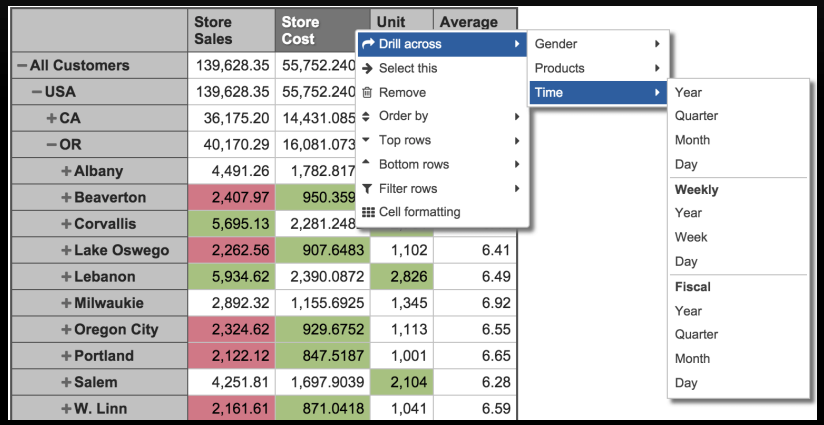


DATA VISUALIZATION. BEST PRACTICES



TABLES

- Tables are essentially the source for all the charts. They are best used for **comparison, composition, or relationship analysis** when there are only few variables and data points. It would not make much sense to create a chart if the data can be easily interpreted from the table.



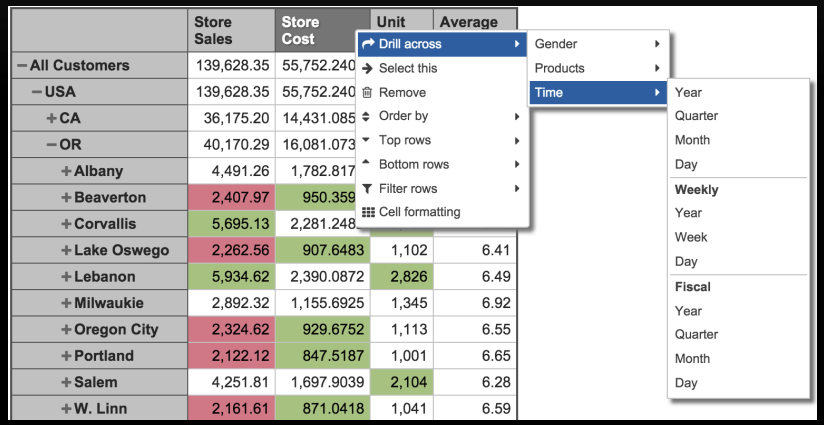
TABLES

Use tables when:

- You need to compare or look up individual values.
- You require precise values.
- Values involve multiple units of measure.
- The data has to communicate quantitative information, but not trends.

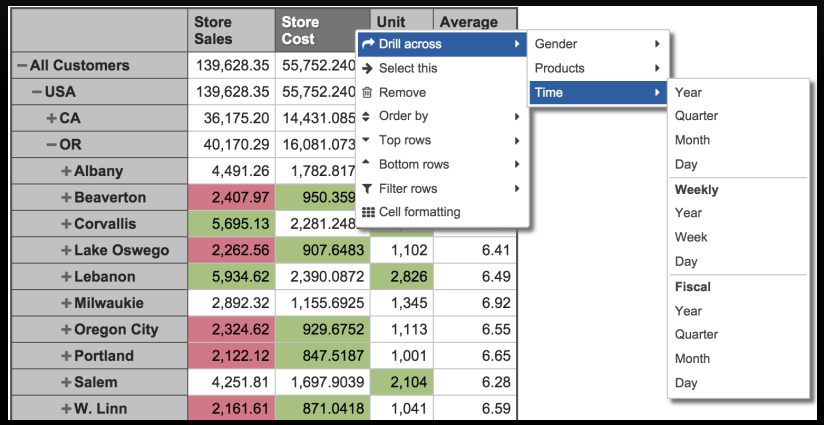
Use charts when the data presentation:

- Is used to convey a message that is contained in the shape of the data.
- Is used to show a relationship between many values.



TABLES

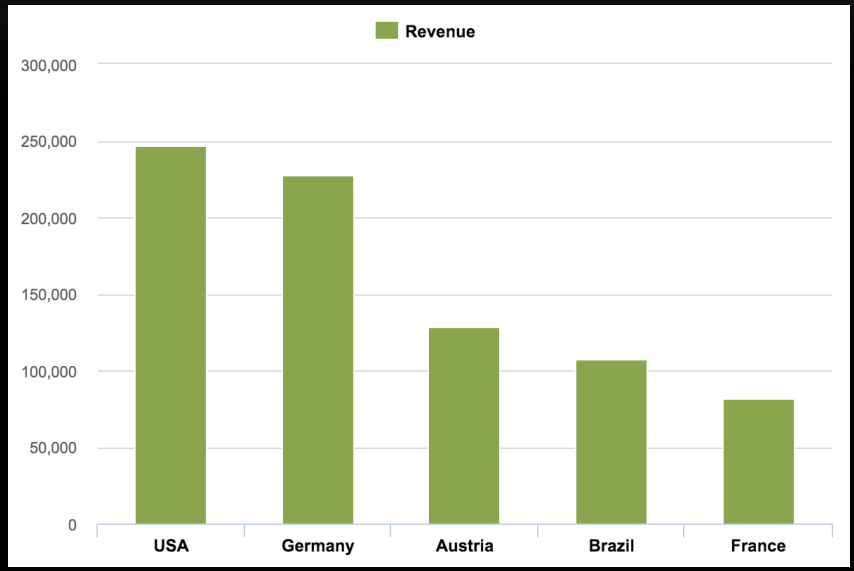
For example, if you want to show the **rate of change**, like sudden drop of temperature, it is best to use a chart that shows the slope of a line because rate of change is not easily grasped from a table.



COLUMN CHARTS

- The column chart is probably the most used chart type. This chart is best used to compare different values when specific values are important, and it is expected that users will look up and compare individual values between each column.

- With column charts you could compare values for different categories or compare value changes over a period of time for a single category.

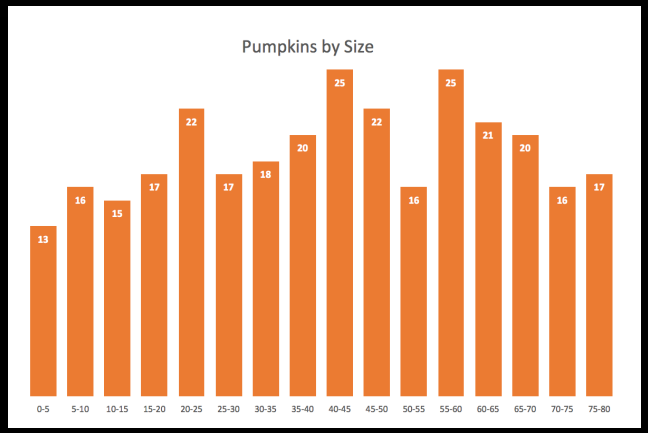


COLUMN CHARTS. BEST PRACTICES

- Use column charts for comparison if the number of categories is quite small — **up to five**, but not more than seven categories.
- If one of your data dimensions is time — including years, quarters, months, weeks, days, or hours — you should always set time dimension on the horizontal axis.
- In charts, time should always run from left to right, never from top to bottom.
- For column charts, the numerical axis must start at zero. Our eyes are very sensitive to the height of columns, and we can draw inaccurate conclusions when those bars are truncated.
- Avoid using pattern lines or fills. Use border only for highlights.
- Only use column charts to show trends if there are a reasonably-low number of data points (less than 20) and if every data point has a clearly-visible value.

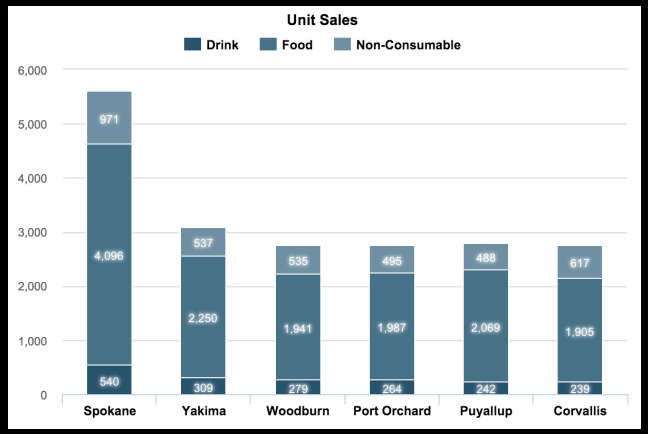
COLUMN CHARTS. HISTOGRAMS

Column Histograms



Histogram is a common variation of column charts used to present distribution and relationships of a single variable over a set of categories.

Stacked Column Charts



Use stacked column charts to show a composition.

BAR CHARTS

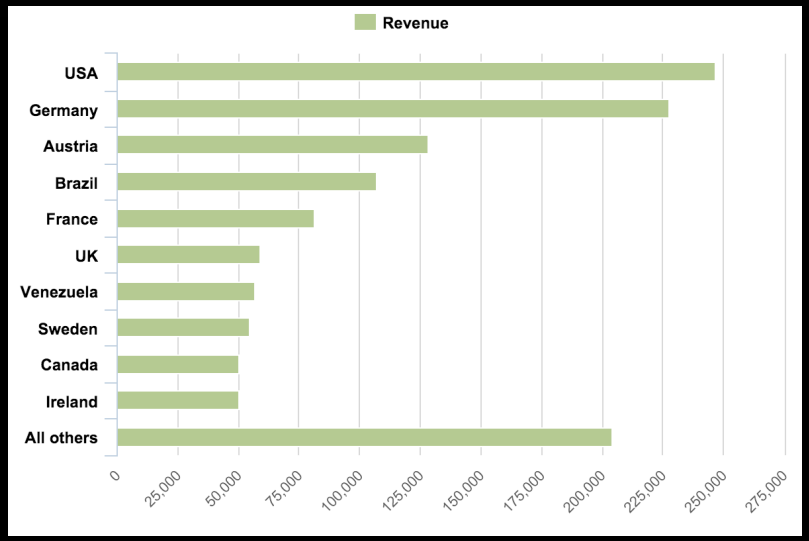
Bar charts are essentially horizontal column charts.

If you have long category names, it is best to use bar charts because they give more space for long text. You should also use bar charts, instead of column charts, when the number of categories is **greater than seven (but not more than fifteen)** or for displaying a set with negative numbers.

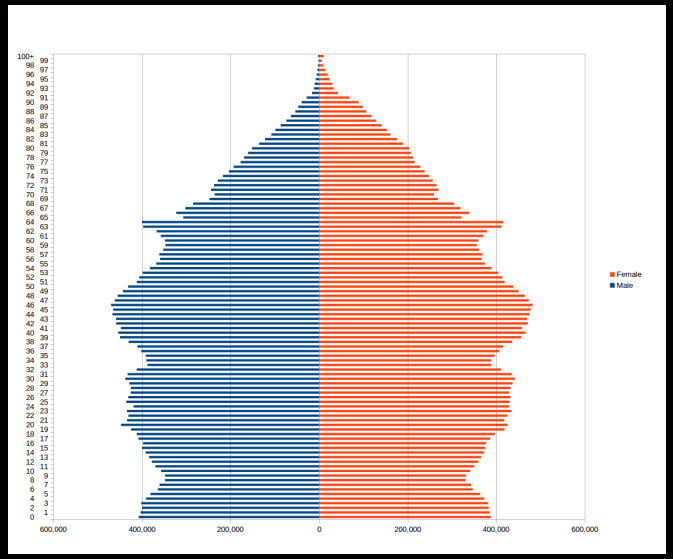
- A typical use of bar charts would be visitor traffic from top referral websites. Referring sites are usually more than five to seven sites and website names are quite long, so those should be better horizontally graphed.
- Another example could be sales performance by sales representatives. Again, names can be quite long, and there might be more than seven sales reps.

BAR CHARTS

Bar chart



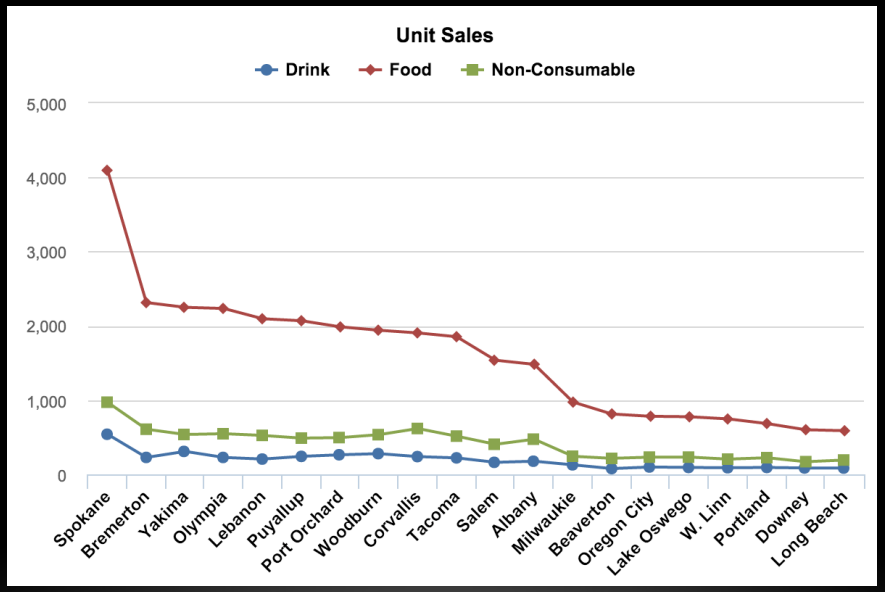
Bar Histogram Charts



Just like column charts, bar charts can be used to present histograms.
A good histogram example is a population distribution by the age (and sex). Remember those Christmas-tree graphs?

LINE CHARTS

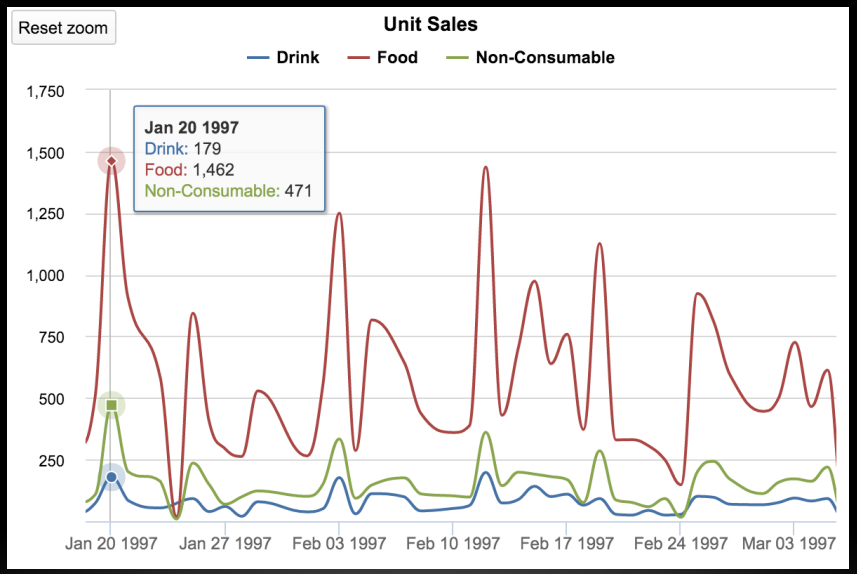
- Use lines when you have a continuous data set. These are best suited for trend-based visualizations of data over a period of time, when the number of data points is very high (more than 20).
- A line chart is also a good alternative to column charts when the chart is small.



LINE CHARTS. TIMELINE CHARTS

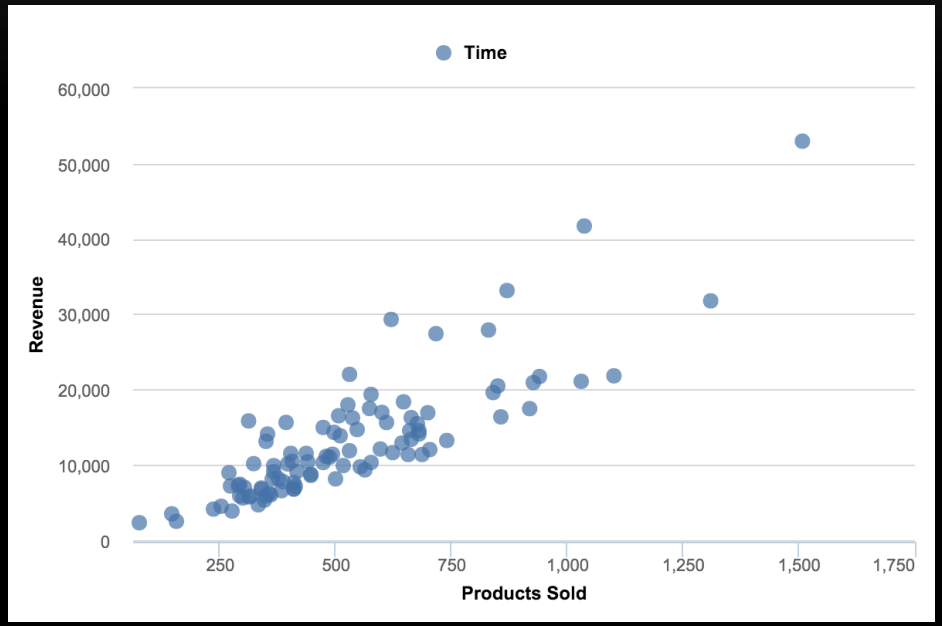
The most common examples of a time-line chart might be:

- stock market price changes over time,
- website visitors per day for the past 30 days,
- sales numbers by day for the previous quarter.



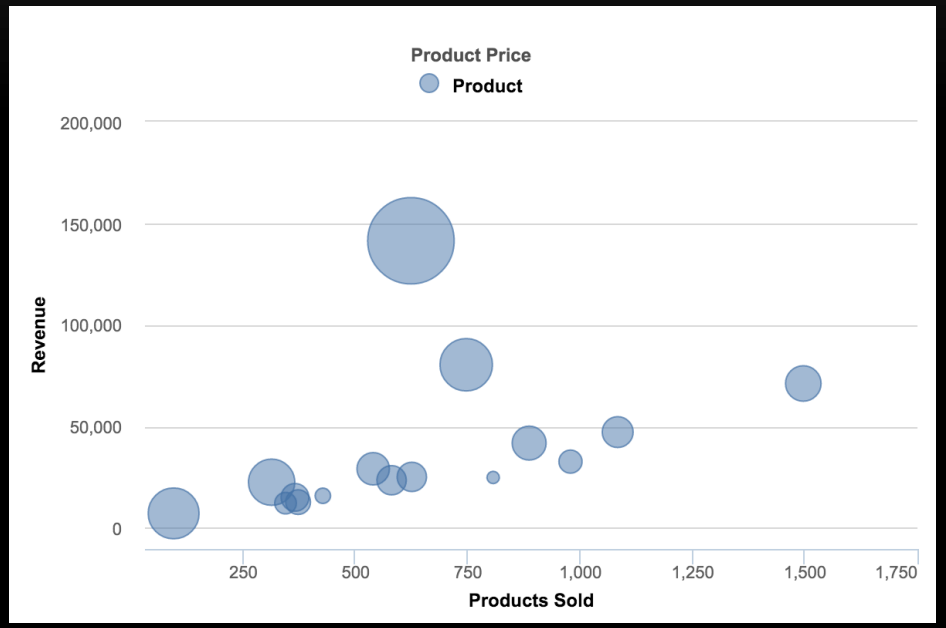
SCATTER CHARTS

- Scatter charts are primarily used for correlation and distribution analysis. Good for showing the relationship between two different variables where one correlates to another (or doesn't).
- Scatter charts can also show the data distribution or clustering trends and help you spot anomalies or outliers.
- A good example of scatter charts would be a chart showing marketing spending vs. revenue.



SCATTER CHARTS. BUBBLE CHARTS

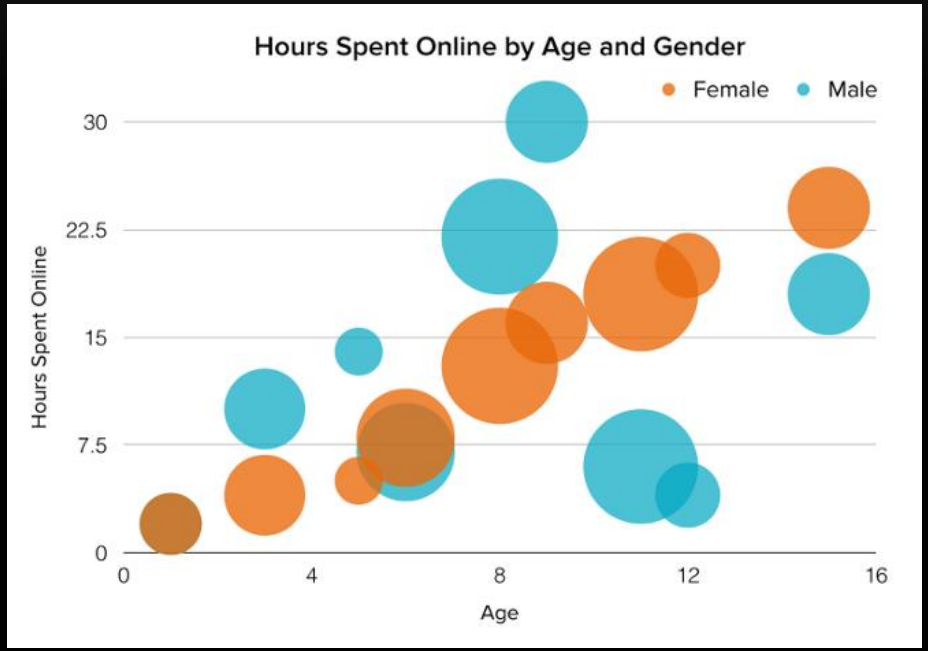
- A bubble chart is a great option if you need to add another dimension to a scatter plot chart. Scatter plots compare two values, but you can add bubble size as the third variable and thus enable comparison. If the bubbles are very similar in size, use labels.
- We could in fact add the fourth variable by color-grading those bubbles or displaying them as pie charts, but that's probably too much.



SCATTER CHARTS. BUBBLE CHARTS

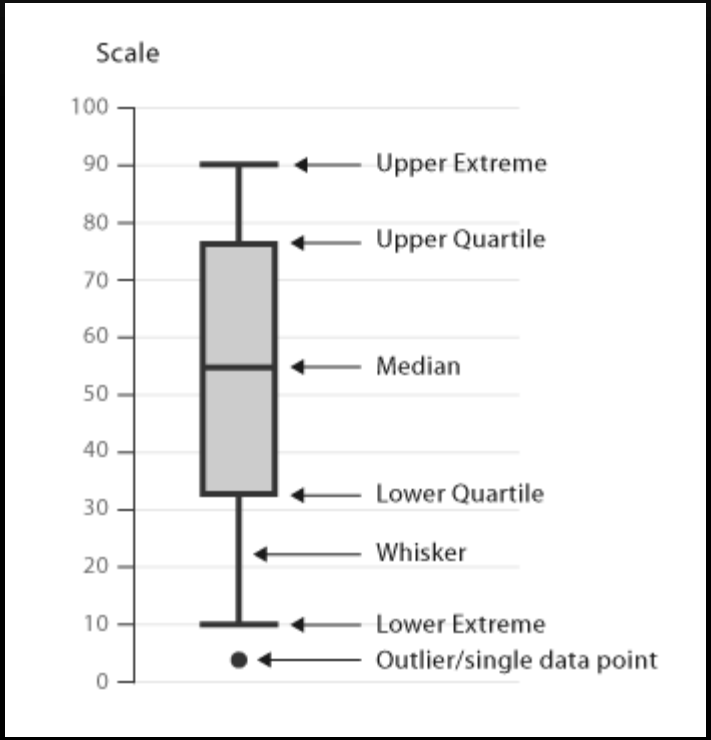
Use Scatter and Bubble charts to:

- Present relationships between two (scatter) or three (bubble) numerical variables,
- Plot two or three sets of variables on one x-y coordinate plane,
- Turn the horizontal axis into a logarithmic scale, thus showing the relationships between more widely distributed elements.
- Present patterns in large sets of data, linear or non-linear trends, correlations, clusters, or outliers.
- Compare large number of data points without regard to time. The more data you include in a scatter chart, the better comparisons you can make.
- Present relationships, but not exact values for comparisons.

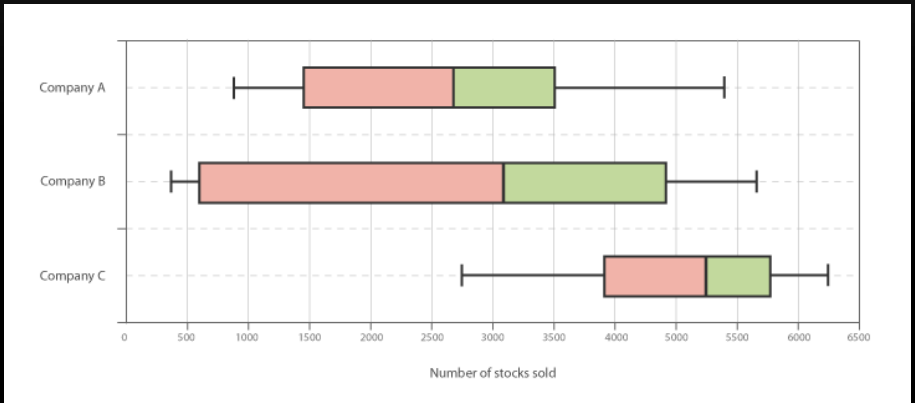


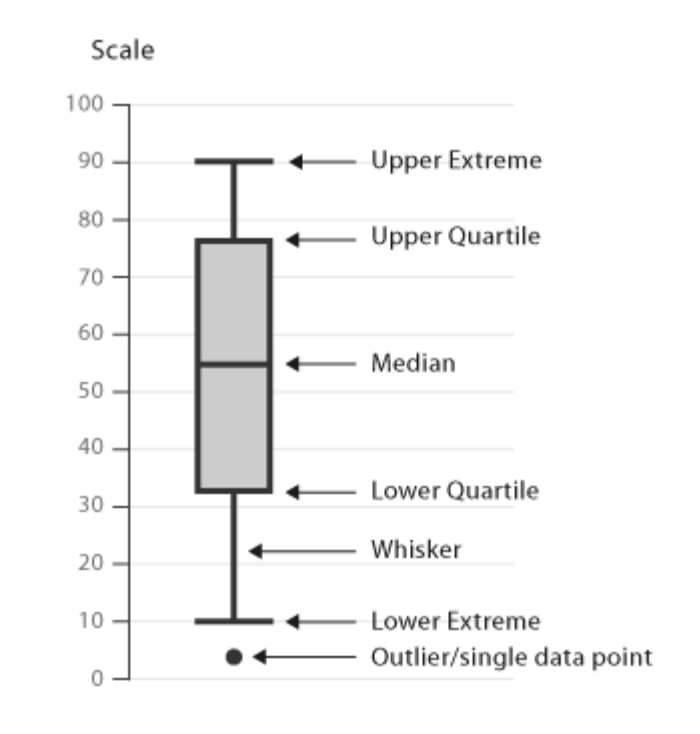
BOX AND WHISKER PLOT

- A Box and Whisker Plot (or Box Plot) is a convenient way of visually displaying the data distribution through their quartiles (четверти).
- The lines extending parallel from the boxes are known as the “whiskers” (бакенбарды), which are used to indicate variability outside the upper and lower quartiles. Outliers are sometimes plotted as individual dots that are in-line with whiskers. Box Plots can be drawn either vertically or horizontally.



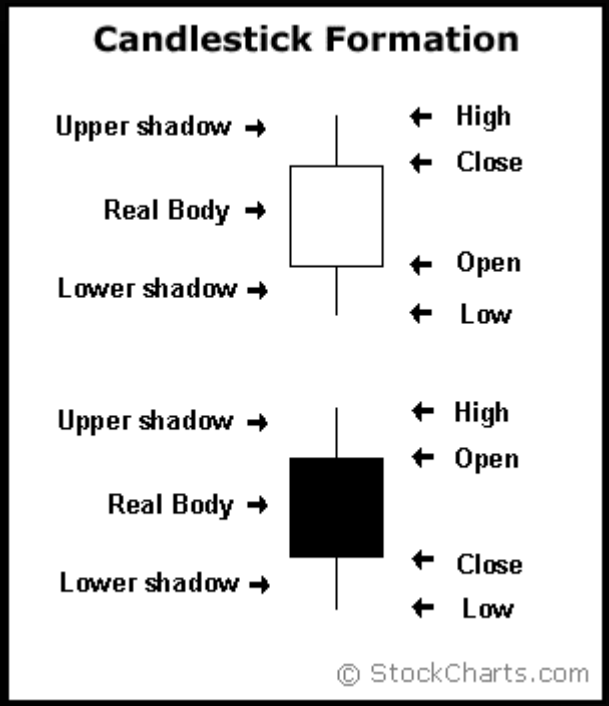
BOX AND WHISKER PLOT



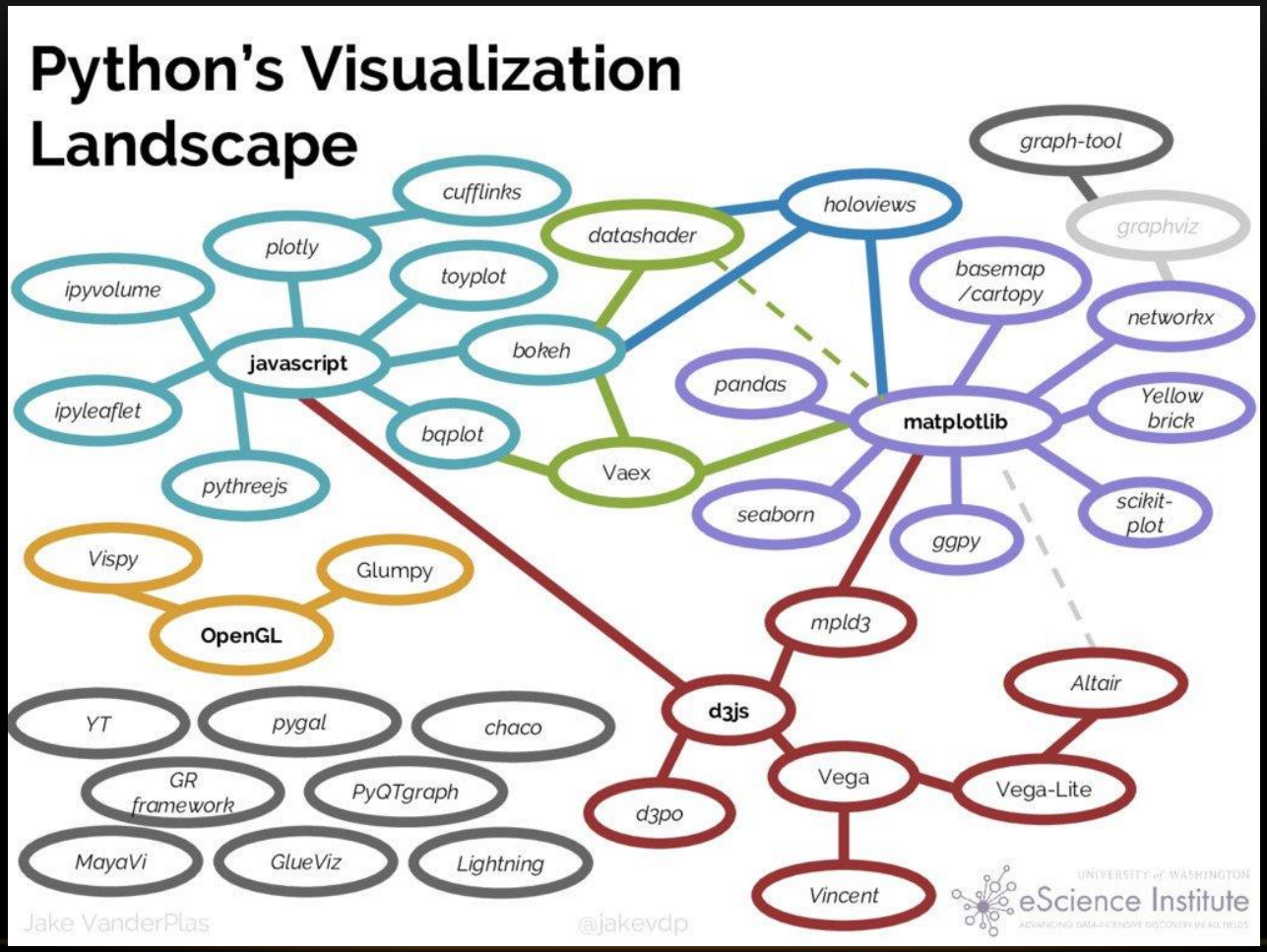


CANDLESTICKS

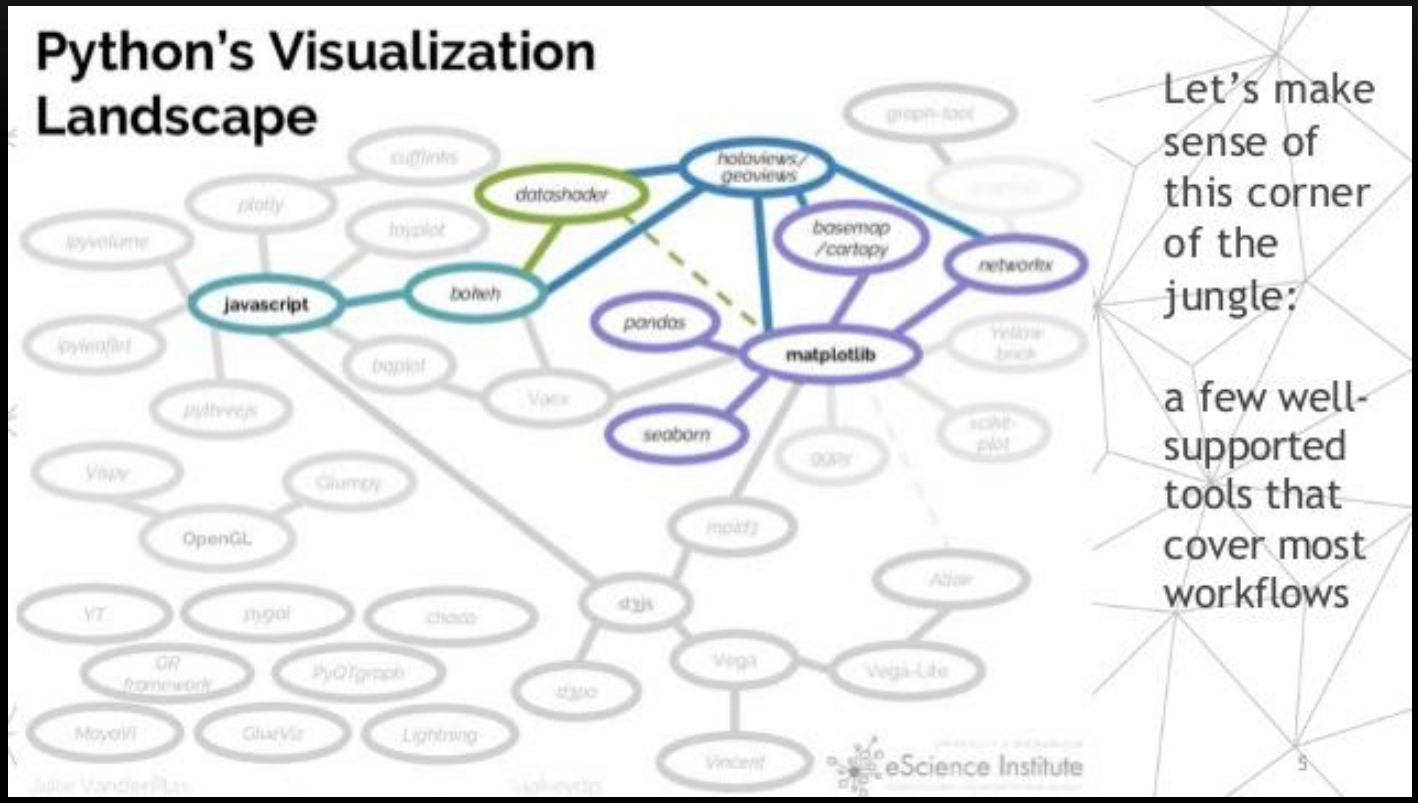
- Are used to represent stock prices



PYTHON VISUALIZATION LANDSCAPE



PYTHON VISUALIZATION LANDSCAPE



RESOURCES

- https://matplotlib.org/index.html - official website
- https://matplotlib.org/gallery/shapes_and_collections/scatter.html#sphx-glr-gallery-shapes-and-collections-scatter-py – scatter plot example

CHEAT SHEETS FOR PLT

- https://www.cheatography.com/gabriellerab/cheat-sheets/matplotlib-pyplot/
- https://www.datacamp.com/community/blog/python-matplotlib-cheat-sheet

REFERENCES

- https://eazybi.com/blog/data_visualization_and_chart_types/ article by Jānis Gulbis
- https://datavizcatalogue.com/methods/box_plot.html
- https://stockcharts.com/school/doku.php?id=chart_school:chart_analysis:introduction_to_candlesticks