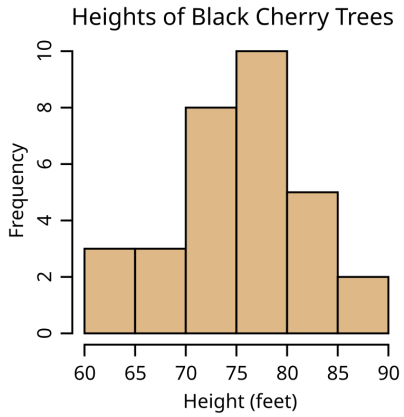
3-Data Visualization 2

By: Nourah Almutlaq

Univariate Distribution Plots

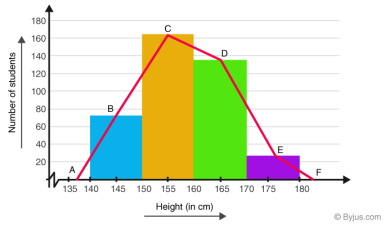
Histogram Plot.



histogram - Wiktionary

- رسم بياني عبارة عن مجموعة من المستطيلات، كل مستطيل يسمى bin.
- يُتيح لنا معرفة:
  - توزيع البيانات أو frequency.
  - المتوسط.
  - وجود تشتت في البيانات.

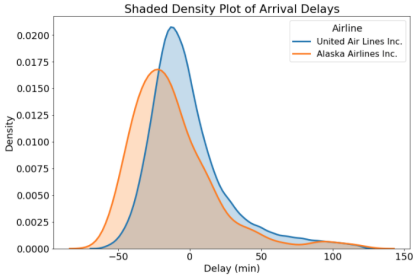
Frequency Polygon Plot.



Frequency Polygon - Definition,Steps and Solved Examples

- a line graph of class frequency plotted against class midpoints.
- It basically gives the same information as the histogram but is plotted in another style.
- Grouped data (histogram & Frequency polygon) destroys much of the original detail of the data but highlights the main features contained.

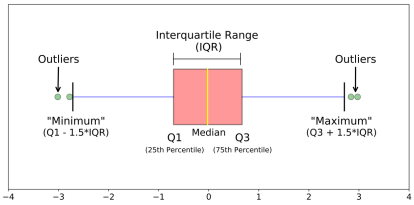
Density Plot.



How To Find Probability From Probability Density Plots | by Admond Lee | Towards Data Science

- It visualizes the distribution of our data over a continuous interval.
- The vertical axis tells us what is called the probability density function.
- the probability of the area under the curve should sum up to 1.
- It better tells the shape of the distribution than the histogram, because it is not affected by the number of bins.

Box Plot.



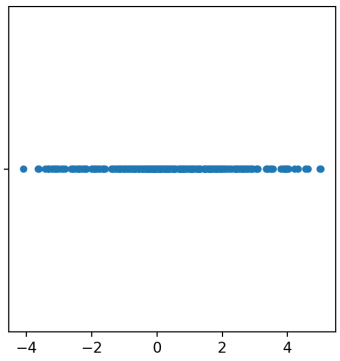
Understanding Boxplots - KDnuggets

- It shows the summary of central tendency measures and dispersion.
- it's useful for spotting outliers.
- The outliers are points above and below max and min.
- If they are too far, they are called extreme outliers.
  - if > Q3 + 3 IQR (**up the max**)
  - if < Q1 - 3 IQR (**under the min**)

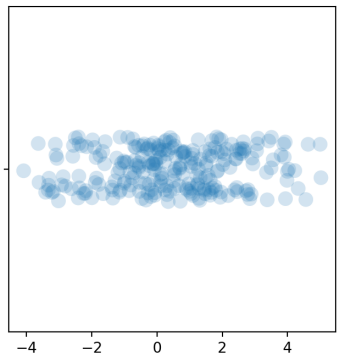
Violin Plot.

- it is the same as the box plot but instead shows the probability density plot.
- The strength of a violin plot is that it solves the full distribution.
- This is useful when data is multi-modal because in this case, a box plot wouldn't be enough.

Strip Plot.



Strip plot without jitter



Change the transparency and the size of the markers of a seaborn strip plot

- a plot of distorted values along one axis.
- In the above graph, the vertical axis doesn't represent any value.
- all these points overlap, so we add a bit of jitter.
- It is usually preferred over a histogram or density plot when we have a small dataset.

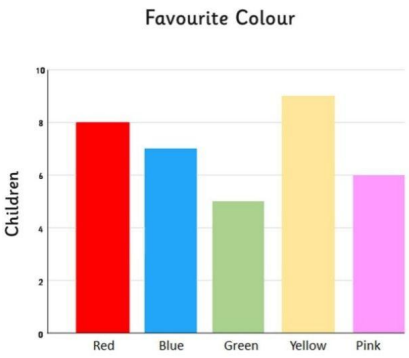
Swarm Plot.

Seaborn swarmplot: Bee Swarm Plots for Distributions of Categorical Data • datagy

- same as a strip plot, but observations are adjusted to avoid overlapping, providing a better visualization.
- It is better recommended for small datasets.
- one axis doesn't represent any value.

Univariate Comparison Plots

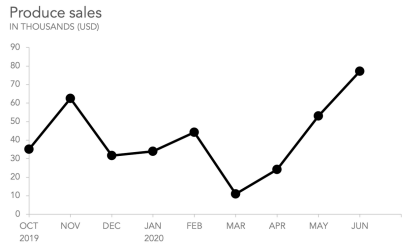
Bar Plot.



What is a Bar Chart? - Twinkl

- It presents categorical data with rectangular bars.
- The width must be the same for every rectangle.
- One axis represents the categories and the other axis their value.
- There is no established order in which to plot the different categories.
- helps us compare the value of a numeric variable among different categorical variables of our dataset.

Line Plot.

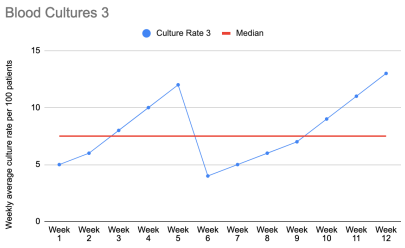


what is a line graph, how does a line graph work, and what is the best way to use a line graph? — storytelling with data

- It is somehow halfway between univariate and multivariate analysis.
- We are representing a chronological variable. This is time-dependent.

- This kind of graph helps us to understand how a variable evolves over time.

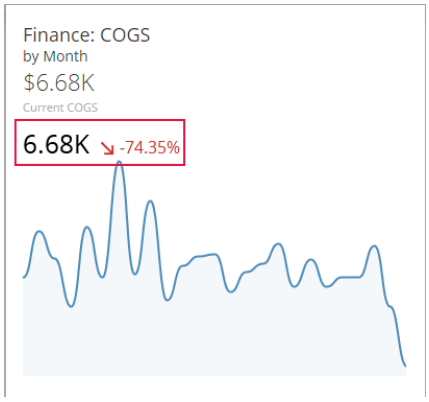
Run Plot.



Reporting QI Results Part 1 – Run Charts - CanadiEM

- It is like the line chart but shows some measure of like mean or median.

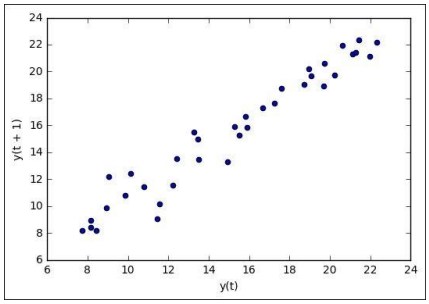
Sparkline Plot.



Spark Line Chart

- It represents the general shape and variation of some value over time.
- e.g.: stock market prices, humidity, temperature.

Lag Plot.



Lag plots | Python Data Analysis - Second Edition

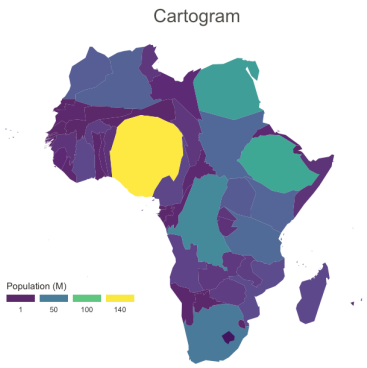
- it represents a chronological variable.
- lagged points(horizontal axis) and observations that happened after (vertical axis).
- for example, air pollution values at t-1 against the values one hour later.

Circular Area (Summary) Plot.

javascript - Is it possible to produce circular (round) shaped radar chart in Chart.js? - Stack Overflow

- It looks like a spider web or radar.
- It's like if we get a line chart and bend it until it has a circular shape.
- to compare the value of a variable among several categorical variables.
- to see the evolution in time of one variable.

Cartograms Plot.



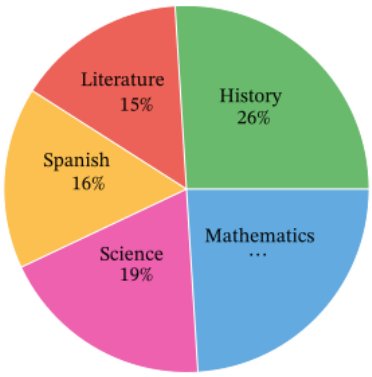
Cartogram – from Data to Viz

- It's a geographical comparison of a variable.
- value of a variable according to the locations represented by different colors or patterns.

Univariate Composition Plots

Composition plots are similar to comparison plots in the sense that they serve for comparing variable magnitudes, but in a way in which you visually join areas representing those magnitudes.

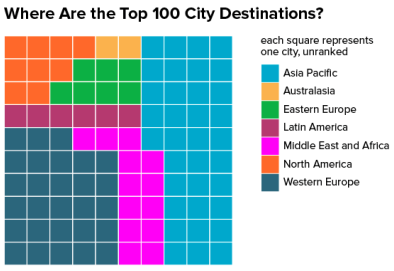
Pie Plot.



Lesson Explainer: Pie Charts | Nagwa

- to visualize different proportions (only with a few items).
- it is not recommended to use unless you want to show that one is greater than the other.

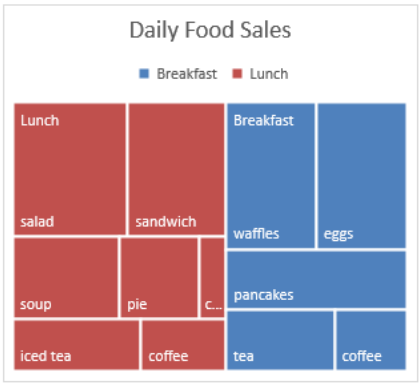
Waffle (Square Pie) Plot.



Waffle Chart | Data Viz Project

- showing equally sized atomic square areas.
- you can get a sense of which group with the same color is bigger or smaller than the other one.

Tree Map Plot.



What is a Treemap Chart? | TIBCO Software

- the variable quantities are represented by rectangle areas.
- It helps us to get an idea about the proportions of each variable.

Waterfall (flying bricks) or (Mario) or (bridge) Plot.

How to Create and Customize a Waterfall Chart in Microsoft Excel

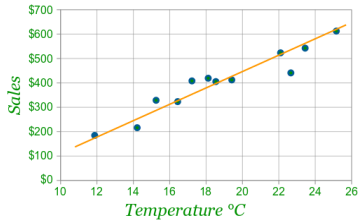
- helps understand the cumulative effect of positive and negative values.
- data is represented sequentially.
- It helps us to see the cumulation effects to a certain quantity.
- for example, sales, credits, and later purchases.

Multivariate Analysis

- to find relationships in data.
- to show the statistical dependence between variables: variation, coefficient, etc.

Multivariate Distribution Plots

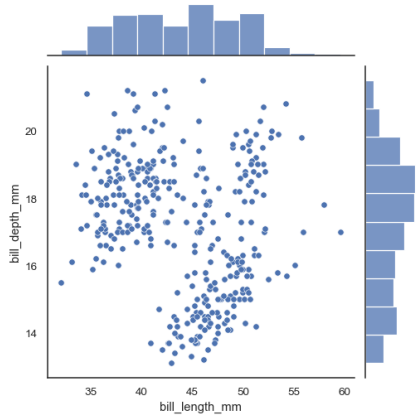
Scatter (dispersion) Plot.



Scatter (XY) Plots

- used to analysis of two quantitative variables.
- it highlights the dispersion of points.
- it shows the correlation between both variables.

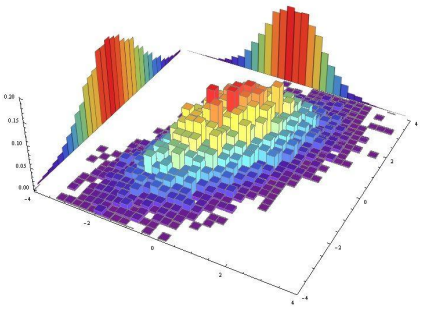
Joint Distribution Plot.



seaborn.jointplot — seaborn 0.12.2 documentation

- it represents the scatter plot with the histograms of both variables at the sides.
- the histograms are called marginal distributions.

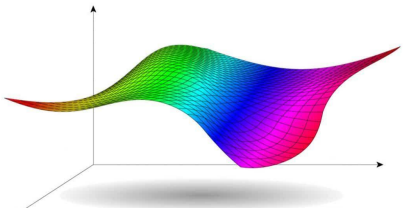
Stereogram or bivariate histogram Plot.



plotting - Visualization of Bivariate Distributions - Mathematica Stack Exchange

- It is a joint distribution but plots the variables as a 3D histogram.

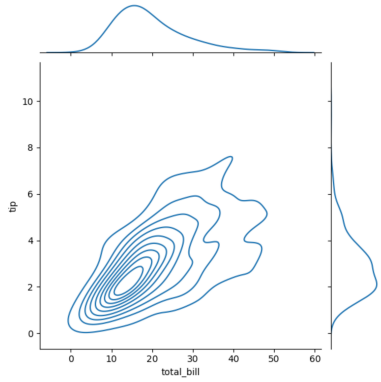
Surface Area Plot.



Surface Chart in Excel

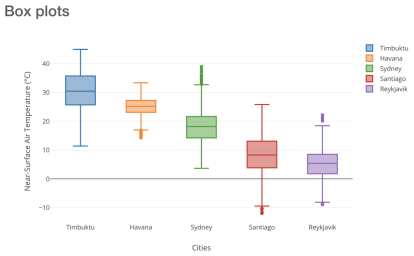
- we pick only the top values of a bivariate histogram and smooth the transitions between these values.
- it is used when the summation of quantitative data is to be communicated rather than the individual data values.

Level Curves Plot.



- it shows the density function instead of absolute values.
- it shows high slope area (total_bill = 5, tip =1) and low slope area (total_bill = 25, tip =4).

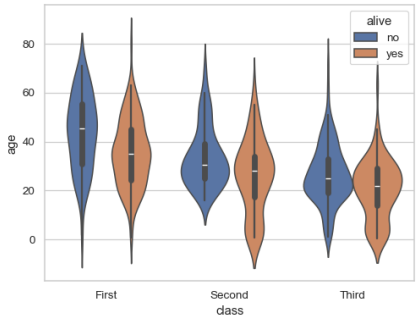
Box Plot.



Box plots — Climate Data Store Toolbox 1.1.5 documentation

- In univariate analysis, it shows the summary of central tendency measures and dispersion.
- In multivariate analysis, help us compare them among different classes.

Violin Plot.

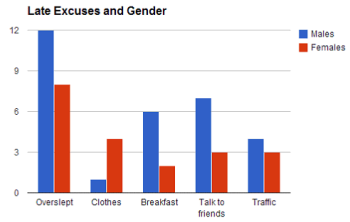


seaborn.violinplot — seaborn 0.12.2 documentation

- Same as the box plot, but with extra information (distribution shape among classes).

Multivariate Comparison Plots

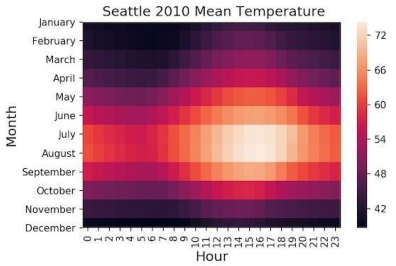
Side by Side Bar Plot.



How to Use Spreadsheets: Side by Side Bar Graphs in Google Sheets

- it lets us compare quantitative variables among classes.

Heatmap Plot.



How to Make Heatmaps with Seaborn in Python? - Data Viz with Python and R

- composed of small squares that represent the value of some variable related to the other.
- represented different colors within a range of shades on a bright/dark scale.
- It helps us to see if there might exist some kind of tendency or value concentration.

Cartograms Plot.

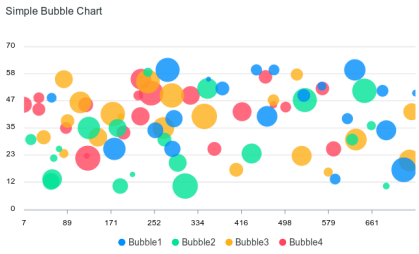
- it's possible to represent more than one variable.
- In this example, we can see the population size of each country in Africa given by the size of the country's shape, and at the same time, we represent its wealth with color, being the brighter colors related to richer countries.

Multivariate Relationship Plots

Scatter Plot.

- it not only gives an idea about the variables' distribution but also about the relationship.

Bubble Plot.

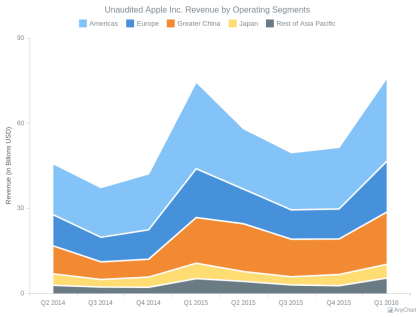


JavaScript Bubble Charts Examples – ApexCharts.js

- it is a scatter plot but instead of representing equally sized points, plot them with a size related to some quantity.

Multivariate Composition Plots

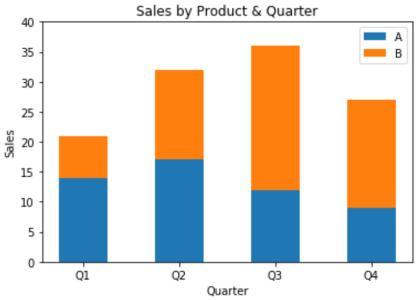
Stacked Area Plot.



Stacked Area Chart | Area Charts

- It is a composition of different series added, each one on top of the others.
- This way, it gives a graphical representation of the extent of the contribution of each one.
- If related to time, we can analyze the proportional contribution evolution.

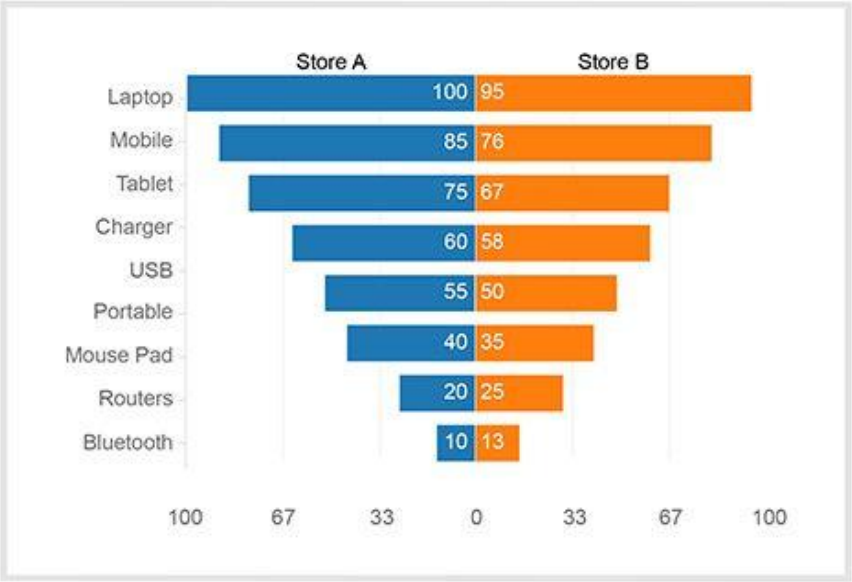
Stack Bars Plot.



How to Create Stacked Bar Charts in Matplotlib (With Examples) - Statology

It is similar to a stack area chart, but instead of series, we represent the contribution of different classes.

Tornado Plot.



How to Create a Tornado Chart in Excel? A Complete Guide

- It's a bar chart composition, usually represented in an ordered manner related to the numerical variable we want to study.
- This way we can compare the value related to two classes among all the classes.