Data Analysis and Visualization

CentraleDigitalLab@Nice

Deborah Dore - PhD - ddore@i3s.unice.fr
MARIANNE, Université Côte d'Azur, CNRS, INRIA, I3S

Basic Visualisations

Overview

- Bar charts
- Scatter Plots
- Histograms
- Density Plot
- Box Plot
- ....

Basic Visualisations

Overview

- Bar charts
- Scatter Plots
- Histograms
- Density Plot
- Box Plot
- ....

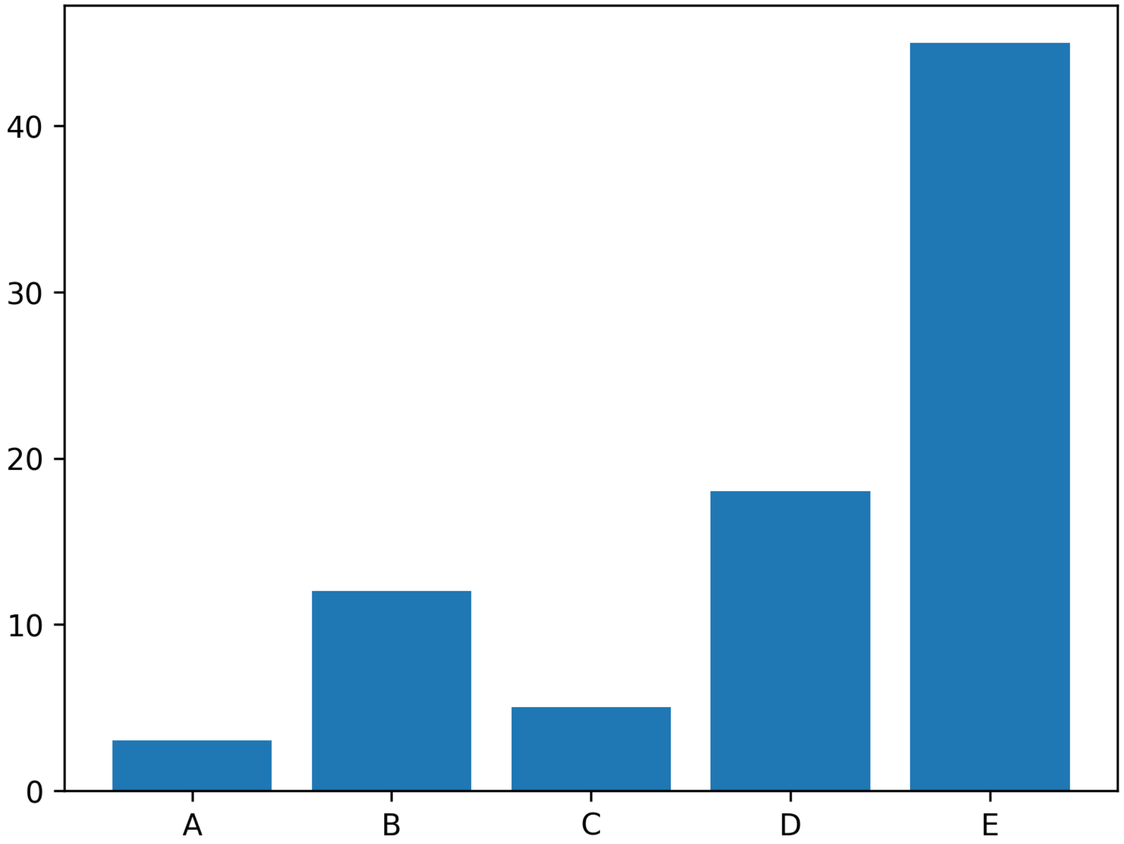
NEVER USE PIE PLOTS OR 3D PIE PLOTS

Basic Visualisations

Bar Charts

A **bar chart** **displays distinct, numerical comparisons across categories using either vertical or horizontal bars (column chart).** The particular categories under comparison are displayed on one axis of the chart, while a discrete value scale is represented on the other

Histograms and bar charts differ in that the former do not show continuous changes over time. Rather, the discrete data in a bar chart is categorical, which means it provides an answer to the query "how many?" in each category.

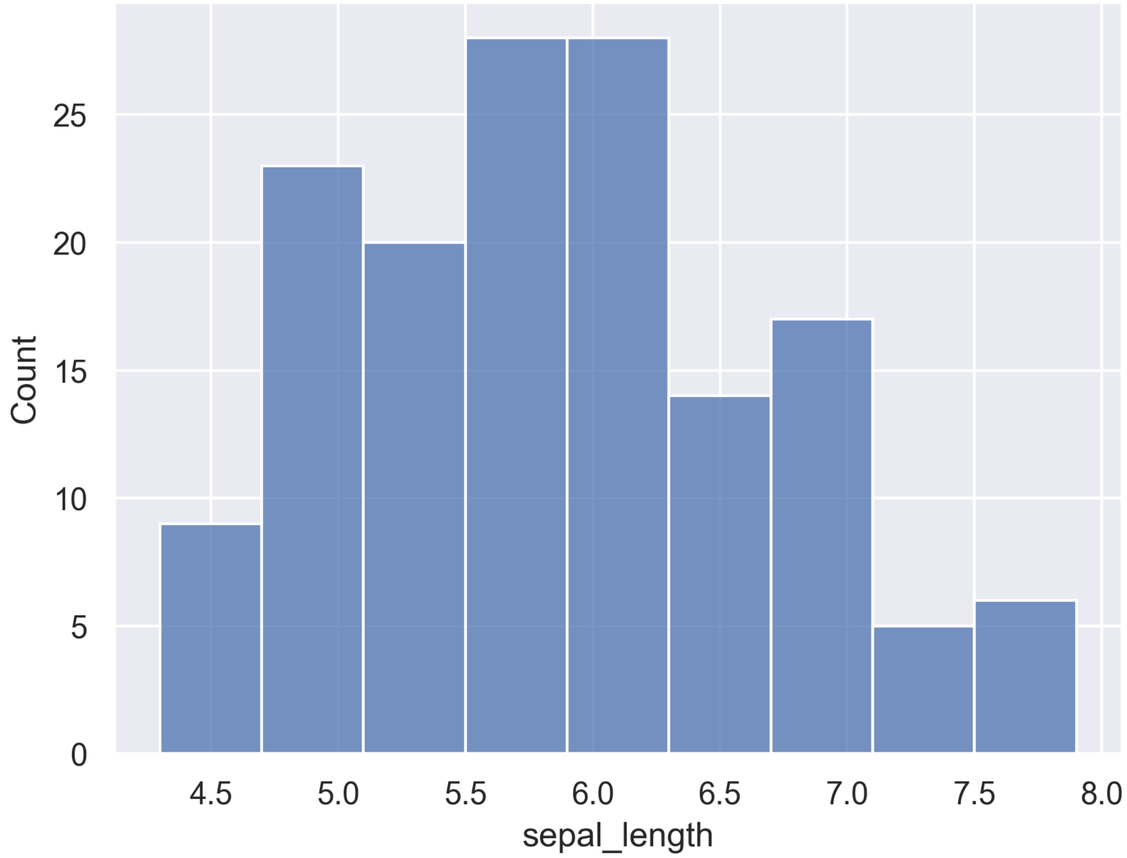


Basic Visualisations

Histograms

A histogram shows how data is distributed over a continuous period of time. The tabulated frequency at each interval/bin is represented by each bar in a histogram.

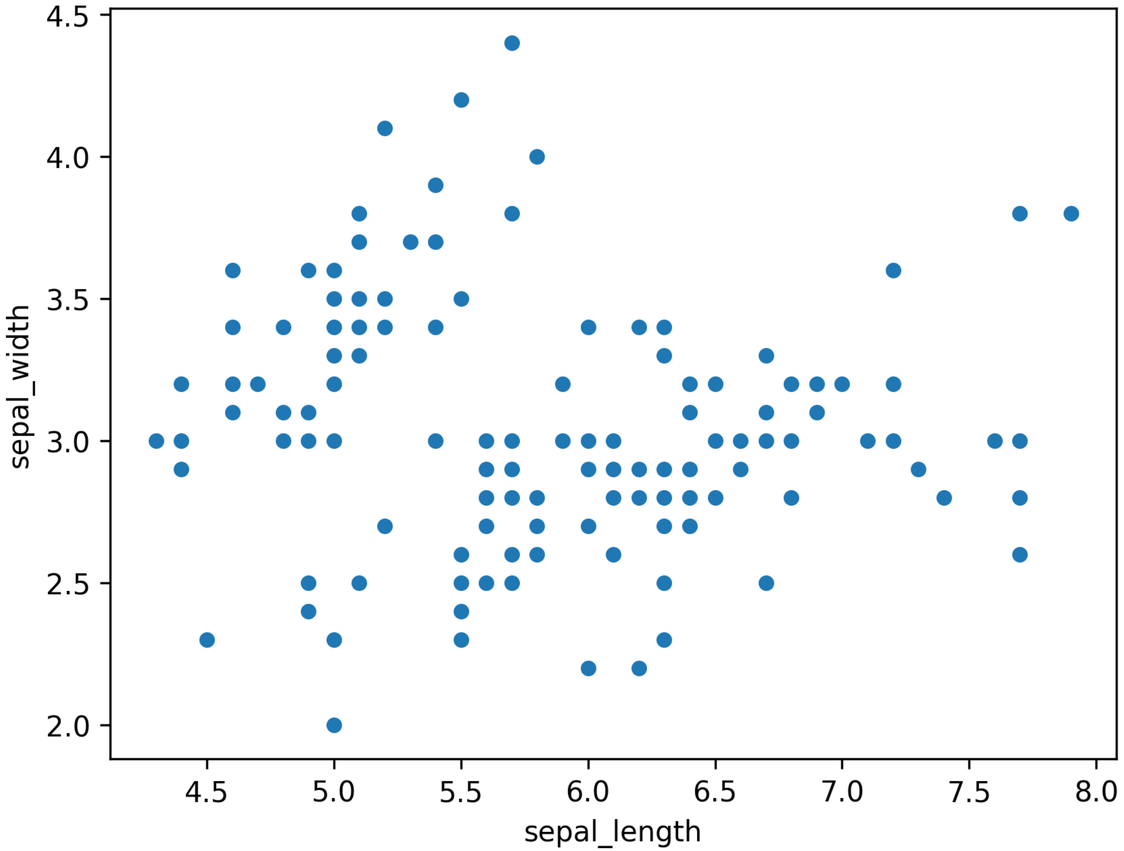
Histograms are useful for estimating the concentration of data, identifying the extremes, and determining whether there are any gaps or anomalous values. Additionally, they are helpful in providing a general overview of the probability distribution.



Basic Visualisations

Scatter Plots

A **scatterplot** shows all of the values between two variables by placing points on a system of cartesian coordinates. You may determine **whether there is a relationship or correlation between two variables** by assigning an axis to each one.



Basic Visualisations

Box Plot

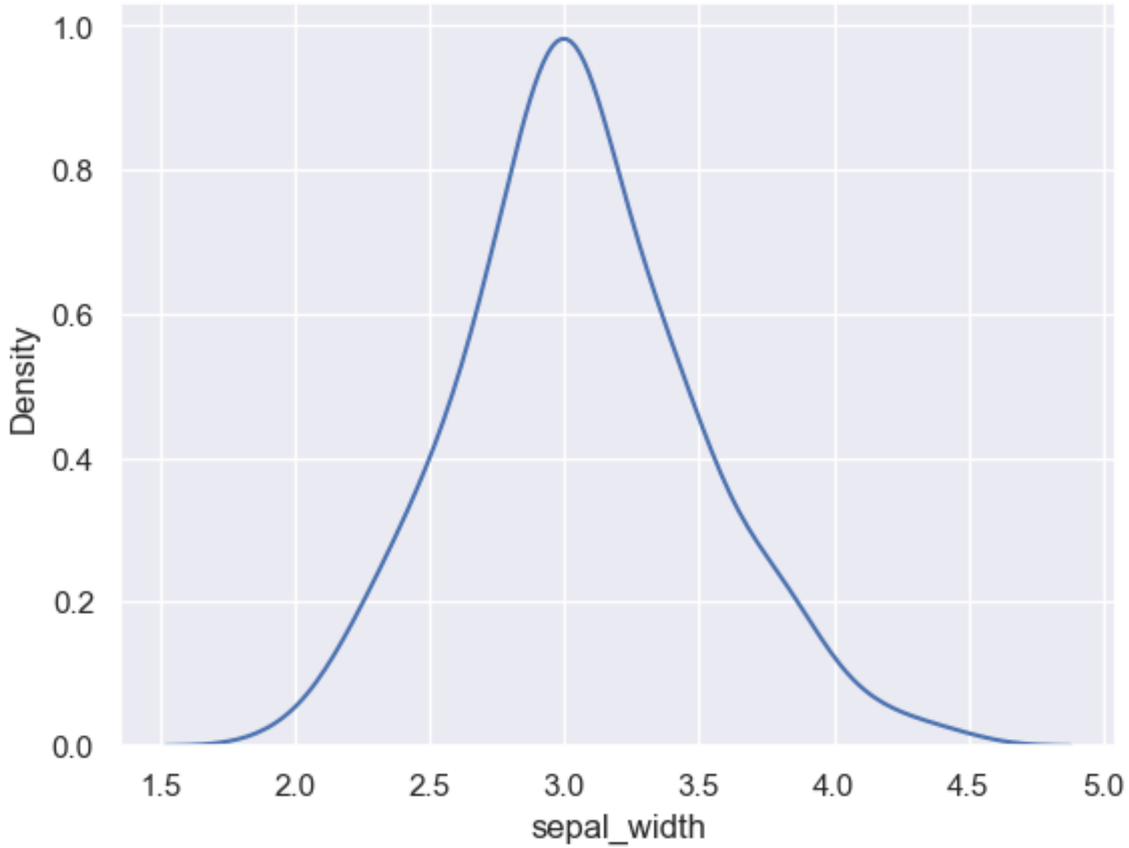
A handy method for **showing the data distribution** through its quartiles is to use a **Box and Whisker Plot**, often known as a Box Plot.

The "whiskers," or lines that extend parallel from the boxes, are used to show variability outside of the upper and lower quartiles. Occasionally, outliers are plotted as lone dots aligned with whiskers. There are two ways to draw box plots: vertically and horizontally.

Basic Visualisations

Density Plot

A **density plot** shows **how data is distributed over a continuous time period or interval**. By smoothing down the noise, this histogram-style chart plots values using kernel smoothing, which enables smoother distributions. Where values are concentrated across the interval is shown by the peaks of a density plot.



Advanced Visualisations

Overview

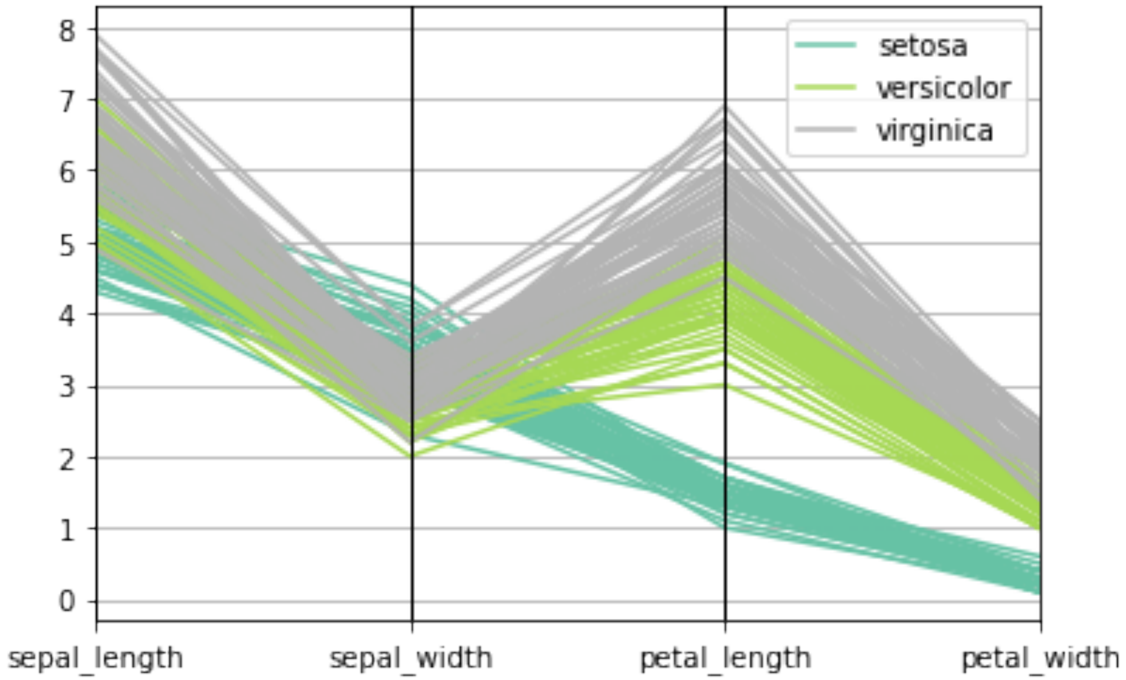
- Parallel Coordinates
- Treemaps
- Violin plot
- Sankey Diagrams
- Radar Charts
- …

Advanced Visualisations

Parallel Coordinates Plot

Numerical data that is multivariate is plotted using **parallel coordinates plot.**

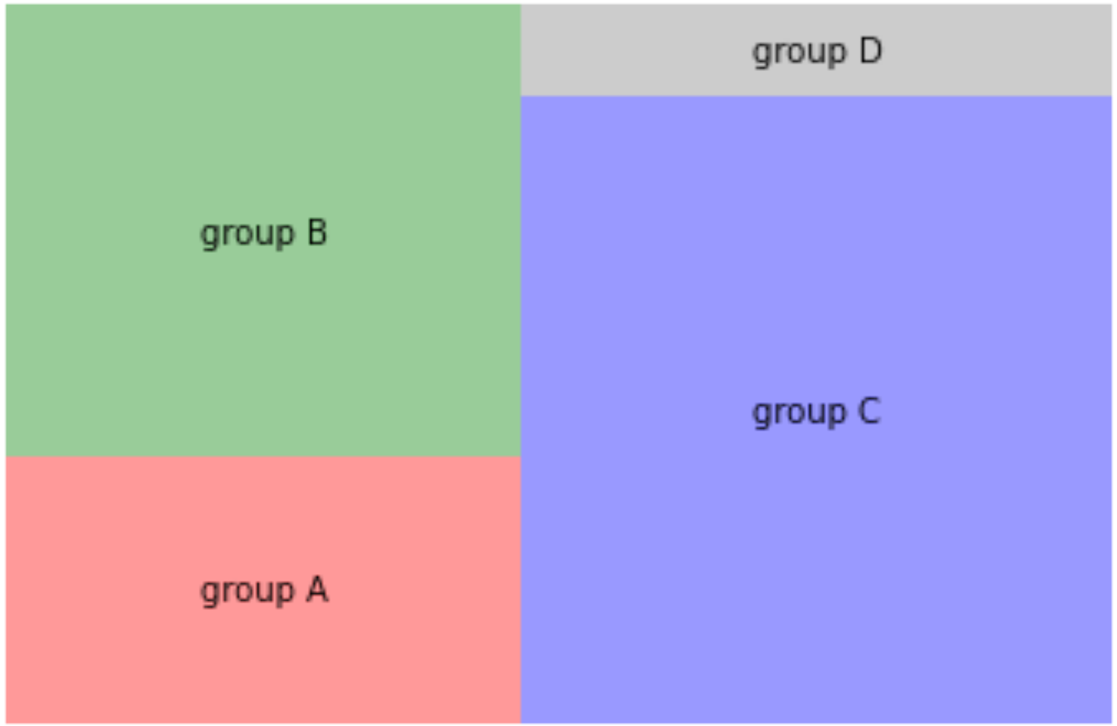
Plots with parallel coordinates are perfect for **comparing a large number of variables at once and examining their relationships.** For instance, comparing computer or car characteristics across models would be an example of comparing a variety of items with the same attributes



Advanced Visualisations

Treemaps

In addition to showing quantities for each category by area size, **treemaps** offer an alternate method of **visualizing a tree diagram's hierarchical structure**. Rectangle areas are allocated to each category, and the subclass rectangles are nestled within them

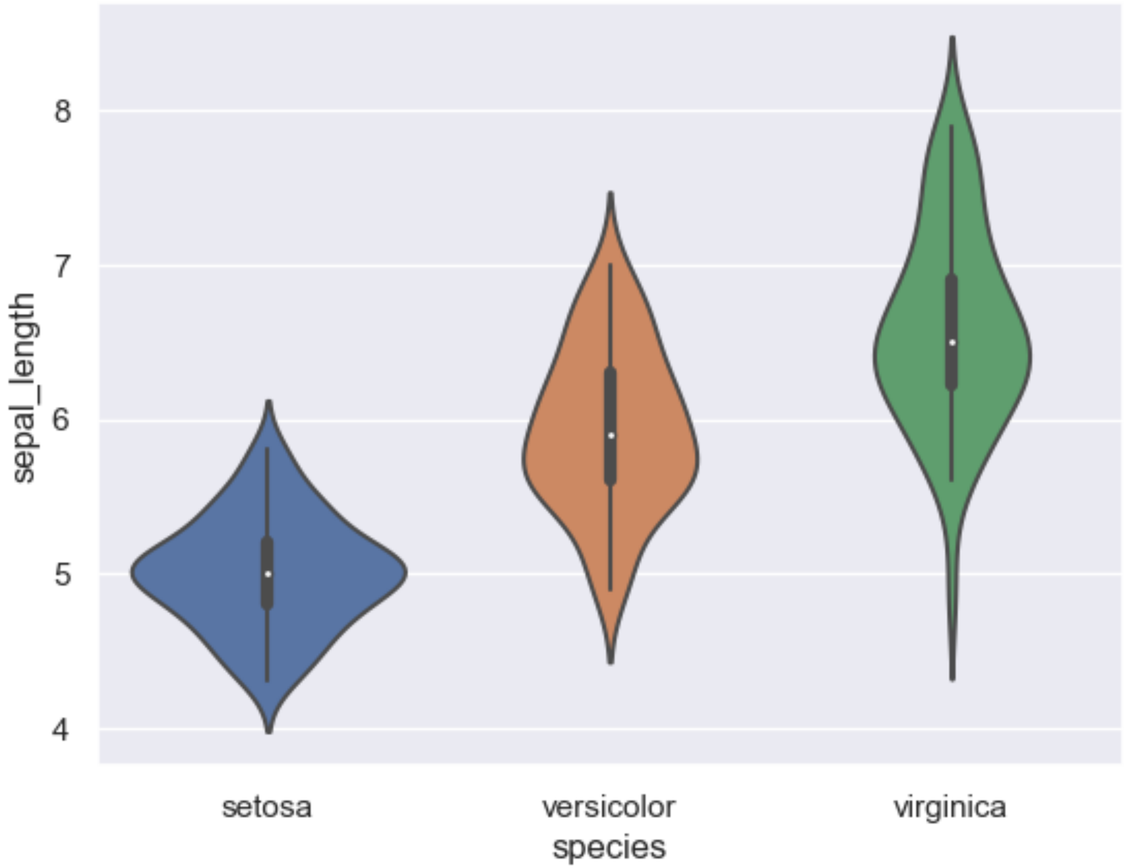


Advanced Visualisations

Violin Plot

The **data distribution and probability density** are visualized using a **violin plot.**

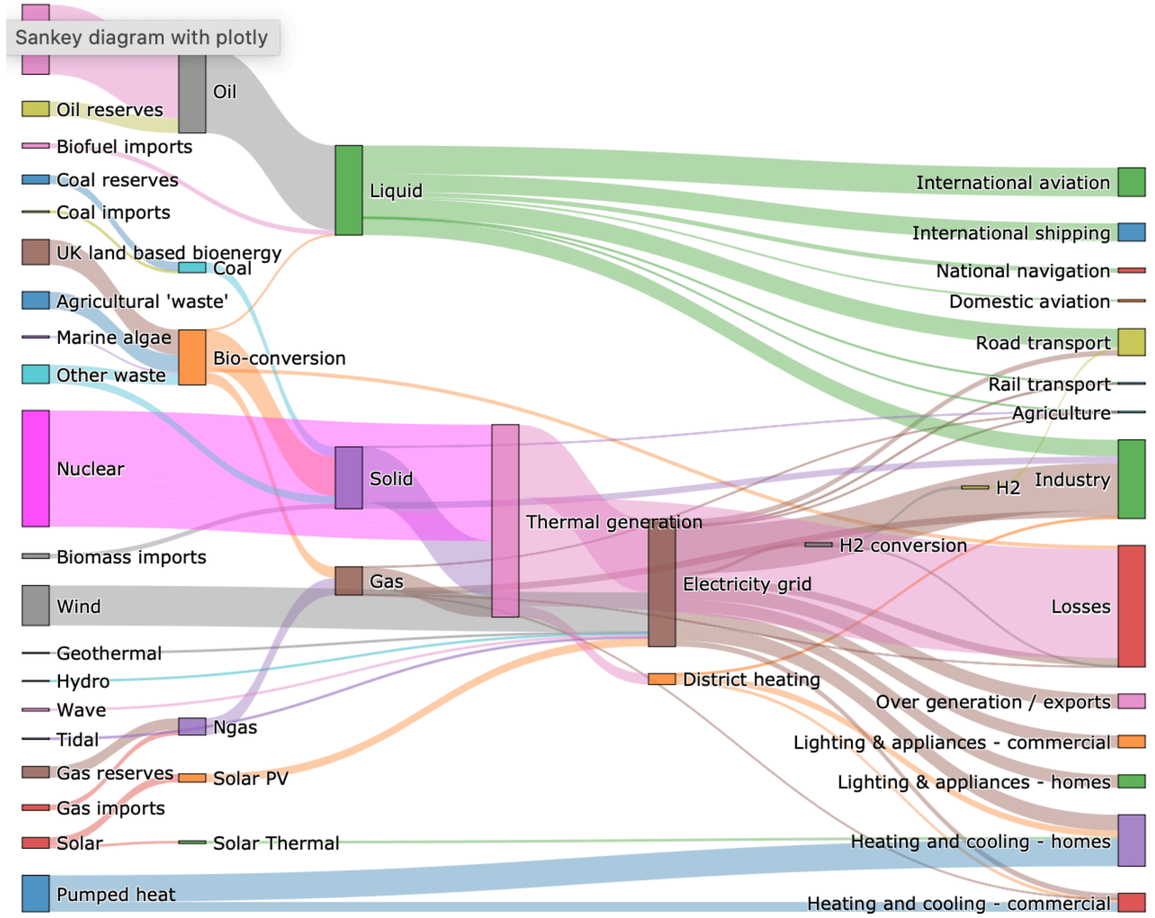
Rotated and positioned on each side, this chart combines a Box Plot and a Density Plot to display the data's distribution form. The thick black bar in the center denotes the interquartile range, while the white dot in the center represents the median value. It is accompanied by a thin black line that shows the data's upper (max) and lower (min) neighboring values. The graph marker is occasionally cut from the end of this line



Advanced Visualisations

Sankey Diagrams

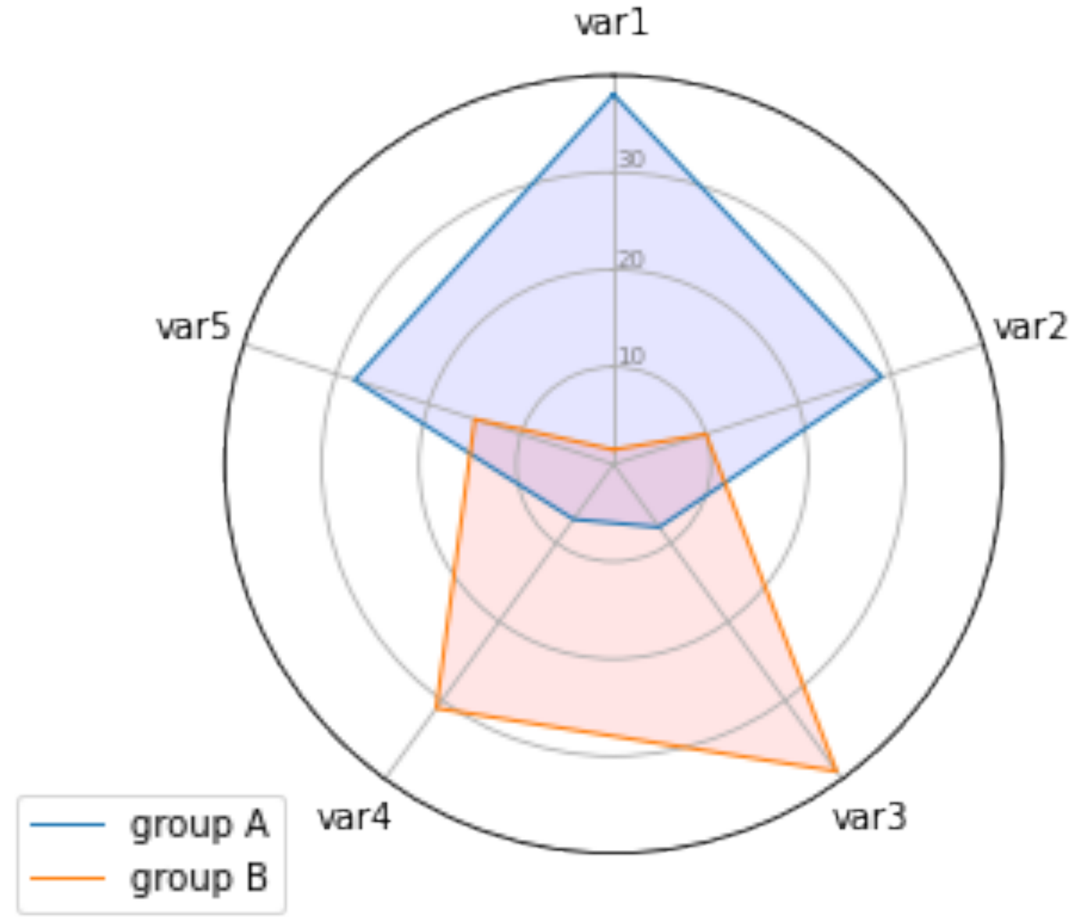
Sankey diagrams show **the quantities of flows in relation to each other**. Sankey diagrams are typically used to graphically depict the flow of any isolated system or process, as well as the movement of energy, money, or materials



Advanced Visualisations

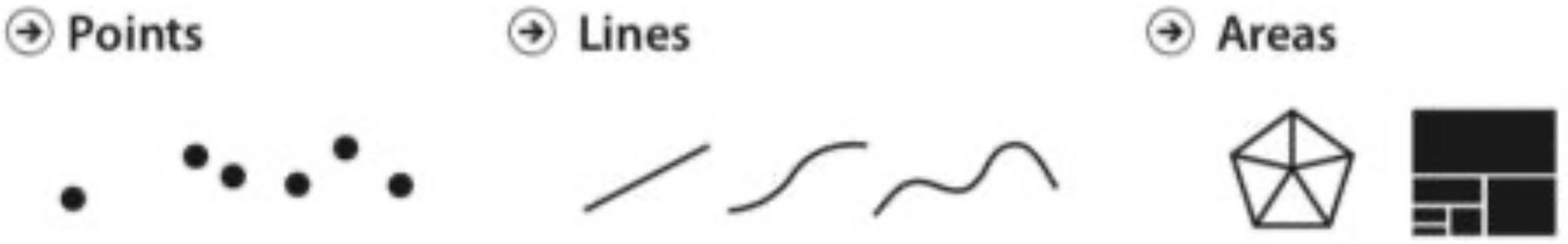
Radar Charts

One tool for **comparing several quantitative data** is a **radar chart.** They are therefore helpful for determining which variables have comparable values or whether any outliers exist within each variable. In addition to showing which variables in a dataset are scoring highly or poorly, radar charts are also a good way to show performance



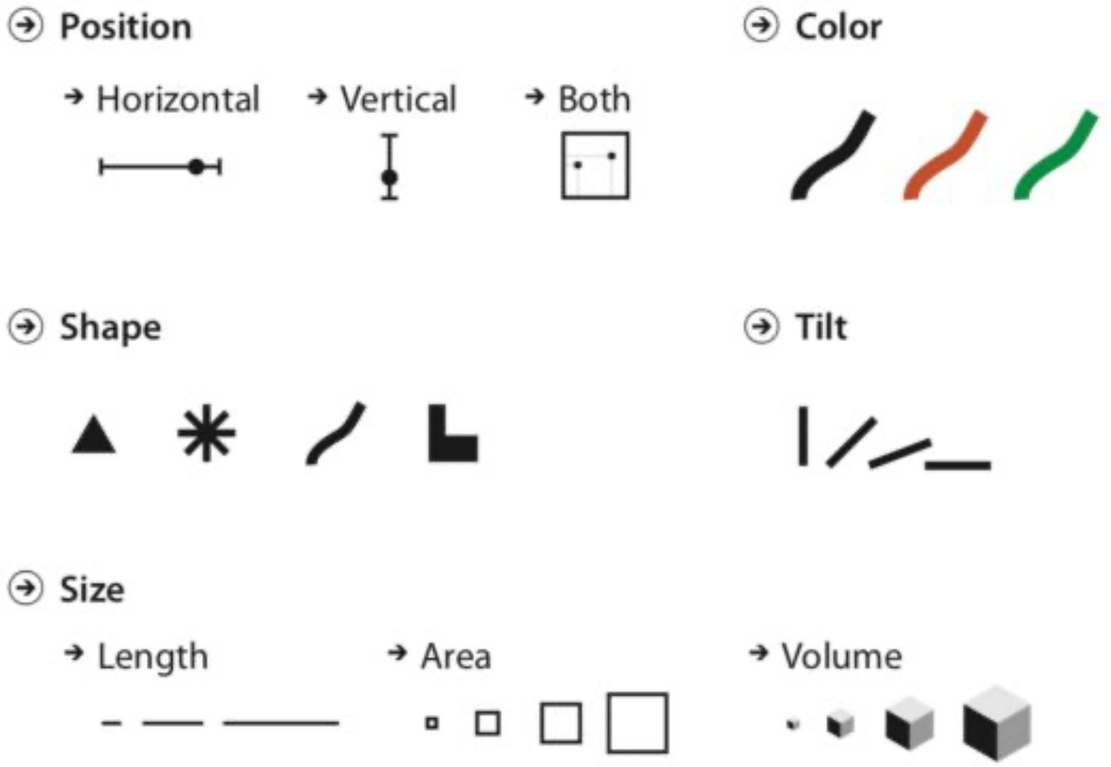
Marks and Channels

Marks are basic geometric elements that depict items or links (point, line, area)

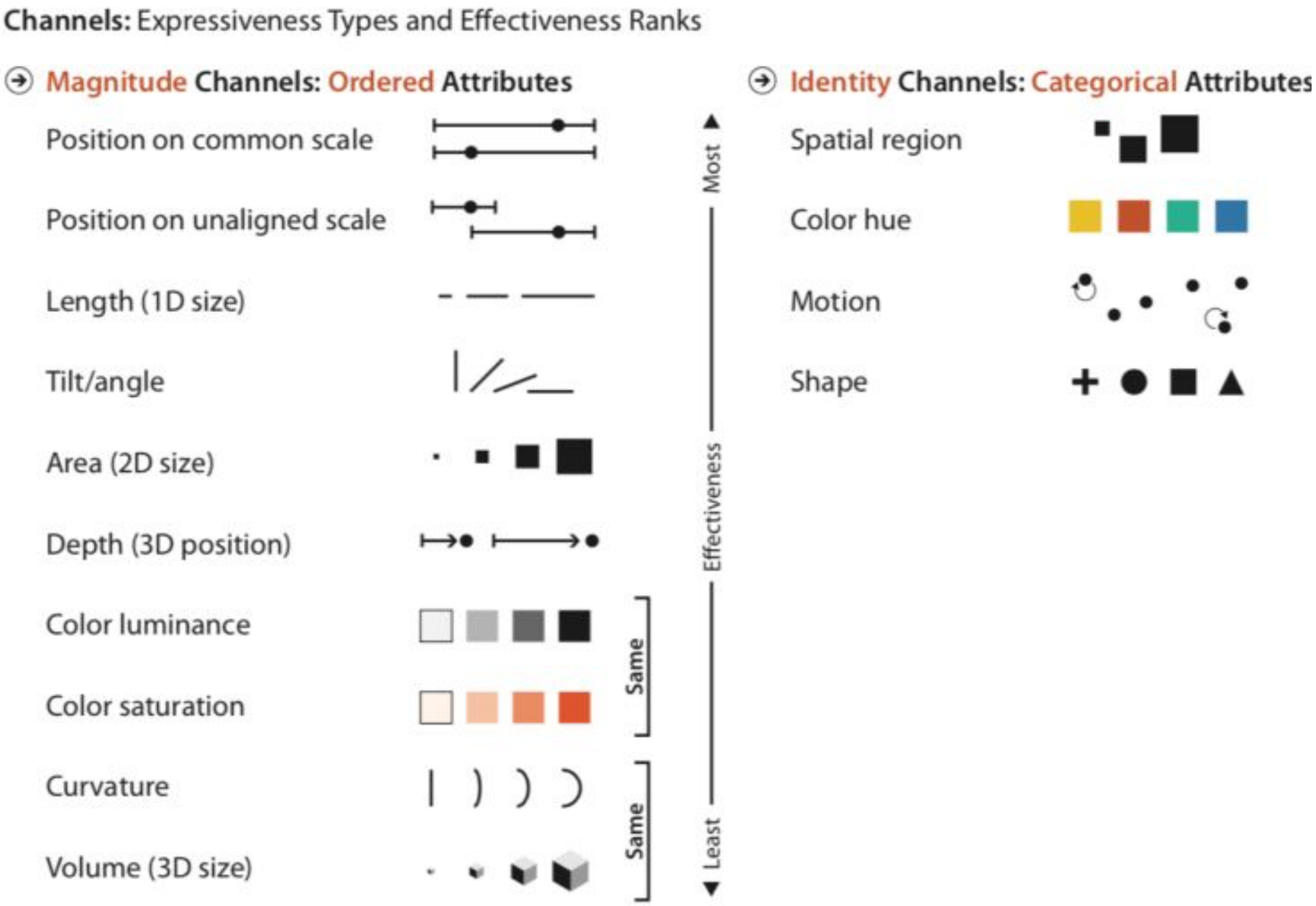


Marks and Channels

Channels control the appearance of marks, independently of the dimensionality go the geometric primitive. (position, color, shape, tilt (angle), size)



What channel is effective?



LINKS

- **Mandatory reading:** https://maelfabien.github.io/machinelearning/Dataviz/#famous-tools

- https://datavizcatalogue.com/index.html

- https://python-graph-gallery.com/