







From data to chart: Selecting a chart type

**I can read a chord chart, but I
couldn't read a note chart.**

**Glen Campbell
Musician**

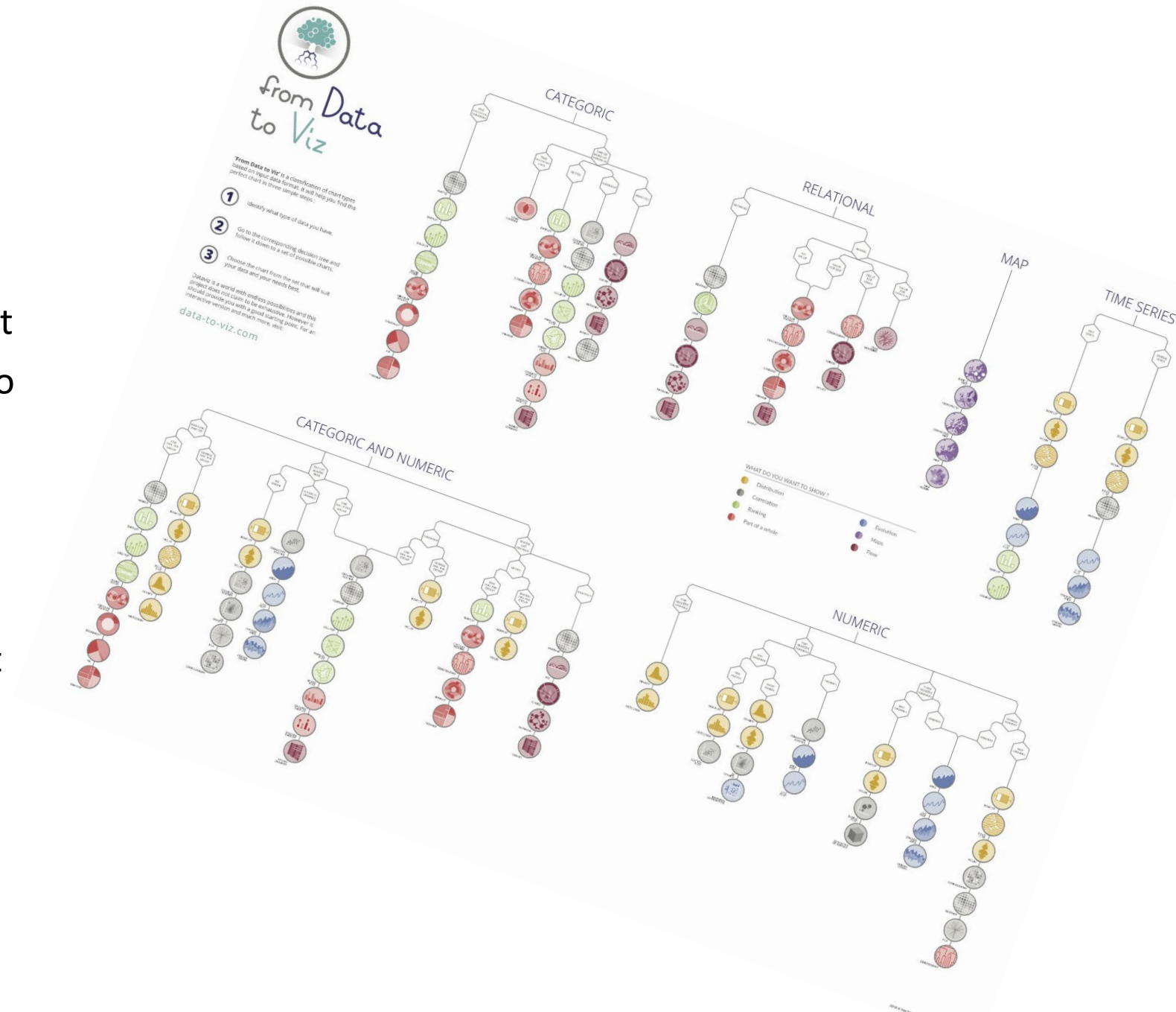
Objectives...

- 1 Relate a chart to the data types to be codified.

- 2 Know a taxonomy of graphs by function.

- 3 Have tools to identify and describe a graph.

- 4 Relate graphs to perception properties.

- 5 Know how to choose the appropriate chart for a data type.

- 6 Know how to choose the right graphic for a task.


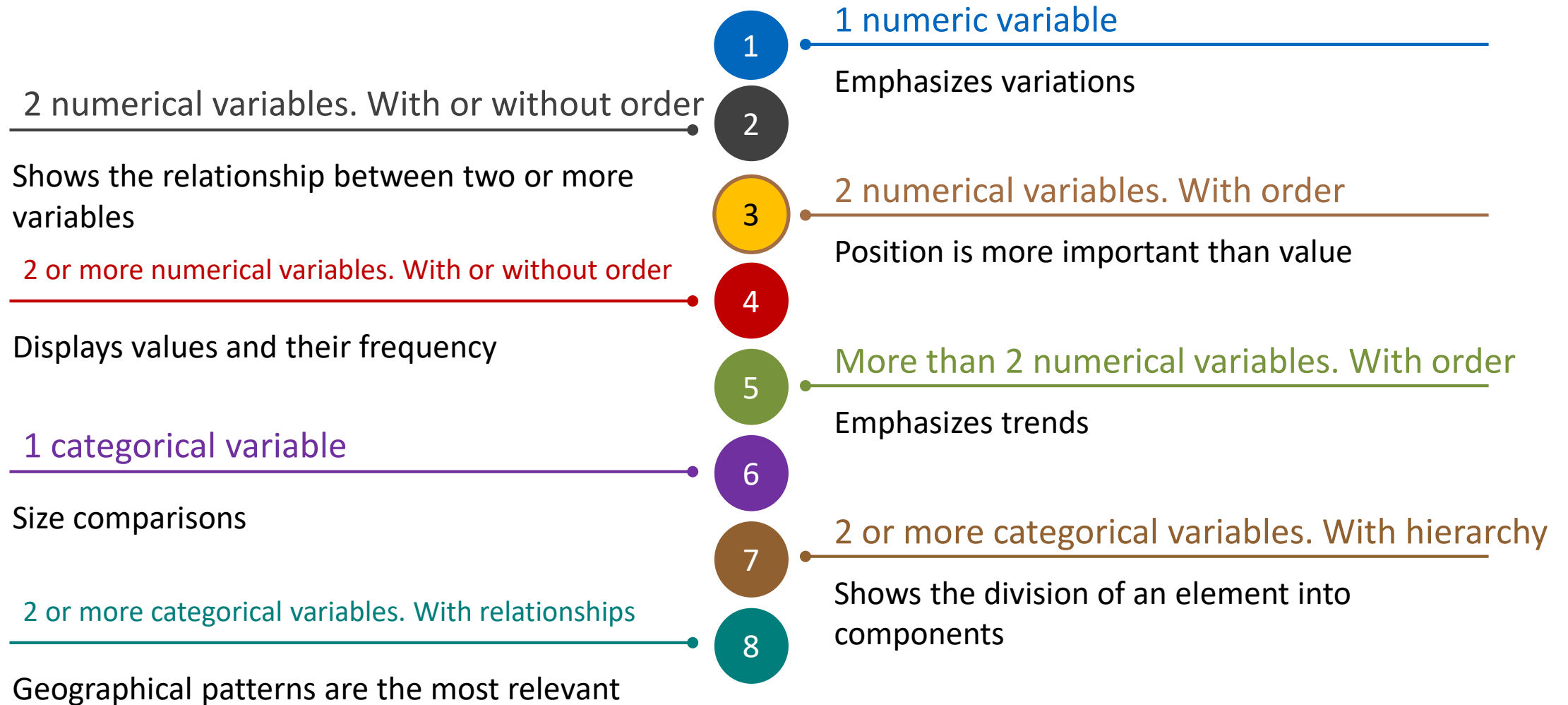
Charts by data type

In the bottom-up approach we start from the data and its possibilities to start exploring different types of graphs and encodings.

This section is based on the project classification “from Data to Viz”



Overview (data)



1

1 numerical variable

Histogram

The variable is grouped into bins and the chart displays the number of observations.



Density diagram

Displays the number of observations without groupings.



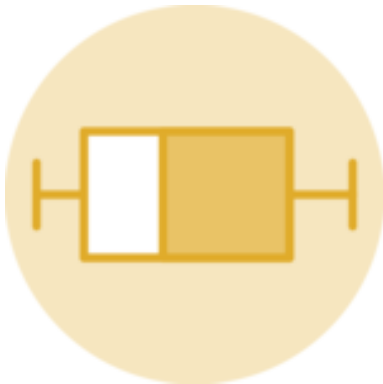
In these graphs we play with the values of the variable (x-axis) and with the count of observations (y-axis). They are based on the property of length.

2

2 numerical variables. With or Without order

Boxplot

It shows the distribution and basic descriptive statistics: median, Q1, Q3 and outliers.



Violin chart

Similar to the boxplot, it is ideal if there are many observations.



Scatter plot

Relates 2 variables.
The X-axis can be categorical.



In boxplot and violin we play with the values of the variable and with the count of observations, its design allows to compare 2 or more variables.

In the scatter plot we show the relationship between 2 variables. They are based on the position property.

3

2 numerical variables. With order

Line chart

Shows the evolution of numerical values.



Area chart

Shows the evolution of numerical values.
Reports on the volume of the value y.
The y-axis must start at 0.



Both graphs allow you to combine observations of 2 or more numerical variables, facilitating their comparison.
They are based on the principle of continuity.

4

2 or more numerical variables. With or without order

Bubble chart

Adds a third dimension to the scatter plot with the size of the points.



Heat map

Adds a third dimension to the scatter plot with the color of the points.



These graphs allow you to display 3 variables with the preattentive properties of position, color and size.

5

More than 2 numerical variables. With order

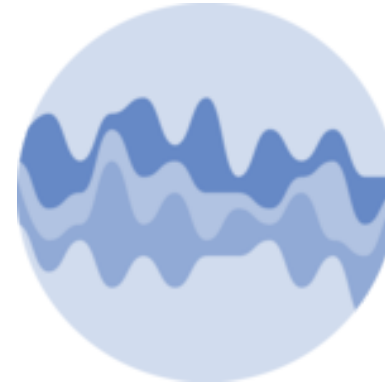
Stacked area chart

Visualizes several variables, one on top of the other. Differences are perceived, not absolute values.



Stream graph

Variation that emphasizes evolution



The x-axis reflects the order in one of the numerical variables and the y-axis its magnitude. They allow the comparison of several variables. They are based on the principle of continuity.

6

1 categorical variable

Word Cloud

Displays the frequency of each word with size or color saturation.



In this graph we play with the values of the variable (words) and with the count of observations (size and/or saturation).

7

2 or more categorical variables. With hierarchy

Tree map

Categories are displayed as overlapping rectangles. The size reflects a numeric attribute.



Dendrogram

Visually displays hierarchy



In these graphs the emphasis is on the partition or grouping relationship. They use the principles of closure and the principle of connection.

8

2 or more categorical variables. With relationships

Network or Graph

Shows the interconnections between various entities.



Sankey diagram

Shows transitions through key steps.



Each categorical variable is represented by an entity, its numerical attributes are encoded with size, length... The emphasis is on their relationship. They use the principle of continuity.

Charts by function

In the top-down approach we start from the audience of the visualization and its objective to select the most appropriate graphic.

In this approach it is important to know the function of each type of chart.

This section is based on the classification of the [Visual Vocabulary](#), [de Financial Times](#).



Overview (function)

Correlation

Shows the relationship between two or more variables

Distribution

Displays values and their frequency

Magnitude

Size comparisons

Spatial

Geographical patterns are the most relevant aspect to display

1

Deviation

Emphasize variations

2

3

Ranking

Ordinal position is more important than value

4

5

Evolution over time

Emphasize trends

6

7

Part of a whole

Shows the division of an element into components

8

9

Flow

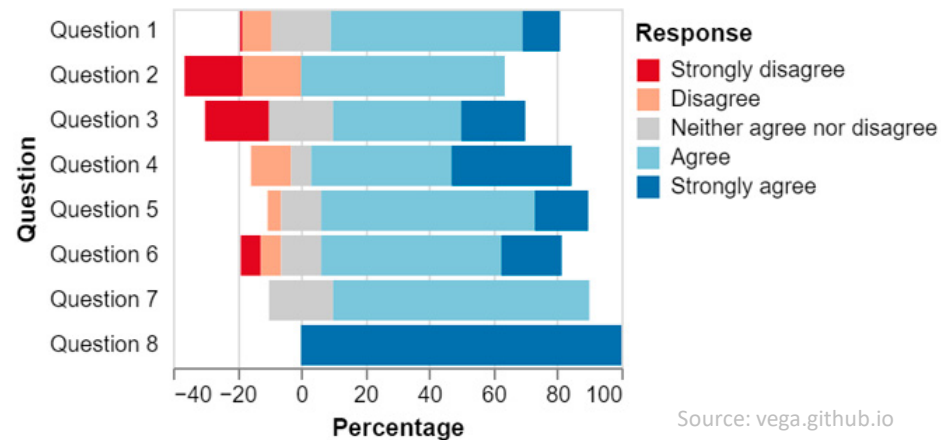
Intensity between two states

1

Deviation

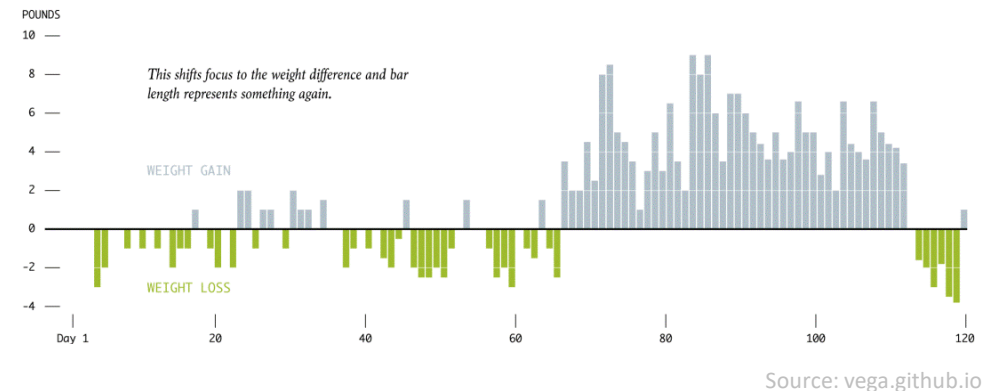
Stacked divergent bars

Perfect for opinion polls



Differential bar chart

A measure separates values into less or greater than it



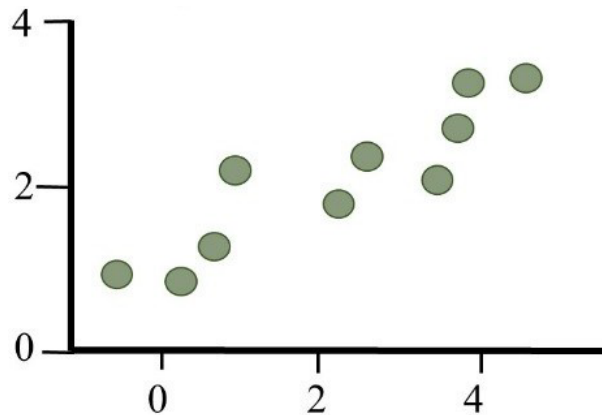
Emphasizes variations (+/-) from a reference point. The reference point is usually 0 but it can also be another relevant measure. Also used to show opinion (positive/neutral/negative)

2

Correlación

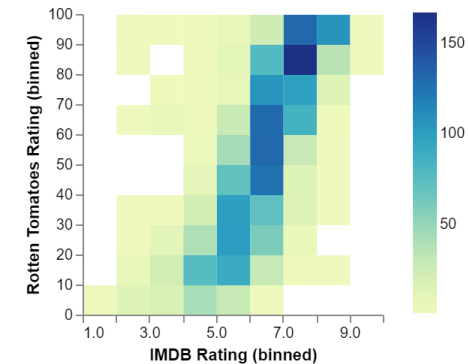
Scatter plot

Shows correlation types: positive, negative, nonexistent



Heat map

Ideal for showing patterns.
Enables high data density.



Source: vega.github.io

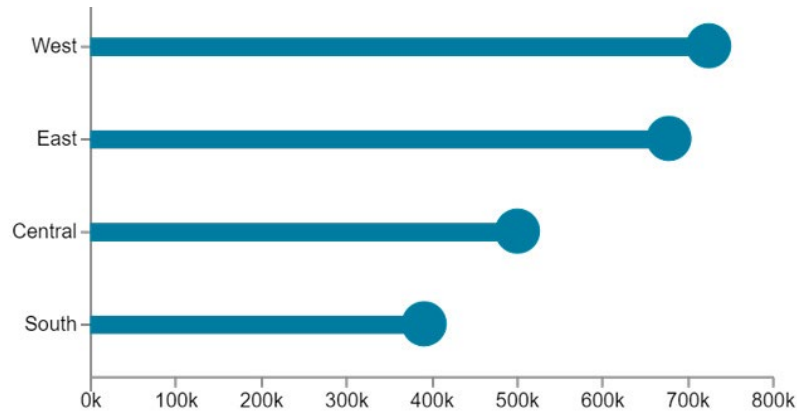
Correlation plots show the relationship between 2 or more variables. Often readers will assume that it is causal, so it must be made explicit that only correlations are shown.

3

Ranking

Lollipop chart

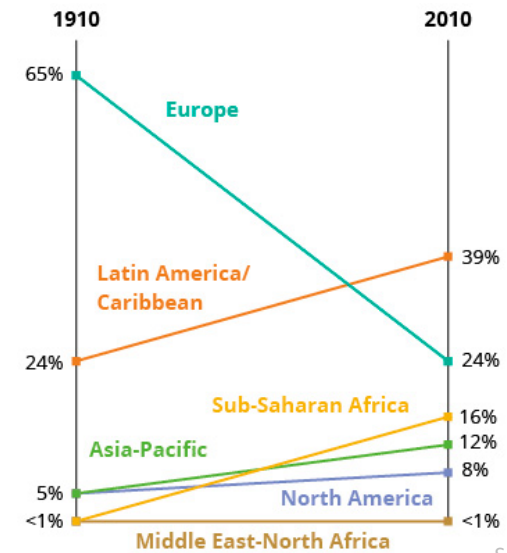
Ordered values with emphasis on the final value



Source: Vega Authors & Community, Gramener

Slope chart

Variation in ranking over time



Source: The Huffington Post

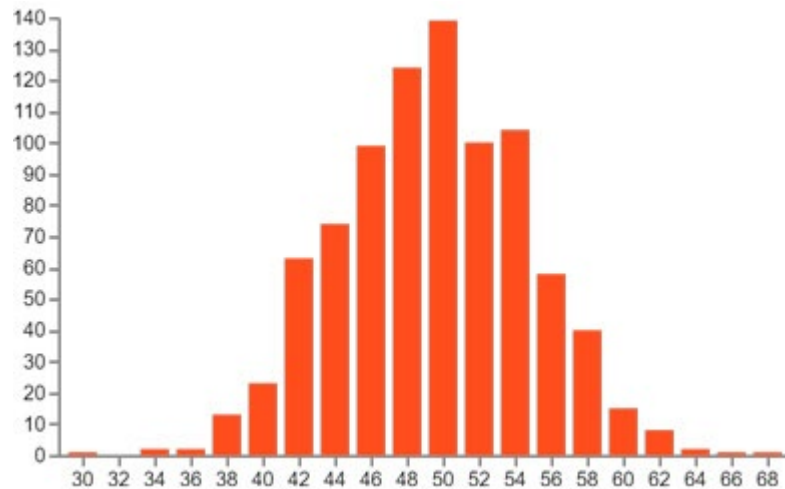
They are used if the ordinal position of an item relative to others is more important than its value. Points of interest can be highlighted

4

Distribution

Histogram

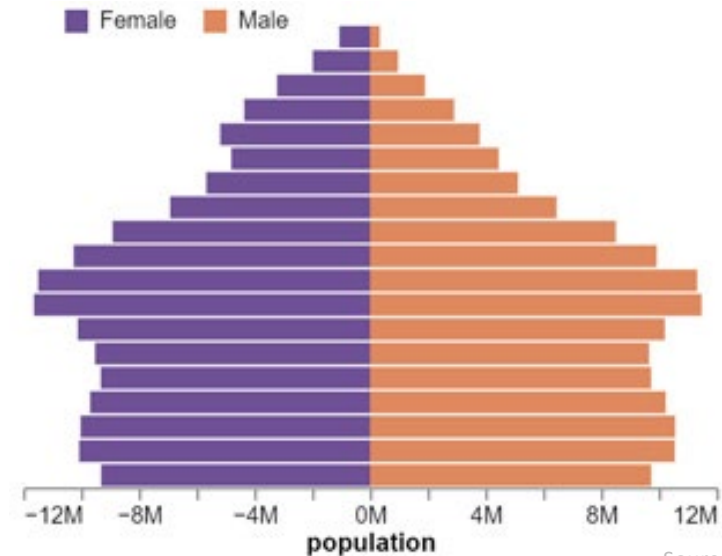
Displays statistical distributions



Source: Vega Authors & Community, Gramener

Butterfly chart

Displays a value according to 2 categories



Source: vega.github.io

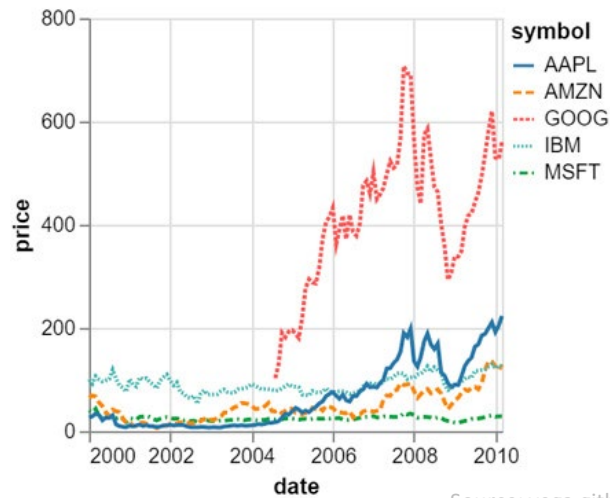
Displays the values and their frequency. The form (or bias) of the distribution is a way of highlighting the heterogeneity of the data and its normality.

5

Evolution over time

Line chart

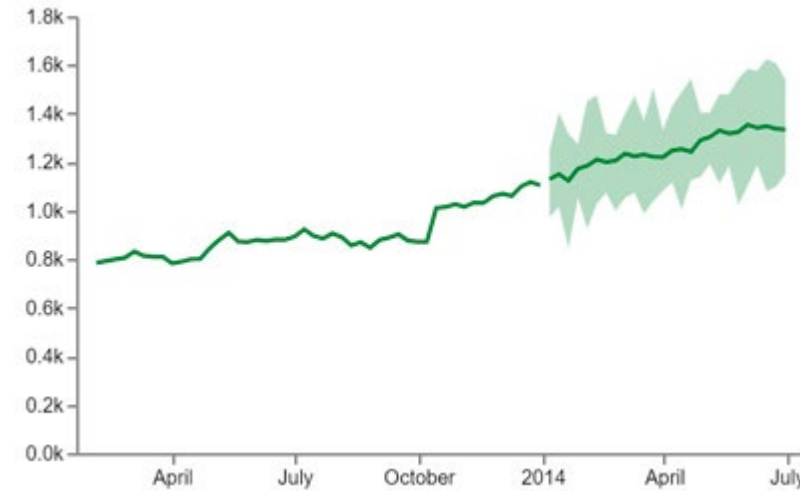
The standard chart for showing changes in a time series



Source: vega.github.io

Fan chart

To show uncertainty



Source: Vega Authors & Community, Gramener

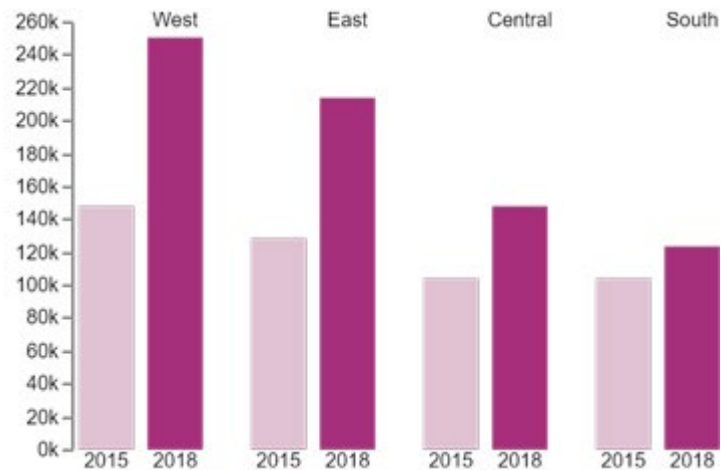
Charts to emphasize trends. Choosing the right period of time is key to providing the reader with the right context.

6

Magnitude

Parallel columns

Perfect for comparison. Maximum 2 categories



Source: Vega Authors & Community, Gramener

Bullet chart

Displays a value in the context of various reference measures



Source: Vega Authors & Community, Gramener

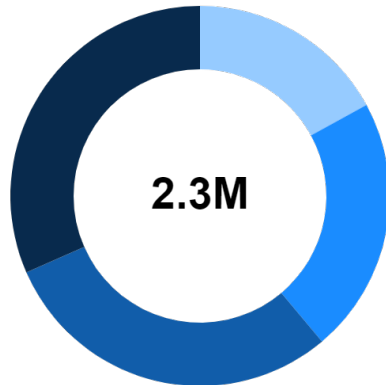
They show relative or absolute size comparisons.
Typically, with discrete values

7

Part of a whole

Doughnut chart

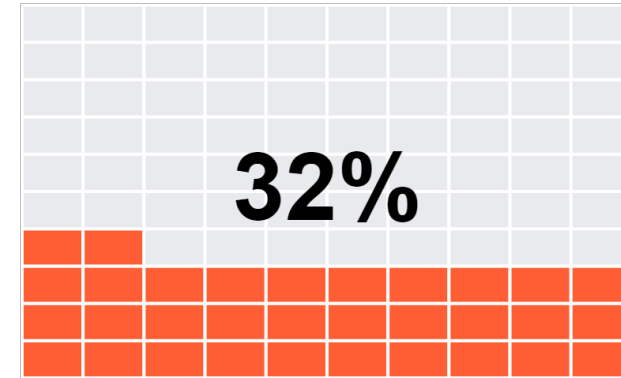
Do not use with many categories. The emphasis is on proportion, not exact measure.



Source: Vega Authors & Community, Gramener

Waffle chart

Ideal for displaying percentages



Source: Vega Authors & Community, Gramener

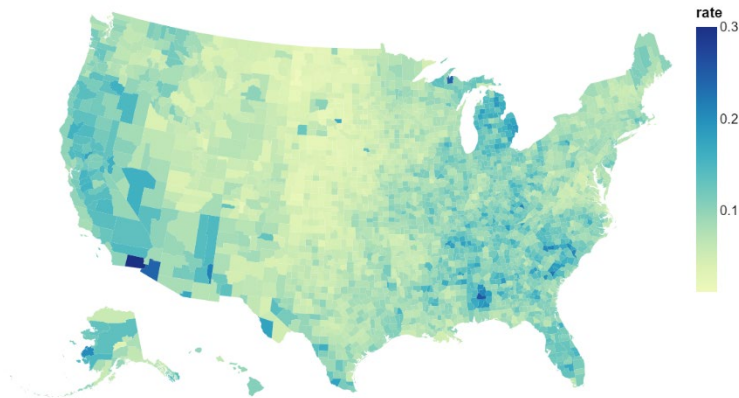
They show how a single element is divided into different components. If the most important thing is the size of the components, better to use a magnitude graph.

8

Spatial

Choropleth map

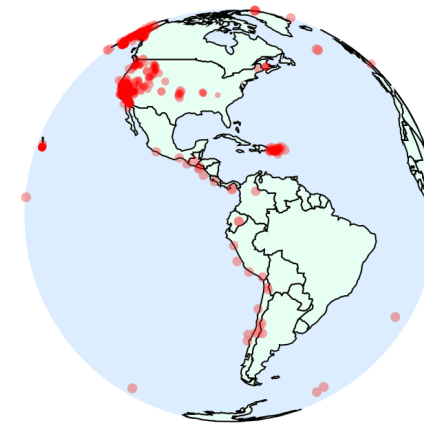
Use with ratio



Source: Vega-Altair

Symbol Map

Absolute values with their exact location



Source: vega.github.io

Use these charts only when there are precise locations or when geographic patterns in the data are most important to the reader.

9. Flow

Some well-known graphics

Bar chart

It relies on length to show magnitude. Ideal for comparisons, it should always start at 0.

Cannot display a lot of data at once

1 numeric attribute

1 or more categorical attributes

Line chart

Similar to the scatter plot, it emphasizes the connection of points. The x-axis must be a continuous variable, often time. It's a widely adopted chart.

Scatter plot

It is based on position to show magnitude.





You can add other attributes with size and colors, and be enriched with the regression line. It's a well-known graphic.

1 numeric attribute

1 categorical attribute

If our audience is not very familiar with graphics, we will prioritize highly adopted graphics.

Key ideas

-  1 For a visualization to be effective, it will be necessary to know the data, the audience and the data coding techniques in interactive graphic language.
-  2 The selection of a chart depends on the type of data and its function.
-  3 Adoption and familiarity is also a criterion to keep in mind.
-  4 The graphics rely on preattentive properties and Gestalt principles to be very efficient.

Sources

- ★ Financial Times Visual Vocabulary [Internet] <https://github.com/Financial-Times/chart-doctor/tree/main/visual-vocabulary>
- ★ From Data to Viz [Internet] <https://www.data-to-viz.com/about.html>
- ★ Visual Vocabulary in Tableau:
<https://www.tableau.com/solutions/gallery/visual-vocabulary>

Thanks you for your attention



MIREIA RIBERA | ACCESSIBILITAT DIGITAL
EXPERIÈNCIA D'USUARI
VISUALITZACIÓ DE DADES



UNIVERSITAT DE
BARCELONA