

DATA VISUALISATION METHODS

DR DANNY POO

BIG DATA ANALYTICS AND VISUALISATION

Taxonomy

Classification	Communication Purpose
COMPARING CATEGORIES	To facilitate comparisons between the relative and absolute sizes of categorical values. The classic example would be the bar chart.
ASSESSING HIERARCHIES AND PART-TO-WHOLE RELATIONSHIPS	To provide a breakdown of categorical values in their relationship to a population of values or as constituent elements of hierarchical structures. The example here would be the pie chart.
SHOWING CHANGES OVER TIME	To exploit temporal data and show the changing trends and patterns of values over a continuous timeframe. A typical example is the line chart.
PLOTTING CONNECTIONS AND RELATIONSHIPS	To assess the associations, distributions, and patterns that exists between multivariate datasets. This collection of solutions reflects some of the most complex visual solutions and usually focuses on facilitating exploratory analysis. A common example would be the scatter plot.
MAPPING GEO-SPATIAL DATA	To plot and present datasets with geo-spatial properties via the many different mapping frameworks. A popular approach would be the choropleth map.

Factors for Considerations

Does it **accommodate** the physical properties of your data?

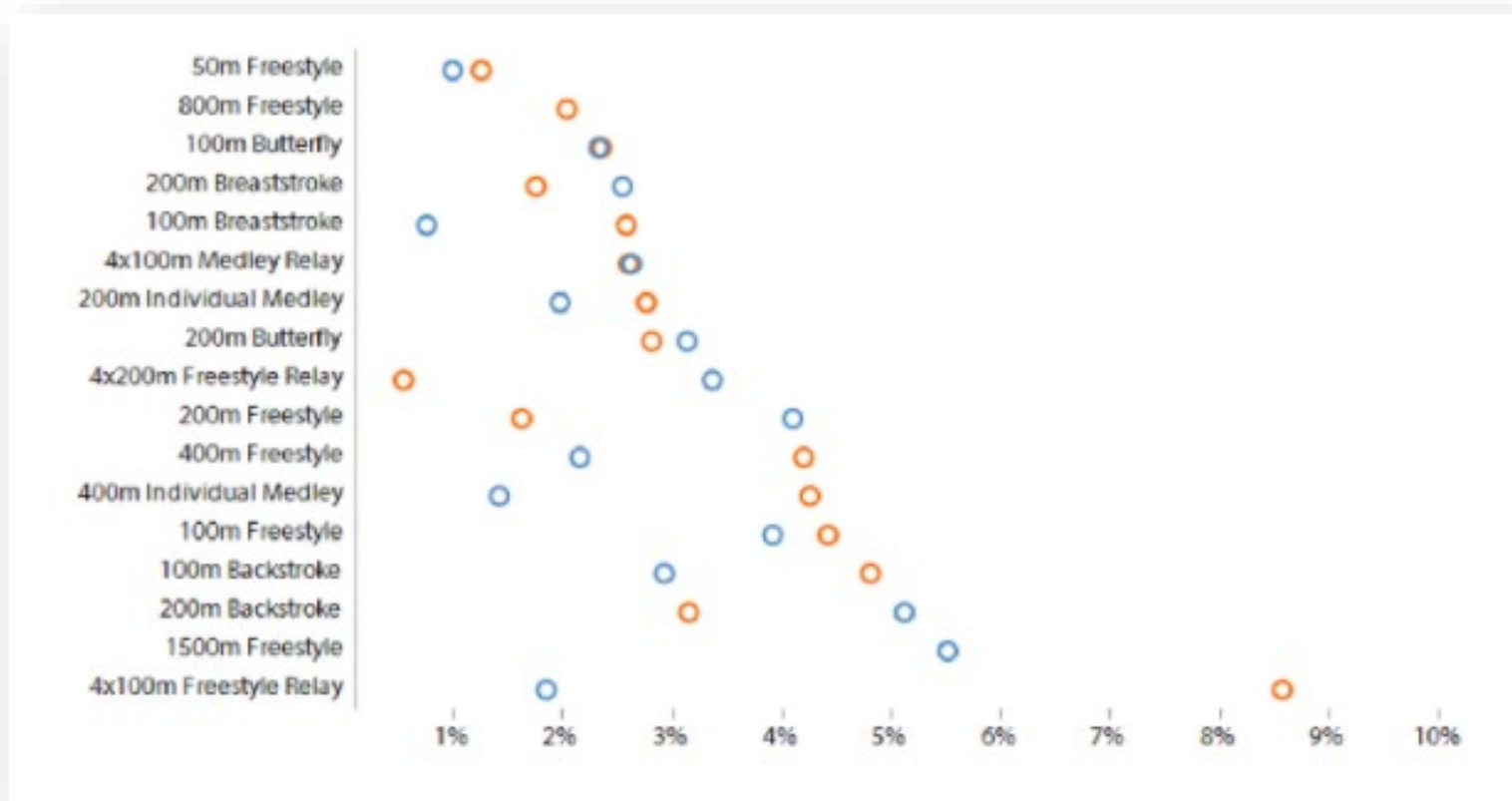
Does it **facilitate** the desired degree of accuracy?

Is it potentially capable of **conveying** a certain metaphorical and design consistency with our subject matter?

COMPARING CATEGORIES

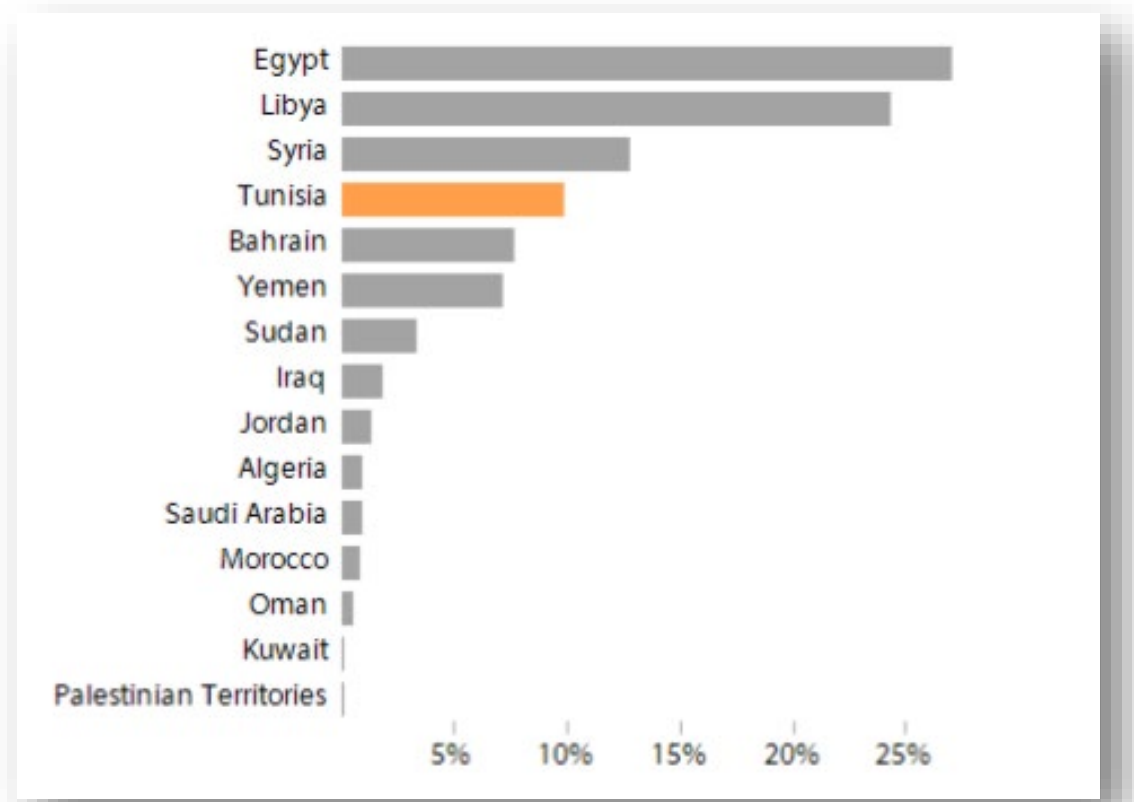
Dot Plot

Description: A dot plot **compares categorical variables** by representing quantitative values with a single mark, such as a dot or symbol.



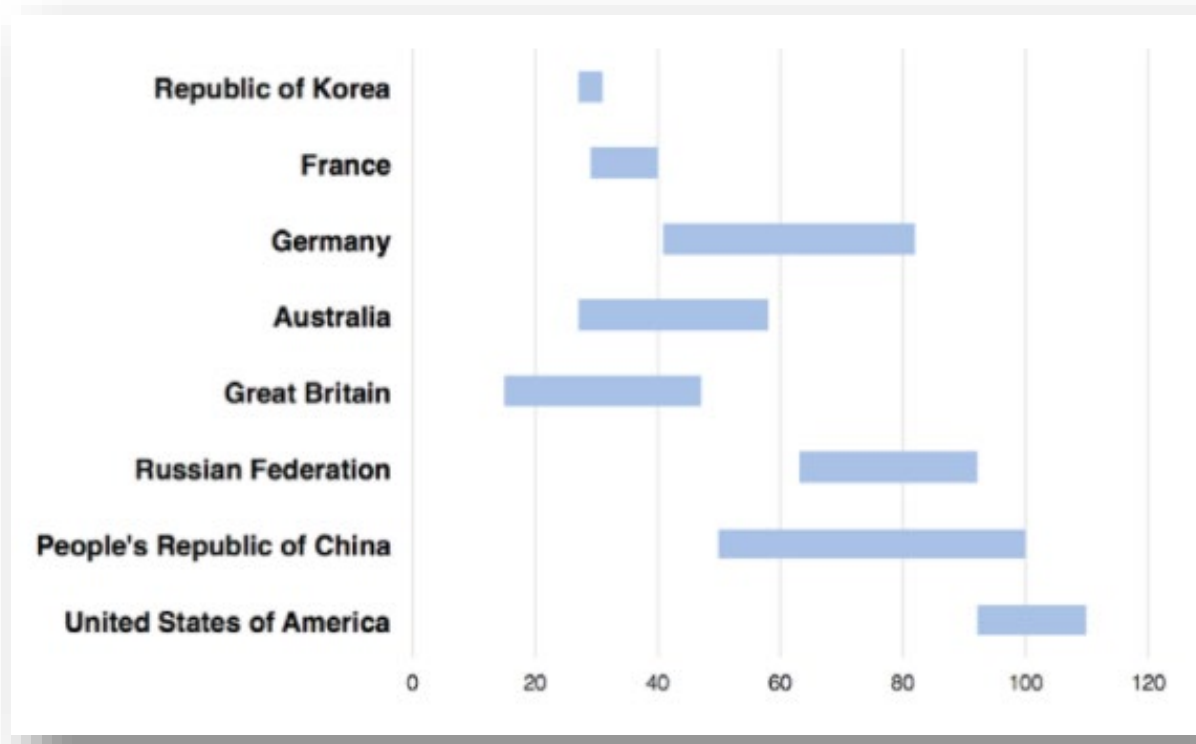
Bar Chart (aka Column Chart)

Description: Bar charts convey data through the length or height of a bar, allowing us to draw accurate comparisons between categories for both relative and absolute values.



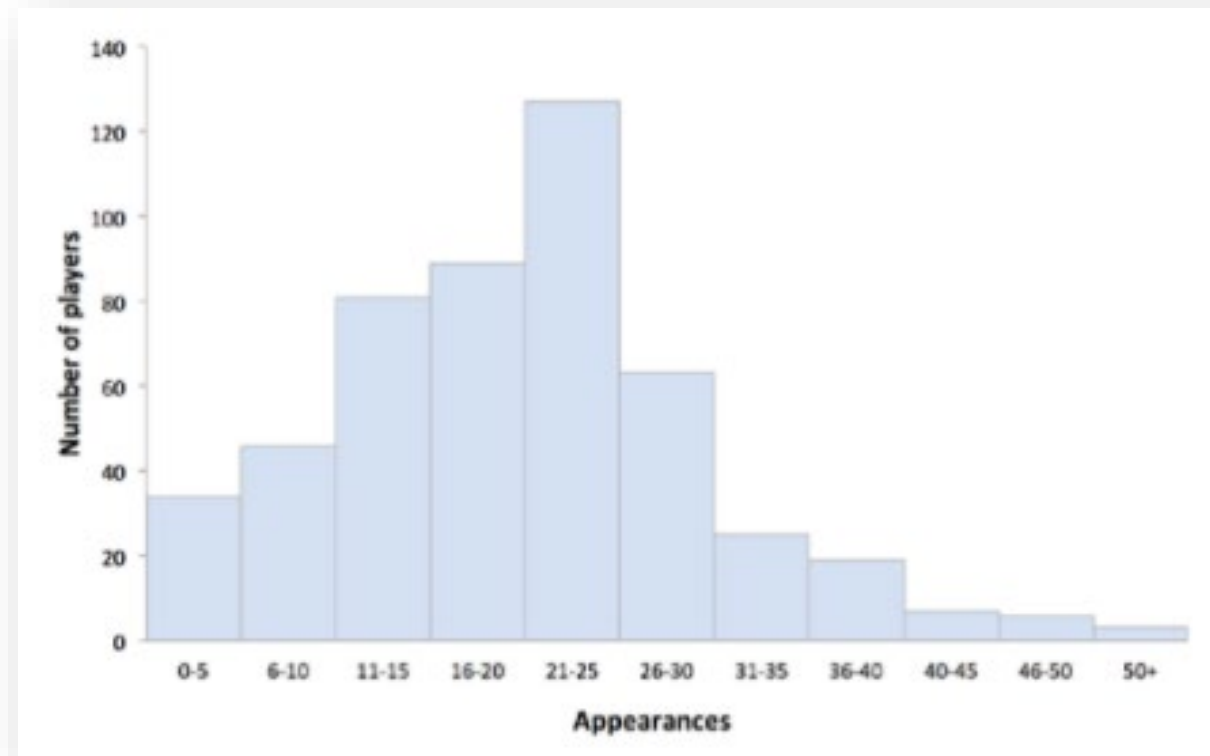
Floating Bar

Description: A floating bar chart – sometimes labeled a Gantt chart because of similarities in appearance – helps to **show the range of quantitative values**. It presents a bar stretching from the lowest to the highest values (therefore the starting position is not the zero point).



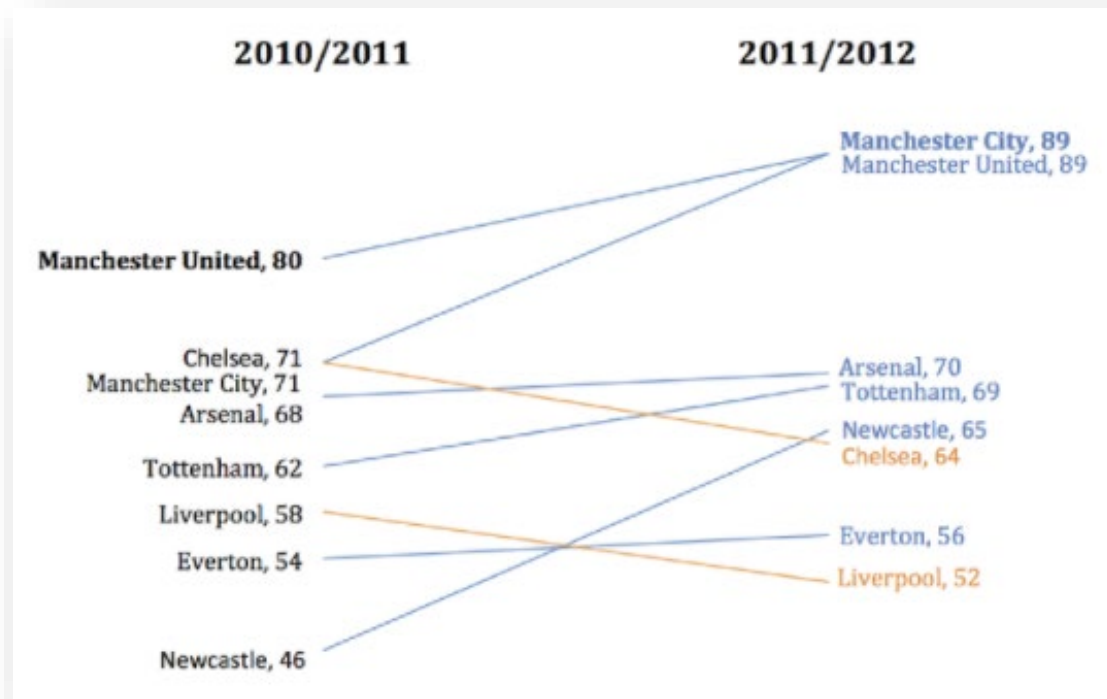
Histogram

Description: Histograms show distribution through the frequency of quantitative values (y axis) against defined intervals of quantitative values(x axis). By contrast, bar charts facilitate comparison of categorical values.



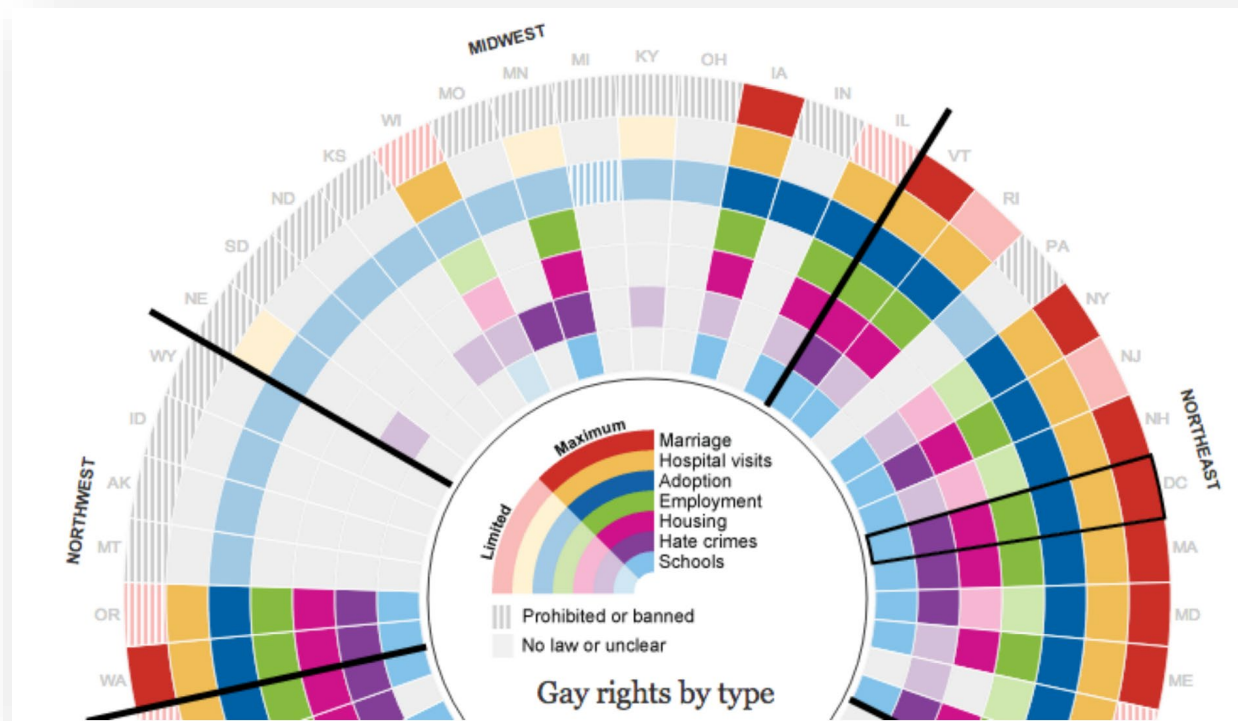
Slopegraph

Description: A slopegraph creates an effective option for **comparing two (or more) sets of quantitative values when they are associated with the same categorical value**. They especially provide a neat way of showing a before and after view or comparison of two different points in time. Colour can be used to further emphasize upward or downward changes.



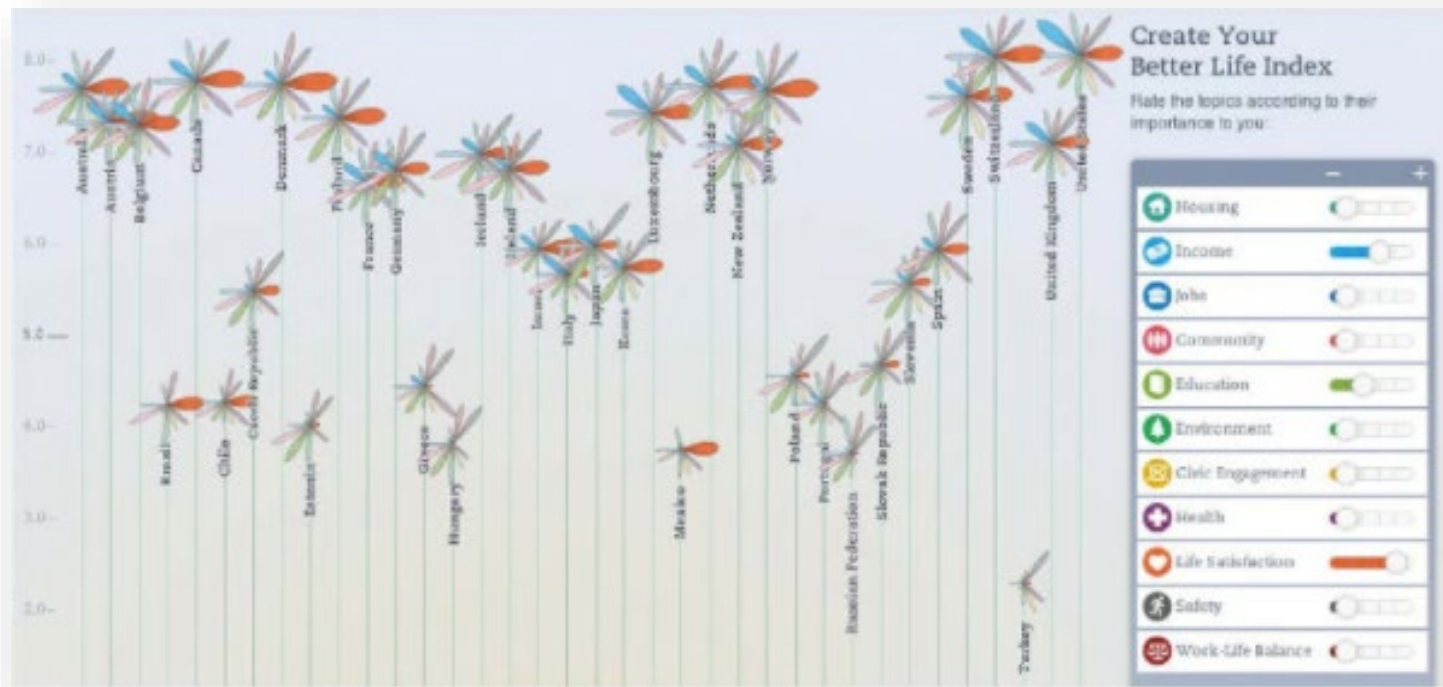
Radial Chart

Description: A radial chart displays data around a concentric, circular layout. A slight visual shortcoming associated with a radial chart is the fractionally distorted prominence of the segments on the outside rings which end up being larger (due to arc length) than those on the inside.



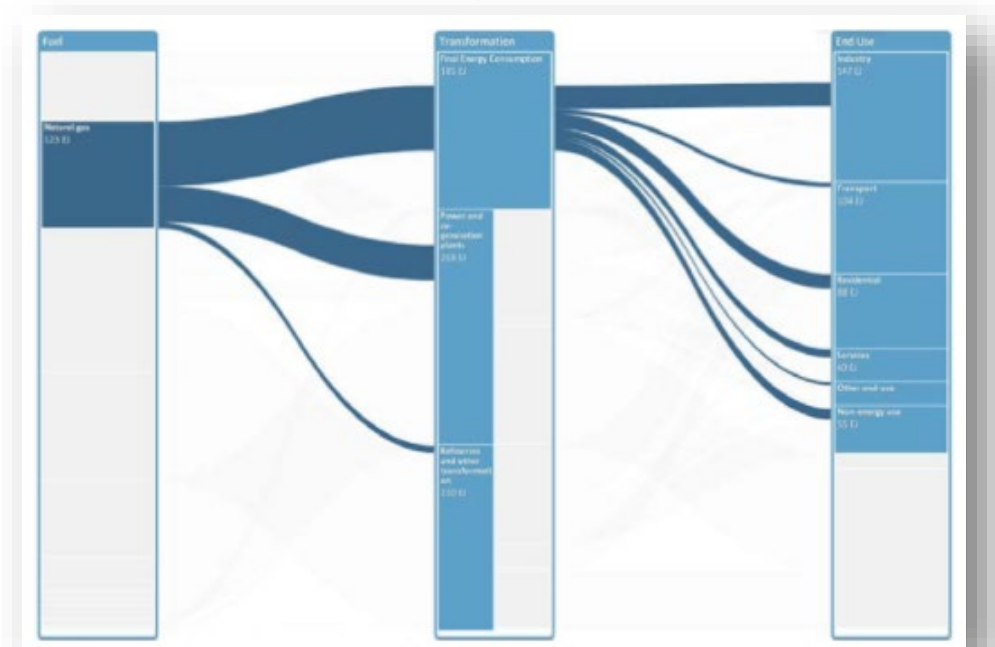
Glyph Chart

Description: A glyph chart is based on a shape (e.g. a flower) being the main artefact of representation. The physical properties of the shape (through a feature such as a petal) represent different categorical variables; they are sized according to the associated quantitative value and distinguished through colour.



Sankey Diagram

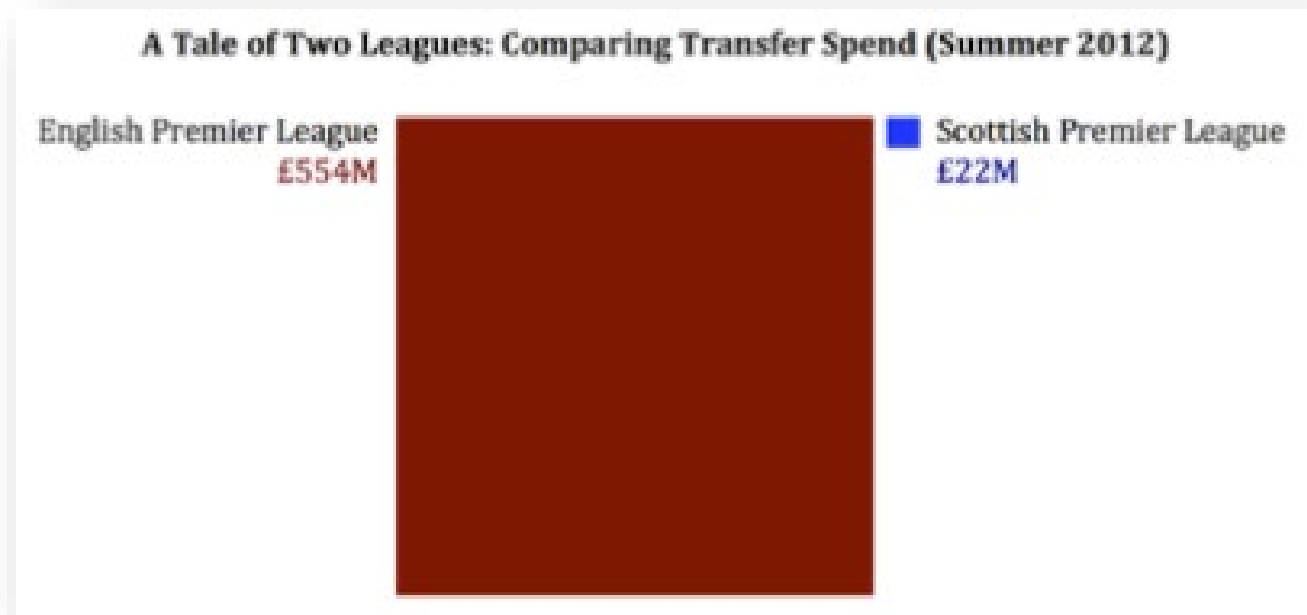
Description: Sankey diagrams are used to **convey the idea of flow**. They portray constituent quantities of a series of associated categorical values, across a number of “stages”, with the ongoing associations represented by connecting bands. The width of these links indicates the proportional flow from one stage to another. **They are useful for showing situations where elements transform and divide over key events.**



Breakdown of different fuels, how they are transformed and then ultimately used.

Area Size Chart

Description: It is a very simple visual device that **deploys the visual variable of area** – normally a rectangle or circle – to **compare two (or maybe several) values**. Normally these values will vary in size quite dramatically to convey a certain shock at the disparity. The subject matter may relate to a part-of-a-whole comparison (portion judgment) but more typically involves separate, independent categories (comparative judgment).



Word Cloud

Description: Word clouds depict the frequency of words used in a given set of text. The font size indicates the quantity of each word's usage. Colour is often just used as decoration. They can be useful for exploring datasets for the first time in order to identify key terms being used. Need to ensure the underlying text being used is carefully prepared in advance to reduce the noise.



ASSESSING HIERARCHIES AND PART-TO-WHOLE RELATIONSHIPS

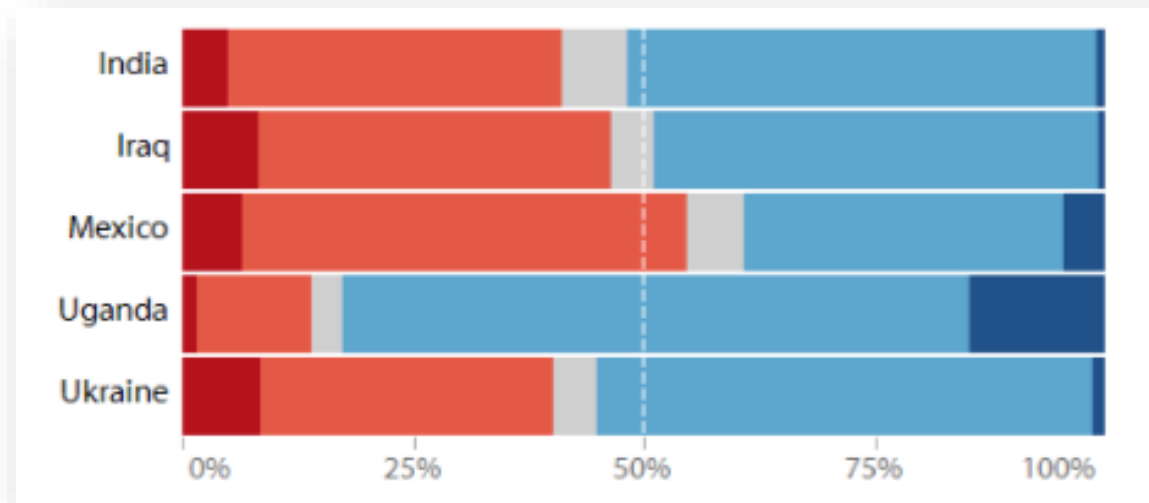
Pie Chart

Description: Always start the first slice from the vertical position (to establish a sense of baseline), minimize the number of categories being displayed (ideally maximum of three), and arrange the segments as logically as possible.



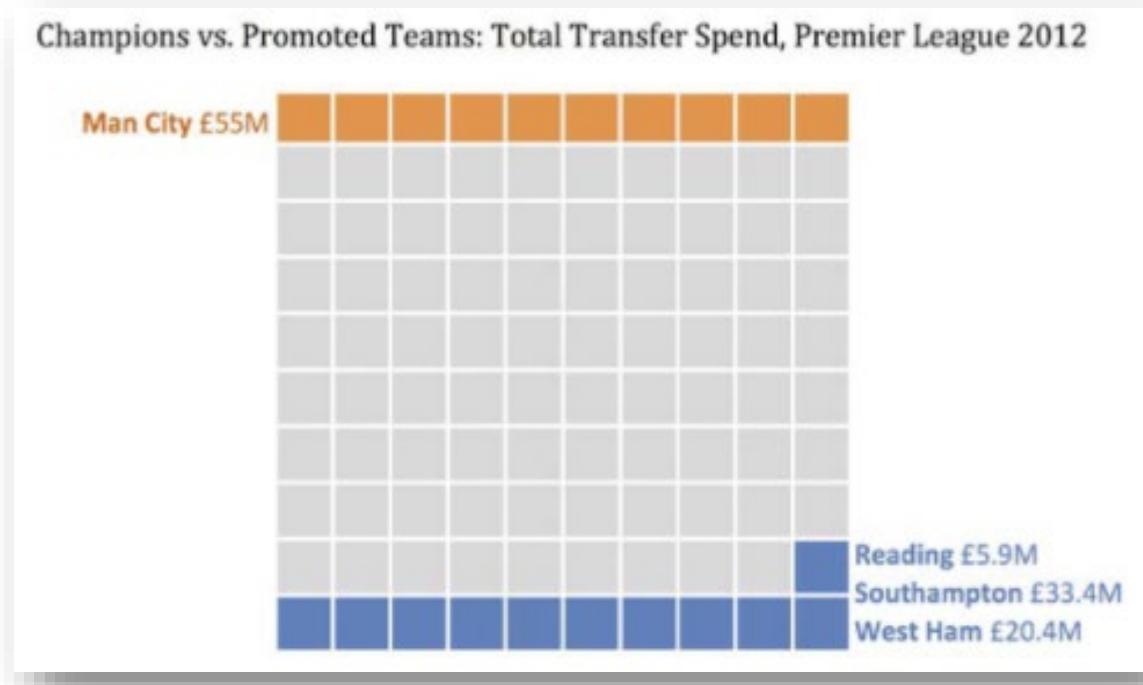
Stacked Bar Chart (aka Stacked Column Chart)

Description: Stacked bars are based on the **stacks of absolute values or standardized to show part of a whole breakdown**. Colours and position differentiate the value categories. Where the categorical values are ordinal in nature, it helps to sequence the values logically, for example when you have sentiment data such as the Likert scale of disagree (**reds**) through to agree (**blues**). This sequencing helps draw out the contrasting composition of the sentiment from all categories. The only drawback of a stacked chart is the difficulty in being able to accurately read bar lengths, as there is no common baseline.



Square Pie (aka Unit Chart, Waffle Chart)

Description: This technique involves a grid of units (may be squares or symbols) to **represent parts of a whole**. This may be for a percentage comparison (square pie) or an absolute quantity (unit chart, waffle chart). The use of colour and symbol establishes the visual composition of the categorical and quantitative values.



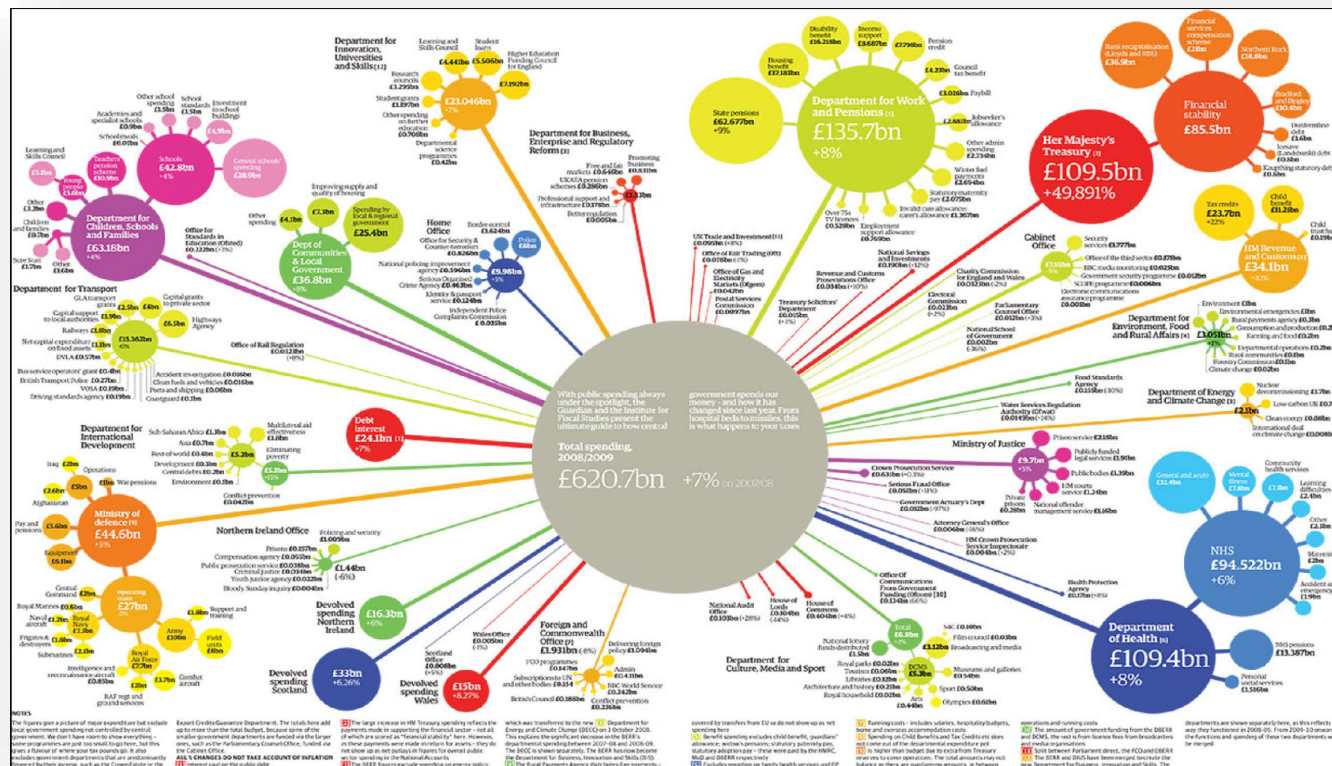
Tree Map

Description: Tree maps take the concept of a whole population and divide up portions of rectangular spaces within to represent organized, clustered constituent units sized according to their relative value. As well as arrangement, various properties of colour are typically used to provide additional layers of quantitative or categorical insight.



Bubble Hierarchy

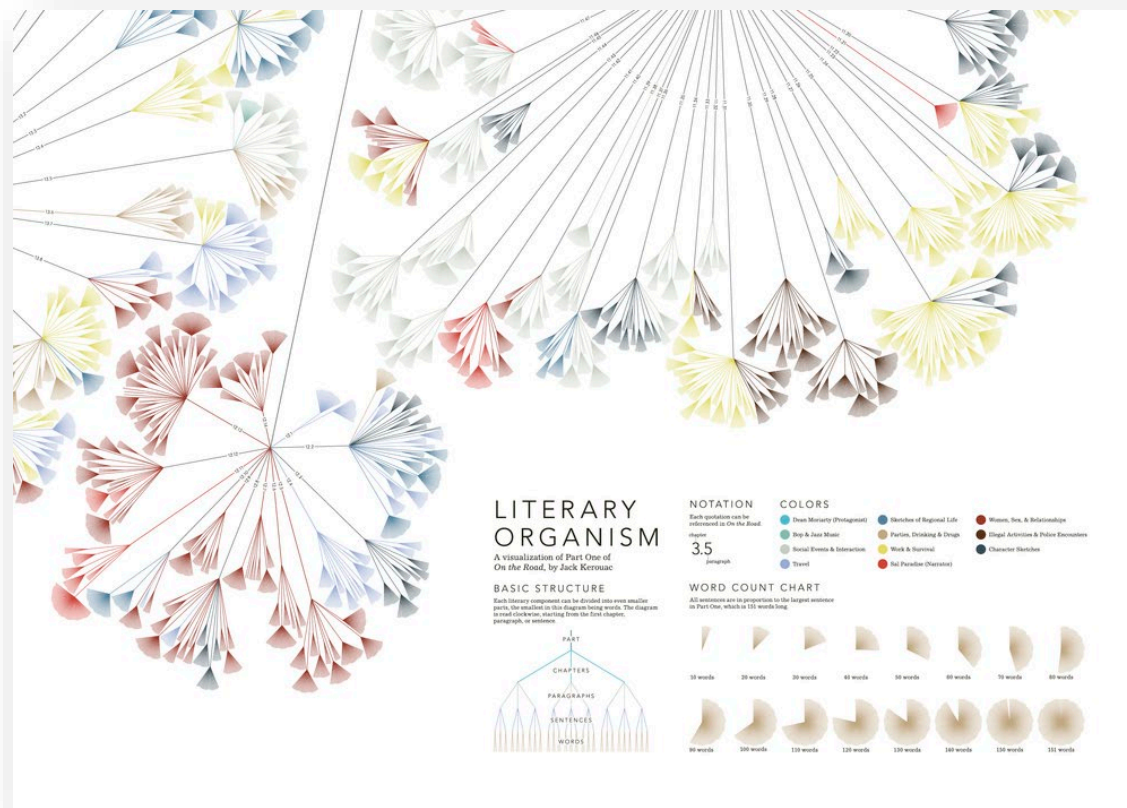
Description: This technique is used to portray organization and structure through a hierarchical display.



The use of circles to represent the constituent departments, sized according to their quantitative value and coloured to visually distinguish the different departments.

Tree Hierarchy

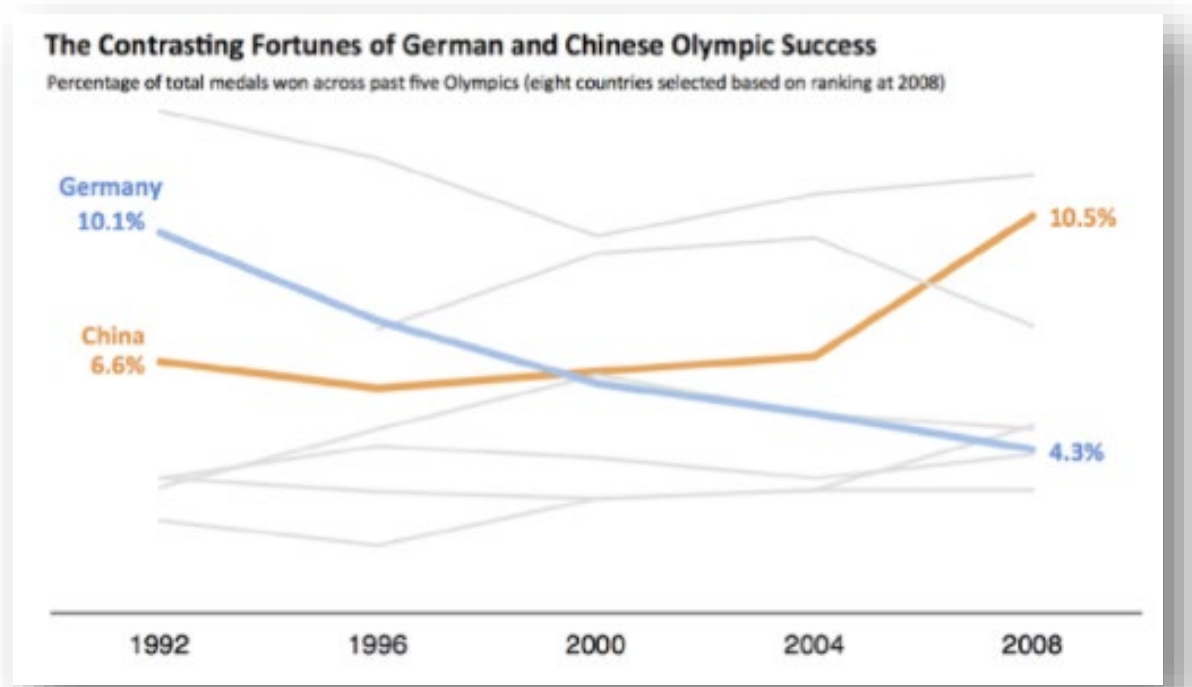
Description: Similar to the bubble hierarchy, this technique presents the organization and structure of data through a hierarchical tree network.



SHOWING CHANGES OVER TIME

Line Chart

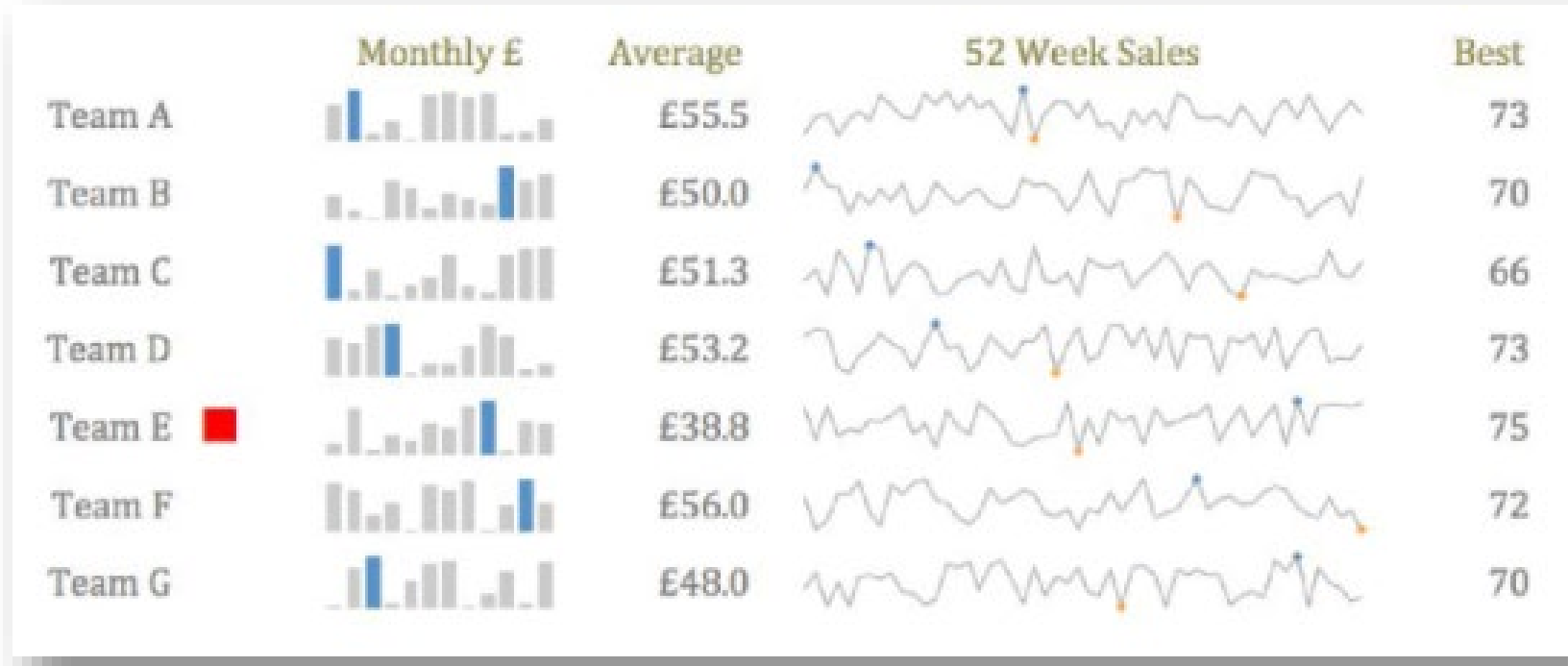
Description: Line charts are used to compare a continuous quantitative variable on the x axis and the size of values on the y axis. The vertical points are joined up using lines to show the shifting trajectory through the resulting slopes. Line charts can help unlock powerful stories of the relative or (maybe) related transition of categorical values.



Unlike bar charts, the y axis does not need to start from zero because we are looking at the relative pattern of the data journey.

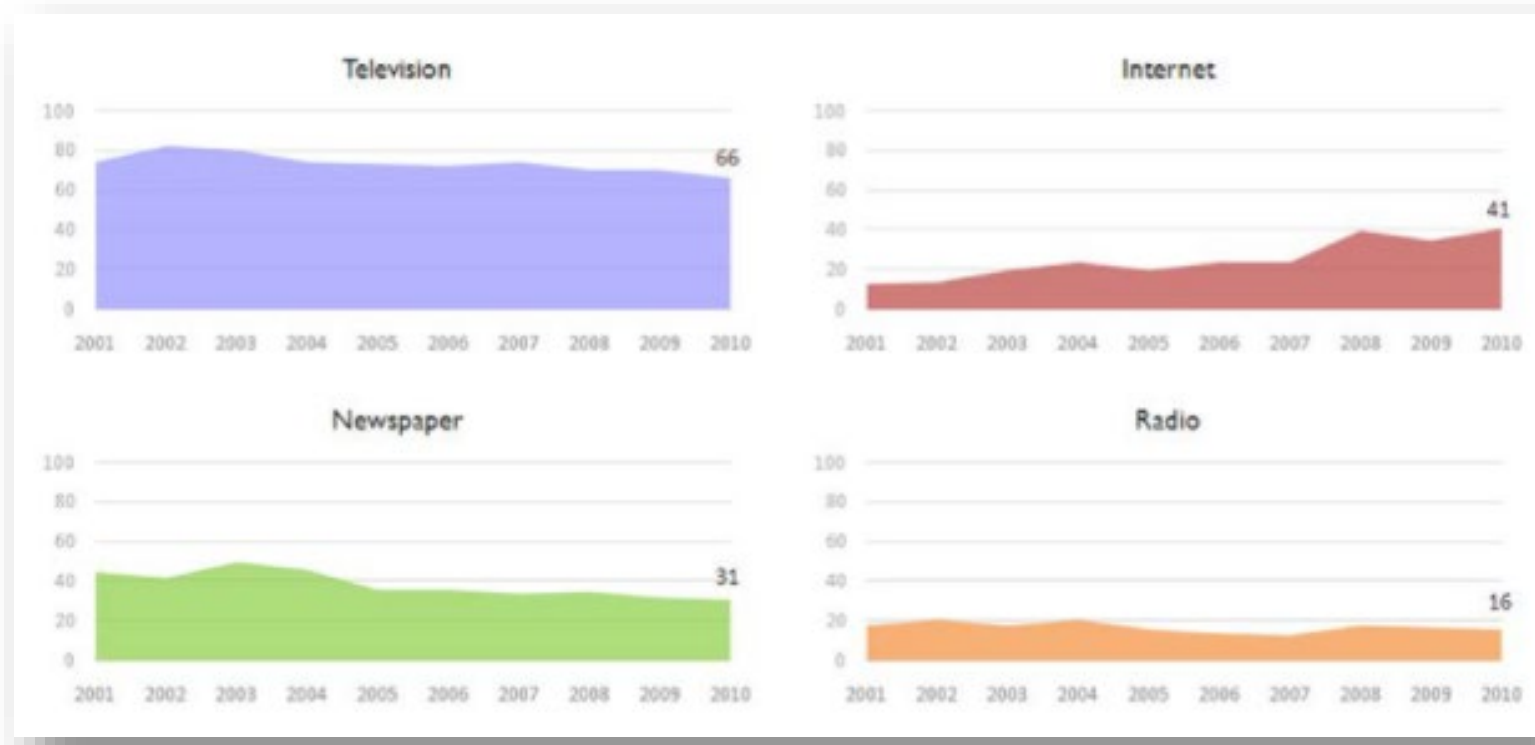
Sparklines

Description: Sparklines are not necessarily a variation on the line chart, rather, a clever use of them. They take advantage of our visual perception capabilities to discriminate changes even at such a low resolution in terms of size. They facilitate opportunities to **construct particularly dense visual displays of data in small space and so are particularly applicable for use on dashboards.**



Area Chart

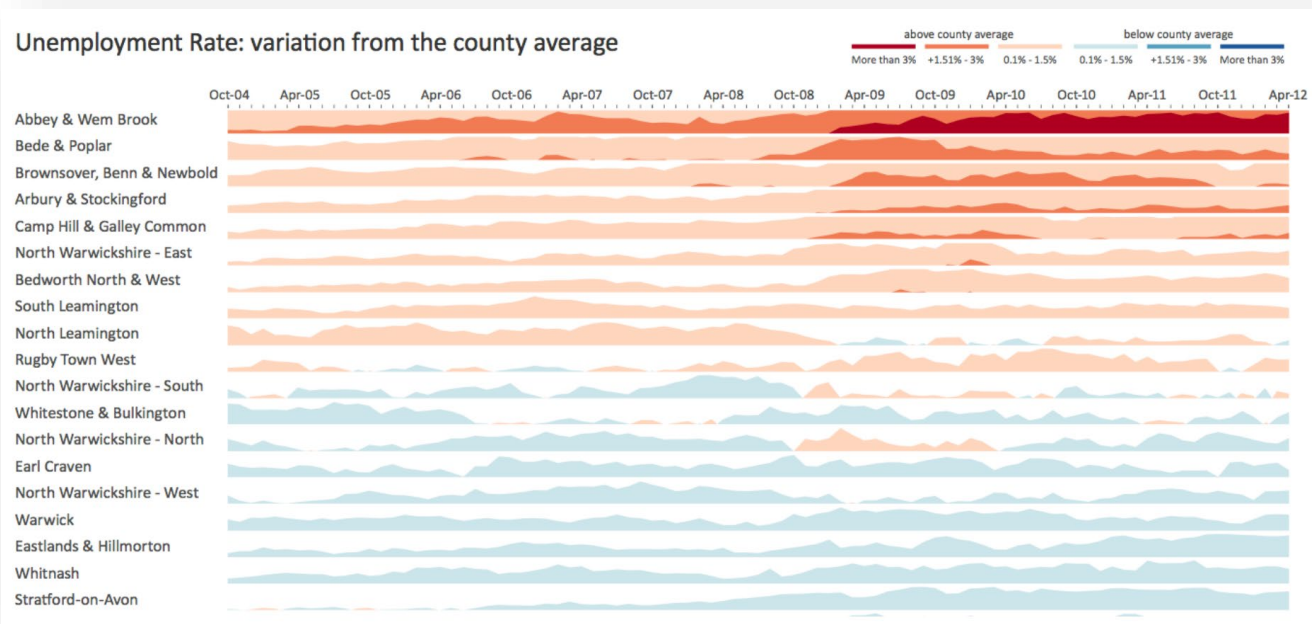
Description: A number of visual properties are involved in area charts. The vertical position and connecting **slope of the horizon** (like a line chart) **shows the progression of the values over time** and the **colour area underneath the chart helps to emphasize these changes**.



Unlike a standard line chart, an area chart should have the y axis starting at zero to ensure the area judgment is being interpreted accurately.

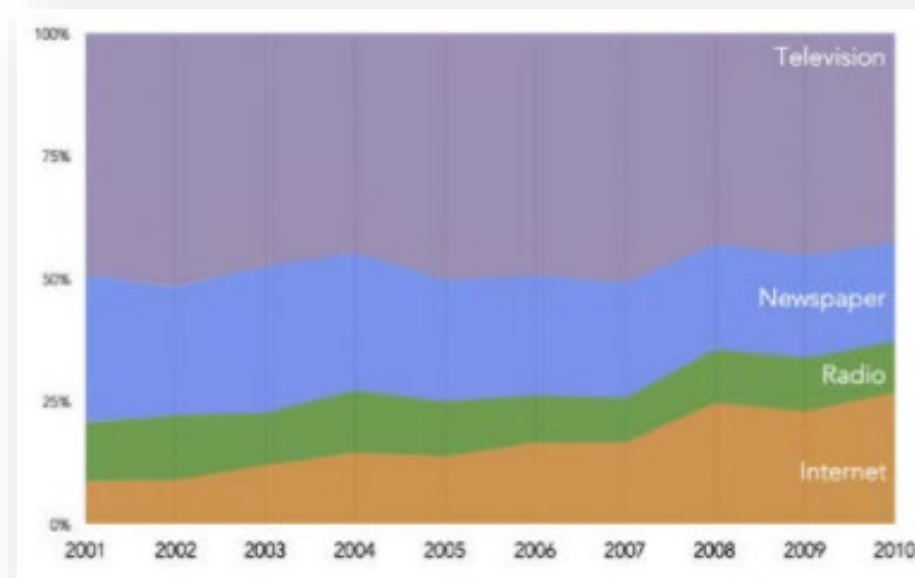
Horizon Chart

Description: This is a variation on the area chart, **modified to include both positive and negative values**. Rather than presenting negative values beneath the x axis, the negative area is mirrored on to the positive side and then coloured differently to indicate its negative polarity. The result is a chart that occupies a single row of space, which helps to accommodate multiple stories onto a single display and facilitates comparison to pick out local and global patterns of change over time.



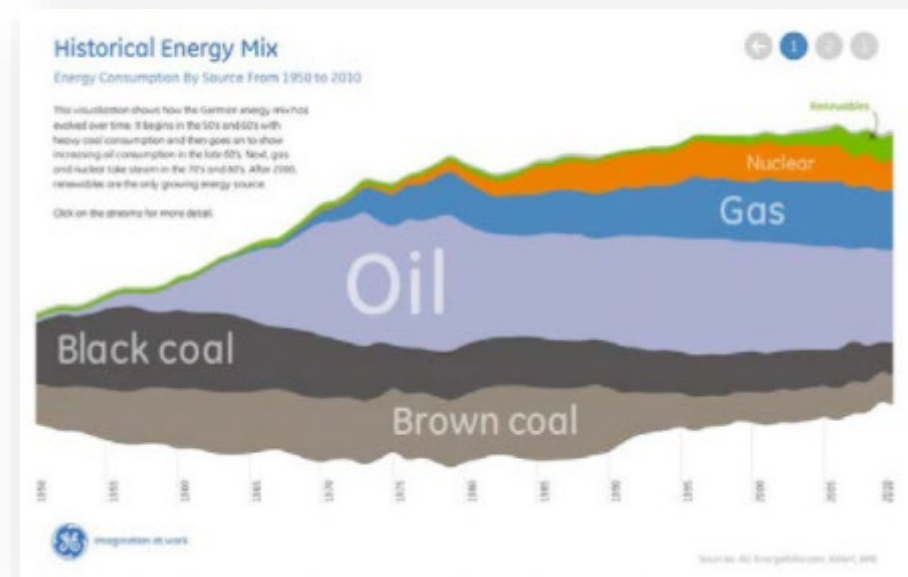
Stacked Area Chart

Description: A stacked area chart provides a **compositional view of categories to show their changes over time**. These are based on stacks of area charts differentiated by colour and present either absolute aggregates or percentage aggregates. Note that the quantitative values are represented by the height (derived from top and bottom positions) of the area stacks at any given point. Sometimes the resulting shapes of the middle sections can be slightly misleading and misinterpreted due to the lack of a common baseline position.



Stream Graph

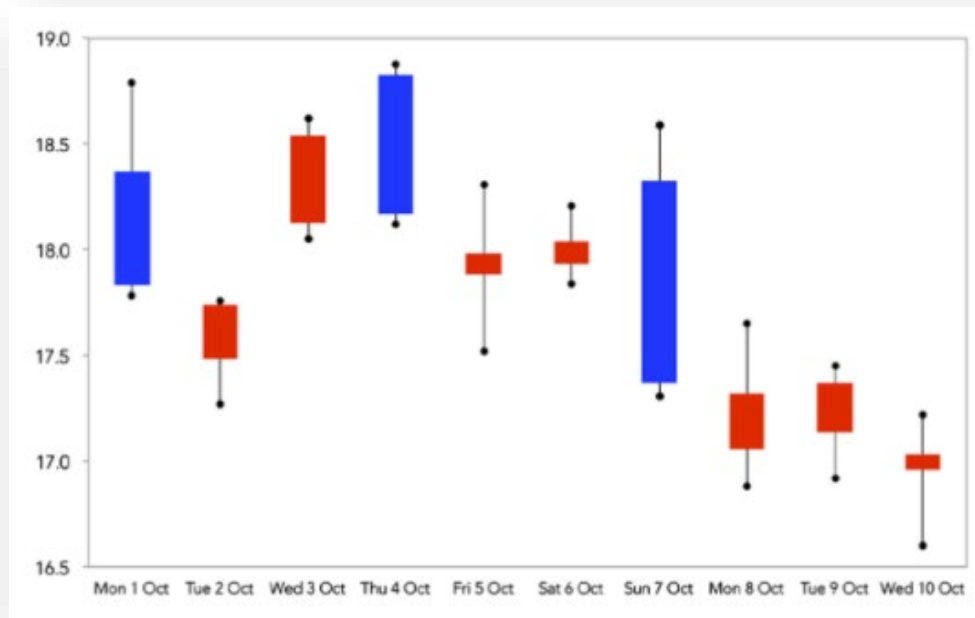
Description: The stream graph operates in a similar fashion to a stacked area chart, **allowing multiple values series to be layered as streams of area with quantitative values expressed through the height of the individual stream at any given time**. It has no baseline x axis and so there is no concept of negative or positive values, purely aggregates. Its functional purpose is really to highlight peaks and troughs – it has a particularly organic feel and is suited to displays intended to show “ebb and flow” stories.



Many stream graphs will offer interactivity to allow you to explore and isolate individual layers.

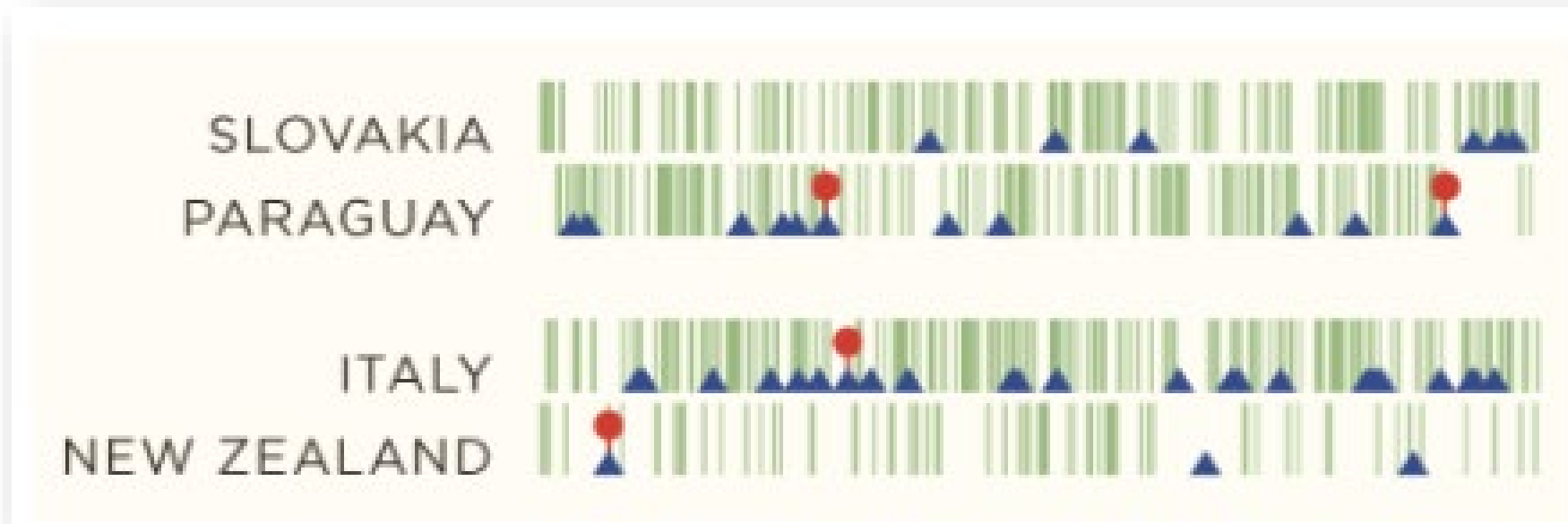
Candlestick Chart

Description: The candlestick chart is commonly used in financial contexts to **reveal the key statistics about a stock market for a given timeframe**. The height of the central bar indicates the change from the opening to closing price and the colour tells us if this is an increase or decrease. They are similar in concept to the “**box and whiskers plot**”, which focus on the statistical distribution of a set of values (showing upper and lower quartiles as well as the median).



Barcode Chart

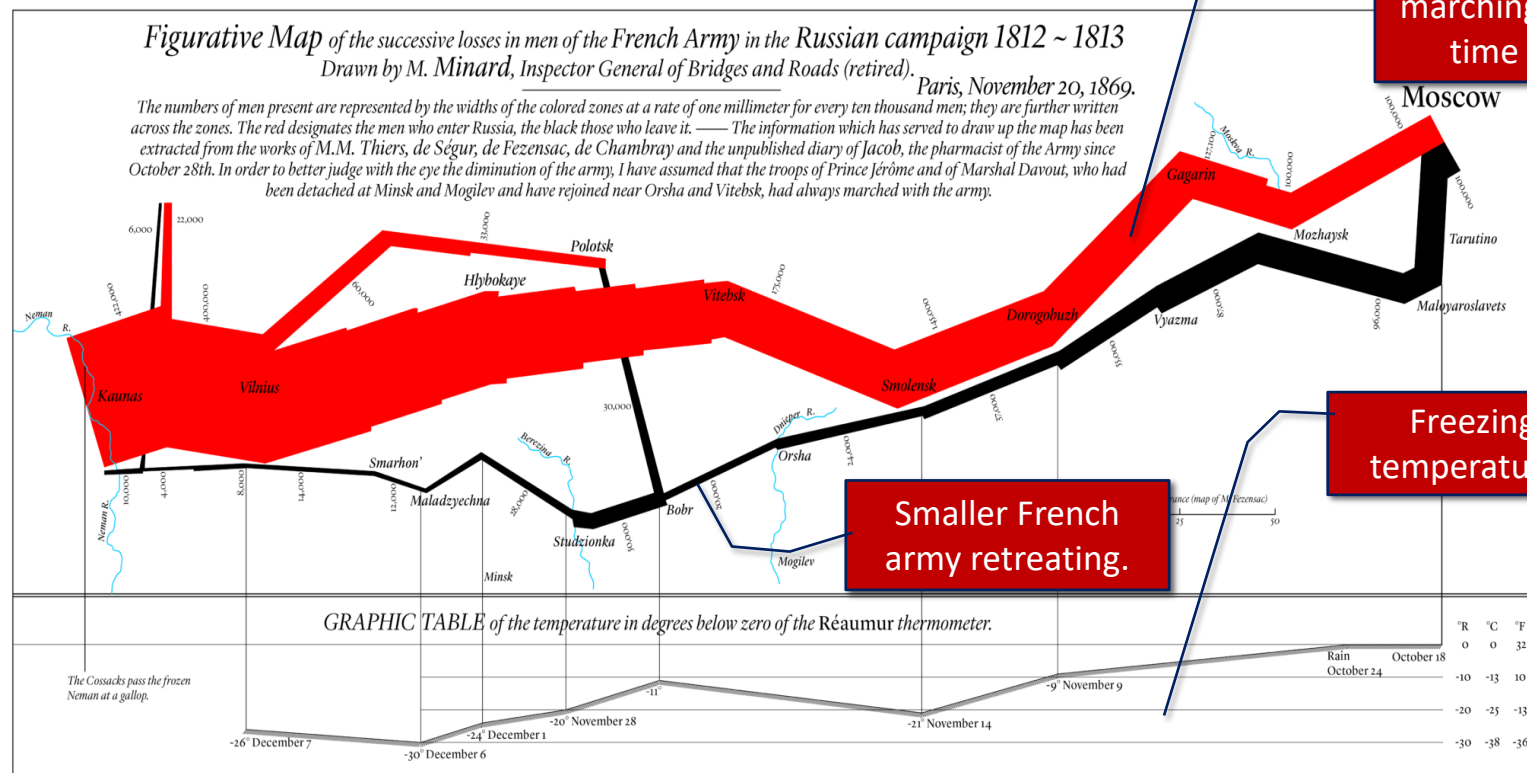
Description: These are very compact displays that depict a sequence of events or milestones over the course of time using a combination of symbols and colour. Demonstrating similar qualities to those of a sparkline, barcode charts convey a significant amount of data packed into a small space.



Key events during two football matches.

Flow Map

Description: Similar in many ways to the Sankey diagram, a flow map portrays the flow of a quantitative value as it is transformed over time and/or space.



Large French army marching towards Moscow time and geography.

Captures 6 multivariate complexity: size of army, location, direction, distance travelled, temperature, and time.

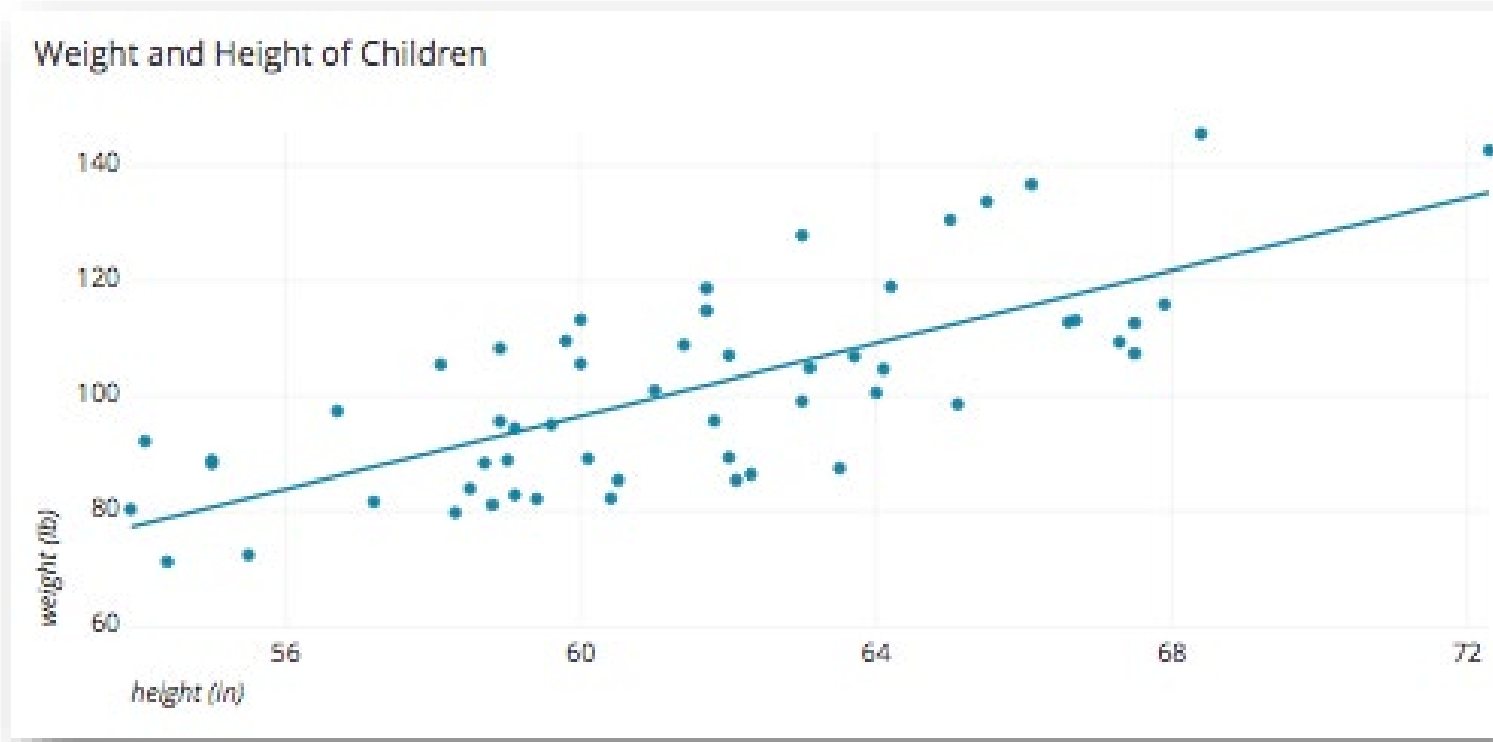
Freezing temperatures

Smaller French army retreating.

PLOTTING CONNECTIONS AND RELATIONSHIPS

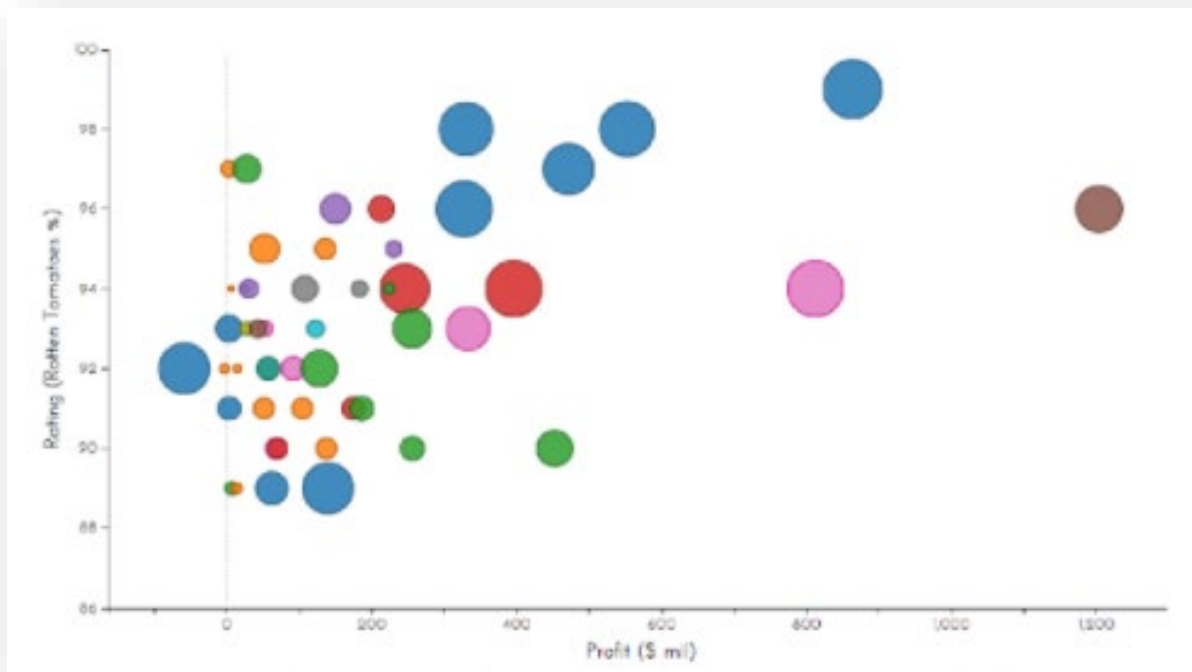
Scatter Plot

Description: A scatter plot is a combination of two quantitative variables plotted on to the x and y axes in order to reveal patterns of correlations, clustering, and outliers. This is a very important chart type, in particular, for when we are familiarizing with and exploring a dataset.



Bubble Plot

Description: A bubble plot extends the potential of a scatter plot through multiple encoding of the data mark. In the following example, the marks become circles of varying size and then coloured according to their categorical relationship. Often, you will see a further layer of time-based data applied to convey motion with the plot animated over time.

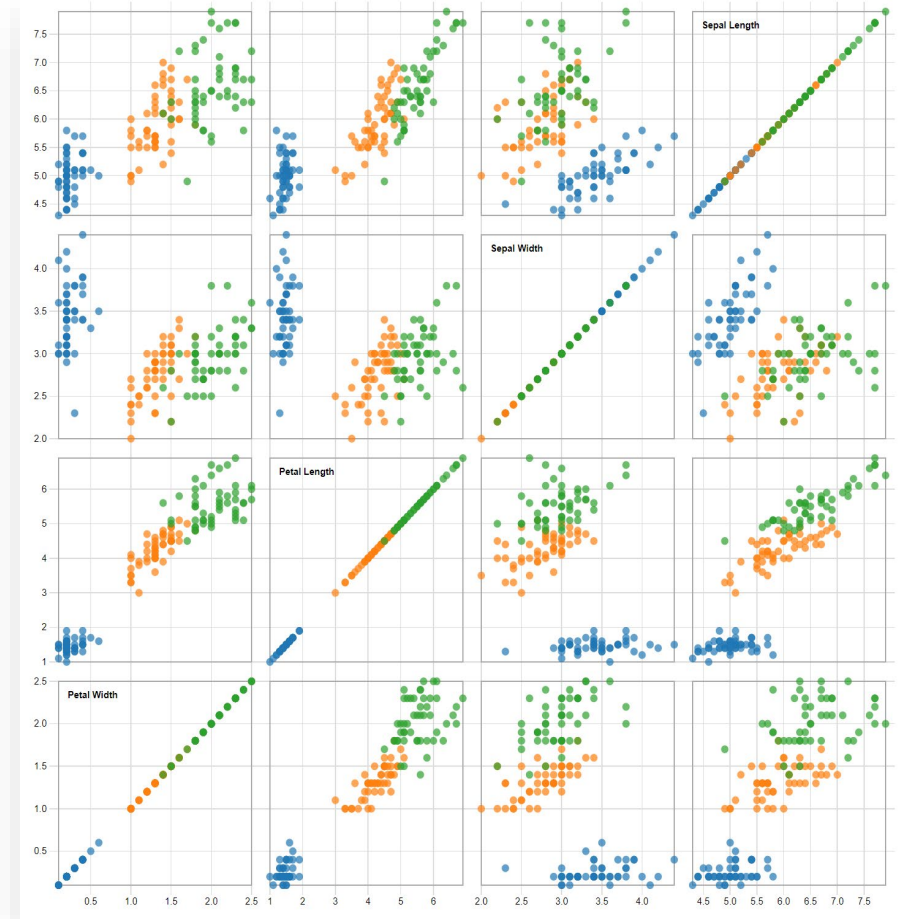


How much money
do the movies we
love make?

Scatter Plot Matrix

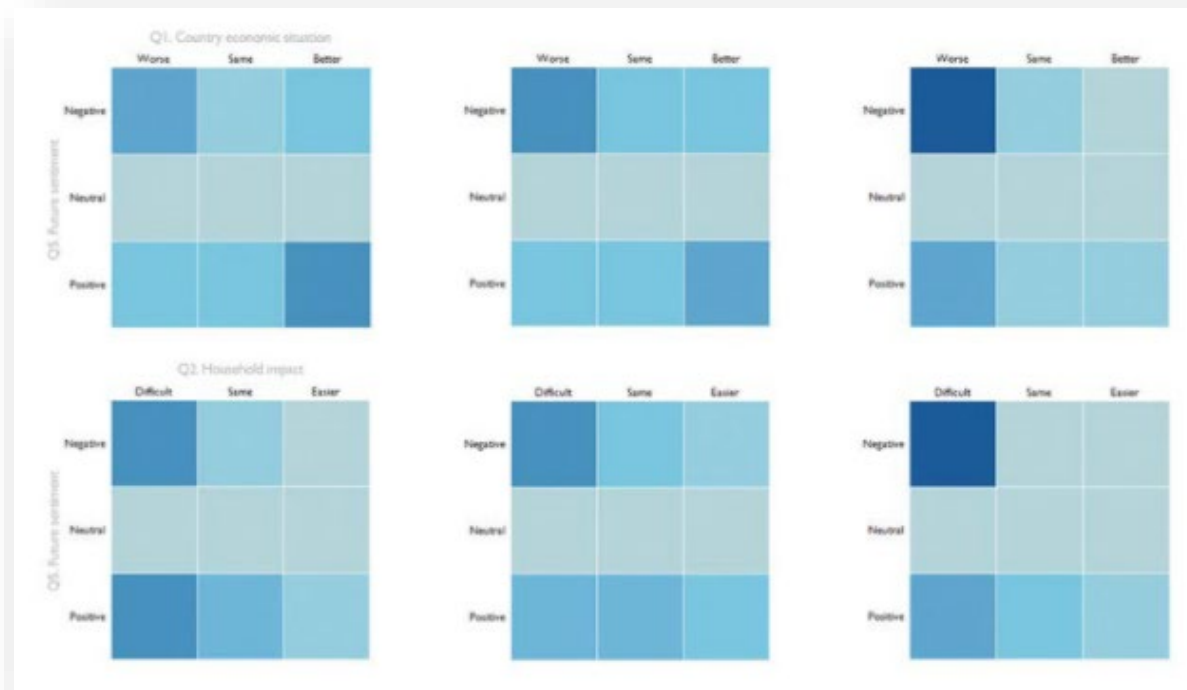
Description: Similar to the small multiples chart that we saw earlier, a scatter plot matrix takes advantage of the eye's rapid capability to spot patterns across multiple views of the same type of chart. In the following case, we have a panel of multiple combined scatter plots.

A panel of
multiple combined
scatter plots.



Heat Map

Description: With further similarities to small multiples, heatmaps enable us to perform rapid pattern matching to detect the order and hierarchy of different quantitative values across a matrix of categorical combinations. The use of a colour scheme with decreasing saturation or increasing lightness helps create the sense of data magnitude ranking.

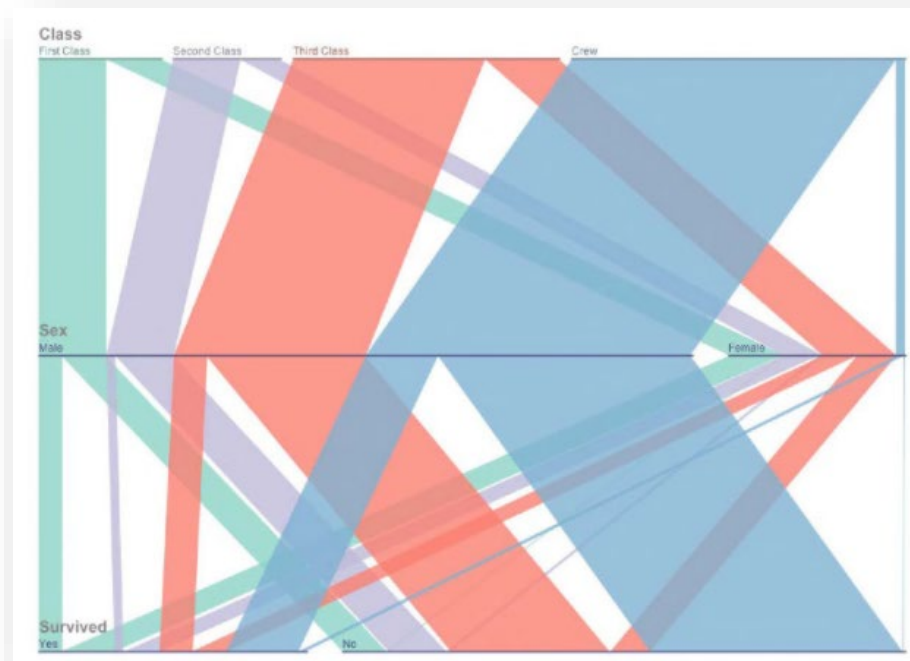


Future Sentiment vs
Country Economic
Situation.

Future Sentiment vs
Household Impact.

Parallel Sets (aka Parallel Coordinates)

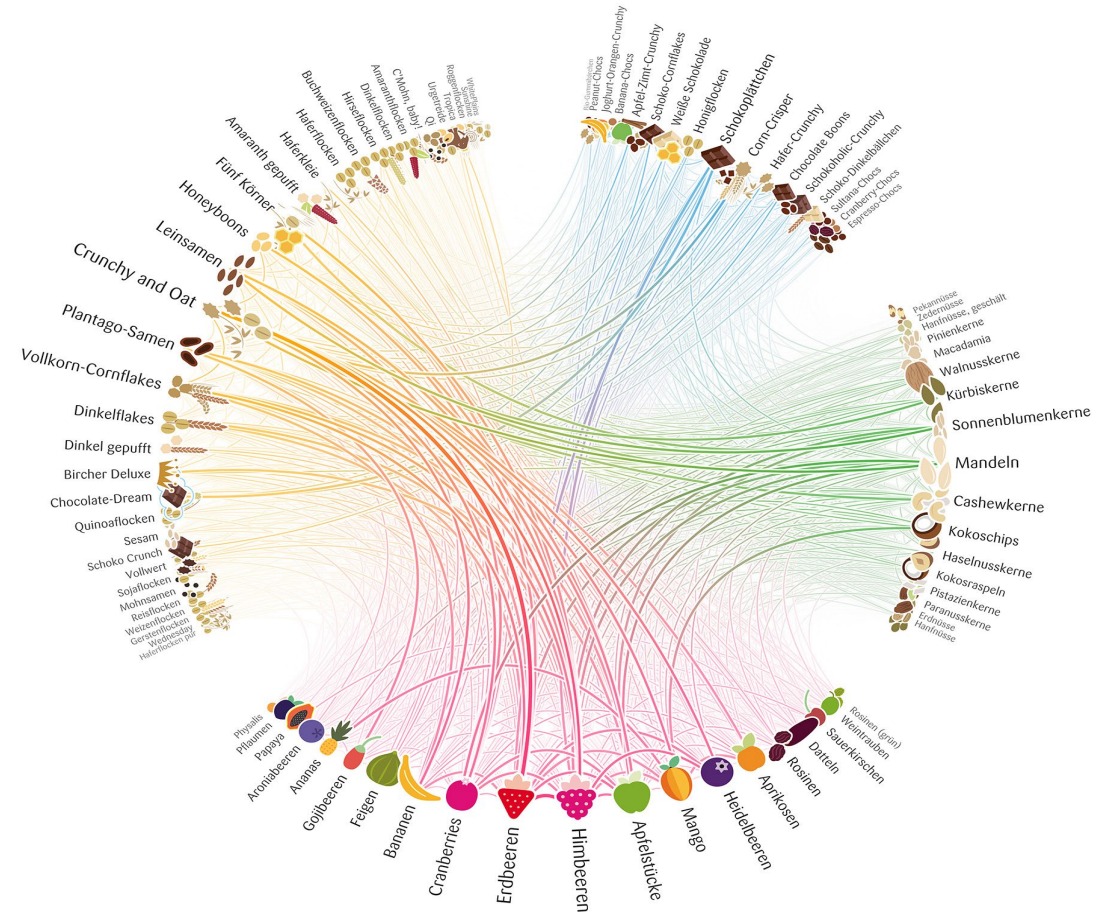
Description: Parallel sets offer a unique way of **visually exploring and analysing datasets**. The technique involves plotting all your data on to a series of axes, one for each of the variables you are interested in examining. This creates pathways that show the connections between the breakdown of values contained within your data for each variable. They are useful for learning about the potential correlations and consistencies that exist in our datasets. **Certain similarities with the function of Sankey diagrams.**



Class of Honours (1st, 2nd, 3rd) vs Sex (male, female) vs Survived (yes, no).

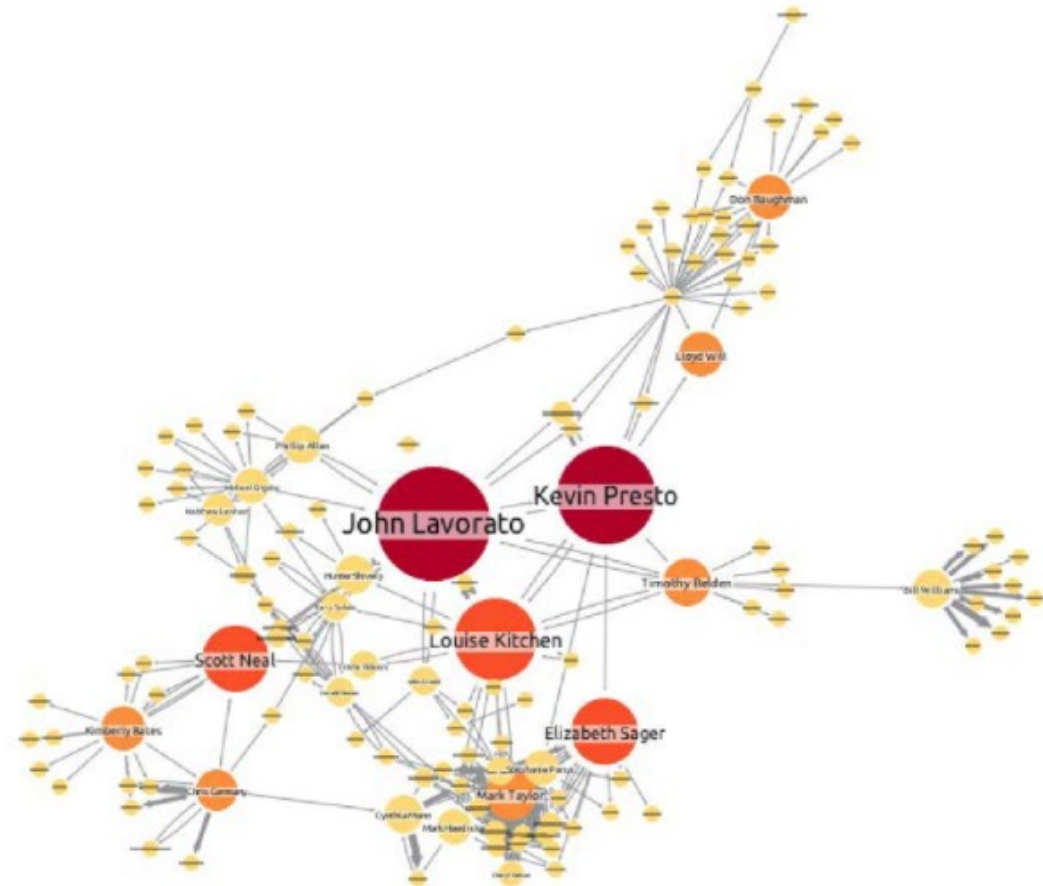
Radial Network (aka Chord Diagram)

Description: A radial network or chord diagram creates a framework for comparing complex relationships between categorical values. The use of a radial layout offers the opportunity to move beyond the restrictions of an x and y axis pairing. The key explanatory property is the connections that exist between components, sometimes sized (thickness) and coloured to incorporate extra layers of detail.



Network Diagram (aka Force-directed Network, Node-link Network)

Description: Network diagrams can look quite daunting through their visual complexity and apparent clutter (indeed, often they are described as “hairballs”). Their intention and value is **to facilitate exploration of complex data frameworks based on the existence or quantifiable strength of relationships, connections, and logical organization.** The typical purpose of these graphs is to enable the viewer to get a sense of patterns – picking out the elements that are of interest, observing clusters and gaps, dominant nodes and sparse connections.

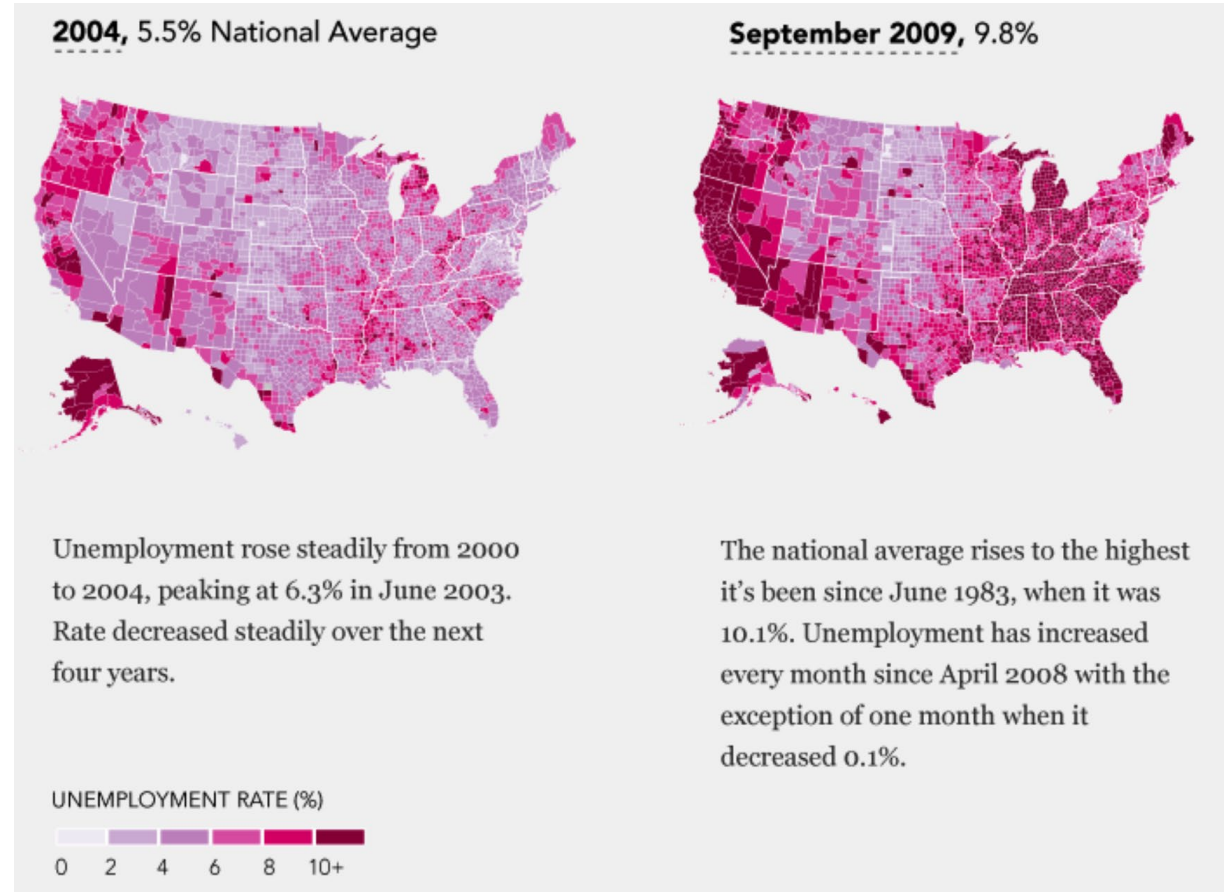


MAPPING GEO-SPATIAL DATA

Choropleth Map

Description: Choropleth maps colour the constituent geographic units (such as states or counties) based on quantitative values using a sequential or diverging scheme of saturation/lightness. While these are popular techniques, there is a recognized shortcoming caused by the fact that populations are not uniformly distributed. There is a potential distorting effect created by the prominence of larger geographic areas which may not be proportionately representative of the population of data.

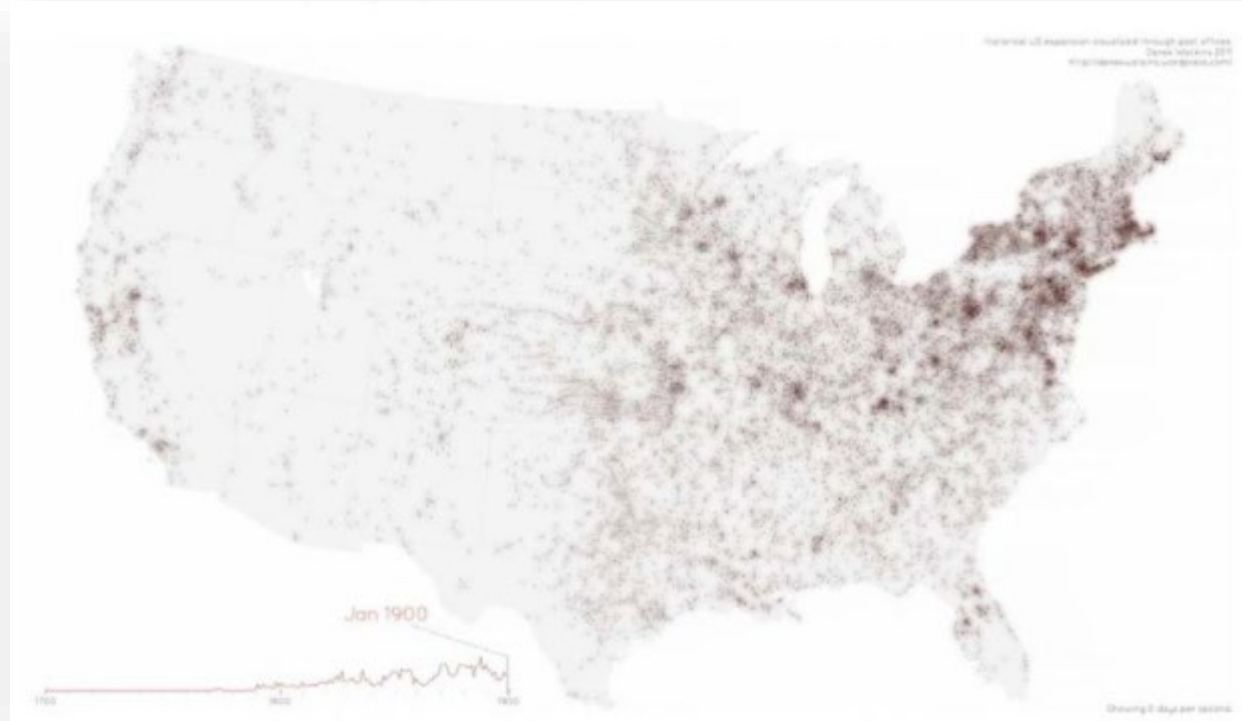
Need to choose colour classifications carefully to ensure accurate representation of chronological prominence of increasing quantities.



Dot Plot Map

Description: A dot plot map essentially displays a geographical scatter plot of records, combining the longitude and latitude to position marks on the map.

Animation: [Visualising US expansion through post offices](https://vimeo.com/27376376) (https://vimeo.com/27376376)

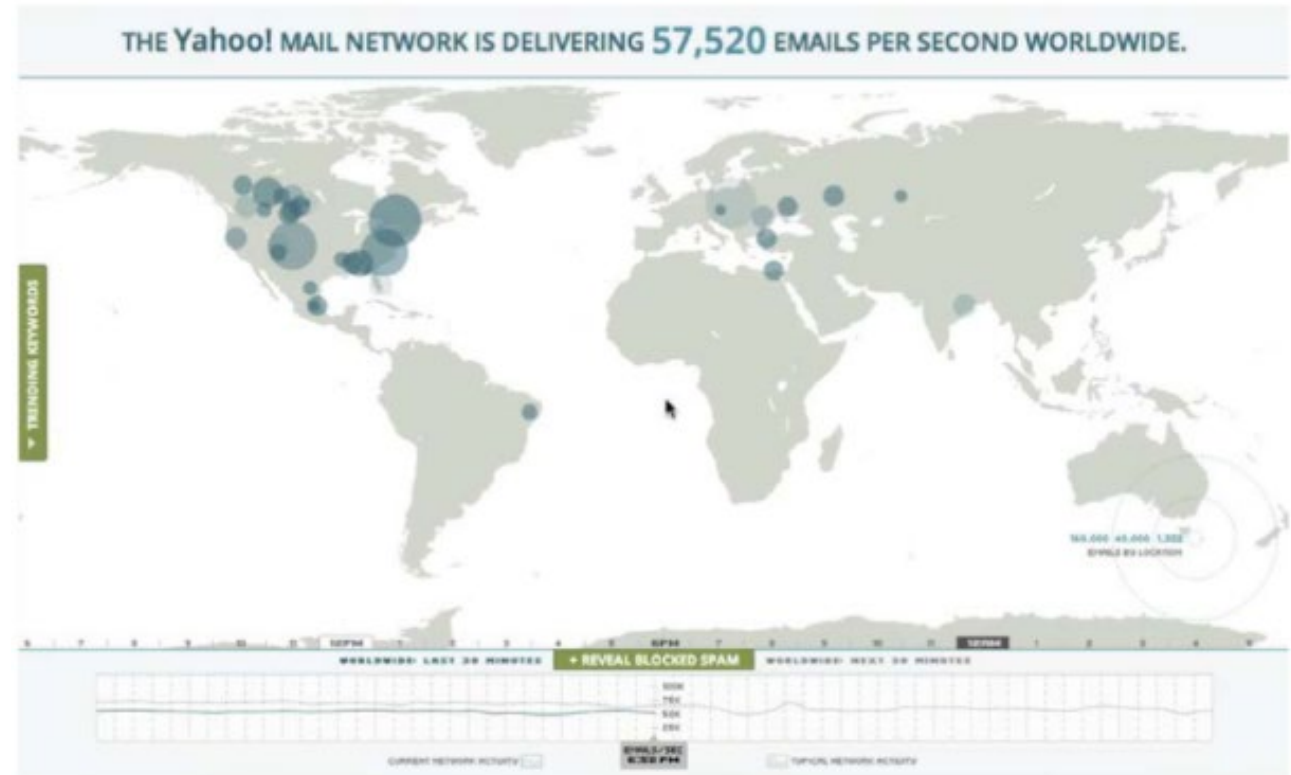


Visualizing US expansion
through post offices.

The data is gradually plotted
over time to reveal a story of
geographical spread.

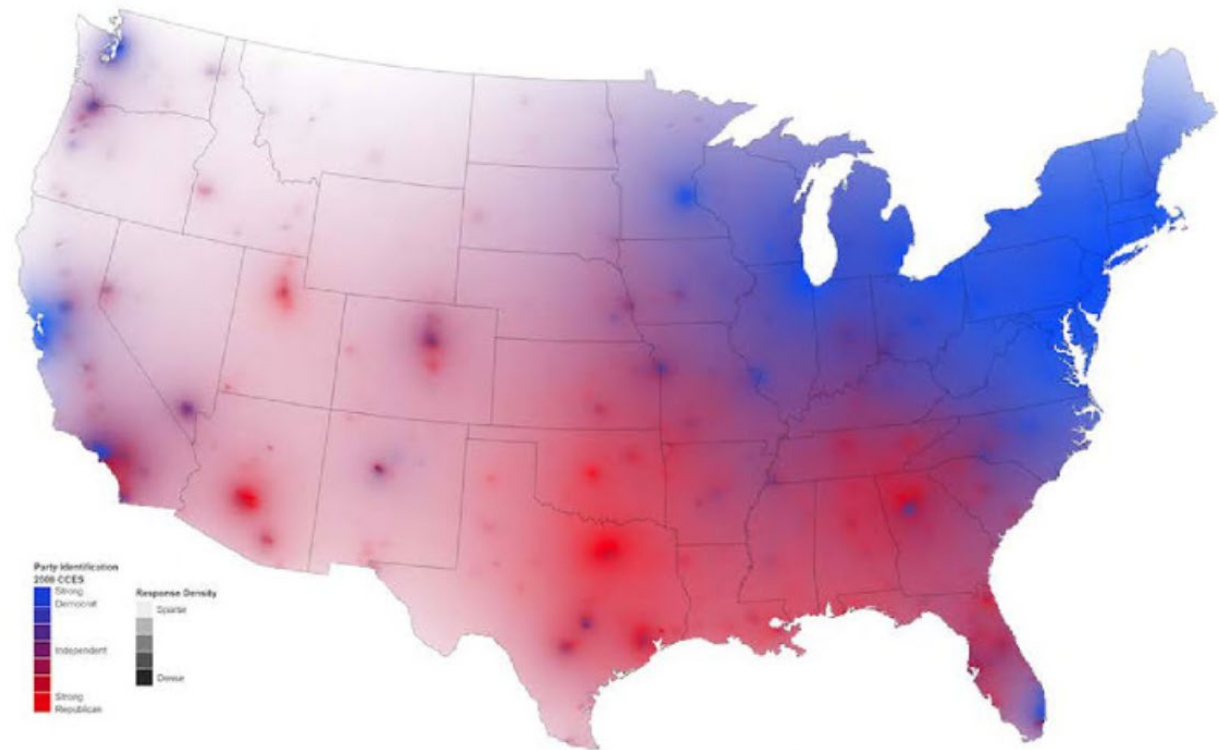
Bubble Plot Map

Description: This type of mapping plots differently-sized circular markers over given geographical coordinates to indicate the magnitude of a quantitative value. Whereas the dot plot maps were like geographical scatter plots, these are essentially bubble charts overlaid on to a map. The main contention with these designs tend to be that the spread of bubbles, depending on their size, can reach far beyond their geographical point and end up bleeding into other circles.



Isarithmic Map (aka Contour Map, Topological Map)

Description: This is a technique for overcoming the flaws associated with the **choropleth map** and involves combining **colour-hue** (to represent a political party), with **colour saturation** (to represent the dominance of party persuasion), with a final dimension of colour-darkness to represent the density of population. Algorithms are applied to help smooth the representation through the contour effect and this creates an elegant end result.

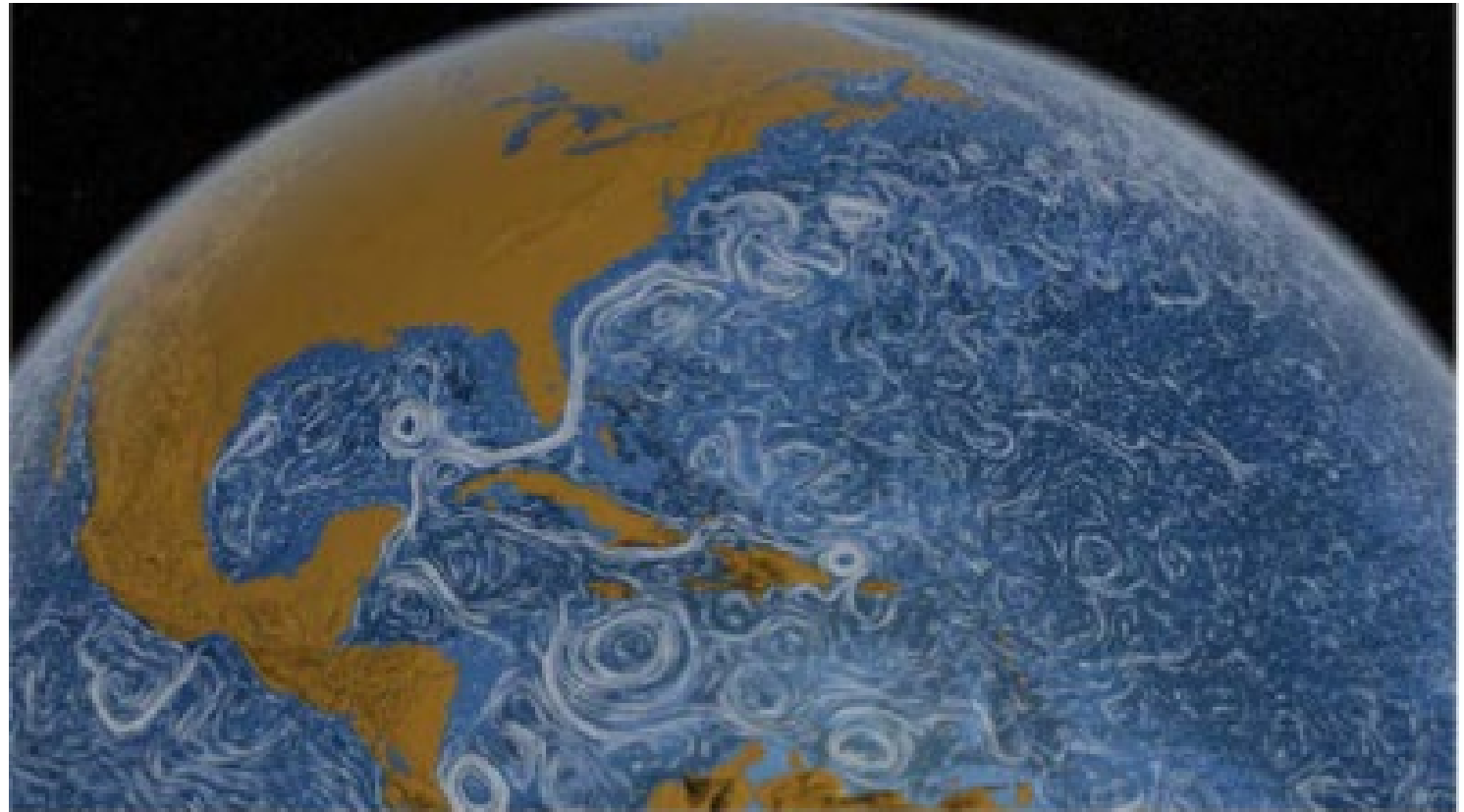


Particle Flow Map

Motion of currents that drive the world's oceans.

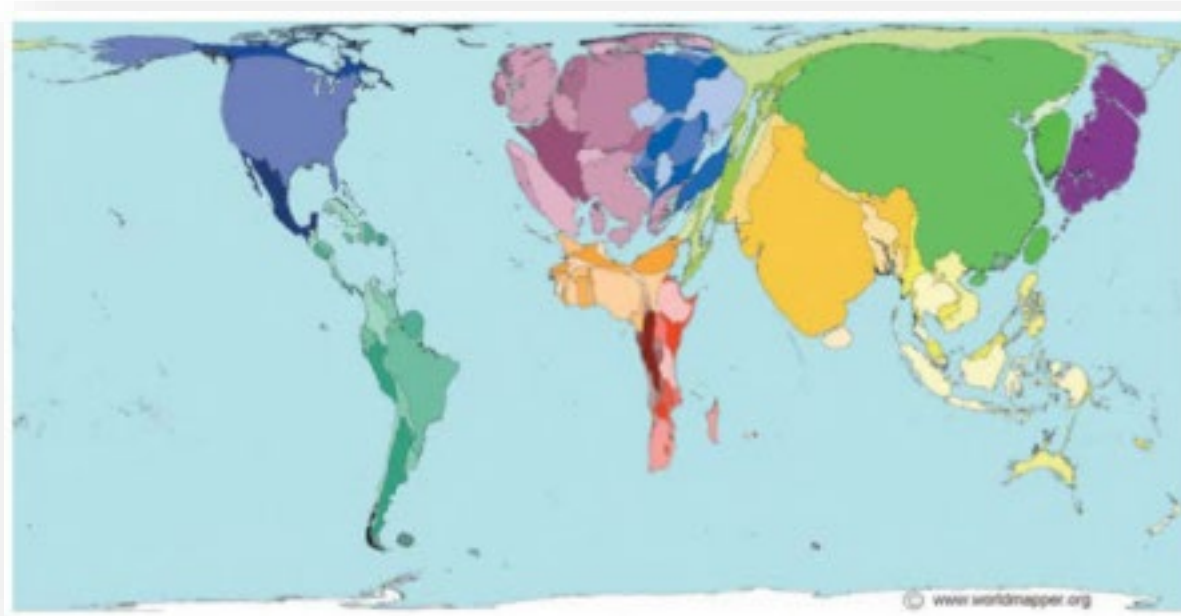
Combine multiple variables of location, size, speed, and direction to create a compelling design that perfectly captures the nature of the subject matter.

Description: A particle flow map uses animation to portray the motion of data across geography and over time. These careful and highly sophisticated constructions combine multiple variables of location, size, speed, and direction to create a compelling design that perfectly captures the nature of the subject matter.



Cartogram

Description: Where a choropleth map takes a location and gives it a shade of colour to represent a value, a cartogram takes a location and resizes the geographic shape to represent a value. The result is a distorted and skewed view of reality in the form of a reconfigured atlas. As with many of the chart types outlined here, the purpose is not to enable exact readings, rather to highlight the highly inflated, deflated, and unchanged shapes and sizes.



Network Connection Map

Description: Similar to the network diagram, the intention of a network connection map is **to facilitate the exploration of complex geographical connections**. A connection map joins up related locations to form a pattern that enables discovery of hubs, overlaps, clusters, and gaps – pretty much the same focus as that of the network diagram but this time with the platform of geographical coordinates.

World's flight paths
with those
involving Toronto
highlighted in
orange.



DATA VISUALISATION METHODS

DR DANNY POO

BIG DATA ANALYTICS AND VISUALISATION