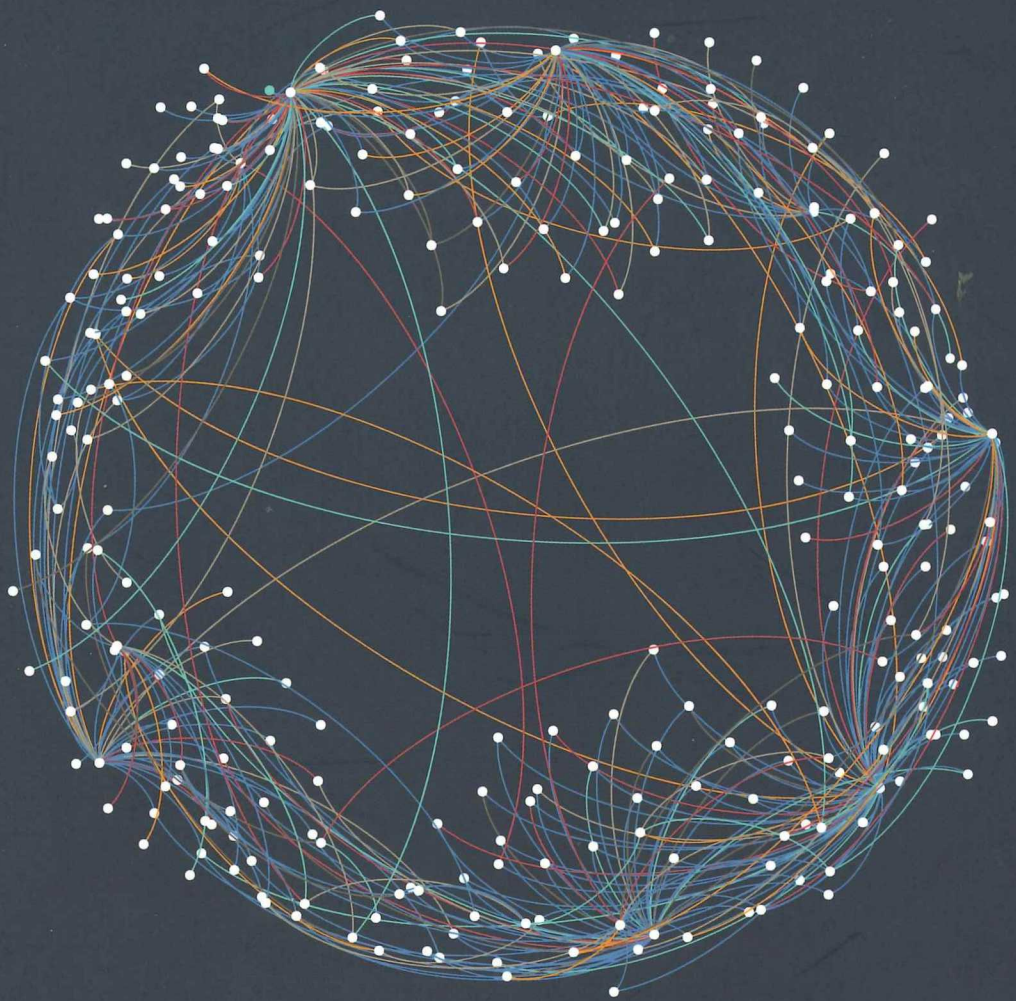
2nd Edition

Data Visualisation

A Handbook for Data Driven Design



Andy Kirk

6

Data Representation

In Chapters 3, 4 and 5 you have been working through activities that embody what I consider to be the hidden thinking of a visualisation project. These preparatory stages have helped you define the requirements and aims of your work, given you steps to become acquainted with your data, and, most recently, provided a structure for defining your editorial intent.

This chapter commences the fourth stage of the design process and represents a shift in focus towards design thinking. 'Developing your design solution' begins with arguably the most significant element of the visualisation design anatomy, namely data representation. How will you visually portray your data?

We start the discussion by looking at the fundamentals of visual encoding, exploring the building blocks that underpin all data representation thinking. From this bottom-up viewpoint we will switch to the more pragmatic perspective of selecting chart types. To close the chapter, you will learn about the influencing factors that will inform the choices you make.

6.1 Visual Encoding and Charts

Representing your data visually involves the act of visual encoding. As visualisers, we encode our data using two main visual properties, marks and attributes. *Marks* are visual placeholders representing data *items*, such as distinct records or discrete groupings, depending on the form of your tabulation. These are the four main types of marks, as shown in Figure 6.1.

Attributes are variations in the visual appearance of marks to represent the values associated with each data item. The main attributes you will encounter include those given in Figure 6.2.

The creative scope of some projects may use variation in attributes around the auditory (sound), haptic (touch), gustatory (taste) and olfactory (smell) senses, otherwise these visual attributes are the most commonly used options.

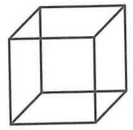
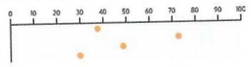
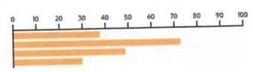
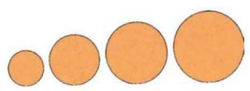
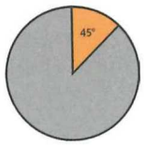
MARK	EXAMPLE	DESCRIPTION
Point		The *point* mark is commonly used as a marker to represent quantitative values through position on a scale, forming the basis of, for example, the scatter plot.
Line		The *line* mark is commonly used to represent quantitative values through variation in size (length), forming the basis of, for example, the bar chart.
Shape		The *shape* mark is commonly used to represent quantitative values through variation in size and position, forming the basis of, for example, the bubble plot.
Form		The *form* mark is used to represent quantitative values through variation in size (volume), forming the basis of charts that encode 3D representations.

Figure 6.1 A Classification of Different Types of 'Mark' Encodings

ATTRIBUTE	EXAMPLE	DESCRIPTION
Position		Variation in position along a scale is used to indicate a quantitative value, often using a point mark.
Size (Length)		Variation in size (length) is used to represent quantitative values based on proportional scales where the larger sizes mean larger quantities. The line mark has a single 'linear' spatial dimension, i.e. it shows quantities through either height or width but not both.
Size (Area)		Variation in size (area) is used to represent quantitative values based on proportional scales where the larger sizes mean larger quantities. The shape mark has two ('quadratic') spatial dimensions i.e. it shows quantities through a combination of both height and width.
Size (Volume)		Variation in size (volume) is used to represent quantitative values based on proportional scales where the larger sizes mean larger quantities. The form mark has three ('cubic') spatial dimensions i.e. it shows quantities through a combination of height, width, and depth.
Angle		Variation in the size of an angle is used to represent quantitative values where larger angles mean larger quantities or, more specifically, larger parts of a whole.

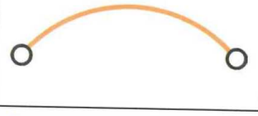
ATTRIBUTE	EXAMPLE	DESCRIPTION
Quantity		Variation in the quantity of a set of point marks (such as symbols) can be used to represent a single or aggregated quantitative value.
Colour: Hue		Variation in colour hue is typically used for distinguishing categorical data values.
Colour: Saturation		Variation in colour saturation can be used, often in conjunction with other colour properties, to represent ordinal scales; typically, the greater the saturation, the greater the hierarchical emphasis.
Colour: Lightness		Variation in the lightness of colour can be used to represent quantitative scales; typically, the darker the colour, the higher the quantity.
Pattern		Variation in pattern (sometimes also described as pattern *texture* or *density*) can be used to represent ordinal scales or distinguish categorical values, perhaps indicating degrees of certainty.
Symbol		Variation in symbols are commonly used for distinguishing categorical data values. The scope of this attribute could extend to images and illustrations explicitly representative of data values.
Connection		Connection (also known as *edge*) indicates a relationship between two nodes established by a connecting line. The shape and size of the connection is usually meaningless but sometimes arrows or variation in line thickness may be used to encode some notion of direction in the relationship.
Containment		Containment (also known as *enclosure*) is a way of encoding a hierarchical relationship between categories that belong to a related 'parent' category grouping.

Figure 6.2 A Classification of Different Types of 'Attribute' Encodings

It is worth noting that sometimes you do not need to encode data. Displaying values in their original numeric or textual form may suffice, perhaps as presented in a table or through callout statistic headlines.

Understanding visual encoding is of fundamental importance and is of particular relevance when representing data using tools that adopt a bottom-up approach. However, for most people's needs, it can often be more pragmatic to think about data representation techniques through selecting chart types.

CATEGORICAL | Comparing categories and distributions of quantitative values

HIERARCHICAL | Revealing part-to-whole relationships and hierarchies

RELATIONAL | Exploring correlations and connections

TEMPORAL | Plotting trends and intervals over time

SPATIAL | Mapping spatial patterns through overlays and distortions

Figure 6.3 The 'CHRTS' Families of Chart Types

If marks and attributes are the ingredients, chart types are the recipes. Different charts offer different established ways of representing data, each one comprising combinations of marks and attributes. As the field has matured, and practitioners have developed new recipes of marks and attributes, the range of established chart-type options has grown.

To acquaint you with a broader repertoire of charting options, over the coming pages I will present a collection of some of the common and useful chart types being used across the field today. This gallery aims to provide you with a valuable reference that will help you to decide how best to show what it is you want to say. I have organised each chart into five main families (Figure 6.3) based on the primary editorial relationship you are trying to understand. The five-letter mnemonic CHRTS provides a useful taxonomy for organising your thinking about which chart(s) to use for your data representation needs.

Each chart-type profile is presented with supporting details that will help you fully understand the role and characteristics of each option, including:

- The primary name used to label each chart type as well as some further alternative names that are often used.
- An indication of which CHRTS family each chart belongs to, based on their specific primary role, as well as a sub-family definition for further classification.
- A description of the chart's representation method, detailing what it shows and what each mark and attribute encoding it deploys.
- An applied example of the chart type in use with a description of what it specifically shows.
- Presentation tips about the potential interactivity, annotation, colour or composition design choices you might consider.
- 'Variations and alternatives' that describe further derivatives to understand other uses and different purposes.

This gallery of charts is by no means an exhaustive list and I have excluded some options because they were not different enough from other charts that have been profiled. I have mentioned some charts that are legitimate derivatives or alternative applications of other similar charts, but have assigned a whole page to profile these separately. For example, the *Voronoi treemap* is really just a variation on the treemap that is profiled. It uses a different algorithm to arrange its constituent pieces within different spatial layouts, like circles. The appearance and method of making this might be slightly different, its usage is not.

I have wrestled with the value of including some of the charts presented, often due to limitations and shortcomings in aspects of their usage. Some charts have merit for specific contexts, but can be quite narrow in scope. Therefore, by including certain partially flawed charts I am attempting to signpost relevant shortcomings, so you know how to use them sparingly. A word cloud, for example, is a chart with absolutely quite limited value, but nonetheless it does have a role, as does the often-derided pie chart. All chart types offer value for different situations; you just need to use discretion to select them only under specific circumstances.

Although I have excluded several charts on grounds of demonstrating only a slight variation on profiled charts, there are some types included that do exhibit small derivations from other charts (such as the bar chart and the clustered bar, or the scatter plot and the bubble plot). In these cases I felt there was sufficient difference in their practical application, and they were in common usage, to merit their separate inclusion, despite the similarities.

Another point to make is that certain charts do not just fit into a single family. All charts that belong to the hierarchical, relational, temporal and spatial families can include features of categorical breakdown. Using a line chart to show how quantitative values have changed over time for different categories could warrant being classified in either the temporal or the categorical families. However, the change over time dimension is the primary dimension of analysis and enables comparison between categories as a secondary perspective, so it is assigned to the temporal family. I have therefore concentrated the taxonomy around the angle of analysis each chart primarily conveys.

Finally, the spatial family of charts often relates to thematic maps that would not normally be considered charts in purist terms. For convenience, though, I am badging them all as charts. It is worth noting too that not all spatial analysis is geographic. Any of the spatial methods presented could be used for non-geographic contexts, such as the anatomy of the body, the layout of a building, the seat plan of an airliner.

BAR CHART

ALSO KNOWN AS Column chart, histogram (wrongly), lollipop chart

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A bar chart displays quantitative values for different category items. The chart comprises line marks (bars) with the size attribute (length or height) used to represent the quantitative value for each item.

EXAMPLE Comparing the number of unique land neighbours for countries with at least seven.

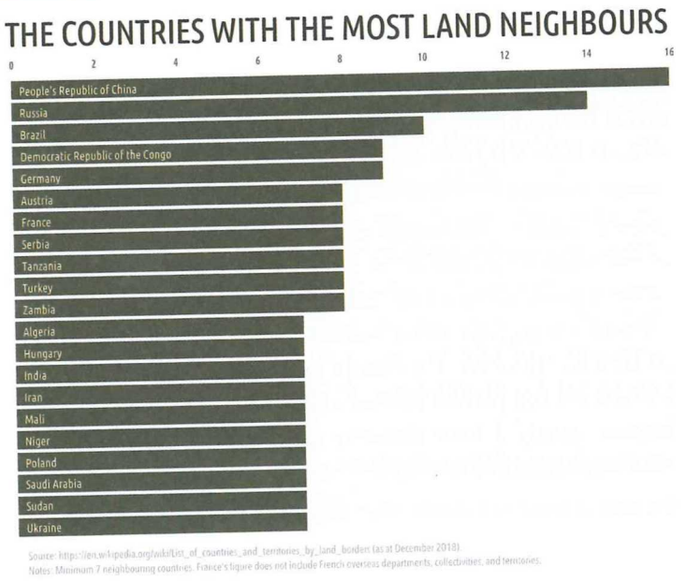


Figure 6.4 The Countries with the Most Land Neighbours

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label directly each bar value, as this will lead to label overload.

COMPOSITION: The bars should be proportionally sized according to the associated quantitative value – nothing more, nothing less – otherwise the perception of the bar sizes will be distorted. Most commonly, this means setting the quantitative value scale to an origin of zero. There is no significant difference in perception between vertically or horizontally arranged bar charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each bar. Including a small gap between each bar will help to preserve a clear distinction between each category item. Aim to make the sorting of values in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

A variation in the application of a bar chart would be to show quantitative values over time. This would be an option to consider over the line chart when you have quantities for discrete periods (such as totals over a monthly period) rather than a purely continuous series of point-in-time measurements. 'Spark bars' are mini bar charts that aim to occupy only a word's length amount of space. They are often seen in dashboards where space is at a premium and there is a desire to optimise the density of the display. If you want to include further categorical subdivisions, an alternative might be the 'clustered bar chart', to compare two or more or adjacent values, or the 'stacked bar chart', if there is a part-to-whole relationship. 'Dot plots' offer a useful alternative for situations where you have large quantitative values with a narrow range of difference and this difference is important to make visible. For contexts where you have diverse value sizes and many categorical items, the 'proportional symbol chart' is an option to consider.

CLUSTERED BAR CHART

ALSO KNOWN AS Clustered column chart, paired bar chart, grouped bar chart

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A clustered bar chart displays quantitative values for different primary category items with a secondary categorical breakdown enabling local comparisons. The chart comprises line marks (bars) with the size attribute (length or height) used to represent the quantitative value for each item. An attribute of colour is also used to distinguish further the secondary categorical groupings.

EXAMPLE Comparing the number of Oscar nominations with the number of Oscar awards won for the 10 actors who have received the most nominations for acting.

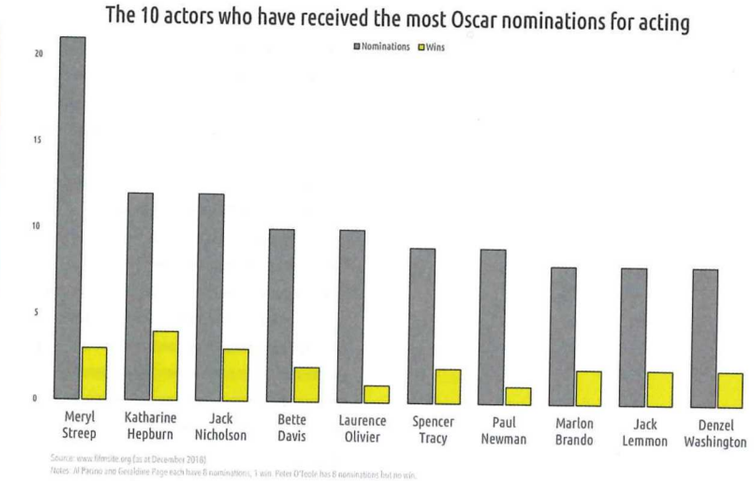


Figure 6.5 The Ten Actors Who Have Received the Most Oscar Nominations for Acting

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label directly each bar value, as this will lead to label overload. Any colours used must be explained through the inclusion of a legend.

COMPOSITION: The bars should be proportionally sized according to the associated quantitative value – nothing more, nothing less – otherwise the perception of the bar sizes will be distorted. Most commonly, this means setting the quantitative value scale to an origin of zero. There is no significant difference in perception between vertically or horizontally arranged clustered bar charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each cluster of bars will help to preserve a clear distinction between each primary category item. Sometimes one bar might be slightly hidden behind the other if the display concerns a before and after relationship. Aim to make the sorting of values in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

Like the bar chart, clustered bar charts can also be used to show how values have changed over time. Alternatives would include the 'connected dot plot', particularly to compare the quantitative size of two categories across a number of major category items. If your clusters comprise many distinct categories, the display might become too busy. You therefore might consider creating separate bar charts for each category item or using a 'matrix chart' structure to show the quantitative values at the intersection of two categorical dimensions.

BULLET CHART

ALSO KNOWN AS {No other names}

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A bullet chart is effectively a bar chart displaying quantitative values for different categories, but incorporating additional bandings to assist with interpreting the bars. The chart comprises line marks (bars) with the size attribute (length or height) used to represent a quantitative value for each item. An attribute of colour (usually the lightness property) is commonly used to distinguish contextual bandings behind each bar to aid interpretation.

EXAMPLE Comparing the batting averages for the current top 20 ranked batsmen in international Test cricket.

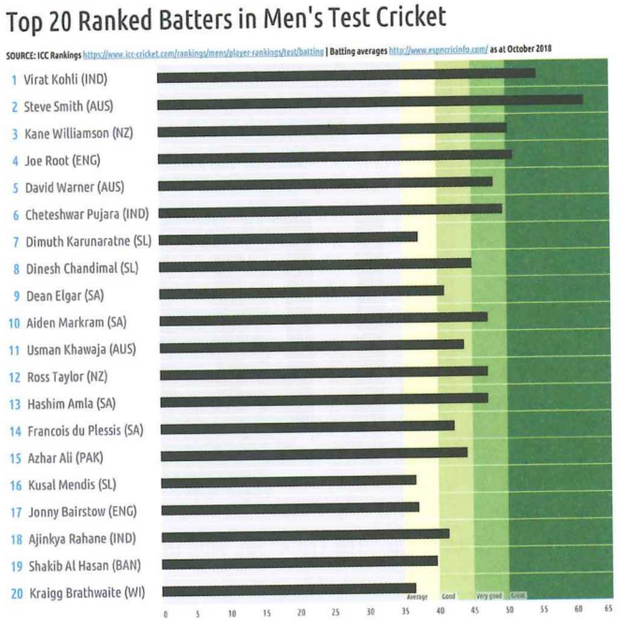


Figure 6.6 The Top 20 Ranked Batters in Men's Test Cricket (October 2018)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label each bar value directly, as this will lead to label overload. Any colours used to indicate meaningful bandings or markers should be explained through the inclusion of a legend.

COMPOSITION: The bars should be proportionally sized according to the associated quantitative value – nothing more, nothing less – otherwise the perception of the bar sizes will be distorted. Most commonly, this means setting the quantitative value scale to an origin of zero. There is no significant difference in perception between vertically or horizontally arranged bullet charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each bar. Aim to make the sorting of values in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

Like the bar chart, bullet charts can also be used to show how values have changed over time. Further point markers (usually small circles or thin lines) can be included in the bullet chart to offer further useful comparisons and to optimise the interpretation.

WATERFALL CHART

ALSO KNOWN AS Cascade chart

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A waterfall chart provides details of how a total or net quantitative value has been formed through an ordered sequence of bars representing quantitative values for discrete categorical components. It is essentially a visual calculation showing different components of positive and negative values, represented by size and direction, to establish a running total. A common application of a waterfall chart would be to break down the calculation of profit as formed by different categories of income and expenditure.

EXAMPLE Comparing the number of arriving and departing migrants from different regions that form the net level of migration in New Zealand.

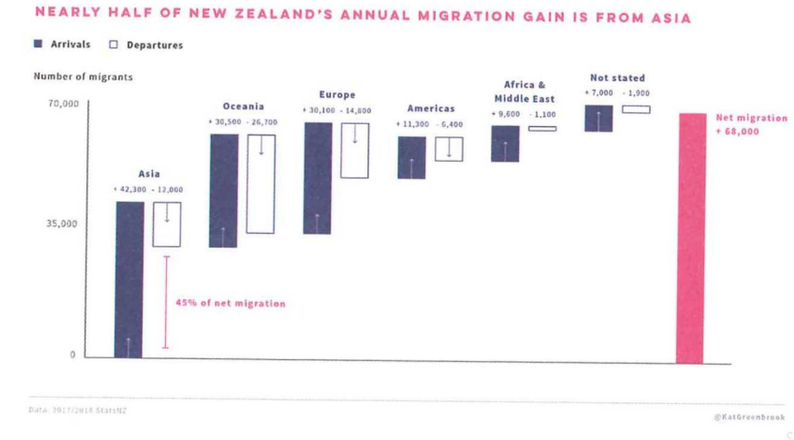


Figure 6.7 Nearly Half of New Zealand's Annual Migration Gain is From Asia, by Kat Greenbrook

PRESENTATION TIPS

ANNOTATION: Direct value labelling is usually applied to each step describing what each relates to as well as the quantitative amount. A dotted line is sometimes added to make more discernible what the running total is at each stage. Any colours used must be explained through the inclusion of a legend.

COLOUR: Attributes of colour are often established to classify visually each categorical stage or to distinguish further the positive and negative direction of quantitative values.

COMPOSITION: Most commonly a waterfall will be presented in landscape form with a left-to-right sequence arriving at a final total or net amount at the final right-side position.

VARIATIONS & ALTERNATIVES

One alternative would be to consider the stacked bar chart, as long as there were no negative quantitative values and all components are included that comprise a total, which represents a meaningful whole. The clustered bar chart may also be used to split the categorical parts of the final total in close proximity, but with all bars sized from a common baseline.

RADAR CHART

ALSO KNOWN AS Filled radar chart, star chart, spider diagram, parallel coordinates

COMPARISONS

REPRESENTATION DESCRIPTION

A radar chart plots values across multiple quantitative variables for one or several categorical items to enable general pattern forming. It uses a radial (circular) layout comprising several axes emerging from the centre-like spokes on a wheel, one for each variable. The quantitative values are then plotted along each scale using the attribute of position and then joined by connecting lines to form a unique geometric shape. Sometimes the lines or the shape fill is coloured for emphasis or for categorical differentiation when more than one item is plotted.

EXAMPLE

Comparing the global competitiveness scores across 12 'pillars' of performance for the UK versus Europe and North America.

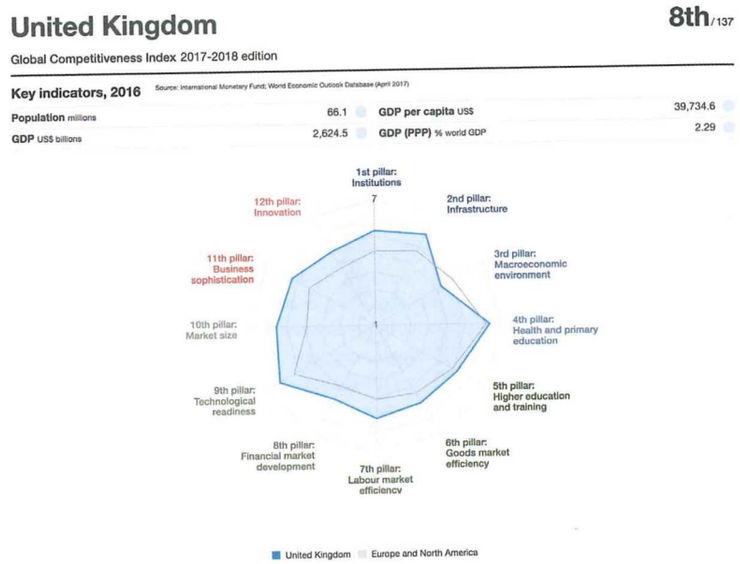


Figure 6.8 Global Competitiveness Report 2017–2018, by the World Economic Forum

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. Gridlines are only relevant if there are common scales across each quantitative variable. If so, the gridlines must be presented as straight lines, not concentric arcs, because the connecting lines joining up the values are themselves straight lines. If your quantitative values are on different scales, do not forget to display the values ranges on each. Any colours used must be explained through the inclusion of a legend.

COLOUR: When radar shapes are filled with a colour, sometimes a degree of transparency is applied to allow the chart apparatus to be still partially visible.

COMPOSITION: The cyclical ordering of the quantitative variables should be as meaningful as possible and consistent, as the shape formed will change for any ordering permutation. This will have a major impact on the readability and meaning of the resulting chart shape. A radar chart works best when the neighbouring pairings have some significant comparable value (such as values being plotted around the face of a clock or compass).

VARIATIONS & ALTERNATIVES

If you have common scales across the quantitative variables, a 'polar chart' is an alternative, should the radial layout be important to preserve. Otherwise, a 'bar chart' or 'dot plot' would be better options. While not strictly a variation, 'parallel coordinates' display a similar technique for plotting several independent quantitative measures in the same chart. The main difference is that parallel coordinates use a linear layout. If you have multiple category items, rather than plot them all on the same radar chart, consider using small multiples formed of distinct radars for each individual item instead.

POLAR CHART

ALSO KNOWN AS Coxcomb plot, polar area plot, circular barplot

COMPARISONS

REPRESENTATION DESCRIPTION

A radar chart plots values across multiple quantitative variables for one categorical item to enable general pattern forming. It uses a radial (circular) layout comprising several equal-angled circular sectors – like slices of a pizza, one for each variable. In contrast to the radar chart (which uses position along a scale), the polar chart uses variation in the size of the sector areas to represent values for each quantitative variable. It is, in essence, a radially arranged bar chart. Colour is an optional attribute, sometimes used to differentiate between different quantitative variables.

EXAMPLE

Comparing the capabilities of the four teams competing in Group F of the 2018 World Cup across seven distinct attributes.

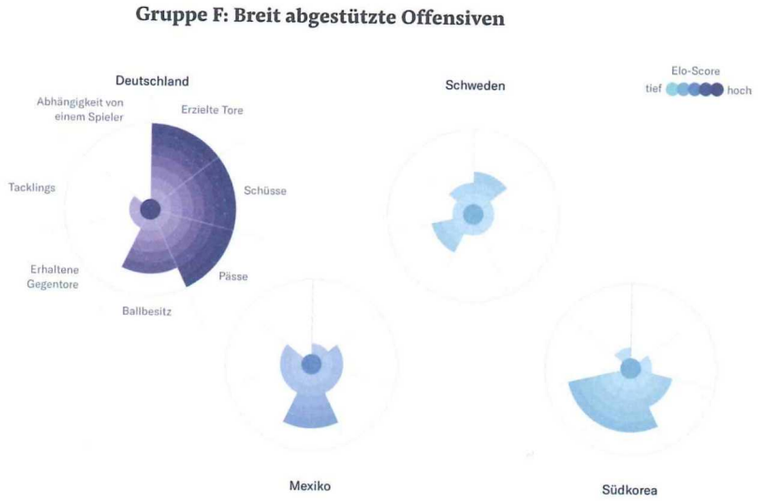


Figure 6.9 How Do National Teams Play? All 32 World Cup Participants in Direct Comparison [Translated], by NZZ Visuals

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. Gridlines are only relevant if there are common scales across each quantitative variable. If so, the gridlines must be presented as arcs reflecting the outer shape of each sector. Connecting lines joining up the values are themselves straight lines. Each sector typically uses the same quantitative scale for each quantitative measure but, on the occasions when this is not the case, do not forget to display the values ranges on each. Any colours used must be explained through the inclusion of a legend.

COMPOSITION: The cyclical ordering of the quantitative variables should be as meaningful as possible and consistent, as the shape formed will change for any ordering permutation. This will have a major impact on the readability and meaning of the resulting chart shape. A polar chart works best when the neighbouring pairings have some significant comparable value (such as values being plotted around the face of a clock or compass). The sizing of the sectors needs to be carefully calculated. Each sector should have a proportionally consistent angle of the whole and, to encode the quantitative values, the area of the sector, not the radius length, should be used.

VARIATIONS & ALTERNATIVES

If you have inconsistent scales across the quantitative variables, a 'radar chart' is an alternative should the radial layout be important to preserve. Otherwise, a 'bar chart' or 'dot plot' would be better options.

CONNECTED DOT PLOT

ALSO KNOWN AS Dot plot, dumbbell chart, range chart, dot chart, arrow chart

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A connected dot plot displays quantitative values for different primary category items with a secondary categorical breakdown enabling local comparisons. The plot is typically formed of two point marks plotting the quantitative value positions for each secondary categorical grouping. Joining the two points together is a connecting line which effectively represents the 'delta' (difference) between the two values through its size. Attributes of colour or variation in symbol are commonly used to distinguish the secondary categorical groupings.

EXAMPLE Comparing the typical salaries of women and men across a range of different job categories in the UK.

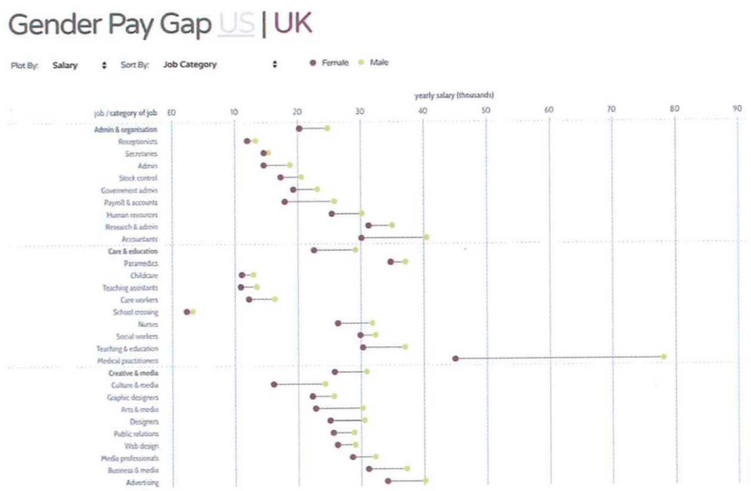


Figure 6.10 Gender Pay Gap UK, by David McCandless, Miriam Quick (Research) and Philippa Thomas (Design)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload. Any colours used must be explained through the inclusion of a legend.

COLOUR: Colour may be used to help emphasise the directional basis of the connecting lines.

COMPOSITION: As the representation of the quantitative values is encoded through position along a scale and not size, the quantitative axis does not need to have a zero origin. However, a zero origin may be helpful to establish the scale of the differences depending on the subject matter being portrayed. If you do not commence from an origin of zero, this will need to be clearly annotated. Aim to make the sorting of values in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

A variation in the application of the 'connected dot plot' would be to plot and compare values representative of two different points in time for the same measure. An alternative would be to use a variation of the 'Gantt chart', and rather than a single line starting from a minimum date and extending to a maximum date, you would just use this line to show the position and difference between quantitative values. An 'arrow chart' is an extension of this whereby the arrowhead is used to emphasise the directional basis of the line. Similarly, the 'carrot chart' uses line width tapering to indicate direction. If the number of secondary categories grows in number, the 'dot plot' would be useful to show the distribution of values rather than attempting to compare differences between just two values. A 'clustered bar chart' offers a further alternative for showing comparisons between secondary categorical dimensions.

PICTOGRAM

ALSO KNOWN AS Isotype chart, pictorial bar chart, symbol chart

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A pictogram displays quantitative values for different primary category items with the option for secondary categorical breakdown. The basis of the pictogram is the repetition in use of point marks, in the form of symbols or pictures, to represent an associated quantitative count. Each point mark may be representative of one or many quantitative units (e.g. a single symbol may represent 100 people). Secondary categorical dimensions can be incorporated through differentiation in the attribute of colour or symbol.

EXAMPLE Comparing the grouped earning levels of the top 100 highest paid athletes during 2017.

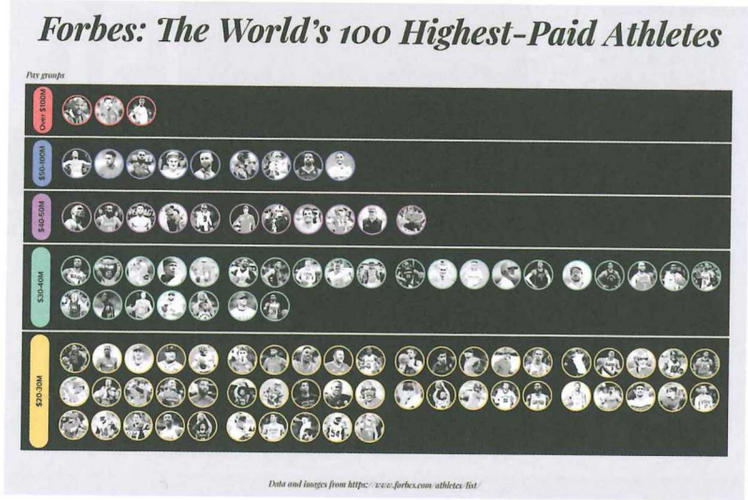


Figure 6.11 Forbes: The World's 100 Highest-paid Athletes, by Andy Kirk

PRESENTATION TIPS

ANNOTATION: The choice of symbols should be as recognisably intuitive as possible. If not, any legends should be presented close to the display to enable quick reference for determining the categorical and quantitative association of each symbol variation used.

COMPOSITION: If the quantities of markers exceed a single row, try to make the number of units per row logically 'countable', such as displaying in groups of 5, 10 or 100. To aid readability, make sure there is a sufficiently noticeable gap between clusters of grouped units. Aim to make the sorting of values in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

When showing a part-to-whole relationship, the 'waffle chart' is similarly formed using point marks and symbol or colour attributes to differentiate the constituent parts of a whole.

PROPORTIONAL SYMBOL CHART

ALSO KNOWN AS Proportional shape chart, graduated symbol plot, bubble chart, circle packing diagram

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A proportional symbol chart displays quantitative values for different category items. The chart comprises shape marks with the size attribute (area) used to represent the quantitative value for each item. An attribute of colour may be used to accentuate the quantitative scale or organise marks by the distinct categories. Estimating and comparing the size of areas with accuracy is not as easy, so this chart type works best when you have a diverse range of quantitative value sizes.

EXAMPLE Comparing the market capitalisation ($) of companies involved in the legal sale of marijuana across different industry sectors.

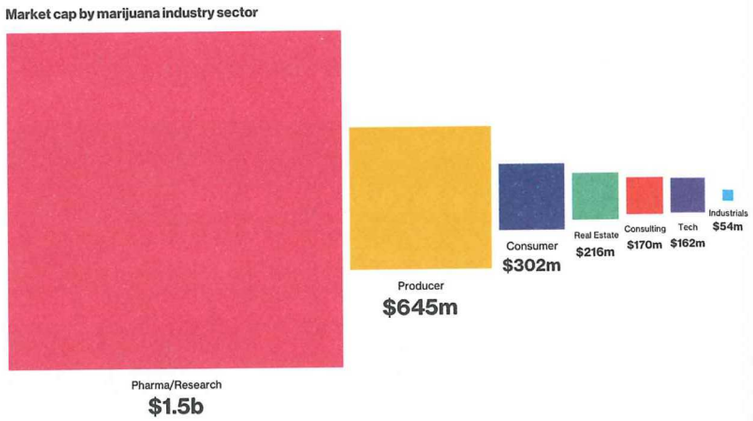


Figure 6.12 For These 55 Marijuana Companies, Every Day is 4/20, by Alex Tribou and Adam Pearce (Bloomberg Visual Data)

PRESENTATION TIPS

INTERACTIVITY: Proportional symbol charts may be accompanied by interactive features that let users select or mouseover individual shapes to reveal annotated values of the quantity and category.

ANNOTATION: If interactivity is not achievable, a quantitative size key should be included, or direct labelling incorporated. Though labelling can make a display cluttered (and be hard to fit when working with small-sized shapes) it will help overcome some of the limitations of judging area size. Any colours used must be explained through the inclusion of a legend.

COMPOSITION: The geometric accuracy of the shape mark size calculation is paramount: it is the area you are modifying, not the diameter/radius. Typically, the layout is quite free-form with no baseline or central gravity binding the display together. Otherwise, you might employ clustering or containers to help organise the categorical distinctions, though the colouring of each shape may already achieve this. Aim to make the sorting of the shapes in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

Often, the data shown represents many parts of a whole. A 'circle packing diagram' uses circular shapes and packs the contents into a neat circular layout representing a whole. The 'bubble plot' also uses differently sized shapes (usually circles) but the position is meaningful across two quantitative variable dimensions. By removing the size attribute (and effectively replacing the shape mark with a point mark) you could use the quantity of points clustered together for different categorical totals to create a variation of the 'pictogram'.

WORD CLOUD

ALSO KNOWN AS Tag cloud, proportional symbol chart

C H R T S

COMPARISONS

REPRESENTATION DESCRIPTION

A word cloud shows the frequency of individual word items within a passage of textual data. Each item is represented by words and then the font size of each is scaled according to the frequency of its usage. Words already have varied lengths, so it is important to remember that it is effectively the area of the word, not its length, that encodes its quantitative measure.

EXAMPLE Comparing the frequency of words used in Chapter 1 in the first edition of this book.

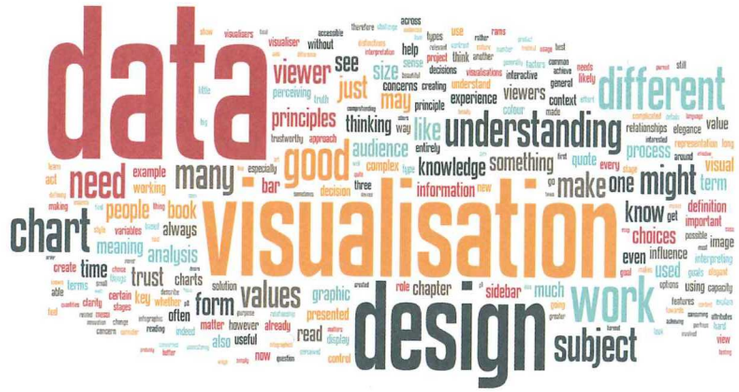


Figure 6.13 Comparing the Frequency of Words Used in Chapter 1 in the First Edition of this Book

PRESENTATION TIPS

INTERACTIVITY: Interactivity that lets users interrogate, filter and scrutinise the words in more depth, perhaps presenting examples of their usage in a passage, can be quite useful features to enhance the value of a word cloud.

ANNOTATION: Word clouds are most useful when you are trying to form a quick sense of some of the dominant keywords used in the text. Relative comparisons can be aided by including a key to explain how the font size scales equate to word frequency. Any colours used must be explained through the inclusion of a legend.

COMPOSITION: The arrangement of the words within a word cloud is typically based on a layout process that calculates the best placement of each word to occupy the optimum space.

VARIATIONS & ALTERNATIVES

Variations may include colours being used as a second form of quantitative encoding to accentuate the larger frequencies further or to organise useful groupings categorically. You might also consider using containers to separate out different clusters. Any alternative method from this categorical family of charts would more usefully display the counts of text, such as a bar chart or a proportional shape chart where the word label sits inside a sized shape mark.

HEAT MAP

ALSO KNOWN AS Matrix chart, mosaic plot, table chart, XY heatmap, 2D density plot

COMPARISONS

REPRESENTATION DESCRIPTION

A heat map displays quantitative values across the intersections of two categorical and/or discrete quantitative dimensions. The chart comprises two categorical axes with each distinct value presented across the row and column headers of a tabular layout. The corresponding cells effectively house a point mark with the attribute of colour (usually, colour lightness) used to represent the associated quantitative value.

EXAMPLE Comparing the average number of daily births across England and Wales between 1995 and 2014.

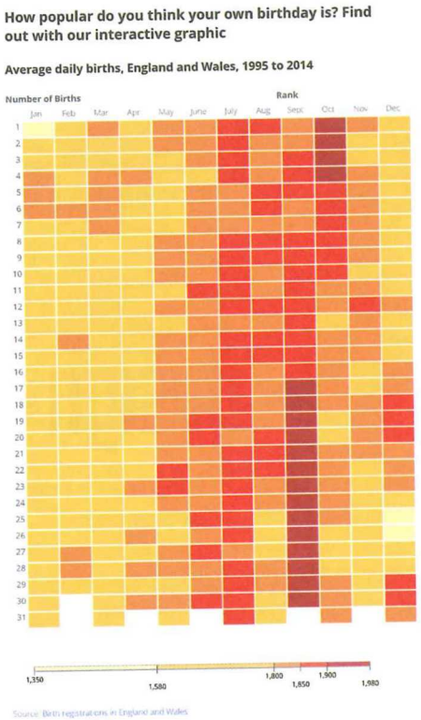


Figure 6.14 How Popular is Your Birthday?, by ONS Digital Content team

VARIATIONS & ALTERNATIVES

A 'radial heat map' offers a structure variation whereby the table may be portrayed using a circular layout. As with any radial display, this is really of value only if the cyclical ordering means something for the subject matter. A variation would see the colour lightness replaced by a categorical colouring approach if the values plotted were not quantitative in nature. An alternative chart approach would be the 'matrix chart' using the size of a shape or the frequency of clustered point marks to indicate a quantitative value.

PRESENTATION TIPS

ANNOTATION: Direct value labelling is possible but normally a clear legend to indicate colour associations will suffice. It is not easy for the eye to determine the exact quantitative values represented by the colours, even if there is a colour scale provided; heat maps mainly facilitate more a gist of the order of magnitude.

COLOUR: Decisions need to be made about whether to use a smooth colour gradient or employ discrete classifications for different value intervals. Different approaches will affect the patterns that emerge. There is no single right answer – you will arrive at it largely through trial and error/experimentation – but it is important to consider, especially when you have a diverse distribution of values.

COMPOSITION: Logical sorting and/or even grouping of the categorical values along each axis will aid readability and may help to surface key relationships.

MATRIX CHART

ALSO KNOWN AS Table chart, correlogram

COMPARISONS

REPRESENTATION DESCRIPTION

A matrix chart displays quantitative values across the intersections of two categorical and/or discrete quantitative dimensions. The chart comprises two categorical axes with each distinct value presented across the row and column headers of a tabular layout. The corresponding cells effectively house a geometric shape with scaled area size or clusters of point marks repeated in quantity to represent the associated quantitative value. Attributes of colour are often used visually to distinguish further categorical detail.

EXAMPLE Comparing the number of Nobel Laureates by award category and country of birth.

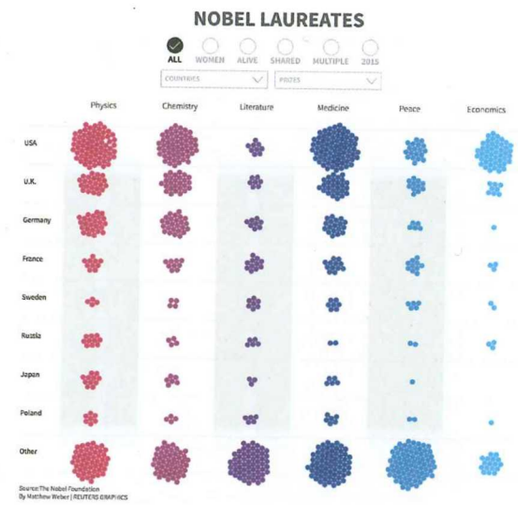


Figure 6.15 Nobel Laureates, by Matthew Weber (Reuters Graphics)

PRESENTATION TIPS

INTERACTIVITY: When using point mark clusters, interactive features can be useful to enable users to discover the labels of each item through tooltips.

ANNOTATION: When shape marks are used, direct value labelling is possible but normally a clear key to indicate the size associations will suffice. Any colours used must be explained through the inclusion of a legend.

COLOUR: Colours may not be necessary because the tabular layout already establishes separation across the two categorical dimensions. However, employing an additional attribute of colour can help to distinguish further the horizontal or vertical categorical values.

COMPOSITION: If there are diverse value sizes with some especially large outliers, it may be necessary for the size of the shape marks or the quantity of point clusters to outgrow the space of the relevant cell. This might help to emphasise editorially the outlier status. Controlling this may not be possible, in which case the largest quantitative value will usually fill no more than the maximum space available. Logical sorting (and maybe even sub-grouping) of the categorical values along each axis will aid readability and may help surface key relationships. The geometric accuracy of the shape mark size calculation is paramount: it is the area you are modifying, not the diameter/radius.

VARIATIONS & ALTERNATIVES

A variation may involve the intersecting cells being representative of categorical values (nominal or ordinal), and therefore you might substitute quantitative attributes of size or quantity with variation in symbols and/or colour attributes.

An alternative chart type might be the 'heat map', which similarly indicates quantitative values at the intersections of two categorical and/or discrete quantitative dimensions.

DOT PLOT

ALSO KNOWN AS Univariate scatter plot, 1D scatter plot, instance chart, strip plot, barcode chart

C H R T S

DISTRIBUTIONS

REPRESENTATION DESCRIPTION

A dot plot displays the distribution of quantitative values for data items, sometimes broken down by a categorical dimension, to show the range and shape of quantities. The plot is typically formed of point marks positioned along a quantitative scale. The point marks may be small circles or thin lines ('strips'). If categorical differentiation is necessary, attributes of colour or variation in symbol may be employed within a single plot, otherwise several separate plot views will be created for each discrete category grouping.

EXAMPLE Comparing the ranking distribution of the top 200 billionaires by industry.

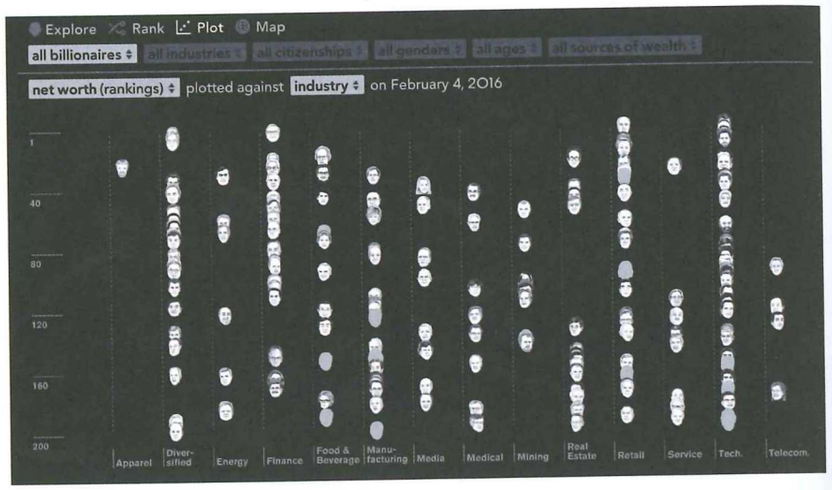


Figure 6.16 Bloomberg Billionaires, by Bloomberg Visual Data (Design and Development), Lina Chen and Anita Rundles (Illustration)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload. Direct labelling will normally be restricted to noteworthy points only. Any colours used must be explained through the inclusion of a legend.

COLOUR: Colour may be used to establish the focus of certain points and/or distinction between different sub-category groups to assist with interpretation. To overcome occlusion caused by plotting several marks at the same value position, you might use unfilled or semi-transparent filled circles to convey value frequency.

COMPOSITION: As the representation of the quantitative values is encoded through position along a scale, the quantitative axis does not need to have a zero origin, unless this is meaningful to the subject. If you do not commence from an origin of zero, this will need to be clearly annotated.

VARIATIONS & ALTERNATIVES

A variation in the encoding of the dot plot may see the point marks replaced by shape marks (usually circles) in order to represent a second quantitative measure through size variation. This might be a useful method to represent the frequency of observations when several items share a similar value. The variation in the role of the dot plot would be through the 'instance chart', which plots events over a temporal axis rather than a quantitative scale. An alternative chart type would be the 'beeswarm plot', especially when you have a non-uniform distribution of values that cluster around similar quantities. You could also use a 'scatter plot' with its second axis offering the scope to plot two data quantitative variables with the items spread across the associated coordinate positions.

BEESWARM PLOT

ALSO KNOWN AS Jitter plot

C H R T S

DISTRIBUTIONS

REPRESENTATION DESCRIPTION

A beeswarm plot displays the distribution of quantitative values for data items to show the range and shape of quantities. The plot is typically formed of point marks, usually small circles, positioned along a quantitative scale. The points are then evenly distributed using a second dimension of space above and below the quantitative axis baseline, not to represent any quantitative measure, but to accommodate closely packed points that have similar value positions. If categorical differentiation is necessary, attributes of colour or variation in symbol may be employed within a single plot, otherwise several separate plot views will be created for each discrete category grouping.

EXAMPLE Comparing the distribution of household incomes for a simulated population of Chicago residents broken down by ethnic group.

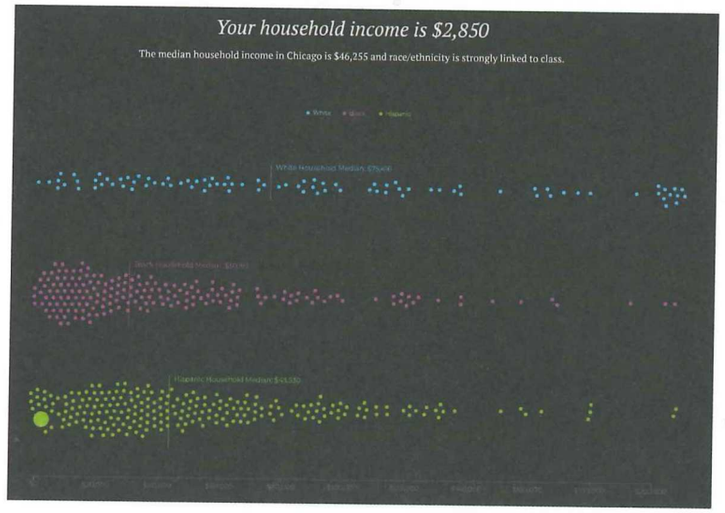


Figure 6.17 Is Your Child Ready for School?, by Gabrielle LaMarr LeMee

PRESENTATION TIPS

INTERACTIVITY: Interactive features can be useful to enable users to discover the value labels of each item through tooltips.

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload. Direct labelling will normally be restricted to noteworthy points only. Any colours used must be explained through the inclusion of a legend.

COLOUR: Colour may be used to establish the focus of certain points and/or distinction between different sub-category groups to assist with interpretation.

COMPOSITION: As the representation of the quantitative values is encoded through position along a scale, the quantitative axis does not need to have a zero origin, unless this is meaningful to the subject. If you do not commence from an origin of zero, this will need to be clearly annotated.

VARIATIONS & ALTERNATIVES

A variation in the encoding of the beeswarm plot may see the point marks replaced by shape marks (usually circles) in order to represent a second quantitative measure through size variation. An alternative chart type would be the 'dot plot', which removes the second dimension spread of values and overlays similar values. You could also use a 'histogram' to show the frequency and distribution of values in discrete quantitative groupings.

HISTOGRAM

ALSO KNOWN AS Bar chart (wrongly), population pyramid

C H R T S

DISTRIBUTIONS

REPRESENTATION DESCRIPTION

A histogram displays the frequency and distribution of quantitative measurements across grouped values for data items. Whereas bar charts compare quantities for discrete nominal categories, a histogram uses discrete quantitative 'bins' to form ordinal value groupings. The representation is formed using variation of line size (if the value groupings have equal intervals) or of shape area (if the value groupings have unequal value intervals) to represent the frequency of measurements.

EXAMPLE Comparing the distribution of lightness range among common foundation products sold in four regions.

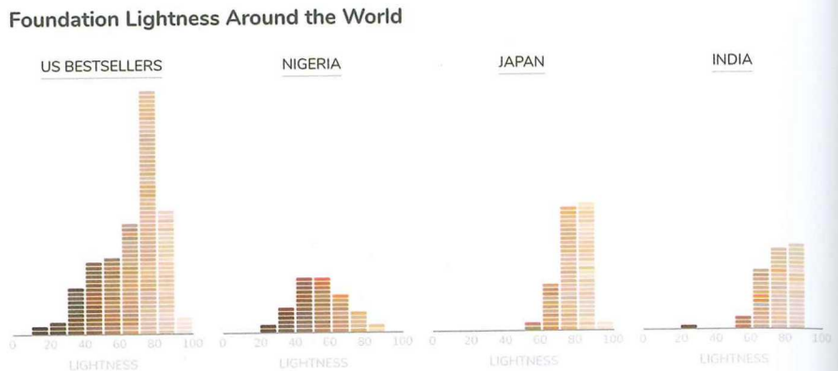


Figure 6.18 Beauty Brawl: How Inclusive are Beauty Brands Around the World?, by Amber Thomas, Jason Li and Divya Manian for 'The Pudding'

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload.

COMPOSITION: Unlike the bar chart there should be no (or, at most, a very thin) gap between bars to help the collective shape of the frequencies emerge. The sorting of the quantitative value bins must be presented in ascending order so that the reading of the overall shape preserves its meaning. The number of value bins and the range of values covered by each have a prominent influence over the appearance of the histogram and the usefulness of what it might reveal: too few bins may disguise interesting nuances, patterns and outliers; too many bins and the most interesting shapes may be abstracted by noise above signal. There is no singular best approach, the right choice simply arrives through experimentation and iteration.

VARIATIONS & ALTERNATIVES

For an analysis that looks at the distribution of values across two dimensions, such as populations by age group across binary gender categories, you might consider a 'back-to-back histogram' also commonly known as a 'population pyramid'. A 'box-and-whisker plot' is an alternative approach that reduces the display of the distribution of values to just five key statistical measures. To reveal more granular detail, the 'dot plot' and 'beeswarm plot' display all items individually across a quantitative scale.

DENSITY PLOT

ALSO KNOWN AS Ridgeline plot

C H R T S

DISTRIBUTIONS

REPRESENTATION DESCRIPTION

Density plots display the frequency and distribution of quantitative values for data items. Whereas histograms compare quantities using discrete quantitative 'bins' to form ordinal value groupings, a density plot can be considered a smoothed histogram. The plot is typically formed of a quantitative scale along which a line connects measurements of the frequency of each quantitative value. The line gets higher as the frequency gets higher. The connected line is then smoothed using various statistical techniques (that will depend on the subject context) and the area below is filled with colour to help visibility of the resulting shape. This creates the appearance of an 'area chart'. Often the density plot comprises multiple rows to separate observations across discrete category groupings.

EXAMPLE Comparing the distribution of scores allocated to a selection of words or phrases indicating the perceived level of positivity or negativity.

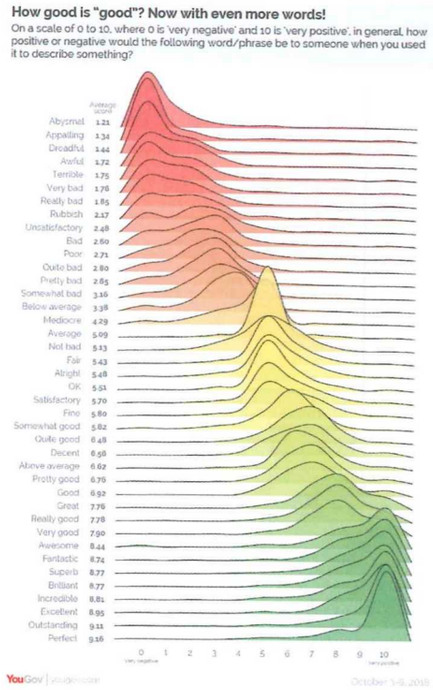


Figure 6.19 How Good is 'Good'?, by Matthew Smith

VARIATIONS & ALTERNATIVES

A variation in the design of a density plot is the 'violin plot' whereby the shape of distribution is plotted symmetrically creating a two-sided violin-like, rather than the one-sided shape of the density plot. An alternative role for the density plot would be in the form of an 'area chart', which plots quantitative trends over a temporal axis rather than a quantitative scale. An alternative chart type would be the 'beeswarm plot' to show the quantitative values of individual data items or a 'histogram' to show the frequency and distribution of values in discrete quantitative groupings.

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines is not usually necessary with density plots as they are more about getting a sense of the shape and patterns.

COMPOSITION: Depending on the nature of the quantitative measurements, and in particular the presence of outlier shapes in the distribution of values, the density plot is often presented in a way whereby high-value area 'spikes' intrude into and over the row space occupied by categories above. The arrangement of discrete categories is important to avoid too much occlusion and/or wasted empty space.

BOX-AND-WHISKER PLOT

ALSO KNOWN AS Box plot, candlestick chart, OHLC chart

C H R T S

DISTRIBUTIONS

REPRESENTATION DESCRIPTION

A box-and-whisker plot displays the distribution and shape of a series of quantitative values for different categories. The display is formed by a combination of lines and point markers to indicate (through position and length), typically, five different statistical measures. Three of the statistical values are common to all plots: the first quartile (25th percentile), the second quartile (or median) and the third quartile (75th percentile) values. These are displayed with a box (effectively a wide bar) positioned and sized according to the first and third quartile values with a marker indicating the median. The remaining two statistical values vary in definition: usually the minimum and maximum values or the 10th and 90th percentiles. These statistical values are represented by extending a line beyond the bottom and top of the main box to join with a point marker indicating the appropriate position. These are the whiskers. A single plot will be produced for each relevant, discrete category grouping.

EXAMPLE Comparing the distribution of annual earnings 10 years after starting school for graduates across the eight Ivy League colleges.

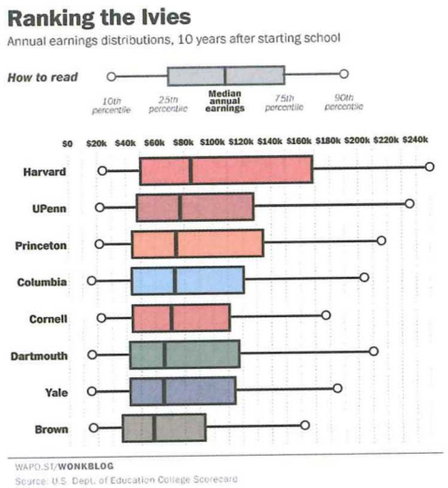


Figure 6.20 This Chart Shows How Much More Ivy League Grads Make Than You, by Christopher Ingraham (*Washington Post*)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload. Direct labelling will normally be restricted to noteworthy points only.

COMPOSITION: The quantitative value axis does not need to commence from zero, unless it means something significant to the interpretation, as the ranges of values themselves do not necessarily start from zero and the focus is more on the statistical properties between the outer values. There is no significant difference in perception between vertically or horizontally arranged box-and-whisker plots; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each bar. When you have several plots in the same chart, where possible try to make the categorical sorting meaningful, perhaps by organising values in ascending or descending order based on the median value.

VARIATIONS & ALTERNATIVES

Variations mainly concern changing the number of statistical measures included in the display. Sometimes you might remove the 'whiskers' to show just the 25th and 75th percentiles through the lower and upper parts of the 'box'. The 'candlestick chart' (or 'OHLC chart' used in stock market analysis to track the opening, highest, lowest and closing prices of stocks) uses a similar method and is often used to show the distribution and milestone quantitative values for events that encounter constant change, such as stock market analysis over a given time frame based on showing the opening, highest, lowest and closing prices.

PIE CHART

ALSO KNOWN AS Pizza chart, donut chart (wrongly)

C H R T S

PART-TO-WHOLE

REPRESENTATION DESCRIPTION

A pie chart shows how proportions of quantities for different constituent categories make up a whole. It uses a circular display divided into sectors for each category, with the angle representing the percentage proportions and attributes of colour to separate the discrete categories. The resulting size of the sector (in area terms) is a spatial by-product of the angle and so offers an additional means for judging values. The total of all sector values must be 100% and the constituent parts must be exclusive and representative of a meaningful 'whole', otherwise the chart will be corrupted.

EXAMPLE Comparing the proportion of Michael Schumacher's F1 races by result group.

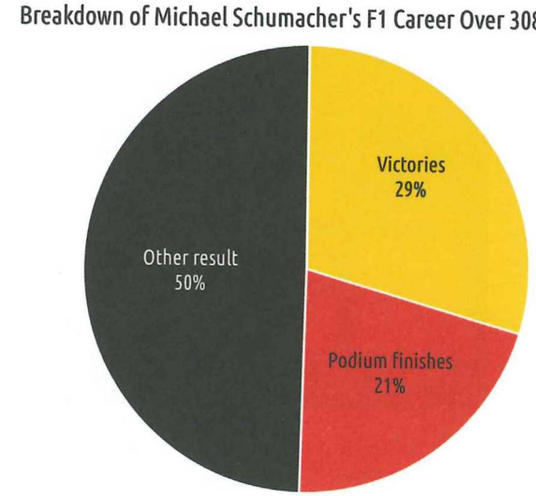


Figure 6.21 Breakdown of Michael Schumacher's F1 Career Over 308 Races

PRESENTATION TIPS

ANNOTATION: Directly labelling each category and associated value can enhance readability but may create inelegant clutter depending on the shape of the data and the size of the label values. Any colours used must be explained through the inclusion of a legend.

COLOUR: Colour is used to classify the categorical associations of each sector, so aim to vary the hue property of each colour to maximise the visible difference. When you have multiple sectors, you might choose to emphasise only two or three parts through editorial selection.

COMPOSITION: Positioning the first slice at the vertical 12 o'clock position gives a useful baseline to help judge the first sector angle value. The ordering of sectors using descending values or ordinal characteristics helps with the overall readability and allocation of effort.

VARIATIONS & ALTERNATIVES

The principal variation of the pie chart would be the 'donut chart'. Its function is exactly the same, but the donut has the centre removed, often to accommodate a labelling property. This removes the possibility of judging the sector angles at the circle origin, so the reading is formed by the arc lengths. The role of a pie chart is primarily about being able to compare a 'part to a whole' than being able to compare one part to another part. If you want to display and compare multiple parts, the 'bar chart' will offer a better option. For showing many parts, especially if they are organised into hierarchical groupings, the 'treemap' is a good option. Depending on the allocated space, a 'stacked bar chart' may provide an alternative layout to the pie chart, especially if your categorical values have an ordinal relationship. A 'nested shape chart', typically based on square or circle marks, enables comparisons across a series of one-part-to-whole relationships showing absolute values (through size) and proportions (through relative size).

WAFFLE CHART

ALSO KNOWN AS Square pie, unit chart, grid plot

C H R T S

PART-TO-WHOLE

REPRESENTATION DESCRIPTION

A waffle chart shows how proportions of quantities for different constituent categories make up a whole. It uses a square display divided typically into 100 points arranged in a grid layout. Each constituent proportion is displayed through colour coding the relevant number of points. The role of the waffle chart is to simplify the counting of proportions in contrast to the angle judgements of the pie chart, though the display is limited to only showing integer values. The total of all sector values must be 100% and the constituent parts must be exclusive and representative of a meaningful 'whole', otherwise the chart will be corrupted.

EXAMPLE Comparing the percentage of hours of sun by month during 2017 in Nuorgam, Finland.

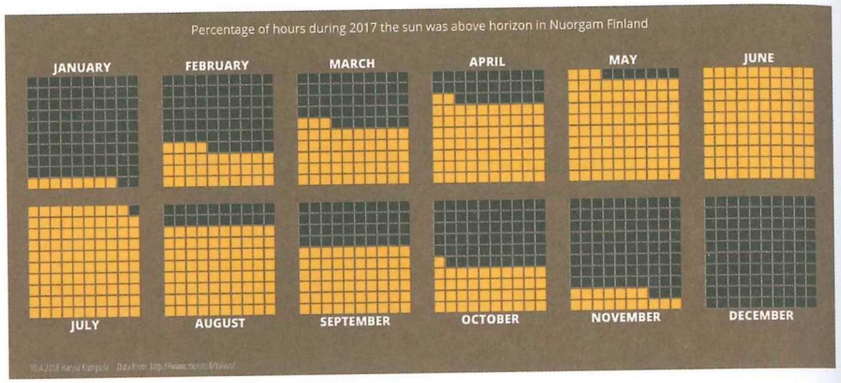


Figure 6.22 Percentage of Hours During 2017 the Sun was Above the Horizon in Nuorgam, Finland, by Hanna Kumpula (@kumpulahanna)

PRESENTATION TIPS

ANNOTATION: Chart apparatus is rarely applied to a waffle chart, though direct labelling may be included, perhaps using a nearby caption to indicate a category and quantitative label. Any colours used must be explained through the inclusion of a legend.

COLOUR: Adding outlines to each point mark (grid cell or circle) can be useful to help discern individual units.

COMPOSITION: A waffle chart is quicker to read when clusters of units, such as groups of five or ten, can be easily recognised. You may therefore seek to arrange the cells in groups to facilitate this. When you have several parts in the same waffle chart, where possible try to make the categorical sorting meaningful.

VARIATIONS & ALTERNATIVES

Rather than using colour, sometimes variations in symbols will be used to classify different categories or groupings. For example, you might see figures or gender icons used to show the makeup of a given sample population. A variation in the role of a waffle chart is to show quantitative counts rather than proportions of a whole, and this approach somewhat overlaps with applications of the 'pictogram'. A 'nested shape chart' using sized rectangular shapes may provide an alternative way of showing a part-to-whole relationship while also occupying a squarified layout.

STACKED BAR CHART

ALSO KNOWN AS Stacked chart, packed bars

C H R T S

PART-TO-WHOLE

REPRESENTATION DESCRIPTION

A stacked bar chart shows how quantitative values for different constituent categories make up a whole across major category items. The proportion of each constituent categorical 'part' is represented by separate bars that are sized according to their quantitative proportion and then stacked to create the whole. Sometimes the whole is standardised to represent 100%, otherwise it will be representative of an absolute total. Colour attributes are used to classify the discrete categorical parts. Stacked bar charts often work best when the categories are ordinal in nature, and it is the overall pattern of spread across the whole that is important. If the parts are representative of nominal categories, judging and comparing the size of individual stacked parts become quite hard without a common baseline, so you might seek to reduce the number of discrete values.

EXAMPLE Comparing the degree of trust held by Australians for different institutions.

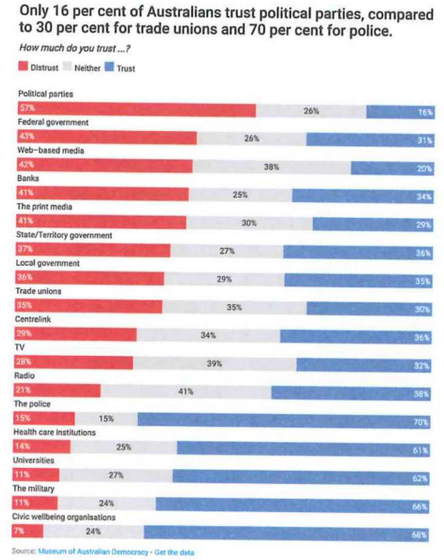


Figure 6.23 In a Nation of Cynics, We're Flocking to the Fringe, by ABC

VARIATIONS & ALTERNATIVES

The main alternative would be to create multiples of bar charts each showing the quantitative values for just a single constituent part for each major category item. The 'waterfall chart' splits out the individual constituent parts to create a step-by-step breakdown of a single stacked whole. Like their unstacked siblings, stacked bar charts can also be used to show how value proportions have changed over time.

PRESENTATION TIPS

ANNOTATION: Direct value labelling can become very cluttered when there are many parts, so you may choose to focus only on labelling noteworthy values. Axis scales using logical intervals will be helpful, as will the inclusion of gridlines, especially highlighting key features such as the 50% position when your data is displaying a 100% stacked total. Any colours used must be explained through the inclusion of a legend.

COLOUR: If you are representing categorical ordinal data, colour can be astutely deployed to give a sense of the general balance of values within the whole, but this will only work if their sorting arrangement within the stack is logically applied. For categorical nominal data, ensure the stacked parts have sufficiently different colour hues so their distinct bar lengths can be easily detected.

COMPOSITION: The bars should be proportionally sized according to the associated quantitative value – nothing more, nothing less – otherwise the perception of the bar sizes will be distorted. Most commonly, this means setting the quantitative value scale to an origin of zero. There is no significant difference in perception between vertically or horizontally arranged stacked bar charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each cluster. Including a noticeable gap between each stack of bars will help to preserve a clear distinction between the primary category items. Aim to make the sorting of values in the chart as meaningful as possible.

DIVERGING BAR CHART

ALSO KNOWN AS Back-to-back bar chart, paired bar chart, spine chart

C H R T S

PART-TO-WHOLE

REPRESENTATION DESCRIPTION

A diverging bar chart shows how quantitative values for different constituent categories make up a whole across major category items. The proportion of each constituent categorical 'part' is represented by separate bars that are sized according to their quantitative proportion and then stacked to create the whole. Sometimes the whole is standardised to represent 100%, otherwise it will be representative of an absolute total. In contrast to the stacked bar chart, the diverging bar chart arranges constituent categorical parts either side of a common baseline depending on the discrete nominal or ordinal relationships that benefit from such separation. Colour attributes are commonly used to classify the discrete categorical parts.

EXAMPLE Comparing the responses to a survey question asking for opinions about the legality of abortion across different demographic categories.

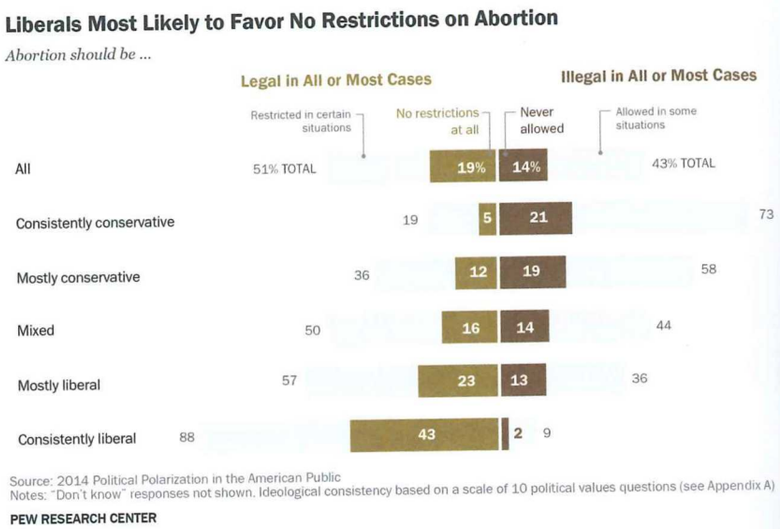


Figure 6.24 Political Polarization in the American Public, Pew Research Center, Washington, DC (February, 2015) (http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/)

PRESENTATION TIPS

ANNOTATION: Direct value labelling can become cluttered when there are many constituent parts, so you may choose to focus only on labelling noteworthy values. Axis scales using logical intervals will be helpful, as will the inclusion of gridlines. Any colours used must be explained through the inclusion of a legend.

COLOUR: If you are representing categorical ordinal data, colour can be astutely deployed to give a sense of the general balance of values within the whole, but this will only work if their sorting arrangement within the stack is logically applied. For categorical nominal data, ensure the stacked parts have sufficiently different colour hues so their distinct bar lengths can be easily detected.

COMPOSITION: There is no significant difference in perception between vertically or horizontally arranged diverging bar charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each cluster. Including a noticeable gap between each stack of bars will help to preserve a clear distinction between the primary category item. Aim to make the sorting of values in the chart as meaningful as possible.

VARIATIONS & ALTERNATIVES

A variation would be a 'diverging histogram' whereby the major categories have ordinal qualities, like increasing age groups, and the resulting shape of the chart has meaning about the distribution of values. If you want to facilitate direct comparison, a 'clustered bar chart' showing adjacent bars may offer a better alternative solution.

MARIMEKKO CHART

ALSO KNOWN AS Mekko chart, mosaic plot, proportional stacked bar

C H R T S

PART-TO-WHOLE

REPRESENTATION DESCRIPTION

A Marimekko chart is effectively a two-dimensional stacked bar chart with variation in size for both height and width to display parts of a whole simultaneously across two dimensions. It is often used to contextualise percentage part-to-whole comparisons of major categories with a second dimension of absolute numbers that make up a total. Attributes of colour are commonly used to provide categorical classifications.

EXAMPLE Comparing the proportion and number of competitors by gender across all Summer Olympic Games.

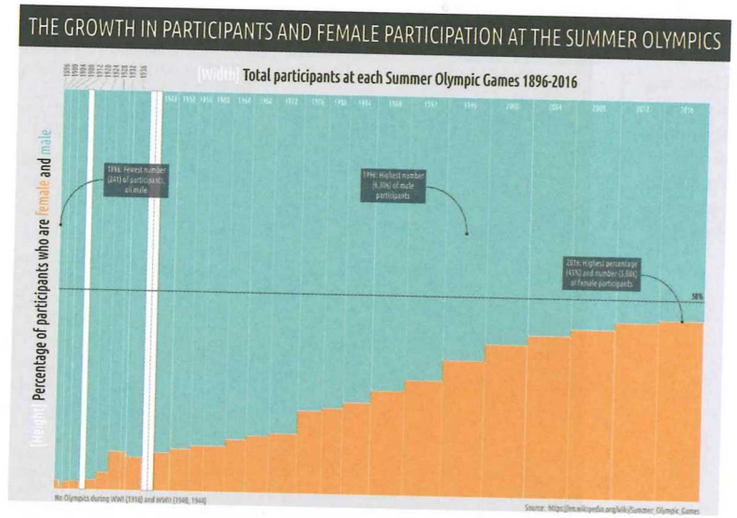


Figure 6.25 The Growth in Participants and Female Participation at the Summer Olympics

PRESENTATION TIPS

ANNOTATION: With two quantitative axes and two dimensions of categorical division, labelling Marimekko charts can become quite cluttered. At the very least, the two axes should be clearly titled, and some size scales provided, either through axis interval labelling or direct labelling of noteworthy items. Any colours used must be explained through the inclusion of a legend.

COLOUR: It will usually be possible to distinguish classifications visually across only one of the categorical dimensions.

VARIATIONS & ALTERNATIVES

An alternative to the Marimekko chart would be the treemap which shows part-to-whole relationships when there are many category parts and there is some hierarchical organisation of those categories.

DENDROGRAM

ALSO KNOWN AS Radial tree, layout tree, cluster tree, tree hierarchy

REPRESENTATION DESCRIPTION

A dendrogram is a node–link diagram that displays hierarchical relationships across multiple tiers of categorical dimensions. It displays a hierarchy based on multi-generational 'parent-and-child' relationships. Starting from a singular origin root node (or 'parent') each subsequent set of constituent 'child' nodes, a tier below and represented by points, is connected by lines (curved or straight) to indicate the existence of a relationship. Each constituent node may then have further constituencies represented in the same way, continuing through to the lowest tier of detail.

EXAMPLE Showing a breakdown of the job skills required by workers in the UK job market organised by a hierarchical taxonomy of skill clusters.

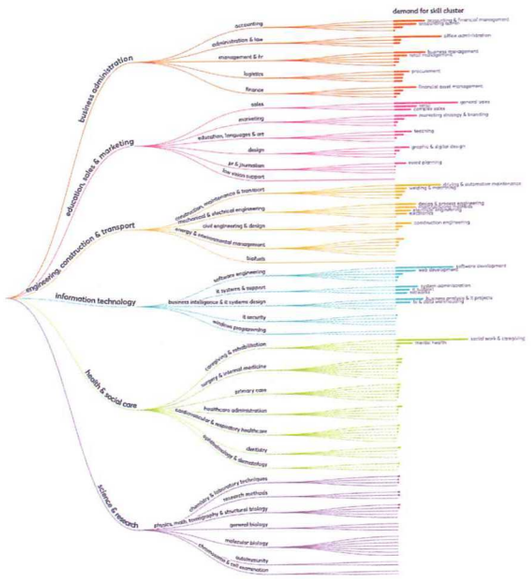


Figure 6.28 Making Sense of Skills: A UK Skills Taxonomy, by Dr Cath Sleeman

VARIATIONS & ALTERNATIVES

Variations of the dendrogram involve incorporating some additional quantitative representation such as using the length or width of connecting lines and, on replacing the point marks for each node, varying the size of node shapes. An alternative approach would be to consider the 'sunburst chart' which would show a part-to-whole relationship across the constituent categories in each hierarchical tier.

PRESENTATION TIPS

INTERACTIVITY: A useful interactive feature would be to enable filtering or highlighting of branches of interest and selection options for revealing tooltips if labelling is too difficult to accommodate elegantly.
ANNOTATION: If labelling is required, depending on the number of tiers and nodes, the size of the text will need to be carefully considered to ensure readability and minimise the effect of clutter.
COLOUR: Colour would be an optional attribute for accentuating certain nodes or applying further detail of categorisation. The colour of the connecting lines is usually based on a neutral option like black or grey.
COMPOSITION: The layout can be based on either a linear tree (typically left to right, top to bottom) or radial tree (outwards from the centre) structure. Sometimes, the hierarchical tiers do not necessarily have a parent–child relationship, so their ordering can be legitimately switched around. Therefore, careful decisions are needed about the most logical hierarchical sequencing given the subject matter and enquiry. There is also scope for arranging the sequencing of constituent parts within each tier in a meaningful way.

VENN DIAGRAM

ALSO KNOWN AS Set diagram, Euler diagram (wrongly)

REPRESENTATION DESCRIPTION

A Venn diagram shows collections of and relationships between multiple sets. It typically uses circular containers to represent all independent and intersecting permutations with value labels or point marks used to place all category items in the appropriate container. The size of the contained area is (typically) not important; what is important is in which containing region an item resides. Variations in the attributes of colour or symbol may be commonly used to represent further unique distinction among the items displayed.

EXAMPLE Comparing sets of permutations for legalities around marijuana usage and same-sex marriage across states of the USA as at 2014.

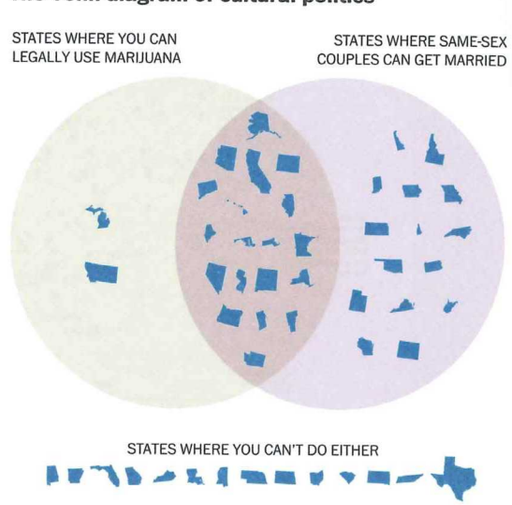


Figure 6.29 This Venn Diagram Shows Where You Can Both Smoke Weed and Get a Same-sex Marriage, by Phillip Bump (*Washington Post*)

PRESENTATION TIPS

ANNOTATION: The main annotation feature required will be to make clear which containers relate to which set or membership grouping. When the permutations increase in number (e.g. three- or four-way Venns) it can be hard to accommodate reasonable labels in each possible container.
COLOUR: Colour is often used to create more immediate distinction between the intersections and independent parts or members of each container, especially when multi-way Venns are being used.
COMPOSITION: As the attributes of size and shape of the containers are of no significance there is more flexibility to manipulate the display to modify the layout to accommodate the number or size of items in each container group. While it is theoretically possible to exceed five-way Venn diagrams, the ability of readers to make sense of such displays diminishes significantly.

VARIATIONS & ALTERNATIVES

A common variation or alternative to the Venn (but often mistakenly called a Venn) is the 'Euler diagram'. The difference is that an Euler diagram does not need to present every possible intersection and independency from all categorical sets. A different approach to visualising sets (especially larger numbers) can be achieved using the 'UpSet' technique, which uses a matrix layout to present all possible set combinations and then a second, aligned method like a bar chart to reveal a quantitative count for each set.

SCATTER PLOT

ALSO KNOWN AS Scatter graph, scatter chart

CORRELATIONS

REPRESENTATION DESCRIPTION

A scatter plot displays the relationship between two quantitative variables for different category items. The display is formed by point marks for each item, plotted positionally along each quantitative axis. Sometimes attributes of colour hue are used to distinguish categorical dimensions across all items.

EXAMPLE Exploring the relationship between life expectancy and the percentage of healthy years across all countries.

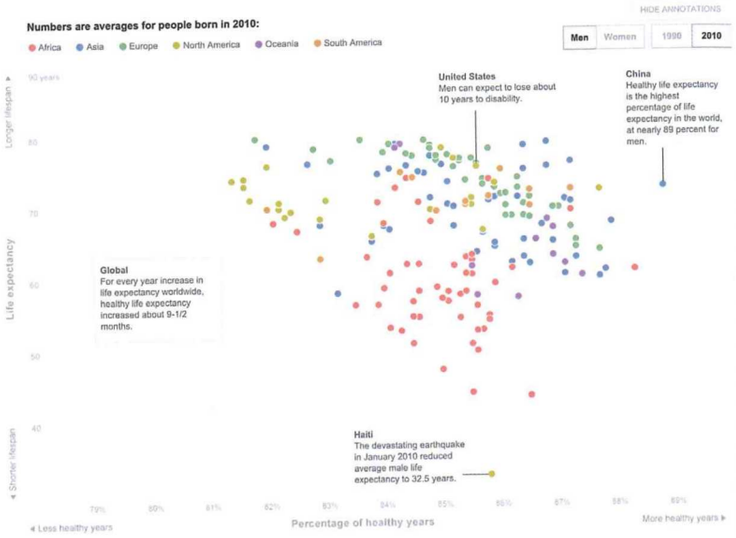


Figure 6.30 How Long Will We Live – And How Well?, by Bonnie Berkowitz, Emily Chow and Todd Lindeman (*Washington Post*)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. Reference lines, such as a trend line of best fit, might also aid interpretation. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload. Direct labelling will normally be restricted to noteworthy points only. Any colours used must be explained through the inclusion of a legend.

COLOUR: To overcome occlusion caused by plotting several marks at the same value position, you could use unfilled or semi-transparent filled circles to help convey value frequency.

COMPOSITION: As the representation of the quantitative values is encoded through position along a scale, the quantitative axis does not need to have a zero origin, unless this is meaningful to the subject. If you do not commence from an origin of zero, this will need to be clearly annotated. Ideally a scatter plot will have a squared aspect ratio (equally tall as it is wide) to help patterns surface more evidently. If one quantitative variable (e.g. weight) is likely to be affected by the other variable (e.g. height), it is general practice to place the former on the y-axis and the latter on the x-axis. If you have to use a logarithmic quantitative scale on either or both axes, you need to make this clear to viewers.

VARIATIONS & ALTERNATIVES

A 'ternary plot' is a variation of the scatter plot through the inclusion of a third quantitative variable axis. The 'bubble plot' also incorporates a third quantitative variable, this time through encoding the size of a geometric shape (replacing the point marker).

A 'scatter plot matrix' involves a single view of multiple scatter plots presenting different combinations of plotted quantitative variables, used to explore possible relationships among larger multivariate datasets.

A 'connected scatter plot' compares the shifting state of two quantitative measures over time.

BUBBLE PLOT

ALSO KNOWN AS Bubble chart

CORRELATIONS

REPRESENTATION DESCRIPTION

A bubble plot displays the relationship between three quantitative variables for different category items. In contrast to the scatter plot, the bubble plot uses shape marks (usually circles) for each category item, plotted positionally along each quantitative axis with the variation in size of each mark representing a third quantitative measure. Sometimes attributes of colour hue are used to distinguish categorical dimensions across all items.

EXAMPLE Exploring the relationship between ecological footprint and biocapacity by country.

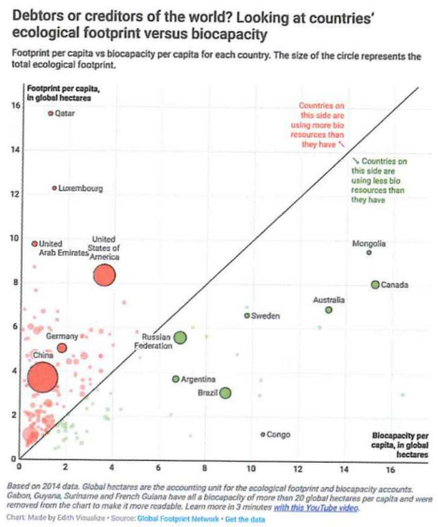


Figure 6.31 Debtors or Creditors of the World? Looking at Countries' Ecological Footprint Versus Biocapacity, by Lisa Rost and Edith Maulandi

VARIATIONS & ALTERNATIVES

If the third quantitative variable is removed, the chart type would revert to a 'scatter plot'. Variations on the bubble plot might see the use of different geometric shapes as the markers, maybe introducing extra meaning through the shape and dimensions used.

PRESENTATION TIPS

INTERACTIVITY: A useful interactive feature would be to enable filtering or highlighting of certain categorical items, especially if there are several distinct categories and lots of items to make sense of. Furthermore, selection options for revealing tooltips can be helpful if direct labelling is too difficult to accommodate elegantly.

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. Reference lines, such as a trend line of best fit, might also aid interpretation. If you include axis-scale labels you should not need to label each value directly, as this will lead to label overload. Direct labelling will normally be restricted to noteworthy points only. Any colours used must be explained through the inclusion of a legend.

COLOUR: If colours are being used to distinguish the different categories, ensure these are as visibly different as possible. When data values are especially diverse in range, the size of shapes may vary from very small to quite large. The largest shapes may overlap, in spatial terms, with other values or even hide them completely. The use of outline borders and semi-transparent colours can help avoid the effect of total occlusion.

COMPOSITION: As the representation of the quantitative values is encoded through position along a scale, the quantitative axis does not need to have a zero origin, unless this is meaningful to the subject. If you do not commence from an origin of zero, this will need to be clearly annotated. Ideally a bubble plot will have a squared aspect ratio (equally tall as it is wide) to help patterns surface more evidently. If one quantitative variable (e.g. weight) is likely to be affected by the other variable (e.g. height), it is general practice to place the former on the y-axis and the latter on the x-axis. If you have to use a logarithmic quantitative scale on either or both axes, you need to make this clear to viewers. The geometric accuracy of the shape mark size calculation is paramount: it is the area you are modifying, not the diameter/radius.

NETWORK DIAGRAM

ALSO KNOWN AS Node–link diagram, graph, hairball graph, social network

CONNECTIONS

REPRESENTATION DESCRIPTION

Node–link diagrams display relationships through the connections between data items. The common version of this type of diagram displays items as nodes, represented by point marks, with links or edges (represented by lines) depicting the existence of a connection. These connecting lines will sometimes encode an attribute of direction to indicate the influencer relationship. Replacing point marks with a geometric shape and using attributes of size is a further variation. In some versions a further quantitative weighting is applied to show the relationship strength, maybe through increased line width. Attributes of colour may be used to indicate a quantitative value (e.g. number of connections) and/or some categorical classification.

EXAMPLE Exploring the connections of voting patterns for Democrats and Republicans across all members of the US House of Representatives from 1949 to 2012.

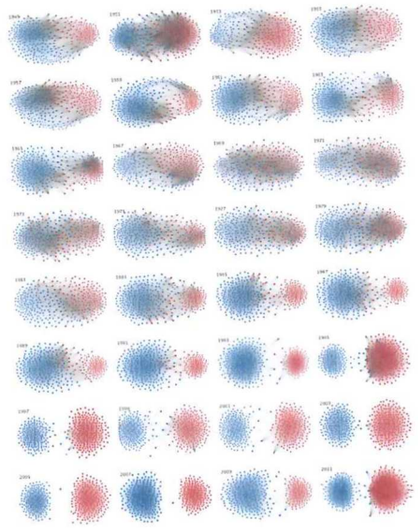


Figure 6.32 The Rise of Partisanship and Super-cooperators in the U.S. House of Representatives, visualisation by Mauro Martino, authored by Clio Andris, David Lee, Marcus J. Hamilton, Mauro Martino, Christian E. Gunning, and John Armistead Selde

VARIATIONS & ALTERNATIVES

There are many derivatives of the node–link diagram, as explained, based on variations in the use of different attributes. 'Hive plots' and 'BioFabric' offer alternative approaches based on replacing nodes with vertices.

PRESENTATION TIPS

INTERACTIVITY: Node–link diagrams are particularly useful when offered with interactive features, enabling the user to interrogate and manipulate the display to facilitate visual exploration. The option to apply filters to reduce the busyness of the display, and maybe enable isolation of individual node connections, can help viewers to focus on specific parts of the network rather than face the whole system at once.

ANNOTATION: The complexity revealed by these diagrams is often a reflection of the underlying subject complexity, so it can be helpful to use annotation to surface key observations about significant clusters or label those items with the most connections.

COLOUR: Aside from the possible categorical colouring of each node, decisions need to be made about the colour of the connecting lines, especially with regard to how prominent these links will be in contrast to the nodes.

COMPOSITION: Composition decisions will be so varied for any network diagram depending on the complexity and volume of the data and the output constraints around space and consumption format. There are several common algorithmic treatments used to compute custom arrangements to optimise network displays, such as force-directed layouts using the physics of repulsion and springs to amplify relationships. There are also simplifying techniques such as edge bundling to aggregate or summarise multiple similar links.

SANKEY DIAGRAM

ALSO KNOWN AS Alluvial diagram

CONNECTIONS

REPRESENTATION DESCRIPTION

Sankey diagrams display categorical composition and flows of quantitative relationships between different major ordinal dimensions. The original application of the Sankey diagram displayed flow relationships over many discrete ordinal stages, but it would be reasonable to say most common forms involve a two-sided parallel display. The two sides represent different states of a paired, ordinal relationship, such as input vs output, or time A vs time B. On each side there is effectively a stacked bar chart displaying proportionally sized and differently coloured (or spaced apart) constituent parts of each whole. These might show categorical breakdowns of income vs categories of expenditure or the categorical composition of some whole in a before and after state. Curved bands join each side of the display to represent the connecting categories (origin and destination) with proportionally sized thickness representing the quantitative flow of this relationship.

EXAMPLE Showing a breakdown of reasons for animals being brought into a shelter and a breakdown of the related outcomes of each animal after one month.

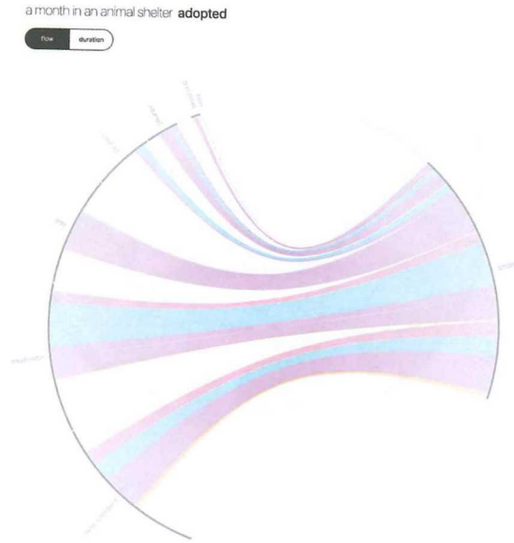


Figure 6.33 A Month in an Animal Shelter, by Sarah Campbell

VARIATIONS & ALTERNATIVES

The Sankey is closely related to the 'alluvial diagram', which tends to show changes in composition and flow over time and often across multiple stages (rather than just the common paired structure of the Sankey). These days, the labels are often applied interchangeably. A 'chord diagram' is a variation that uses a radial display to enable certain origins and destinations to be one and the same. Showing how component parts have changed over time could just be displayed using a 'stacked area chart'. Plotting composition and flow can also be applied to a spatial display to create a variation of the geographical 'flow map'.

PRESENTATION TIPS

INTERACTIVITY: Sankey diagrams are particularly useful when offered with interactive features, enabling the user to interrogate and manipulate the display to facilitate visual exploration. The option to apply filters to reduce the chaos of the visual and enable isolation of individual or groups of flows helps users to focus on specific relationships of interest. Interactively enabled labelling can also be beneficial as direct labelling is difficult to incorporate elegantly.

ANNOTATION: Direct labelling will normally be restricted to noteworthy points only. Any colours used must be explained through the inclusion of a legend.

COLOUR: Colouring is often used visually to indicate the categories of the connecting bands, though this can become complicated by an origin categorical colour joining to a different destination colour. A sense of this change can be conveyed by blending the origin and destination colours half-way across.

COMPOSITION: The main arrangement decisions concern sorting. Firstly, by deriving as much logical meaning from the categorical values within each stack and, secondly, by deciding on the sorting of the connecting lines in the z-dimension – if many lines are crossing, there is a need to think about which will be on top and which will be below. There is no significant difference between a landscape or portrait layout; it will depend on the subject-matter 'fit' and the space within which you have to work. Sometimes the stacks on each side are curved and appear more like stacked arcs.

CHORD DIAGRAM

ALSO KNOWN AS Radial Sankey diagram, radial network, arc diagram

C H **R** T S

CONNECTIONS

REPRESENTATION DESCRIPTION

A chord diagram displays relationships through connections between and within category items. The diagram is formed around a radial display with categories located around the edge: either shown as individual nodes or as arc segments proportionally sized around the circumference to represent a part-to-whole breakdown. Emerging inwards from each origin position are curved lines that join with related categorical destinations around the edge. The connecting lines are proportionally sized according to a quantitative measure and a directional or influencing relationship is often indicated. Attributes of colour hue are commonly used to distinguish different category groupings visually.

EXAMPLE Exploring the connections of migration between and within ten world regions based on estimates across five-year intervals between 1990 and 2010.

Figure 6.34 The Global Flow of People, by Nikola Sander, Guy J. Abel and Ramon Bauer

PRESENTATION TIPS

INTERACTIVITY: Chord diagrams are particularly useful when offered with interactive features, enabling the user to interrogate and manipulate the display to facilitate visual exploration. The option to apply filters to reduce the chaos of the visual and enable isolation of individual or groups of flows helps users to focus on specific relationships of interest. Interactively enabled labelling can also be beneficial as direct labelling is difficulty to incorporate elegantly.

ANNOTATION: Direct labelling will normally be restricted to noteworthy points only. Any colours used must be explained through the inclusion of a legend.

COLOUR: Aside from the categorical colouring of each node, decisions need to be made about the colour of the connecting lines, especially on deciding how prominent these links will be in contrast to the nodes. Sometimes the connections will match the origin or destination colours, or they will combine the two (with a start and end colour blend to match the relationship).

COMPOSITION: The main arrangement decisions concern sorting. Firstly, by deriving as much logical meaning from the categorical values within each stack and, secondly, by deciding on the sorting of the connecting lines in the z-dimension – if many lines are crossing, there is a need to think about which will be on top and which will be below. Showing the direction of connections can be achieved using arrowheads or colour lightness variation. One common, subtle solution is to pull the destination join away from the edge of the destination arc to contrast with connecting lines that emerge directly from an origin.

VARIATIONS & ALTERNATIVES

Variations of the chord diagram would be to consider using a single baseline axis, placing all category items along it and forming connections between using semi-circular arcs. Additionally, a 'Sankey diagram' would be relevant if there are distinct origins and destination relationships to reveal.

LINE CHART

ALSO KNOWN AS Fever chart, stock chart

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

A line chart shows how quantitative values have changed over time for different categorical items. Line charts are typically structured around a continuous temporal x-axis and quantitative y-axis with values plotted using point marks at relevant coordinates. Connecting lines join up adjacent and related categorical items to form slopes which are then extended along the full timescale to display a complete change over time. Multiple categories can be displayed in the same view, each represented by a discrete line often with categorical or editorial colour attributes. The connecting lines are typically straight, but sometimes curved line 'interpolation' may be applied to help emphasise a general trend above precise point reading.

EXAMPLE Showing cumulative runs scored in Test matches by English Test batsmen between 1947 and 2018.

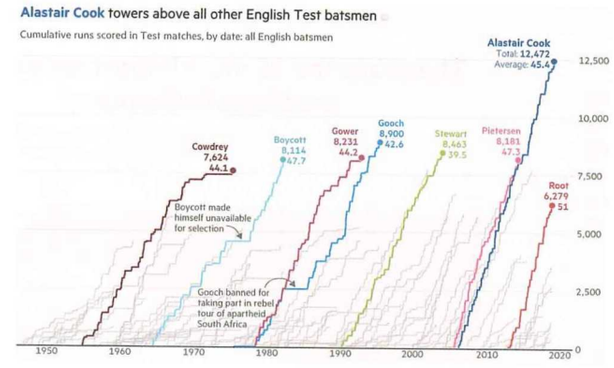


Figure 6.35 Cricketer Alastair Cook Plays His 161st and Final Test Match, by *Financial Times* / John Burn-Murdoch

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful if you have many discrete categorical lines and wish to enable the user to isolate a certain category of interest, either through filtering to exclude the others or using a contrasting colour to emphasise its shape among the rest.

ANNOTATION: Sometimes the point mark is quite pronounced, to aid value judgements and possibly to provide space for a value label, but on most occasions only the position of the connecting lines is displayed. Ranking labels can be derived from the vertical position along the scale so direct labelling is usually unnecessary. The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. You might choose to annotate specific values of interest (highest, lowest, specific milestones). Any colours used must be explained through the inclusion of a legend. If the shape of the data presents an opportunity, you might consider directly labelling each or specific category lines, at the first or last point mark position.

COLOUR: When many categories are shown, rather than colouring each line with a distinct categorical classification, it may only be viable to emphasise lines of interest using colour hue or saturation.

COMPOSITION: The chart's dimensions will need to be carefully considered, specifically the aspect ratio formed by its height and width. The upward and downward slopes can seem more significant if the chart width is narrow and less significant if it is more stretched out. There is no single practical rule to obey other than using common sense to ensure you do not overly amplify or underplay features of your data. The sequencing of values tends to follow a chronological left-to-right direction for the time-based x-axis and low values rising up to high values on the y-axis; you will need a good (and clearly annotated) reason to break this convention. Line charts do not always require the quantitative axis origin to start from zero, as the size of a value is represented through position along a scale rather than the size of a line or shape. If zero has significance for the interpretation of the trends portrayed, given the subject matter, then you should start the baseline at zero.

VARIATIONS & ALTERNATIVES

Variations of the line chart may include the 'bump chart', to show rankings over time, and the 'slope graph', to compare trends over two points in time. 'Spark lines' are mini line charts that aim to occupy only a word's length amount of space. They are often seen in dashboards where space is at a premium and there is a desire to optimise the density of the display. An alternative would to use the 'bar chart' when you have quantities for discrete periods (such as totals over a monthly period) rather than a purely continuous series of point-in-time measurements.

BUMP CHART

ALSO KNOWN AS Bumps chart, rank chart

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

A bump chart shows how quantitative values have changed over time for different categorical items, where the quantitative values are ranking measurements. These charts are typically structured around a continuous temporal x-axis and quantitative y-axis with values plotted using point marks at relevant coordinates. Connecting lines join up adjacent and related items to form slopes which are then extended along the full timescale to display a complete change over time. A common extension of the bump charts uses variation in the size (width) of each line to represent a quantitative measure, usually the absolute value that informs the ranking measurement. Multiple categories are commonly displayed in the same view, each represented by a discrete line often with categorical or editorial colour attributes. The connecting lines are typically straight, but sometimes curved line 'interpolation' may be applied to help emphasise a general trend above precise point reading.

EXAMPLE Showing changes in rank of the most politically important topics for Germans between 1998 and 2017.

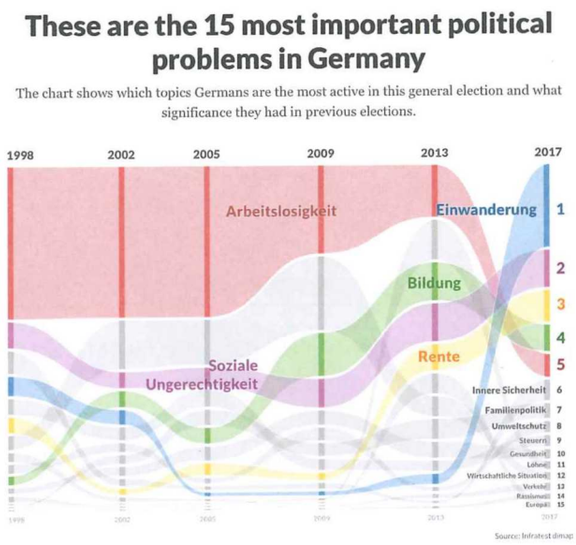


Figure 6.36 These are the 15 Most Important Political Problems in Germany, [Translated] by *Berliner Morgenpost*

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful if you have many discrete categorical lines and wish to enable the user to isolate a certain category of interest, either through filtering to exclude the others or using a contrasting colour to emphasise its shape among the rest.

ANNOTATION: Sometimes the point mark is quite pronounced, to aid value judgements and possibly to provide space for a value label, but on most occasions only the position of the connecting lines is displayed. Ranking labels can be derived from the vertical position along the scale so direct labelling is usually unnecessary. You might choose to annotate specific values of interest (highest, lowest, specific milestones). Any colours used must be explained through the inclusion of a legend. If the shape of the data presents an opportunity, you might consider directly labelling each or specific category lines, at the first or last point mark position, or even both.

COLOUR: When many categories are shown, rather than colouring each line with a distinct categorical classification, it may only be viable to emphasise lines of interest using colour hue or saturation.

COMPOSITION: The sequencing of values tends to follow a chronological left-to-right direction for the time-based x-axis and highest ranking values dropping to lowest ranking values on the y-axis; you will need a good (and clearly annotated) reason to break this convention.

VARIATIONS & ALTERNATIVES

Consider alternatives like 'line charts' and 'area charts' if the ranking measurement is of secondary interest to plotting absolute quantitative values.

SLOPE GRAPH

ALSO KNOWN AS Slope chart, parallel coordinates

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

A slope graph shows how quantitative values have changed over two points in time for different category items. The display is based on (typically) two parallel quantitative axes with a common value range. A line is plotted for each category connecting the two axes together with the vertical position on each axis representing the respective quantitative values. These connecting lines form slopes that indicate the upward, downward or stable trend between the two temporal axes. Attributes of colours are often used to distinguish different categorical lines or to surface major trends among the items plotted.

EXAMPLE Showing changes in the share of power sources across all US states between 2004 and 2014.

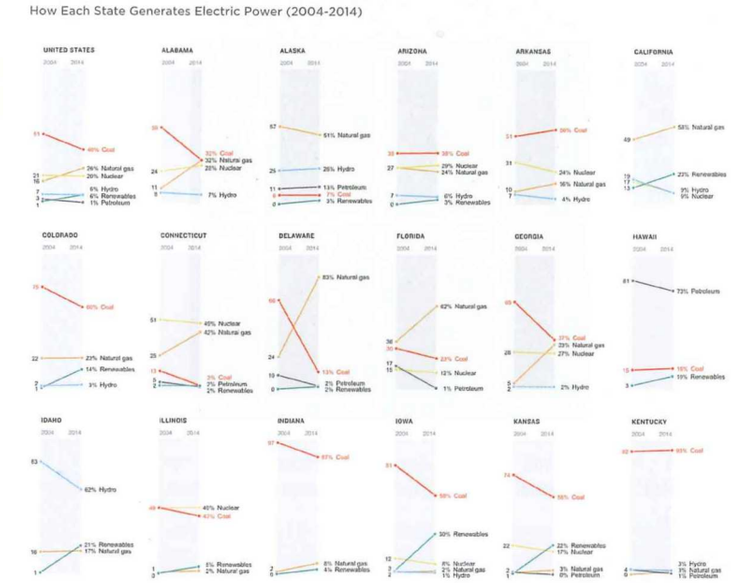


Figure 6.37 Coal, Gas, Nuclear, Hydro? How Your State Generates Power

Source: U.S. Energy Information Administration, Credit: Christopher Groskopf, Alyson Hurt and Avie Schneider (NPR)

PRESENTATION TIPS

INTERACTIVITY: Depending on the number of category values being presented, slope graphs can become quite busy, especially if there are bunches of similar quantitative values with slope transitions. This also causes a problem with accommodating multiple labels on the same value. On these occasions you might find interactive features useful to enable filtering of certain items, to exclude others or to highlight a selection. Discovering value labels of each item through interactive tooltips can also be beneficial.

ANNOTATION: Labelling of each category item on both sides will often be necessary, though this can be challenging composition-wise when there are several items positioned in close proximity. You might therefore choose to annotate only specific values of interest (highest, lowest, of editorial interest). The parallel axes will need clear labels to explain the respective points in time. Any colours used must be explained through the inclusion of a legend.

VARIATIONS & ALTERNATIVES

Rather than comparing two points in time, some variations in the application of a slope graph are used to show the relationship between discrete quantitative variables for related category items. In this case the connecting line is not indicative of a trend, rather a join to connected related items. This approach can also lead to the slope graph being extended beyond just two parallel axes and thus evolving into the technique known as 'parallel coordinates'. An alternative chart type would be to consider the 'connected dot plot' which can also show comparisons of quantities for two points in time across multiple category items.

CONNECTED SCATTER PLOT

ALSO KNOWN AS Trail chart, comet chart

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

A connected scatter plot displays the relationship between two quantitative measures over time. The display is formed of two quantitative x- and y-axes and with the values represented by point marks at the respective coordinates, one for each measurement over time. The individual points are then connected (think of a dot-to-dot drawing puzzle) using lines joining each consecutive point in time to form a sequence of change.

EXAMPLE Showing changes in the daily price and availability of Super Bowl tickets on the secondary market four weeks prior to the event across five Super Bowl finals.

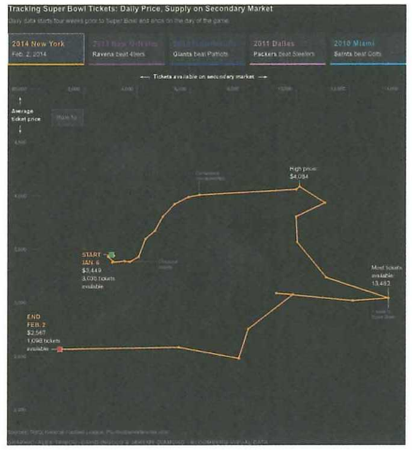


Figure 6.38 Holdouts Find Cheapest Super Bowl Tickets Late in the Game, by Alex Tribou, David Ingold and Jeremy Diamond (Bloomberg Visual Data)

VARIATIONS & ALTERNATIVES

The 'comet chart' is to the connected scatter plot what the 'slope graph' is to the 'line chart' – a summarised view of the relationship between two quantitative measures over two points in time. The comet aspect is demonstrated through the cone shape of the connecting line, with the more recent period of time generally having a thicker width. A variation of the connected scatter plot is simply the 'scatter plot' where there is no time dimension or elements of connectedness between points.

PRESENTATION TIPS

INTERACTIVITY: The biggest challenge is making the connections and the sequence as visible as possible. This becomes much harder when values change very little and/or they loop back to previous positions, crossing back over themselves. It is especially hard to label the sequential time values elegantly. One option to overcome this is through animated sequences which might build up the display, connecting one line at a time and unveiling the date labels as time progresses. It is often the case that only one series will be plotted. However, interactive options may allow the user to overlay one or more for comparison, switching them on and off as required.

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. If you can elegantly include direct labels to each point value, indicating the time period it relates to, then this can be helpful. Connected scatter plots are unfamiliar to many audiences and it can be quite demanding to learn how to read them. It may be necessary to provide a 'how to read' guide illustrating what the axis values represent and what it means when connecting lines are moving in different directions. Also consider labelling parts of the chart region that carry particular meaning, so if a connecting line moves into that region, the interpretation of what this means can be accelerated.

COLOUR: Colour might be used to explain the temporal status of the connecting lines, for instance using a faded colour for the past and a more vivid colour for the present. Otherwise, you might use attributes of colour to accentuate certain sections of a sequence that might warrant particular attention.

COMPOSITION: As the representation of the quantitative values is encoded through position along a scale, the quantitative axis does not need to have a zero origin, unless this is meaningful to the subject. If you do not commence from an origin of zero, this will need to be clearly annotated. Ideally a scatter plot will have a squared aspect ratio (equally tall as it is wide) to help patterns surface more evidently. If one quantitative variable (e.g. weight) is likely to be affected by the other variable (e.g. height), it is general practice to place the former on the y-axis and the latter on the x-axis.

AREA CHART

ALSO KNOWN AS Density plot

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

An area chart shows how quantitative values have changed over time for a single categorical item. The charts are typically structured around a continuous temporal x-axis and quantitative y-axis with values plotted using point marks at relevant coordinates. Connecting lines join up adjacent and related items to form slopes which are then extended along the full timescale to display a complete change over time. The connecting lines are typically straight, but sometimes curved line 'interpolation' may be applied to help emphasise a general trend above precise point reading. To accentuate the shape of the trends, the area beneath the line is filled with colour, which means the height of the area at any given point also reveals its quantity.

EXAMPLE Showing changes in the average weekly price ($ per barrel) of Brent crude oil between 2008 and 2018.

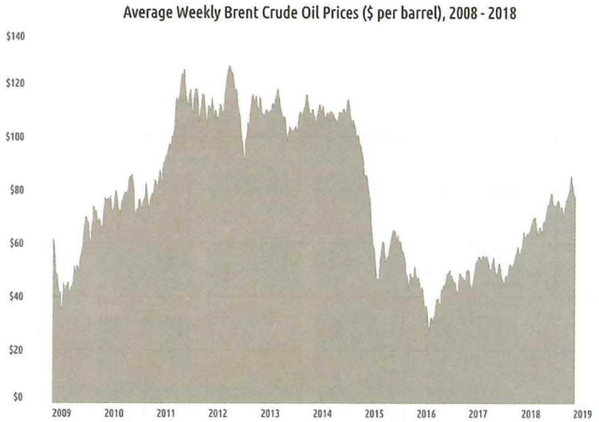


Figure 6.39 Crude Oil Prices for Brent (Dollars per Barrel) 2008–2018

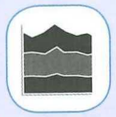
PRESENTATION TIPS

ANNOTATION: Sometimes the point mark is quite pronounced, to aid value judgements and possibly to provide space for a value label, but on most occasions only the position of the connecting lines is displayed. The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. You might choose to annotate specific values of interest (highest, lowest, specific milestones).

COMPOSITION: The chart's dimensions will need to be carefully considered, specifically the aspect ratio formed by its height and width. The upward and downward slopes can seem more significant if the chart width is narrow and less significant if it is more stretched out. There is no single practical rule to obey other than using common sense to ensure you do not overly amplify or underplay features of your data. The sequencing of values tends to follow a chronological left-to-right direction for the time-based x-axis and low values rising up to high values on the y-axis; you will need a good (and clearly annotated) reason to break this convention. Unlike the line chart, the quantitative axis for area charts must have an origin of zero as it is the height of the coloured area under the trend line that is used to perceive the quantitative values.

VARIATIONS & ALTERNATIVES

A variation of the area chart is the 'stacked area chart', which can be used to show how multiple categories form a whole and how this composition changes over time. The stacks may amount to an absolute total or form a 100% proportion view. A 'density plot' appears like an area chart but is used to show the distribution of values across a quantitative axis, rather than a time axis. Another variation is the 'horizon chart', which is based on an area chart but for space-limited contexts. Values that exceed an imposed fixed maximum y-axis range are coloured to indicate different bands of magnitudes, with the extremes usually darker. Like slicing layers off a mountain, each distinct band of values above the maximum y-axis range is chopped off and dropped down to the baseline in front of its foundation base. The final effect shows overlapping layers of increasingly darker colour-shaded areas occupying the same vertical space. An alternative may be simply to consider the 'line chart', especially if you want to compare against several discrete categorical items.



STACKED AREA CHART

ALSO KNOWN AS Area chart, horizon chart

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

A stacked area chart shows how quantitative values have changed over time for multiple categorical items. These charts are typically structured around a continuous temporal x-axis and quantitative y-axis with values plotted using point marks at relevant coordinates. Connecting lines join up adjacent and related items to form slopes which are then extended along the full timescale to display a complete change over time. The connecting lines are typically straight, but sometimes curved line 'interpolation' may be applied to help emphasise a general trend above precise point reading. To accentuate the shape of the trends, the area beneath the line is filled with colour, which means the height of the area at any given point also reveals its quantity. When there are multiple discrete categories, separate stacked areas, sized in height proportionally to their shifting values, are distinguished through distinct stacked regions often coloured to establish their categorical association. The resulting display reveals how a part-to-whole relationship changes over time.

Example
Showing the average trends of bike share usage across the bike share stations of Philadelphia.

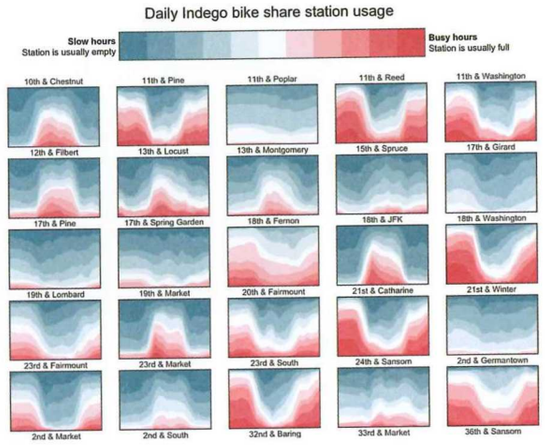


Figure 6.40 Daily Indego Bike Share Station Usage, by Randy Olson (@randal_olson)(www.randalolson.com)

VARIATIONS & ALTERNATIVES

The main variation in stacked area charts will be based on the quantitative values plotted and whether they are representative of an absolute total or a proportional total forming a 100% whole. Rather than stacking categories you might consider using small multiples of single-category area charts, especially as this will display each from a common baseline and therefore make judgements of shape and size a little easier. An alternative may be simply to consider the 'line chart' formed of multiple lines for discrete categorical items.

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful if you have many discrete categorical stacks and wish to enable the user to isolate a certain category of interest, either through filtering to exclude the others or using a contrasting colour to emphasise its shape among the rest. Revealing the quantitative value, time and category label at any point on the chart through a selectable tooltip can also be useful.

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the quantitative values. You might choose to annotate specific values of interest (highest, lowest, specific milestones). Directly labelling the discrete category stacks can be helpful, otherwise use a clear colour legend to explain associations.

COMPOSITION: The chart's dimensions will need to be carefully considered, specifically the aspect ratio formed by its height and width. The upward and downward slopes can seem more significant if the chart width is narrow and less significant if it is more stretched out. There is no single practical rule to obey other than using common sense to ensure you do not overly amplify or underplay features of your data. The sequencing of values tends to follow a chronological left-to-right direction for the time-based x-axis and low values rising up to high values on the y-axis; you will need a good (and clearly annotated) reason to break this convention. Unlike the line chart, the quantitative axis for stacked area charts must have an origin of zero as it is the height of the coloured areas used to perceive the quantitative values. Try to make the sorting of the categorical stacks as meaningful as possible, perhaps placing the most important values on the bottom stack to give it a consistent baseline.

STREAM GRAPH

ALSO KNOWN AS Theme river

C H R **T** S

TRENDS

REPRESENTATION DESCRIPTION

A stream graph shows how quantitative values have changed over time for multiple categorical items. The graphs are generally used when you have many concurrent, constituent categories at any given point in time and these categories may start and stop at different points in time rather than continue throughout the presented time frame. As befitting the name, the appearance of the graphs is characterised by a flowing, organic display of meandering layers. They are structured around a temporal x-axis with quantitative values plotted to quantify height above a local baseline, which is not a stable zero baseline but rather a shifting shape formed out of other category layers. Connecting lines join up adjacent and related items to form slopes which are then extended along the relevant time frame to create a unique categorical layer. The area occupied by this layer is filled with an attribute of colour to represent a further quantitative value scale or to associate with categorical classifications. The stacking arrangement of the multiple categorical layers can shift up and down the implied y-axis dimension, in order to optimise the layout, but not to indicate any notion of positive or negative values.

EXAMPLE
Showing changes in the total domestic gross takings ($US) and the longevity of all movies released between 1986 and 2008.

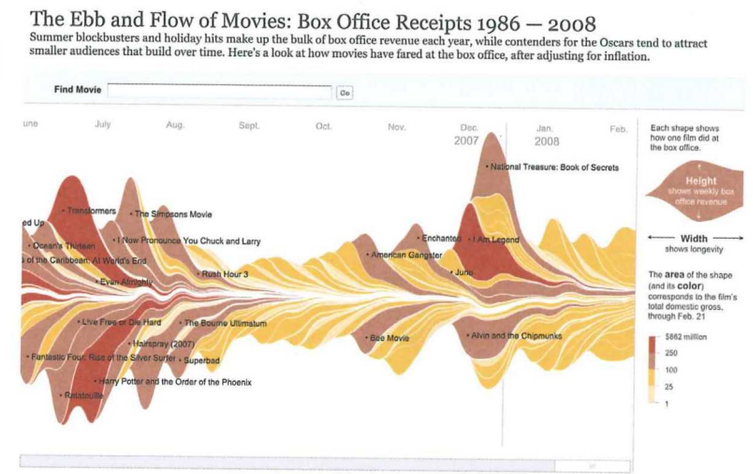


Figure 6.41 The Ebb and Flow of Movies: Box Office Receipts 1986–2008, by Mathew Bloch, Lee Byron, Shan Carter and Amanda Cox (*New York Times*)

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful if you have many discrete categorical layers and wish to enable the user to isolate a certain category of interest, either through filtering to exclude the others or using a contrasting colour to emphasise its shape among the rest. Revealing the quantitative value, time and category label at any point on the chart through a selectable tooltip can also be useful.

ANNOTATION: Chart apparatus devices are generally of limited use in a stream graph with the priority being more on offering a general sense of pattern above precision of value reading. Direct labelling of the discrete categorical layers may be possible, depending on the shape of the data, otherwise use a clear colour legend to explain associations.

VARIATIONS & ALTERNATIVES

If you have relatively few discrete categorical items, you might consider using an alternative chart like the 'stacked area chart' or small multiples of individual 'area charts'. A 'stacked bar chart' would be a consideration, again if there are relatively few categories to include and the quantitative measurements are based on discrete periods (such as totals over a monthly period) rather than a purely continuous series of point-in-time measurements.

GANTT CHART

ALSO KNOWN AS Range chart, floating bar chart, Priestley timeline

C H R **T** S

INTERVALS

REPRESENTATION DESCRIPTION

A Gantt chart displays time-based intervals for different categorical items. The charts are typically structured around a continuous temporal x-axis with a separate row allocated to each major categorical item. Intervals are formed by line marks positioned according to a starting point and sized through length according to a closing point in time. Point marks at each end of this line are sometimes included and presented with discrete symbols or attributes of colour to highlight their distinction. The line may also display an attribute of colour to relate to some categorical status.

EXAMPLE Showing the timeline of all current and former US national parks based on when they were officially established or designated.

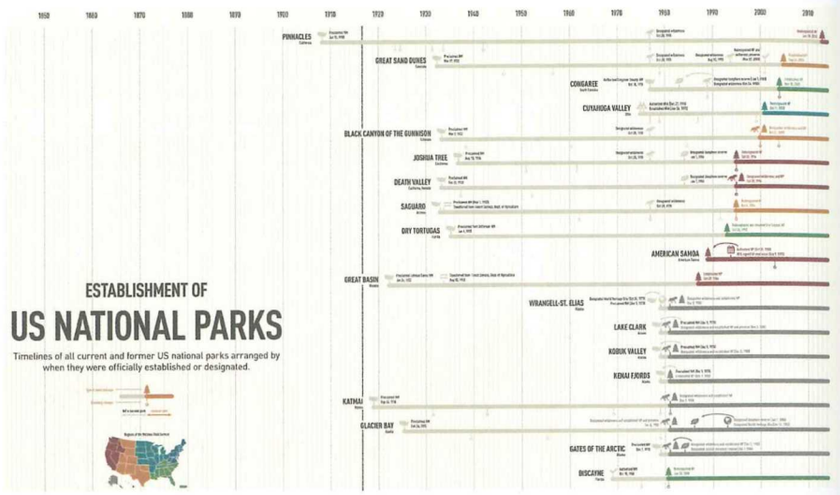


Figure 6.42 Establishment of the U.S. National Parks, by Nicholas Rougeux (www.c82net)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the date values and durations. If you include axis-scale labels you should not need to label each bar value directly, as this will lead to label overload.

COMPOSITION: The bars should be proportionally sized according to the associated duration length – nothing more, nothing less – otherwise the perception of the bar sizes will be distorted. There is no significant difference in perception between vertically or horizontally arranged Gantt charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each category. Landscape layouts with time chronologically sequenced from left to right would be the most common arrangement. Where possible, try to sequence the categorical items in a way that makes for the most logical reading, organised by either the start/finish dates or maybe the durations (depending on which has most relevance).

VARIATIONS & ALTERNATIVES

Gantt charts share many characteristics with the 'connected dot plot', but the measurement dimension here is of time duration rather than quantitative difference. If duration between points in time is less important than individual milestones or events, the 'instance chart' would be worth considering. Sometimes interval lines join up with other adjacent categories, rather than being bound by discrete rows. This might be representative of the merging of activities or the absence of 'discreteness' between activities, and the technique may therefore evolve more towards being a 'connected timeline'.

INSTANCE CHART

ALSO KNOWN AS Dot plot, barcode plot, strip plot

C H R **T** S

ACTIVITIES

REPRESENTATION DESCRIPTION

An instance chart displays time-based events for different categorical items. It is typically structured around a continuous temporal x-axis with a separate row allocated to each major categorical item. Events are represented by point markers, plotted along the timeline, using combinations of symbols and colours to represent different types.

EXAMPLE Showing the instances of different 'Avengers' characters appearing in Marvel's comic book titles between 1963 and 2015.

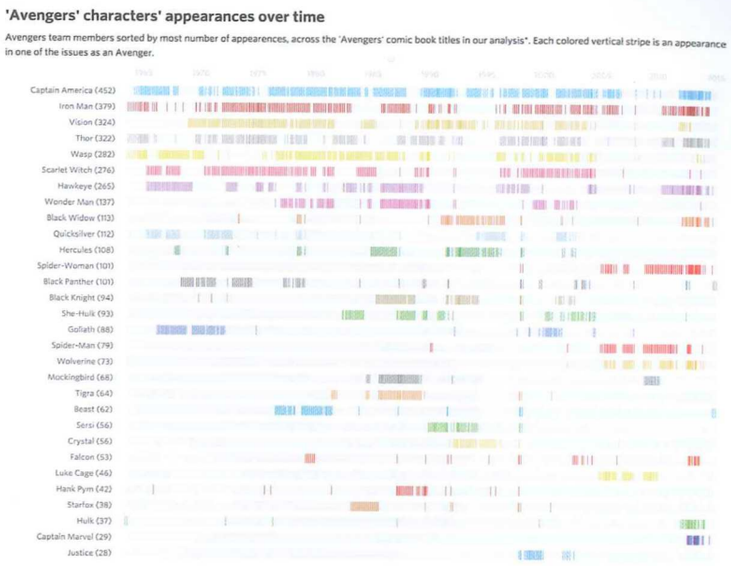


Figure 6.43 How the 'Avengers' Line-up Has Changed Over the Years, by Jon Keegan (*Wall Street Journal*)

PRESENTATION TIPS

ANNOTATION: The inclusion of chart apparatus devices like tick marks and gridlines can help increase the precision of judging the date values. If you include axis-scale labels you should not need to label each bar value directly, as this will lead to label overload. Any colours, symbols or size attributes used must be explained through the inclusion of a legend.

COMPOSITION: There is no significant difference in perception between vertically or horizontally arranged instance charts; it will depend on which layout makes it easier to accommodate the range of values and to read the item labels associated with each category. Landscape layouts with time chronologically sequenced from left to right would be the most common arrangement. Where possible, try to sequence the categorical items in a way that makes for the most logical reading, organised by either the earliest or latest points in time or maybe some measure of quantity (such as which category has the most recorded events).

VARIATIONS & ALTERNATIVES

Instance charts share many characteristics with the 'dot plot' but the measurement dimension here is of time rather than quantitative value. Variations may see the point mark replaced by a geometric shape sized to represent a quantitative measure associated with each event. If the data is more about durations and intervals between events, the 'Gantt chart' will be the best-fit option.

CHOROPLETH MAP

ALSO KNOWN AS Heat map

C H R T **S**

OVERLAYS

REPRESENTATION DESCRIPTION

A choropleth map displays quantitative values for distinct, definable spatial regions. Each region is represented by a defined polygonal shape, with each distinct shape collectively arranged to form the entire landscape. An attribute of colour is used to represent a quantitative measurement. Choropleth maps should only be used when the quantitative measure is directly associated with and continuously relevant across the spatial region on which it will be displayed. If the quantitative measure is related to the consequence of more people living in an area, consider transforming your data by standardising it as per capita or per acre (or other spatial denominator) accordingly.

EXAMPLE Mapping the results of the 2017 German general election showing the winning party for each electoral location.

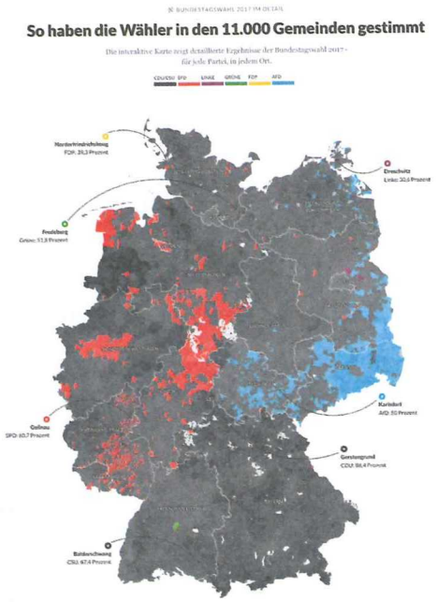


Figure 6.44 How Voters in the 11,000 Municipalities Voted, [Translated] by *Berliner Morgenpost*

VARIATIONS & ALTERNATIVES

Some choropleth maps may be used to indicate categorical association rather than quantitative measurements. Alternative thematic mapping approaches to representing quantitative values might include the 'proportional symbol map', using sized shapes over locations, and the 'dot density map', which plots a representative quantity of dots equally (but randomly) across and within a defined spatial region. 'Dasymetric mapping' is similar in approach to choropleth mapping but breaks the constituent regional areas into much smaller, more specific sub-regions better to represent the realities of the distribution of human and physical phenomena within a given spatial boundary. This might include details of individual buildings, for example.

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful to offer selectable tooltips to view quantitative values and category or location labels for any region on the display.

ANNOTATION: Depending on the shapes of the regions displayed, direct labelling may be limited to just a number of noteworthy values. Any colours used must be explained through the inclusion of a legend. If you choose to include a detailed map image in the background, do not include any unnecessary geographic details that add no value to the spatial orientation or interpretation (e.g. roads, building structures).

COLOUR: The outline colour and stroke width for each spatial area should be distinguishable enough to define the shape but not so prominent as to dominate. Usually, a light-grey or white-coloured stroke will suffice. Sometimes variation in pattern may be included, as well as colour, to represent values that may be uncertain or incomplete. When background map images are included, consider making them semi-transparent or light in colour to avoid competition for attention with the more important data layer.

COMPOSITION: There are many different mapping projections for spatially representing the regions of the world on a plane surface. Be aware that the transformation adjustments made by some of these projections can distort the size of regions of the world, inflating their size relative to other regions, so you will need to pick a projection that is appropriate to the spatial view you are providing.

ISARITHMIC MAP

ALSO KNOWN AS Contour map, isopleth map, isochrone map

C H R T **S**

OVERLAYS

REPRESENTATION DESCRIPTION

An isarithmic map displays distinct spatial surfaces on a map that shares the same quantitative classification. The spatial definition here is not framed by geopolitical boundaries, rather it is organic regions that share a certain quantitative value or interval scale. The regions are formed by interpolated 'isolines' connecting points of similar measurement to form distinct surface areas. Each area is then colour coded to represent the relevant quantitative value.

EXAMPLE Mapping the degree of dialect similarity across the USA.

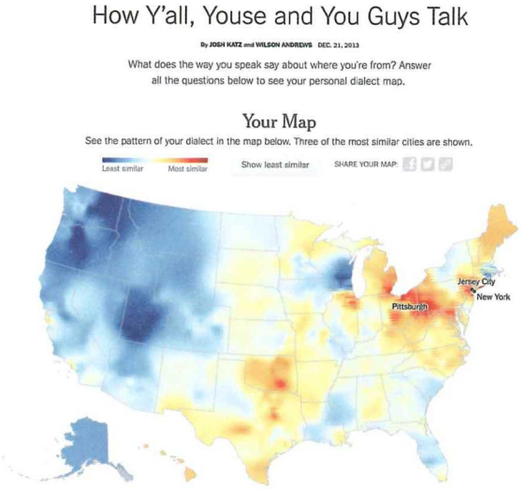


Figure 6.45 How Y'all, Youse and You Guys Talk, by Josh Katz (*New York Times*)

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful to offer selectable tooltips to view quantitative values and category or location labels for any region on the display.

ANNOTATION: Depending on the shapes of the regions displayed, direct labelling may be limited to just a number of noteworthy values. Any colours used must be explained through the inclusion of a legend. If you choose to include a detailed map image in the background, do not include any unnecessary geographic details that add no value to the spatial orientation or interpretation (e.g. roads, building structures).

COLOUR: The outline colour and stroke width for each spatial area should be distinguishable enough to define the shape but not so prominent as to dominate. Usually, a light-grey or white-coloured stroke will suffice. Sometimes variation in pattern may be included, as well as colour, to represent values that may be uncertain or incomplete. When background map images are included, consider making them semi-transparent or light in colour to avoid competition for attention with the more important data layer.

COMPOSITION: There are many different mapping projections for spatially representing the regions of the world on a plane surface. Be aware that the transformation adjustments made by some of these projections can distort the size of regions of the world, inflating their size relative to other regions, so you will need to pick a projection that is appropriate to the spatial view you are providing.

VARIATIONS & ALTERNATIVES

There are specific applications of isarithmic maps used for showing elevation ('contour maps'), atmospheric pressure ('isopleth maps') or travel-time distances ('isochrone maps'). Sometimes you might use isarithmic maps to show a categorical status (perhaps a binary state) instead of a quantitative scale. 'Choropleth maps' will be the method used if your data is organised by bound regions.

PROPORTIONAL SYMBOL MAP

ALSO KNOWN AS Graduated symbol map

C H R T **S**

OVERLAYS

REPRESENTATION DESCRIPTION

A proportional symbol map displays quantitative values for locations on a map. The values are represented via proportionally sized shapes (usually circles), which are positioned with the centre mid-point over a given location coordinate. Colour is sometimes used to introduce further categorical distinction.

EXAMPLE Mapping the origin and size of funds raised across the USA for Democrat candidate Hillary Clinton during the first half of 2015.

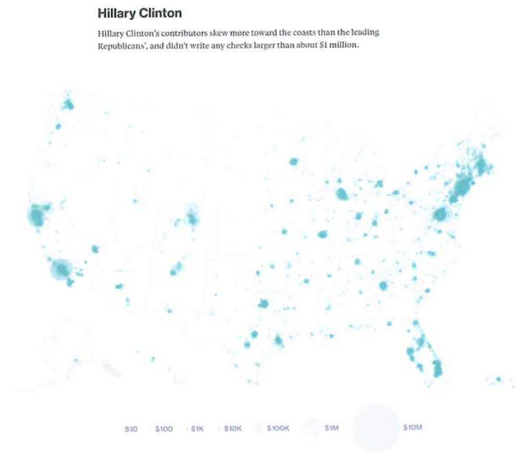


Figure 6.46 Here's Exactly Where the Candidates' Cash Came From, by Zach Mider, Christopher Cannon, and Adam Pearce (Bloomberg Visual Data)

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be especially helpful to offer selectable tooltips to view quantitative values and category or location labels for any region on the display.

ANNOTATION: Depending on the size and overlapping of shapes displayed, direct labelling may be limited to just a number of noteworthy values. Any size scales and colours used must be explained through the inclusion of a legend. If you choose to include a detailed map image in the background, do not include any unnecessary geographic details that add no value to the spatial orientation or interpretation (e.g. roads, building structures).

COLOUR: The outline colour and stroke width for each spatial area should be distinguishable enough to define the shape but not so prominent as to dominate. Usually, a light-grey or white-coloured stroke will suffice. The largest shapes may overlap, in spatial terms, with other nearby locations and sometimes even hide them completely. The use of semi-transparent colours can help avoid the effect of total occlusion. When background map images are included, consider making them semi-transparent or light in colour to avoid competition for attention with the more important data layer.

COMPOSITION: The geometric accuracy of the shape mark size calculation is paramount: it is the area you are modifying, not the diameter/radius. There are many different mapping projections for spatially representing the regions of the world on a plane surface. Be aware that the transformation adjustments made by some of these projections can distort the size of regions of the world, inflating their size relative to other regions, so you will need to pick a projection that is appropriate to the spatial view you are providing.

VARIATIONS & ALTERNATIVES

The main variations usually involve different geometric shapes being used. Alternatives include the 'choropleth map', which colour codes regions, or the 'dot map', which uses dots to represent all items across a spatial region.

PRISM MAP

ALSO KNOWN AS Isometric map, spike map, datascape

C H R T **S**

OVERLAYS

REPRESENTATION DESCRIPTION

A prism map displays quantitative values for locations on a map. The values are represented via proportionally sized lines, appearing as 3D bars, that typically cover a fixed surface area of space and are then sized through height to proportionally represent the quantitative value at each location. Attributes of colour are sometimes used to emphasise large values in particular.

EXAMPLE Mapping the population of trees for each 180 square km of land across the globe.

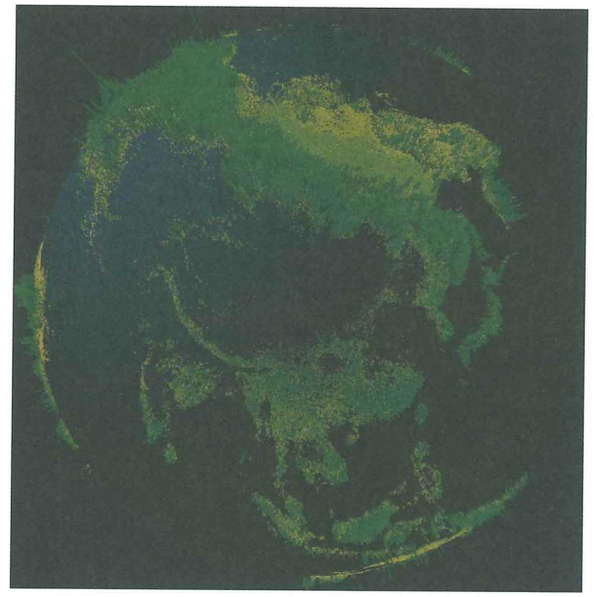


Figure 6.47 Trillions of trees, by Jan Willem Tulp

PRESENTATION TIPS

INTERACTIVITY: Ideally prism maps would be accompanied with interactive features that allow panning around the map region to offer different viewing angles that overcome the perceptual difficulties of judging the 3D presentations of data in a 2D view. Otherwise, smaller values can find themselves hidden behind larger forms, just as small buildings are hidden by skyscrapers in a city.

ANNOTATION: Direct labelling is usually impractical, so the most important feature of annotation is to indicate the size scales used in the map display. If you choose to include a detailed map image in the background, do not include any unnecessary geographic details that add no value to the spatial orientation or interpretation (e.g. roads, building structures).

COLOUR: When background map images are included, consider making them semi-transparent or light in colour to avoid competition for attention with the more important data layer.

VARIATIONS & ALTERNATIVES

Alternatives to the prism map, especially to avoid the 3D form, include the 'proportional symbol map', which uses proportionally sized geometric shapes, and the 'choropleth map', which colour codes regional shapes.

DOT MAP

ALSO KNOWN AS Dot distribution map, pointillist map, location map, dot density map

C H R T **S**

OVERLAYS

REPRESENTATION DESCRIPTION

A dot map displays the distribution of phenomena on a map. It uses point marks to plot data items at specific geographic coordinates. Items might be representative of instances of people, notable sites or incidences. The point marks are usually small circles with attributes of colour used to distinguish categorical classifications. Sometimes a dot represents a one-to-one phenomenon (i.e. a single record at that location) or one-to-many phenomena (i.e. for an aggregated statistic whereby the location represents a logical mid-point), usually depending on the potential relevance and/or sensitivity of directly plotting phenomena at precise locations.

EXAMPLE Mapping each resident of the USA based on the location at which they were counted during the 2010 Census across different ethnicities.

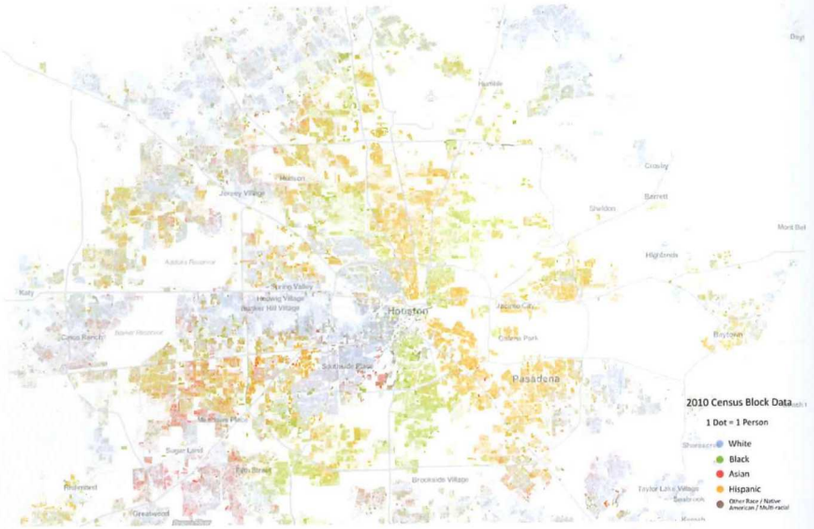


Figure 6.48 The Racial Dot Map: Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia (Dustin A. Cable, creator)

PRESENTATION TIPS

INTERACTIVITY: One method for dealing with viewing high quantities of observations is to provide interactive semantic zoom features, whereby each time a user zooms in by one level of focus, the unit quantity represented by each dot decreases, from a one-to-many towards a one-to-one relationship. Filtering options to exclude or highlight certain selections may also aid the process of understanding.

ANNOTATION: Direct labelling is rarely applied. Clear legends explaining the dot unit scale and any colour associations should ideally be placed as close to the map display as possible. If you choose to include a detailed map image in the background, do not include any unnecessary geographic details that add no value to the spatial orientation or interpretation (e.g. roads, building structures).

COLOUR: If colours are being used to distinguish the different categories, ensure these are as visibly different as possible. When background map images are included, consider making them semi-transparent or light in colour to avoid competition for attention with the more important data layer.

COMPOSITION: Dot maps should be displayed using an equal-area projection, as the precision of the plotted locations is usually paramount. From a readability perspective, try to find a balance between making the size of the dots small enough to preserve their individuality but not too tiny as to be indecipherable.

VARIATIONS & ALTERNATIVES

A 'dot density map' is a variation that involves plotting a representative quantity of dots equally (but randomly) across and within a defined spatial region. The position of individual dots is therefore not to be read as indicative of precise locations but used to form a measure of quantitative density. This offers a useful alternative to the choropleth map, especially when categorical separation of the dots through colour is of value.

FLOW MAP

ALSO KNOWN AS Connection map, route map, stream map, particle flow map

C H R T **S**

OVERLAYS

REPRESENTATION DESCRIPTION

A flow map shows the characteristics of movement or connections between phenomena across spatial regions. There is no fixed recipe for a flow map, but it generally displays characteristics of origin and destination (positions on a map), route (using organic or vector paths), direction (using arrow or tapered line width), categorical classification (colour) and quantitative measurement (line weight or, if animated, motion speed).

EXAMPLE Mapping the average number of vehicles using Hong Kong's main network of roads during 2011.

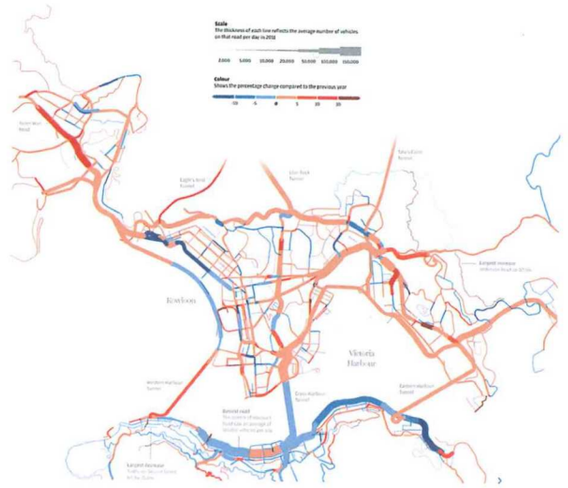


Figure 6.49 Arteries of the City, by Simon Scarr (*South China Morning Post*)

PRESENTATION TIPS

INTERACTIVITY: Animated sequences may provide a useful presentation method when the phenomena are characteristic of some notion of movement.

ANNOTATION: Annotation needs will be unique to each approach and the inherent complexity or otherwise of the display. Often the general patterns may offer the sufficient level of readability without the need for imposing amounts of value labels, but clear legends explaining the associations with any attributes used will be important to include. If you choose to include a detailed map image in the background, only include any relevant geographic details that offer spatial orientation or interpretation to the nature of flow being represented (e.g. roads, rivers, oceans).

COLOUR: If colours are being used to distinguish the different categories, ensure these are as visibly different as possible. When background map images are included, consider making them semi-transparent or light in colour to avoid competition for attention with the more important data layer.

COMPOSITION: Some degree of geographic distortion or smoothing of flow routes may be required. Decisions about the degree of interpolation applied to line smoothing or the merging of relatively similar pathways may be entirely legitimate, but ensure that this is made clear to the viewer. There are many different mapping projections for spatially representing the regions of the world on a plane surface. Be aware that the transformation adjustments made by some of these projections can distort the size of regions of the world, inflating their size relative to other regions, so you will need to pick a projection that is appropriate to the spatial view you are providing.

VARIATIONS & ALTERNATIVES

There are several variations for how you might label different applications of displaying flow. It generally depends on whether you are showing point A to point B journeys ('connection maps'), more intricate pathways ('route maps') or organic phenomena ('particle flow maps').

AREA CARTOGRAM

ALSO KNOWN AS Contiguous cartogram, density-equalising map

DISTORTIONS

REPRESENTATION DESCRIPTION

An area cartogram displays the quantitative values associated with distinct, definable spatial regions on a map. Each geographic region is represented by a polygonal area based on its outline shape with the collective regional shapes forming the entire landscape. Quantitative values are represented by proportionately distorting (inflating or deflating) the relative size of and, to some degree, shape of the respective regional areas. Traditionally, area cartograms strictly aim to preserve the neighbourhood relationships between different regions. Attributes of colour are often used to represent the quantitative measurements and/or to associate the region with a categorical classification. Area cartograms require the reader to be relatively familiar with the original size and shape of regions in order to be able to establish the degree of relative change in their proportions.

EXAMPLE Mapping the measures of climate change responsibility compared with vulnerability across all countries.

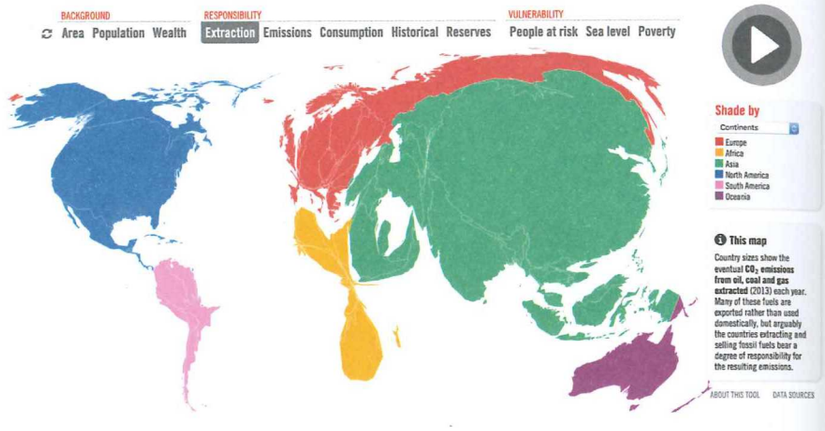


Figure 6.50 The Carbon Map, by Duncan Clark and Robin Houston (Kiln)

PRESENTATION TIPS

INTERACTIVITY: Animated sequences enabled through interactive controls can help to better identify instances and degrees of change, but usually only over a small set of regions and only if the change is relatively smooth and sustained. Manual animation will help to provide more control over the experience. Selectable tooltips to view quantitative values and category or location labels for any region on the display may also prove useful.

ANNOTATION: Directly labelling the regional areas with geographic details and the value they hold is likely to lead to too much clutter. As it is difficult to assess the degree of distortion and, indeed, often to identify the regions themselves, it can be useful to present a thumbnail view of the undistorted original geographic layout to help readers orient themselves with the changes. Additionally, a limited number of regional labels might be included to provide direct spatial context and orientation. Any colours used must be explained through the inclusion of a legend.

COLOUR: The outline colour and stroke width for each spatial area should be distinguishable enough to define the shape but not so prominent as to dominate. Usually, a light-grey or white-coloured stroke will suffice.

VARIATIONS & ALTERNATIVES

Unlike contiguous cartograms, non-contiguous cartograms tend to preserve the shape of the individual polygons but modify the size and the neighbouring connectivity to other adjacent regional polygon areas. The best alternative ways of showing similar data would be to consider using the 'choropleth map' or 'Dorling cartogram'.

DORLING CARTOGRAM

ALSO KNOWN AS Demers cartogram

DISTORTIONS

REPRESENTATION DESCRIPTION

A Dorling cartogram displays the quantitative values associated with distinct, definable spatial regions on a map. Each geographic region is represented by a circular mark which is proportionally sized to represent a quantitative value. The placement of each circle loosely resembles the region's geographic location with general preservation of neighbourhood relationships between adjacent shapes. Attributes of colour hue are often used to associate each spatial region with a categorical classification.

EXAMPLE Mapping the share of people using the Internet by country as at 2015.

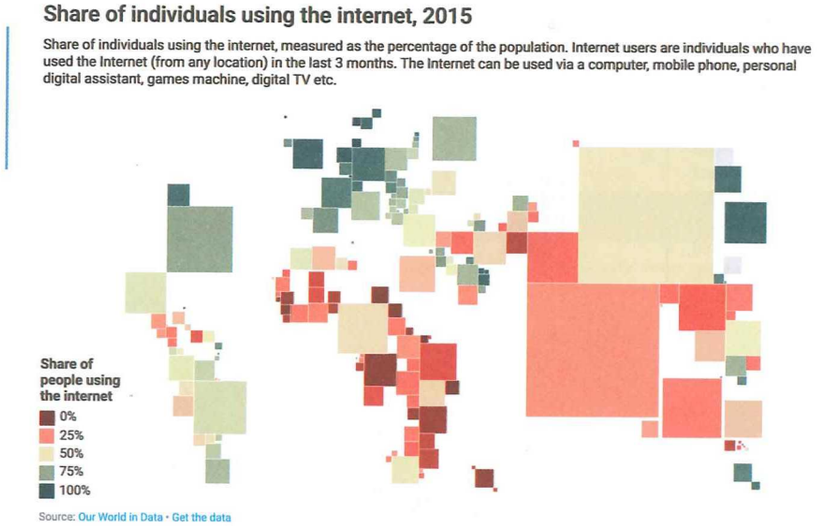


Figure 6.51 Share of Individuals Using the Internet, 2015, by Lisa Rost

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be helpful to offer selectable tooltips to view quantitative values and category or location labels for any region on the display.

ANNOTATION: Directly labelling the shapes with geographic details and the values they hold is common, though you might restrict this to only circles that are of sufficient size to hold such annotations. Any colours used must be explained through the inclusion of a legend.

COMPOSITION: Preserving the layout adjacency with neighbouring regions is important. Dorling cartograms tend not to allow circles to overlap or occlude, so some accommodation of large values might result in some location distortion.

VARIATIONS & ALTERNATIVES

A variation on the approach, called the 'Demers cartogram', involves the use of rectangular marks instead of circles. This offers an alternative way of connecting adjacent shapes. Other alternative chart types to consider would be the 'area cartogram' or the 'choropleth map'.

GRID MAP

ALSO KNOWN AS Cartogram, bin map, equal-area cartogram, hexagon bin map

C H R T S

DISTORTIONS

REPRESENTATION DESCRIPTION

A grid map displays the quantitative values associated with distinct, definable spatial regions on a map. Each geographic region (or a statistically consistent interval of space, known as a 'bin') is represented by a fixed-size uniform shape, sometimes termed a **tile**. The marks used tend to be squares or hexagons, though any tessellating shape might help to arrange all regional tiles into a collective shape that roughly fits the real-world geographic adjacency. Attributes of colour are applied to each regional tile either to represent a quantitative measurement or to associate the region with a categorical classification.

EXAMPLE Showing the percentage of people voting to leave and remain across the UK electoral seats during the EU referendum in 2016.

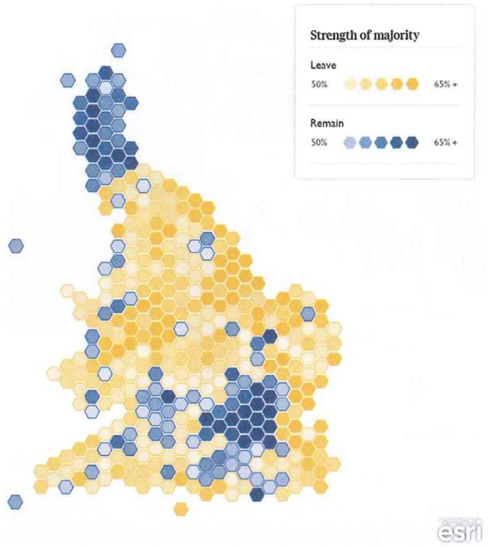


Figure 6.52 Share of People Voting to Leave and Remain During the EU Referendum in 2016, by Ben Flanagan

VARIATIONS & ALTERNATIVES

'Hexagon bin maps' are specific deployments of the grid map that offer a layout formed by a high resolution of smaller hexagons to preserve localised details.

PRESENTATION TIPS

INTERACTIVITY: Interactivity may be helpful to offer selectable tooltips to view quantitative values and category or location labels for any region on the display.

ANNOTATION: Directly labelling the shapes with geographic details is usually impractical due to the small size of each point mark, unless short abbreviated values can suitably represent the location label. Legends explaining the colour associations must be included.

COMPOSITION: The main composition challenge is to determine the right geographic level for each constituent tile to be representative of, and to optimise, the best-fit collective layout that preserves as many neighbouring relationships as possible.

6.2 Influencing Factors and Considerations

You have now been through the gallery of chart-type options learning more about their specific roles and what design features may enhance their particular deployment. Even if you have a fairly clear idea about which chart(s) you might choose, there are other factors that may influence your final decision of how to represent your data. There is a blend of considerations to draw from your progress through the first three preparatory stages of the design process, supplemented by the enduring need to satisfy the three principles of good visualisation design, as presented in Chapter 2.

Technological: What charts you can actually make and how easily you can personally create them is a big factor. Data visualisation technologies offer different chart-making capabilities and it can be hard navigating through the options that exist. To assist with this, you might consider consulting the 'Chartmaker Directory' (chartmaker.visualisingdata.com/). This digital resource organises a huge catalogue of useful references that will offer an answer to the most common of questions: 'Which tool do you need to make that chart?'

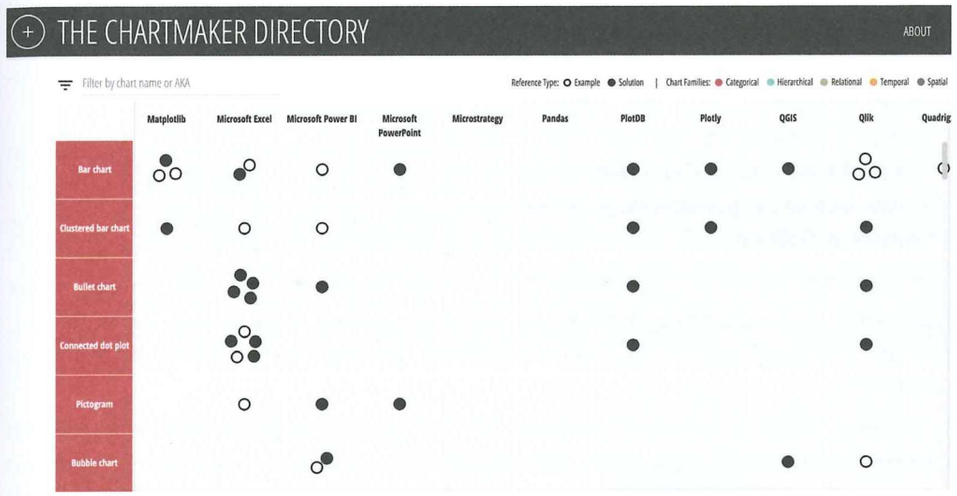


Figure 6.53 Screenshot of the 'Chartmaker Directory'

The directory's content is presented through a tabular layout (Figure 6.53). Across the top of the table are a selection of around 40 chart-making tools. A comprehensive list of different chart types is presented down the side matching the gallery you have just explored. Inside the intersecting cells, you will find unfilled and filled circular markers representing a reference in the directory:

- An *unfilled* mark represents a link to an example, providing evidence that a given chart can be made in a given tool. Read it as 'here's a link to a bar chart made using Excel', for example.

- A *filled* mark represents a link to a solution, which provides guidance on how to create a given chart with a given tool. Solutions might exist as 'how-to' tutorials with step-by-step instructions, video demonstrations, downloadable workbooks/templates or reusable code.

The directory is constantly growing as more chart-making solutions and examples are discovered for each tool. In particular, many valuable references present smart workarounds or 'out of the box' thinking that employ unconventional techniques to create a chart in a tool that normally would not seem possible.

'The capability to cope with the technological dimension is a key attribute of successful students: coding – more as a logic and a mindset than a technical task – is becoming a very important asset for designers who want to work in Data Visualisation. It doesn't necessarily mean that you need to be able to code to find a job, but it helps a lot in the design process. The profile in the (near) future will be a hybrid one, mixing competences, skills and approaches currently separated into disciplinary silos.'
Paolo Ciuccarelli, discussing students on his Communication Design Master Programme at Politecnico di Milano

Purpose: Having the technical ability to create a broad repertoire of chart types is the *vocabulary* of this discipline; judging when to use them is the *literacy*. The first question to ask yourself is if you even need to represent your data in chart form. Will this enable new qualities and relationships in your data to be seen? Do not rule out the value of a table if providing a means for your viewer to look up and reference values. It might be a more suitable solution option.

This brings us back to the discussion about the importance of defining the tone of your project. Are you aiming to facilitate the reading of data or the feeling of data? Is it more important to offer precise value judgements or should more emphasis be placed on general sense-making about the big, medium and small values? Were there emotional qualities you wanted to emphasise or suppress?

In his book *Semiology Graphique*, published in 1967, Jacques Bertin proposed the idea that different ways of encoding data might offer varying degrees of accuracy in the perception of data values. In 1984, William Cleveland and Robert McGill published a seminal paper, 'Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods'. This offered more empirical evidence of Bertin's thoughts. From this study they developed a general ranking that explained which attributes used to encode quantitative values would facilitate the highest degree of perceptual accuracy. In 1986, Jock Mackinlay's paper, 'Automating the Design of Graphical Presentations of Relational Information', further extended this to include proposed rankings for encoding categorical (nominal and ordinal) data, as well as quantitative values. The table shown in Figure 6.54 presents the 'Ranking of Perceptual Tasks'.

What this ancestry of studies reveals is that the use of certain attributes to encode certain types of data may make it quicker, easier and more accurate to judge the values portrayed. Two classic illustrations of this notion are shown below. Looking at Figure 6.55, if A is 10, how big is B?

Qualitative Nominal	Qualitative Ordinal	Quantitative Interval, Ratio
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Containment	Pattern (Texture)	Size (Volume)
Pattern (Density)	Connection	Pattern (Density)
Colour (Lightness)	Containment	Colour (Lightness)
Symbol	Size (Length)	Colour (Hue)
Size (Length)	Angle	Pattern (Texture)
Angle	Size (Area)	Connection
Size (Area)	Size (Volume)	Containment
Size (Volume)	Symbol	Symbol

Note that the attribute of 'Motion' was not included in this study. For the purposes of this display, 'Angle' and 'Slope' are combined whereas they were distinguished as separate in the study.

Figure 6.54 The Ranking of Perceptual Tasks, adapted from Mackinlay (1986)

In both cases the answer is B equals 5. Although B in the bar chart being of size 5 feels about right, the idea that circle B is also 5 feels less so. Our visual system is superior in its accuracy when performing relative judgements for a line, in comparison with a shape. This is explained by the fact that judging the variation in size of lines involves detecting change in a linear dimension (length), whereas the variation in size of a geometric shape like a circle happens across a quadratic dimension (area). If you look at the rankings in Figure 6.54 in the 'Quantitative' column, you will see the encoding attribute of *Length* is ranked higher than the attribute of *Area*.

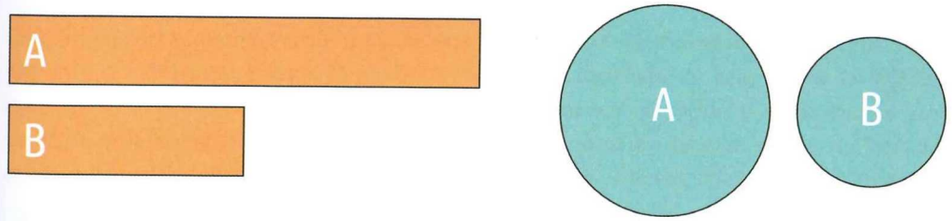


Figure 6.55 Comparison of Judging Line Size vs Area Size

Now let's consider a demonstration of perceptual accuracy when using different dimensions of colour variation to represent nominal values. In the charts shown in Figure 6.56 you can see that different attributes are used to represent the categorical groupings in the two scatter plots.

On the left you see variation in the attribute of colour hue (blue, orange and green) to classify the distinct categories; on the right you see the attribute of symbol (diamond, circle and square) applied similarly.

What you will find is a more immediate, effortless and accurate experience in identifying the groupings of the coloured category markers compared with the symbol-based equivalents. It is easier to observe classifications through variation in colour than it is using variation in symbol, as supported by *colour hue* being ranked higher than *shape* for nominal data types as shown in the table in Figure 6.54.

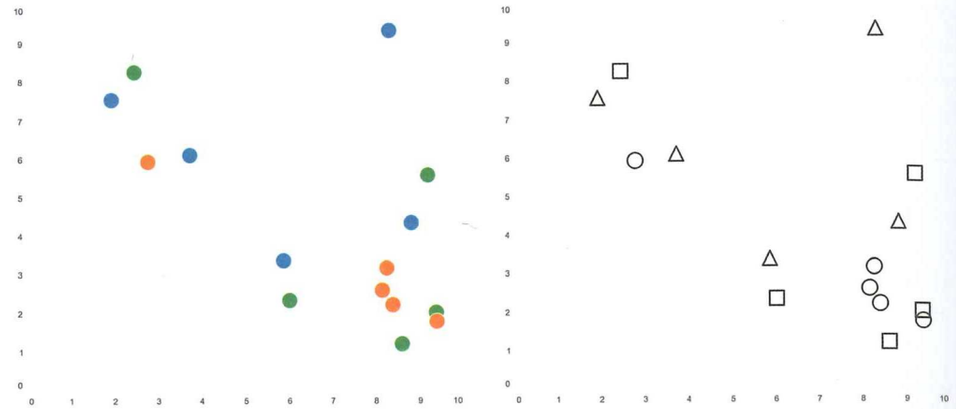


Figure 6.56 Comparison of Judging Categorical Associations Using Variation in Hue vs Variation in Shape

You can see from these simple demonstrations that there are clearly ways of encoding data that will make it easier to read values accurately and efficiently. However, as Cleveland and McGill stress in their paper, this should only be taken as guidance, commenting that the ranking of attributes 'does not result in a precise prescription for displaying data but rather is a framework within which to work'.

This is important to acknowledge because you have to weigh up whether precise perceiving is actually what you wish to offer your viewers. As stated in Chapter 3, sometimes getting the 'gist' of data values is sufficient. You might therefore determine that selecting a chart that uses the attribute of size through variation in area, which is lower down the quantitative attribute rankings, offers a suitable balance. Judging the hierarchy of large, medium and small features may be sufficient for your needs. It depends on your purpose.

Sometimes, you will have scope in your encoding choices to incorporate a certain amount of visual immediacy in accordance with your topic. I warned earlier about the need to be driven by your data and not by your ideas, but sometimes there is scope to squeeze out extra stylistic associations between the visual and the content. The flowers of the Better Life Index feel consistent in metaphor with the idea of better life: the more in bloom the flowers, the more colourful and prouder each petal appears and the better the quality of life in that country.

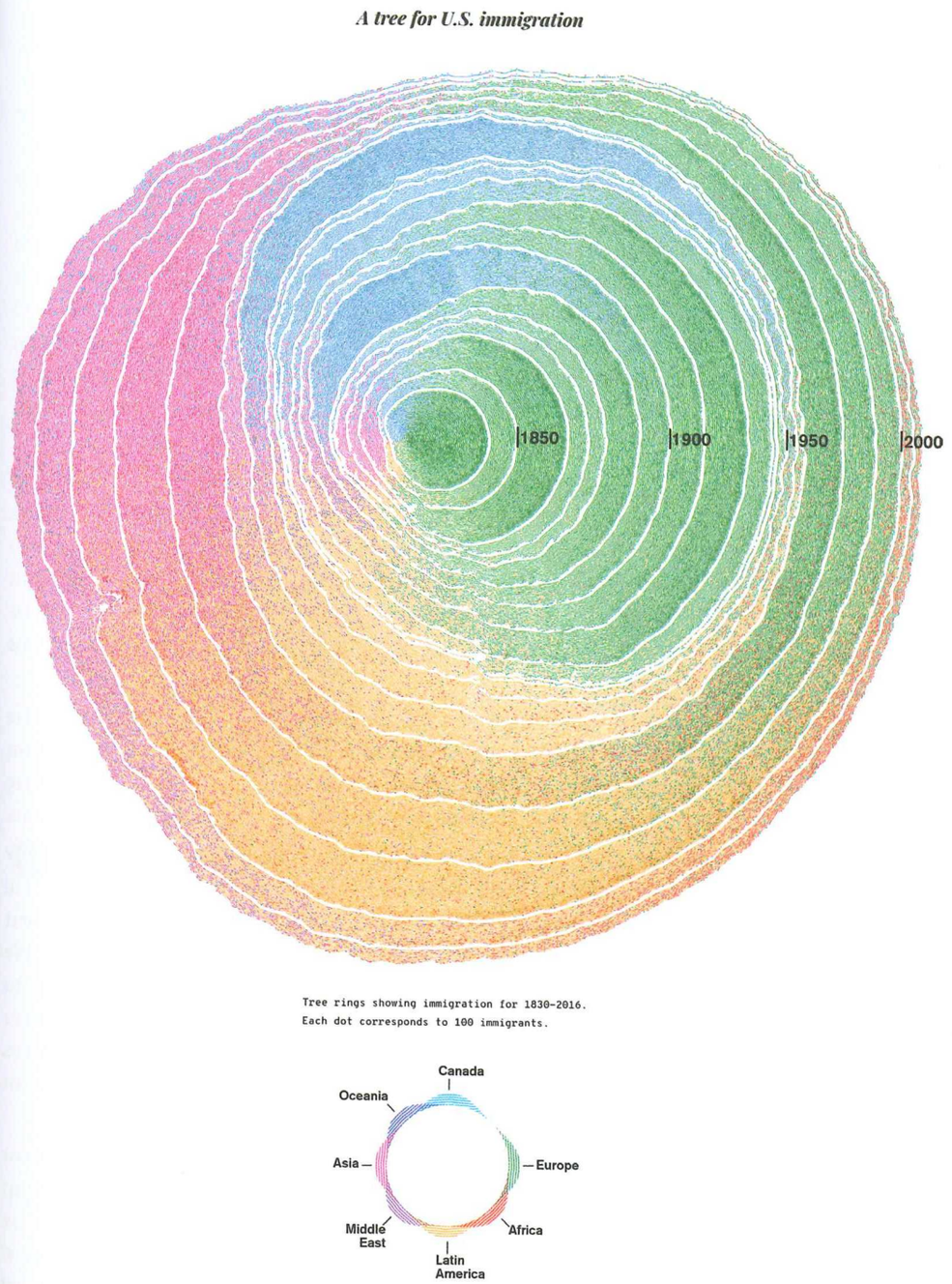


Figure 6.57 Simulated Dendrochronology of U.S. Immigration, by Pedro Cruz, John Wihbey, Avni Ghael and Felipe Shibuya

> 'I've come to believe that pure beautiful visual works are somehow relevant in everyday life, because they can become a trigger to get people curious to explore the contents these visuals convey. I like the idea of making people say "oh that's beautiful! I want to know what this is about!" I think that probably (or, at least, lots of people pointed that out to us) being Italians plays its role on this idea of "making things not only functional but beautiful".' **Giorgia Lupi, Co-founder and Design Director at Accurat**

What you are trying to represent may not be possible using a conventional chart. Another example that draws from nature is shown in Figure 6.57. This piece uses the notion of dendrochronology – the study of tree rings – to create a compelling portrayal of the history of immigration into the USA. Each ring is a decade container, working outwards through chronological decades. Within each container ring are dots corresponding to 100 immigrants. The colours indicate the origin continents or major regions. The outcome is a stunning concept that perfectly aligns subject matter and visual encoding.

Data type and shape: The types of data and range of values you are trying to display will have a bearing on which charts you can use, and, of those, which will best portray what you want to say. Any chart type will only accommodate certain types of data. For example, if you want to use a line chart, you will need one or more continuous series of quantitative values that have a dimension of temporal data. Additionally, the viability of any chart choice will be determined by how well it accommodates the range of values you wish to include. As ever, this depends on what it is you want to say.

Let's suppose you are producing some simple analysis about the market share of browsers. The first chart you consider is the pie chart. To use this you will need quantitative values, in the form of percentages, for different categories that aggregate to a true 'whole' (nothing more, nothing less, than 100%). The data shows there is a market share breakdown across 30 discrete browsers.

As you can see in Figure 6.58, there are a few issues with the pie chart (A). If it is important for a viewer to judge values for each of the 30 browsers with a certain degree of accuracy, the pie chart will not be fit for purpose. It gets harder to perceive the size of each slice after the first three or four. Furthermore, the colour associations as shown in the legend are indiscernible. We need a plan B. In this case you might switch to a bar chart. Even though this chart belongs to a different 'family' you can still use it to represent parts-of-a-whole percentages for each browser item. This will offer an improved option to make it more readable, both in judging the values and through the proximity of the category labels to each bar. It does, though, result in a lot of empty space due to the skewed shape of the data values.

If you are really seeking to enable the readability of each value, you may try to convey just how dominant Chrome is as one part of this whole. You might therefore revert to using a pie chart (C) to include all the discrete browser categories, but label only the Chrome part and summarise the rest as a single 'All others' value. You only need the Chrome value to be seen as 'biggest' compared with the many other competitors battling for but losing out on the dominant market share. If the visibility of the many other parts is not important, group them into a single 'All others' value so now you have a simple two-slice pie (D) or a donut chart (E) if you wish to exploit the empty centre to accommodate the summary value labels.

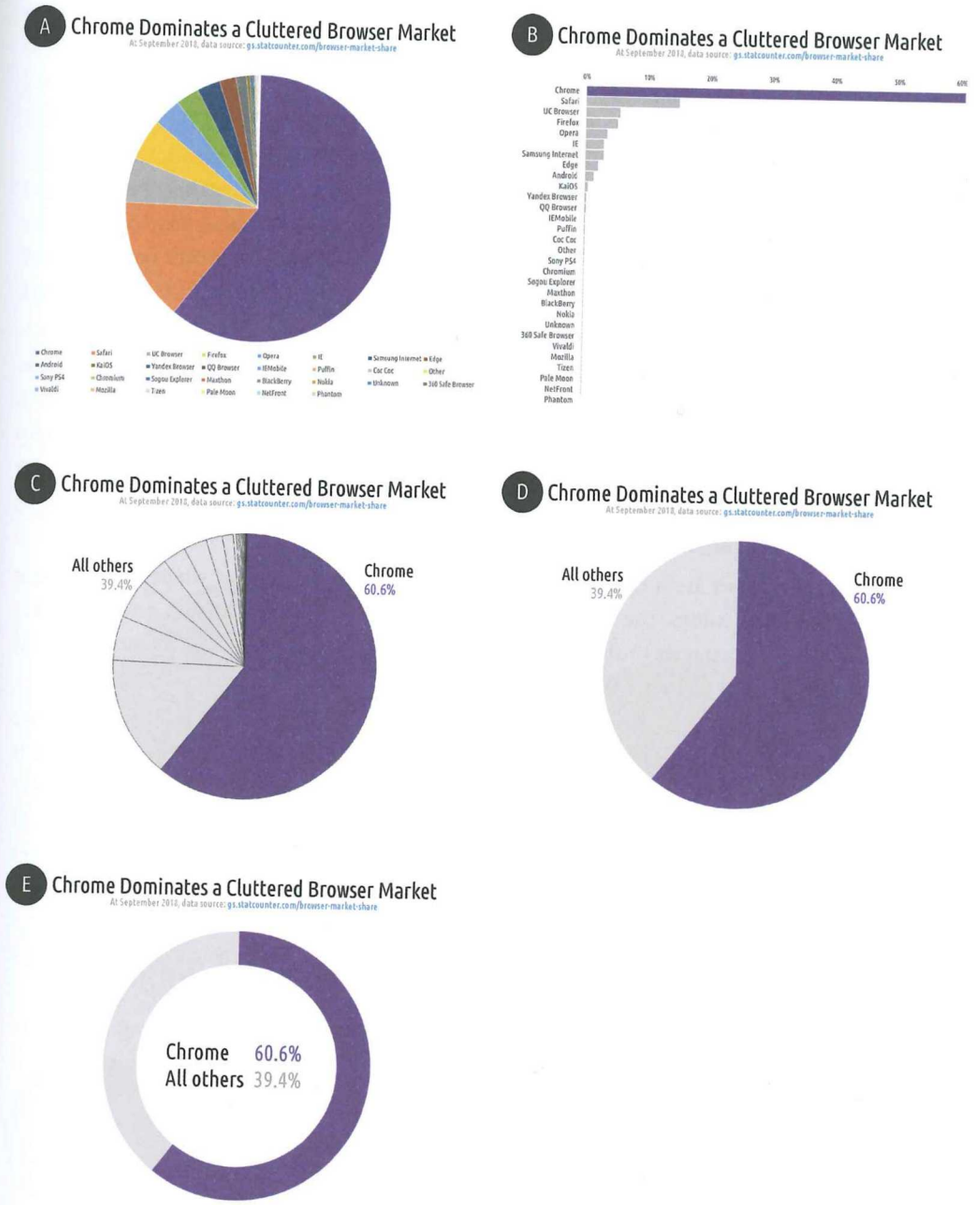


Figure 6.58 Iterations of Different Chart Options to Show the Same Data

This illustration demonstrates how you only know if a chart will serve your purpose once you try it out with real data. After that, consider variations in chart and/or transformations of your data to find the best way to show what you really want to say.

Data exploration: One consistently useful pointer to how you might visually communicate your data is to consider which techniques helped *you* to unearth key insights when you were

visually exploring the data. What chart types have you already tried out and maybe found to reveal interesting patterns? Exploratory data analysis, in many ways, offers this bridge to visual communication: the charts you use to see data for yourself often represent prototype thinking about how you might communicate real data to others. The way you style the chart may differ, but if a method is already working, why not utilise the same approach again?

Editorial angle: When defining your editorial angle(s) you are expressing what specific aspect of understanding you are attempting to portray to your viewers. This helps you to determine which chart type might be most relevant or at least which family across the CHRTS taxonomy will provide the best option to pick from. Always give yourself time to spend on the editorial stage, carefully articulating *what* you want to say before you get too carried away with picking *how*.

Trustworthy design: In the discussion about tone I explained how you might sacrifice precision in the perception of values to suit the purpose of your work. Precision in perception is one thing, but precision in design is a different matter and one for which there should be no compromise. Being accurate in your portrayal of data is a fundamental obligation. There are many ways in which viewers can be deceived through incorrect and inappropriate encoding choices, whether they are intended or not.

Geometric miscalculations are a common mistake. When using the area of shapes to represent different quantitative values, the underlying geometry needs to be calculated accurately. For example, using circular shapes to show a quantitative value of 20 compared with another of 10, you would just half the diameter of the second, right? Wrong.

The illustration in Figure 6.59 shows the incorrect and correct ways of encoding two quantitative values through circle size, where value A is twice the size of B. The orange circle for B has half the *diameter* of A, the green circle for B has half the *area* of A. Using variation in diameter distorts the perceived size of circle B as being far smaller than the value actually is. Viewers base estimates of quantitative size through the area of a circle, not its diameter. Therefore, the green circles demonstrate the correct way to encode these values.

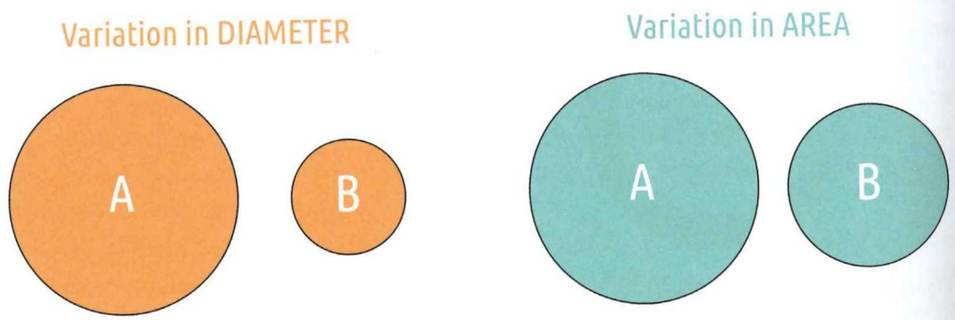


Figure 6.59 The Correct and Incorrect Way to Encode Variation in Shape Size

Another representation accuracy issue causing problems for size judgements concerns truncated axis scales. When quantitative values are encoded through the height or length of size (e.g. for bar charts), truncating the value axis (not starting the range of quantitative values from the

origin of zero) distorts the size judgements. I will revisit this issue in Chapter 10 because it is ultimately a consideration about the sizing of chart-scale ranges, which I deem to be a matter of composition.

Another design issue that can distort data is 3D decoration. In the majority of cases, the use of 3D charts is, at best, unnecessary and, at worst, hugely distorting. Though I concede that there can be a certain appeal to the physical appearance of 3D charts, it is not an effective choice for trustworthy practices. It is often seen applied to a chart when the visualiser is motivated by a desire to demonstrate technical competence with a tool or encouraged by stakeholders who want to see charts made to look 'fancy' or 'cool'.

Using psuedo-3D decoration, when you have only two dimensions of data, is gratuitous and will distort the viewer's ability to judge values with any degree of acceptable accuracy. As illustrated in Figure 6.60, when forming value estimates of the angles and sectors in the respective pie charts, the 3D version makes it much harder to form accurate judgements. The tilting of the isometric plane amplifies the front part of the chart and diminishes the back. It also introduces a raised 'step' which is purely decorative, thus embellishing the judgement of the sector sizes.

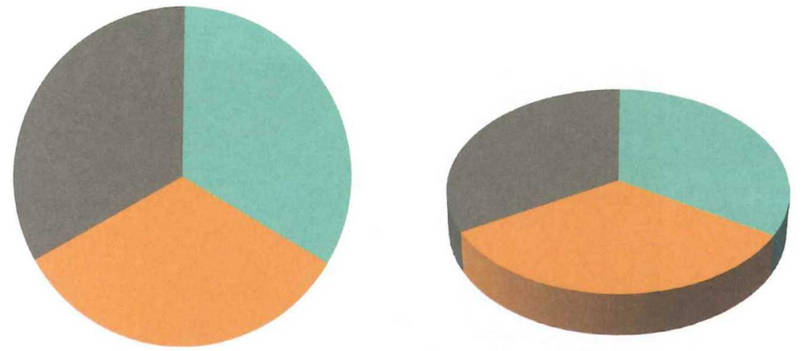


Figure 6.60 Illustrating the Distortions Created by 3D Decoration

For charts genuinely based on three dimensions of data, a 3D representation should only be considered reasonable if the viewer is provided with the means to adjust the field of view. This will help to overcome the distortion of distance and perspective, creating multiple potential 2D viewing angles. 'All the Buildings in Manhattan', Figure 6.61, offers a slick interactive experience that lets users navigate around a 3D view of New York City to observe the size of the buildings around Manhattan. This means you can change your field of view to determine properly the height of the modelled building shapes and make comparisons across the city.

Another legitimate application of 3D visualisation is through the potential of physical displays, perhaps using 3D printing techniques, as demonstrated by the piece shown in Figure 6.62. This portrays trajectories for every home run scored by Kris Bryant of the Chicago Cubs during 2017, including the height, distance and landing position of each shot.

The final matter related to trustworthiness concerns thematic mapping, specifically the often contentious matter of choosing a map projection. The Earth is not flat. Although advances

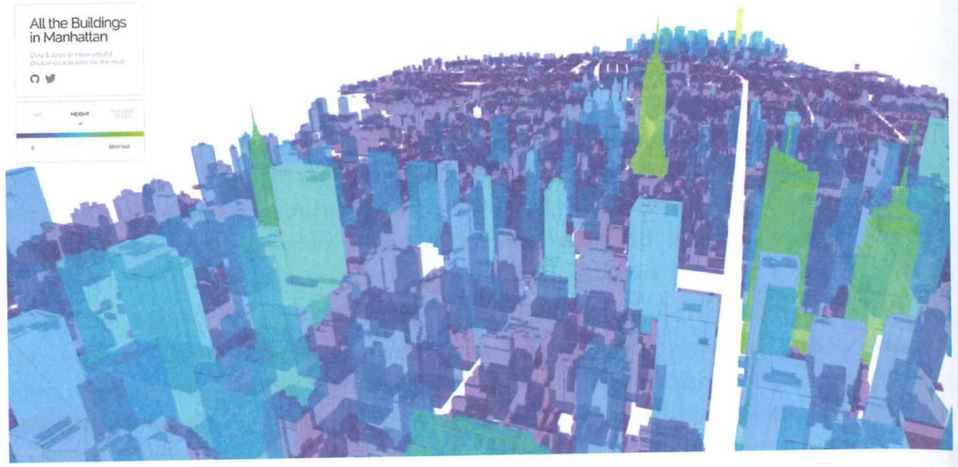


Figure 6.61 All the Buildings in Manhattan, by Taylor Baldwin (tbaldwin.net, @taylorbaldwin)

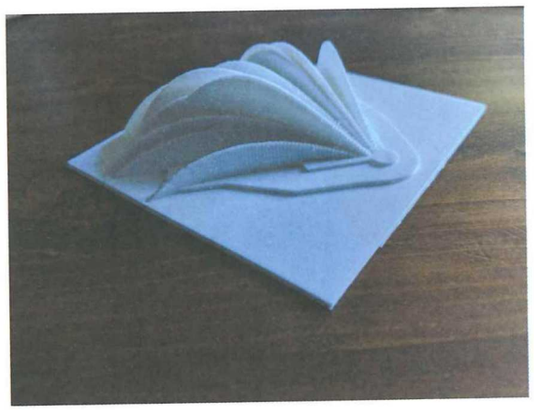


Figure 6.62 Representing Three Dimensions of Data (Baseball Home Run Trajectories) in a 3D Space

in technology are enabling interaction with 3D portrayals of the Earth within a 2D space, the dominant form through which maps are presented portrays the Earth as a flat surface. Features such as size, shape and distance can be measured accurately on Earth, but when projected onto a flat surface a compromise has to occur. Only some of these qualities can be preserved and represented accurately. Although there are exceptionally complicated calculations attached to each spatial projection, the main features most of us need to know about are that:

- every type of map projection has some sort of distortion;
- the larger the area of the Earth portrayed as a flat map, the greater the distortion;
- there is no single right answer – it is often about choosing the least-worst case.

Thematic mapping (as opposed to mapping spatially for navigation or reference purposes) is generally best carried out using mapping projections based on 'equal-area' calculations (so the sacrifice is more on the shape, not the size). This ensures that the phenomena per unit – the values you are typically plotting – are correctly represented by proportion of regional area. For choosing the best specific projection, in the absence of perfect, the decision is usually based on which one will distort the spatial truth the least given the level of mapping required. There are many variables in play, however, based on the scope of view (world, continent or country/sub-region), the potential distance from the equator of your region of focus and whether you

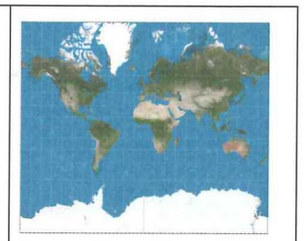
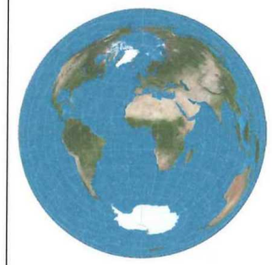
Mercator While the Mercator has been widely discredited in its role as a means of portraying the world (due to the vast distortions at the poles) it is still the most common projection found in mapping tools (where it is often termed Web Mercator). This is largely because of its rectangular dimensions that support seamless zooming. If you are determined to use this projection, you should not use it for a global view; stick to a lower regional level so the distortions are minimised, especially for regions around the equator.	
Equal Earth The Equal Earth map projection is an equal-area pseudo-cylindrical projection for world maps. It was developed in order to create a world map showing continents and countries at their true sizes relative to each other.	
Lambert Azimuthal Equal-area This spherical projection is most commonly recommended for hemisphere- or continent-level views. The European Environment Agency, for example, recommends its usage for any European mapping purpose.	
Winkel–Tripel Most of the important people who are far better informed about mapping projections than I often describe the Winkel–Tripel projection as being one of the best choices for viewing the world. Indeed, it represents the modern standard world map adopted by National Geographic.	
Mollweide In contrast to the Winkel–Tripel, the Mollweide (equal-area) projection offers greater emphasis on the accuracy of ocean areas and can be useful for atmospheric mapping (e.g. flight paths).	

Figure 6.63 A Selection of Commonly Deployed Mapping Projections. Images from Wikimedia Commons published under the Creative Commons Attribution-Share Alike 3.0 Unported Licence

are focusing on land, sea or sky (atmosphere), to name but a few. As with many other topics in this field, a discussion of mapping projections requires a dedicated text, but let me at least offer a brief outline of five different projections (Figure 6.63).

Summary: Data Representation

Visual Encoding and Charts

This chapter introduced the act of visual encoding, the fundamentals of how you represent data visually. All charts are based on a combination of marks and attributes:

- Marks: Visual placeholders representing data *items*, such as distinct records or discrete groupings.
- Attributes: Variations in the visual appearance of marks to represent the values associated with each data item.

Expanding on this introduction, you were then introduced to a wide gallery of chart types, including profiles of 49 distinct approaches, to give you a sense of the common options that exist. The charts were organised into five family groupings, based on what each type is primarily used to show:

- Categorical: Comparing categories and distributions of quantitative values.
- Hierarchical: Revealing part-to-whole relationships and hierarchies.
- Relational: Exploring correlations and connections.
- Temporal: Plotting trends and intervals over time.
- Spatial: Mapping spatial patterns through overlays and distortions.

Influencing Factors and Considerations

If these were the options, how did you make your choices? The influencing factors included:

- Technological: What charts can you make and how efficiently?
- Purpose: What is the intended 'tone' of voice your representation should convey? Where is the emphasis between reading and feeling data?
- Data type and shape: The types of data and range of values you are trying to display will have a bearing on which charts you can use.
- Data exploration: What charting methods did you use to explore your data and did any of those represent possible means for communicating to your audience?
- Editorial angle: What is the specific angle of enquiry that you want to portray visually? Is it relevant and representative of the most interesting analysis of your data?
- Trustworthy design: Avoid deception through mistaken geometric calculations, 3D decoration, truncated axis scales, corrupt charts.

General Tips and Tactics

- Do not arrive at this stage with fixed, preconceived ideas about wanting to use certain chart types: be driven by your data and by your editorial thinking.
- Do not be precious: acknowledge when you have made a wrong call or gone down a dead end.

What now? Visit book.visualisingdata.com

EXPLORE THE FIELD Expand your knowledge and reinforce your learning about working with data through this chapter's library of further reading, references, and tutorials.

TRY THIS YOURSELF Revise, reflect, and refine your skill and understanding about the challenges of working with data through these practical exercises.

SEE DATA VISUALISATION IN ACTION Get to grips with the nuances and intricacies of working with data in the real world by working through this next instalment in the narrative case study and see an additional extended example of data visualisation in practice. Follow along with Andy's video diary of the process and get direct insight into his thought processes, challenges, mistakes, and decisions along the way.