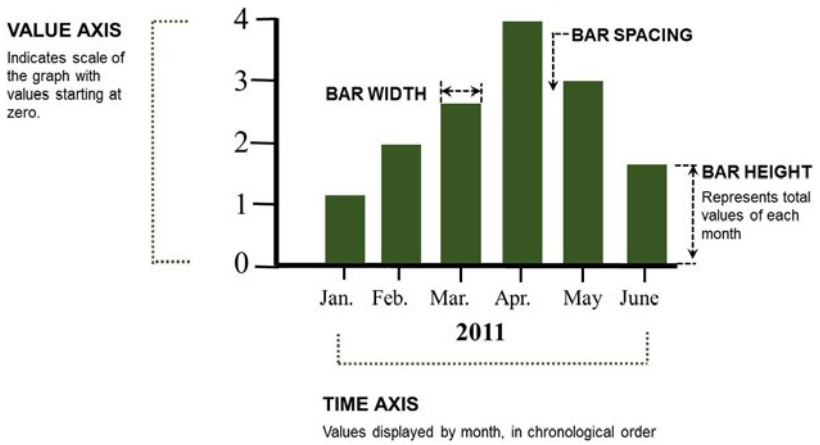


# COMMON CHARTS

# 2021

InfoVis

## DESCRIPTION AND USAGE OF MOST COMMON CHARTS

Which charts are most suitable for what data and user goals? Selection criteria, design and interaction recommendations.

Mireia Ribera
Data Science Master Degree

# Types of graphs

This chapter describes the most adopted types of graphs, and describe them in terms of data, interaction, and utility.

## Charts by task

Based on the Visual Vocabulary created by Financial Times there is a description of most common types of charts organized by task. Afterwards, every one is described in more detail. This list does not cover all charts included in the visual vocabulary but gives you a starting point.

### Deviation

Emphasize variations (+/-) from a fixed reference point. Typically, the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative)

Difference bar-chart, diverging stacked bar, spine and surplus/deficit area chart are in this category.

### Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationship you show them to be causal (i.e. one causes the other)

Scatterplot, Bar chart+ Line chart, Line chart, Bubble chart, HeatMaps are in this category

### Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the point of interest.

Bar chart(horizontal or vertical) with sort, Slope chart, lollipop, Step chart are in this category

### Distribution

Show values in a dataset and how often they occur. The shape (or skew) of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Histogram, BoxPlot, violin plot, population pyramid are in this category

### Change over time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader

Line chart, Bar chart(, , Bar chart+ Line chart, Slope chart, Area chart, streamgraph are in this category

### Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent

Bar chart(horizontal or vertical), Isotype, Lollipop, Radar, Parallel coordinates, Bullet chart, are in this category

## Part to whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Stacked bar chart, Pie charts and donut charts, Treemap and Waffle chart are in this category

## Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Choropleth, Flow maps and Proportional symbol maps are in this category

## Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Sankey diagram, chord and Network graphs are in this category

# Bar chart

Bar chart relies on length to show data. Shorter bars represent lower values, and longer bars represent greater values. Compare bar lengths to compare values.

Bar charts are basic statistical graphics, they do not include a wealth of data, but they are well known by every user, and very tuned to our perception system.

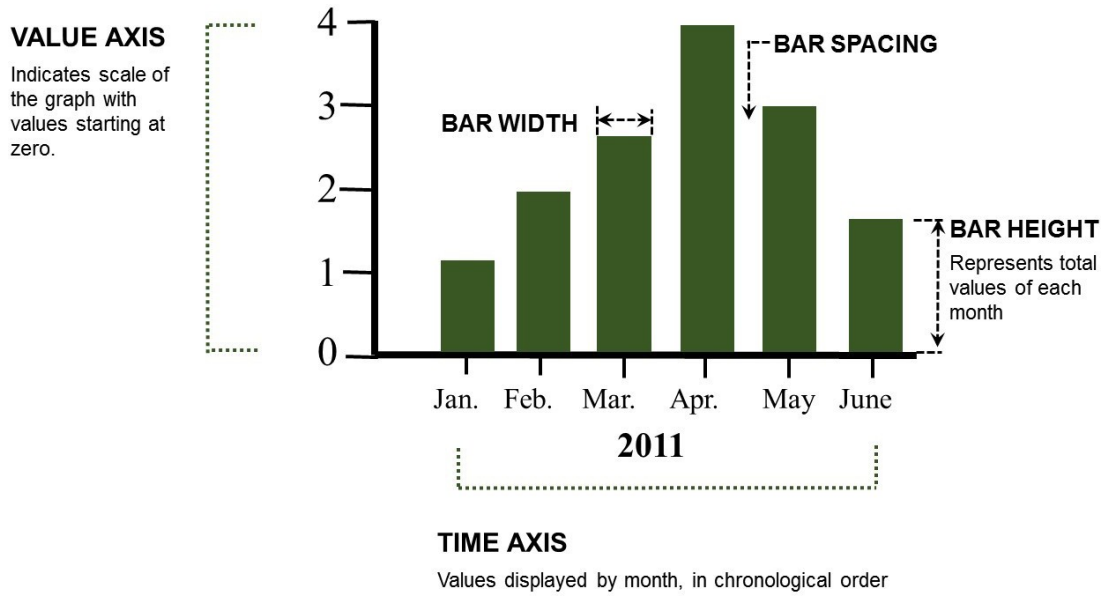


Figure 1. Schema of a bar chart. Source (Yau 2011)

## Taxonomy

- One dimension dependent attribute, quantitative. Often coded in the y-axis
- One or more independent attribute, categorical or ordinal. Often the independent attribute is time.
- Suitable for comparison queries
- It does not fit many data, as the axis puts a limit on its physical space.

## Adoption

- Adopted by all tools, even the most simplest
- Well known by every user
- Khan Academy has a pre-algebra content dedicated to read bar graphs and histograms

## Variant1. Horizontal bar chart.

In horizontal bar graphs, the eye can measure length with more precision. If there are many categorical values, vertical bar charts offer a horizontal view more adept for the eye.

In fact more than 2 series can become tricky to reaD

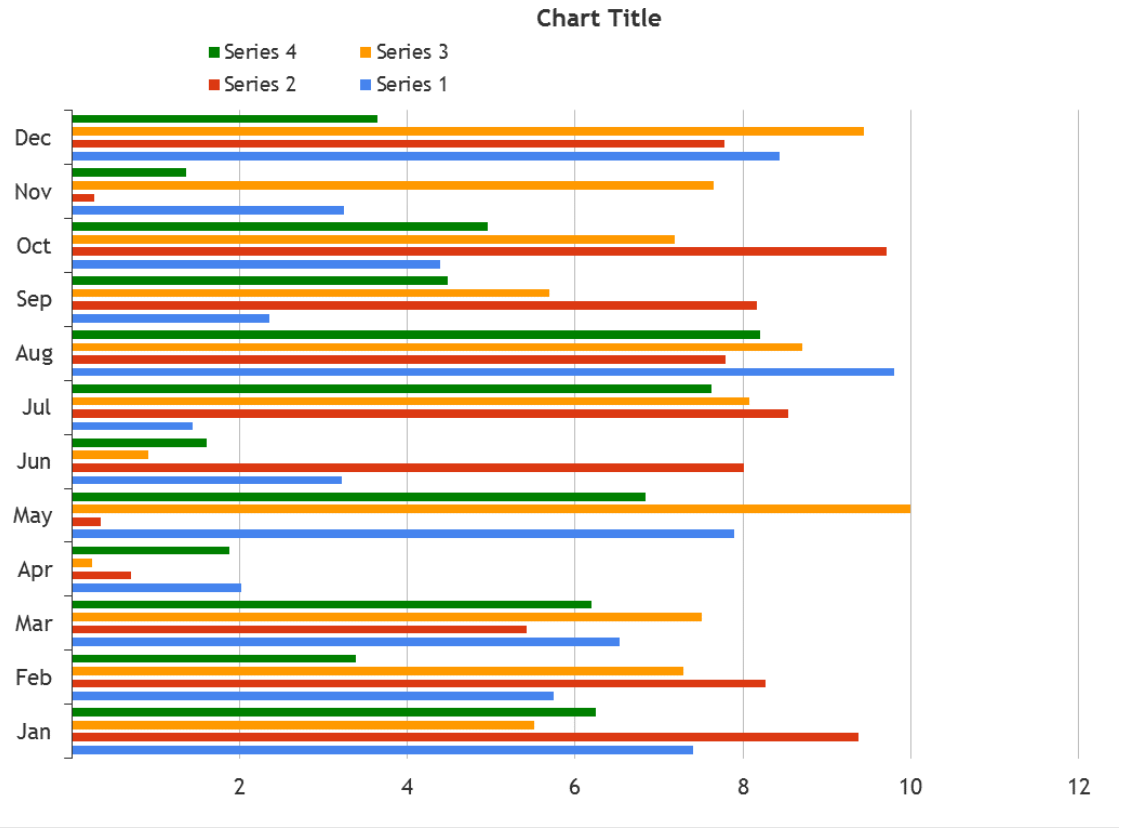


Figure 2. Horizontal bar chart

This type of chart is also useful to show rankings between different categories, when ordered.

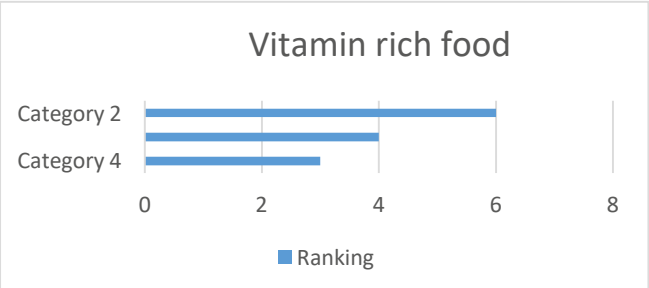


Figure 3. Sorted horizontal bar chart showing a ranking

## Variant2. 3D or more dimensions bar chart

The first independent attribute is coded on the axis not used by the dependent attribute, and if there are more than one they are usually color or texture coded.

## Variant3. Difference bar-chart

A threshold measure is used to separate values lower or higher than it, converting the dependent attribute into a diverging value.

It is the paradigmatic chart for deviation.

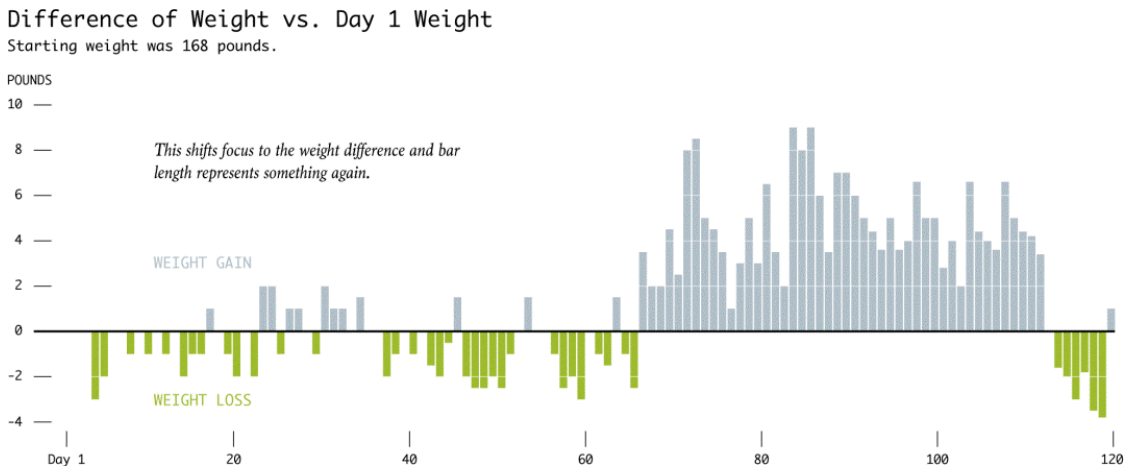


Figure 4. Difference bar chart

## Variant4. Stacked bar chart

A typical variant of the bar chart is the stacked bar chart, were several measures are “stacked” in each bar and the sum makes sense. Stacked bar chart can also be painted as difference bar charts.

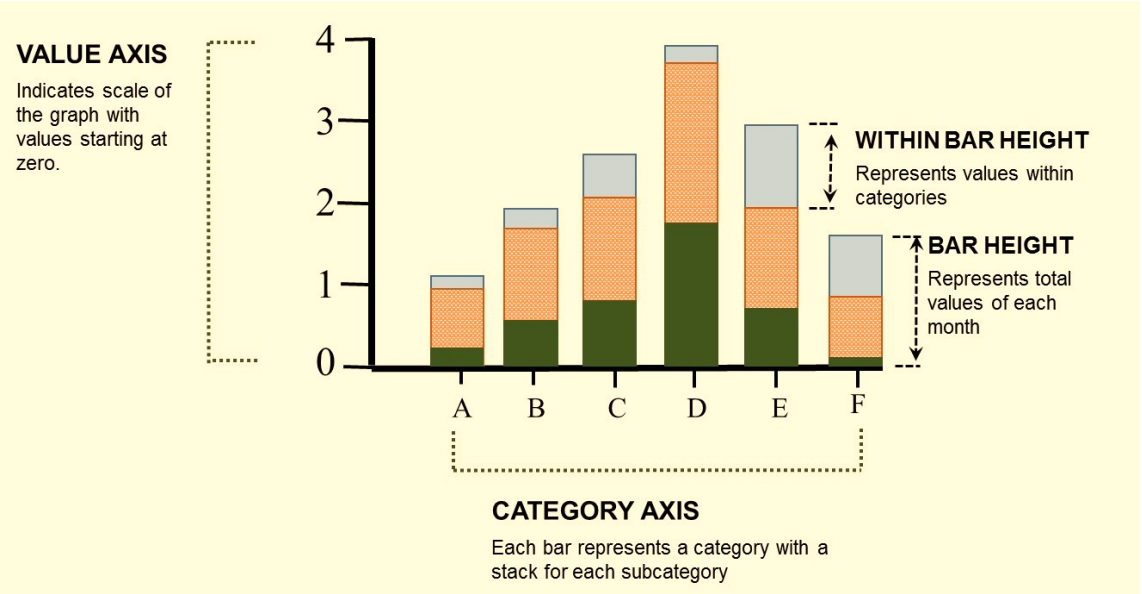


Figure 5. Stacked bar chart. Use sparingly

The problem with stacked bar charts is that it is not easy to compare values for the categories up to the first one, as their base is not the same. The recommendation is to use it with care, when the important things to show are the sum and the first value. If you provide interaction to change the order of the values, then every value could be compared when it occupies the first position, still not ideal for complete comparisons. A solution may be small multiples, with each chart depicting one of the categories.

Sometimes a one column stacked bar chart is a very handy solution to show a wealth of related values in a compressed form. A variant is to present not values but percentages, being the bar a total of 100%. It is a simple way of showing part-to-whole realationship but can be difficult to read with more than a few components.

Equivalent data tables

| | Series 1 | Series 2 | Series 3 | Series 4 |
|---|---|---|---|---|
| Jan | 7,41 | 9,38 | 5,52 | 6,25 |
| Feb | 5,74 | 8,27 | 7,29 | 3,39 |
| Mar | 6,52 | 5,42 | 7,51 | 6,20 |
| Apr | 2,02 | 0,70 | 0,24 | 1,88 |
| May | 7,90 | 0,35 | 9,99 | 6,84 |
| Jun | 3,22 | 8,01 | 0,91 | 1,61 |
| Jul | 1,43 | 8,54 | 8,08 | 7,62 |
| Aug | 9,80 | 7,79 | 8,71 | 8,21 |
| Sep | 2,36 | 8,17 | 5,70 | 4,48 |
| Oct | 4,39 | 9,71 | 7,19 | 4,96 |
| Nov | 3,24 | 0,26 | 7,65 | 1,37 |
| Dec | 8,44 | 7,78 | 9,44 | 3,65 |

Sometimes bar charts may be combined with line charts over time to show the relationship between an amount (bars) and a rate (line).

## Perception principles

Length: length is one of the most powerful preattentive properties, and it is perceived with high acuity. Our perception is relative, especially when comparing two lengths. Our perception is more acute for horizontal lengths than for vertical ones.

## Design recommendations

[Tufte] Data-ink ratio: reduce the information shown to the minimum without losing semantics

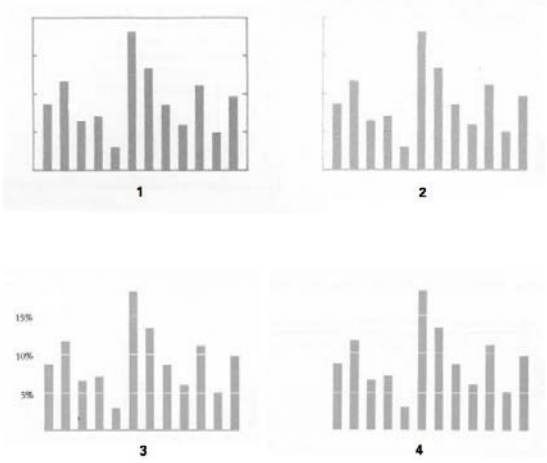


Figure 6. Redesign of a bar chart for minimal ink overhead

[Yau] Start the bar at zero.

As our perception of length is relative it is misleading to show a bar chart not starting at zero as we do not perceive absolute values but relative ones.

### Interaction patterns

- **Filtering**: With bar charts, you can use filters and blur or reduce intensity of some of the data to show more clearly the bars selected.
- **Highlighting**: With bar charts, you can use highlighting and colorize or emphasize some of the data to bring the attention on it. Filtering and highlighting techniques are complementary and can be used reversely.
- **Changing over time**: if the categorical attribute is not time, you can add a third dimension by interactivity observing the change of the bars through time.
- **Sorting:** as bar charts are perfect for comparison, being able to sort the data in any direction helps detect the best / worst in a category and to create ranks.

## BoxPlot

The box-and-whisker plot is an exploratory graphic, created by John W.Tukey, used to show a summary of the distribution of a dataset (at a glance) by showing the median (centre) and range of the data.

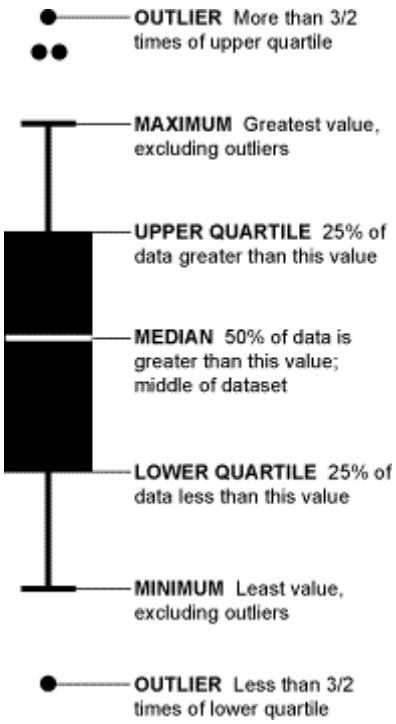


Figure 7. Boxplot. Use it for statistically oriented audiences only.

Khan academy has published a guide on how to read and create boxplots.

### Taxonomy

- 1 dimension quantitative variable. The boxplot shows the distribution
- Sometimes you can show several boxplots in the same graph for comparison purposes.

## Adoption

- Boxplot is not a laymen chart, but it is very well known among mathematicians and statisticians. You should be careful about your audience when using it.

## Variants

- A simplified boxplot extends the whiskers to the outliers and do not mark the ends horizontally.

## Perception principles

**Length**: length is one of the most powerful preattentive properties, and it is perceived with high acuity. Our perception is relative, especially when comparing two lengths. Our perception is more acute for horizontal lengths than for vertical ones.

# Line chart

Line charts are very similar to a Scatterplot with the difference that the dots are connected, so the x-axis attribute must be continuous. For time series (sometimes the chart is called time series chart), the use of a bar graph or a line chart will depend on what aspect of the data you want to focus on: trends and patterns or individual values.

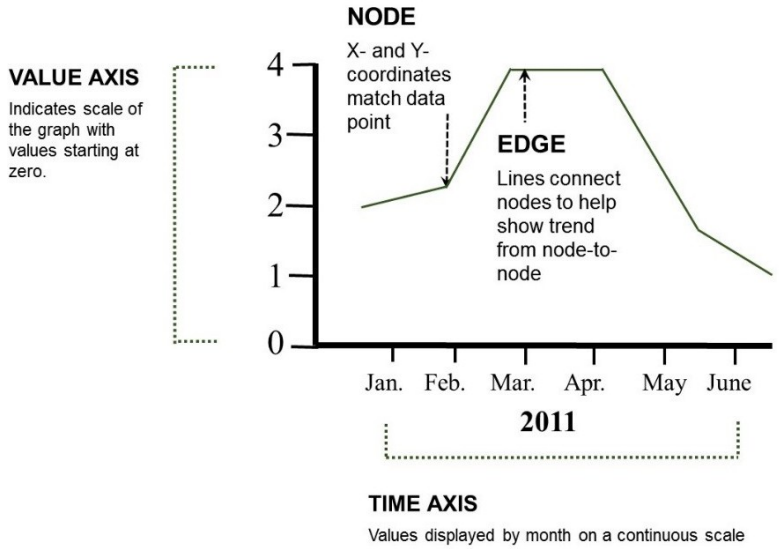


Figure 8. Line chart. Source (Yau, 2011)

In line charts the y-axis scale not necessarily starts at zero, although it is still a good practice. But sometimes it is better to start later and offer a more precise scale. Other times the scale will be logarithmic, but this shall be clearly indicated and it is more recommended for expert users.

See the comment on combining bar charts + line charts in the bar chart section.

## Taxonomy

- One dimension dependent attributes, quantitative. Often coded in the y-axis
- One or more independent attributes, categorical. Often the independent attribute is time. In line charts this variable must be continuous.
- Suitable for comparison variables and to detecttrends.
- It does not fit a lot of data, as it is limited by the physical space of the axis, but as already said, we can show only a small proportion of the axis, if we clearly indicate it.

## Adoption

- Line charts are basic charts mostly adopted and well known

## Perception principles

- Line charts rely on the Gestalt principle of connectedness, a most powerful perception principle than position.
- They are also based on the preattentive property of position, of course.

## Design recommendations

In case of combined line charts (when there are more than one line), it is recommended to use colors together with patterns to distinguish them. Also sometimes is good to label each line near the line, not in a legend.

Be careful with the y-axis scale, as it can magnify or minify the effect size of the shown variable.

## Variant1: Multiple line chart

Consists in using the same axis for several dependent variables. With this variant, you increase the dimensionality of the chart, and facilitate comparisons between the different depicted variables.

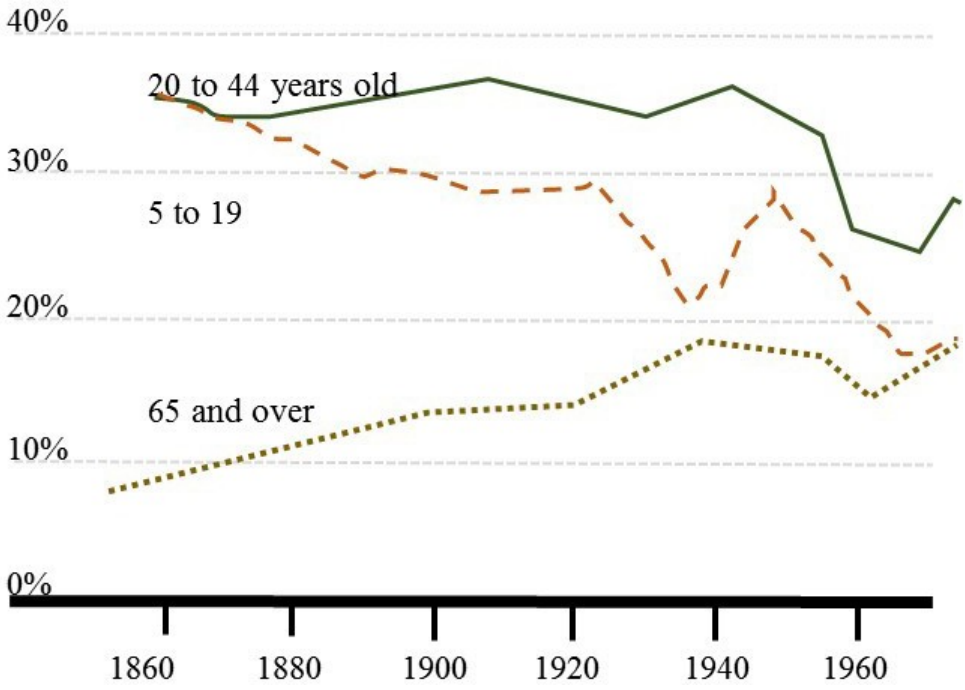


Figure 9. Multiple line chart. Source (Yau, 2011)

## Variant2: Stacked line chart / Area chart

As stacked bar charts, it is also possible to show line charts as stacked line charts. The advantage over regular line charts is that they show areas and you can evaluate totals. The disadvantage is that the perception of every variable is not clear, and the recommendations are similar to stacked bar charts.

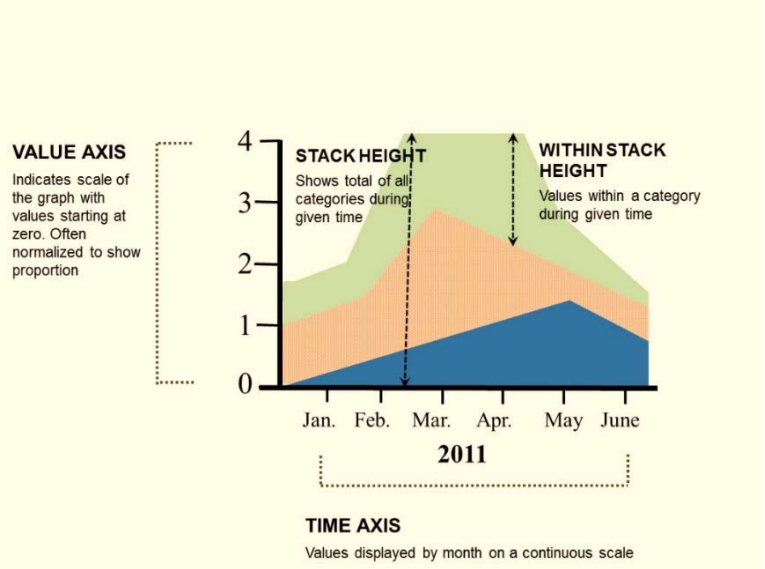


Figure 10. Stacked line chart. Use with care (Yau, 2011)

When baseline is the middle of the graph we talk about steamgraphs, they prioritize the perception of proportions over time over individual values.

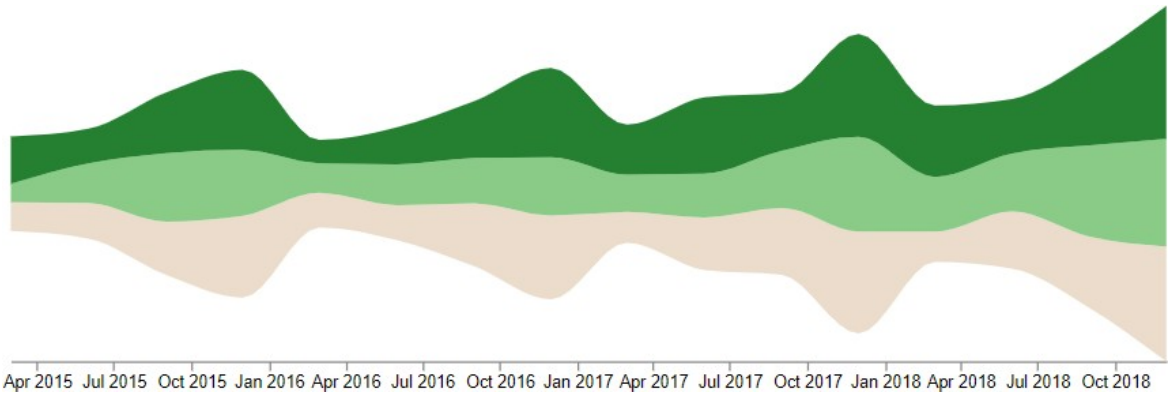


Figure 11. Steamgraph. Source https://gramener.github.io/visual-vocabulary-vega/#/Change-over-Time/

One value area charts can be used as surplus/deficit filled area for deviation purposes. The shaded area of these charts allows a balance to be shown; either against a baseline or between two series.

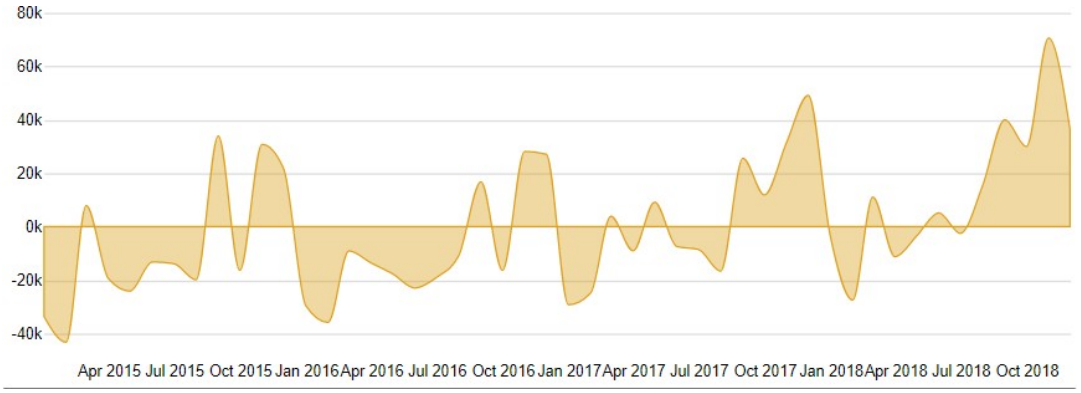


Figure 12. Surplus/deficit filled area. Source https://gramener.github.io/visual-vocabulary-vega/#/Deviation/

## Variant3: step chart / bump

This is very specific for some kind of dependent variables that don't change in a steady way, but change abruptly from one value to another.

When used for many values, it is effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using color.

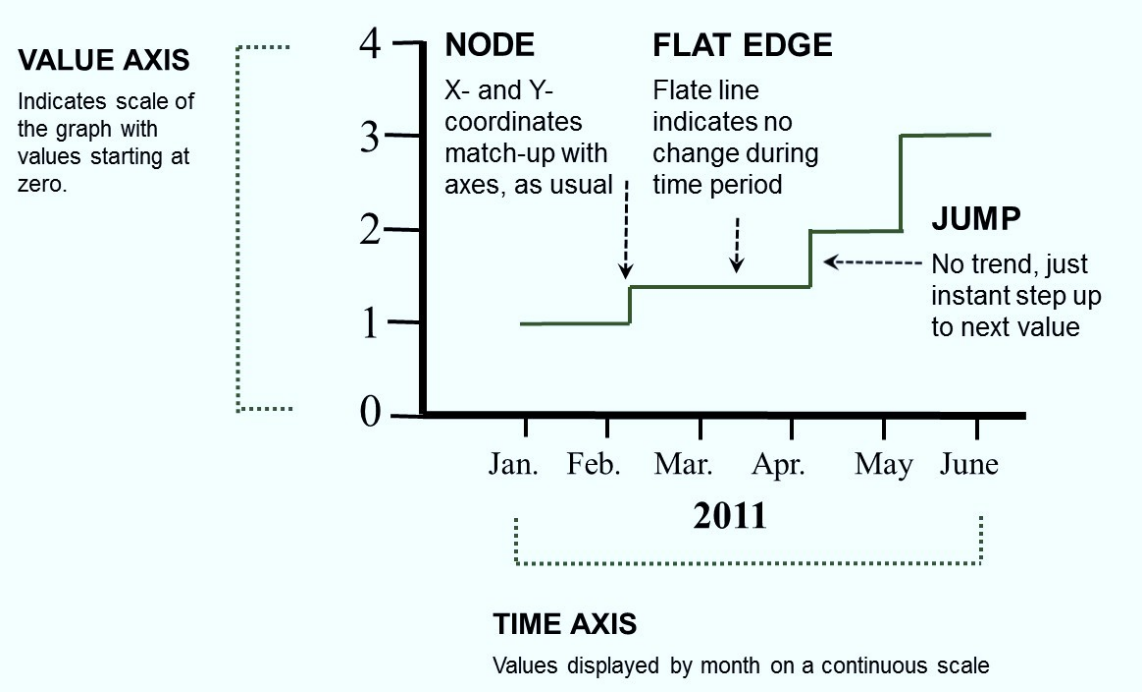


Figure 13. Step chart. Source (Yau, 2011).

# Sparkline

A very small unique line chart intended to be blended between text and numbers, the x-axis is usually time. Occupying a very small space gives a lot of information. Designed by Tufte as a thumbnail version of a line chart. It is an ideal chart for dashboards.


Figure 14. Simple sparkline side-by-side with text information

It can be improved with some context indicating where is located the described measure and where the normal limits of values are


Figure 15. Sparklines with a context bar indicating normal measures. Source (Tufte)

Sometimes instead of a superior and inferior line, a zero value line is indicated to put it in context.

If included in a table the header can indicate the time scope.


Figure 16. Sparklines with start and end marks. Source (Tufte)

The height should be enough to perceive differences in values.

For adoption, design, interaction and variants you can look at the line chart section. You can always use small line charts as sparklines.

# Scatterplot

A scatterplot show correlations. We can see whether or not, in what direction, and to what degree two paired sets of quantitative attributes are correlated [Yau]. Readers will often assume causality not correlation, tell them otherwise.

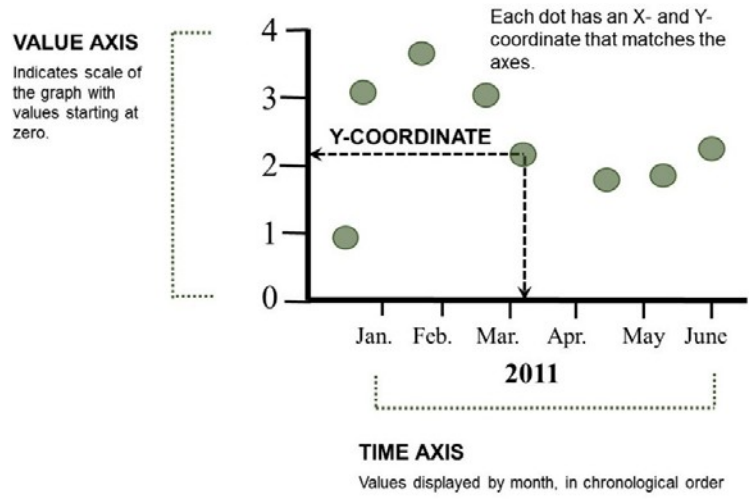


Figure 17. Scatterplot. Source (Yau, 2011)

This graph is good for finding correlation patterns in exploratory analysis. It is the standard way to show the relationship between two variables, each of which has its own axis.

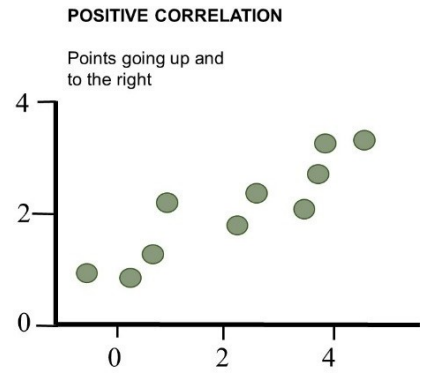


Figure 18. Scatterplot showing a positive correlation. Source (Yau, 2011)

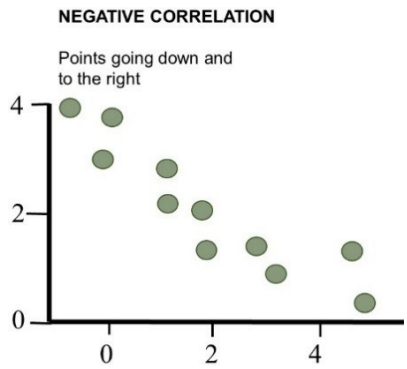


Figure 19. Scatterplot showing a negative correlation. Source (Yau, 2011)

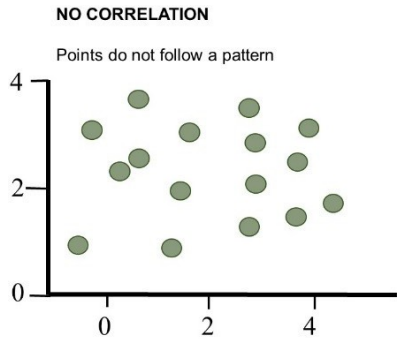


Figure 20. Scatterplot showing no correlation. Source (Yau, 2011)

It also show grouping in variables and with it outliers are easily detected

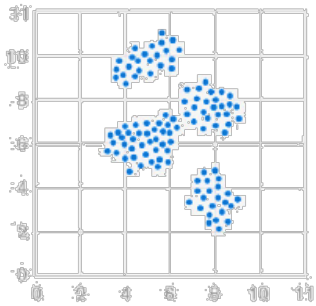

Figure 21. Scatterplot showing variable groups

## Taxonomy

- One dimension dependent attribute, quantitative. Often coded in the y-axis
- One independent attribute, ordered or categorical. Often the independent attribute is time.
- Optionally other quantitative attributes, often coded by size; or other categorical attributes coded by color or shape.

## Adoption

Scatterplots are very common graphics and every tool offer the possibility to create one. Sometimes they are related to regression analysis and advanced statistical tools offer the possibility to create best-fit lines.

## Variant1: Draftman's plot

For more dimensions, there is a generalized Draftman's plot, where each attributes could be compared to the others, but it is not easy to see multiple relations at a time. (Ware, 2013)

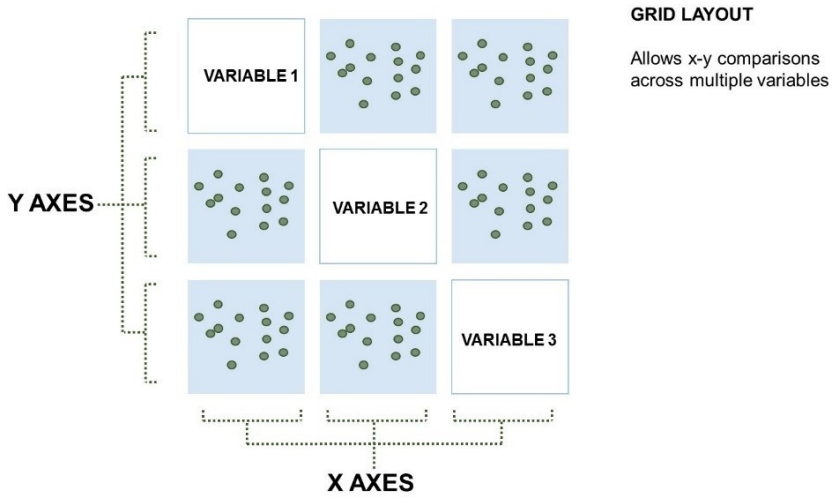


Figure 22. Draftman's plots. Source (Yau, 2011)

As an alternative for multivariate data you can use the parallel coordinates plot. User studies show best results with generalized Draftman's plots than with parallel coordinates.

G6.19 To display discrete data with more than four dimensions, consider using color-enhanced generalized draftsman's plots in combination with brushing (Ware, 2013)

## Variant2: Quadrants

In business meetings it is usual to show a scatterplot dividing variables in two bivalued variables. Being one of the most well known the Gartner Magic quadrant.

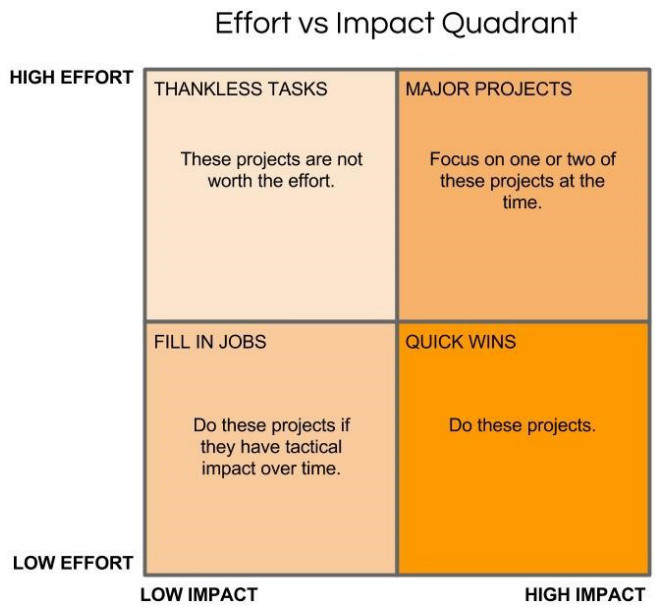


Figure 23. Magic quadrant example

## Variant3: Bubble chart

Another way to encode a third dimension in a scatterplot is to size the points depending on their value, then we have got a bubble chart.

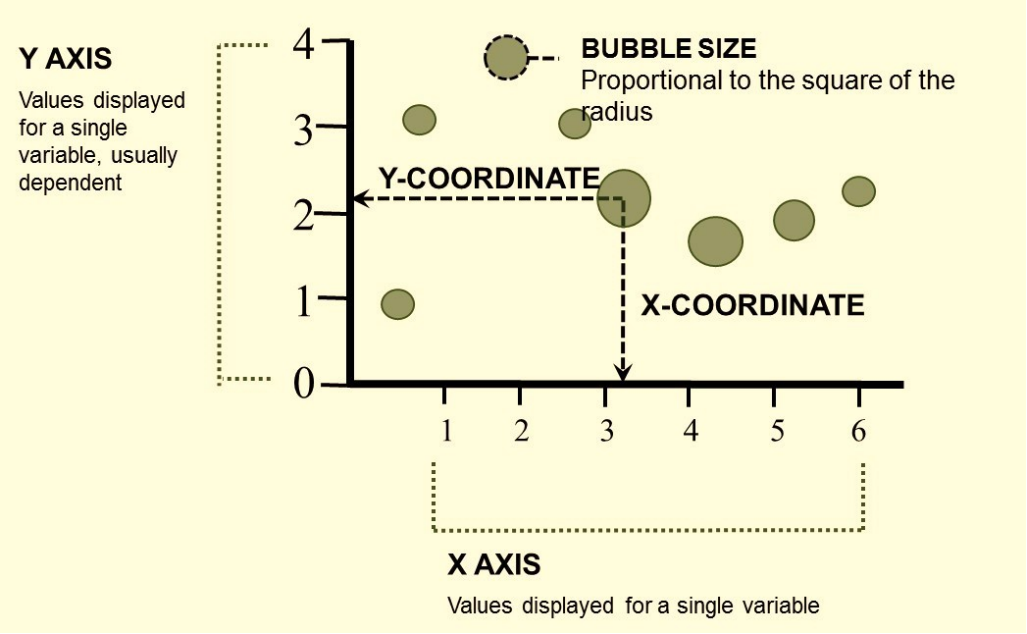


Figure 24. Bubble chart. Do not use. Source (Yau, 2011)

The problem with bubble charts is that the perception of area is not quite acute in our visual system, and we tend to change area by radius.

## Variant4: Scatterplot with regression best fit line

In this variant we are putting together observed values and estimation, and so you must clearly indicate your calculus for estimation and the error margin.

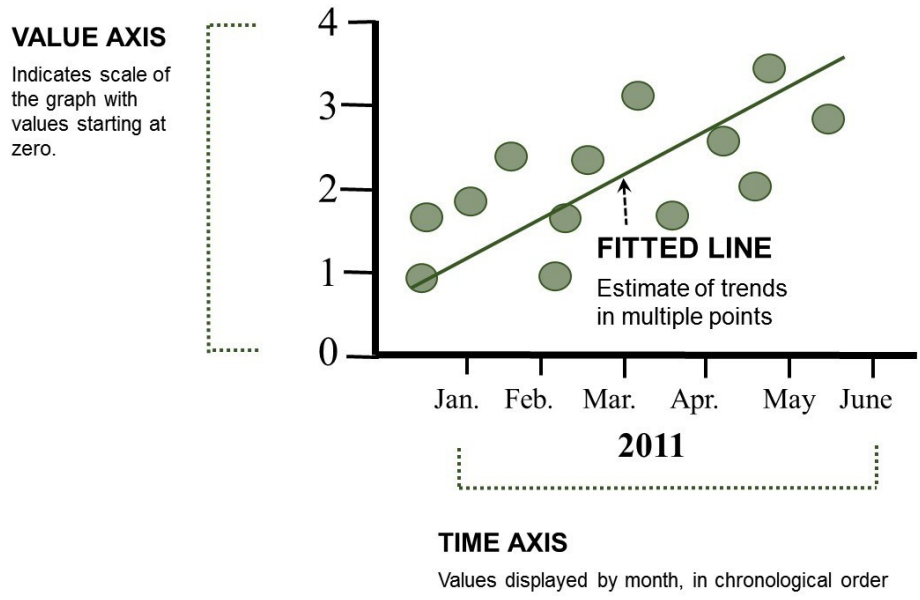


Figure 25. Scatterplot with regression estimate. Source (Yau, 2011)

## Perception principles

Position: position is a powerful preattentive property, and it is perceived with high acuity. Position is related to the Gestalt principle of proximity.

## Design recommendations

Scatterplots do not need to start axis on 0. It is better to have a good use of space, centering the points within the graph and offering proper resolution.

## Interaction patterns

- **Brushing**: As scatterplots are a quite dense chart, it is common to offer brushing interactivity for one, multiple points or for a region to see this selection by different views, specially in Draftman's plot.
- **Linked charts**: related to brushing it is also common to offer several views of the same data, linked by the brushing selection. So you can see a bar chart with aggregate data of selected points, or totals in numbers, etc.

# Pie charts and donut charts

A common way of showing part-to-whole data.

Our perception of dimensions and angles is not very precise. But still pie charts are a well known and very adopted chart. The recommendation is to use them to facilitate gross comparison among very few values (2 to 4), and give additional information on quantities or percentages. The order of the values may affect the perception.

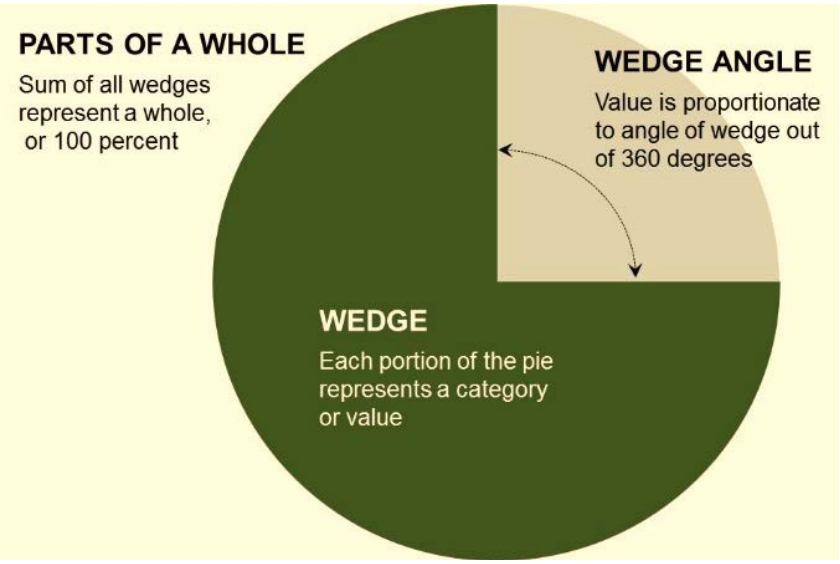


Figure 26. Pie chart. Use with very few values for gross comparison, give additional information. Source (Yau, 2011)

As an improvement to pie charts, donut charts reduce the effect of the angle and focus the attention more in the length. They are a way to show in a very small area some values. A common use is to show the progress of a unique attribute towards the goal (for example, consumption of internet data).

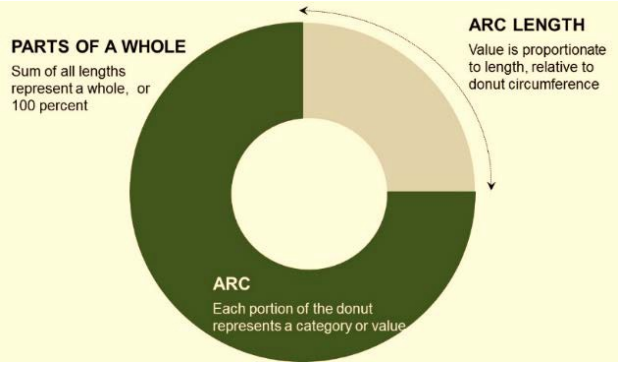


Figure 27. Donut chart. Don't use it. Source (Yau, 2011)

The recommendation is to use them to facilitate progress of one attribute or gross comparison among very few values (2 to 4), and give additional information on quantities or percentages using the central space of the donut. The order of the values may affect the perception. The center can be a good way of making space to include more information about the data.

A variant of donut charts are arcs, often used for visualizing political results

# Waffle chart

Good for showing % information, they work best when used on whole numbers.

Waffle charts are commonly used in infographics, mainly to show percentages. It is useful to show a value within context. They are useful as indicators.


Figure 28. Simple waffle chart. Source: Guia, 2018

To animate more the waffle charts often the points are gifs depicting the quantified category (as persons, cars…)

# Parallel coordinates

Parallel coordinates are an extension of line charts to represent multivariate data. Each item is represented by a vertical line, and every attribute by a line. A drawback is that the order of variables affects the detection of patterns (Ware, 2013).

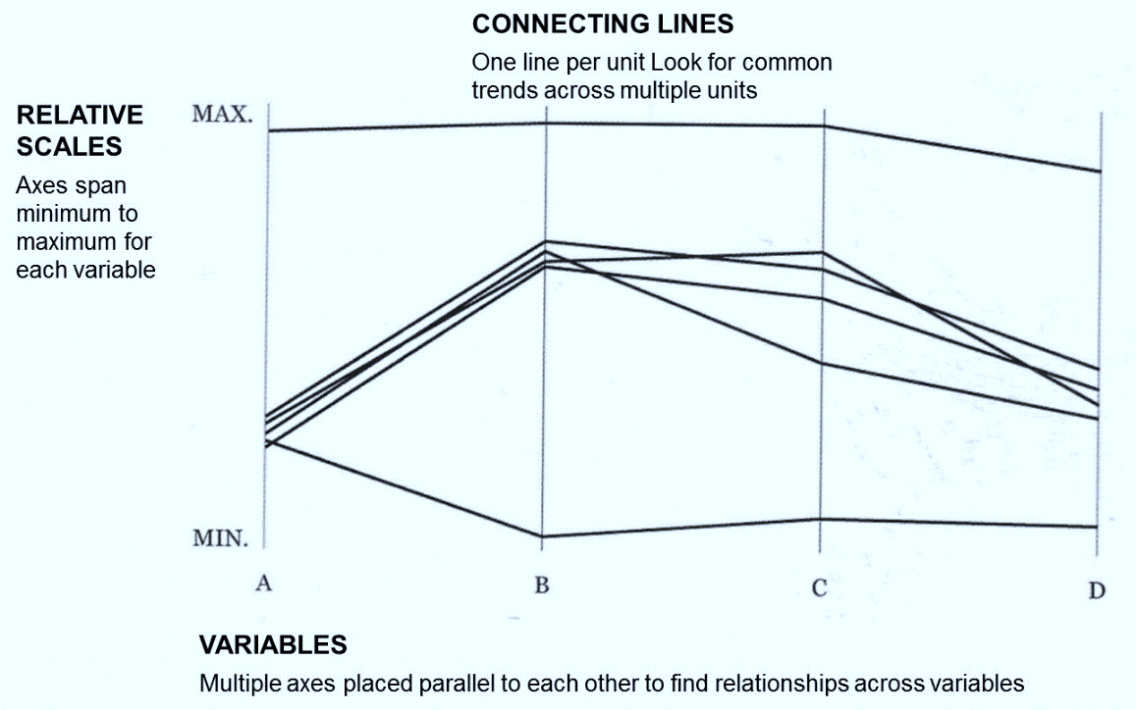


Figure 29. Scatterplot. Source: (Yau, 2011). Use sparingly

As a more usable graph, look at the generalized draftman's scatter plot.

## Taxonomy

- Multiple independent items, categorical
- Multiple dependent attributes, quantitative. They must have a value for every item, and share units and range values.

## Adoption

- Not highly adopted. It is complicated to read.

## Perception principles

- It is based on Gestalt principle of connectedness

## Design recommendations

- Use sparingly, and always with interaction for filtering.

## Interaction patterns

It is intended to be used with brushing. The user selects one of the axes, as part of an exploratory process. It is also good to offer some filtering option to reduce the attributes shown, or a selection option to highlight a particular attribute.

## Variant1: Slope chart (created by Edward Tufte)

In a slope chart only two attributes (one of them, usually time) are plotted for comparison, and extensive labelling of each value is given. Alberto Cairo uses this type of graph very often.

It is perfect for showing how ranks have changed over time or vary between categories.

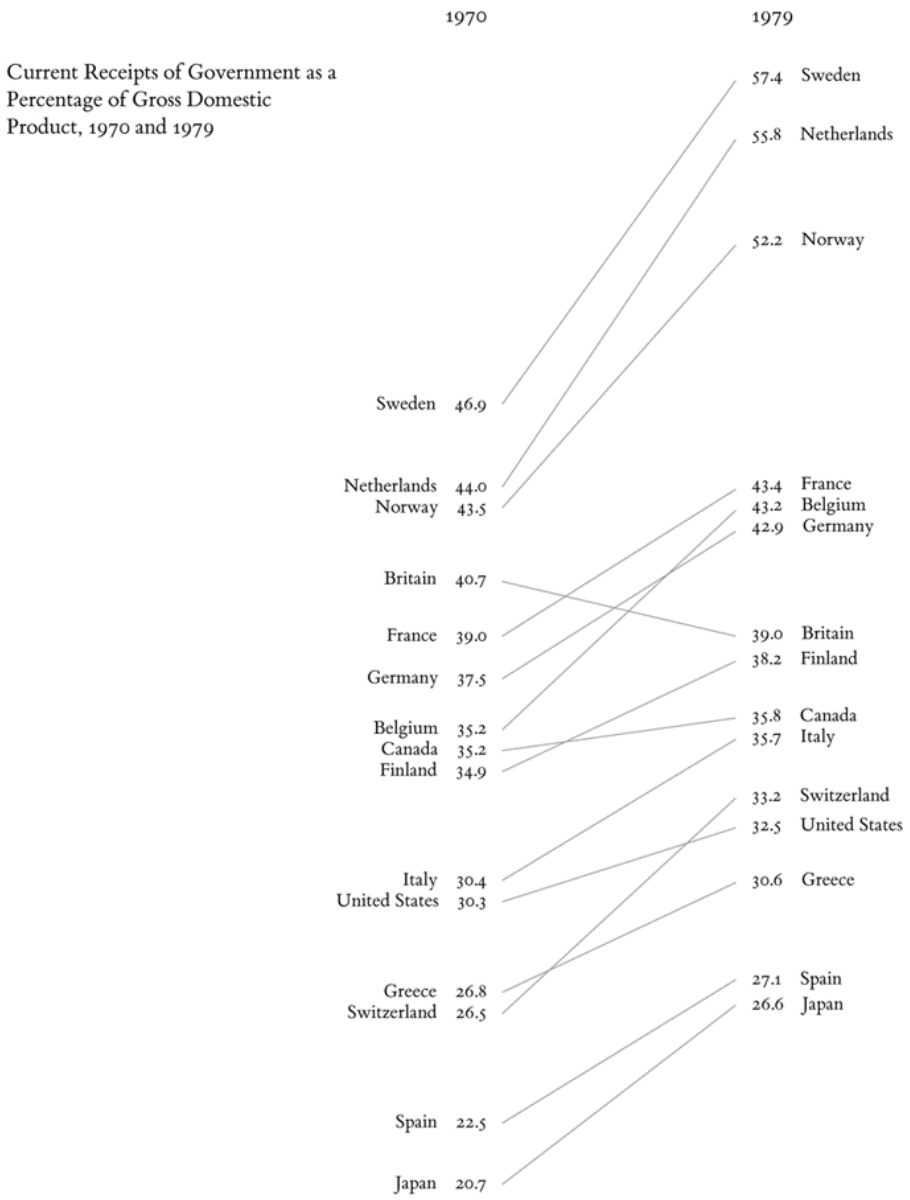


Figure 30. Slope chart. Source: (Tufte, 1983). A recommended variant of parallel coordinates

Axis must be same scale and same start point.

# Histogram

Histograms are a very classical statistics graph showing distribution of a variable.

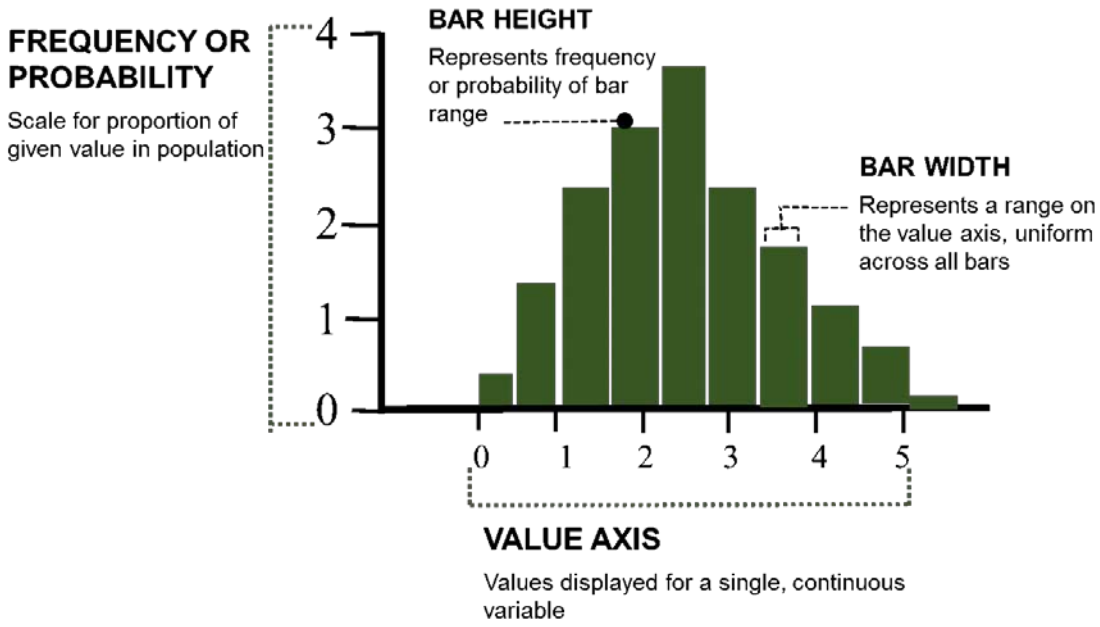


Figure 31. Histogram. Source: Yau, 2011

## Taxonomy

- One dimension quantitative attribute. The histogram shows the distribution instead of frequency. A probability scale chart shows the estimated probability of the corresponding value occurring.

## Variant: Probability scale

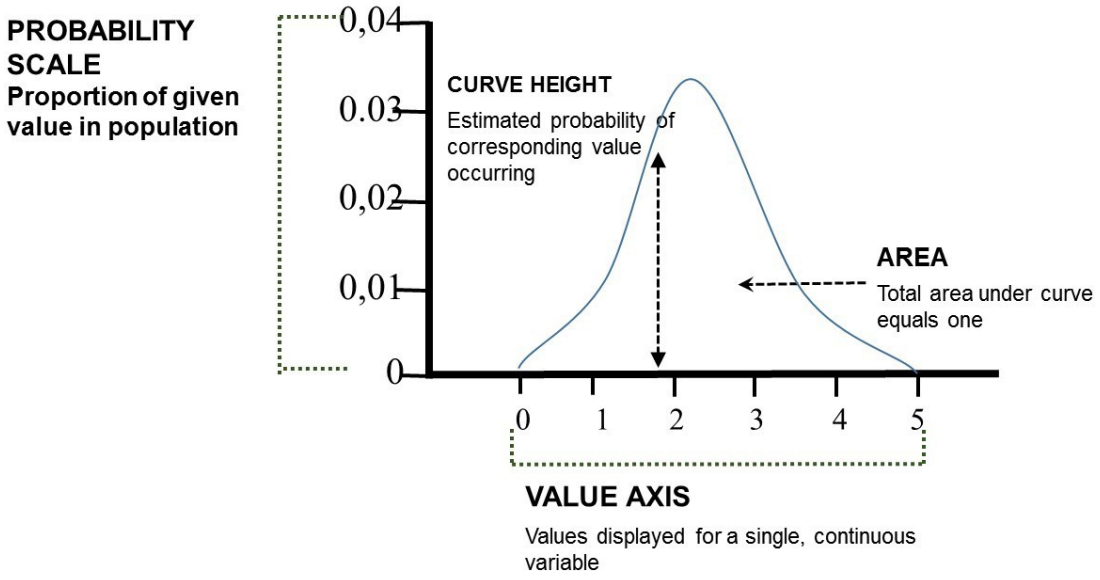


Figure 32. Probability scale

## Adoption

- Histogram is a very common graph, based on bar chart and it is very well known. For more experienced audiences you can use boxplots instead.

## Design principles

The gaps between columns should be small to highlight the “shape” of the data.

If you have to show several histograms at once, you can use the small-multiples strategy in order to allow quick comparison. In a very small space, you can compare many years distributions and perceive general trends.

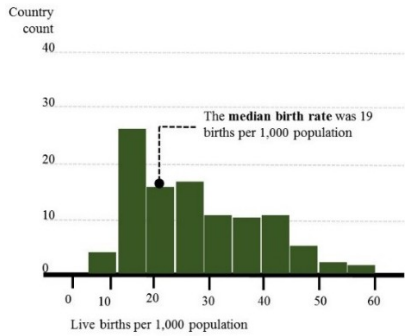


Figure 33. One of the histograms to show

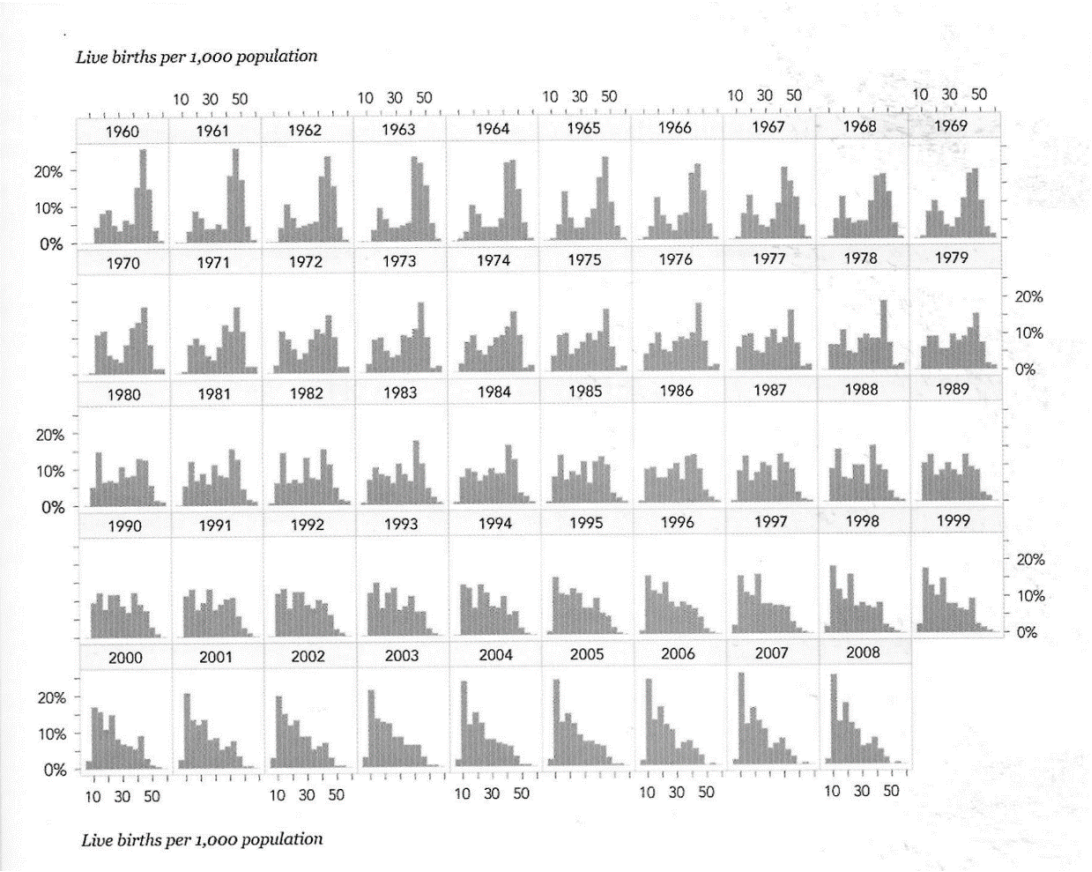


Figure 34. Several histograms shown on a small-multiples display.

With interaction, you can allow to select one of the histogram and zoom at it to get a more detailed view. For additional interaction, design principles... you can refer to the bar chart section.

# Bullet graph

Bullet graph is a graph designed by Stephen Few to substitute gauges in dashboards.

Designed to display, in a compact manner, a single key measure along with a comparative measure and qualitative ranges to instantly signal whether the measure is good, bad, or in some other state.

A bullet graph is just a bar graph with a single bar and an additional mark for a quantitative comparison. Shades of color in the background give context to the different values. The author published a complete specification in his blog.

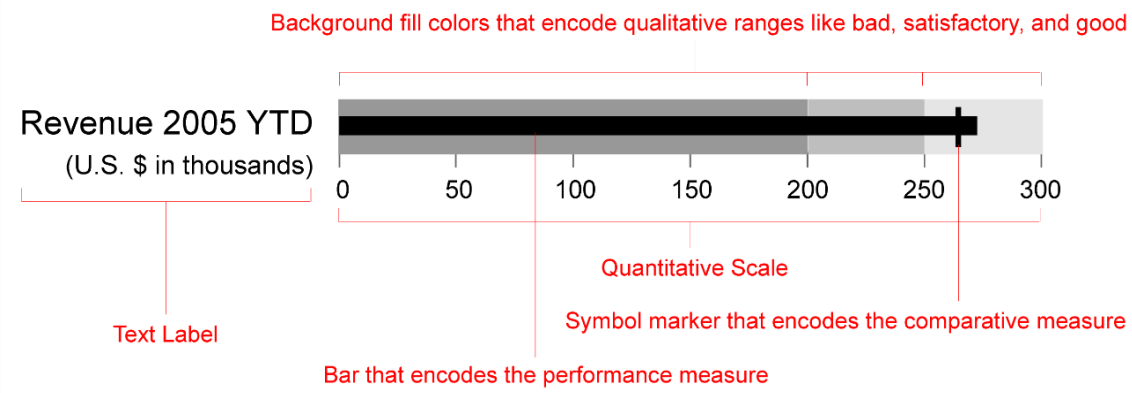


Figure 35. Bullet graph. Source Few 2013

You can combine different colors for background bars and foreground bar. On the background it is recommended to use a unique hue and changes of light/saturation.

## Taxonomy

- 1 dimension quantitative variable
- Several values are shown for the same variable: bad, satisfactory and bad levels; expected value and real value.

## Adoption

- Although quite a new chart it is being increasingly adopted by many tools and in practice.

## Perception principles

**Length**: length is one of the most powerful preattentive properties, and it is perceived with high acuity. Our perception is relative, especially when comparing two lengths. Our perception is more acute for horizontal lengths than for vertical ones.

## Design recommendations

- [Yau] Start the bar at zero.

As our perception of length is relative it is misleading to show a bar not starting at zero as we do not perceive absolute values but relative ones.

# Treemap

Use for hierarchical part-to-whole relationships; can be difficult to read when there are many small segements.

A treemap is a graph created by Shneiderman to visualize the occupation of his hard drive. This graph shows a hierarchy and two attributes per each leaf node, one quantitative represented by size and one categorical represented by color.

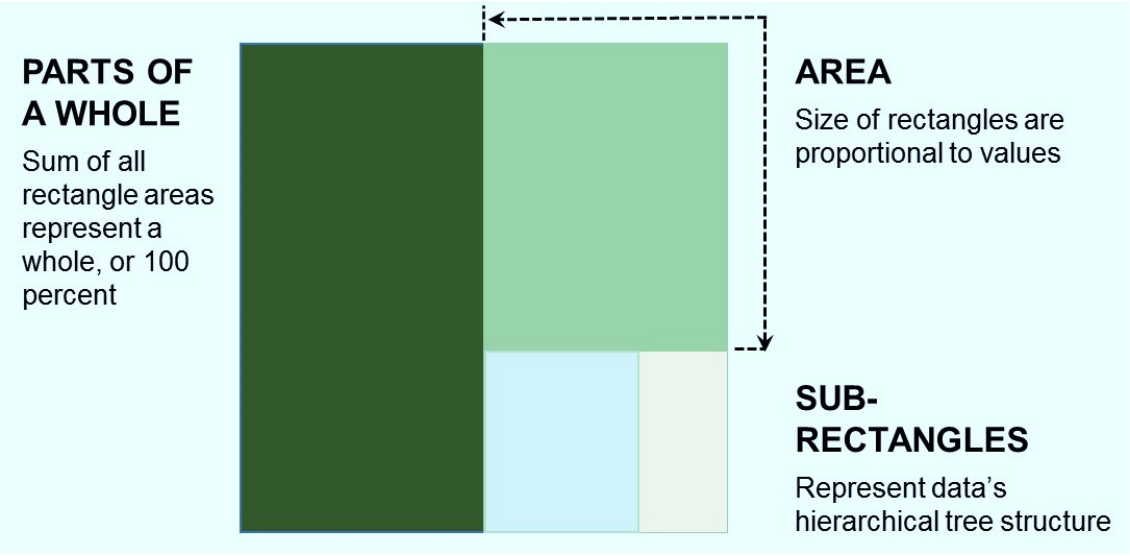


Figure 36. Treemap. Use sparingly and always with interaction. Source: (Yau, 2011)

The main advantage is its capacity to cover lots of data in a very compressed view (Few, 2009). The disadvantage is that non-leaf nodes are not shown and the hierarchical structure is not as clear as in conventional drawing. You may instead use tree network graphs.

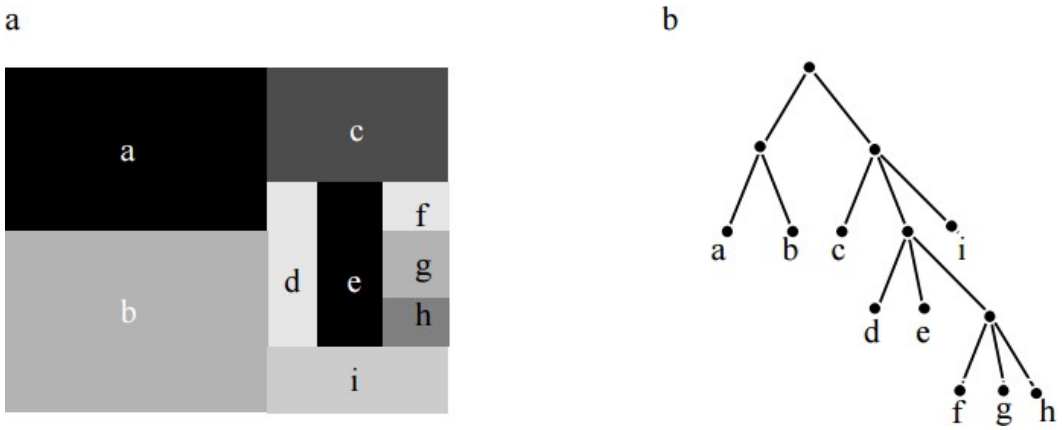


Figure 37. Hierarchy as a treemap and as a classical tree graph. Source (Few, 2013)

## Taxonomy

Hierarchical structure and two attributes for every leaf-node: one quantitative represented by size, one categorical represented by color.

## Adoption

- Not very adopted, but common in press when summarizing lots of data such as elections or market trends.

## Perception principles

- Size, color and position are preattentive properties, easily perceived. Area perception is not very good in our visual system, so we can get a gross idea of trends but no acute detail.

## Design recommendations

Colin Ware does not recommend to use more than 8 variants of color in a graph (Ware, 2013).

Be aware of how to label different leafs, as they can easily become cluttered.

## Interaction patterns

Zoom through the different levels of the hierarchy

Details on demand. When hovering or selecting a category more information is shown.

# HeatMaps

When there are many variables to look at, looking at all at once can point out what to pay attention to. This is the main use of heatmaps, to show a big deal of numbers by coloring them.

It is a good way of showing the patterns between 2 categories of data, less good at showing fine differences in amounts. Often the value is painted onto each tile.

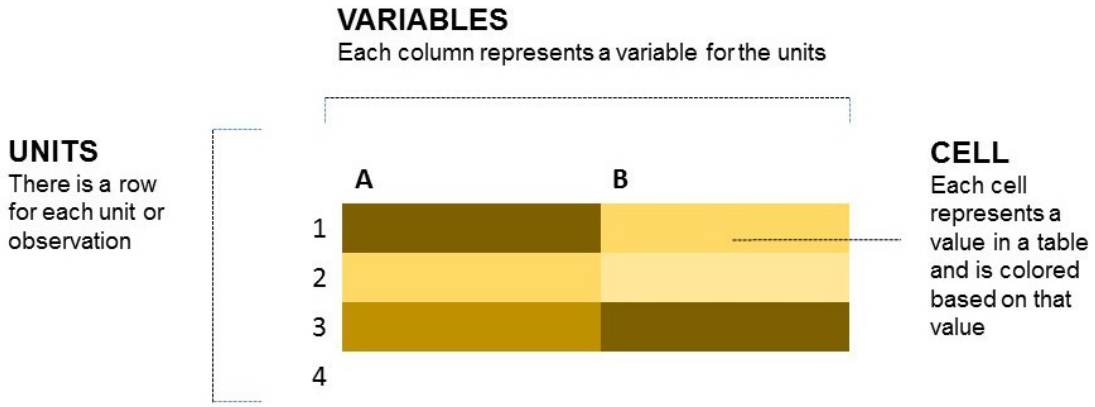


Figure 38. Heatmap. Source (Yau, 2011)

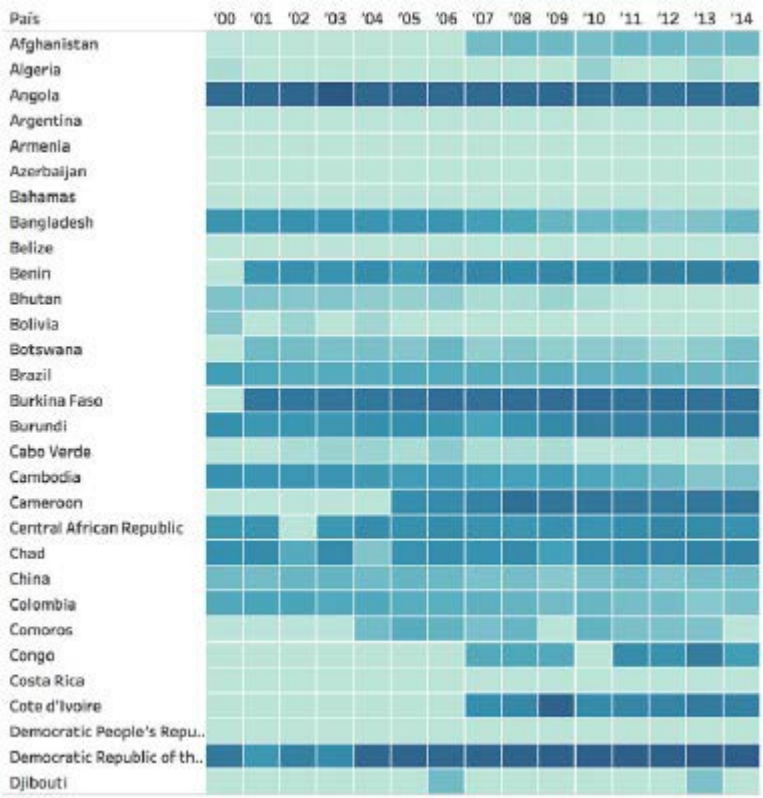


Figure 39. Heatmap example, source (Guia 2018 https://public.tableau.com/views/malaria_5/Dashboard1). People with Malaria.

## Taxonomy

- One dimension dependent attribute, quantitative. Coded in color.
- Two independent attributes, categorical. Coded in the axis.
- Suitable for getting an overview of very dense data, not to get a very detailed information. Suitable for comparison.

## Adoption

- Choropleth maps are the geographical cousins of heat maps and are much adopted, but the tabular version is not very generalized. Even so, many visualization software allow creating them.

## Perception principles

**Color**: Saturation and Lightness code the quantitative value of the dependent variable. Usually only a hue is used.

## Design recommendations

When it is necessary to have precise numbers you can complement the heatmap with the real table or put both numbers and colors in the same space being aware of contrast issues.

## Interaction patterns

- **Hovering**: it is possible to indicate the quantitative value of a cell when hovering it.
- **Sorting**: You may be able to sort rows by its value

# Choropleth

Choropleth is a kind of thematic map that charts phenomena that are evenly distributed within a specific area.

The shown attributes must be indexed by area to avoid a highlight effect on big areas

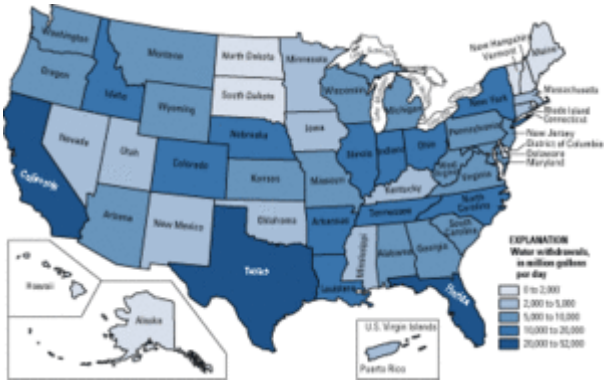

Figure 40. Chorepleth map of water use.
Source (http://wiki-1-1930356585.us-east-1.elb.amazonaws.com/wiki/index.php/Thematic_map)

## Taxonomy

Geographical areas and one ordered attribute (rate), represented by color. Depending on the nature of values, the colors could be sequential or bivariate.

## Adoption

- Is the most widely used thematic map. The projection should be known by the user.

## Perception principles

- Color and position are preattentive properties, easily perceived.

## Design recommendations

- Don't use more than five value levels for sequential values, seven to nine for bivariate, and include a legend. Experiment with threshold scales.

## Interaction patterns

Geographical and Semantic zoom

Details on demand

# Proportional symbol map

A proportional symbol is a type of thematic map that uses map symbols that vary in size to represent a quantitative variable. The symbol could be of any kind and a graph in itself and often is representative of its meaning. Very often only one attribute is depicted as a bubble icon.

Use for totals rather than rates – be wary that small differences in data will be hard to see.

Stephen Few proposed a solution for bubble icons, the bricks (Few & Edge, 2013) which more clearly depict quantities. See his proposal and compare the map at the left to the map at the right.

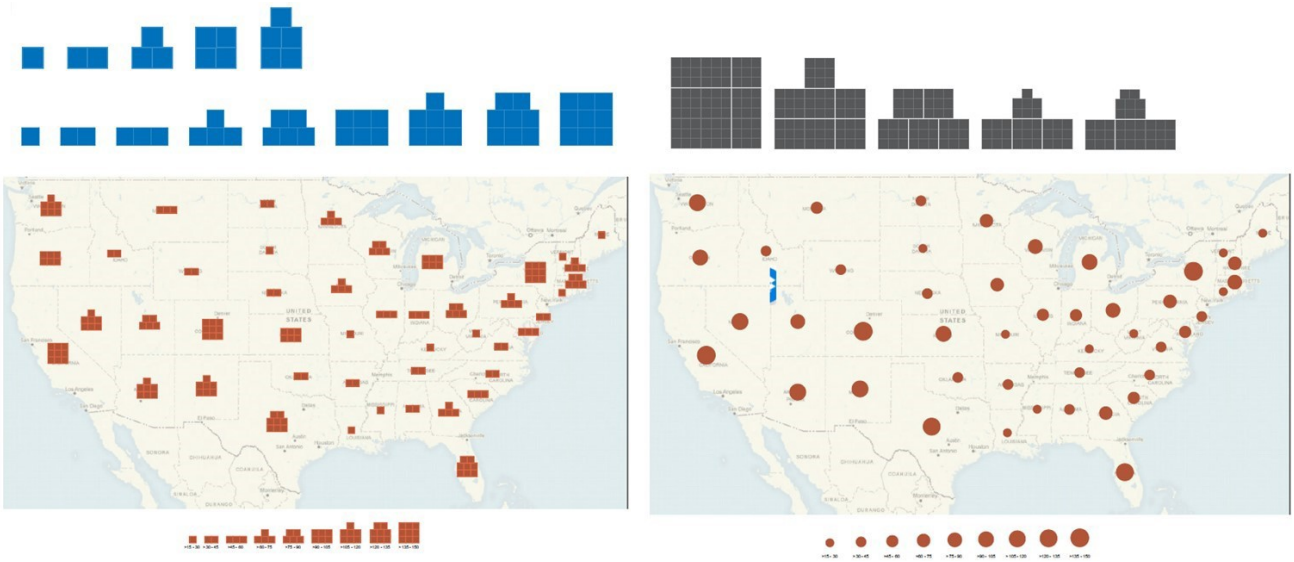

Figure 41. Few bricks. Source (Few & Bridge 2013)

## Taxonomy

Geographical areas and a few ordered attributes, categories represented by the symbol, quantities by size. Not many symbols could be included or it will look cluttered.

## Adoption

- Is the very common thematic map. The projection should be known by the user.

## Perception principles

- Color and position are preattentive properties, easily perceived.

## Design recommendations

- Use clearly differentiate icons. Group values into some few categories.

## Interaction patterns

Geographical and Semantic zoom

Details on demand

Layers

# Flow maps

A Flow map is a type of thematic map that hybridizes maps and flow charts, showing the movement of objects from one location to another such as the number of people in a migration, the amount of goods being traded, or the number of packets in a network [wiki]


Figure 42. NY Taxi visualization. Source (https://taxi.imagework.com/)

## Taxonomy

Geographical areas and paths. For not very dense area different categories could be shown by different color paths.

## Adoption

- Not as common as the previous maps, they are increasingly used thanks to GPS.

## Perception principles

- Color and position are preattentive properties, easily perceived. Connectedness is a strong Gestalt principle

## Design recommendations

- Offer several color combinations to choose.

## Interaction patterns

Geographical and Semantic zoom

Highlight

Details on demand

# Network graphs

Network graphs show relations between items. They have two basic elements: the node and the link. The node can be colored, have a shape or even a real picture; they also can be labeled. The link can be straight or curved (better for perception), can have direction (with an arrow), different widths or colors and can be labeled. This graph easily becomes cluttered, but with zoom and pan can show a lot of data.

It is a very good graph to detect patterns and communities.

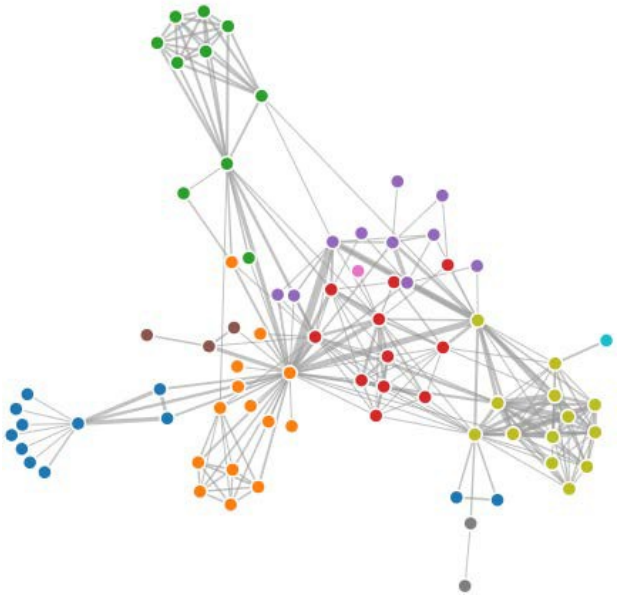

Figure 43. Network graph in D3 by Mike Bostock [https://beta.observablehq.com/@mbostock/d3-force-directed-graph]

## Taxonomy

One categorical value, coded in nodes. One quantitative attribute relating two items, coded in link. This can be enriched by many other categorical values though color, shapes and labels. Quantitative relations can be also coded in link width.

## Adoption

Although not offered by many visualization tools, this is a very common graph to show relations, as it is unique in its kind.

## Variant: Tree graph

Tree graphs are a particular type of network graph with hierarchical structure. All nodes depart from a root and they are increasingly divided.

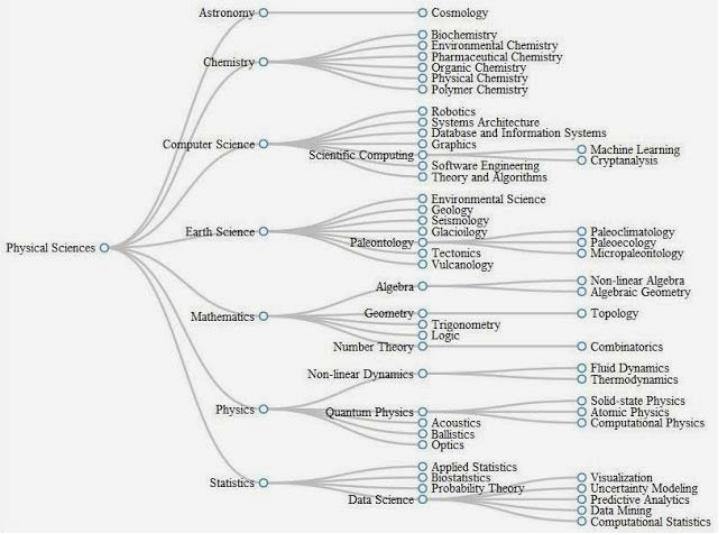


Figure 44. Collapsible tree. Source: http://2centsapiece.blogspot.com/2013/12/hierarchical-tree-diagram-with.html

They are very common in text analysis. And often they offer interaction facilities to deploy or collapse nodes.

## Perception principles

Network graphs rely on the Gestalt principle of connectedness, a very powerful perception principle. Curved links show this relation in a smoother way that is best for perception.

## Design recommendations

Be careful to not clutter the graph. Do not include animations for the sake of impact.

## Interaction patterns

Zoom and pan: you can move through the visualization space through zoom and pan. Usually semantic zoom is also offered.

Layers: In order to avoid clutter, it is good to offer layers of information on demand.

Detail on demand: It is common to offer detailed information of a node when selecting it by a tooltip or information box.

# Sankey diagram

Sankey diagram are used to show the flow and relation of some variables through different situations. It is more a design diagram than a data driven one.

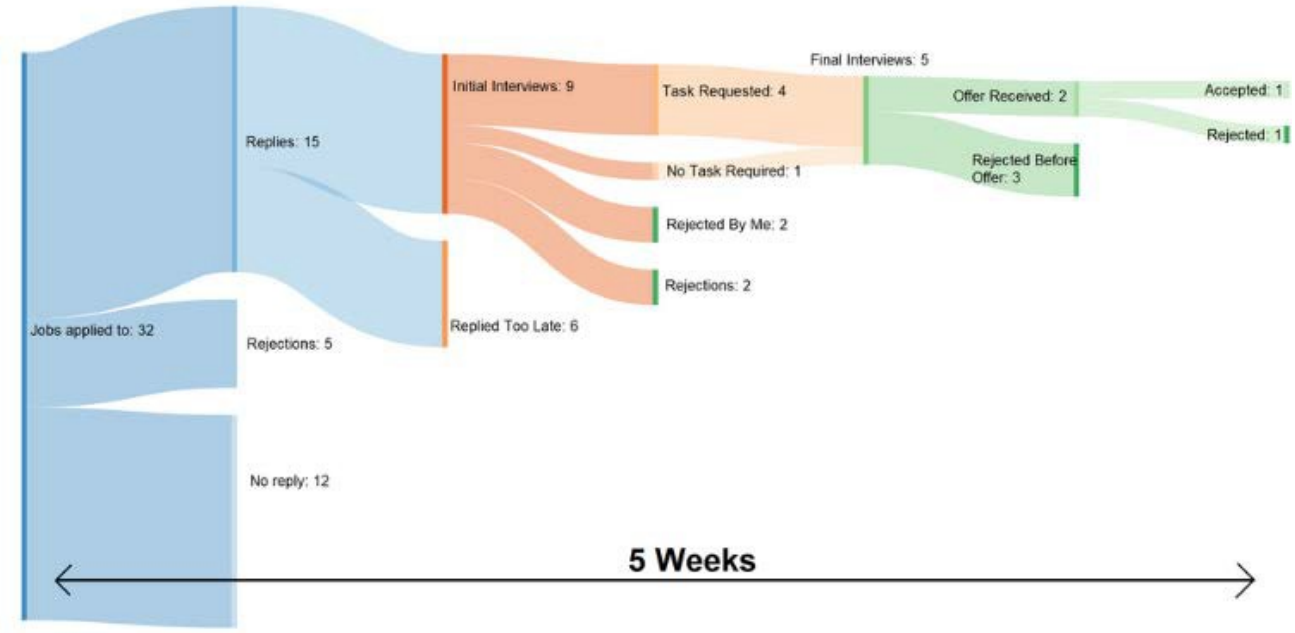


Figure 45. Sankey diagram showing evolution of job seeking efforts. Source: https://www.reddit.com/r/dataisbeautiful/comments/992s4c/a_detailed_sankey_diagram_of_my_recent_job_search/

## Taxonomy

Several categorical variables. Time or process flow as horizontal axis.

## Adoption

Common in infographics. D3, Highcharts and Tableau offer tools to create Sankey charts.

## Perception principles

Sankey diagrams, as network graphs, rely also on the Gestalt principle of connectedness.

## Design recommendations

As it is not a clearly defined chart, it is recommended to iterate drafts and rely on user centered design techniques. Do not overlap flows.

# Bibliographic references


- Arno Klein “CTHRU: a composition-based taxonomy of information graphics” in *Medium* published online at https://medium.com/@binarybottle/cthru-a-composition-based-taxonomy-of-information-graphics-c57dd419e8b4. [Consulted on 29-august-2016]
- Few, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press.
- Few, S. (2013). *Information dashboard design: displaying data for at-a-glance monitoring* (2nd ed.). Analytics Press.
- Few, S., & Edge, P. (2013). *Building Insight with Bricks: A New , More Perceptible Method for Encoding Quantitative Values in Geospatial Displays*.
- Guia de visualització de dades. Generalitat de Catalunya. 2018. ISBN: 978-84-393-9734-2
- Tufte, E. R. (1983). *The visual display of quantitative information*. Graphics Press.
- Ware, C. (2013). *Information visualization: perception for design* (3rd ed.). Morgan Kaufmann.
- Yau, N. (2011). *Visualize this: the flowingdata guide to design, visualization, and statistics*. Wiley Publishing.
- Wiki.GIS.com “Proportional symbol map” [http://wiki-1-1930356585.us-east-1.elb.amazonaws.com/wiki/index.php/Proportional_symbol_map], “Thematic Map Design” [http://wiki-1-1930356585.us-east-1.elb.amazonaws.com/wiki/index.php/Thematic_map]
- Financial Times Visual vocabulary [ft.com/vocabulary]