Government Statistical Service

Effective tables and graphs in official statistics

Guidance for producers

Second Edition
Version 2.0
January 2017

Contacts

Good Practice Team

Government Statistical Service

1 Drummond Gate

London SW1V 2QQ

Twitter: @GSSGoodPractice, @UKGSS

Email us: goodpracticeteam@statistics. gov.uk

There are more resources on this topic on the GSS website: http://bit.ly/goodpracticeresources

Acknowledgements

This is the second edition of "Effective tables and graphs in official statistics". It was first released in January 2015. We have updated it to include a new section on maps and to include other minor changes based on our experience delivering the associated course, feedback from colleagues and finding out more about data visualisation.

This guidance was developed in collaboration with the Office for National Statistics Data Visualisation Centre (ONS DVC). It was based on original source material developed by ONS DVC for the one-day course "Data visualisation – an introduction". The Good Practice Team would like to thank Alan Smith, Steve Rogers, John Nixon, Zoe Hartland and Rob Fry for developing the original content and advising and supporting us in writing this guidance.

We draw on material developed by Adam Little, Greg Wye and colleagues from NHS Digital in their document "Presenting information visually" and Full Fact's internal "Charts in brief" guide. Other source material (including web content written by other authors) is cited in the text and described in the references section at the end of the document. The guidance on tables was devised originally by Nicole Choong during her Fast Stream summer internship with the Good Practice Team, while the guidance on maps was devised and collated by Sylvia Bolton (née Kwan) during her 2016 secondment to the Good Practice Team.

Finally, we would like to thank colleagues from the GSS and wider government who have participated in the "Effective Tables and Graphs" course and provided feedback and constructive comments about this material.

Contents

Introduction

Who is this guidance for?

This guidance is primarily for producers of official statistics who must design data visualisations that are clear, consistent, informative and easy to use.

What is its aim?

The aim of the guidance is to ensure that the visualisations we produce convey key messages quickly and effectively.

The guidance is not a set of standards. Instead, it sets out some principles to think about when visualising and presenting statistical data and illustrates them with examples.

Why do we need it?

It is our responsibility to ensure that important patterns and trends in statistics are clearly described and easy to see.

This helps informed decision making and debate.

Statistical tools provide all sorts of techniques for graphing and tabulation. Some are useful, but many of these features can encourage bad practice. For example, it is simple to introduce 3D effects and shadows which reduce clarity and obscure the message. It is vital that we get the basics right.

What does it cover?

The guidance explores principles and approaches for the effective presentation of statistical data in tables, graphs and maps. It also discusses principles around the effective use of colour.

It brings together good practice from a range of existing sources and provides references for further reading.

While the guidance does not cover dynamic or interactive visualisation directly, the principles outlined here are also generally applicable in those contexts.

Don’t forget...

Think clearly and carefully about the information you are trying to convey when creating tables, graphs and maps. Consider the requirements of your audience. Choose and tailor your visualisation method accordingly.

“The use of tables and graphs to communicate quantitative information is common practice, yet few of us have learned the design practices that make them effective.”

Stephen Few

Show Me the Numbers [1]

The big picture

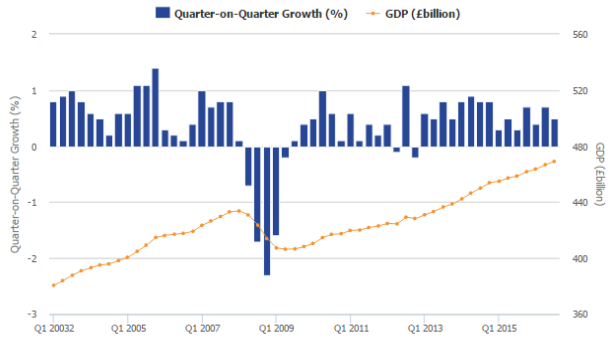
The UK Statistics Authority requires that producers of official statistics ensure that charts, tables and maps conform to good practice standards. Graphs and tables must be presented impartially, so that they avoid favouring a particular viewpoint.

We must also follow good practice in meeting requirements for accessibility. This ensures that the information we produce is helpful to the widest possible audience.

As members of the Government Statistical Service, it is our role to show people what the numbers mean – clearly and consistently. We must also think carefully about choosing appropriate visualisations to convey the messages in the numbers. Careful design underpins and supports this role.

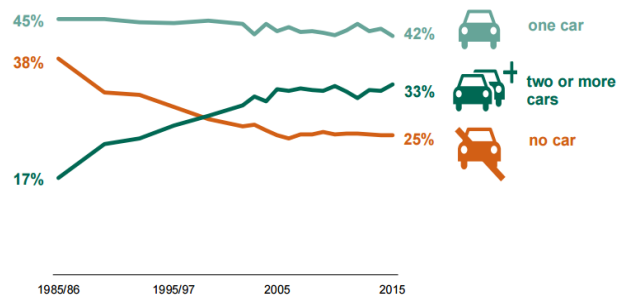
Figure 2: GDP (£ billions) and quarter-on-quarter growth[1], Quarter 3 (July to Sept) 2016

UK, 2003 to 2016



Source: Office for National Statistics

Household car availability: England 1985/86 to 2015 [NTS0205]



Principle 8 of the Code of Practice for Official Statistics requires that official statistics are readily accessible to all users. [14]

Designing graphs, tables and maps that enhance clarity, interpretability and consistency is essential to meeting this requirement.

Tables

Tables should be used to present numbers in a clear and systematic way.

Reference and demonstration tables

When we want to make statistics available for future reference, we supply them in a *reference table*. These typically have a large number of entries covering a wide variety of different statistics broken down into different categories. They are usually supplied away from the main commentary in an appendix or an accompanying spreadsheet. Because of this volume of information, it is important that the design of the tables allows the user to identify the right statistics with minimum effort.

If we are using a table to demonstrate a point that we are making in the text, we create a *demonstration table.* These use statistics, extracted from reference tables, laid out so as to quickly reinforce the point. The design should ensure that "the patterns, and exceptions should be obvious at a glance, at least once one knows what they are." [2], [3]

When to use a demonstration table instead of a graph

Messages may be illustrated using demonstration tables or graphs.

Use a graph when you want to show patterns, trends and relationships in the statistics, where the actual values are not required to make the point and all values share the same units.

Use a demonstration table:

- If you are asking the reader to compare individual values.
- If you want to include both the values and derived measures such as percentages or indices.
- If you want to include summary statistics such as means or totals.

"Nation sizes will look different in the future

Any differences in growth rate accumulate into significant changes. Note how the United States and the Soviet Union grow relative to Western Europe.

(Source: United Nations.)" [2]

Table: Predicted future population sizes of different countries
Table 2.4 from Samuelson's Economics

Percentages and millions

	Annual Growth (% per year)	1970	1980	1985
United States	*1.3*	205.0	226.0	240.0
United Kingdom	*0.6*	55.1	59.5	61.8
France	*0.8*	50.8	55.3	57.6
Soviet Union	*1.0*	243.0	271.0	287.0
Sweden	*0.7*	8.0	8.6	8.8
Italy	*0.8*	53.7	57.9	60.0
Japan	*1.2*	103.0	116.0	121.0

Adapted from Ehrenberg (1977) *Source: United Nations*

Tables

Presenting numbers for comparison

If you are inviting the reader to compare numbers, try to ensure that those numbers are physically close together.

If this is a series of numbers, it is simpler to make the comparison and discern patterns if the numbers are arranged in a column.

To help the reader make comparisons:

- use the same level of precision in each column,
- use commas to separate thousands, and
- right align the figures.

Decimal fractions less than one should always begin with a zero.

Rounding

Simplifying by rounding assists comparisons by making numbers easier to read and remember [4] [5]. In making decisions, we usually focus on the most significant digits. By presenting too much detail we can make things harder.

The extent of rounding will depend on the intended use: a commentator may be content to report that the population of the UK is 64m, or that this has changed from 63.7m to 64.1m. An analyst performing further calculations will want to work with more precise figures.

Rounding does reduce precision. This also usually means that the reported totals no longer equal the sum of the component parts. While demonstration tables should present suitably rounded numbers to illustrate the point being made, reference tables usually retain most or all of the precision so that users can decide on their own rounding.

Table: Formatting numbers for comparison and rounding to one decimal place or two significant figures

Quarter	Original	Rounded to one decimal place	Rounded to two significant figures
2013 Q1	5617.87	5,617.9	5,600.00
2013 Q2	.304	0.3	0.30
2013 Q3	12.292	12.3	12.00
2013 Q4	844	844.0	840.00
2014 Q1	5515.99	5,516.0	5,500.00
2014 Q2	.272	0.3	0.27
2014 Q3	10.867	10.9	11.00
2014 Q4	769	769.0	770.00

Making a decision on rounding can be difficult when the values show a variety of magnitudes (as in the table above). If so, consider rounding to a fixed number of (for example two) significant figures, as in the third column of the table, to enable rapid comparison, or similarly to two effective digits [6].

Tables

Grid lines

Grid lines can help to separate different parts of a table and group together related items. In this example, lines are used:

- to indicate where the body of the table starts and finishes;
- to separate the headings and bring together the columns for the two sets of statistics;
- to show that two of the columns refer to change rather than level.

Excessive use of grid lines clutters the page, confusing foreground and background objects and interrupting numerical comparisons.

Summary of latest Labour Market Statistics

Thousands and percentages

United Kingdom, May-July 2014 *Seasonally adjusted*

	Total people (thousands)			Rate (%)		
		Change on:			Change on:	
	May-Jul 2014	Feb-Apr 2014	May-Jul 2013	May-Jul 2014	Feb-Apr 2014	May-Jul 2013
Employed	30,609	74	774			
Aged 16-64	29,507	69	677	*73.0*	*0.1*	*1.4*
Aged 65+	1,103	4	97			
Unemployed	2,019	-146	-468	*6.2*	*-0.4*	*-1.5*
Aged 16-64	1,998	-138	-465			
Aged 65+	21	-8	-3			
Inactive	18,714	183	136			
Aged 16-64	8,930	114	-31	*22.1*	*0.3*	*-0.2*
Aged 65+	9,784	69	167			

Footnotes omitted *Source: UK Labour Force Survey*

White space

As well as separating groups, white space can be used to associate items together through continuity.

Here the age group-specific statistics are grouped with the headline statistics, but shown to be together at the lower-level through indentation.

Alignment

The rates on the right of the table are associated with particular age groups through alignment. To achieve this association, it is important that the columns are not too far apart. This ensures that there is no break in the implied continuity when reading across.

Grouping

Objects grouped together are assumed to be associated. Here the different measures are grouped in rows and the different types of estimate in columns.

Fonts

Different fonts can signal related items through similarity. Here, they indicate main headings and differentiate between estimated totals and rates.

Tables

Ordering categories

Ordering the categories in a table is a very effective way to aid rapid interpretation. In the example on the right, the table from page 7 is re-ordered to emphasise annual growth levels.

For some categorical variables, like time or age group, there is a natural order for presentation. Others may have standard or harmonised orderings. Use these whenever possible. An appropriate order may also be obvious from knowledge of the subject matter.

Alternatively, consider ordering categories according to the statistics in one of the columns. In our example, the rows are ordered using the summary column on the left with the largest value at the top. This shows the rankings of the categories on that statistic, and may also show where some of the statistics depart from the overall pattern.

Be aware when ranking a table by outcome that this deliberately emphasises the relative positions of the entries, even in cases where this is largely determined by random variation.

It's interesting to note that Samuelson didn't include Japan in his comparison of growth rates

Table: Predicted future population sizes of different countries

Ranked according to growth rate *Percentages and millions*

	Annual Growth (% per year)	1970	1980	1985
United States	*1.3*	205.0	226.0	240.0
Japan	*1.2*	103.0	116.0	121.0
Soviet Union	*1.0*	243.0	271.0	287.0
Italy	*0.8*	53.7	57.9	60.0
France	*0.8*	50.8	55.3	57.6
Sweden	*0.7*	8.0	8.6	8.8
United Kingdom	*0.6*	55.1	59.5	61.8

Adapted from Ehrenberg (1977) *Source: United Nations*

Positioning summary rows and columns

Summary rows and columns, particularly for totals, were traditionally placed at the bottom or right of the table. To help set the context for the subsequent statistics, it may be more helpful to place the totals at the top or left.

Shading

In reference tables, shading is sometimes used to indicate the degree of uncertainty in an estimate.

For demonstration tables, shading can be used to highlight the point being made in the commentary. We return to this point in the section on colour.

Tables

Titles and labelling

When designing a table, bear in mind that the reader may not read the commentary or that the table could be copied and used in another context.

We cannot be completely prescriptive about what should be included, but you should consider including the following information in the titles, labels, headings and footnotes accompanying the tables:

- Analysis units (people, households, enterprises)
- Types of statistics (totals, rates, means, etc)
- Units (thousands, km, £, etc)
- Classifications used to categorise
- Geographical or sector coverage
- Time periods
- Source of data
- Information about where statistics are not comparable, e.g. over time
- Where to find further guidance

Department for Transport statistics

Traffic (www.gov.uk/government/organisations/department-for-transport/series/road-traffic-statistics)

Table TRA0101

Road traffic (vehicle miles) by vehicle type in Great Britain, annual from 1949

Billion vehicle miles

				Other Vehicles			
	Cars and taxis	Light vans [1]	Goods vehicles [2]	Motorcycles	Buses & Coaches	Total [3]	All motor vehicles
1949	12.6	4.1	7.8	1.9	2.5	4.4	**28.9**
1950	15.9	4.8	6.9	2.7	2.5	5.2	**33.0**
1951	18.2	5.1	7.3	3.4	2.6	6.0	**36.6**
[rows omitted]							
1990	208.7	24.8	15.5	3.5	2.8	6.3	**255.3**
1991	208.3	25.9	15.2	3.4	3.0	6.4	**255.7**
1992	210.0	25.6	14.8	2.8	2.9	5.7	**256.1**
1993 [4]	210.1	25.8	15.1	2.3	2.9	5.2	**256.2**
[rows omitted]							
2013	240.0	42.6	15.7	2.7	2.8	5.5	**303.7**
2014 R	245.0	45.0	16.1	2.8	2.8	5.6	**311.6**
2015	247.7	46.9	16.7	2.8	2.7	5.4	**316.7**

1 Not exceeding 3,500 kgs gross vehicle weight, post 1982
2 Over 3,500 kgs gross vehicle weight, post 1982
3 Total of all other vehicles (i.e. motorcycles, buses, and coaches)
4 Data for 1993 onwards are not directly comparable with the figures for 1992 and earlier
R 2014 traffic estimates were revised as part of the production of 2015 statistics

Source: DfT National Road Traffic Survey
Last updated: May 2016
Next update: February 2017

Telephone: 020 7944 3095
Email: roadtraff.stats@dft.gsi.gov.uk
Notes & definitions (www.gov.uk/transport-statistics-notes-and-guidance-road-traffic)

The figures in this table are National Statistics.

Graphs

When to use a graph

Graphs are an excellent way to tell a memorable story or summarise something complex. They can also reveal insight that would otherwise be hidden if the statistics were only presented in a table.

Use a graph when you want to show patterns, trends and relationships in the statistics, where the actual values are not required to make the point.

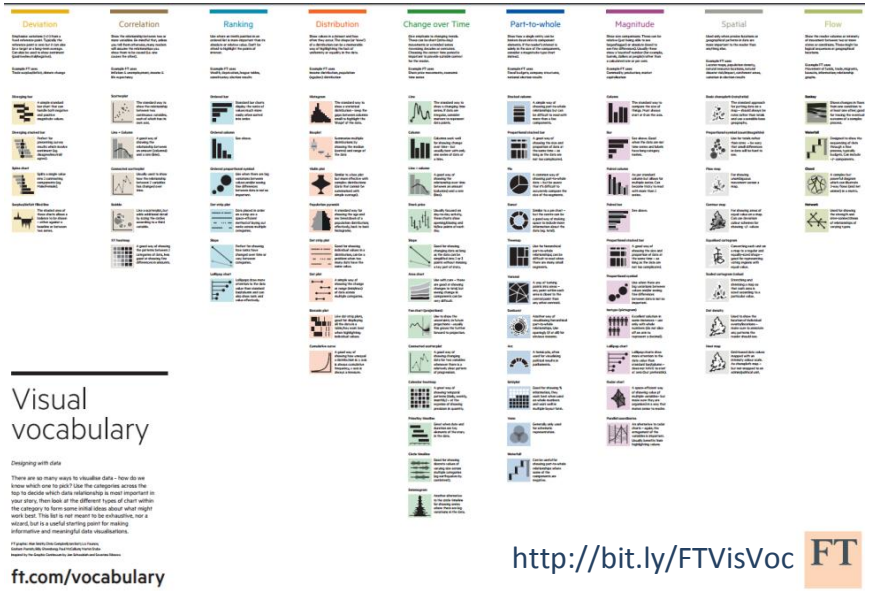
Choose the right graph for the job

What is the statistical story that you want to tell with your graph and what type of relationships does that involve?

We think the statistical relationships you may want to represent can be covered by nine categories [7].

- Magnitude, e.g. average income by region
- Time series
  e.g. price inflation shown over months
- Ranking
  e.g. schools ranked by performance
- Part-to-whole
  e.g. economic production by industrial sector
- Deviation
  e.g. rail company performance compared with target
- Distribution
  e.g. population by age group
- Correlation
  e.g. exploring relationship between weight and height)
- Spatial
  e.g. identifying geographical clusters of notifiable diseases
- Flow
  e.g. value of trade between countries

For any statistical relationship there are usually several graph options. The effectiveness will depend on both the statistical relationship and the particular values in your data. We recommend you try out different options, look critically at how well each graph works for your particular situation and pick the best one.

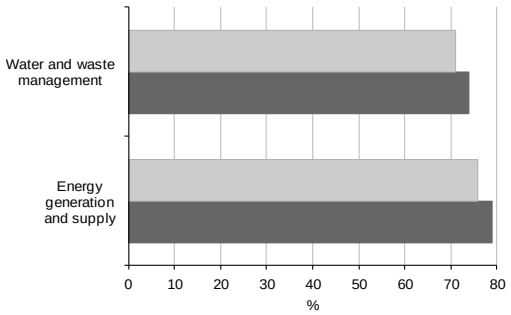


http://bit.ly/FTVisVoc

Graphs

Bar graphs

Bar graphs are versatile, used for comparing magnitude, showing time series, ranking, part-to-whole, deviation and distribution.

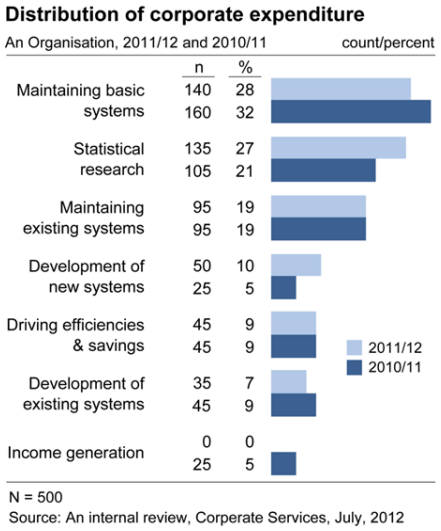


They can be used vertically or horizontally. Horizontal bar graphs are useful when you have long category labels that do not fit under vertical bars.

When you have clustered bars, as in the example above, the gap between the clusters should be slightly wider than a single bar. For non-clustered graphs, the gap should be slightly narrower than a bar.

If you find yourself labelling individual bar values, consider whether a table would be better. If you do add bar values, make sure they are well presented for comparison, as described in the tables section.

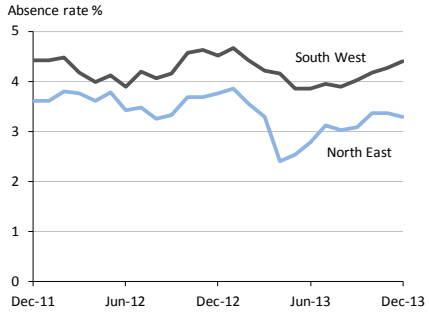
You could also try using *spark lines* [7]. This approach combines table text and a miniature graph, and can be very effective. Here is an example:



Line graphs

Line graphs are the default option for time series. They are also used to represent cumulative distributions and in a simple form as a slope chart to compare ranks.

When you can, label lines directly rather than using a legend. This reduces effort because readers do not have to look across the page. If a legend is unavoidable, place it prominently near the lines.

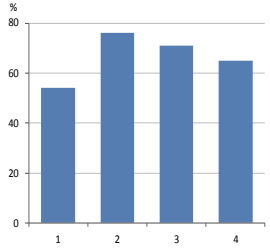


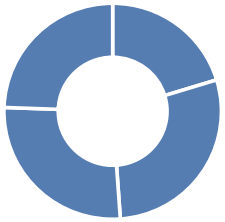
Line graphs with more than four lines are usually hard to follow, even if you vary line textures and point markers. **Consider using *small multiples*** ("panel charts", "lattice plots") to present many series in a line graph. See page 30 for an example.

Graphs

Pie charts

Pie charts (or donut charts) are designed for showing part-to-whole relationships, clearly indicating that the 'parts' add to the 'whole'.





As for many relationships, the bar chart is an alternative, but does not give the same immediate indication of the whole.

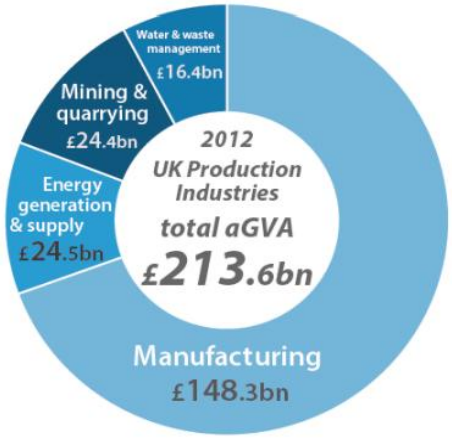
Use a bar graph for part-to-whole:

- To accurately show small variations across categories which aren't obvious in a pie chart (as in the example above)
- When there are more than four or five categories

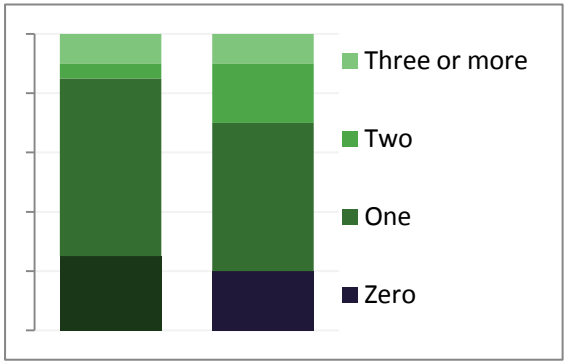
Use a pie chart or donut chart:

- If there is a dominant value in the statistics, which would dwarf the other categories in a bar graph
- If the variations shown are not significant
- To break up a page of bar graphs

Donut charts are an alternative to pie charts. The central space is a convenient place to show the value of the total, as in this example from ONS, or for an icon to indicate the category. It is sometimes argued that it is simpler to compare the sizes in a donut than in a pie chart.



You can also represent a part-to-whole relationship using stacked bars. For the sections at each end of the stacked bars, the size of the section and comparisons with neighbouring bars are clear, but for sections in the middle, without a common base, this is more difficult.



The impact of this will depend on whether the outcome variable is categorical or ordinal: for an ordinal variable, each level of the stack has a meaning and can clearly be compared.

Graphs

Set the scene

A graph should always make sense when seen on its own as it could potentially be copied and re-used elsewhere.

Title and source. Give your graph a meaningful title, so readers know what it shows, even if removed from its original context. Include the source of the statistics underneath your graph.

Axes. Label the axes so it is clear what the graph is showing. Horizontal labels are much easier to read and interpret than vertical or diagonal ones. In particular, it's usually best to place the vertical axis label on the top of the axis.

Annotations. Consider adding annotations to the graph, if this helps to tell the story. Work Programme Statistics [6] by the Department for Work and Pensions includes good examples of annotation, explaining some of the features of the series and avoiding misinterpretation. They also include a description of the overall pattern in an active title, above a more formal subtitle.

Focus on the story

Some graphing packages include numerous chart features by default and these can distract from the story you're trying to tell. Aim to simplify your graph, focusing on the story for your readers.

Maximise the data to ink ratio [8]

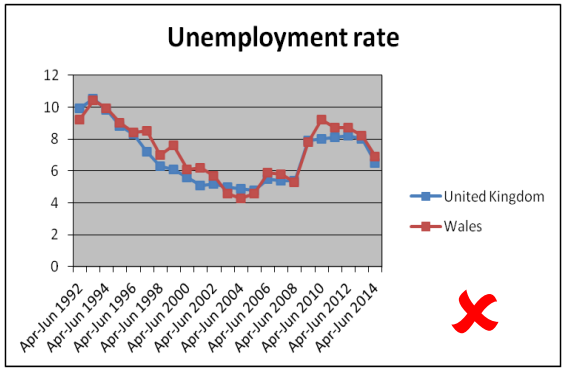
Remove anything from your graph which isn't required. Some common "chart junk" includes:

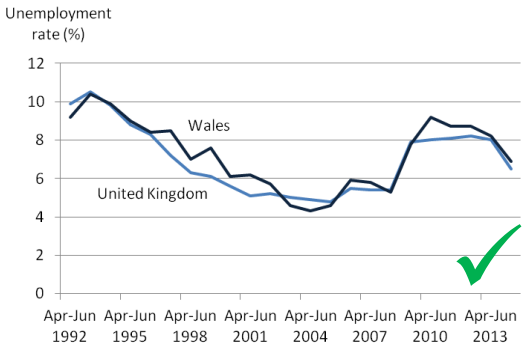
- legends
- shaded backgrounds
- borders
- patterns, textures and shadows
- 3D shapes
- data markers on line charts

Grid lines. Keep them to a minimum, so they don't clutter the graph. Multiples of two, five and ten work well for grid line increments. Make the grid lines pale grey.

Unemployment rate

Wales and the UK, 1992-2014





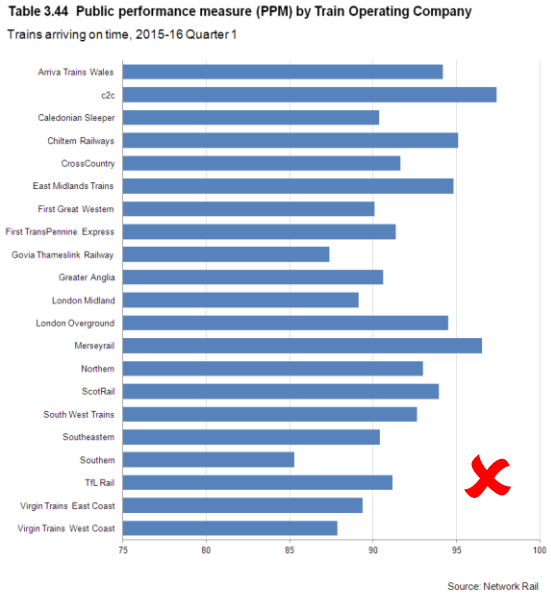
Source: Labour Force Survey

Graphs

Should the y–axis start at zero?

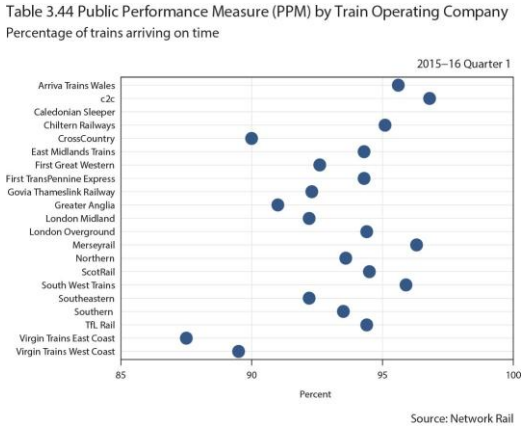
Including zero in the scale for the outcome variable can concentrate the main story in a small part of the plot area. Whether to break the axis to focus in on the story is a point of contention.

On bar charts…



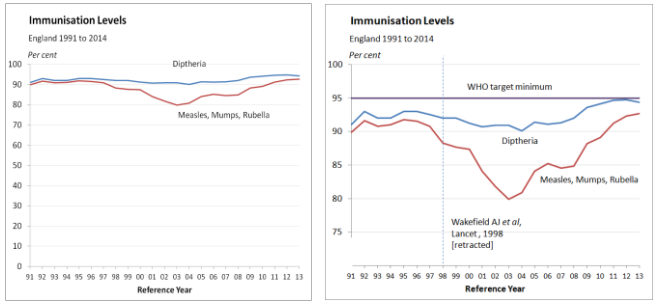
With bar charts, the length of bars gives an immediate indication of the relative magnitude of different quantities. Even if the axis is clearly labelled and a break signalled with a gap, this impression of a ratio scale is still conveyed.

So always include zero on bar charts and, if this prevents telling the story clearly, consider using an alternative like a dot plot, which removes the invitation to the reader to compare relative lengths.



On line charts…

In contrast to bars, lines do not give as strong an implication of a ratio scale. So, we think it is acceptable, with clear labelling, to break the axis here.



With line graphs, the most precise indication of the movement in the series is given when the lines are banked at around 45° [9]. You should also consider the quality of the statistics, to avoid focusing too closely on a volatile series. In some cases there will be a substantive point in the scale, such as a policy target or, for an index, the '100%' line that is important to include in the scale.

Finally, if you *are* describing a substantial relative change in a series, like a halving of a rate, it makes sense to include the zero in the chart to reinforce that point.

Graphs

Graphs should be the same size as a paragraph of text

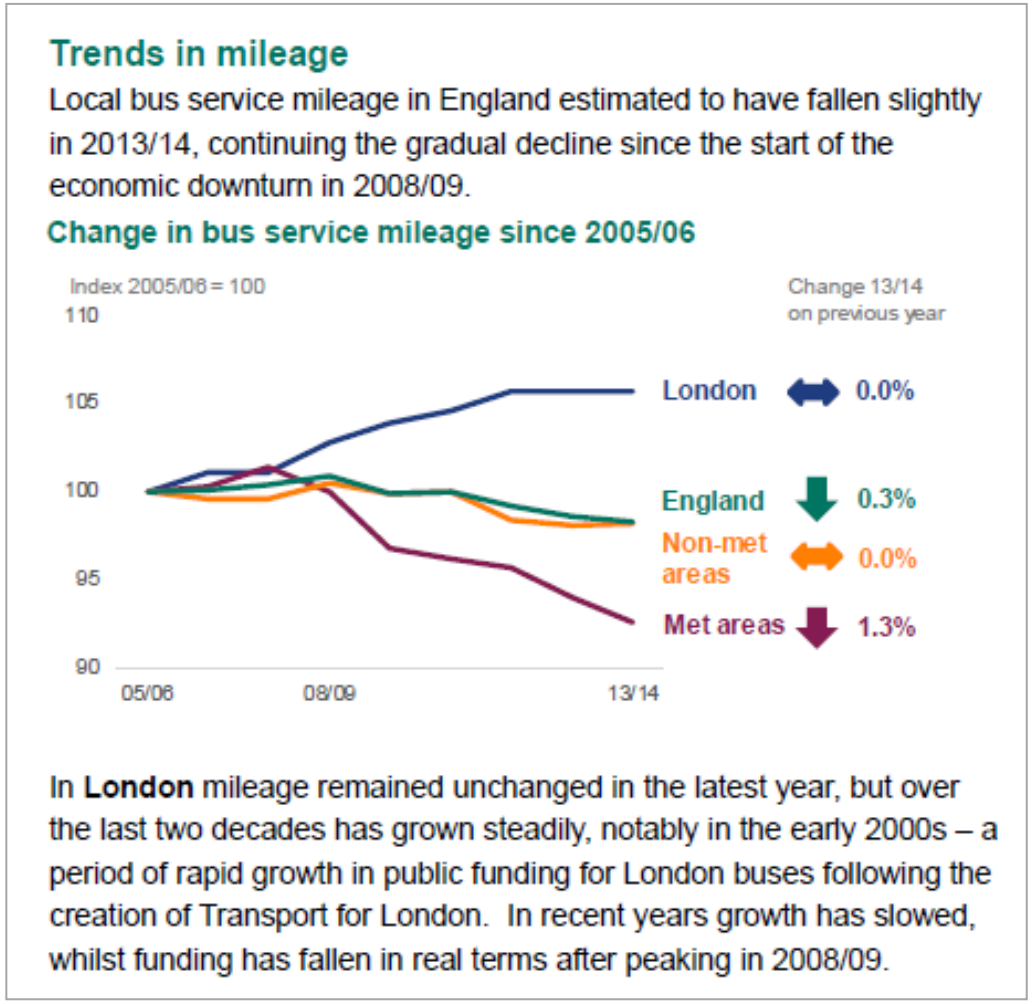
A graph that is about the same size as a paragraph of text becomes part of the natural flow of content.

When a graph is too big, it interrupts the eye's journey through the page. Over-sized graphs are perceived as being unprofessional and lack authority.

If a graph is too small the change to a smaller font disrupts the flow of text and it is hard to see what the graph is showing.

Graph text should be about the same size as the body text in the document, with the title a little larger, as in the example on the right.

The reader should be able to take in a graph at a glance. They should not need to click, scroll or enlarge a graph to view it.



Maps

Thematic maps are a powerful way of visualising geographical variation in official statistics. Before you start mapping, think about whether a map is the right tool for displaying your data. Maps are there to reveal geographical patterns. Success depends on the input data, the geographical areas that are relevant to your message and the level of granularity that you require.

Choropleth maps

Choropleth maps are the most common type of statistical map and the "go to" option in most cases. Statistics are linked to predefined geographical areas and values are represented by shading those areas in different colours.

Choropleth maps are appropriate for:

- Data that follow geographical boundaries.
- Continuous or categorical data.
- Standardised data like rates, percentages or densities, rather than counts or totals. Counts usually increase with area size, and can be misleading.

The level of geographical aggregation may impact on the message conveyed because aggregation can smooth out differences. The position of boundaries can also affect the message (a property that is abused in 'gerrymandering').

- Try the map at different levels of geographical aggregation to assess the impact on the patterns you see.
- Check that you have enough data for your chosen geographical breakdown - ensure that patterns are not just noise.
- Use appropriate measures. If the distribution within each area is skewed, use the median rather than the mean for averages.
- Explore other demographic data at your chosen geographic level to check how this may affect the patterns seen. For example, a map of cancer incidence may look very different once rates have been age-adjusted.
- Qualify your conclusions. Be clear about the geographical level used and the limitations associated with this. This flags to users that different trends may appear at other levels of aggregation.

Maps

Categorising the values

How many categories?

Aim to group your values into no more than six non-overlapping categories. It may be difficult to distinguish between colours if there are more.

Where should the breaks be?

This depends on the distribution of your data. Examine the distribution and consider the patterns you see and what works best to convey the important messages in your data.

Equal intervals divide the data into equal class sizes. Use it with normally distributed data, or to highlight the differences between a small number of areas and the rest. It is not very effective for representing skewed data.

Quantiles assign an equal number of observations per class. This produces a more balanced map for non-uniformly distributed data and works well in most cases. However, categories may have very different ranges and different values may be grouped together. Check that tied values are assigned to the same category.

Natural breaks minimise within-class difference and maximise variation between classes. Results are usually good, but be aware that class intervals are unlikely to be consistent across a series of maps.

Standard deviation measures distance from the mean. It is best used with data that are normally distributed, although the median and suitable quantiles can be used to build a more robust alternative.

Manual adjustments can make intervals more intuitive – but consider the impact on comparability across a series of maps. You might also consider inserting breaks at key target value, to show areas that fall above or below them.

How to deal with outliers?

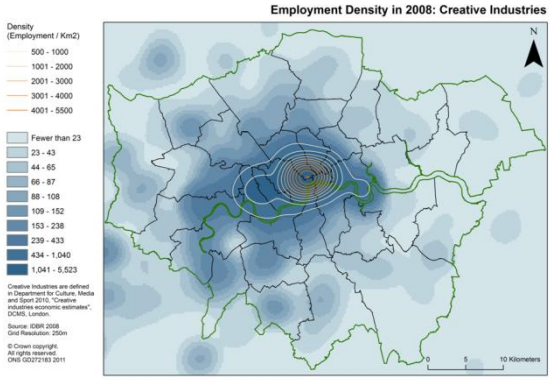
If they are important, outliers should be highlighted in a unique category. If they are simply noise, they can be hidden by widening a range.

Maps

There are other, more specialised mapping options. Here are three common ones.

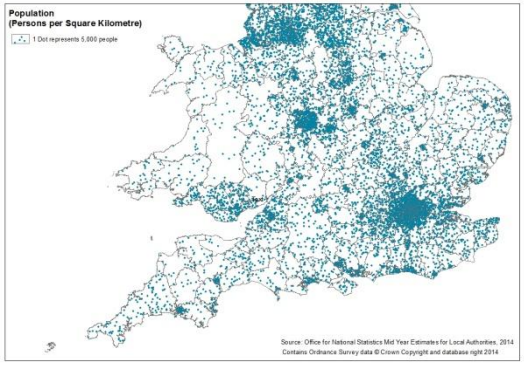
Isolines and grid shading

These maps are designed for visualising measures which vary continuously across space like density, temperature, or rainfall. They are normally based on a surface model built from point measures. They use lines of equal value ("isolines") or equally sized shaded grid cells to visualise density. The two techniques are sometimes combined to enhance clarity.



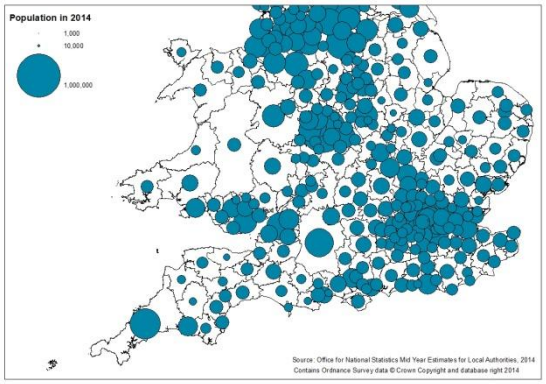
Dot maps

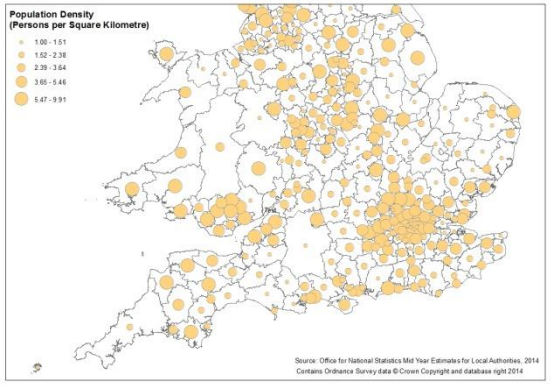
Dot maps can provide a quick visual impression of the density of observations in space. Each dot represents a fixed number of events. Getting this right requires careful experimentation. The size, positioning and number of dots can give very different impressions.



Proportional symbol maps

Here, symbols are placed in each area on the map (or at given data points) and sized according to the variable of interest. A variant (the second map on the right) is the graduated symbol map, where a fixed set of symbols depict the data classes.





Most readers find it hard to compare the relative sizes of proportional symbols, especially circles. Careful legend design can help, but a choropleth map or cartogram is a safer way to convey varying magnitude if your data are area-based.

Maps

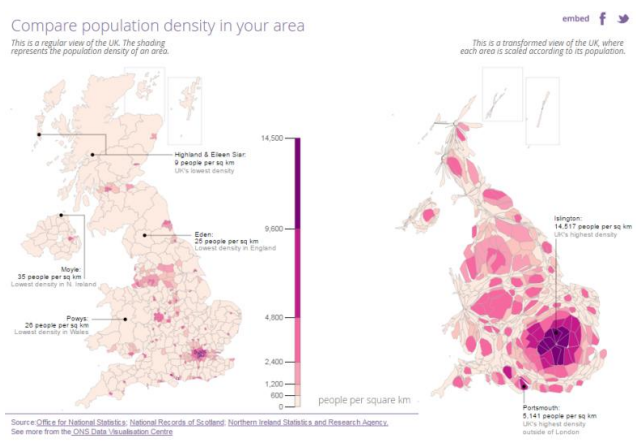
Cartograms

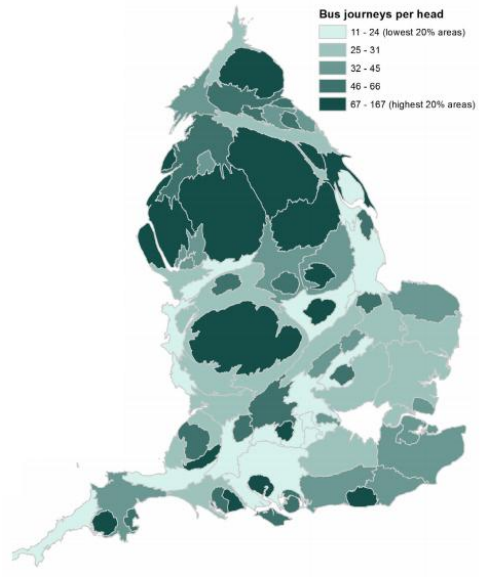
Cartograms distort map geometry to convey a specific message about the data. They can be very effective, especially if displayed interactively alongside a standard map, but are less familiar than other types and will require some supporting explanation.

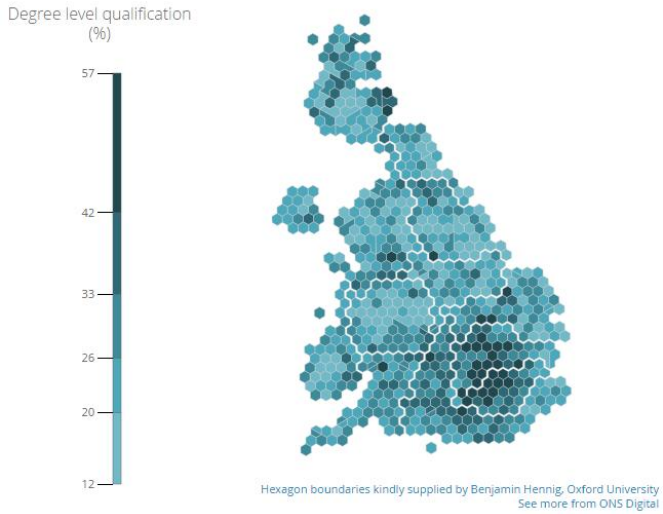
- Areas on cartograms will not be recognisable by their shape. If they need to be identifiable, use labels.
- A standard choropleth map should normally be displayed for reference alongside the cartogram in a static display.

Area cartograms (where the size of an area is proportional to its value) are useful for highlighting contextual information, particularly in small areas. However they cannot display areas with zero, negative, or missing values. The cartogram below showing bus journeys per head across the country is scaled by the number of journeys in each area. This gives a sense of the overall impact for the whole country.

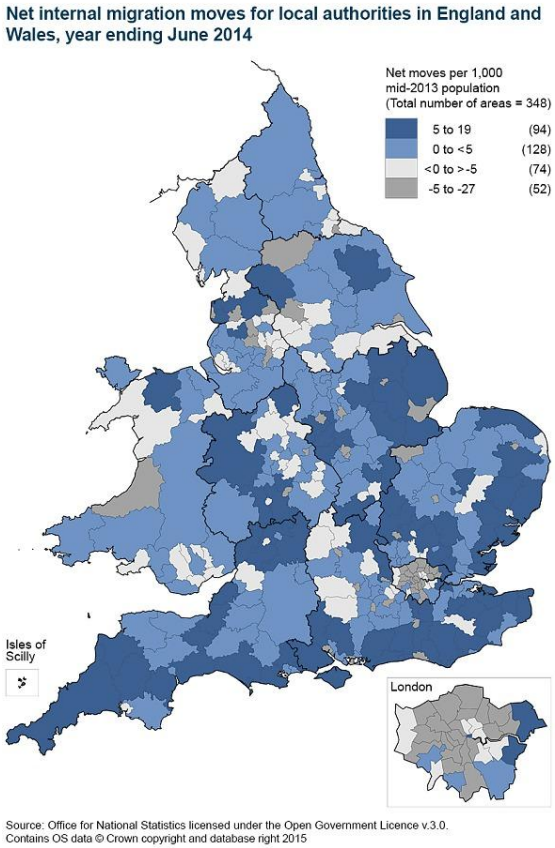
Equal area cartograms use equally sized units to represent areas. They are useful when areas are very different in size but the land area is irrelevant to the message. Larger areas dominate regular choropleth maps. Equal area cartograms solve this problem, but it is harder to identify specific areas. Appropriate labels may be needed.







Maps



Colour on maps

Use a single colour with dark shades for high values and lighter shades for lower values.

Use a diverging colour scale if the purpose is to show deviation from a value, such as a national average.

If there is no natural order to categories, use a different hue for each.

Avoid white as a foreground colour except to represent 0, “no data” or the centre of a diverging scheme.

Check that colours are distinct when viewing on the intended display (see **Colour** section).

Map projections

Map projections affect the size and shape of areas. Maps for printing should use the British National Grid projection and you should also use it when calculating densities or areas.

Legend

Use appropriately rounded numbers.

Show the number of areas contained within each category or include a bar chart or frequency distribution with the key. This is particularly helpful when large areas dominate.

Insets

Use an inset to distinguish small areas, or to highlight details for a particular region or provide context. Insets for Greater London are common, for example.

Labels

Include meaningful titles, footnotes, source statements and annotation, as with tables and graphs. Aim for your map to be self-contained. Include a copyright statement when required, e.g. if the map is based on Ordnance Survey information.

Colour

Colour can fundamentally change how we understand the information in graphs and tables. Colour used well can enhance and clarify statistical content. Colour used poorly will obscure, muddle and confuse [8].

This section sets out some key principles for working with colour in graphs and tables. It also provides examples of their application in practice.

Using colour: what users need

We add colour to make graphs, tables and maps more effective. Users need to be able to tell which colour is which (*identification)* and to tell the difference between distinct colours (*discrimination*).

The way the brain perceives colour can affect the ability to do this, as can the context in which the colours are used.

Colours are usually chosen because of a combination of three factors:

Graphic design. Illustrators may prefer particular colour combinations. House styles might use specific palettes.

Cultural context. Colours can have cultural associations. We react to these consciously and unconsciously and they vary widely across countries and groups. Recent research [9] shows that using colours that people would expect to see when representing familiar concepts (like blue for water) can improve the quality and speed of information processing. Counter-intuitive colours (red for grass) do the opposite.

Think about whether your choice of colours could have a cultural association, and, whether this is appropriate in the context of the information that you are presenting. Take care not to mislead.

Science. Biological and psychological knowledge can help us to design colour schemes that take account of how the human brain and visual system process information. This can really improve the usability of graphs and tables.

"Avoiding catastrophe becomes the first principle in bringing colour to information. Above all, do no harm."

Edward Tufte

Envisioning Information [17]

Colour

Use colour sparingly and with restraint

Never use colour to specify something on its own. People often print documents in black and white and some users are colour-blind.

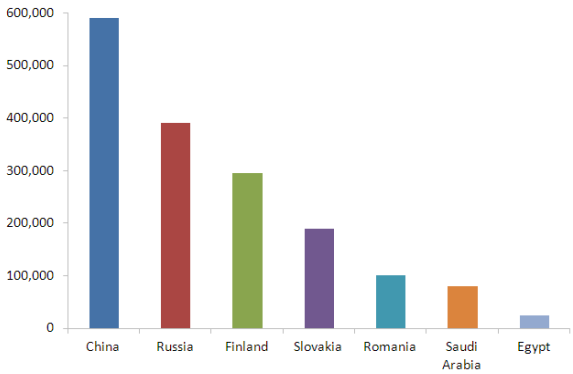
Use different colours only when they represent helpful differences of meaning in the data.

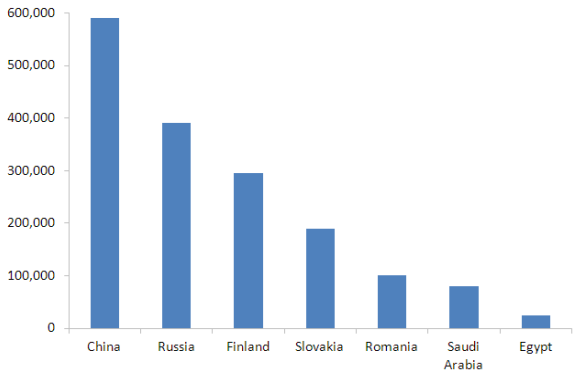
Colours are most effective when they are not overused. Limiting colour increases its impact by drawing on the brain's ability to highlight differences quickly.

Think carefully before you introduce additional colours into a table or graph. Do you really need them? Do they enhance the clarity of the message that you want to get across?

This example, adapted from Stephen Few's Perceptual Edge website [10] illustrates the point. The two graphs show the same data on sales levels by country for a product. The labels along the x-axis tell us which countries the bars represent.

The colours in this first graph add no information value, but their presence suggests that they do.





When people look at a data display like this one and see visual differences, they try to determine the meaning of those differences. Suggesting meanings which aren't there makes the reader waste time and effort trying to understand them.

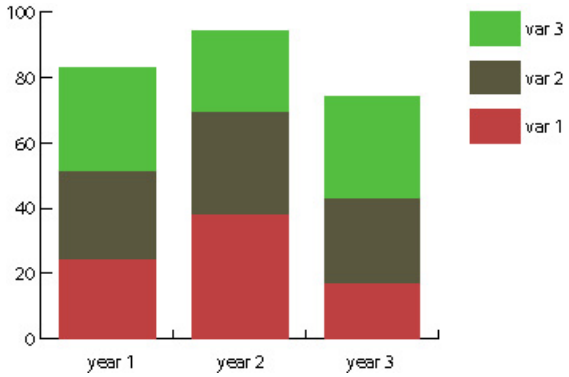
The second version of the graph, on the right, is more effective. The reader is much more likely to compare the bars when they look alike than when they look different.

Colour

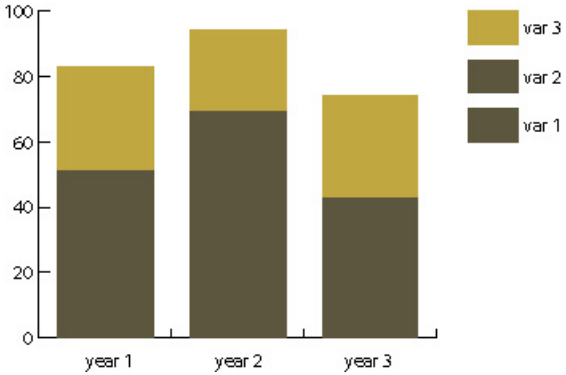
Consider accessibility

Colour blindness affects the ability to distinguish between some groups of colours, especially reds and greens. It affects about 1 in 12 men and 1 in 200 women – 4.5% of the UK population – with varying levels of severity [12].

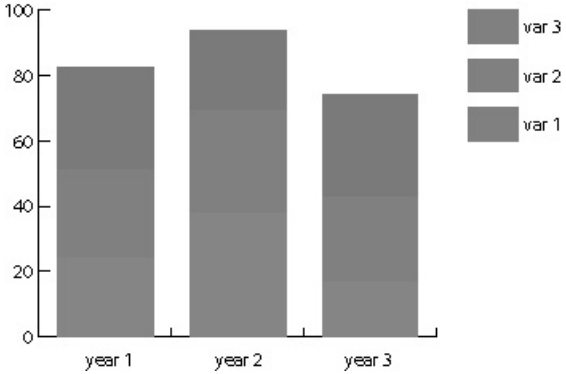
Consider this graph, which uses a red, green and grey colour scheme:



While the graph is reasonably clear for a reader with normal vision, it is much less effective for a reader with red-green colour blindness, as the next graph shows.

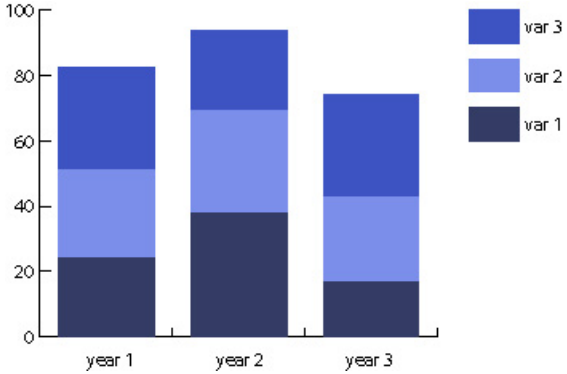


This scheme has the additional issue that it is not helpful to anybody in greyscale:

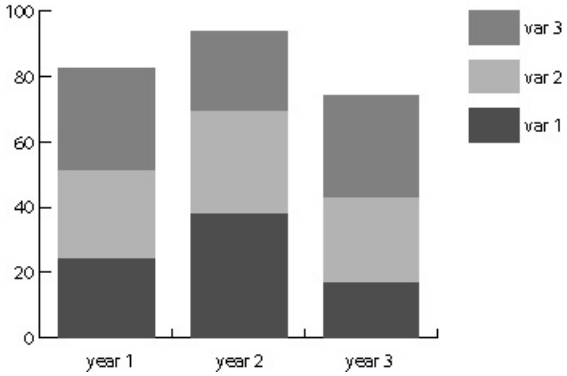


To minimise the impact on colour blind readers, avoid using greens and reds in the same display.

A safe starting point is a blue palette:



This one also works better in greyscale because the colours have been chosen to be optimally distinct from one another.



Colour

Defining colour digitally

Colours are represented digitally using several common schemes. For our purposes, the most useful of these is the HSL model [11]. HSL allows us to define colours uniquely using three properties which are fairly intuitive:

Hue

Hues are colours like red, blue or yellow. They are not generally perceived as having an agreed sequence, so readers may have difficulty in assigning a logical order to them. Small changes in hue are easy to detect – but colour blindness can have an impact here. This is discussed later.

Saturation (Chroma)

Saturation is the intensity of colour. It varies from grey or white (no saturation at all) to rich, glowing colour. Saturation is perceived on a continuous scale, but small changes are hard to detect.

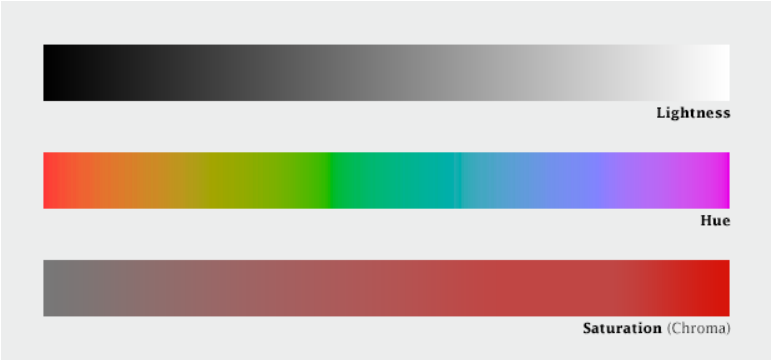
Luminance (Lightness)

Luminance is the brightness of colour. It is perceived as a continuous, ordered scale from dark to light. Changes in luminance are easy to detect, and humans can rank levels of lightness quite well unless the change is very subtle.

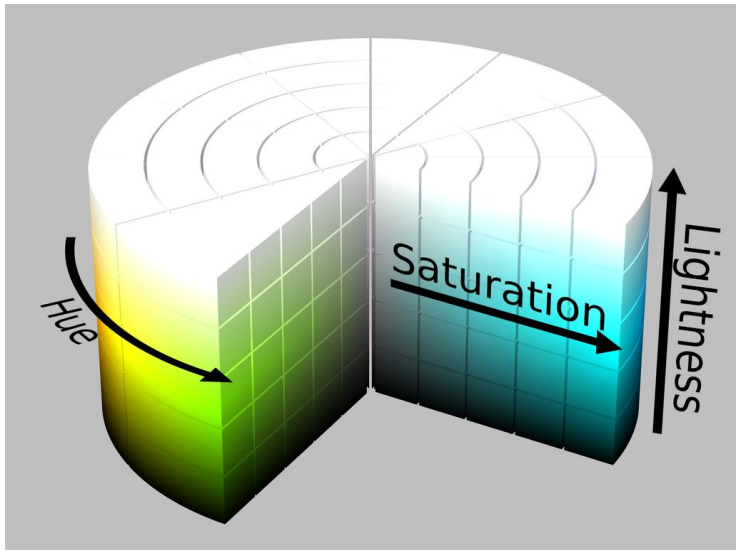
How we perceive luminance depends on hue. Consider these coloured squares:

Although they all have the same luminance and saturation, the yellow and green squares look lighter than the blue, red and purple ones. Colour palettes can be designed to adjust for this effect.

The diagrams on the right provide a visual representation of hue, saturation and luminance and how they are inter-related.



Source: Robert Simmon / NASA [11]

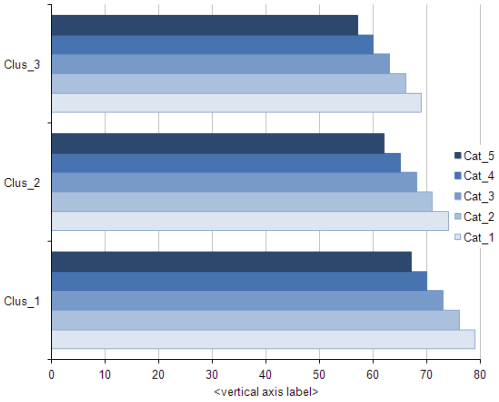


Source: Wikipedia [16]

Colour

Because luminance is perceived as having a natural order, it can help us to optimise colour schemes for maximum distinction and differentiation.

Consider this chart. It is easy to distinguish between the bars. The only changing colour parameter is luminance.

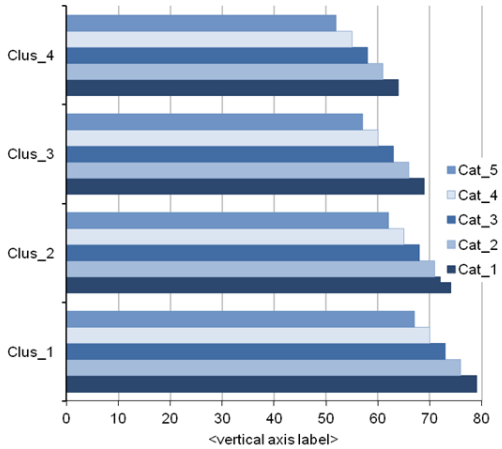


Here is the colour sequence from the bar chart above, with the percentage changes in luminance shown.

Changes in luminance of 10-20% are enough to distinguish shades in bar graphs, pie and donut charts. Changes of 30-40% are needed to achieve the same effect in line graphs.

Alternating colours

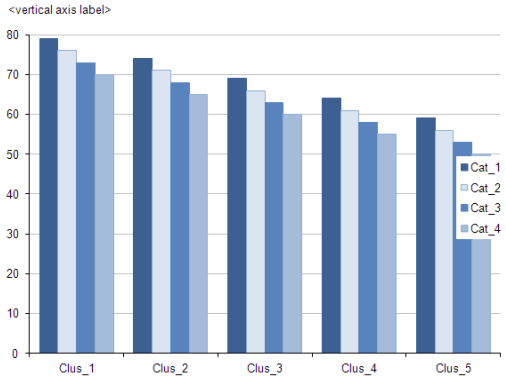
Consider alternating dark and light colours for categorical data to improve clarity. This graph uses the same palette as before, but alternates the dark and light colours to improve differentiation.



Shading edges to improve contrast

Edges can also enhance clarity. Using a dark tint for the edges of light bars makes them stand out more.

The dark bars below have the same outline and fill colour, while the light bars have darker outlines than the fill colour to improve contrast.



Colour

Do not overuse saturated colours

Use bold, saturated colours only when you want to draw attention to a specific piece of information, rather than for all of the colours in a graph.

Do not use saturated colours to highlight information in a table.

Bold, saturated colours have a powerful and dramatic impact. This can include unsettling visual side-effects. They may appear to glow for many readers, can generate after-images and their presence can affect how colours viewed subsequently or nearby appear.

Lots of saturated colour actually reduces impact and clarity. If all the colours in a graph are bold, this can destroy any logical visual hierarchy in the data.

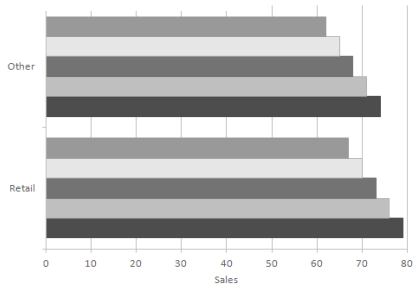
Mid to low levels of saturation are easy on the eye. High levels are bright and vibrant.

Consider the three graphs on the right. Hue and luminance are held constant as the colours move from no saturation at all in the top graph through to full saturation in the bottom graph.

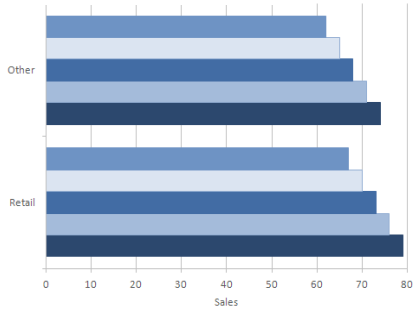
The third graph is actually quite uncomfortable to look at on a screen. Saturated colours like this are best left for highlighting key messages or to draw attention to small, hard to see elements like points on a graph. Avoid using them to cover large expanses in a graph or table.

For point and line graphs, experiment with colours of medium saturation to see if you can achieve an effective result before resorting to bold, saturated ones.

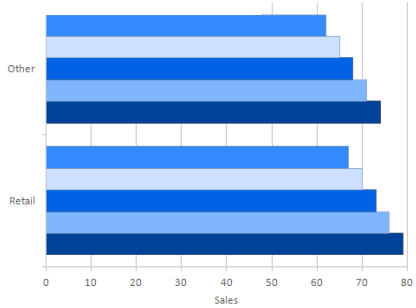
No Saturation



Medium Saturation



Full Saturation



Colour

Be consistent in your use of colour

Use the same colour to mean the same thing in a series of graphs.

Make sure that your use of colour is consistent and logical. Where possible, use colours that users would expect to see to represent familiar concepts.

Changing what colours represent in a sequence of graphs or tables increases the reader's cognitive workload. It can also cause them to mistake one data series for another, especially if skim reading.

Using unexpected colours to represent familiar concepts (such as red for grass) slows down information processing [9] and forces the reader to work harder. These effects are small and subtle, but do accumulate.

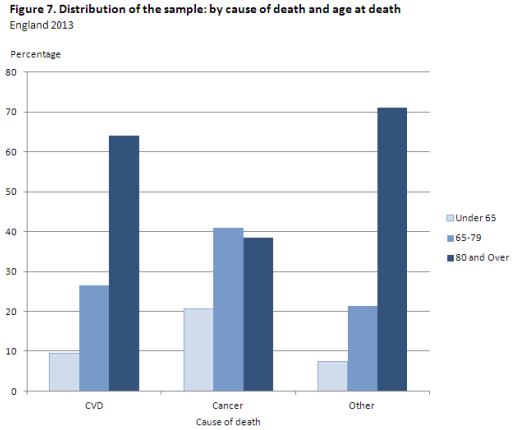
Use colour logically in sequences

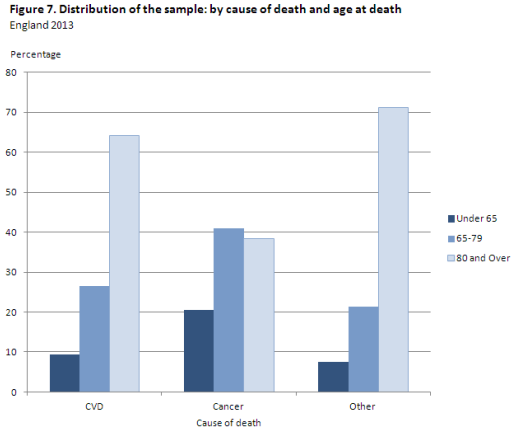
For sequences of colours, ensure that these progress in a way that the user would expect (e.g. in luminance order).

When representing a sequence, use a single hue (or small set of closely related hues) and vary lightness from pale colours to dark colours, rather than alternating.

This example uses a sequential blue palette to portray ascending age bands. The order of the colours is a matter of experimentation. A logical sequence here might be to go from light (young) to dark (old). However, the reverse means that the small bars are easier to pick out and the biggest are less overly dominant.

Use colours that are clearly distinct from one another given your choice of hue.





Colour

Colours on line graphs

We have already seen that graphs with more than four lines are often hard to follow, even with variations in line texture and shading. Introducing additional colour is unlikely to solve the problem.

The example on this page includes seven lines. We have used Stephen Few's optimised, multi-colour, medium saturation palette from the book "Show Me the Numbers" [1]. A palette like this is sufficient for four lines (although printing it in greyscale may still be problematic), but with seven lines, the graph is hard to follow. Even this palette is not very successful for so many lines.

A better approach to visualise five or more lines is to use a "small multiples" plot (also known as a lattice or panel chart) [13], which picks out variations in the different series at a glance. In this example, the UK data series appears in all of the plots for easy comparison with other countries.

Figure 3: Quarterly international manufacturing output

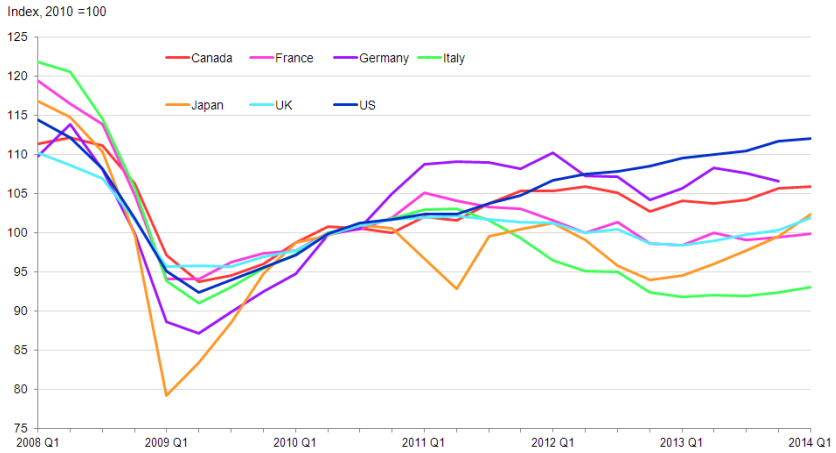
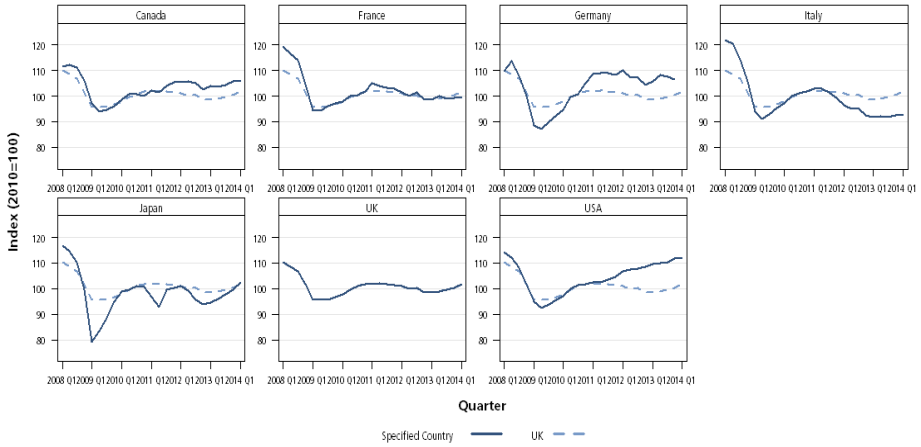


Figure 3: Quarterly international manufacturing output



Colour

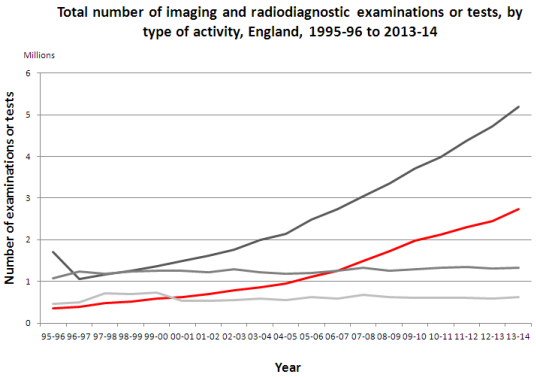
Colour and highlighting

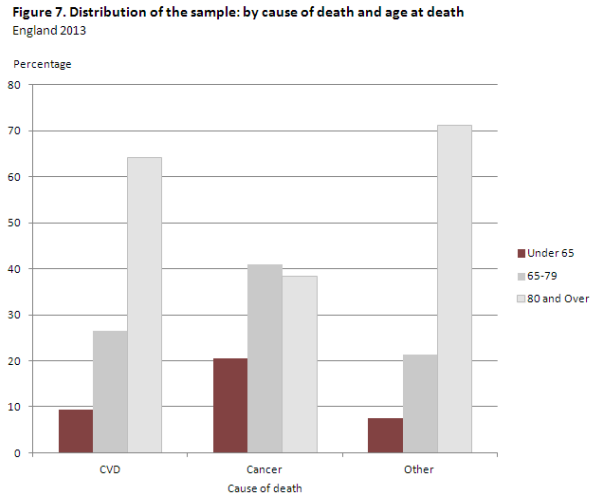
Colour can be used to highlight elements of graphs and tables to aid interpretation.

In graphs, use a distinct foreground colour to draw attention to specific features. Muted pastel or grey shades can be used to reduce the impact of the other elements in the graphic.

The graphs on this page use a red highlight colour, with muted greys for the other elements. Note the use of a more saturated red on the line chart than on the bar chart. Even here, it is not usually necessary to use very high levels of saturation to achieve helpful results.

Carefully chosen background colours can also be used to improve the clarity of tables by highlighting particular rows or columns. Use subtle shades rather than bold, saturated ones for highlighting in tables.





SomeData: By SomeCategory

SomePlace, 2011 SomeUnit

	Cat_1	Cat_2	Cat_3	Cat_4	Cat_5
Variable_1	22.7	56.3	98.5	33.6	87.4
Variable_2	19.6	65.2	78.2	78.9	65
Variable_3	23.7	23.9	12.7	55.2	44.2
Variable_4	12.9	98.5	32.6	25.1	23.9
Variable_5	67.9	73.1	78.3	12.8	17.4
Variable_6	23.9	45.2	93.2	83.8	45.1
Variable_7	45.8	21.8	77.3	87.2	98.9

1 Some footnote about something

Source: SomeSurvey from somew here

The following colours and mixtures of colours work well for this purpose:

- Grey
- Blue
- Grey with any one of blue, purple, red, pink or orange
- Blue with any one of purple, red, pink or orange

Unless you use very light shades green, cyan and yellow should be avoided. Remember not to mix greens and reds.

Don't overdo highlighting in tables. It is best to restrict this to one or two columns.

Colour

Background colour

Effective use of colour applies as much to graph annotation and background as it does to data elements like bars and lines.

Use a white background

Most colour palettes are designed to appear on a white background.

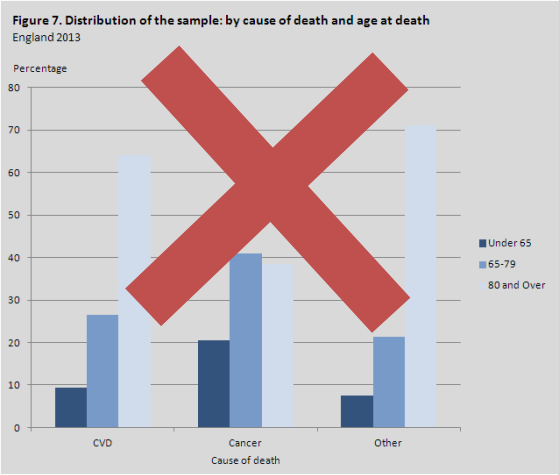
Human vision adopts colour perception relative to the local definition of white. A white background provides a helpful reference “anchor” for the visual system.

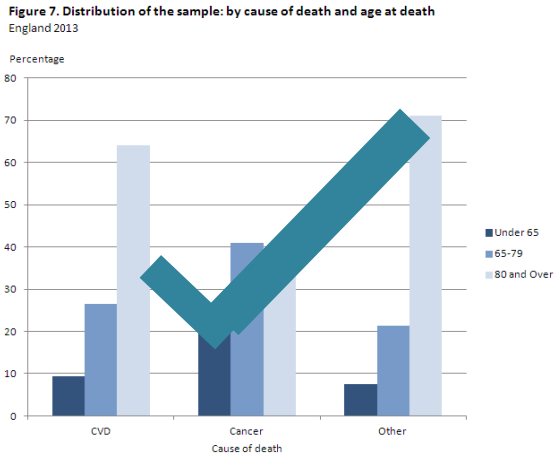
The only functional reason to use a non-white background is for viewing the image in the dark. The use of modern digital projectors, which work well under normal lighting conditions [8], make this issue is largely irrelevant today.

- Confine use of colour to foreground items in graphs. Always use white for background. Use grey palettes for drawing and labelling axes and annotation.
- In general, background colour should be avoided completely in tables unless it is to provide subtle highlighting in a limited subset of cells.
- Avoid using white as a foreground colour in graphs. It should also be avoided on maps unless it represents 0, “no data” or the centre of a diverging distribution.

A word about images

- Never use images as backdrops in graphs or tables. These simply distract the reader and make it more difficult to pick up the important messages.
- Maps sometimes include backgrounds such as aerial photography or Google Map data to provide context. If you do this, take care to ensure that the messages of your map are not obscured or compromised by the additional complexity of the background.





References and resources

References cited in the text

[1] S. Few, Show Me The Numbers - Designing Tables and Graphs to Enlighten (2nd. Edition), Burlingame, CA: Analytics Press, 2012.

[2] A. Ehrenberg, “Rudiments of numeracy,” *Journal of the Royal Statistical Society, Series A,* vol. 140, no. 3, pp. 277-297, 1977.

[3] M. Chapman and C. Wykes, Plain Figures, London: HMSO, 1996.

[4] J. Anseau, “Statistical literacy guide: rounding and significant places,” July 2007. [Online]. Available: http://www.parliament.uk/briefing-papers/sn04443.pdf. [Accessed 11 December 2014].

[5] S. Bigwood and M. Spore, “Rounding figures for clarity,” in *The designer's guide to presenting numbers, figures and charts*, New York, Allworth Press, 2013, pp. 16-21.

[6] E. Swires-Hennessy, Presenting Data: How to Communicate Your Message Effectively, Chichester: John Wiley & Sons, 2014.

[7] A. Smith, “Simple techniques for bridging the graphics language gap,” Financial Times, 17 August 2016. [Online]. Available: https://www.ft.com/content/304419ec-63a3-11e6-8310-ecf0bddad227. [Accessed 16 11 2016].

[8] E. R. Tufte, *The visual display of quantitative information*, Connecticut, Graphics Press, 2001.

[9] M. Stone, “Choosing colors for data visualization,” 17 January 2006. [Online]. Available: http://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf. [Accessed 25 September 2014].

[10] S. Lin, J. Fortuna, K. Chinmay, M. Stone and J. Heer, “Selecting semantically resonant colours for data visualisation,” 2013. [Online]. Available: http://idl.cs.washington.edu/files/2013-SemanticColor-EuroVis.pdf. [Accessed September 2014].

[11] S. Few, “Practical rules for using color in charts,” February 2008. [Online]. Available: http://www.perceptualedge.com/articles/visual.../rules_for_using_color.pdf. [Accessed September 2014].

[12] R. Simmon, “Subtleties of color,” 5 August 2013. [Online]. Available: http://earthobservatory.nasa.gov/blogs/elegantfigures/2013/08/05/subtleties-of-color-part-1-of-6/. [Accessed 5 October 2014].

[13] Colour Blind Awareness, “Colour Blind Awareness,” [Online]. Available: http://www.colourblindawareness.org/colour-blindness/. [Accessed September 2014].

[14] E. R. Tufte, “Small multiples,” in *The visual display of quantitative information*, Connecticut, Graphics Press, 2001, pp. 168-170.

[15] UK Statistics Authority, Code of Practice for Official Statistics, London: UK Statistics Authority, 2009.

[16] J. Camoes, “Is your chart really answering your question?,” 24 November 2008. [Online]. Available: http://www.excelcharts.com/blog/2008/11/. [Accessed 15 08 2014].

[17] Wikpedia, “HSL and HSV,” Wikipedia, 28 November 2014. [Online]. Available: http://en.wikipedia.org/wiki/HSL_and_HSV. [Accessed 3 December 2014].

[18] E. R. Tufte, Envisioning information, Connecticut: Graphics Press, 1990.

[19] Department for Transport, “Annual Bus Statistics: England 2013/14,” Department for Transport, London, 2014.

[20] A. Little and G. Wye, “Presenting information visually,” Health and Social Care Information Centre, Leeds, 2014.

[21] Department of Work and Pensions, “Work Programme Statistics,” 6 October 2014. [Online]. Available: https://www.gov.uk/government/statistics/work-programme-statistical-summary-data-to-30-june-2014 . [Accessed 16 October 2014].

Other useful material

General references

Bigwood, S., and Spore, M., 2013, The designer's guide to presenting numbers, figures and charts, Allworth Press, New York.

Camoes, Jose, Excel Charts blog, http://www.excelcharts.com.

Cleveland, W.S. and McGill, Robert, Graphical perception: theory, experimentation and application to the development of graphical methods, Journal of the American Statistical Association, Vol. 79, No. 387, pp. 531 – 554, 1984.

Evergreen, Stephanie D., 2014, Presenting data effectively – communicating your findings for maximum impact, Sage Publications.

Few, Stephen, 2013, Information Dashboard Design, Analytics Press.

Few, Stephen: Perceptual Edge website: http://www.perceptualedge.com.

Government Statistical Service: "Communicating Statistics" web resources: https://gss.civilservice.gov.uk/statistics/presentation-and-dissemination/.

McCandless, David: Information Is Beautiful website: http://www.informationisbeautiful.net/

Office for National Statistics, Infographic Guidelines, Version 1.0, https://gss.civilservice.gov.uk/wp-content/uploads/2012/12/infographic-guidelines-v1-0.pdf

Office for National Statistics: The Infographic Superhighway blog https://blog.ons.digital/2013/10/29/the-infographic-superhighway/

Office for National Statistics: Style.ONS guidance on data visualisation: http://style.ons.gov.uk/category/data-visualisation/

Robbins, Naomi B., 2005, Creating more effective graphs, Wiley.

Schwabish, J., and Ribecca, S., 2014, The Graphic Continuum, https://policyviz.com/2014/11/11/graphic-continuum-desktop-version/

Swires-Hennessy, E., 2014, Presenting data: how to communicate your message effectively, Wiley.

Tufte, E.R., 1990, Envisioning Information, Graphics Press.

United Nations Economic Commission for Europe, Making data meaningful Part 2: A guide to presenting statistics, available from http://www.unece.org/stats/documents/writing/ .

Visualising data website: http://www.visualisingdata.com/

Yau, N., Flowingdata.com website: http://www.flowingdata.com

Yau, N., 2011, Visualize this: The FlowingData guide to design, visualisation and statistics, Wiley.

Yau, N., 2013, Data points: visualisation that means something, Wiley.

On colour

Brewer palettes website: http://mkweb.bcgsc.ca/brewer/

Colorbrewer website: http://colorbrewer2.org

Colour usage guidance and tools at NASA Ames Research Laboratory: http://colorusage.arc.nasa.gov/guidelines.php

Colour FAQ: http://www.poynton.com/PDFs/ColorFAQ.pdf

I Want Hue website – colours for data visualisation: http://tools.medialab.sciences-po.fr/iwanthue/

Kosara, R., 2013, How the rainbow color map misleads, available at http://eagereyes.org/basics/rainbow-color-map