DATA VISUALISATION

A practical guide to producing effective visualisations for research communication

Rebecca Wolfe, 2014

Research Uptake Manager, RESYST Consortium

London School of Hygiene & Tropical Medicine

Email: Rebecca.wolfe@lshtm.ac.uk
Twitter: @RESYSTresearch

DATA VISUALISATION

A practical guide to producing effective visualisations for research communication

CONTENTS

1. Introduction to data visualisation

1.1 What is data visualisation?

Data visualisation is the presentation of data or information in a graphical format. It is a way of visually communicating information – often quantitative in nature – in an accurate and compelling format.

For centuries, people have used visual representations such as charts, maps and diagrams, to understand and communicate information. One reason for this is that the brain can process and retain visual information more easily than verbal information (words or numbers). We see immediately, with little effort. Thinking (cognition) is much slower and less efficient.

Data visualisations, ranging from simple line charts to elaborate interactive maps, have become an increasingly popular approach used by researchers to communicate complex information and reach audiences outside the research community. Effective visualisations can communicate large amounts of information clearly, and have the potential to alter perceptions, influence people and bring about change.

1.2 The value of data visualisations

- Visualisations can make sense of large amounts of data by revealing patterns and relationships.

- Visualisations help to communicate information and patterns to others: they can explain data using a small space, and can be easily shared with others.

- Visualisations can tell stories in a compelling and immediate way which can be more memorable than words.

- Effective visualisations can alter perceptions, influence people and bring about change.

The use of data visualisations to communicate public health knowledge and influence change is not a recent phenomenon, as shown in the examples below:

**John Snow's map of deaths from Cholera in London**

In 1854 John Snow mapped deaths from a Cholera outbreak in central London to show its cause. The map (marked with dots for a death, and crosses for water pumps) showed that cases were grouped around a water pump on Broad Street.

As a result, the handle of the contaminated pump was removed and further deaths were prevented. His work, promoting the idea that disease can be spread through contaminated water, was a major turning point in the history of public health.

Florence Nightingale's polar area diagram of the causes of mortality during the Crimean War

In 1858 Florence Nightingale published a diagram showing what soldiers in the British army died of at field hospitals during the Crimean War. The graphic shows the number of deaths that occurred from wounds (in red), from preventable diseases (in blue) and from other causes (in black).

Florence Nightingale showed that the biggest cause of death was the result of preventable diseases, and successfully used this to advocate to the government for better medical and sanitary conditions for soldiers.
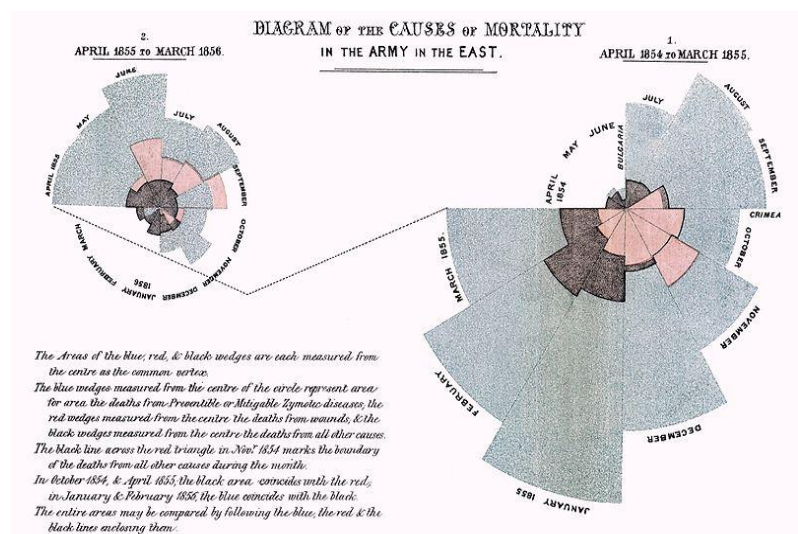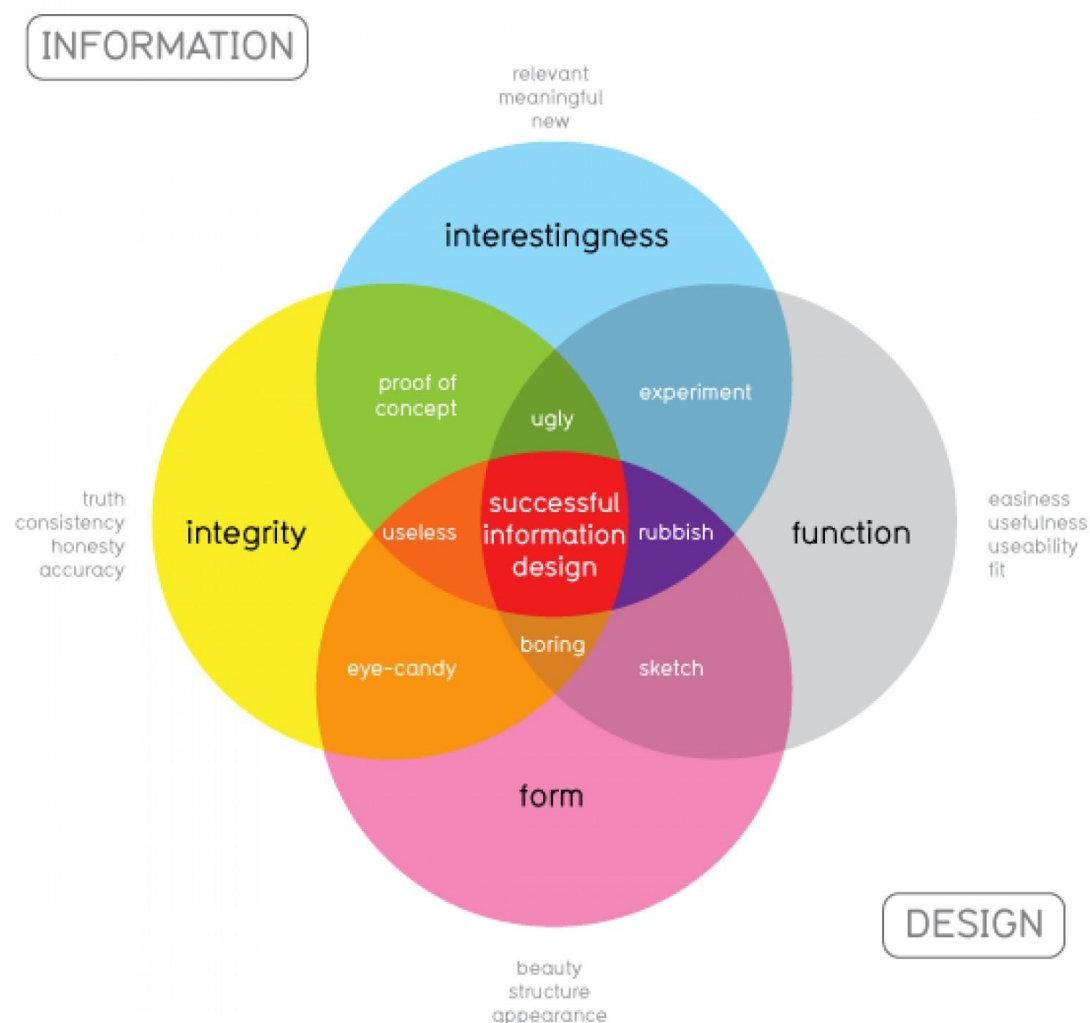
2. Planning an effective visualisation

According to David McCandless (visualisation guru), good visualisations share several key features:

1. Function: they let you see trends and patterns clearly.
2. Form: they are visually appealing and well structured to attract readers and hold their attention.
3. Integrity: they portray the data accurately and honestly.
4. Interesting: they are relevant and meaningful, or reveal new information.

What Makes Good Information Design?

INFORMATION

relevant
meaningful
new

interestingness

proof of
concept

ugly

experiment

truth
consistency
honesty
accuracy

integrity

useless

successful
information
design

rubbish

function

easiness
usefulness
useability
fit

eye-candy

boring

sketch

form

DESIGN

beauty
structure
appearance

David McCandless // v1.0 // Nov 09
InformationIsBeautiful.net

With the many free and low-cost tools that are now available on the internet, artistic talent is not essential to produce a good visualisation and interactive visualisations don't necessarily require programming skills. However, good visualisations do require visual thinking and thorough planning.

2.1 Start with the objective and the audience

Identify the purpose of the visualisation:
- What story do you want to tell?
- Who do you want to reach and what does your audience care about?
- What do you want your audience to do after seeing the graphic?

2.2 Explore and select the data

- What datasets are relevant to the story?
- What are the interesting patterns and trends that should be represented and emphasised?
- Look for outliers or patterns that don't follow your expectations
- What do you think the readers will try to do with the visualisation?
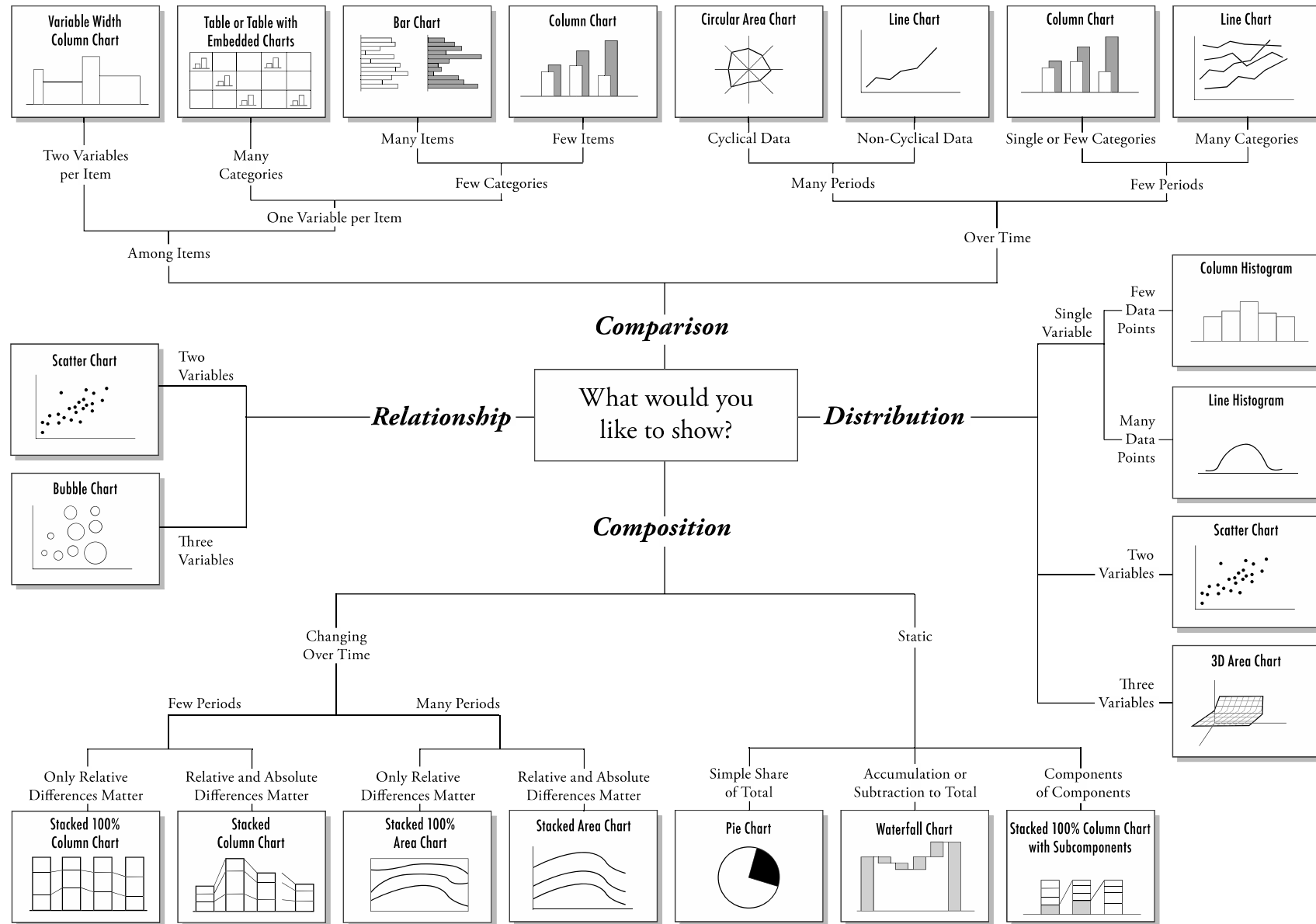
2.3 Choose the right type of chart

Data visualisations will only be successful if they communicate the data easily, efficiently and accurately, and this requires choosing the right type of chart.
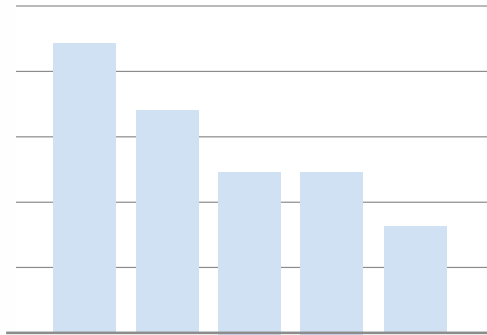
The right chart should:

➢ Clearly indicate how values relate to one another.

➢ Represent the quantities accurately.

➢ Make it easy to compare the quantities.

➢ Make it easy to see the ranked order of values.

➢ Make it obvious how people should use the information.

Source: Stephen Few, 2013

Chart Suggestions—A Thought-Starter

Variable Width Column Chart

Table or Table with Embedded Charts

Bar Chart

Column Chart

Circular Area Chart

Line Chart

Column Chart

Line Chart

Two Variables per Item

Many Categories

Many Items

Few Items

Cyclical Data

Non-Cyclical Data

Single or Few Categories

Many Categories

Few Categories

One Variable per Item

Many Periods

Few Periods

Among Items

Over Time

Comparison

Scatter Chart

Two Variables

Single Variable

Few Data Points

Column Histogram

Relationship

What would you like to show?

Distribution

Many Data Points

Line Histogram

Bubble Chart

Three Variables

Two Variables

Scatter Chart

Composition

Three Variables

3D Area Chart

Changing Over Time

Static

Few Periods

Many Periods

Only Relative Differences Matter

Relative and Absolute Differences Matter

Only Relative Differences Matter

Relative and Absolute Differences Matter

Simple Share of Total

Accumulation or Subtraction to Total

Components of Components

Stacked 100% Column Chart

Stacked Column Chart

Stacked 100% Area Chart

Stacked Area Chart

Pie Chart

Waterfall Chart

Stacked 100% Column Chart with Subcomponents

Column/bar charts

Pros

- Clear and easy to read.
- Represent quantities easily and accurately.
- Ordering bars by value makes comparison easier.

Cons

- Can be dull.
- Difficult to do anything visually interesting with shape or colour without destroying clarity.
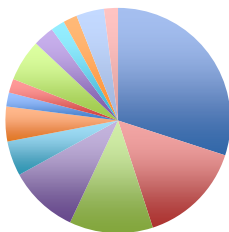
Pie charts

Largest segment starts at 12 o'clock
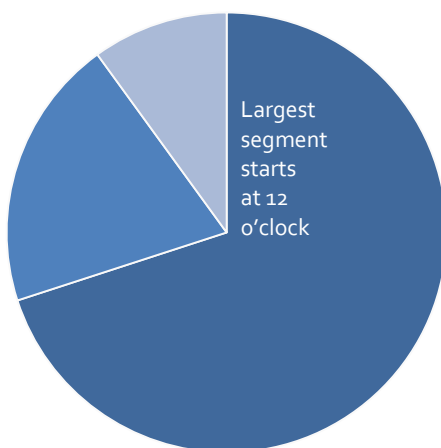
Pros

- Pie charts are good for showing percentages.
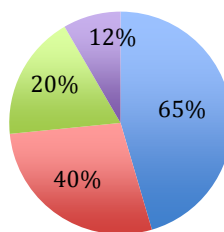- They are most effective when there is one dominant value.

Cons

- Don't represent values accurately, which makes it hard to compare values with in a chart or between charts.
- Simple or stacked bar charts may be a more effective way to show portions of a whole

Examples of bad pie charts

Too many values (and colours)   Values don't add up to 100   Too many similar values

12%

20%   65%

40%
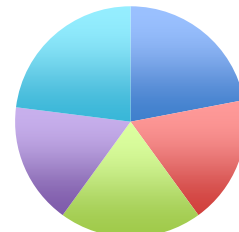
➢ Don't show more than 4 values using a pie chart (group the rest into "other" category).
➢ Consider stacked bar charts as an alternative to pie charts.

3. Design tips for your visualisation

3.1 Represent the data accurately

When producing a graphic, it is important to ensure that it tells the truth about the data. It is possible to (unintentionally) produce a misleading or ambiguous graph.
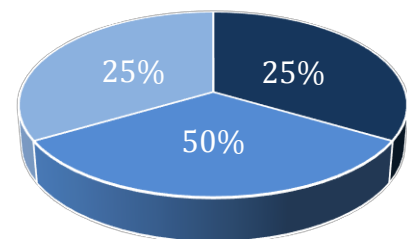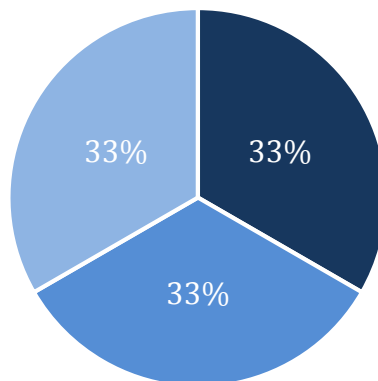
1. Always scale to zero

Bar/line charts without a zero baseline are misleading as they exaggerate differences between values.

These two graphs show the same data.
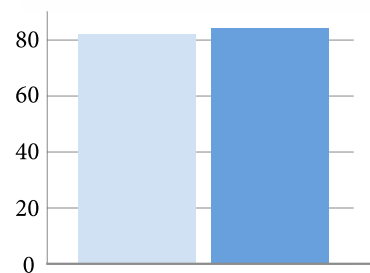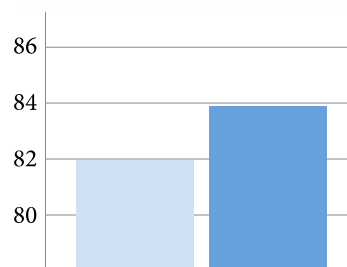
2. Don't use 3D charts

3D charts distort the data by giving prominence to the section closest to the front, and distract viewers from the actual values.

3. Circle size by area, not radius

Incorrect proportion based on radius (rather than area) can lead to differences being exaggerated.

The relative size of these pies is based on their radii. In this example, the largest circle is actually 16x the size of the smallest circle
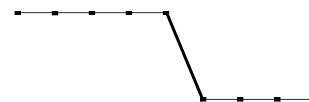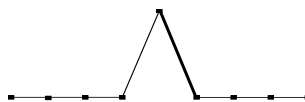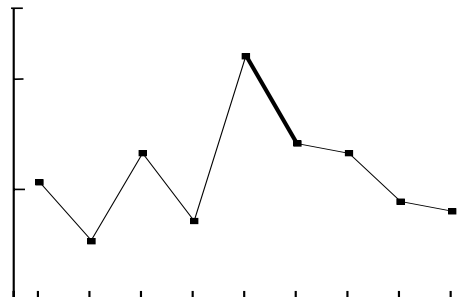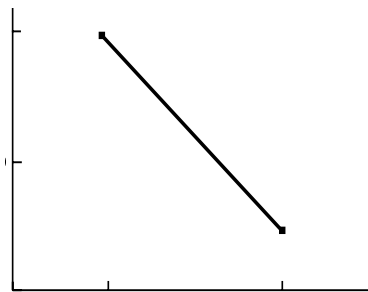
1   2   4
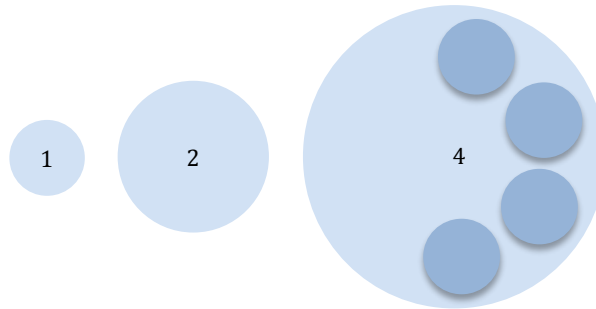
In this example the proportions are correct.
Area = πr^2

1   2   4

3.   Provide context

Lack of context can result in different interpretations of the data.

(Source: diagrams edited from Tufte, 2001)

More tips for representing the data accurately

➢ Use clear, detailed and thorough labelling to defeat any ambiguities.

➢ Write out explanations of the data on the graphic and label important events.

➢ For groups of graphs, use consistent scales, colours and order of categories.

➢ Use deflated and standardised units of money rather than nominal units.

3.2 Style the chart simply

The most important point to consider when styling the chart is that it is clear so that the viewer can see trends and patterns easily. This usually is achieved through simplicity in design and freedom from distractions.

Tips on styling a chart:

1. **Use a limited colour palette**

   The human eye can only distinguish between a few different colour shades on a single visualisation.
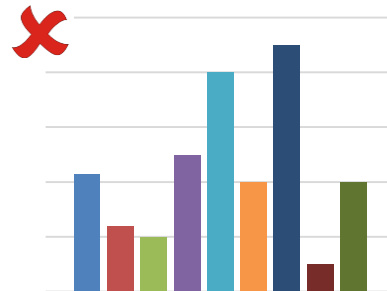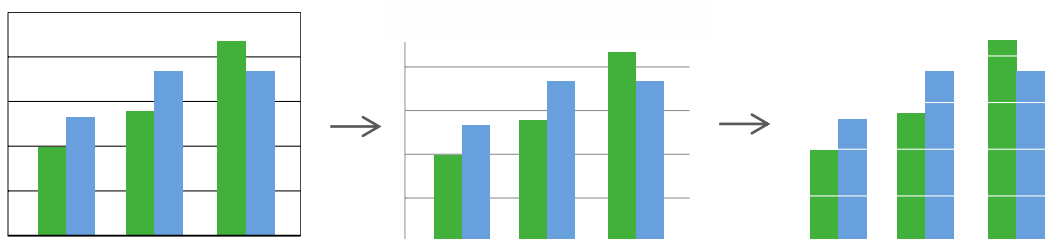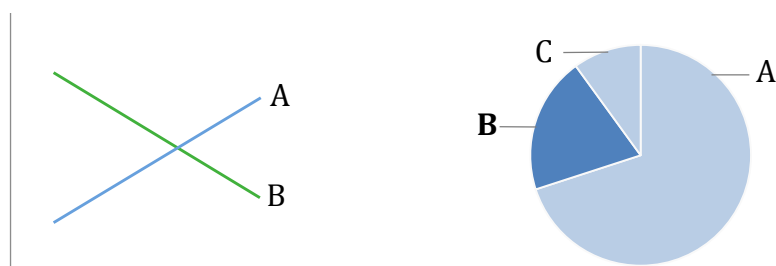
2. **Don't fill with patterns.**

   These can create disorientating visual effects and have no relevance. Instead, use shades of colour.

2. **Mute gridlines,** or remove gridlines completely if they are not necessary.

3. **Label as directly as possible** as this is helpful for all readers

3.3 Use colour with caution

The role of colour in a visualisation should be to compare and contrast one element from another, and to highlight the important points of the graphic. Used well, colour can enhance and clarify visualisations.
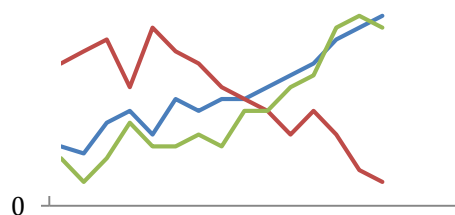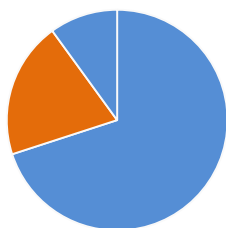
The colour wheel
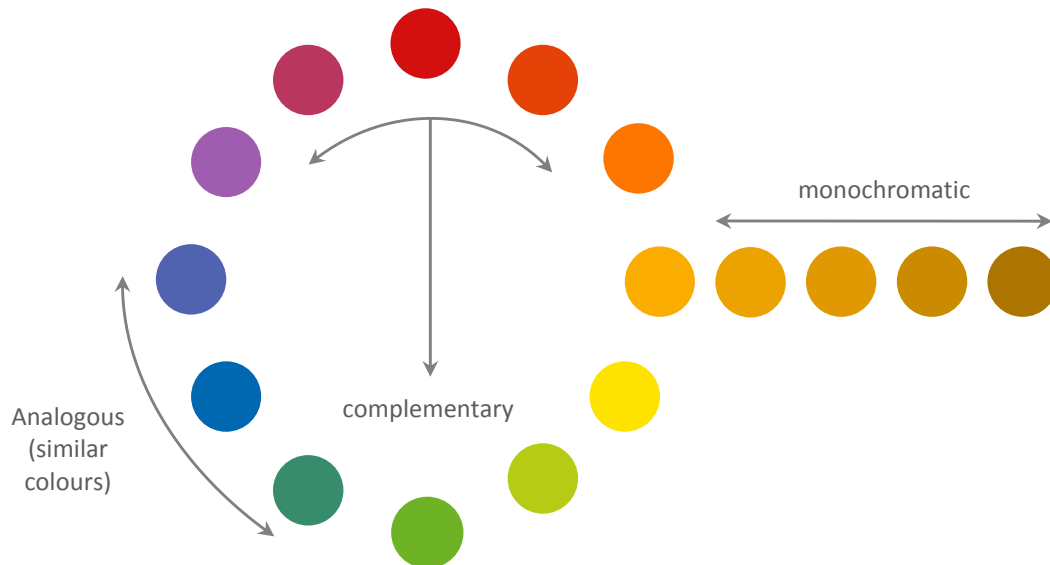
- Monochromatic series of colours can be used to make a heat map.

- Analogous colours can distinguish multiple data series without clashing.
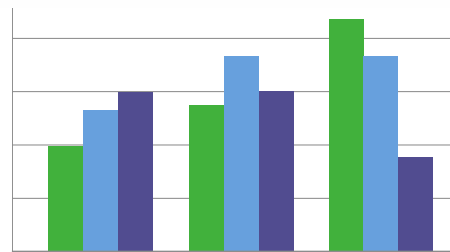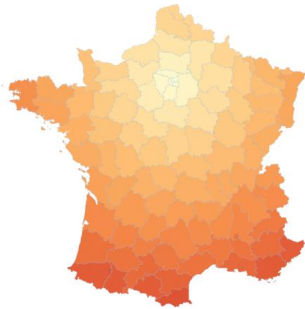
- Complementary colours can be used to show highlights and detail.

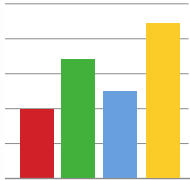- Bright (red) or dark (black) colours can emphasise the important line.

Dos and don'ts of using colour in charts

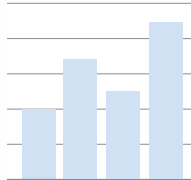(Adapted from *The Wall Street Journal guide to information graphics* by Dona M Wong, 2010)

Don't
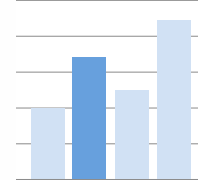Use multiple colours to represent the same kind of data.

Do
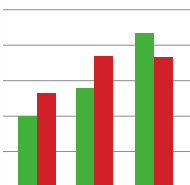Use the same colour to represent the same variable.

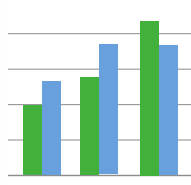Use darker shades or a different colour to highlight the focal point.

Don't
Use complementary colours in a multiple bar chart as the colour contrast is distracting

Do
Use graduating shades of one colour or analogous colours

Things to consider when choosing colours

1. Colours have meaning (and these meanings vary across cultures). For instance, red is associated with loss in business so shouldn't be used for positive numbers in a bar chart.
2. Many people are colour blind, in particular they can have problems distinguishing between red and green. Common alternatives are red-blue or purple-green.
3. Different colours (hues) that are the same lightness (gradient) can also be indecipherable for colour-blind readers and can appear the same if they are printed in grey scale.
4. Warm colours (red, orange, yellow) appear larger than cool colours (blue, green), so red can visually overpower blue even if used in equal amounts.

Web tools for choosing colours

ColorBrewer	http://colorbrewer2.org/	Provides colour-blind safe, print friendly and photocopy friendly colour schemes.
Adobe Kuler	https://kuler.adobe.com/	A searchable database and a tool to adjust the colour palette you are interested in.
COLOURlovers	www.colourlovers.com	A database of palettes that are sorted by themes.

3.4 Choose a clear font

Choosing the right font is important because it determines the tone of the graphic. For charts, the role of the font is to describe the information clearly so it must be legible. Poor font choice can draw attention away from the data and the message.

Which font to use?

- Choose a font that is easy to read (especially on a computer), and with a neutral tone.
- For digital or interactive visualisations sans serif fonts tend to be easier to read on screens than serif fonts. Examples include: Calibri, Myriad Pro, Helvetica, or franklin gothic
- For printed visualisations serif is easier. E.g. Georgia, Minion, Garamond, Baskerville
- Research has found that the use of Baskerville increases the likelihood of a reader agreeing with a statement by 1.5% compared to other fonts.
- Comic Sans MS is more informal and not taken seriously.
- Don't used highly stylized fonts such as Brush Script

Serif	Sans-serif
d	d
Serif fonts have a small line attached to the end of the strokes in a letter.	Sans-serif fonts (meaning without serif) don't have the line.
Serif fonts are easier to read in printed documents.	Small on-screen text is easier to read in a sans-serif font.

Basic rules of type legibility in charts

(Source: Wall Street Journal guide to information graphics)

- Don't set the type too small or too condensed.
- Use **bold** or *italic* only to emphasize a point. Don't use bold and italic at the same time.
- Don't use ALL CAPS as it is hard to read, upper and lower case lettering is easier to read.
- Avoid knocking white type out of black or colour.
- Avoid hyphenation
- Don't set type at an angle

Decreasing legibility

Is this easy to read?

Is this easy to read?

Is this easy to read?

Is this easy to read?

IS THIS EASY TO READ?

Is this easy to read?

3.5 Structure the story

Structure relates to how the data is organised on the page. It guides the narrative of the piece, and is key for understanding the visualisation/s and telling the story.
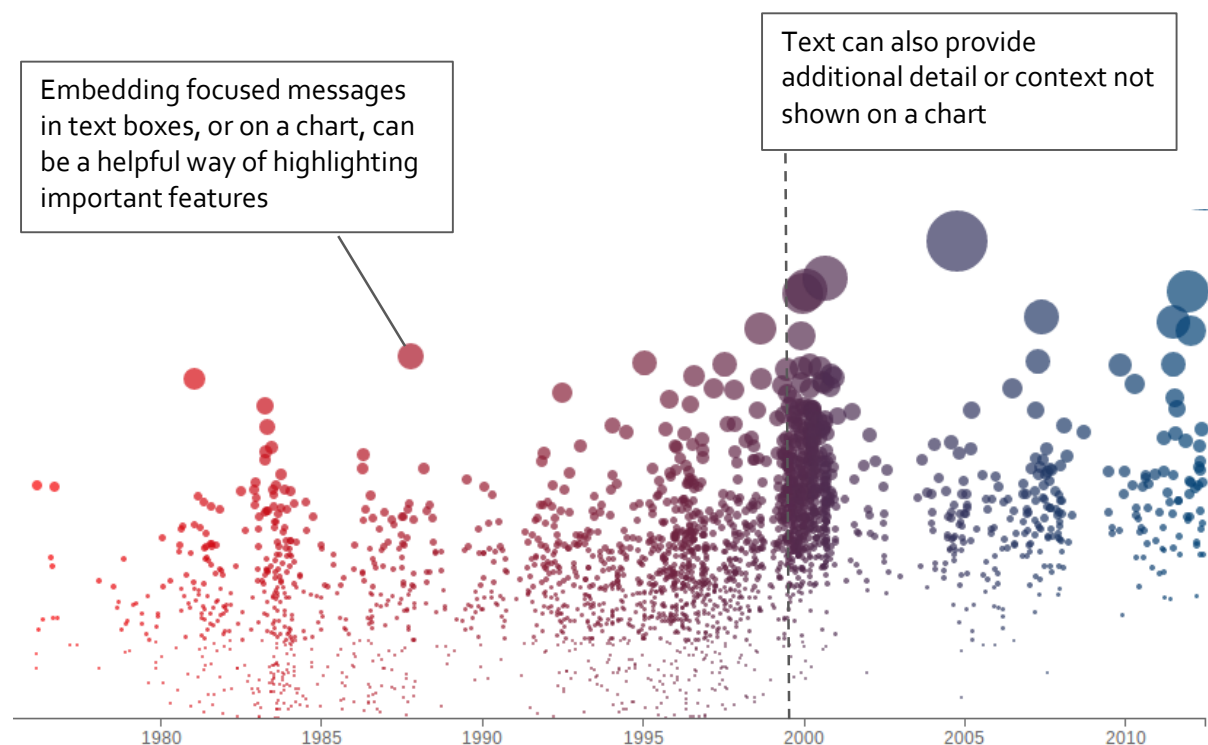
Title
As with storytelling, the beginning (title) is important. The title should make it clear to the viewer what it is they are looking at. The sub-heading could be a key finding.

Layout and ordering

➢ Establish a structure which sets out the messages with a logical flow (top to bottom – left to right).
➢ Organise the information to emphasise what you are trying to say to the audience: emphasise the most important visualisation by making it the biggest object on the page, or by placing it at the centre/top of the page.
➢ If the story has many sections – break it into manageable compartments with sub-headings to guide the user through it.
➢ Make sure that the layout, colours and keys are consistent across charts.

Narrative commentary

Data visualisations can often be enhanced by narrative commentary. Adding text to a visualisation can help to frame the data that is being presented, and also to guide the user through the story.


Embedding focused messages in text boxes, or on a chart, can be a helpful way of highlighting important features

Text can also provide additional detail or context not shown on a chart

4. Tools to create visualisations

- Datawrapper   http://datawrapper.de/
- Infogr.am   http://infogr.am/
- IBM Many Eyes   http://www-958.ibm.com/software/data/cognos/manyeyes/
- Visual.ly   http://visual.ly/
- Google Fusion Tables   https://support.google.com/fusiontables/answer/2571232
  (for maps)
- Tableau Public   http://www.tableausoftware.com/

Blogs and resources on data visualisation

- Gapminder   www.gapminder.org
- Guardian Datablog   http://www.theguardian.com/news/datablog
- Information is Beautiful   www.informationisbeautiful.net
- On Think Tanks   http://ttdatavis.onthinktanks.org
  – On Data Visualisation
- Visualising Data   http://visualisingdata.com
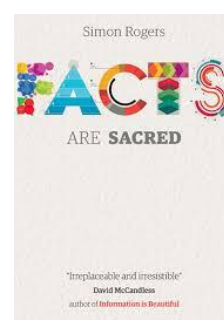- Worldmapper   www.worldmapper.org
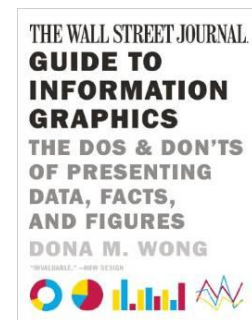

Books, articles and references
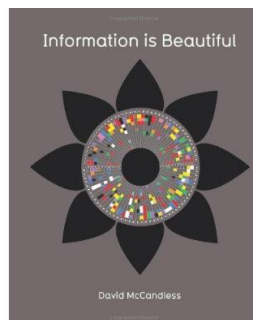
Information is Beautiful by David McCandless. Second Edition, 2012.
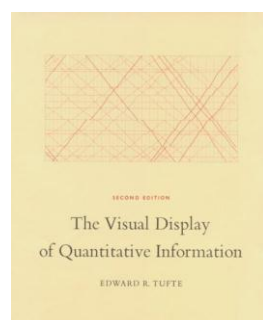
The Visual Display of Quantitative Information by Edward Tufte, Second Edition, 2001.

Facts are Sacred by Simon Rogers, 2013.

The Wall Street Journal Guide to Information Graphics by Dona M Wong, 2010

- **Making visualisations: a survival guide** by Gregor Aisch.
  http://www.slideshare.net/vis4/making-data-visualizations-a-survival-guide
- **Data Visualisation for human perception** by Stephen Few, 2013.
  http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html
- **Visualizing Information for Advocacy – An Introduction to Information Design**
  http://www.opensocietyfoundations.org/reports/visualizing-information-advocacy-introduction-information-design