4imprint

Blue Papers

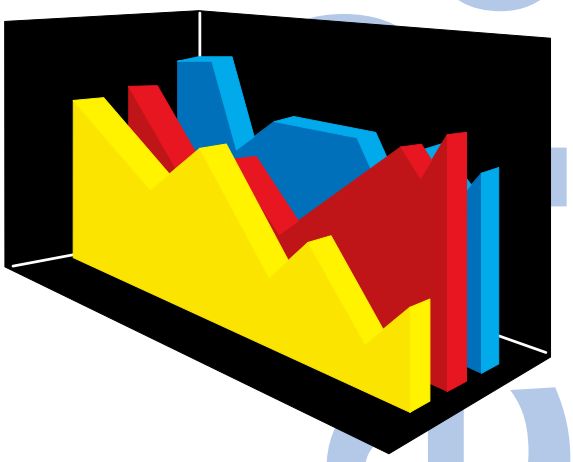
Communicating information by graphics - Data Visualisation

Data Visualisation: The science - and art - of discovering patterns and trends in data

Back in January 1997, The New York Times published a two-part article by a Pulitzer Prize-winning journalist on diseases in the developing world. This is said to have been the inspiration for Bill and Melinda Gates to dedicate a significant proportion of their wealth to the fight against global health challenges.

But according to its author, Nicholas Kristof, it wasn't actually the article itself that ignited the Gates' passion. Instead, it was one of the graphics that accompanied the article that really caught the Gates' attention: a simple two-column list of health problems in the developing world and the number of people those diseases kill. In a note to colleagues on the power of art, Kristof related this story, writing, "No graphic in human history has saved so many lives in Africa and Asia."[1] The graphic accomplished this by showing Times readers, including the Microsoft founder and his wife, the problem of diseases in developing countries in a quick, easy-to-comprehend format.

Generations of young English students have had the 'show, don't tell' principle of composition drilled into their psyche whilst generations of young maths students have plotted graph points, created Venn diagrams and turned data into pie charts in order to illustrate information. Today, these two disciplines are being combined to make connections and 'show' stories in ways that raw data, words and traditional graphic forms could never communicate. Data visualisation - all those graphs, charts and maps - is the new way to harness massive amounts of data, reveal patterns, discover trends and inspire action.



Simply put, data visualisation is the practice of communicating information through graphical means. Combined with emerging software capabilities and design principles, data visualisation is capable of presenting large amounts of information in ways that reveal previously unknown insights and invite more in-depth analysis than was possible with traditional spreadsheets, bar graphs and pie charts. (In fact, many data visualisation experts loathe the pie chart - but we'll get to that later.)

Data visualisation has the potential to be a powerful storyteller. Consider this - for the human brain, visual perception is fast and efficient. Cognitive function

1 *"Talk to the Newsroom: Graphics Director Steve Duenes." The New York Times. 25 Feb. 2008. Web. 5 May 2014.* http://www.nytimes.com/2008/02/25/business/media/25asktheeditors.html?pagewanted=1&_r=2

is slower and less efficient.[2] As data visualisation expert Stephen Few explains, "Traditional data sense-making and presentation methods require conscious thinking for almost all of the work. Data visualisation shifts the balance toward greater use of visual perception, taking advantage of our powerful eyes whenever possible."[3]

This Blue Paper will explore some of the media buzzwords related to data visualisation, delve into a variety of forms of data visualisation, examine best practices and flag up a couple of warnings. If one of the purposes of data visualisation is to reveal previously unknown information, then let's start peeling back the layers.

The buzz

One popular metaphor over the last few years has been that 'data' is the new 'crude oil.' There are massive amounts of it and its considerably valuable but, if left unrefined is virtually unusable. To create value, data must be broken down, analysed and presented in a format that is easily understood by those viewing and using it.

And one way to present a large amount of data is by using visualisation. But before we go any further let's first consider some related buzzwords: big data, data mining and infographics.

Big data: Information technology research company Gartner describes big data as 'high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimisation.'[4]

Where does this data come from? According to IBM 90% of the data in the world today has been created in the last two years, from various sources including social media sites, digital pictures and videos, purchase transaction records and mobile phone GPS signals.[5] And it's going to keep coming. Networking technology firm Cisco predicts that by 2018, there will be nearly five billion global mobile users in

2 *Stephen Few. "Data Visualization for Human Perception." The Encyclopedia of Human-Computer Interaction, 2nd Ed. Soegaard, Mads and Dam, Rikke Friis. The Interaction Design Foundation, 2013. Web. 5 May 2014. http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html*

3 *Ibid.*

4 *Michael Matzer. "What Exactly Is Big Data?" Business Innovation from SAP. 14 Dec. 2012. Web. 5 May 2014. http://blogs.sap.com/innovation/big-data/what-exactly-is-big-data-022976*

5 *"What Is Big Data?" IBM. Web. 5 May. 2014. http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html*

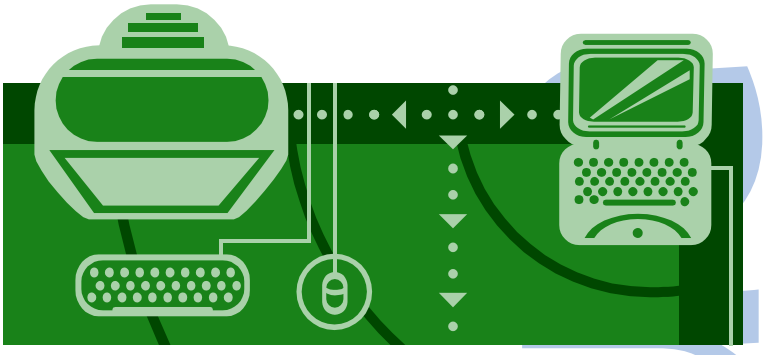
2018, up from more than four billion in 2013.[6] IT journalist Michael Matzer writes that 'while big data certainly presents a problem in terms of storage and analysis, the actual problem, according to Gartner, lies in spotting meaningful patterns within the data that can help companies make better decisions.'[7]

Data mining: One approach to harness big data is data mining, the process of searching through large amounts of data to uncover hidden patterns or trends. Software development firm Albion Research Ltd. explains that computer-assisted data mining can help companies find profitable relationships that were not even suspected to exist, or that would take too long to find by manual means.[8]

According to Albion, data mining has successfully contributed to:

- Identifying unexpected shopping patterns in supermarkets
- Optimising website profitability by making appropriate 'personalised' offers to each visitor
- Predicting customer response rates in marketing campaigns
- Defining new customer groups for marketing purposes
- Predicting customer defections
- Distinguishing between profitable and unprofitable customers
- Improving yields in complex production processes by finding unexpected relationships between process parameters and defect rates.
- Identifying suspicious behaviour as part of a fraud detection process[9]

Data mining and data visualisation are separate disciplines, but data visualisation is sometimes referred to as visual data mining.



Infographic: The terms data visualisation and infographic are often banded together, but there are distinctions between these two types of visual tools. Below is a quick snapshot example of each:

6 *"Visual Networking Index (VNI) – VNI Forecast. Cisco. Web. 5 May 2014. http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html*

7 *Michael Matzer. "What Exactly Is Big Data?" SAP AG. 14 Dec. 2012. Web. 5 May 2014. http://blogs.sap.com/innovation/big-data/what-exactly-is-big-data-022976*

8 *"Why Should I Be Considering Data Mining?" Albion Research Ltd. Web. 5 May 2014. http://www.albionresearch.com/data_mining/why.php*

9 *Ibid.*

Data Visualisation[10]

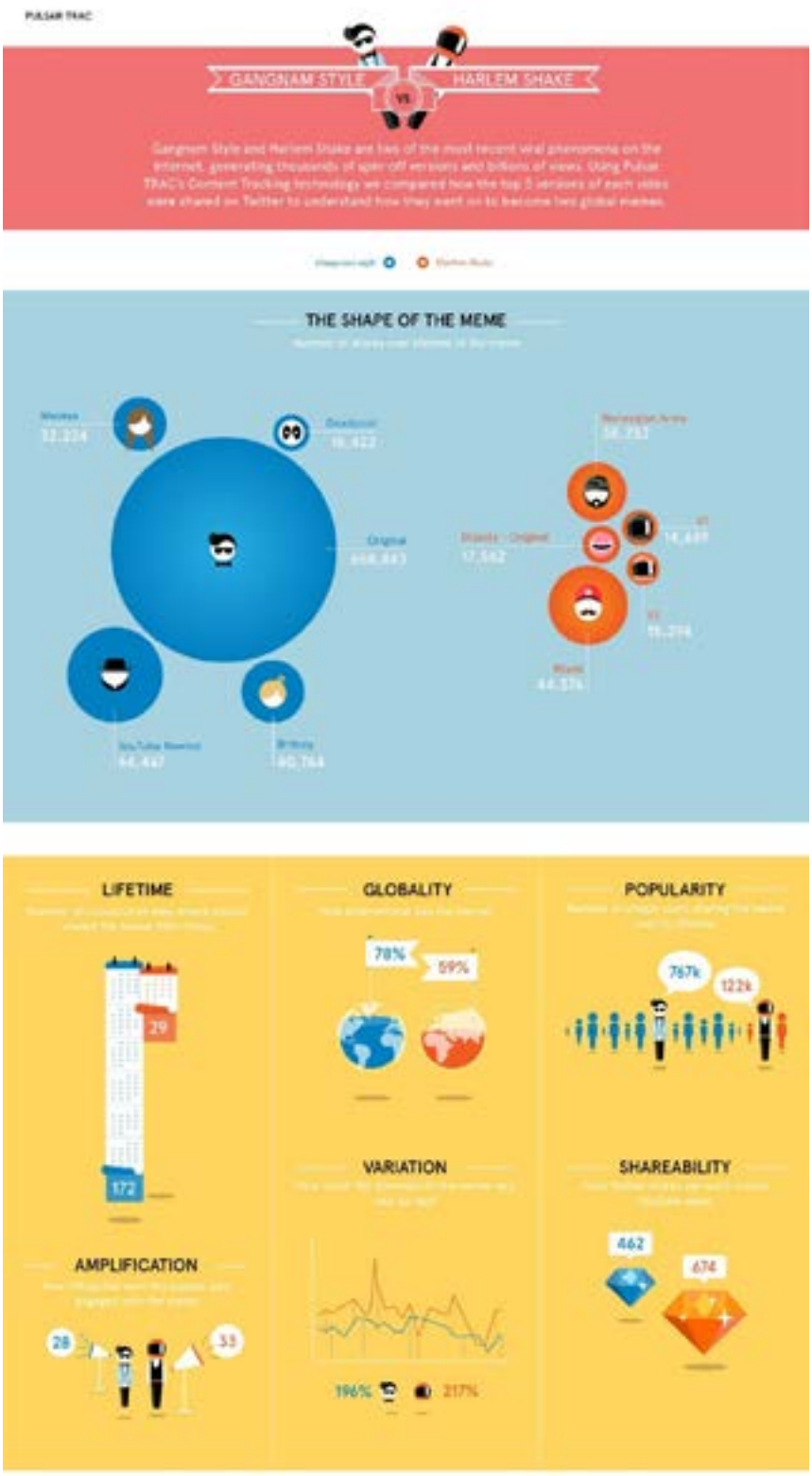
Infographic[11]

In general, an infographic illustrates a small amount of information or a single idea and the conclusions of an infographic are known to its creators before the infographic is designed. Basically, it is a graphic that communicates known information more quickly than words alone (as we all know, a picture is worth a thousand words.) In contrast, data visualisation usually encompasses a larger set of information and the patterns revealed within that visualisation aren't necessarily known to the creators before the graphic is designed. The visualisation itself reveals information that would have been extremely time-consuming or impossible to see by studying numbers on a spreadsheet.

10 *Audree Lapierre. 'Data Visualization for Marketing.' Audre Lapierres Blog. 30 Jan. 2011. Web. 5 May 2014. http://audreelapierre.com/blog*

11 *"The role of Promotional Products in the Marketing Mix." 4imprint Infographic. Web 5 May 2014. http://info.4imprint.co.uk/wp-content/uploads/MarketingMix.pdf*

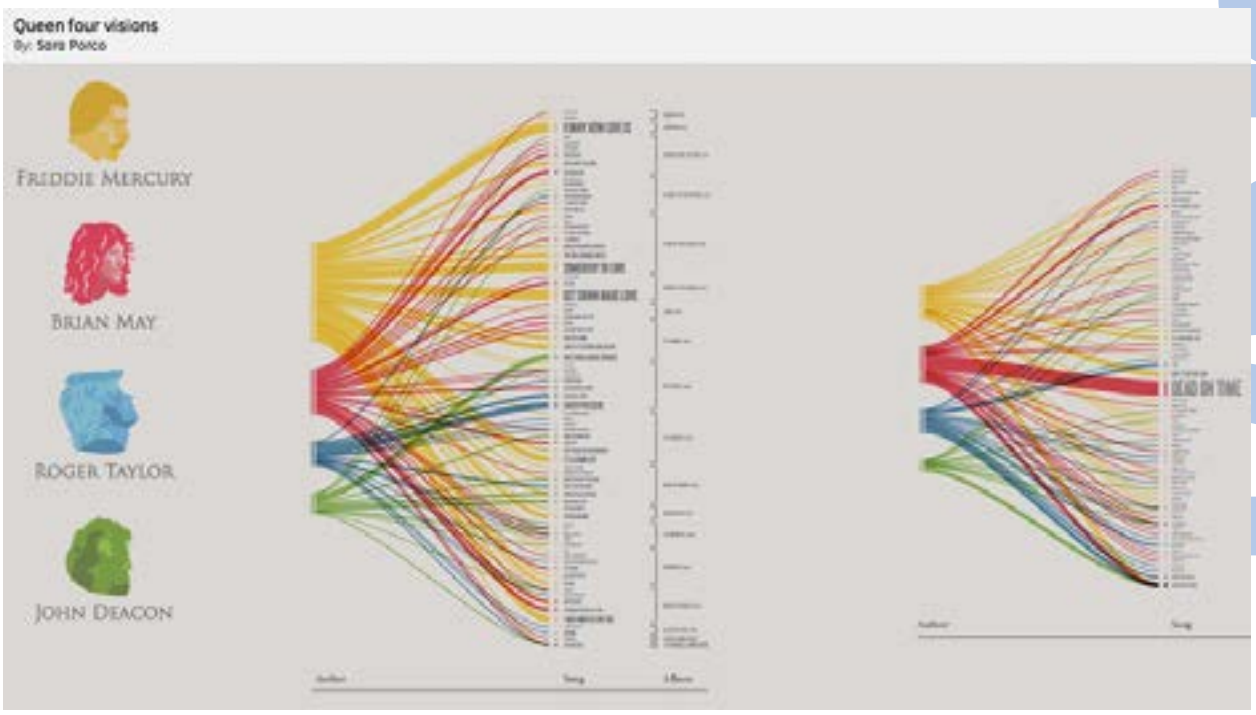
Let's take a look at two examples using music as our common theme. The first, is an infographic[12] comparing the most popular spin offs from the massively popular videos 'Gangnam Style' and 'Harlem Shake.'



12 *'Pulsar' Social Media Insight platform. Sennep. Web. 5 May 2014.* http://sennep.com/work/pulsar

The 'Gangnam Style vs. Harlem Shake' infographic[13] created by UK based design agency Sennep shows how many times the spin offs were shared on Twitter to understand how they went on to become global memes (a meme is an idea which spreads from one individual to another through imitation). Sennep had all the data to hand before creating the infographic to illustrate their findings.

In contrast, Visualizing.org did not know the outcome of their research before the visualisation began for their 'Four Visions' project[14] which analysed the importance of language in songs by Queen, the famous British rock-band. By analysing significant words repeated in the lyrics the visualisation showed both the key themes in Queen's music and the influence of each song's writer.

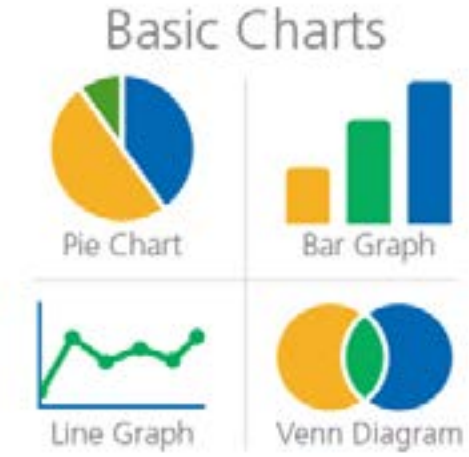


Forms of data visualisation

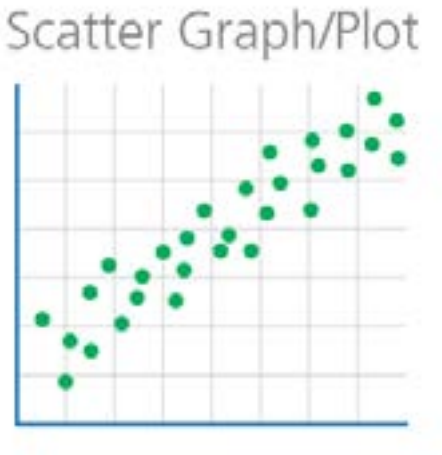
Whether you have a massive amount of data or a reasonable amount that you want to use to help reveal unknown patterns or trends, a key to good data visualisation is choosing the right form of chart, graph or map. The 'Four Visions' project example is one form of a data visualisation chart. Here are other examples:

13 *'Pulsar' Social Media Insight platform. Sennep. Web. 5 May 2014. http://sennep.com/work/pulsar*
14 *Mashable. Web 5 May 2014. http://mashable.com/2013/03/05/data-visualization-projects/ and http://www.visualizing.org/full-screen/46075*

Basic charts: The traditional bar graphs, pie charts, line graphs and Venn diagrams that we all learned in back at school.



Scatter graph or scatter plot: A diagram that plots values with two variables, with one variable represented on a horizontal axis and the other represented on a vertical axis.

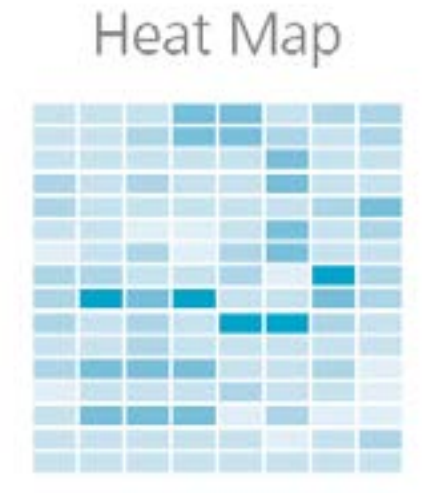


Bubble chart: A chart that applies numeric value to circles whose areas are proportional to the values. A common variant of this is the **bubble map**, which charts proportionally sized circles on a geographic map.

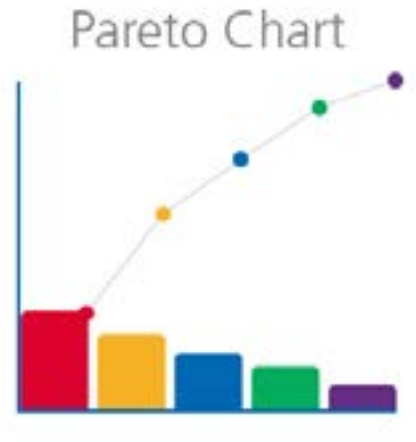
Sparkline chart: A small line chart that displays a general trend in measurement. Often lacking a defined x- and y-axis, these charts are commonly seen in stock market visualisations.

Tree map: A visualisation of hierarchical data.

Heat map: A graphical representation of data values in a colour-coded rectangular array.



Pareto chart: A chart that combines bars and a line graph.



These are only a few of the most common forms. Ralph Lengler and Martin J. Eppler of the e-learning course Visual-Literacy.org identified 100 visualisation methods, grouping them into six general categories:

1. **Data visualisation** (Examples include pie charts, bar charts and line charts.)
2. **Information visualisation** (Examples include timelines, flow charts and Venn diagrams.)
3. **Concept visualisation** (Examples include mind maps and synergy maps.)
4. **Strategy visualisation** (Examples include stakeholder maps, s-cycle diagrams and value chains.)
5. **Metaphor visualisation** (Examples include story templates, icebergs and funnels.)
6. **Compound visualisation** (Examples include cartoons and knowledge maps.)[15]

In addition to the charts themselves, electronic data visualisation often includes interactive elements that add more information and depth to the visualisation if the viewer needs or desires it. For example, Lengler and Eppler illustrated their 100 identified methods as a 'periodic table' of visualisation methods[16] and in the electronic version of this table, an example of each of the 100 methods pops up as the viewer scrolls over the table.

Another great example of interaction is from concerthotels.com with a blog entitled '100 Years of Rock in Under a Minute'[17] which at first glance looks a mind-map or infographic - but by clicking on a genre the viewer becomes the listener as a sample of music is played, along with a credit for the group and song title.

---

15 Ralph Lengler & Martin J. Eppler. "A Periodic Table of Visualization Methods." Visual-Literacy.org. Web. 5 May 2014. http://www.visual-literacy.org/periodic_table/periodic_table.html

16 ibid.

17 "100 Years of Rock." concert.hotels.com. Web. 5 May 2014. http://www.concerthotels.com/100-years-of-rock

And if you think electoral politics can be a little hard going, take a look at the fantastic example of data visualisation produced by the BBC politics team which highlights the key marginal seats for the 2015 general election.[18] Sounds a bit dull? Try it here and discover the different channels and routes that viewers can take which deliver a totally personalised user-experience.

Interaction in visualization enables the fast exploration and discovery of data patterns that the user may not even have expected, and what's more it is possible to reduce the amount of data shown at the same time, providing clearer visualisations whilst still giving the user the option to get further information at any time.[19]

Tooltips are graphical interface elements that appear when a curser hovers over a particular spot on a visualisation. Delegating (or transferring) detailed information into Tooltips helps provide additional information while keeping a chart clean. The experts at FusionCharts provide this example of hotel analysing data related to reservation trends: The hotel experienced a dip in reservations in September because a new hotel opened nearby and then reservations picked up again in October when the hotel offered a discount. Putting all of this information on a line graph of reservations by month would make it very cluttered. Putting the detailed information in Tooltips is more helpful for users, in this case the management team. When a user is interested in getting more information about a data set, they can use their cursor to hover over it to access additional information. The chart is clean, yet still provides all of the necessary information.[20]

Best practices

Data visualisation may be an emerging discipline, but solid visualisations rely on well-established design principles. The Gestalt principles (theories of visual perception developed by German psychologists in the 1920s) help to explain how and why humans see patterns and trends in visualised data. These principles include similarity (when objects look similar, humans perceive them to be a group or pattern), continuation (the human eye tends to follow a line or curve from one object to another) and proximity (when objects are placed close together, humans perceive them as a group).

18 *'The seats which could decide the next election.' BBC Politics. Web. 5 May 2014. http://www.bbc.co.uk/news/uk-politics-25726270#interactive*

19 *Robert Kosara."Data Visualization for Human Perception." The Encyclopedia of Human-Computer Interaction, 2nd Ed. Web. 5 May 2014. http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html*

20 *"6 Tips to Increase the Usability of Your Charts." FusionCharts. Web. 5 May 2014. http://www.fusioncharts.com/resources/charting-best-practices/6-tips-to-increase-the-usability-of-your-charts*

Data visualisation consultant Jorge Camoes argues that good data visualisation functions with an understanding of how human perception works. He writes that:

- Humans are constantly grouping objects based on colour, shape, direction, proximity and closure/enclosure
- Humans like simple, close, smooth, symmetrical, easy-to-process shapes
- If we are aware of these laws, we can take advantage of them to design better charts or dashboards
- We must also be aware of their negative effect. We shouldn't force the reader to see groups that aren't really there; a simple shape is not necessarily the right shape[21]

Good data visualisation gets both the science and the art right. The following steps are adapted from an approach to data visualisation by user interface developer Ryan Bell of EffectiveUI:[22]

Step 1: Understand the problem: Ask questions to understand what specific problem needs to be explored.

Step 2: Get sound data: This includes understanding the meaning of the data, ensuring that it is as accurate as possible and having a sense of how it was gathered and what errors or inadequacies may exist.

Step 3: Show the data and show comparisons: Choose the visualisation method that will best show relationships within the data.

Step 4: Incorporate visual design principles: Sound design elements (line, form, shape, value, colour) and principles will make trends easy to see.

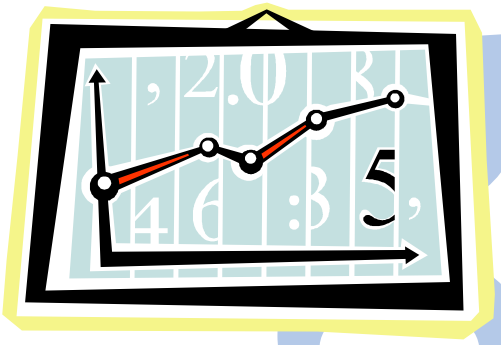
Step 5: Bring in more dimensions with one or more of the following techniques: Add small multiples: Thumbnail charts of related data help users make quick side-by-side comparisons.

- Add layers: Adding extra levels of information, while preserving the high-level summary data, can make a graphic more flexible and useful

21 *Jorge Camoes. "Perception: Gestalt Laws." Web. 5 May 2014. http://www.excelcharts.com/blog/data-visualization-excel-users/gestalt-laws*

22 *Ryan Bell. "Eight Principles of Data Visualization." Information Management. SourceMedia, 17 Aug. 2012. Web. 5 May 2014. http://www.information-management.com/news/Eight-Principles-of-Data-Visualization-10023032-1.html?zkPrintable=1*

- Add axes or coding patterns: Adding an additional variable or coding of information can help reveal new information. However, be careful not to add too much clutter
- Combine metaphors: Add other forms of data visualisation to show additional connections



Finally, all good data visualisation is accompanied by clear text that helps viewers to understand quickly what they are looking at. Use plain and simple language to summarise the data, include units of representation and the time period and avoid using articles ('a', 'an' and 'the') and adjectives.[23]

Statistics and bad data visualisation

There's nothing new about number manipulation. It was Mark Twain, among others, who warned of 'lies, damned lies and statistics'.[24] Whether intentionally misleading or just poorly executed, data visualisation has as much potential for misunderstanding and misuse as earlier iterations of data analysis. In the hands of a statistician with no understanding of how design principles affect the way humans perceive graphic renderings, data visualisation may be accurate but fail to communicate. In the hands of a designer who wants to show off skills but doesn't understand the information, data visualisation may look eye-catching but mislead or fail in its goal to reveal connections and trends.

Vitaly Friedman, editor of Smashing Magazine, an online magazine for designers and developers, writes about the primary goal of data visualisation: [25]

> "[To] communicate information clearly and effectively through graphical means. It doesn't mean that data visualisation needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key aspects in a more intuitive way. Yet designers often fail to achieve a balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose - to communicate information."

23 *"5 Tips for Writing Great Chart Captions." FusionCharts. Web. 5 May 2014. http://www.fusioncharts.com/resources/charting-best-practices/5-tips-for-writing-great-chart-captions*

24 *"Lies, damned lies, and statistics." Wikipedia. Web 5 May 2014. http://en.wikipedia.org/wiki/Lies,_damned_lies,_and_statistics*

25 *Vitaly Friedman. "Data Visualization and Infographics." Smashing Media, 14 Jan. 2008. Web. 5 May 2014. http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics*

Researcher and data educator Dana Griffin offers the following data visualisation 'Don'ts'[26]:

1. Don't assume viewers will automatically 'get' whatever graphic you create: Understanding doesn't happen in a vacuum - it is the researcher's responsibility to distill data in a way that makes clean, intuitive sense to viewers and to provide visual tools that guide the audience to insights or conclusions.
2. Don't overstate or oversimplify the data: The conclusions that viewers reach are highly dependent on how specific the data set is.
3. Don't make viewers look too long and/or think too hard in order to get the point: Time and attention are limited. Poorly designed visuals risk confusing, irritating or alienating viewers.

And while we are on the topic of bad data visualisation, let's return to that ubiquitous pie chart. In an Oracle blog exploring why designers would choose a pie chart as a visualisation metaphor, Tony Wolfram explains that pie charts don't lend themselves well to human perception capabilities. He writes[27]:

> "When comparing data, which is what a pie chart is for, people have a hard time judging the angles and areas of the multiple pie slices in order to calculate how much bigger one slice is than the others."
>
> "Controlled studies show that people will overestimate the percentage that a pie slice area represents. This is because we have trouble calculating the area based on the space between the two angles that define the slice."

Wolfram's preference for data typically applied to a pie chart is a sorted bar chart.[28]

With that in mind, it's probably best you save your pies for lunch and choose a visualisation method that reveals relationships and communicates patterns and trends in a clear and accurate manner. Use basic design principles based on visual

---

26 *Dana Griffin. "Data Visualization Lesson 6: The Ultimate List of Dos and Don'ts." Research Access, 17 Oct. 2012. Web. 5 May 2014. http://researchaccess.com/2012/10/data-visualization-lesson-6-the-ultimate-list-of-dos-and-donts*

27 *Tony Wolfram. "Pie Charts Just Don't Work When Comparing Data - Number 10 of Top 10 Reasons to Never Ever Use a Pie Chart." The Designer Experience. Oracle. 10 May 2010. Web. 5 May 2014. https://blogs.oracle.com/experience/entry/pie_charts_just_dont_work_when_comparing_data_-_number_10_of_top_10_reasons_to_never_ever_use_a_pie*

28 *Ibid.*

perception theories and make it interactive to keep charts simple while providing the maximum amount of information. Whether you need to uncover sales patterns, reveal trends in purchase categories or customise offers to particular shoppers, good data visualisation will help to provide the insights that you need and your customers want. You never know what actions may be inspired!