LESSON HANDOUT

Visualisation

Visualisation is the representation of an object in a graph or other image. By turning numbers or raw data into shapes or patterns, it allows us to quickly make comparisons and comprehend trends. Visualisation is not just the act of making things prettier or smarter looking - visualisation fundamentally changes the way we absorb data.

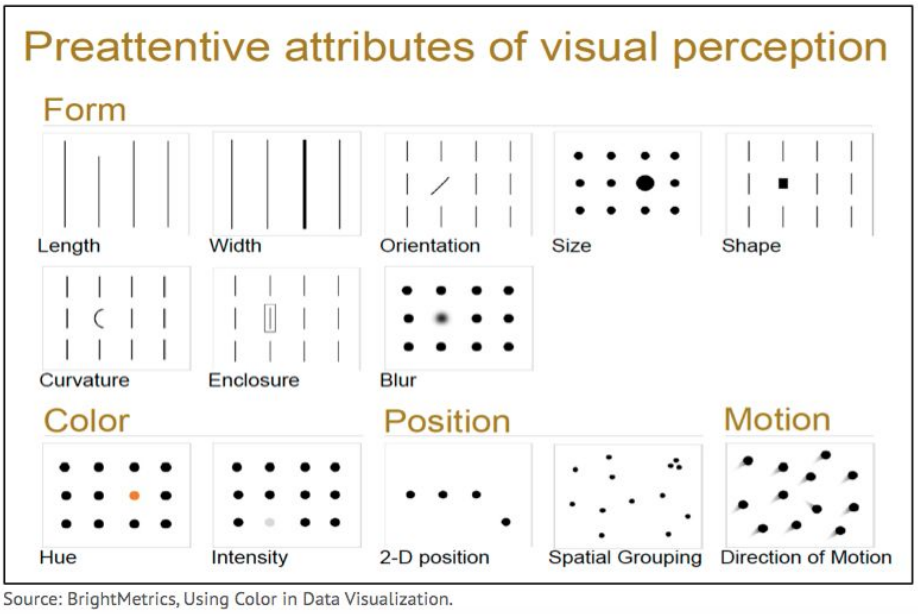
Preattentive processing

On a conscious level, visualization is essential for comparing the scale of data. But visualisation also taps into something very powerful, which is the human ability to subconsciously assimilate information from their environment. This is known as '*preattentive processing*'.

Preattentive features/ attributes

Preattentive features show us where to look. These are features which are processed in our spatial memory with no conscious action. In fact, using spatial memory, it takes an average of fewer than 500 milliseconds for the eye and the brain to make sense of preattentive features.

The diagram below shows different examples of preattentive features:



Source: BrightMetrics, Using Color in Data Visualization.

Figure 1 - Preattentive attributes of visual perception

Using preattentive features - avoid cognitive overload

We should use preattentive features **sparingly** and highlight the most important parts of our visualisations. Overuse of preattentive features can result in what is called *cognitive overload*.

Cognitive overload is the result of where we are presented with too much information. It's when the cognitive ability required to process the information outweighs the cost. The key when using these elements is that **less is more**.

Planning your data visualisations

In the book – storytelling with data, which I highly recommend – the author explains the three key factors to consider when building your visualisations.

1. **Who** – Who is your audience? What is their ability to process information and appetite for visualisations?
2. **What** – What do you need your audience to know. What do you want them to take from your visualisation?
3. **How** – What is the best way to get your message across. Will it be one more complex graph or a series of visualisations?

Keep in mind that you have a responsibility to convey not only your findings but enough **context** as to how your findings sit in comparison to the whole.

Colour

Why use colour?

Colour is a powerful preattentive feature and is one of the easiest ways to draw attention to an element. We also derive emotions from some colours (ie the colour **red** to danger), making their proper use very important.

How to use colour

Colour should be used to:

- Distinguish one variable from another
- Tie similar categories together
- Draw attention to a particular element

In general, you should use a low saturated colour palette. Bright or highly saturated colour should be used **sparingly** and to make an impact (ie show us where to look). High saturation colours such as bright red can spark certain physiological reactions associated with danger.

This LINK will take you to a great website to choose a colour palette. In addition, most programs come loaded with some decent options you can easily add - for example using different themes in Excel or Power BI.

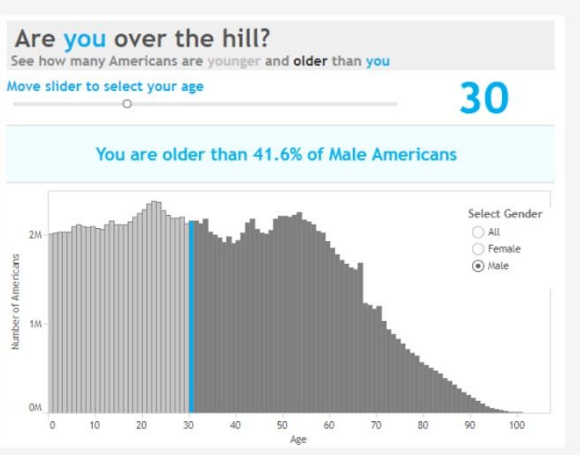


Figure 2 - A good example of the correct use of colour

Clutter

Clutter are elements in your visualisations which aren't worth the space and attention. Clutter more than completely unnecessary elements, it can also include elements that while they do inform, they take up too much space. The more clutter you have on your page, the more you run the risk of cognitive overload and detracting from your point.

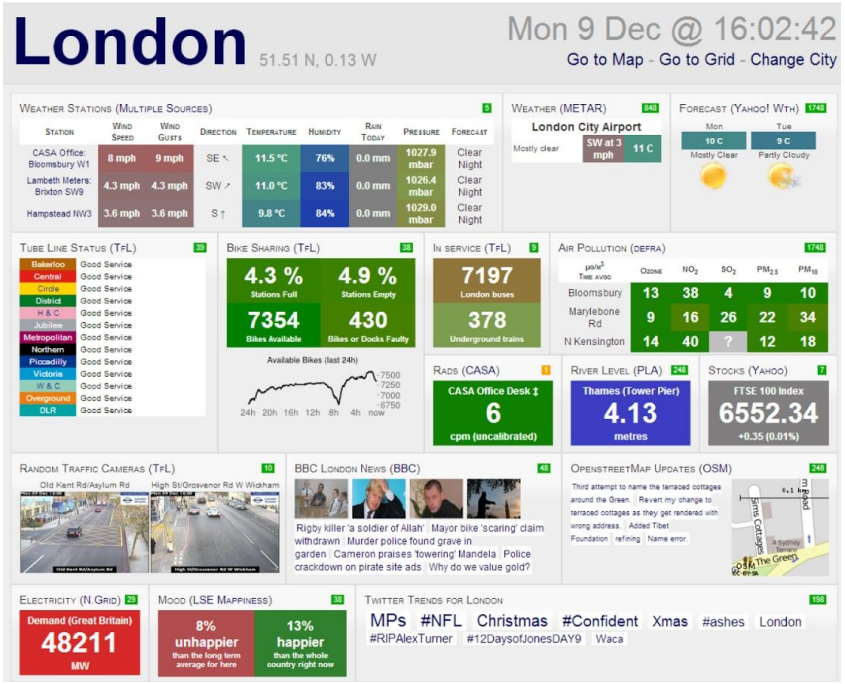


Figure 3 - City of London dashboard: an example of too much clutter

Components of a graph

The below graph will be used as an example for us to understand the different components.

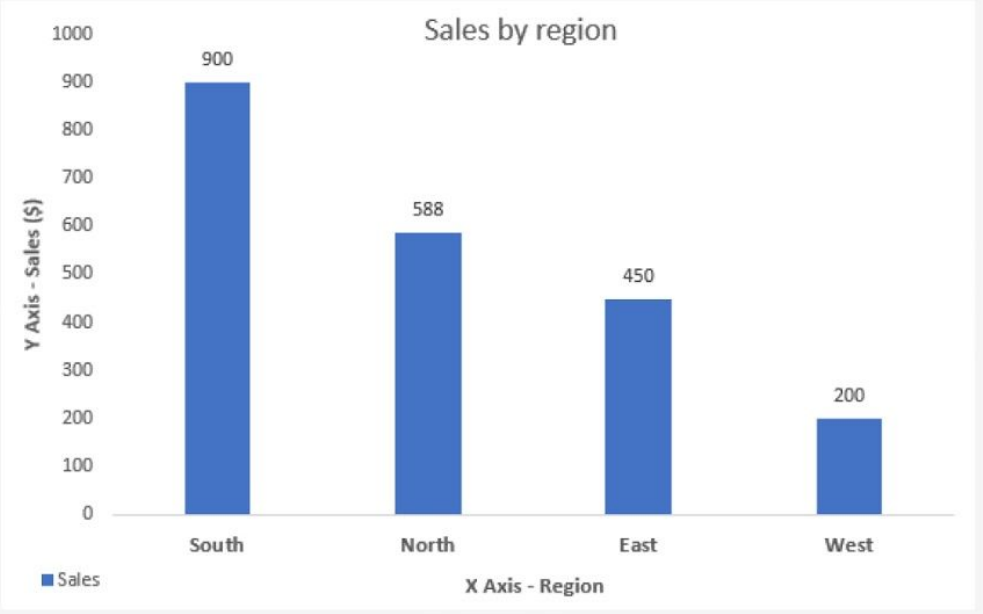


Figure 4 - Example Graph

The Axis

Graphs generally have two Axes, the X-axis or horizontal axis and the Y-axis or vertical axis. These are the scales at which your data is plotted. Generally, the X-axis is used to plot the independent variable and the Y-axis used to plot the dependent variable (in the example above, **sales** are dependent on **region**).

The Axis Titles

These are the titles which correspond with the relevant axis and can be used to specify the variables displayed on each axis. Note that they are not always necessary - it depends on the graph.

The Graph Title

The clear and unambiguous graph title is essential because it explains what is on display. A good convention for naming the chart title is the Y-axis filled by the X-axis - in this example: **Sales by Region**.

Data points

Also known as Data Labels, these are the individual values of data on display. In this example, the sale amount of 450 for the East region is a data point.

Data series

This is a collection of relevant data points. In this call all of the regional sale values comprise of the data series.

Visualization selection

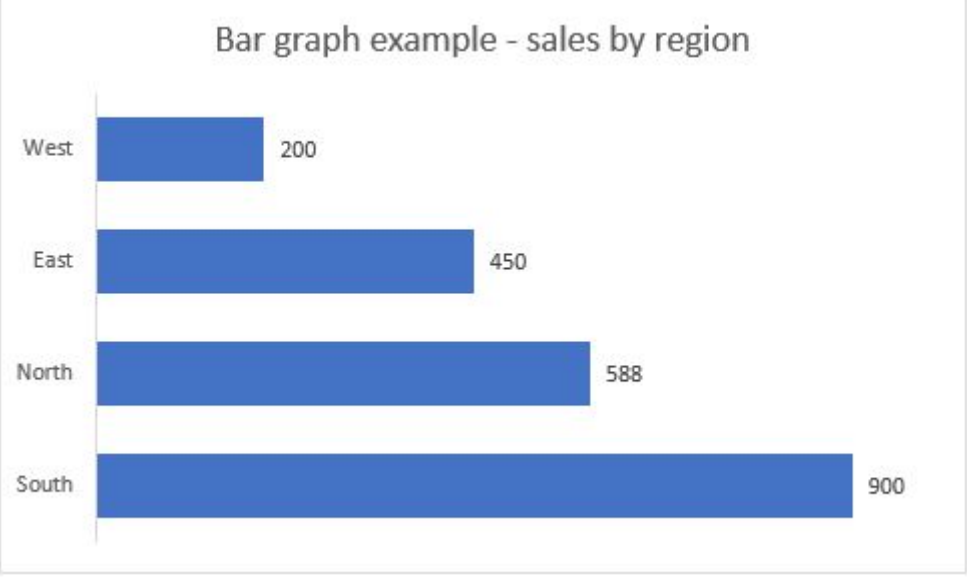
Different graphs are appropriate for different displays of information. This is an extensive and difficult choice.
Below are examples of some of the most common ways of visualising data, and where they should be used.

The line graph

- This graph is usually used to display how variables change over time.
- This is a great example of independent vs dependant variable, with the independent variable (time) always on the X-axis and the dependant variable (sales) on the Y-axis.

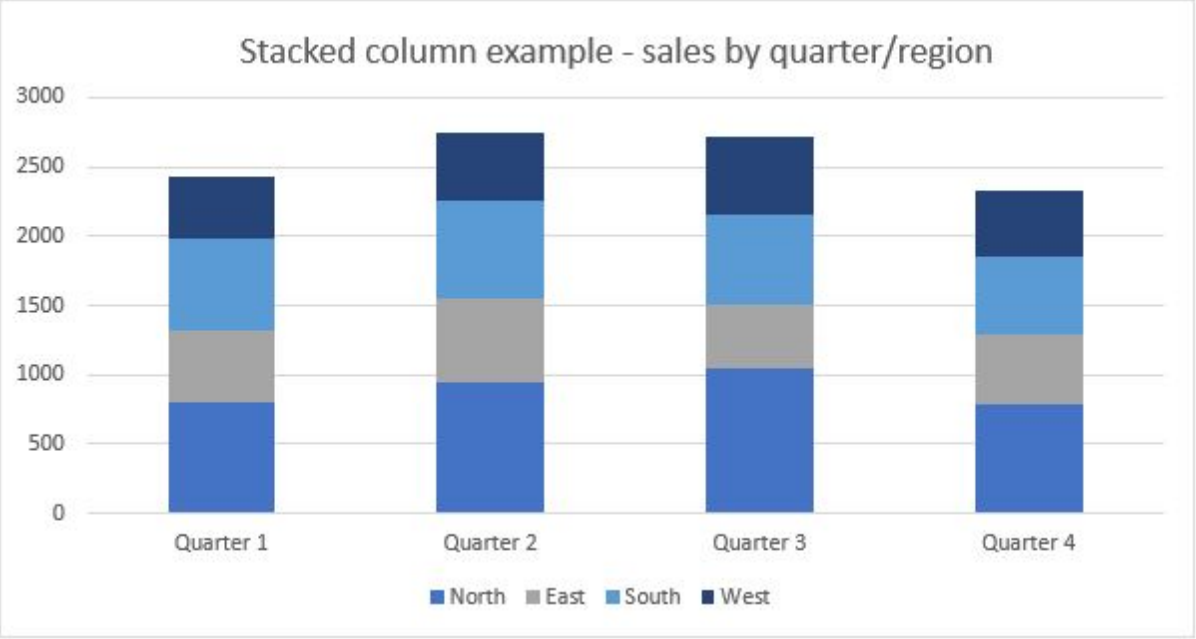
The bar chart or column chart

- A bar chart is the same display as a column chart but with the axis reversed.
- Both bar and column charts are great for comparing variables against each other.

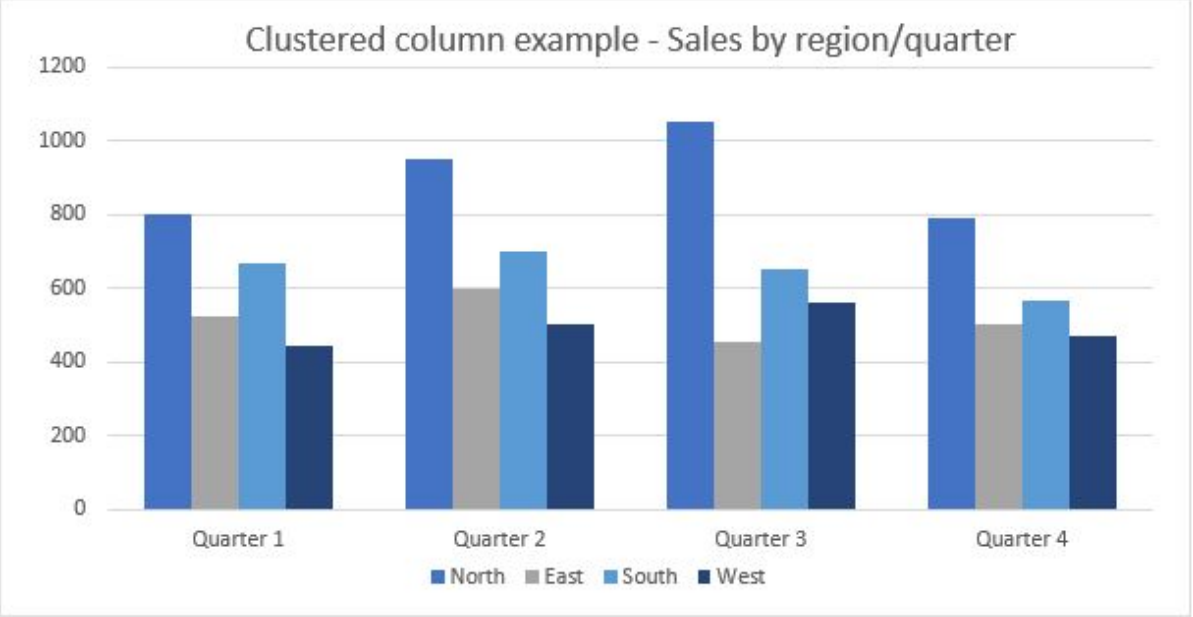


Stacked column and clustered column

- A stacked column is a way of showing the *proportion* of multiple dependant variables against one independent variable.
- In the example below, the Y-axis of Sales is further broken down by regions.

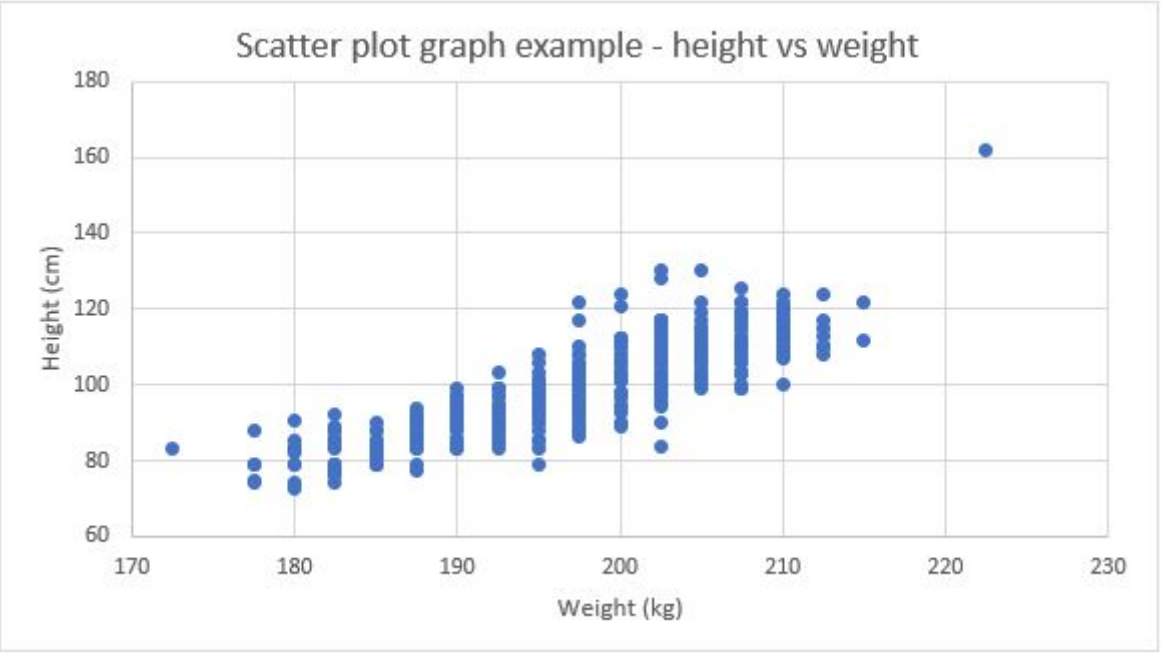


- The clustered column is another way to show multiple variables against a variable.



Scatter plot

- Shows the relationship between two variables



Charts to avoid

- **Pie charts.** Generally speaking, humans aren't very good at measuring slices of a pie, making pie charts difficult to understand. In addition, pie charts generally take up more space on a dashboard than other charts.
- **3D charts.** 3D charts can distort your image, which can make some data look incorrectly larger than others.