Week 1: Types of Charts

What chart to use to communicate?

- Is my data categorical or quantitative? Or multiple types?
- Do you want the audience to remember a single number, a comparison, a trend, a distribution or a breakdown?

> consider the type of data you are graphing and the message you are trying to communicate

'C.H.R.T.S'

Andy Kirk (2016) grouping:

- Categorical: Comparing **categories and distributions** of quantitative values
- Hierarchical: Charting **part-to-whole relationships** and hierarchies
- Relational: Graphing relationships to explore correlations and connections
- Temporal: Showing trends and activities over **time**
- Spatial: Mapping spatial patterns through **overlays and distortions**

Other ways of grouping charts

Knaflic (2016, p37)

: https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600216

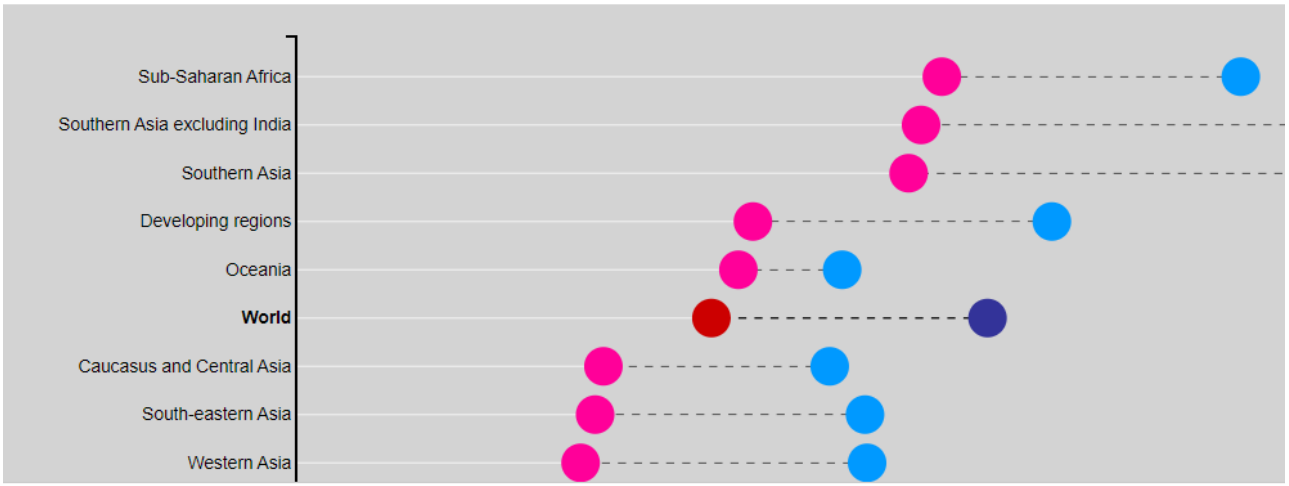
FIGURE 2.1 The visuals I use most

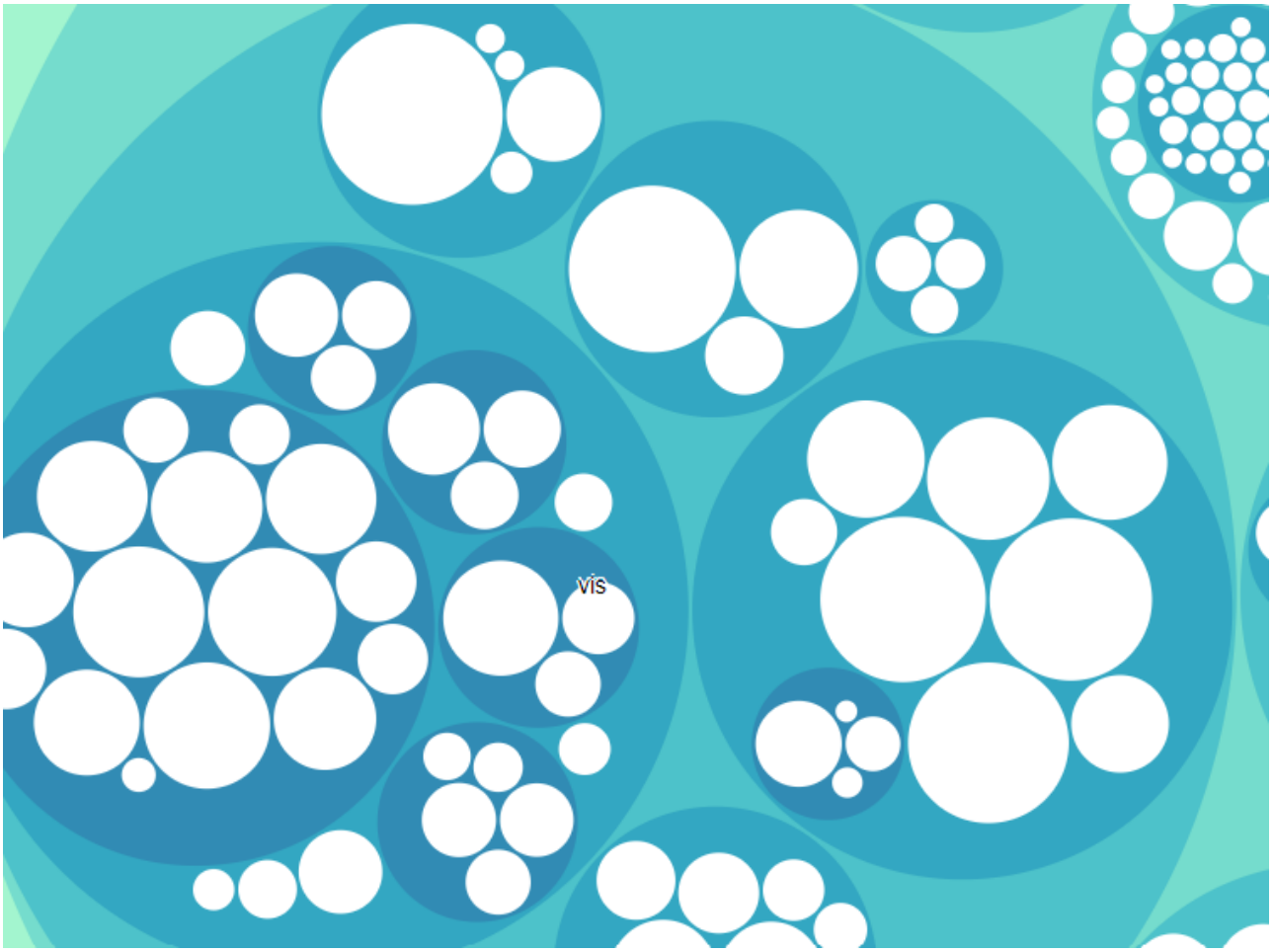
Categorical examples

Comparing categories and distributions of quantitative values.

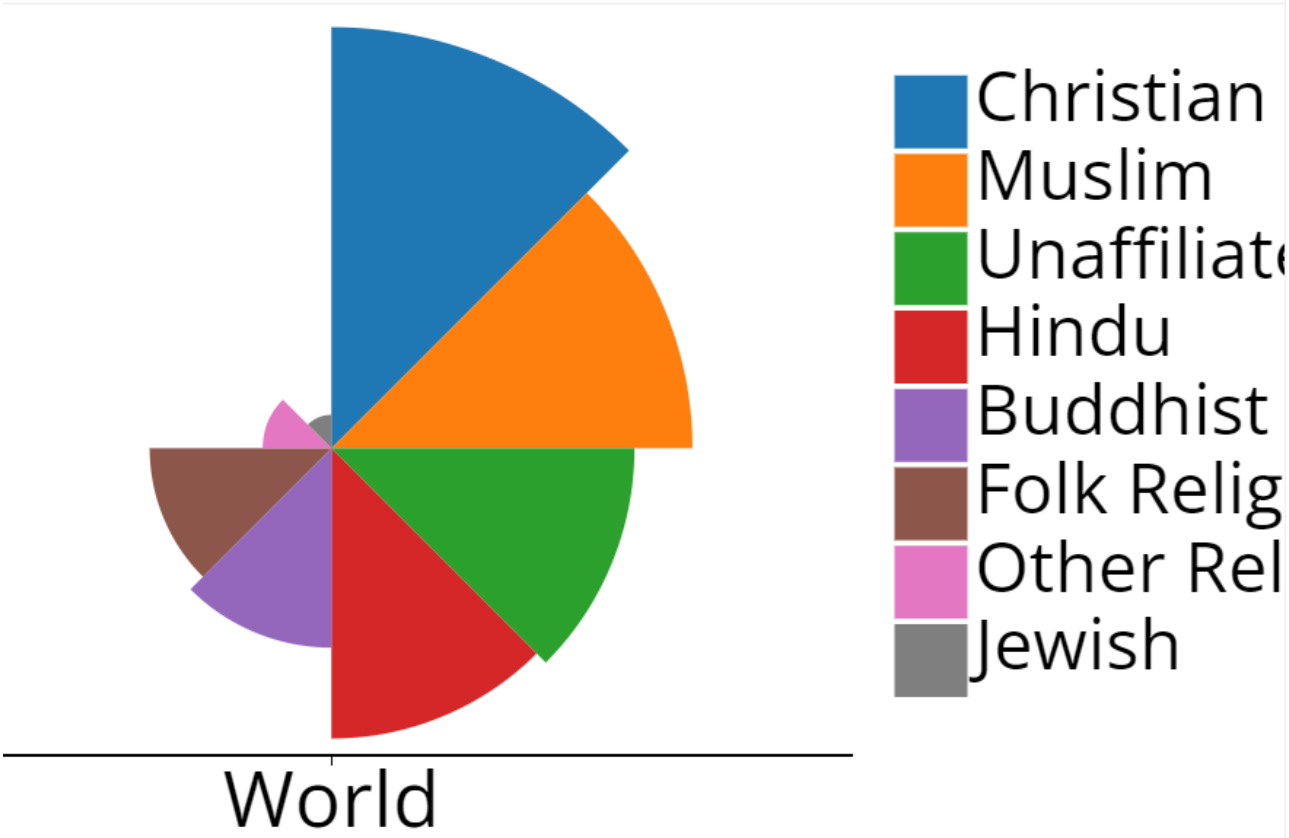
- **Bar chart** (NOT histogram)

Other examples:

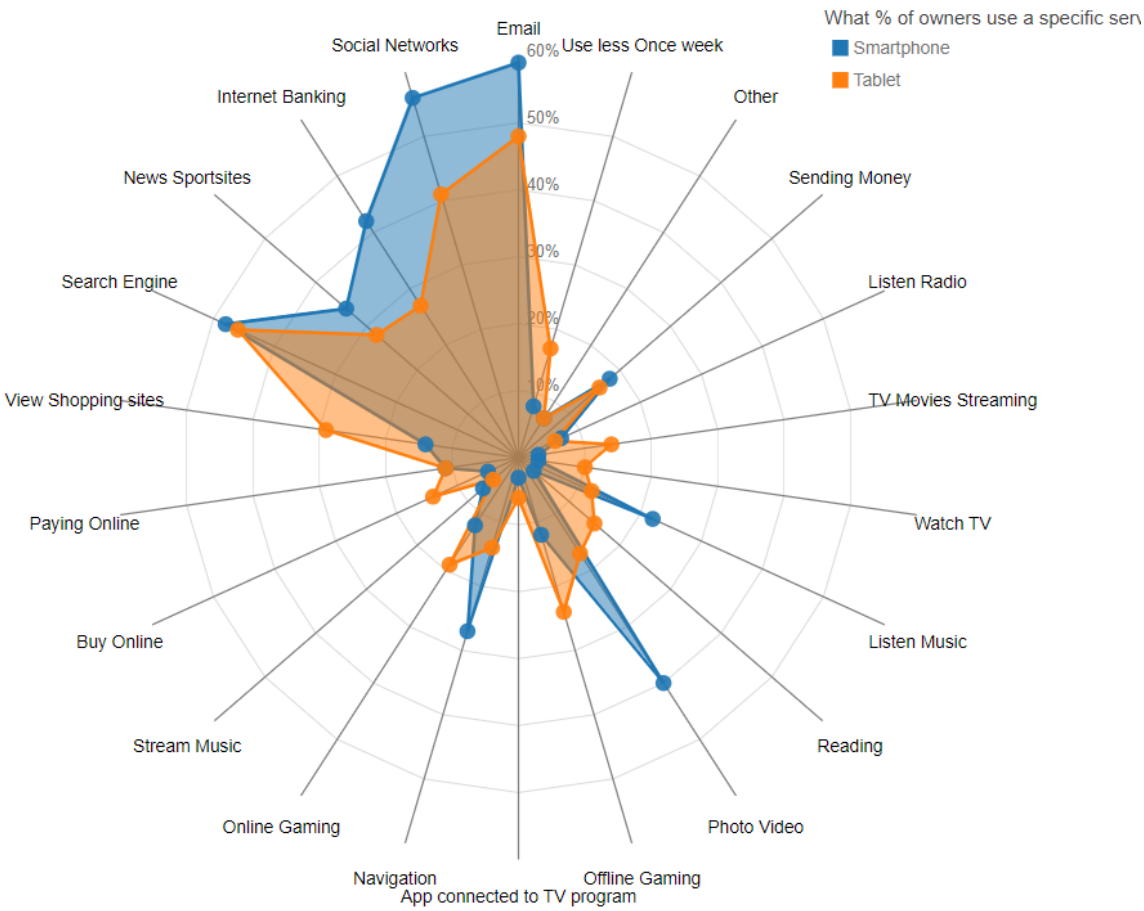
- **Dot plot**: Similar to a bar but use a point or symbol to indicate the value so it can include colour, area, shape to capture extra dimensions. Variation is the connected dot plot. AKA: 1-dimensional scatterplot
- 

- **Circle packing**: comparisons of values using area, shape, colour, layout
- 

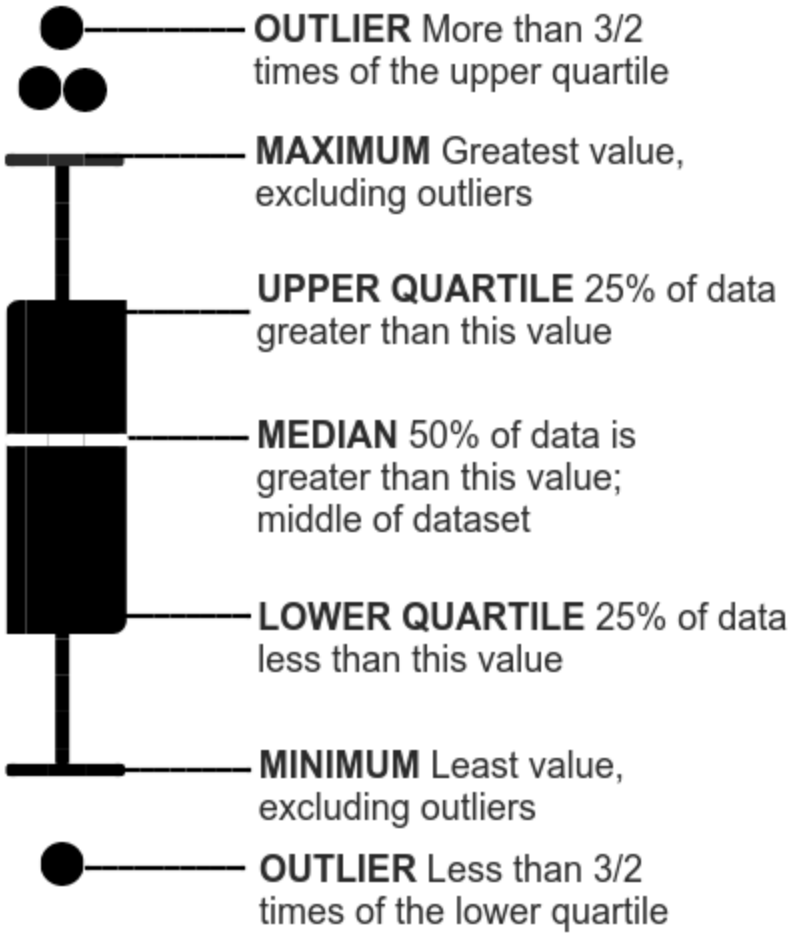
- **Polar chart**: (also radar or spider) radially plotted bar chart showing 3+ quantitative measures. Radar charts are able to display multivariate observations with an arbitrary number of variables.



- also radar or spide



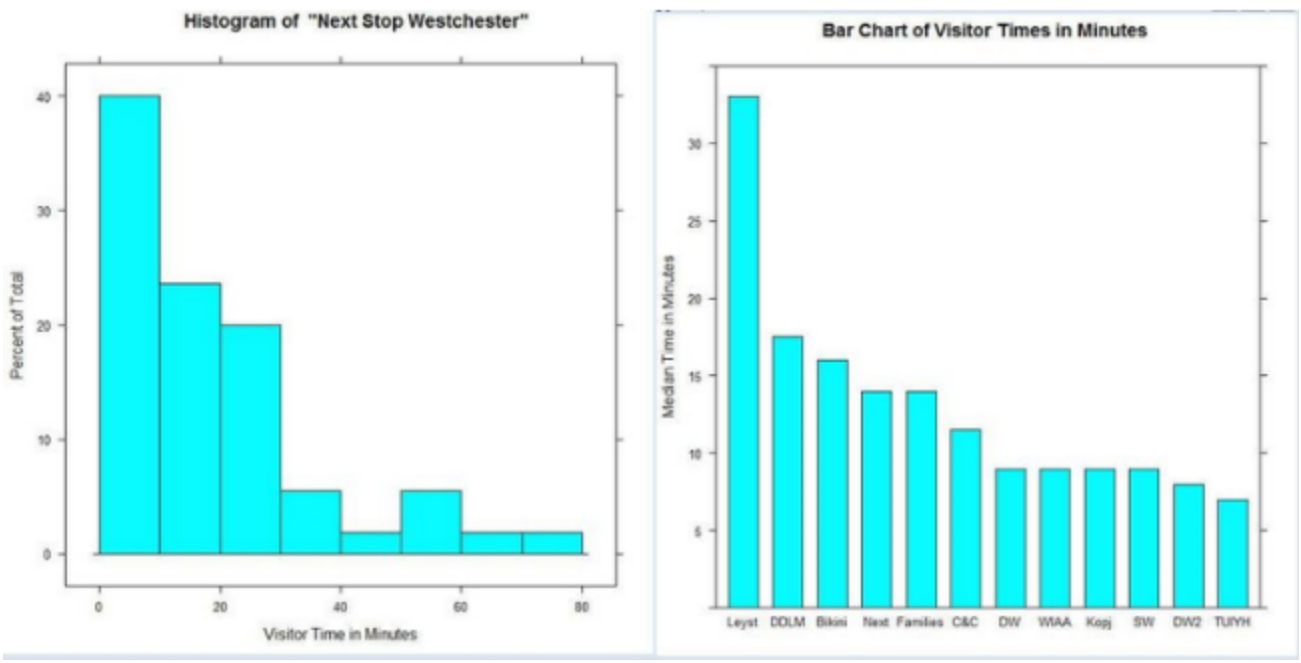
Distribution



- Box-and-Whisker plot: common in statistical analysis
- Histogram (not a bar chart): frequency and distribution
- Word cloud: frequency of concepts

Do not confuse a histogram with a bar chart.

he gap between the bars is often more pronounced in the bar chart and non-existent in the histogram (which uses `bins`)

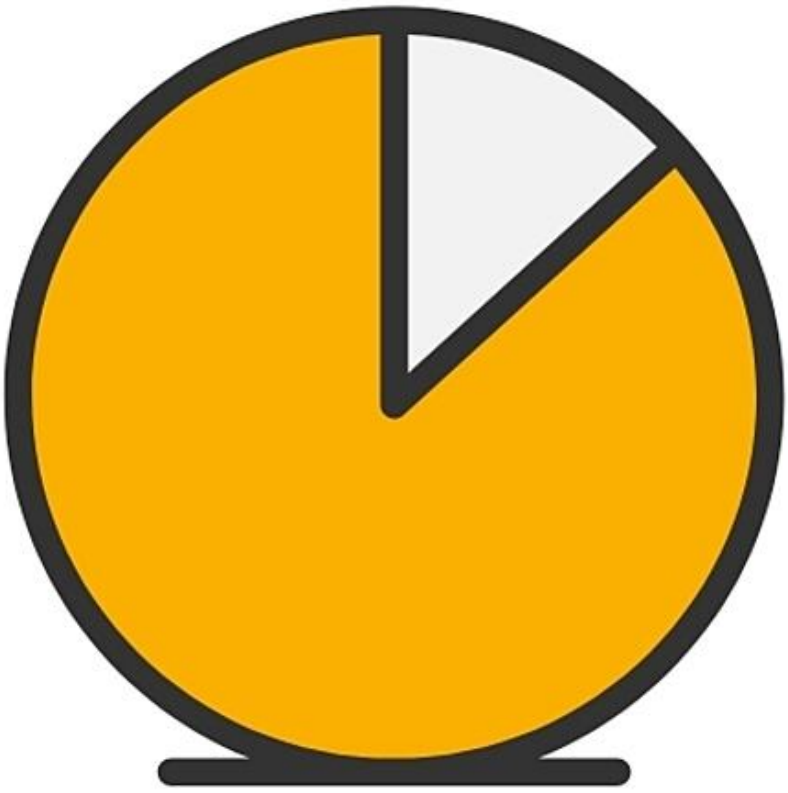


Various adudience more suited for certain charts.

Use these plots when you need:

- Histogram → rich visualisation of distributions
- Boxplot or Boxplots → comparing distributions between several groups

---



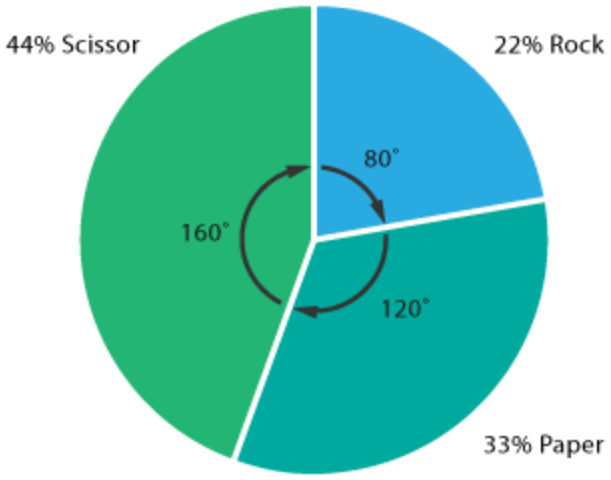
Hierarchical examples

Charting part-to-whole relationships and hierarchies.

- need to show the **composition or break down of something**. Especially where you have a category and percentages of a total (ie, they add up to 100%).

> Used where you need to show the composition or break down of something. Especially where you have a category and percentages of a total (ie, they add up to 100%)

Pie charts break down quantities using the angles (wedges) in a circle as illustrated in Figure 1 below.



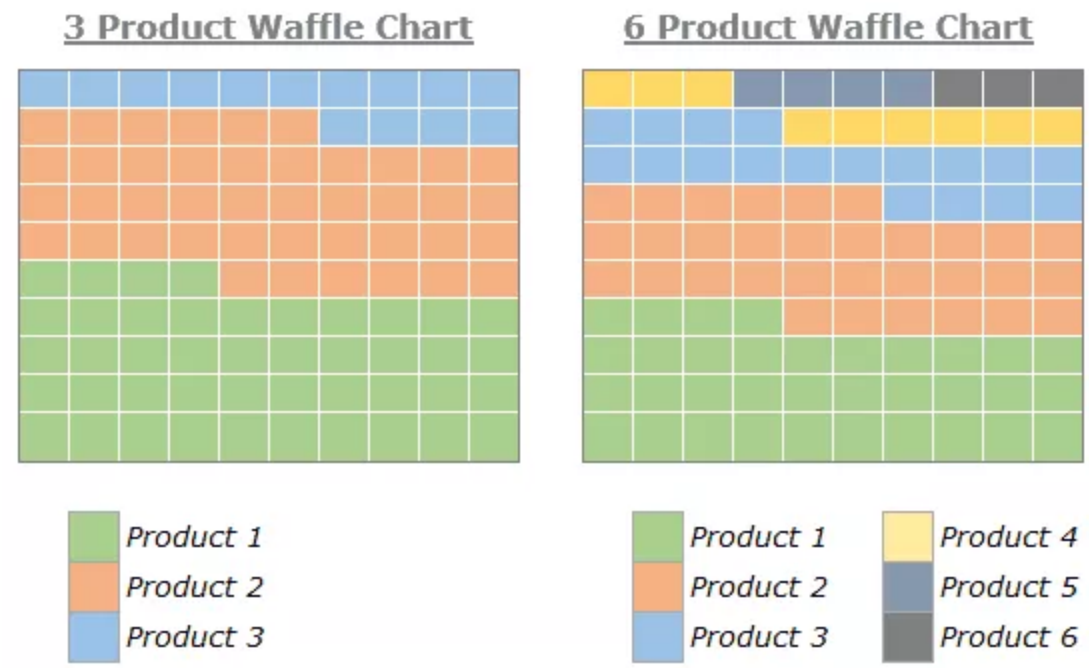
Data			
Rock	Paper	Scissor	TOTAL
2	3	4	9
To calculate percentages			
2/9=22%	3/9=33%	4/9=44%	100%
Degrees for each "pie slice"			
(2/9) x 360 = 80°	(3/9) x 360 = 120°	(4/9) x 360 = 160°	360°

"Pie charts are evil" vs moderating view from Storytelling data comm.

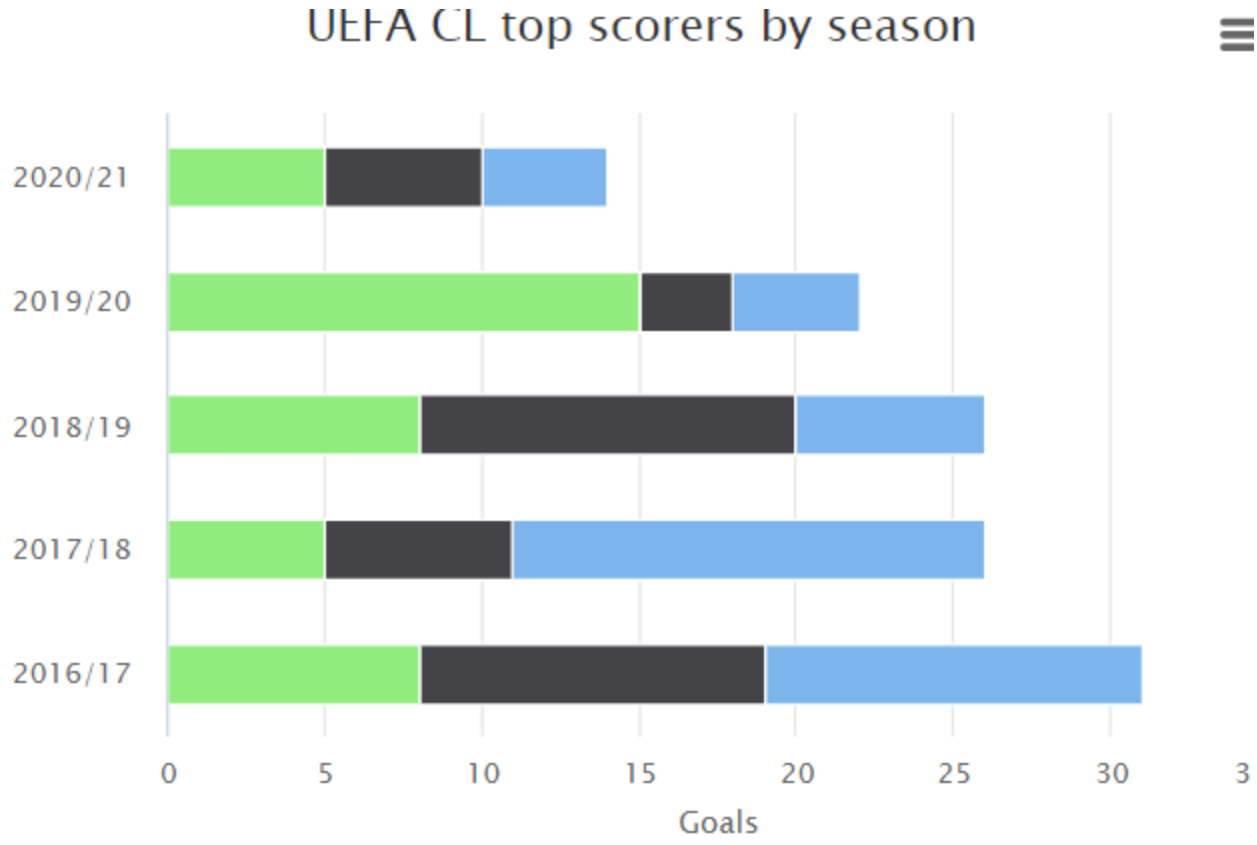
http://www.storytellingwithdata.com/blog/2017/1/10/an-updated-post-on-pies

Other examples:

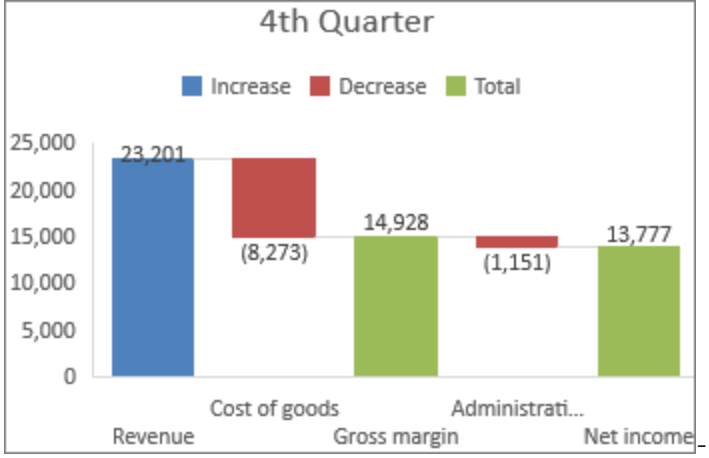
- **Waffle charts**: aka square pie, coloured grid squares to show quantities make it easier to compare or evaluate via counting rather than angles.



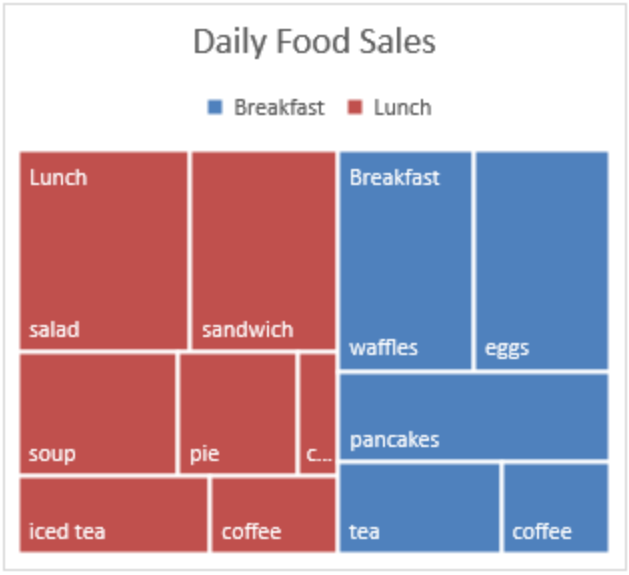
- **Stacked bar chart**: breakdown values within a bar. A variation of the stacked bar chart is the waterfall.



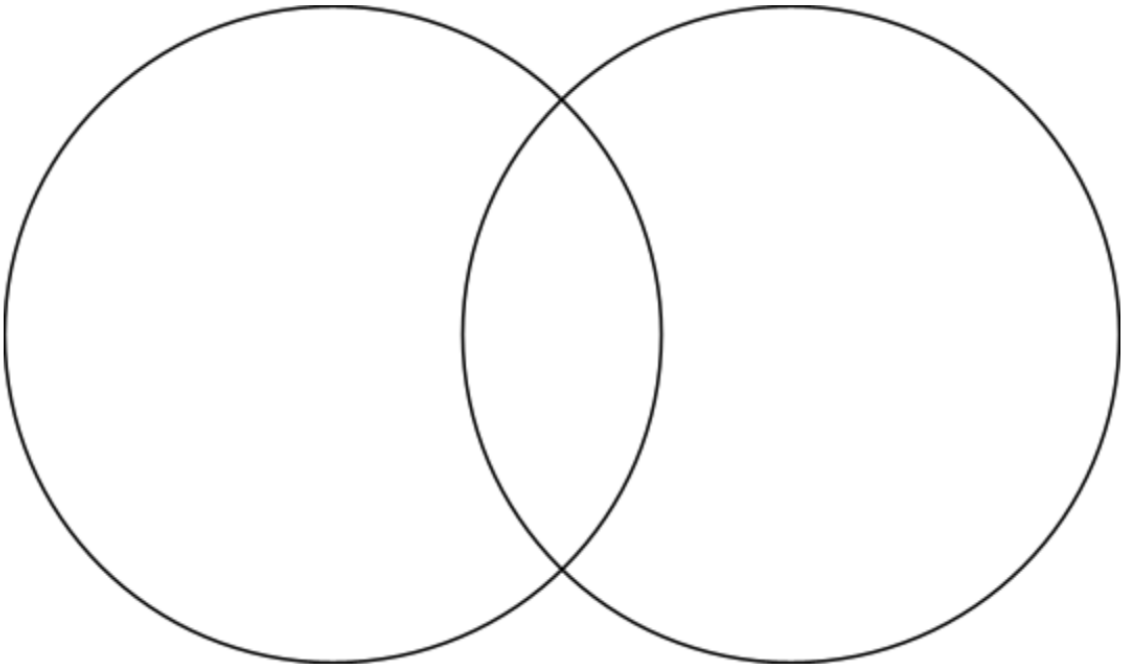
  - waterfall (variation of stacked)

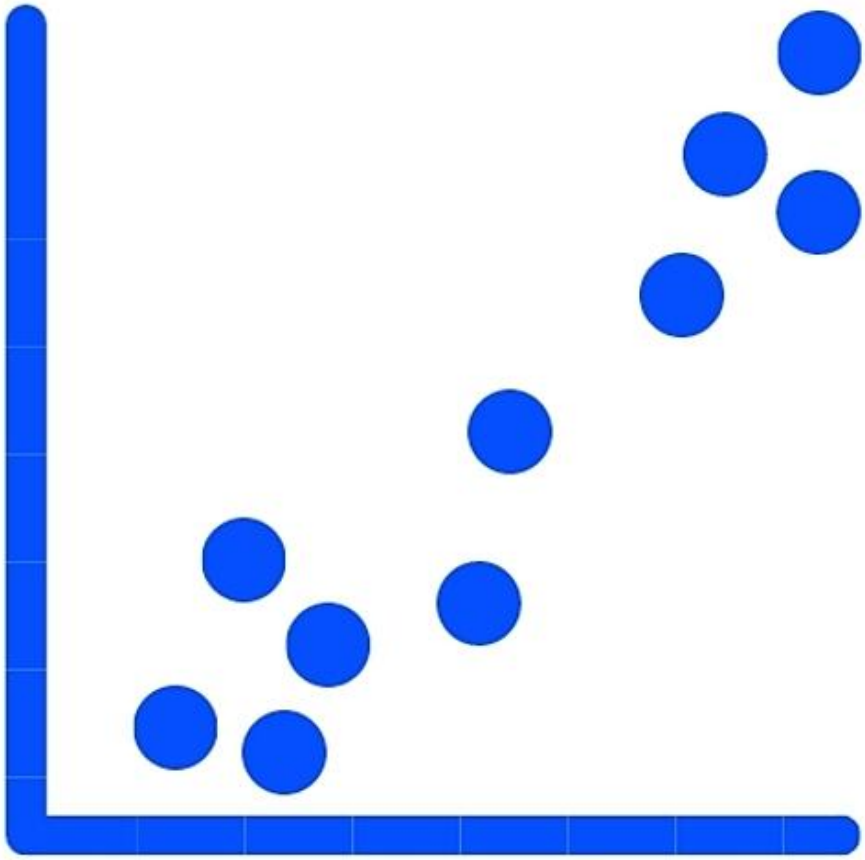


- **Treemap**: enclosed hierarchical display



- **Venn diagram**: relationships between sets and collections

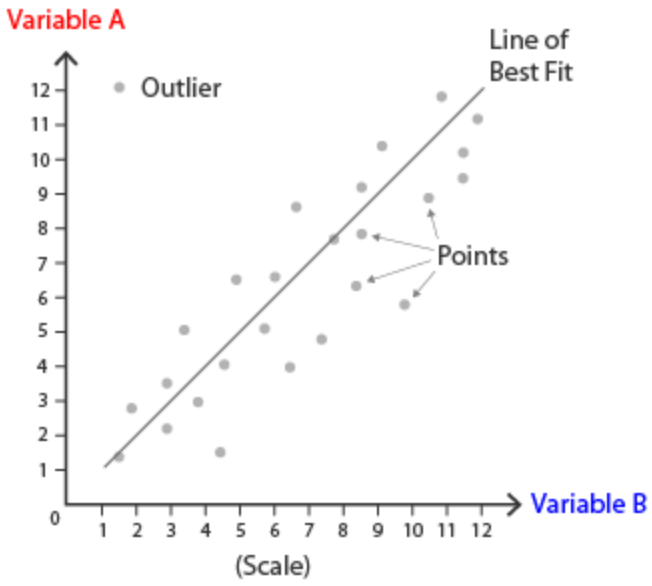


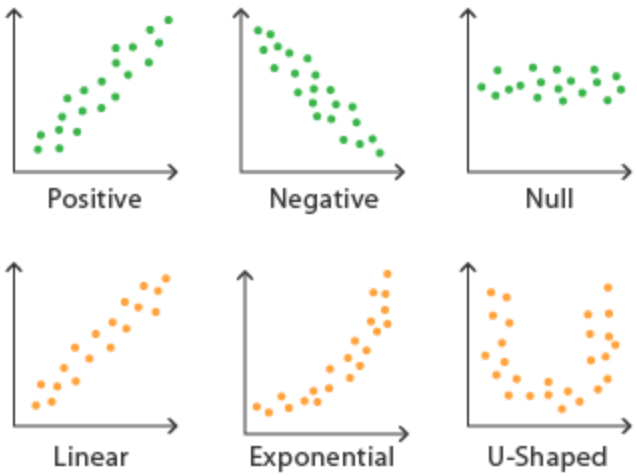


Relational

Relational charts are used to show the connections between data points.

The classic: **scatterplot**

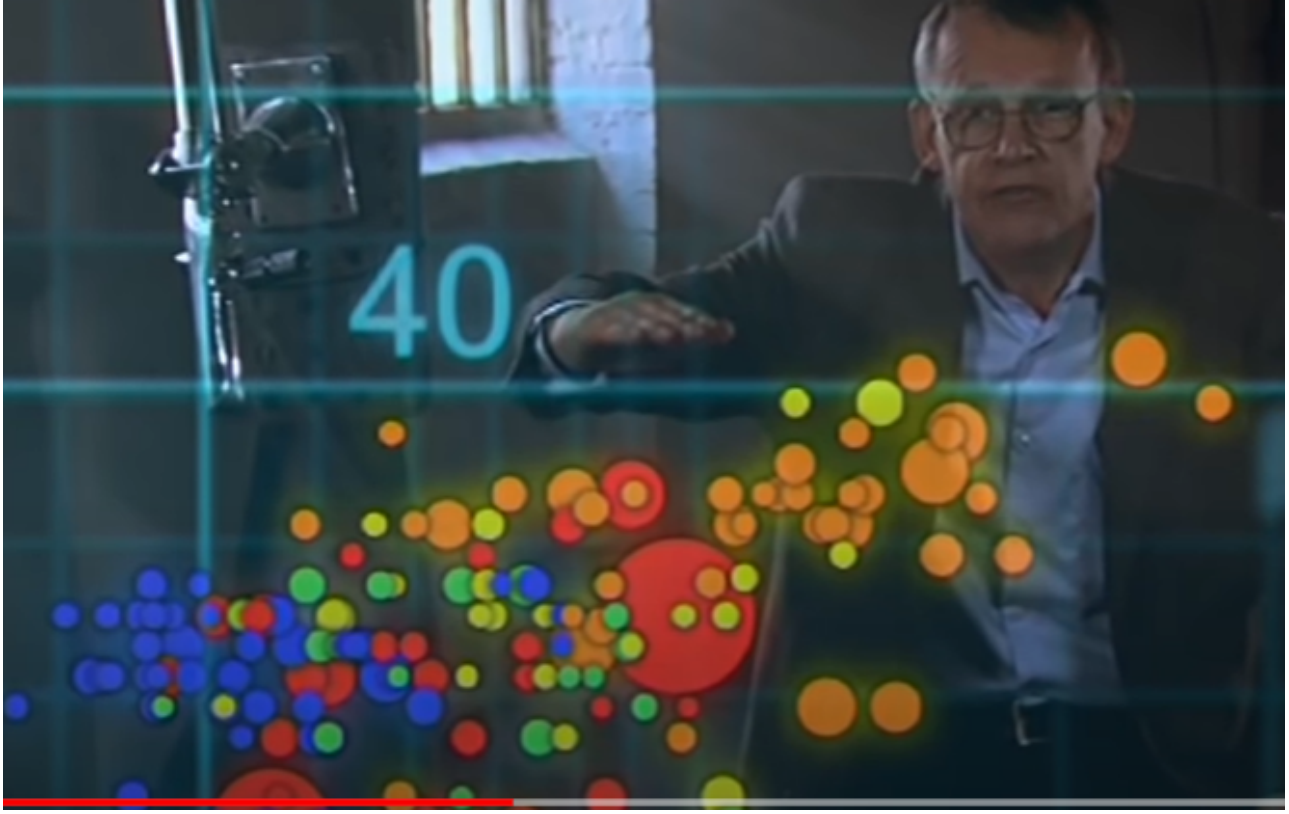




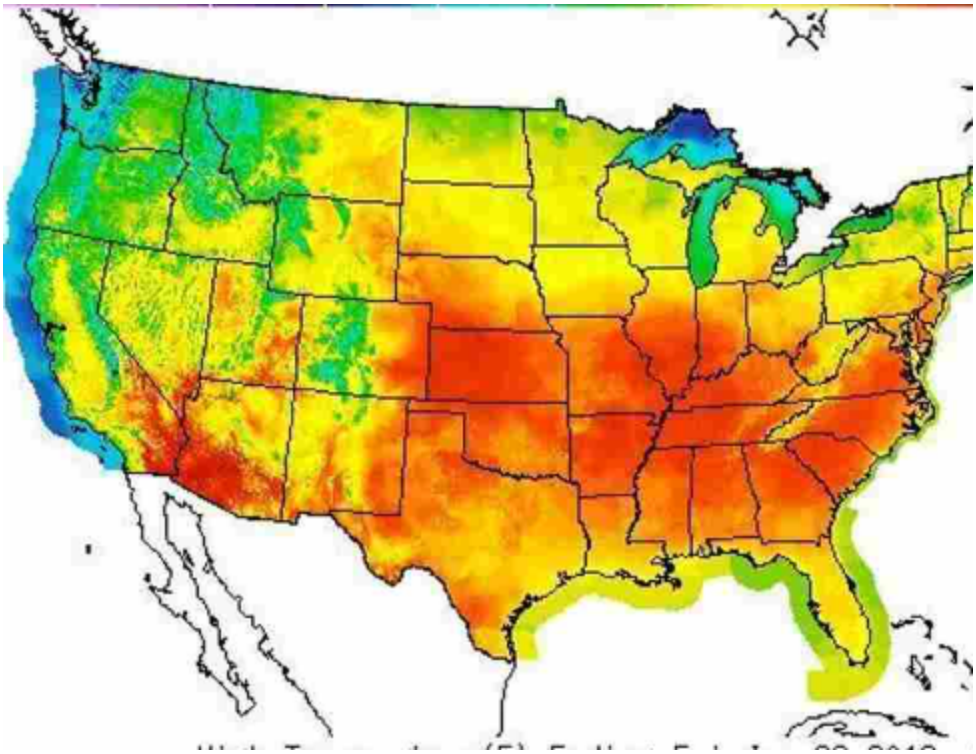
Correlation Strength:

Other examples:

- **Bubble plot**: relationship between 3 qualitative values (area, x position, y position)



- **Heat map**: quantitative values between 2 categorical dimensions (colour coded)



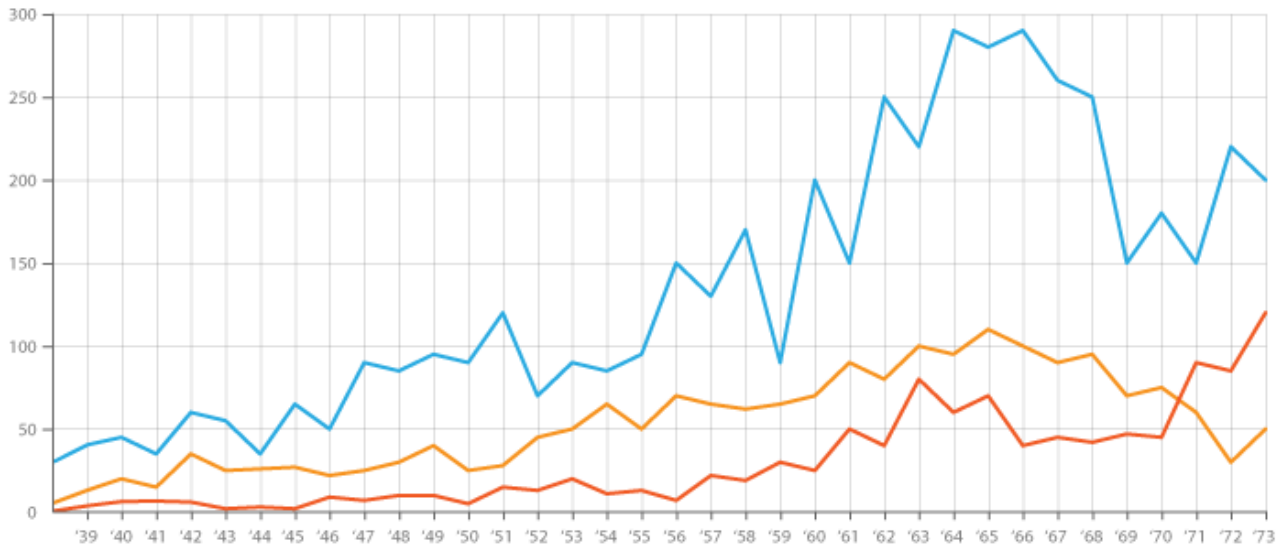
- **Matrix chart**: quantitative values between 2 categorical dimensions
- **Sankey diagram**: categorical composition and qualitative flows

---

Temporal examples

Trends

The classic: **line chart**



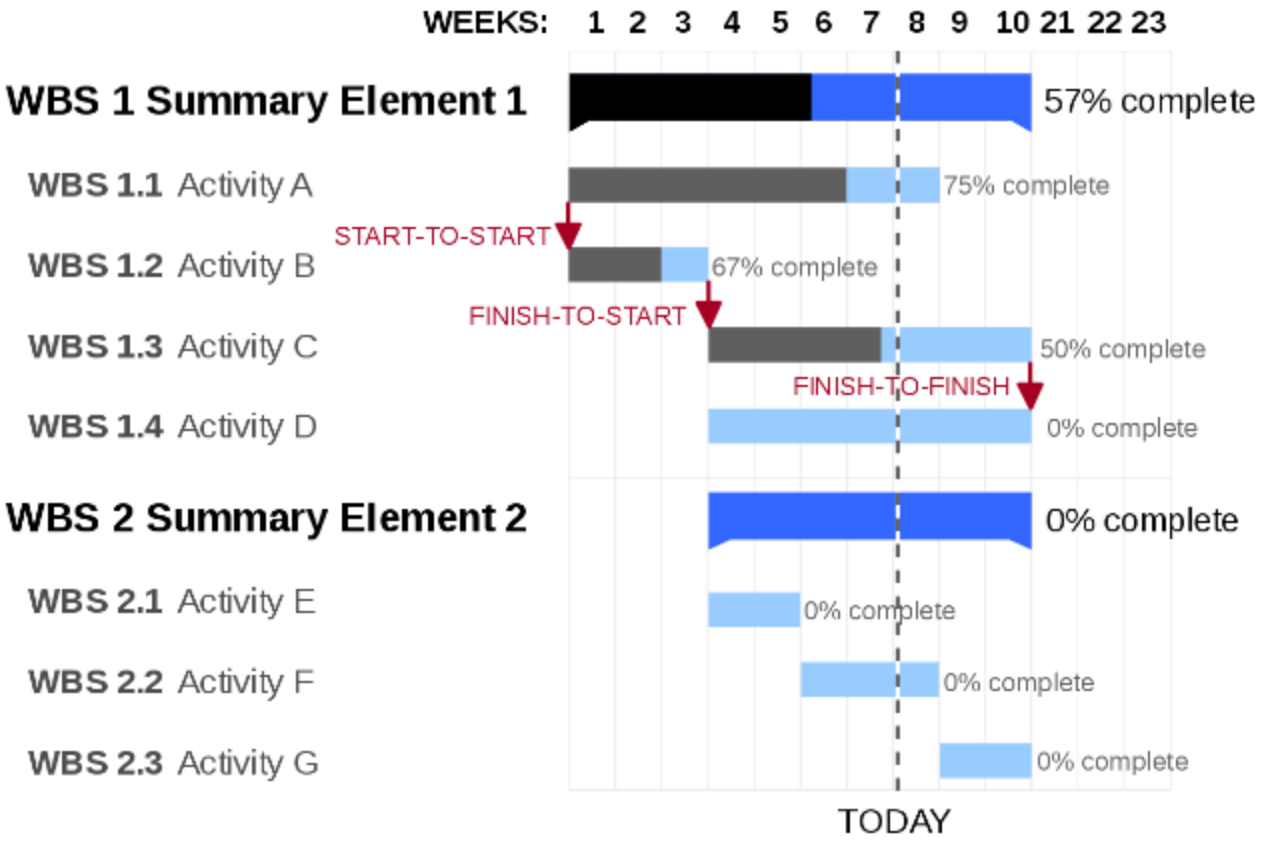
> A line chart shows the change in quantitative values over time (or possibly a sequence of equal intervals). Be careful not to use this where your x-axis has categories! This type of chart is also known as a stock chart.

- Line chart: change in quantitative values over time.
- Area chart: coloured in line chart :-).

- Stream graph: continuous changes in qualitative values in different categories over time.
- Slope graph: before and after changes in quantity over a number of variables.

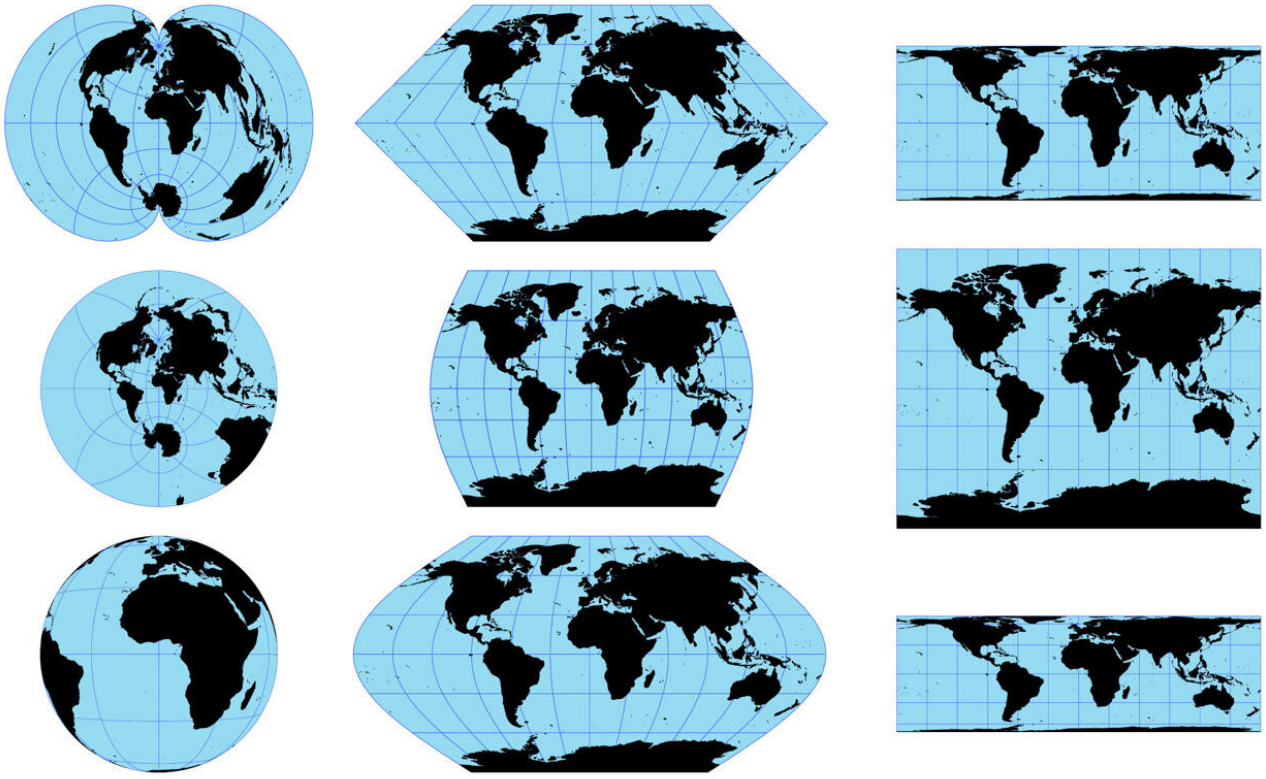
Activities:

This is a special type of temporal chart best represented by the **Gantt chart**



Spatial examples

- A word about geographic maps:

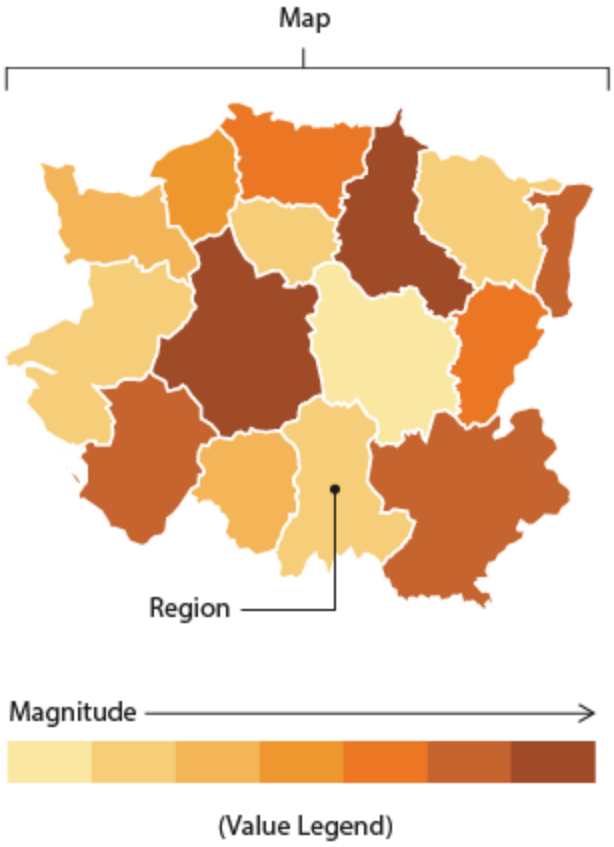


> Any use of a map in a visualisation relies heavily on the audience having a basic understanding of geography. Most people are familiar with the Mercator projection (the

> common map of the world).

There are two main ways to encode spatial data using a map: Overlays and Distortions.

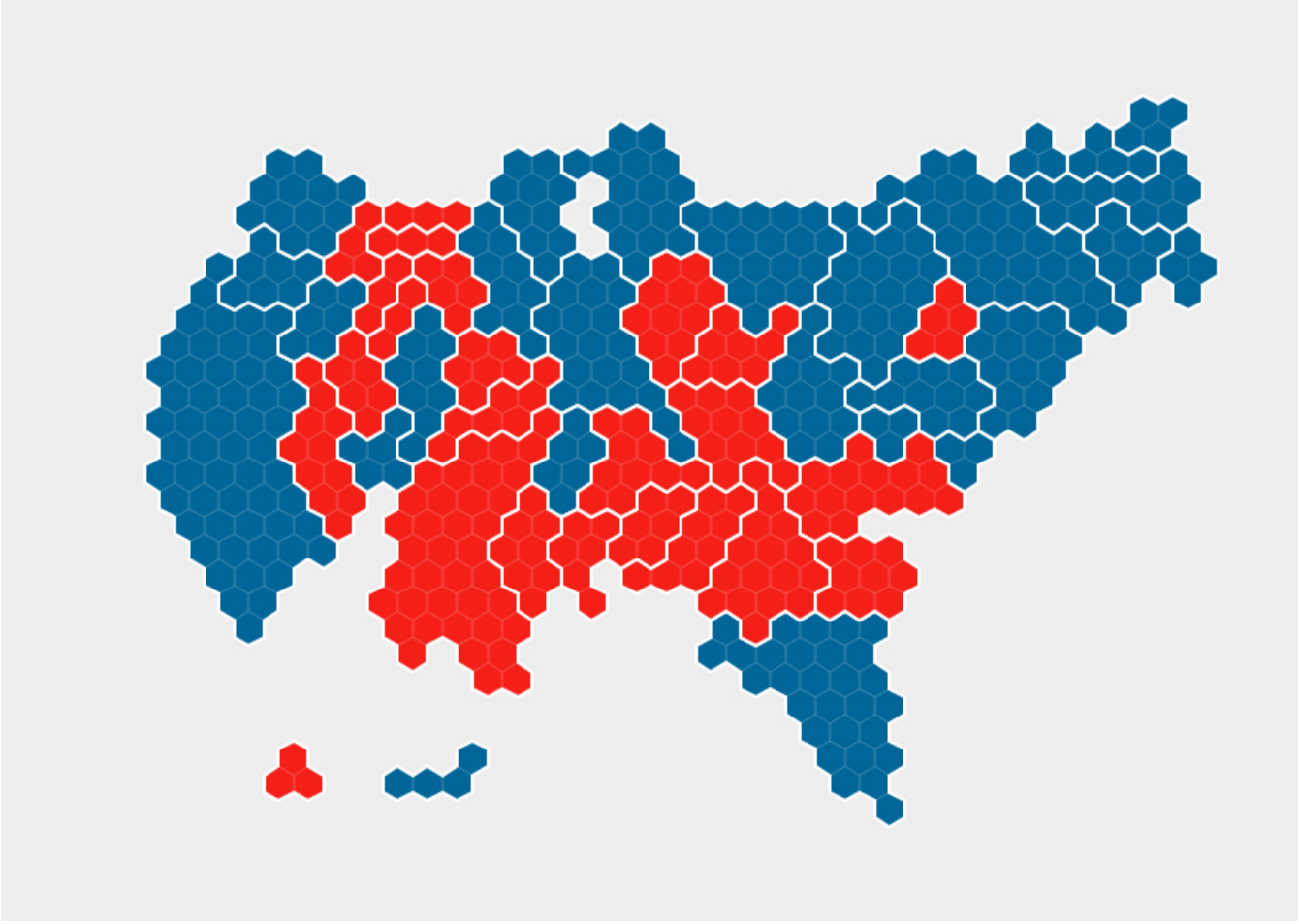
1. Overlays:



- The classic: **Choropleth map** A choropleth map encodes quantitative values for distinct spatial regions often using a colour progression. This is also known as a heat map.
- Isarithmic map: (also known as a contour map) quantitative values linking spatial regions. Commonly seen in **weather maps showing the connections between areas of pressure.**

2. Distortions

> Distortions play with the sizes of regions in the map or build a visual representation of the map layout using symbols. This relies on your audience knowing what the map should look like to understand the data being conveyed.



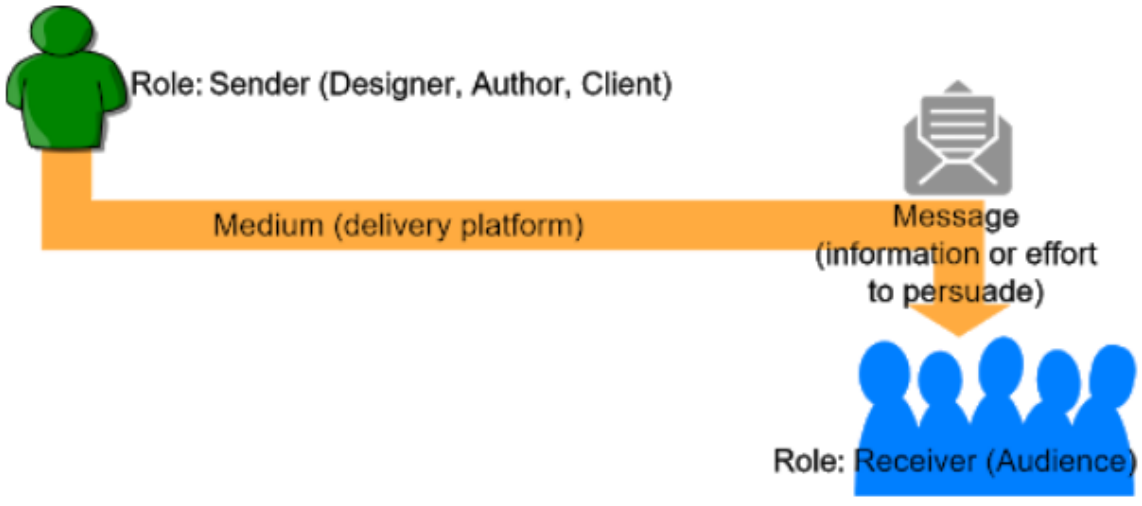
Area cartogram: distort map spatial regions to show value. Dorling Cartogram/Grid map: arrange regular shapes into a map using colour to indicate category.

wk1 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600222/quiz/introduction

What is your message?

> The best communications inspire your audience, engage their emotions and encourage them to act in response to your story. The story helps to create a memorable message. clear and simple messages are easy to follow

- Remember the participants in a visual communication from Data-driven Visualisations, Step 1.5?



- the general communication goals or motivations covered in the third course in Step 2.2: Information, Persuasion, Education or Entertainment

two different approaches to identifying your message:

What's the "big idea"?:

- Let's start with some basics: who, what and how. Knaflic ("Storytelling with data") uses these three questions to help you form "The Big Idea":

Try asking yourself the following **three questions:**

1. To whom are you communicating?
2. What do you want your audience to know or do?
3. How can you use data to help you make your point?

"Storytelling with Data" (Knaflic, 2016)

Editorial Brief: **the reason that the visualisation exists**

Recall that we identified two reasons for visualisation of data;

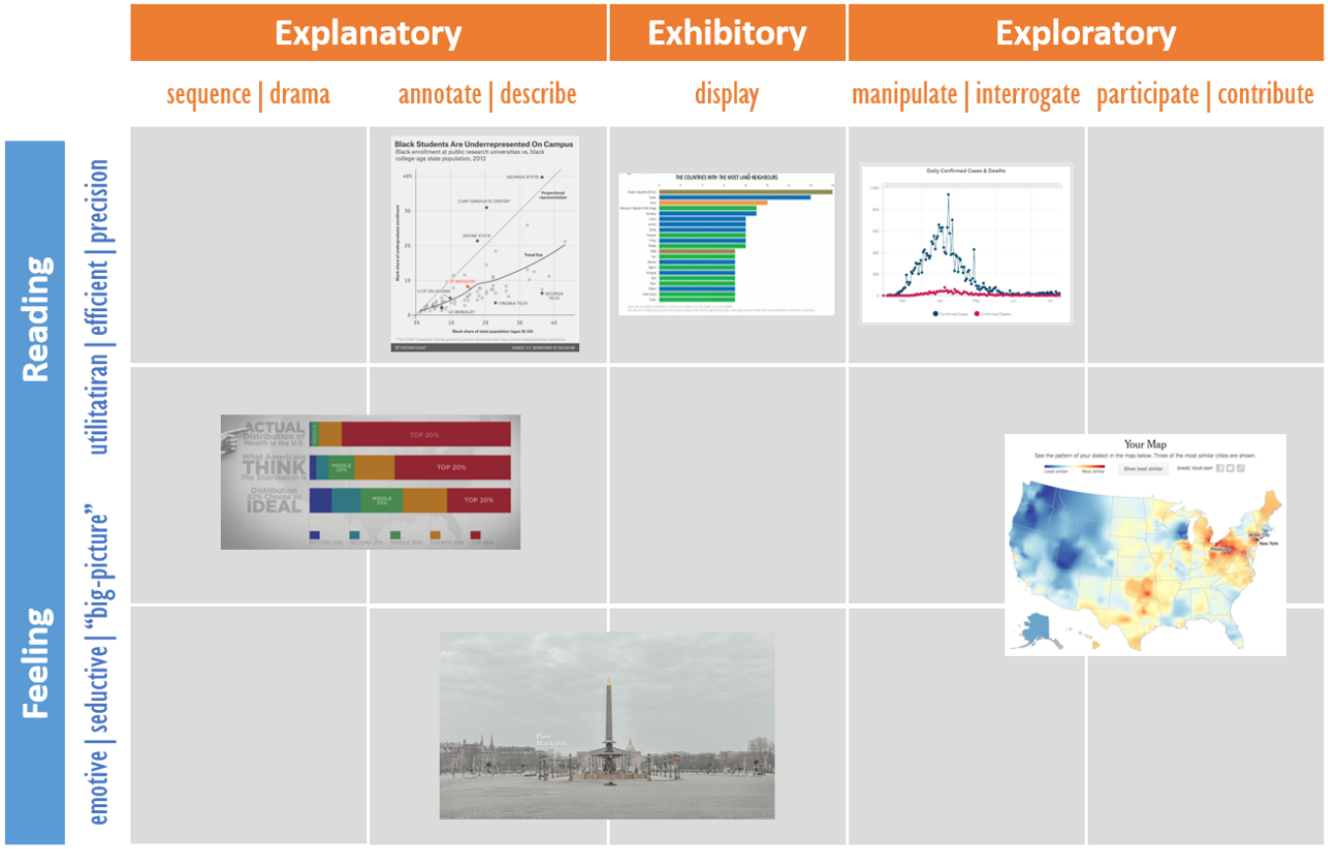
1. exploratory
2. explanatory.

Kirk adds a third option **exhibitory**

1. exhibitory: the visualisation is displayed to the viewer without explicit guided explanations due to either the context (public space) or the amount of text or design highlighting included on the visualisation.

This means that the viewer is doing the work to interpret the visualisation but they don't have the interactive options that an exploratory visualisation would offer.

...skipp... https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600223



1. https://fivethirtyeight.com/features/mizzous-racial-gap-is-typical-on-college-campuses/
2. Simple bar chart showing borders of land locked countries (Kirk Figure 3.7)
3. https://covid19ireland-geohive.hub.arcgis.com/
4. https://www.youtube.com/watch?v=QPKKQnijnsM
5. https://www.nytimes.com/interactive/2014/upshot/dialect-quiz-map.html?_r=0
6. https://www.nytimes.com/interactive/2020/03/23/world/coronavirus-great-empty.html

Summary:

Hopefully looking at the processes from **Knaflic** and **Kirk** will give you some ideas about how to identify and specify the message and motivation behind your visualisations.

To simplify the process try the following four steps when you are tackling an explanatory visualisation project:

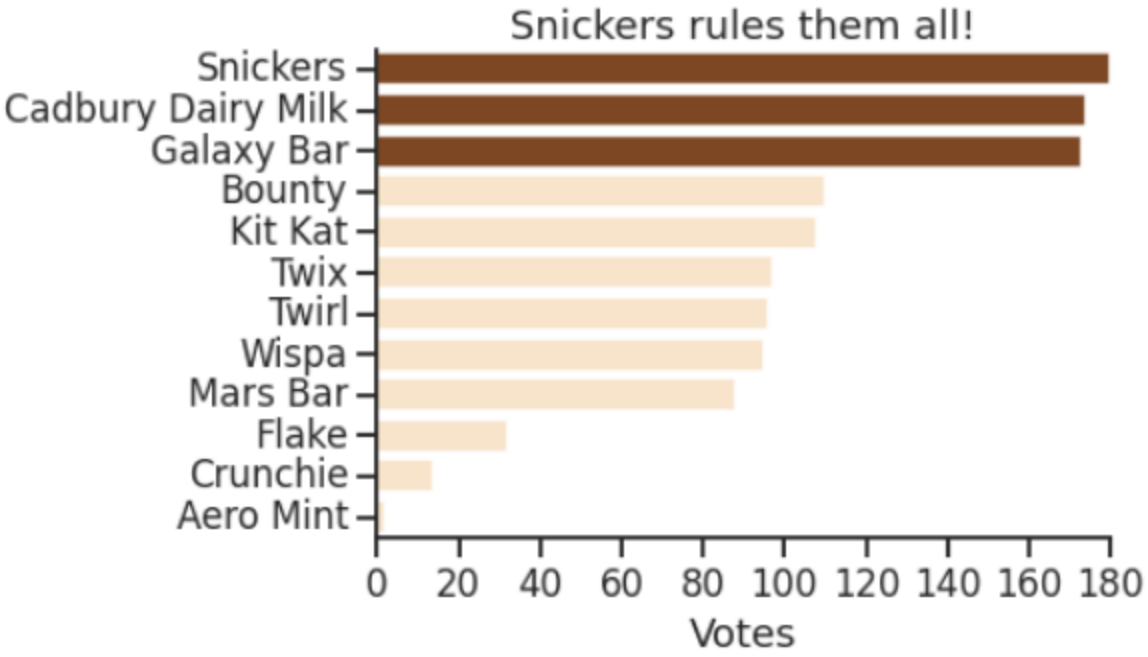
1. **The Message**: Ask yourself some questions to understand the purpose and message of your visualisation.
2. **The Receiver**: Consider the audience, their motivations and their viewing context.
3. **The Medium**: What are you delivering? Static Image, Presentation, Poster, App, Dashboard?
4. (Optionally) Use a **storyboarding or mapping** method to fully create your narrative or the editorial brief.

---

identify and define the message for your visualisation

https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600224

example:EXAM



1. In a single (short) sentence what do you think is the message?
2. Which of the following words would you use to describe the graph? sequence, drama, annotate, describe, manipulate, interrogate, participate, contribute, emotive, seductive, “big-picture”, utilitarian, efficient, precision. Use the words to decide where the visualisation fits on Kirk’s “purpose” map.
3. On a scale of 1 (terrible) to 5 (very effective), how would you rate the graph and why?

Now let’s go the other way. Given the briefing below, how would you approach the task?

1. Use the **Big Idea template** and answer the following questions: (a)Who is your audience? (b) What is at stake? https://drive.google.com/file/d/1WJDLkMVTxlDHyeGWzeKQmSiXdmlxQSAg/view
2. Write a single sentence capturing your big idea.
3. Use **Kirk’s purpose map** and decide where you would aim to place your graph. https://github.com/suzannelittle/ca682i/blob/master/resources/blank-purpose-map.png

---

BIG

the **BIG IDEA** worksheet

storytelling with data®

Identify a project you are working on where you need to communicate in a data-driven way. Reflect upon and fill out the following.

PROJECT _____

WHO IS YOUR **AUDIENCE**?

(1) List the primary groups or individuals to whom you'll be communicating.

(2) If you had to narrow that to a *single person*, who would that be?

(3) What does your audience care about?

(4) What action does your audience need to take?

WHAT IS **AT STAKE**?

What are the *benefits* if your audience acts in the way that you want them to?

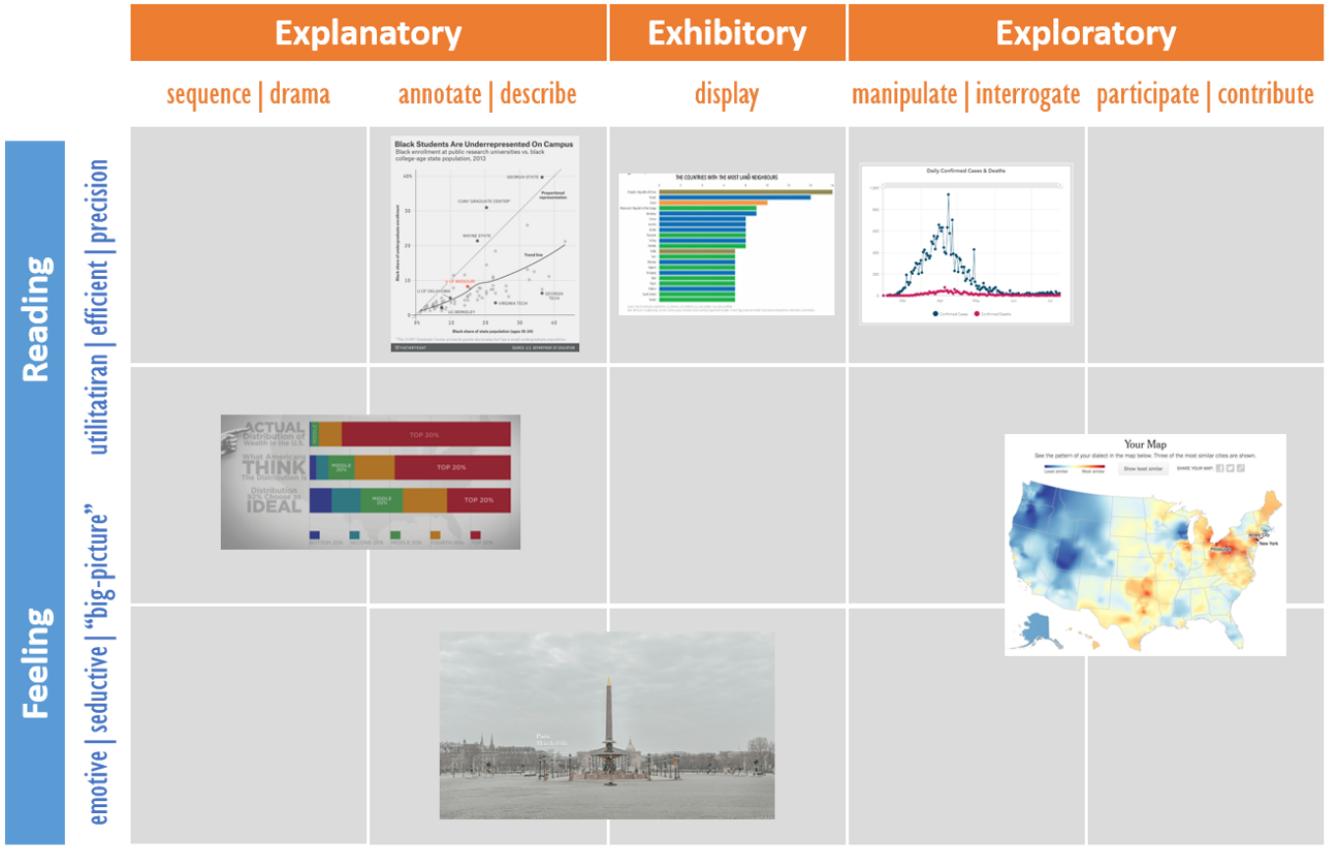
What are the *risks* if they do not?

FORM YOUR **BIG IDEA**

It should:

(1) articulate your point of view,
(2) convey what's at stake, and
(3) be a complete (and single!) sentence.

purpose map



1. https://fivethirtyeight.com/features/mizzous-racial-gap-is-typical-on-college-campuses/
2. Simple bar chart showing borders of land locked countries (Kirk Figure 3.7)
3. https://covid19ireland-geohive.hub.arcgis.com/
4. https://www.youtube.com/watch?v=QPKKQnijnsM
5. https://www.nytimes.com/interactive/2014/upshot/dialect-quiz-map.html?_r=0
6. https://www.nytimes.com/interactive/2020/03/23/world/coronavirus-great-empty.html

Data Visiualisation Guides:

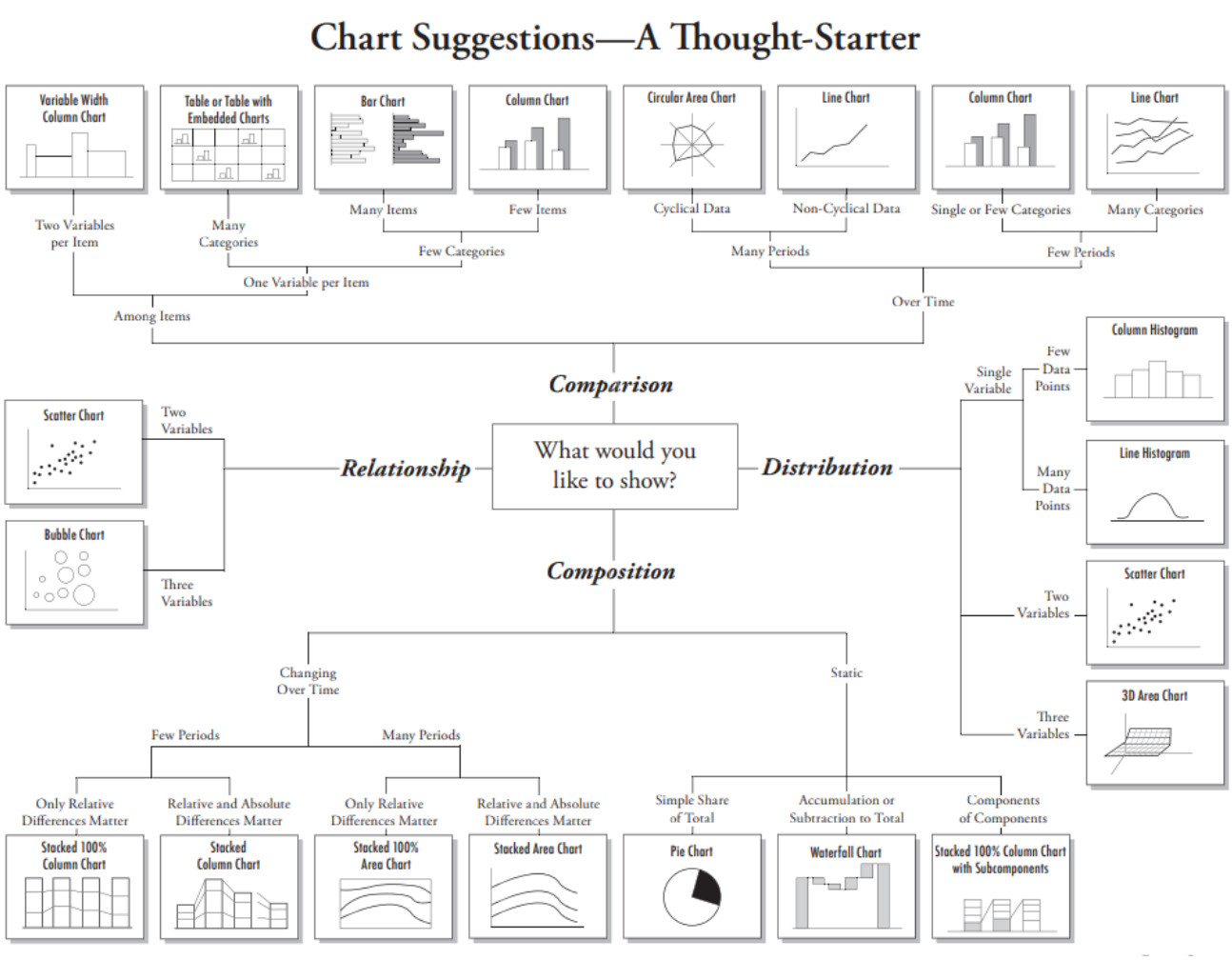
The Chartmaker Directory: based on Kirk (2016) this includes implementations in many different tools for each graph type. http://chartmaker.visualisingdata.com/

The Dataviz Catalogue: an interactive website with a catalogue of different chart types.

The Dataviz Project: browsable graphs indexed by family, input, function and shape.

choosing a chart:

**dont forget to check BOX-PLOT/WHISKER-BOX-PLOT



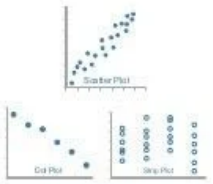
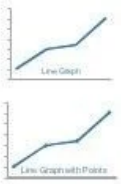
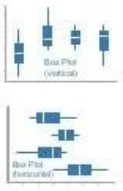
Graph Selection Matrix

If you know the relationship that you'd like to present then Stephen Few links relationships to value-encoding objects (points, lines, bars or boxes).

```
Time series → vertical bars, lines, points, vertical box plots
ranking → bars (vertical or horizontal), points
part-to-whole → bars (vertical or horizontal), stacked bars (vertical or
horizontal)
deviation → bars (vertical or horizontal), lines (when showing through time)
single distribution → bars (vertical or horizontal), lines, points
multiple distribution → lines, boxes (vertical or horizontal), points (strip
plots)
correlation → points, bars (as table lens)
geospatial → points (varying size), points/area (varying colour), lines
nominal comparison → bars (vertical or horizontal) or points
```

perceptual edge

Graph Selection Matrix

Featured Relationships	Value-Encoding Objects: Points	Lines	Bars	Boxes
				
Time Series Values display how something changed through time (yearly, monthly, etc.)	Yes (as a *dot plot*, when you don't have a value for every interval of time)	Yes (to feature overall trends and patterns and to support their comparisons)	Yes (vertical bars only, to feature individual values and to support their comparisons)	Yes (vertical boxes only, to display how a distribution changes through time)
Ranking Values are ordered by size (descending or ascending)	Yes (as a *dot plot*, especially when the quantitative scale does not begin at zero)	No	Yes	Yes (to display a ranked set of distributions)
Part-to-Whole Values represent parts (proportions) of a whole (for example, regional portions of total sales)	No	No	Yes	No
Deviation The difference between two sets of values (for example, the variance between actual and budgeted expenses)	Yes (as a *dot plot*, especially when the quantitative scale does not begin at zero)	Yes (when also featuring a time series)	Yes	No
Distribution Counts of values per interval from lowest to highest (for example, counts of people by age intervals of 10 years each)	Yes (as a *strip plot*, to feature individual values)	Yes (as a *frequency polygon*, to feature the overall shape of the distribution)	Yes	Yes (when comparing multiple distributions)
Correlation Comparison of two paired sets of values (for example, the heights and weights of several people) to determine if there is a relationship between them	Yes (as a *scatter plot*)	No	Yes (as a *table lens*, especially when your audience is not familiar with *scatter plots*)	No
Geospatial Values are displayed on a map to show their location	Yes (as bubbles of various sizes on a map)	Yes (to display routes on a map)	No	No
Nominal Comparison A simple comparison of values for a set of unordered items (for example, products, or regions)	Yes (as a *dot plot*, especially when the quantitative scale does not begin at zero)	No	Yes	No

www.PerceptualEdge.com Derived from the book *Show Me the Numbers, Second Edition*

EXAM

Scenario:

SCENARIO: The Simpsons is your favourite cartoon and you want to see how its popularity has changed over time. Fortunately, someone has created a dataset with the IMDB rating and the number of US viewers for

each episode.

DATA: simpsons_episodes.csv
https://github.com/suzannelittle/ca682i/blob/master/data/simpsons/simpsons_episodes.csv

YOUR TASK:

Use some of the resources from the previous step to identify what you think is the most appropriate chart choice for this scenario and data.

Roughly sketch (using pencil and paper) what you think the graph will look like. Make sure you indicate what data will be used on the different axes and encoded in the marks.

Use whatever tool you like (spreadsheet, python, etc.) to create the graph using the data.

Does your graph fulfil the brief? Could you make it look the way you wanted to?

Tell us about your graph in the comments section.

---

wk1 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600227/quiz/introduction

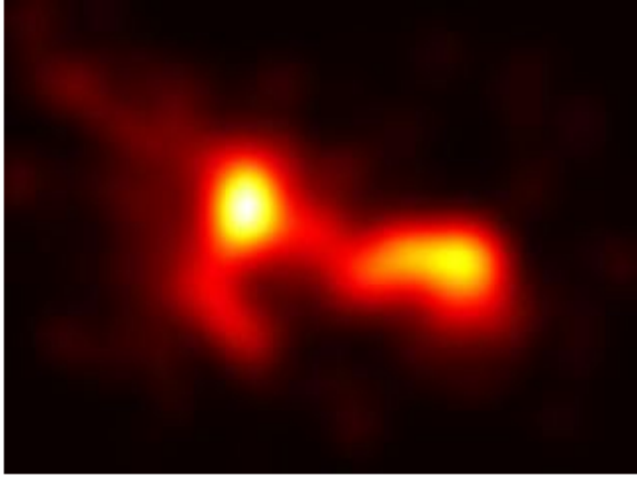
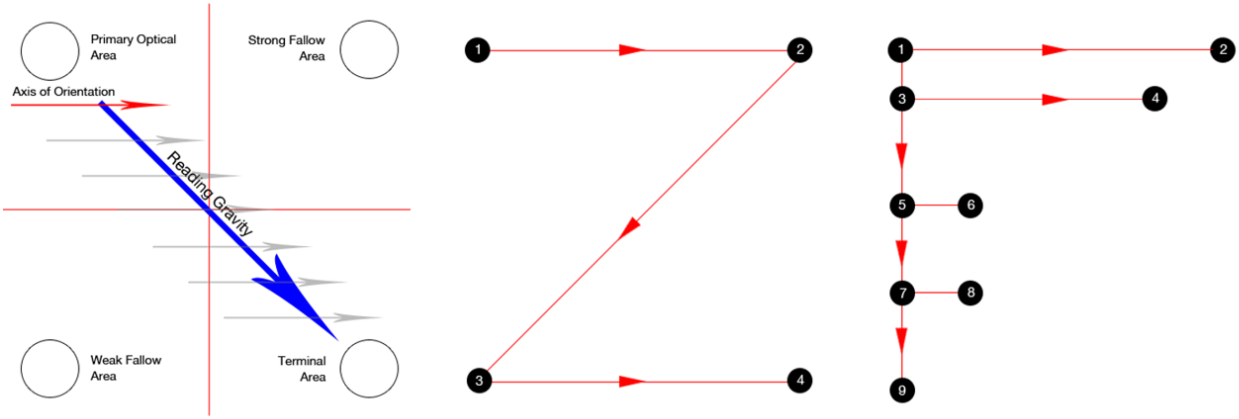
You should now be able to:

- Describe the different categories of graphs/charts.
- Identify specific graph types for each category.
- Select an appropriate graph type given a visualisation scenario.
- Discuss the uses of and issues with box plots and histograms.

Week 2: Attention (or Distraction)

Attention

- visual attention called the spotlight or searchlight model
- The size of the field of view (the spotlight) can vary (a zoom lens metaphor) and this can be influenced by the density of data present and the stress level of the viewer.
- grab attention with popup affects (flashing colour, movement ect.)
- Our eyes move in what is called a series of saccades (rapid movement of the eye between fixed points)
- measure eye-gaze tracker

-  
- how we segment a document, dividing up the visual field and guiding attention such as the Z pattern, Gutenberg and F patterns (see Figure 3 below)
- 

  Gutenberg (L), Z-pattern (M), F-pattern (R) for visual attention
- change blindness (gorilla video)

Preattentive features

- Some visual properties can be detected very very quickly. These are called “preattentive features”.
- This is an example of using a preattentive attribute (intensity of colour) to highlight some data:

035219248730515
708029135238051
337920714842415
119665329034538

-
- vs

0352192487305157080291352380513379207148424151196653290345388

There are a number of attributes that are **preattentive**

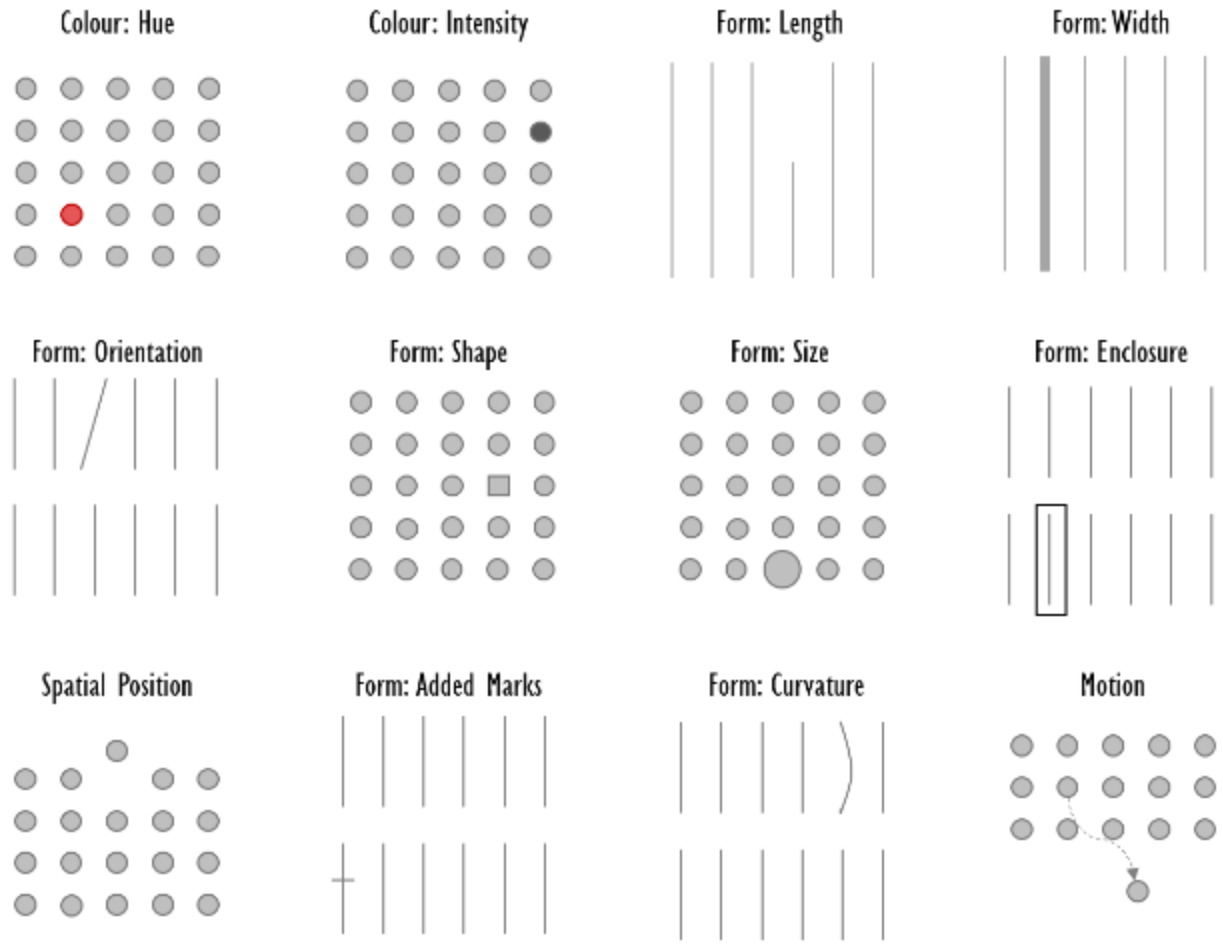
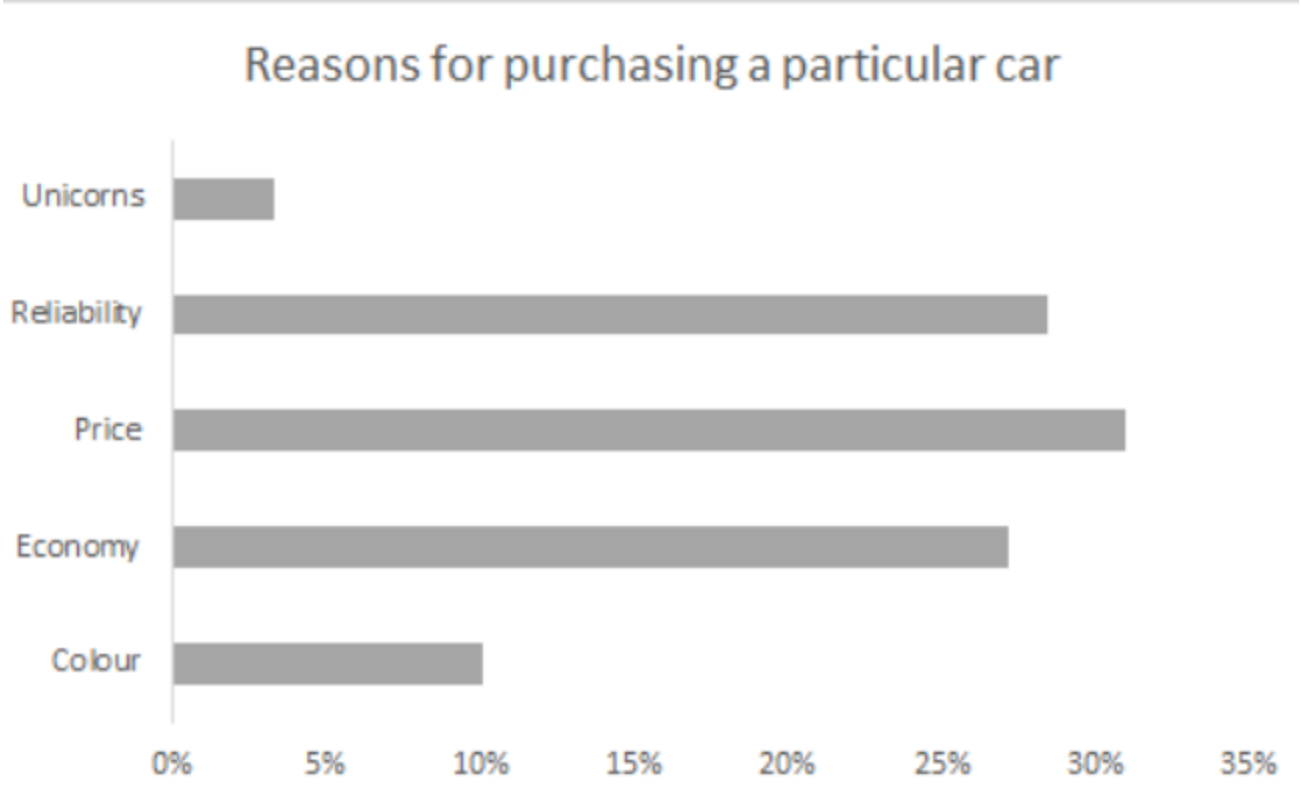


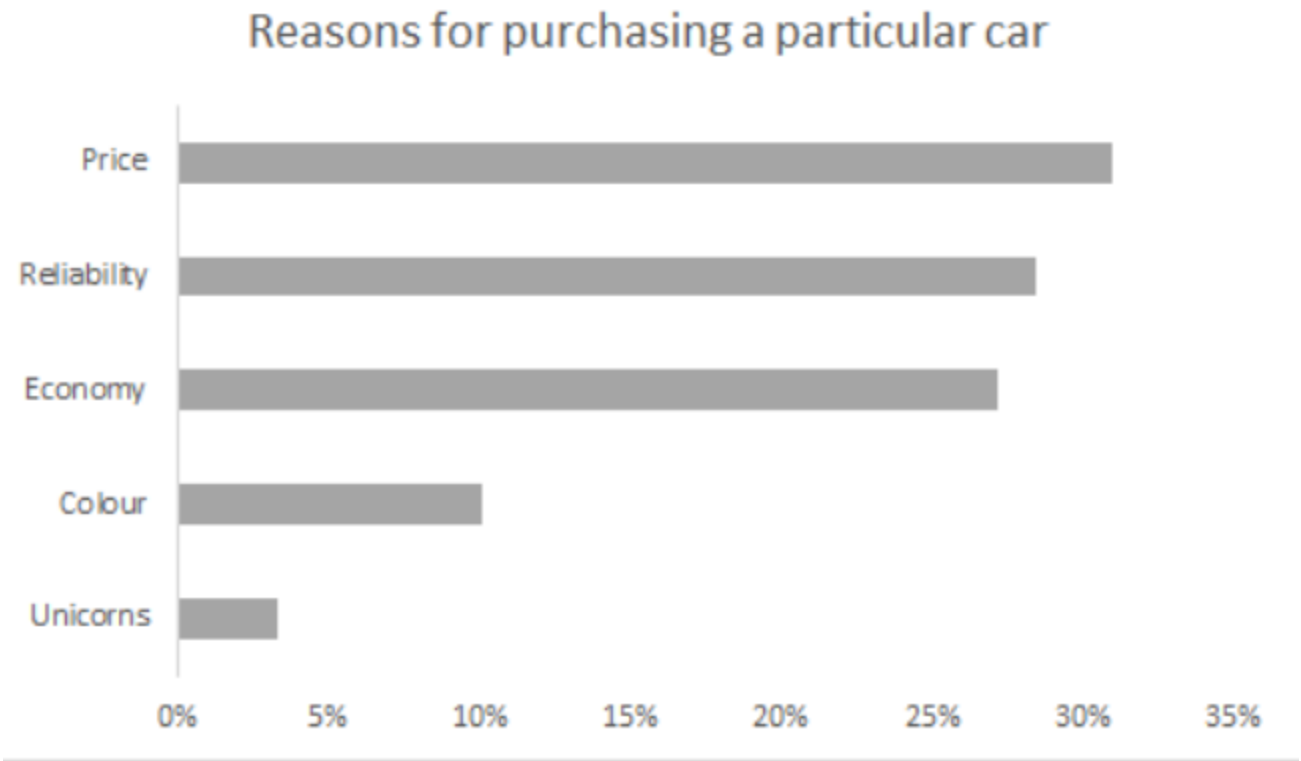
Figure 1: Common preattentive attributes (from Few (2004), Healy and Enns (2012) and Knaflic (2016)).

easy to confuse the preattentive features by overusing them or not making them distinct enough.

how Colour, Form and Spatial Position are used in visualisations

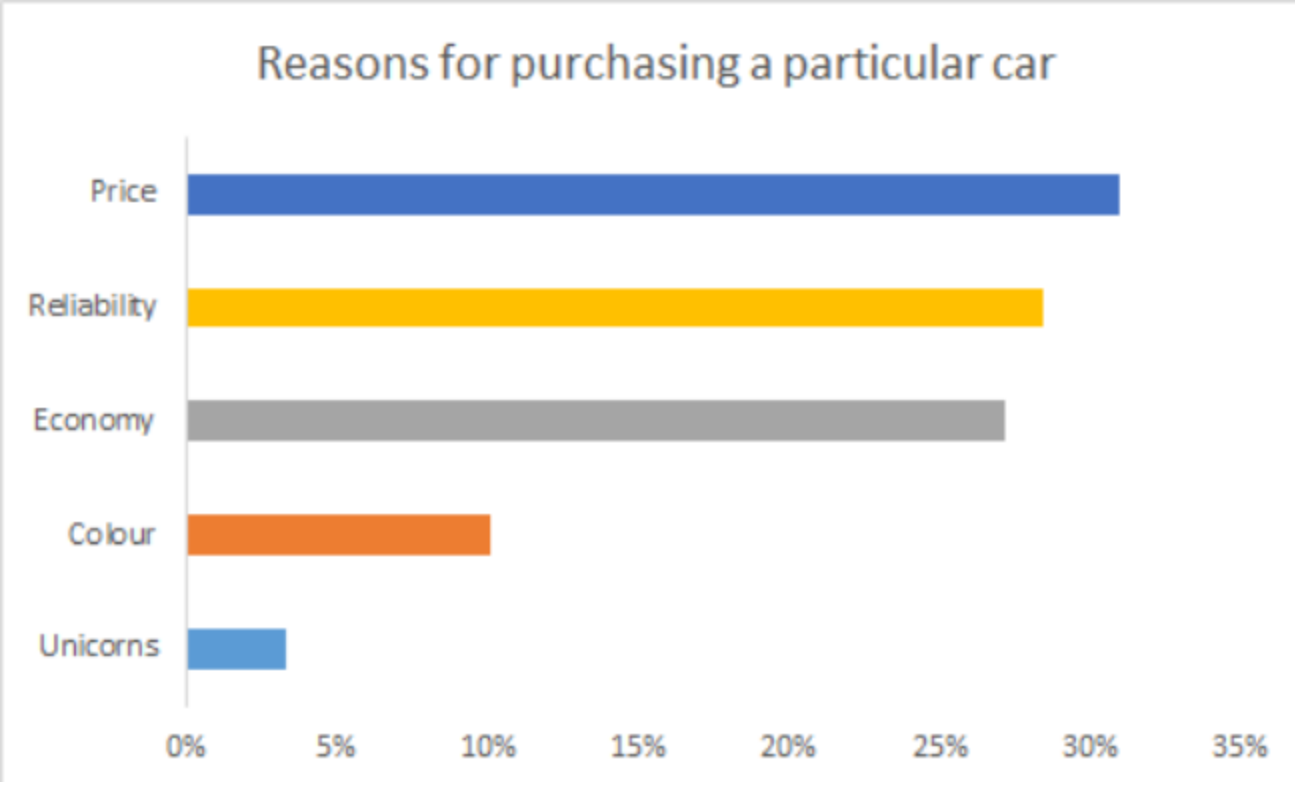


improve with Form

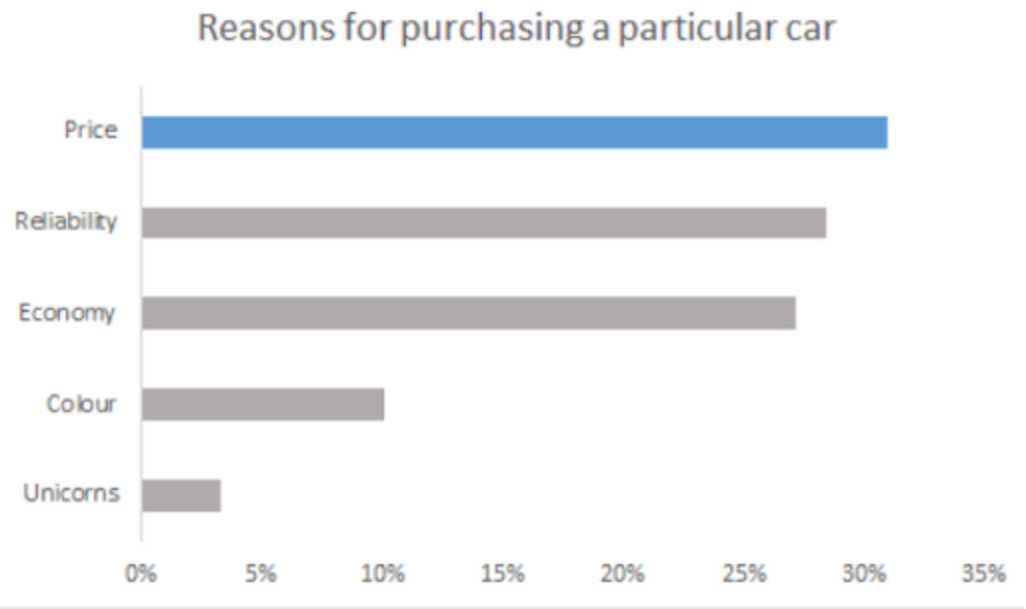


now really obvious the Price has the longest bar and is the top feature.

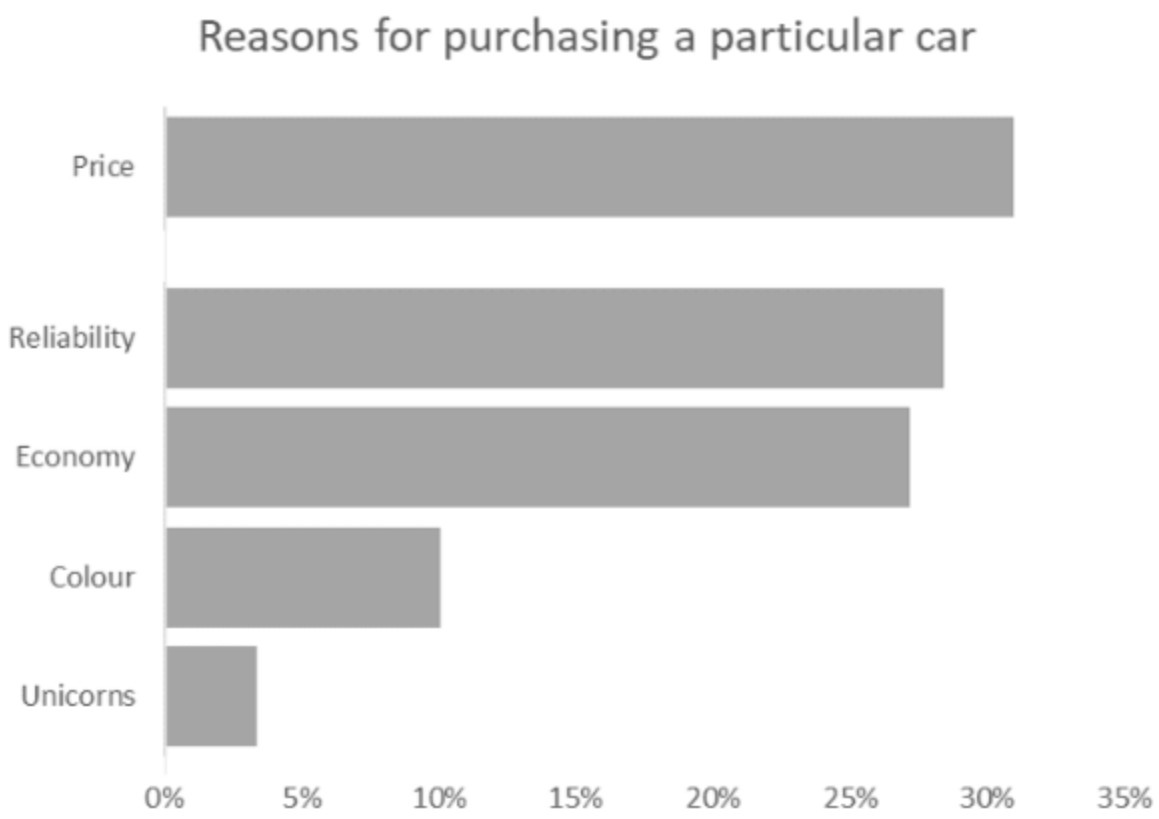
improve with Colour



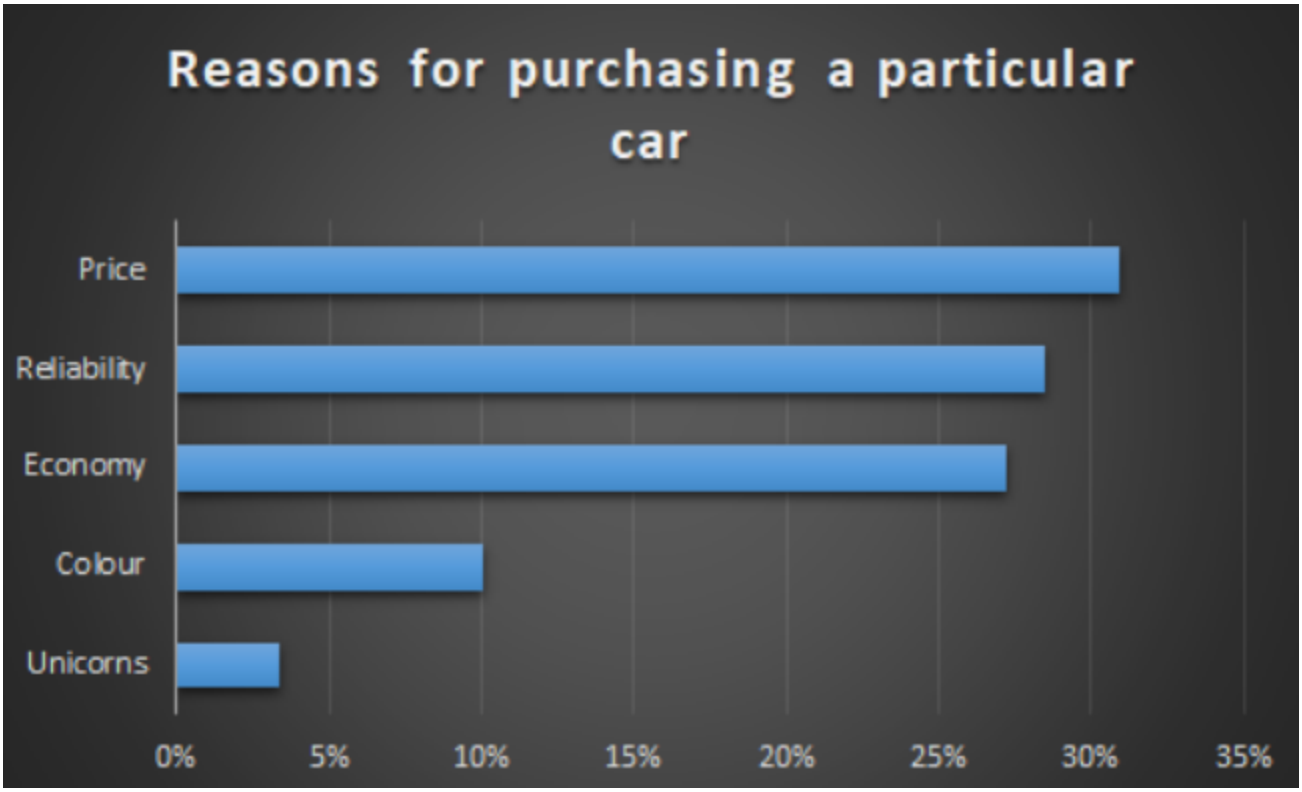
Certainly more colourful but the preattentive feature is **overwhelmed**. Let's try highlighting with a single "action" colour the story that most people consider Price.



imrove with Spatial Positioning



Spatial position highlighting is better seen by looking at the white space (negative space) used to surround the graph. Below is an example where an unnecessarily dark background is distracting.



> The target, the mark or relationship that you want to highlight as part of your message, must stand out in a simple dimension like colour, orientation, size, position or through motion if that's available. Overusing these attributes can lead to confusion and distraction so use wisely.

Gestalt Theory

> The Gestalt Theory of Visual Perception (Max Westheimer, Kurt Koffka and Wolfgang Kohler (1912)) is based on the concept that we view scenes as a combination of the parts and create order from the inputs. The parts that make up the image of a tree (grass, branches, trunk, leaves) are less noticeable than the concept of “tree”.

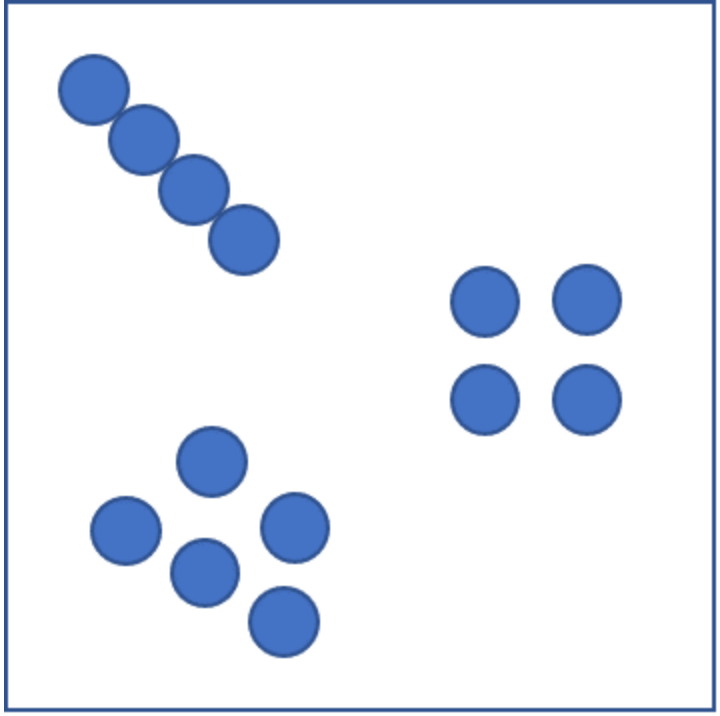
Figure is the foreground object and **Ground** is the surroundings (or background).

The Six Principles

https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600234

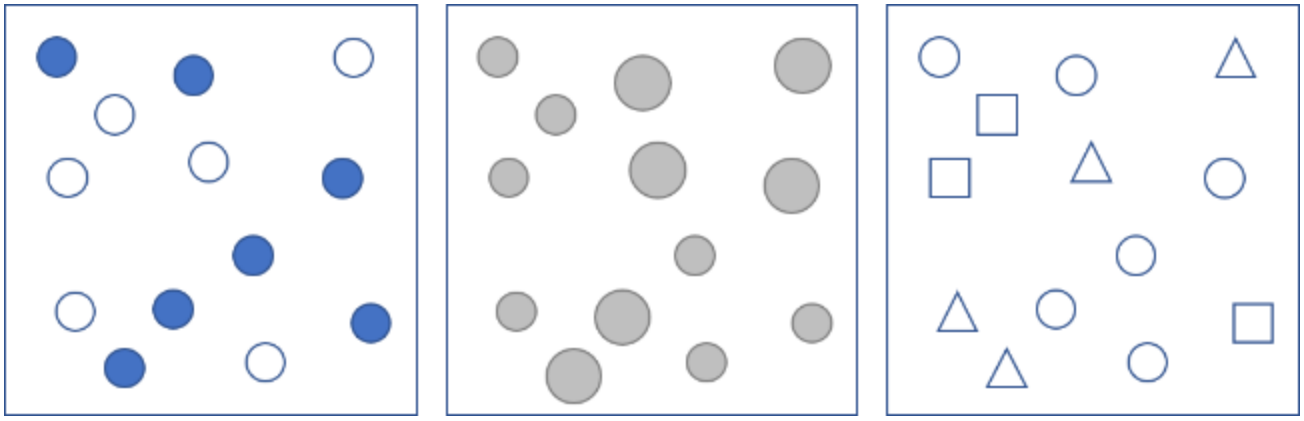
1) Proximity

Elements placed together are seen as a whole



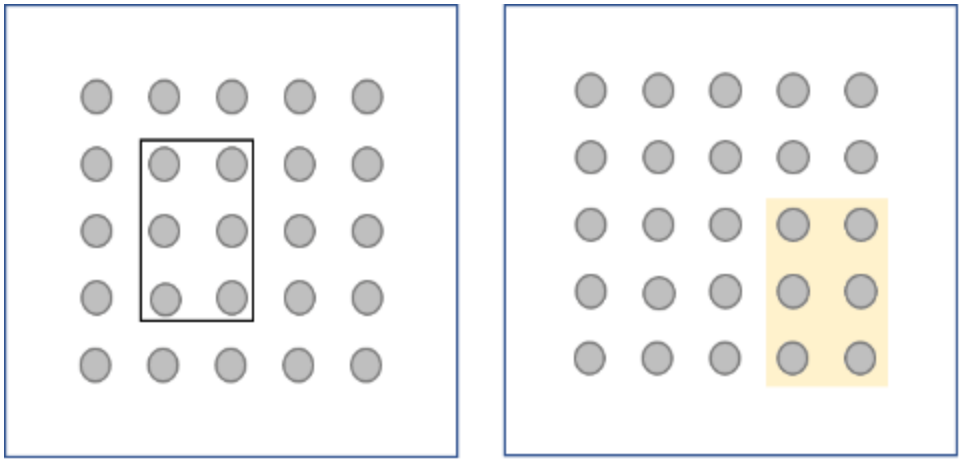
2) Similarity

Elements that share similar attributes (e.g., colour, size, shape, orientation) are seen as a whole



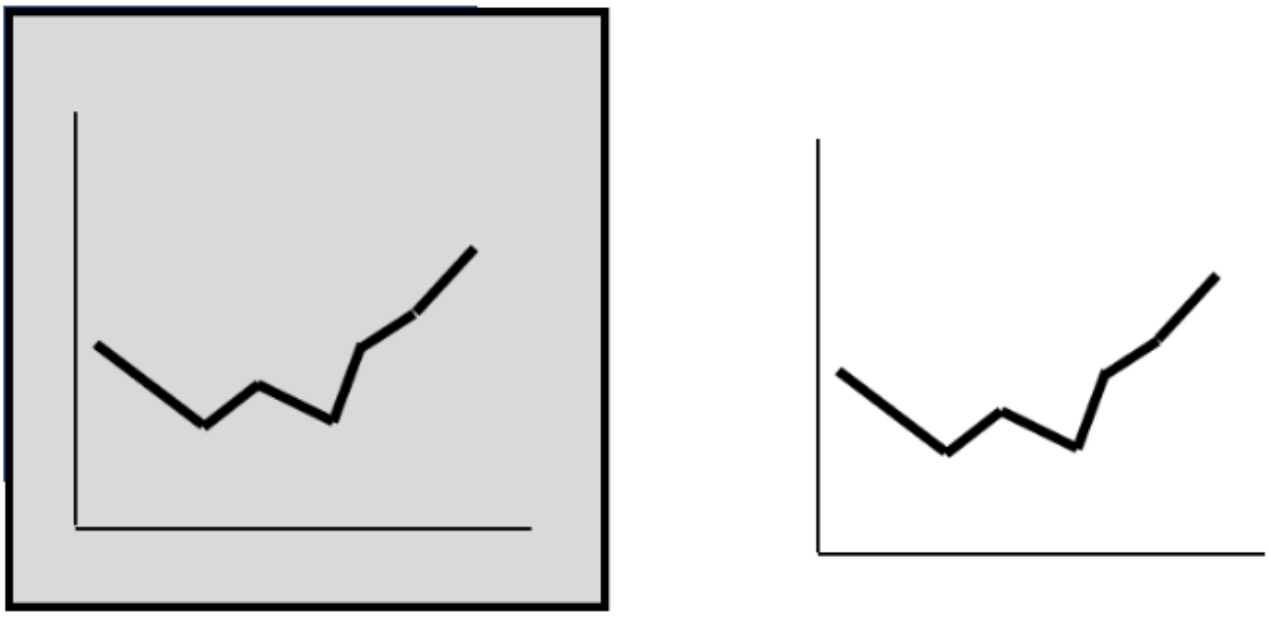
3) Enclosure

Objects that appear to have a boundary around them are perceived as a group

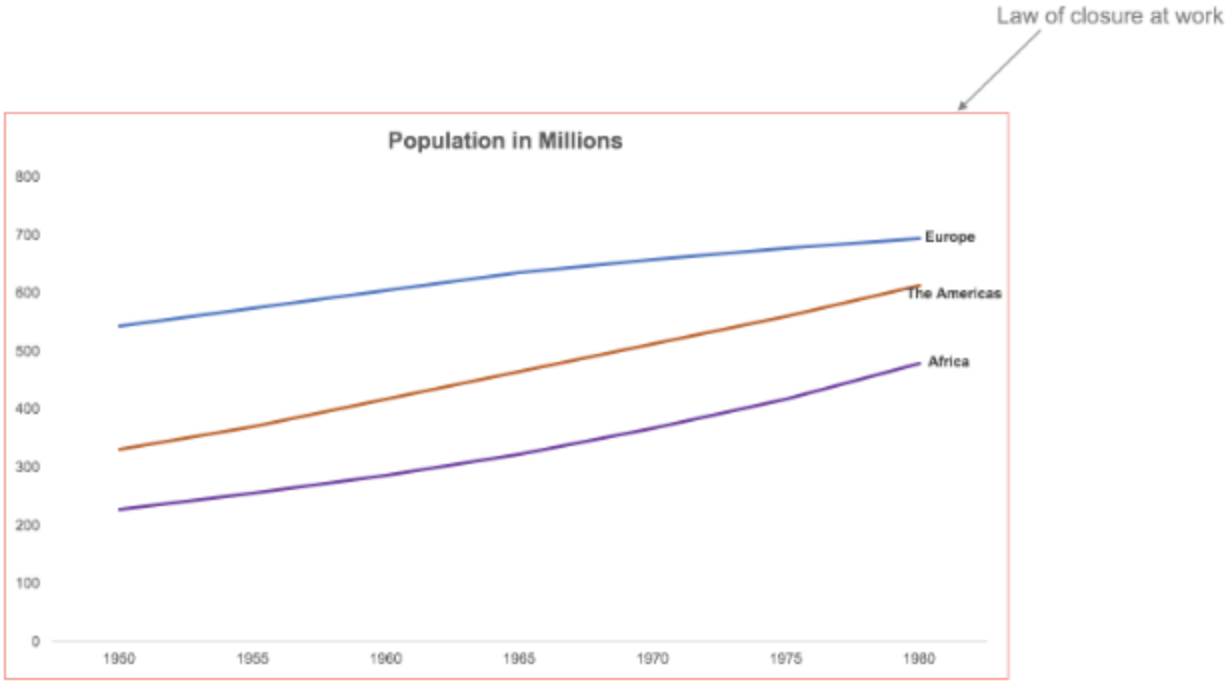


4) Closure

Open structures are perceived as closed and complete wherever possible. Note that people are very good at "joining the dots"!



> The image above shows how even without the explicit background and border enclose we are still able to imply the enclosure of the graph as a grouped object. While including the background and border may be a default option in many graphing applications, removing this lets us reduce clutter in our graphs and improve readability.

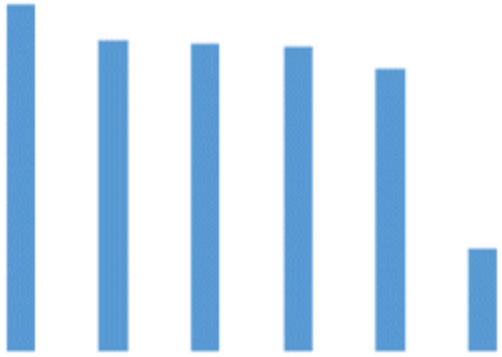


5) Continuity

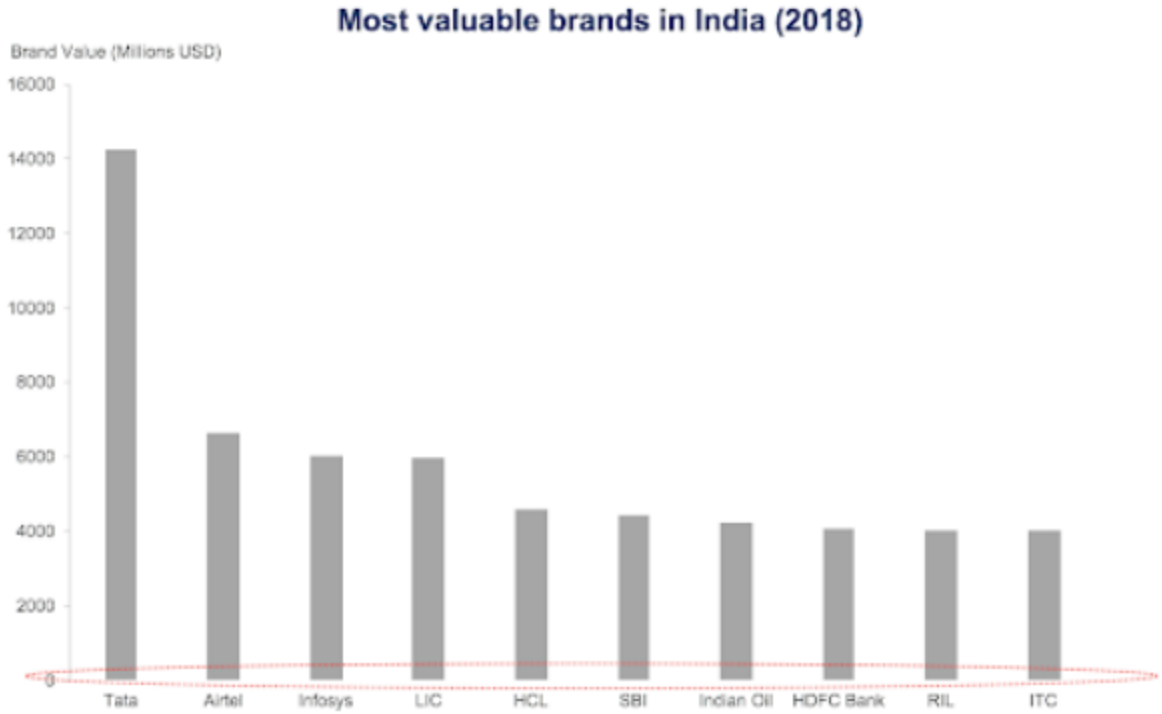
Elements that appear to be sequential and/or forming continuous lines are seen as a grouping.

We tend to imply this continuity to the strongest degree possible, grouping objects into the fewest options.

Similarly to **removing borders and backgrounds by exploiting the principle of closure**, we can use this to remove the explicit axis line and the viewer will fill it in as shown below.

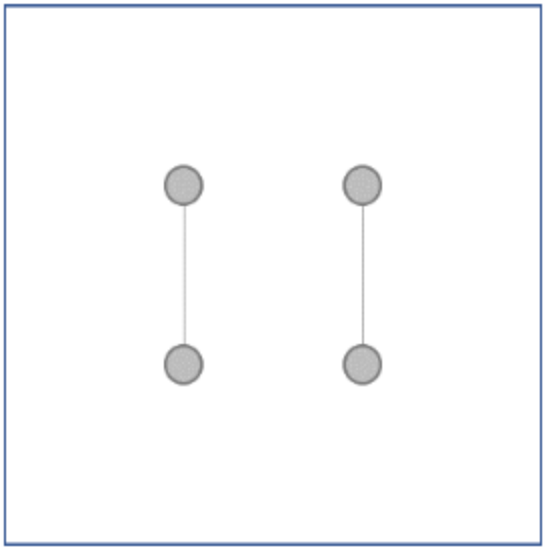


eg:

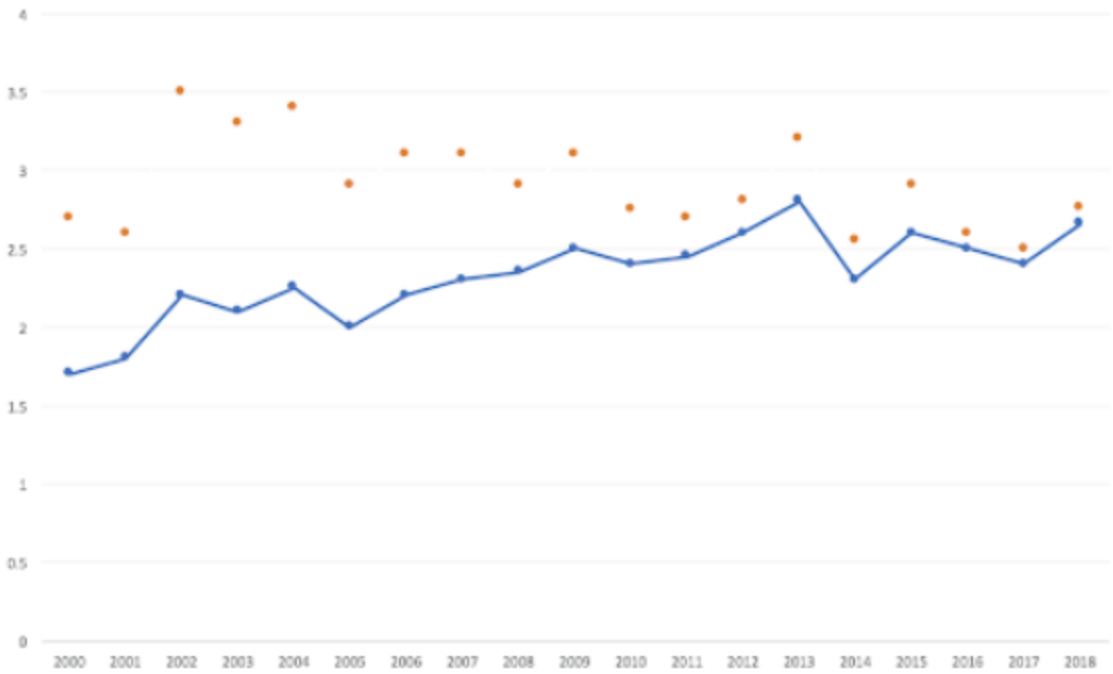


6) Connection

Objects that are connected (e.g., by a line) are perceived as a group. Similar to enclosure we can draw an explicit link between objects using lines to connect them.



eg:



wk2 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600230/quiz/introduction

---

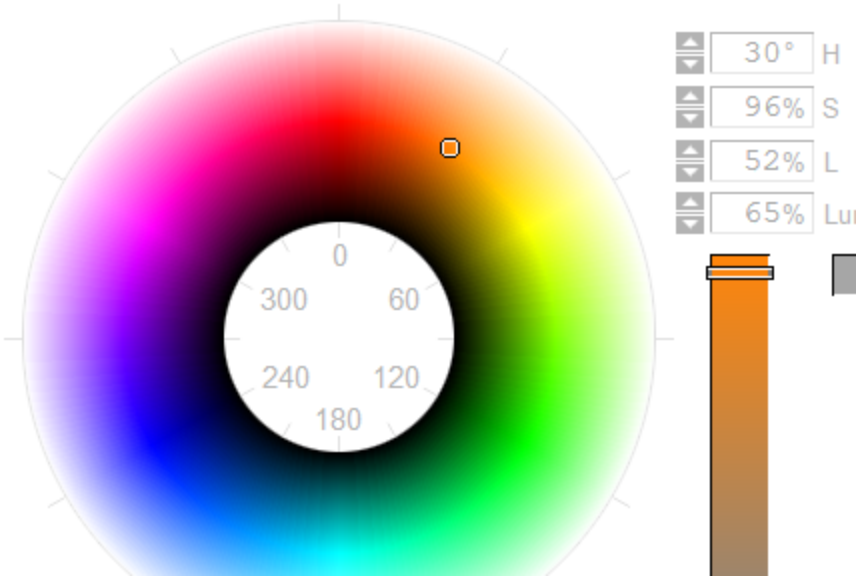
Week 3: Design

Practical Design Tips:

1) Colour

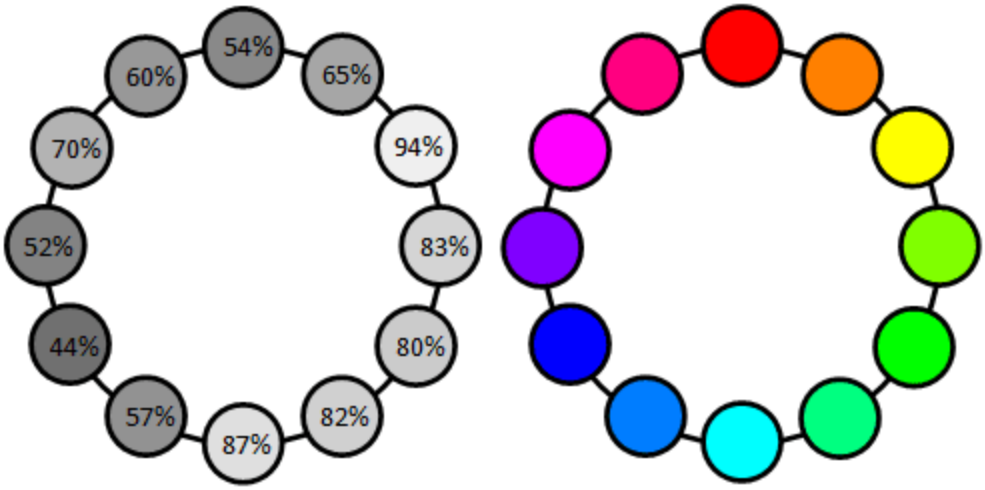
- context matters (percieve candle indoor vs outdoor differently)
- cultural meaning of certain colors

- A **HSL** colour picker will let you see how each of these values changes the colour appearance. Hue, Saturation and Luminance are useful for understanding colour choices and the relationships between colours.



1. Hue: different colours as measured using a colour wheel 0-359°. For example, 0° is Red, 120° is Green and 240° is Blue.
2. Saturation: the vividness or "purity" of a colour, between 100% (pure colour) to 0% (grey). Also called Intensity.
3. Luminance: visually perceived brightness or lightness, between 0% and 100% where white is 100%. Also called Lightness.

The figure below shows the Luminance value for a set of colours (Hue)



The main principle is to use colour sparingly and thoughtfully.

Remember that colour (Hue, Saturation and Luminance) is a **data encoding** attribute. Any use of colour on your graph can represent values from your data.

Colour is also strongly associated with attention and Hue and Saturation (Intensity) are both preattentive attributes.

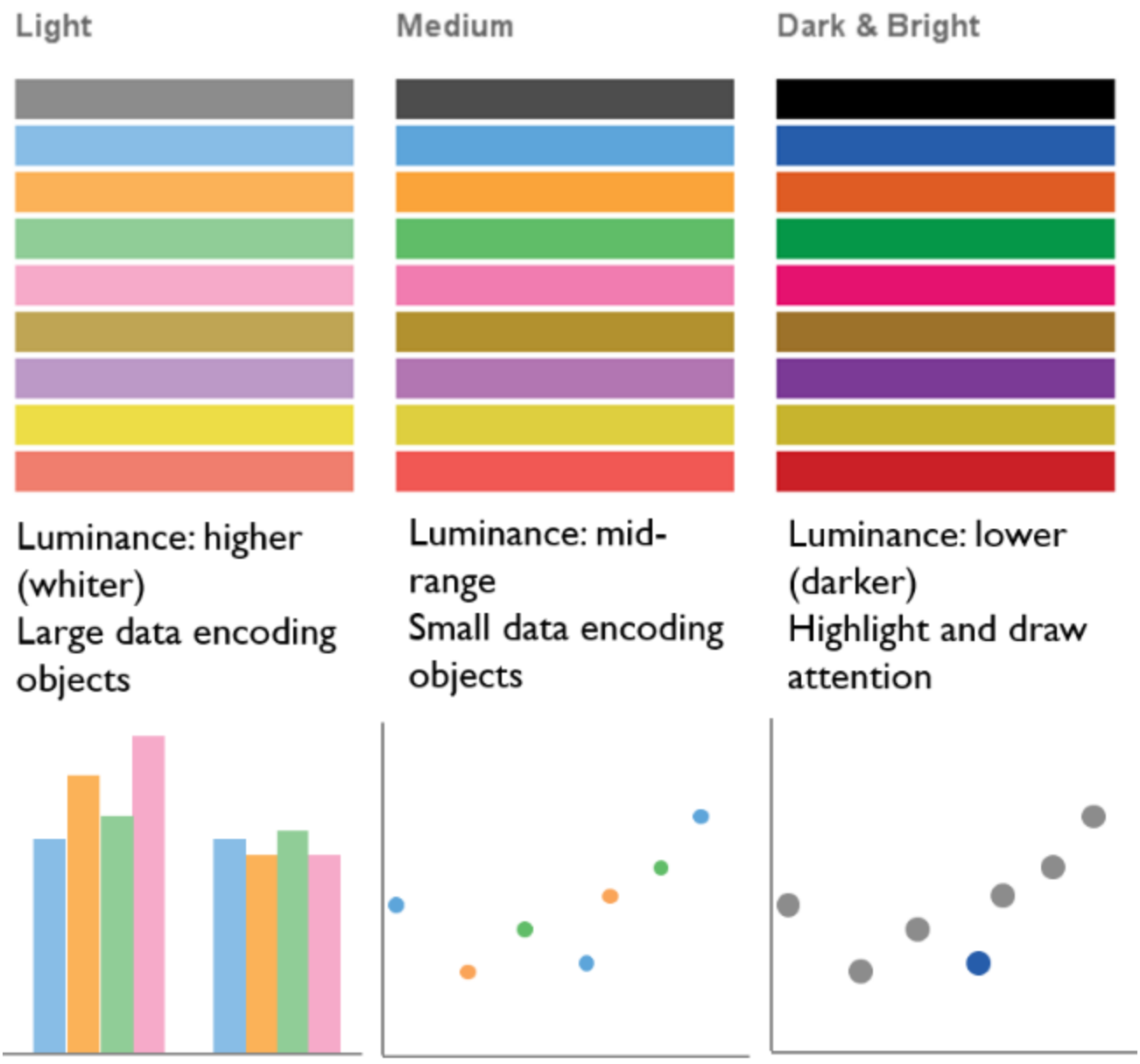
So if you overuse colour you are diluting your message and overwhelming the attention channels of your audience.

Don’t use multiple colours for the sake of being colourful (better to have one hue, varied saturation (Knaflic, 2015))

Reserve the use of bright (highly saturated) colours for items that need attention or those that should be examined first, and use less intense (less saturated) colours for other items on the page. Overuse of hue and saturation will be distracting.

Few 2004

Few (2004) condenses these principles down to use cases for three palettes; Light, Medium and Dark & Bright:



SUMMARY ON COLOUR

Summary of general principles:

- Use colour sparingly and intentionally.
- Vary Luminance to highlight details.
- Vary Hue to indicate different categories.
- Low saturation for large areas of colour (e.g., backgrounds).
- High saturation for small areas of colour (e.g., marks, labels).
- Be aware of potential colour blindness issues and how your graph may be viewed in black and white.
- Be aware of the difference in contrast between foreground (graph marks or labels) and background colours. Large differences in contrast values are easier to read. The figure below shows a high contrast (Left) and low contrast (Right) example.

2) Decluttering

Knaflic (2015) defines clutter as “visual elements that take up space but don’t increase understanding” .

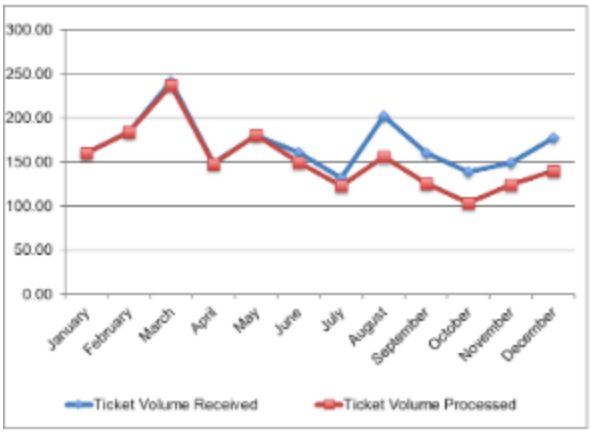
Any elements that don’t contribute to our understanding are adding unnecessary cognitive load on your audience

Knaflic has some good examples of removing clutter from graphs. The figure below has been decluttered by following six steps:

decluttered by following six steps

1. **Remove chart border**: use whitespace effectively instead.
2. **Remove gridlines**: unless your audience needs to look up specific values from your graph
3. **Remove unnecessary data markers**: or at least use them with intent
4. **Clean up axis labels**: you don't need trailing zeros
5. **Label data directly**: why make your audience look up a legend when you can label the line
6. **Leverage consistent colour**: reinforce the relationship between line and label (recall Gestalt principles!).

BEFORE/AFTER



> "The next time you are trying to improve a chart, consider what you can take away rather than what you can add." (Data looks better naked — Darkhorse Analytics | Edmonton, AB)

Using design rules

https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600243

Five Simple Tips to Remember

https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600244

1. Avoid pie charts.

2. Don't use 3D effects in a 2D medium.

3. Be careful about Axes.

4. Watch out for distorted Area.

5. Distraction and Clutter – “Remove to Improve”.

A bonus three tips to keep in mind as motivation for all your data-driven visualisation efforts.

1. Know why you’re visualising – to understand or communicate?
2. Think of being correct & clear then be compelling (beautiful).
3. Be kind to your viewers.

Week 4: Hall of Shame

We’ve now worked our way through multiple layers of aims for good visualisations.

1. High level aims: communication participants and goals
2. Practical and technical constraints: how we see
3. Matching data to graphs: appropriate chart choices and the CHRTS categories
4. Low-level design details: using colour and shape to direct attention and simplifying to reduce distracting clutter

COVID-19 Flatten the vurve

1. Understand the context.
2. Choose an appropriate display.
3. Eliminate clutter.
4. Draw attention to where you want your audience to focus.
5. Think like a designer.
6. Tell a story.

Kirk (2016) defines a visualisation workflow (chapter 2) that has the following stages:

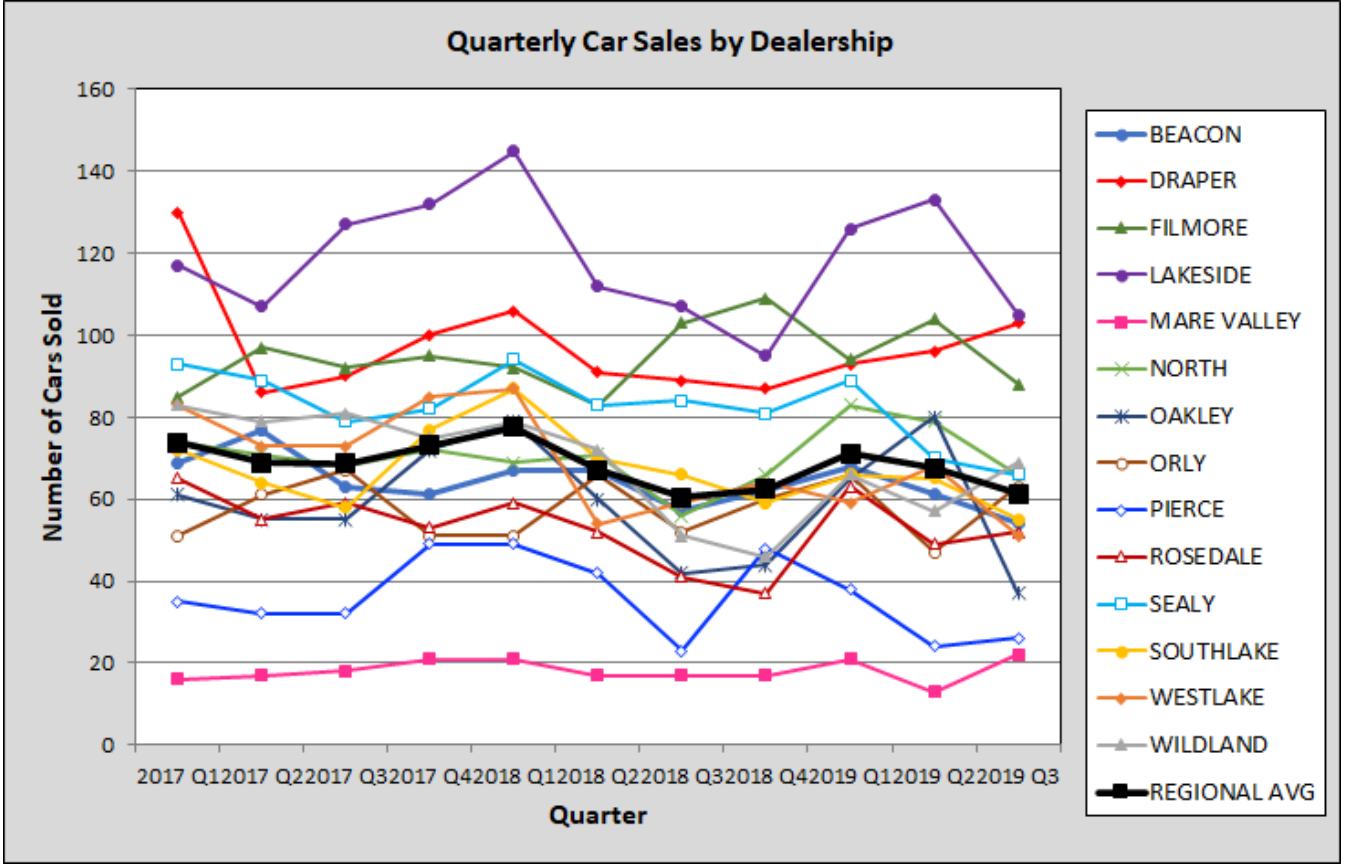
1. Formulating your brief
2. Working with data
3. Establishing your editorial thinking
4. Developing your design solution

Bringing these ideas together, I suggest the following set of steps or questions for you to use in developing your own data-driven visualisations:

1. What is your data? What is your message or story? Why are you visualising?
2. Choose an appropriate chart.
3. Sketch what it might look like. Identify aspects you want to draw attention to.
4. Make your graph.
5. What elements of the graph could be removed? (Data-to-pixel ratio).
6. What attentive elements are you using? Is the attention drawn to your message?
7. Consider design rules or branding requirements.

Critiquing Visiualisations:

Task: practise critiquing a visualisation 0 comments



Recall quote from Kirk (2012: 20) about no single right answer:

> "... data visualization is not an exact science. There is rarely, if ever, a single right answer or single best solution. It is much more about using heuristic methods to determine the most satisfactory solutions."

week 4 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-communicating-with-data/3/steps/1600248/quiz/introduction