wk1: visiual comm

In Data Fundamentals in Step 1.14 the concepts of **exploratory and explanatory** visualisation were introduced. We know from the generic pipeline that there are different reasons to create a data-driven visualisation.

Visualisations can:

- help understand the data
- clean data (show missing vals/poutliers)
- be static, dynamic or interactive
- comm conclusions
- convience your audience

graph vs chart:

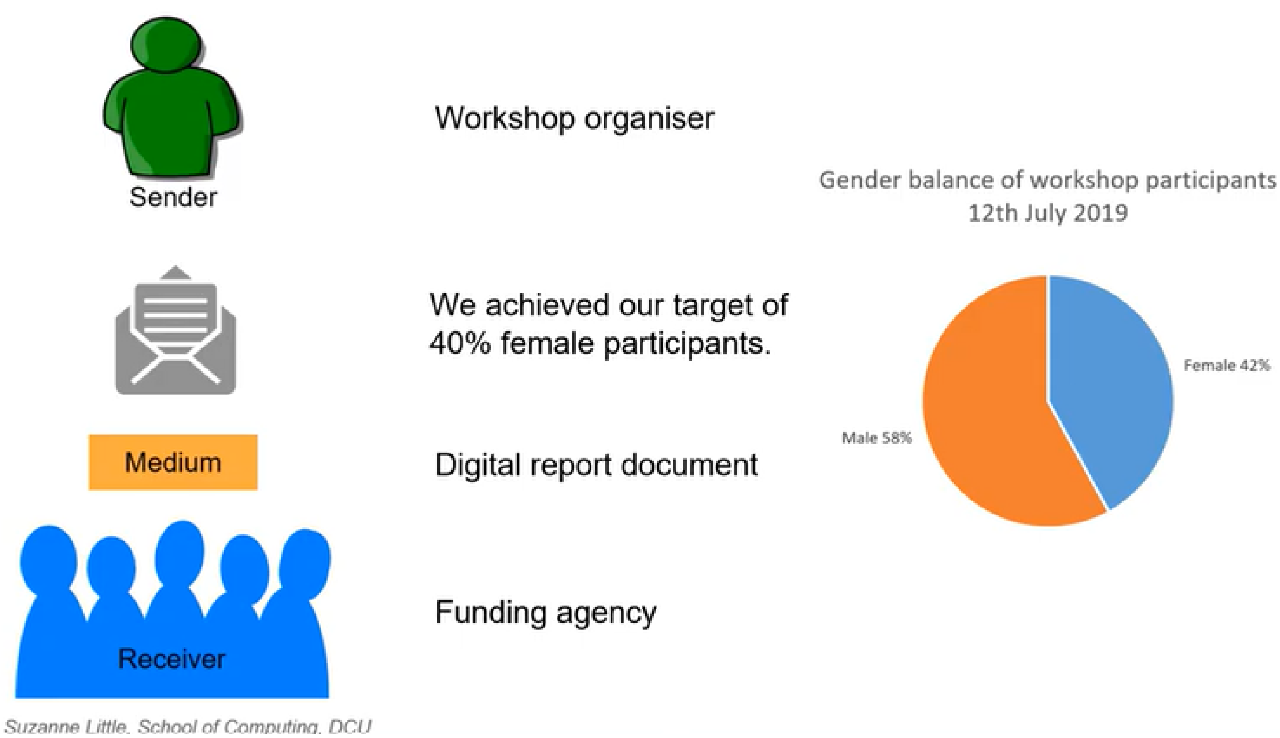
graph is a “specific type of chart, showing the relationships between mathematical data”

participants

sender (client/designer) and second role is reciever (audience, single or multiple - may know lots or little)

third participant is the info/pursuassion and all brought together usng medium (how message is transmitted)

example:



example 2: red candy message, what if we change sender from candy marketer and reciever is consumers - medium can shift based on these participants changing!

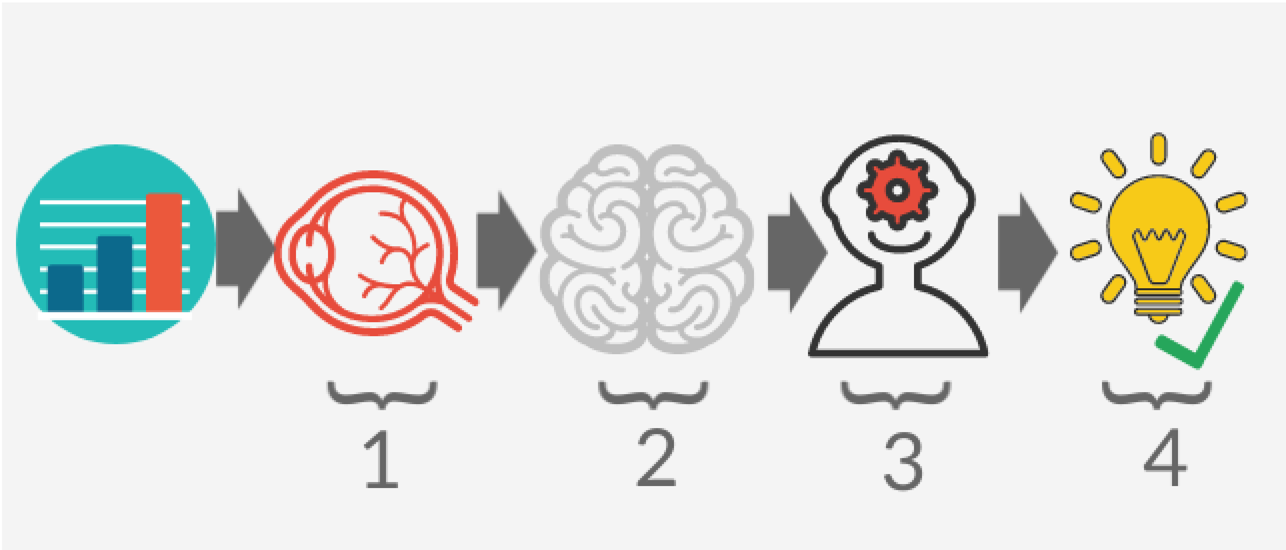
sometimes the medium IS THE MESSAGE

- changing medium can affect trust, reach of comm.

- motivation if sender and reciever affects medium choice (visual comm)

Stages of Understanding

What happens when you view a graphic communication?

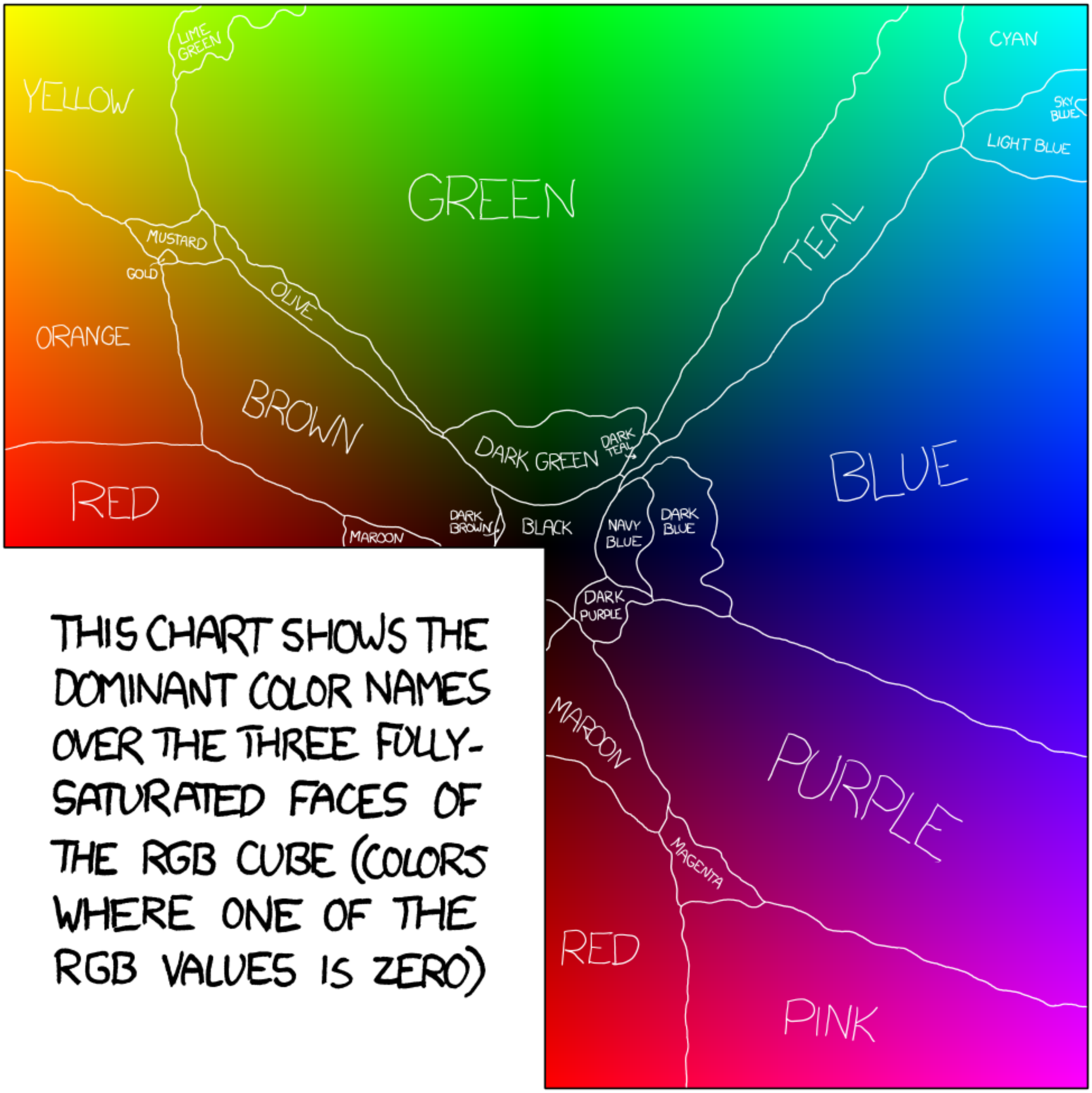


- Sensing: eye and brain see colours and shapes
- Perceiving: assign meaning to the inputs
- Interpreting: highly subjective and influenced by many factors
- Comprehending: what does it mean for me? relevance/consequence

humans could recognise images in as little as 13 milliseconds (Dimarco, 2010).

colour:

- RGB colour space or a Hex code.
- peiople perceptions of colour not uniform:



Python data visualisation libraries

the most important skill is that you understand the visualisation design process, that is, how to:

- make effective visualisations,
- choose the most appropriate chart types,
- tell compelling stories,
- avoid cognitive load on your audience.

Matplotlib

Seaborn

- colab example: https://github.com/suzannelittle/ca682i/blob/master/notebooks/seaborn.ipynb

SCENARIO:

Your mission: create a graph showing the most popular chocolate bars for your boss to use in a presentation about how to stock the vending machines.

Get the notebook outline with the data https://github.com/suzannelittle/ca682i/blob/master/notebooks/3_1_11_Data_Visualisation_with_Python.ipynb. You can open this directly in Google Colab from Github (using the button at the top - shown below) but will need to make a copy of the notebook to your own Google Drive account to be able to save your work.

wk1 quiz https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600185/quiz/introduction

You should now be able to:

Explain the participants in a visual communication. Identify the stages of understanding. Describe the key concepts and terms behind colour and digital images.

Week 2: Why We Communicate

communication goals

> Recall that communication is a process that requires a sender (the designer, author or client), a message (information or an effort to persuade), a medium (the delivery platform) and a receiver of that message (the audience).

There are **four main motivations for communication**:

1. information
2. persuasion
3. education
4. entertainment

https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600188

Priorities: High Comprehension, High Retention, Moderate Appeal (depends on audience motivation)

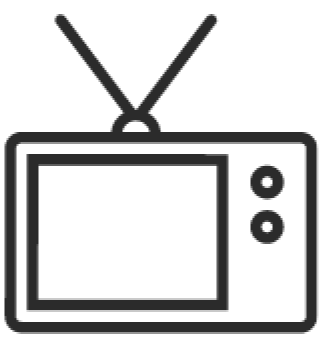
data source is explicitly or formally referenced.

Persuasion

Priorities: High Appeal, Moderate Attention, Low Comprehension

Education

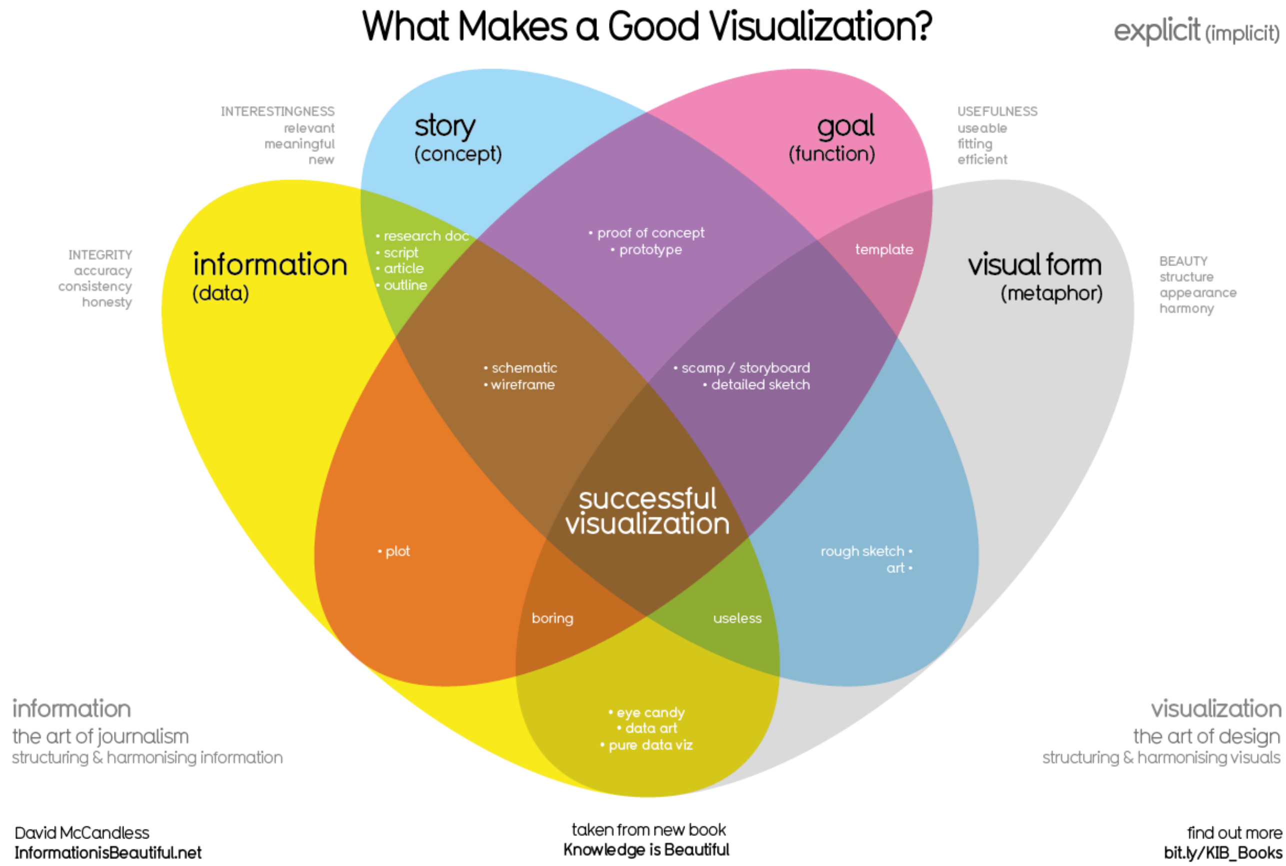
Priorities: High Comprehension, High Retention, Low Appeal. This assumes a mostly captive audience! But unappealing communications tend to have negative consequences for comprehension and retention.



Entertainment

Priorities: High Appeal, Moderate Comprehension, Low Retention (though note that good stories are an excellent aid to memory!)

what makes a good visiualisation?



(David McCandless, 2014). He defines these as the intersection of information (data), story (concept), goal (function) and visual form (metaphor)

https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600190

We will return to this again in the next Topics when we look more at:

1. How humans see and therefore what impact design choices can have,
2. How to encode data (numbers) into visual languages and
3. How to gain and lose attention.

Kirk (2016)

According to the principles of Good Data Visualisations are:

1. **Trustworthy**:
    - A) Don’t use inappropriate colour palettes or fonts.
    - B) Don’t include unnecessary chart junk (decoration)
2. **Accessible**:
    - A) Useful and understandable.
    - B) Reward vs Effort (complexity is sometimes okay!).
3. **Elegant**:
    - A) Thorough (get this little details right).
    - B) Stylish.

> This is why it can be tricky to create a definitive and absolute list of what makes a "good" visualisation. Is it the simplicity of Tufte/Few or the data-driven focus of Cairo or the storytelling of Knaflic?

Why you should never trust a data visualisation

https://www.theguardian.com/news/datablog/2013/jul/24/why-you-should-never-trust-a-data-visualisation#:~:text=Data%20scientists%20are%20typically%20attached,out%20analyses%20of%20their%20owı

responses about above from Andy Kirk and Andy Cosgreave:
https://www.visualisingdata.com/2013/07/should-you-trust-data-visualisations/

SCENARIO

exam?

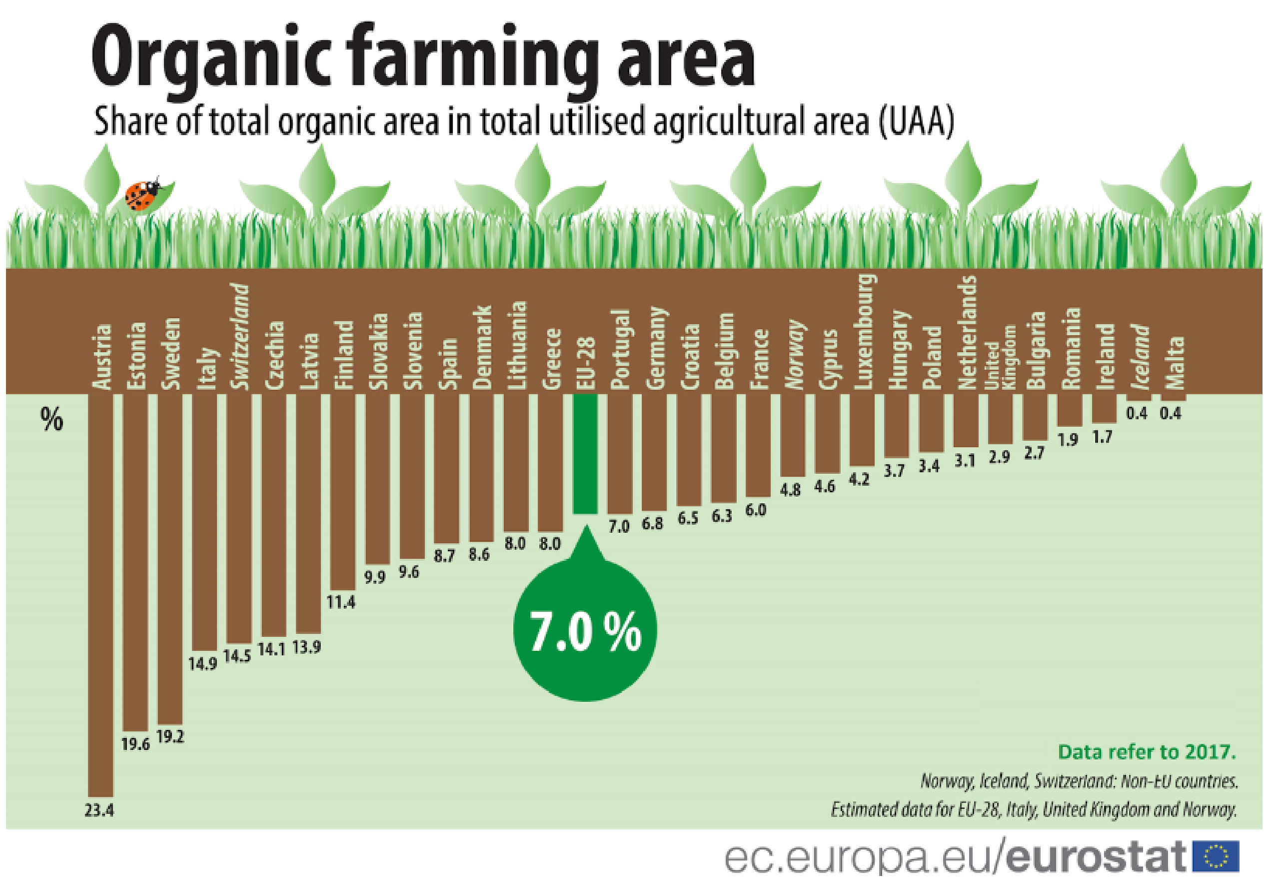
Discuss: How can good visualisation design concepts be applied to avoid misleading and "bad" data visualisations? Consider how much time is needed to fully understand (clean, explore and analyse) a dataset to create an effective, simple visualisation to understand and interpret that visualisation. Where and how can misunderstandings occur?

---

wk2 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600194/quiz/introduction

communication goal examples:

1)



> This is primarily **information** due to its *use of data* and the *upside down bar graph*. You could also say it's partially **persuasion** if the *context* was to encourage an increase (or decrease) in organic farming in the EU. It is **educational** but *lacks the hierarchical structure* that a focussed educational communication would have.

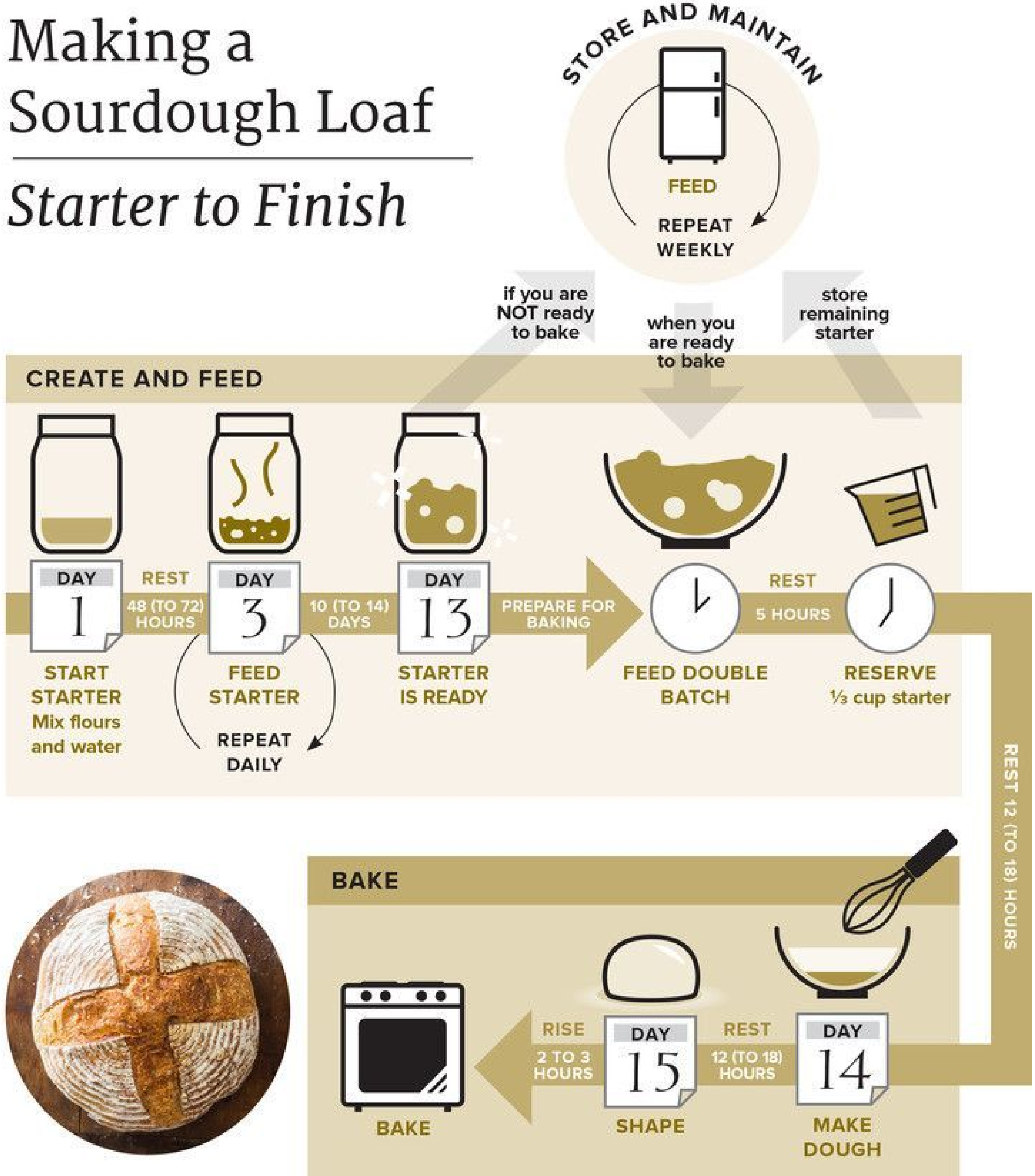
2)

> This is an *advertisement* though it's made to look similar to a public service announcement (Information) with the message about putting your phones away to socialise so the main purpose is persuasion (but Guinness). This is despite the visual not explicitly saying so. The graphic in the middle is recognisably a pint of Guinness even though we recognise the individual components as phones. We'll learn more about why our brains turn this into a single object in the next course.

3) juerassic park movie trailer

> a movie trailer is a visual construction that is entertaining in itself (Entertainment) but it's also intended as an advertisement to encourage you to see the movie so you could also class this a Persuasion. It focuses on the narrative (what's going to happen) and shows you briefly some of the characters. You also get a distinct feel of a Spielberg movie (and he's mentioned by name).

4)



> This visual is **informative and also educational**. While it doesn't have data, it does come from a fairly reliable and *authoritative source* (America's Test Kitchen). There's clear breakdown into a *hierarchy of steps* and use of arrows and visual grouping to make the process easier to understand.

Now that you have completed Topic 2 you should now be able to:

1. Compare the different aims for visual communication
2. Evaluate a visualisation to identify possible communication goals
3. Identify principles of good visualisations
4. Debate the universality of visualisation principles

Week 3: How We See

https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600197

what happens to the signal in your brain

Parallel vs serial processing:

- Parallel processing occurs when objects are processed simultaneously
- Serial processing happens one item at a time
    - So for serial processes both attention (where you are looking) and memory are used
- bottom-up (stimulus-driven) and top-down (knowledge-driven)

> Sensory input might be objective but perception can be subjective so be aware that your audience will not view the message in the same way as the sender. Also, consider the extra cognitive load where serial processing is needed. If you have a complex message then don't use a medium where the audience is involved in another task. For example, displaying your complex graph on a billboard where your target audience may be concentrating on driving.

Perception and depth cues

There are two binocular cues that provide the most powerful inputs for depth perception: binocular disparity and convergence.

binocular disparity :

- images sense by our two eyes are slightly different and the brain can use this distance to determine depth
- Exploiting the binocular disparity is what gives us 3D movies

Convergence:

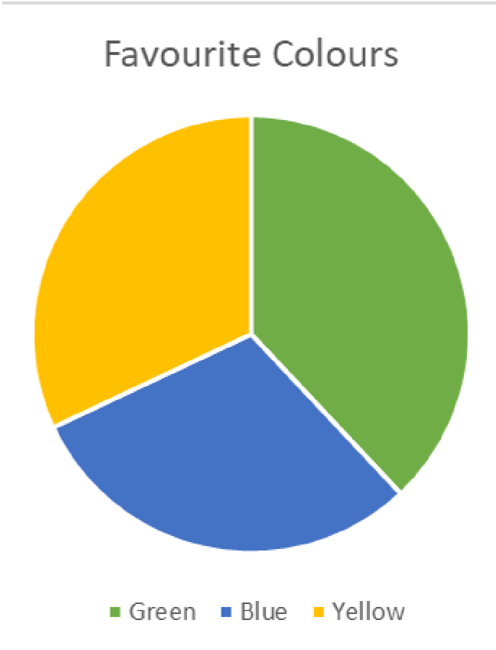
- weaker than disparity
- relates to the difference in the direction of our eyes when looking at an object closer versus one that is far away

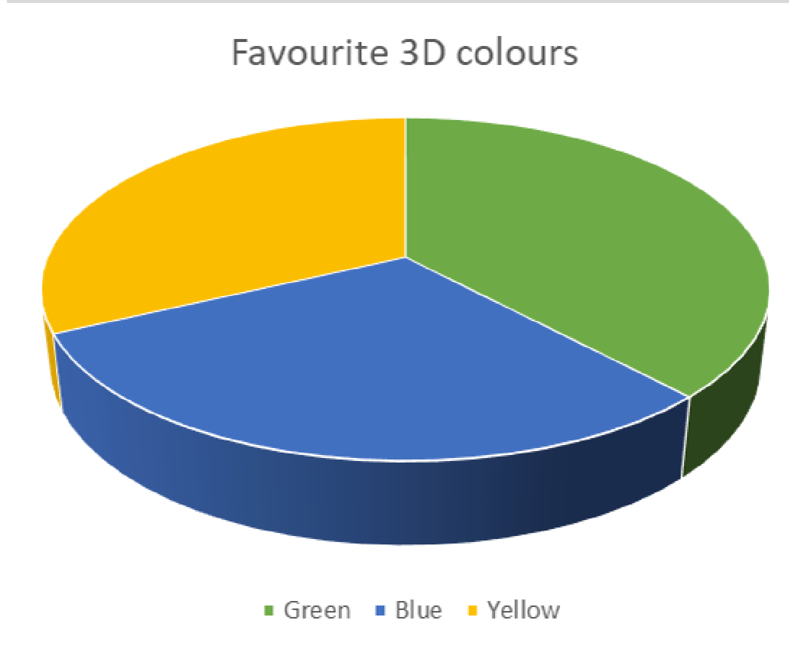
Using 3D effects in your graphs:

exam

If we add 3D effects to our 2D graphics then we are adding cognitive processing load for no benefit,

3D one really makes it look like blue is the most popular colour.:

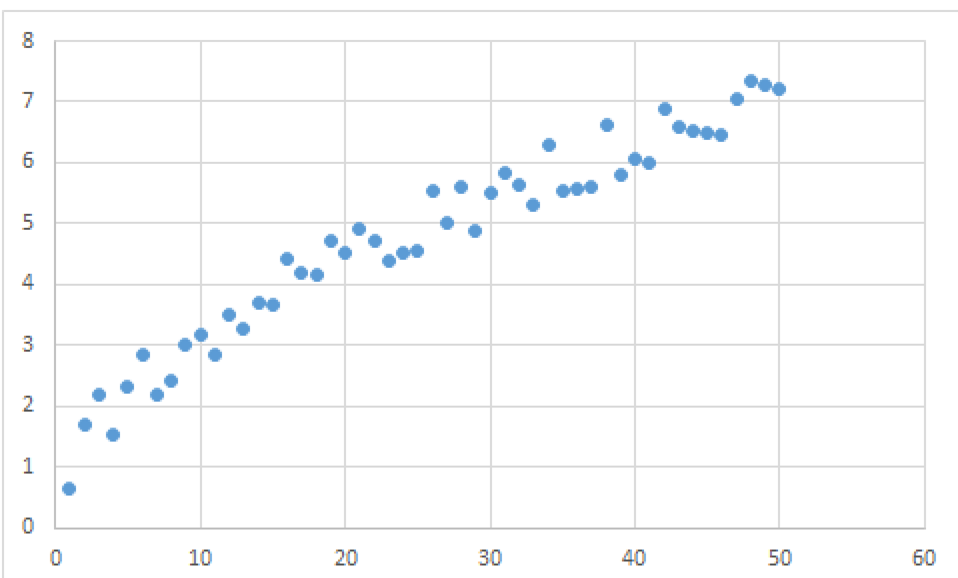




Understanding how we see helps visualisation designers to make choices so that they:

1. Make the most of visual cues.
2. Don't overload the visual information channel.
3. Don't dilute or distort your message.
4. Don't exclude or miscommunicate with your audience (**consider colour blindness**).
5. Simulate effective depth, motion and make judicious use of 3D effects!
6. Engage attention (we'll discuss this in Topic 2 in the fourth course in the program).

SCENARIO:



Discuss: exam?

Why would it be helpful to understand visual perception when creating a graph? Can you give an example using the simple graph above of how your brain interprets the visual signal? What are your eyes drawn to first?

> human visual system is a pattern seeker.....If we disobey these rules, our data will be incomprehensible or misleading Colin Ware (2012)

tools: D3.js (Data-driven Documents) is a Javascript (JS) library for creating data visualisations in the browser and it uses standard web technologies like HTML, DOM, CSS and SVG.

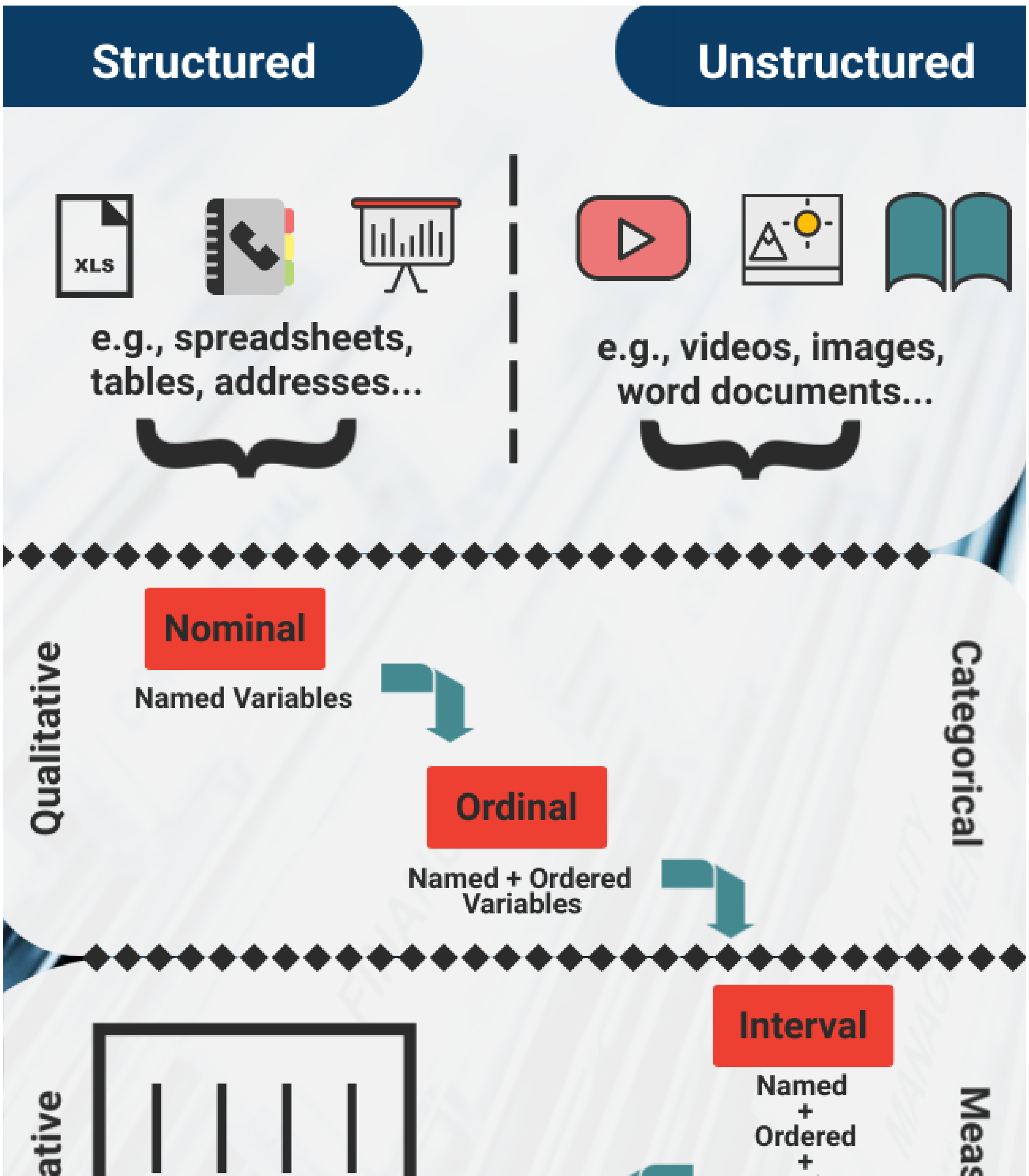
https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600200

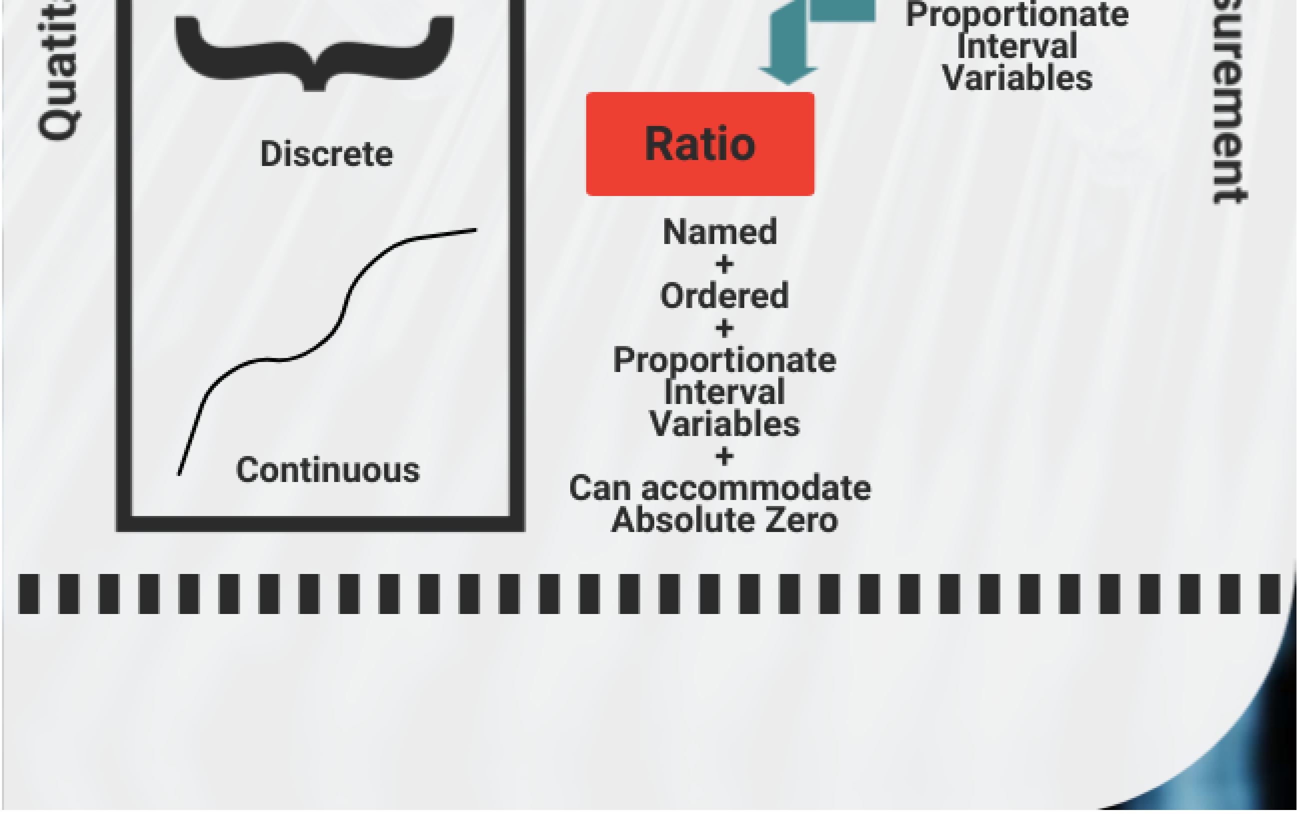
wk3 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600202/quiz/introduction

---

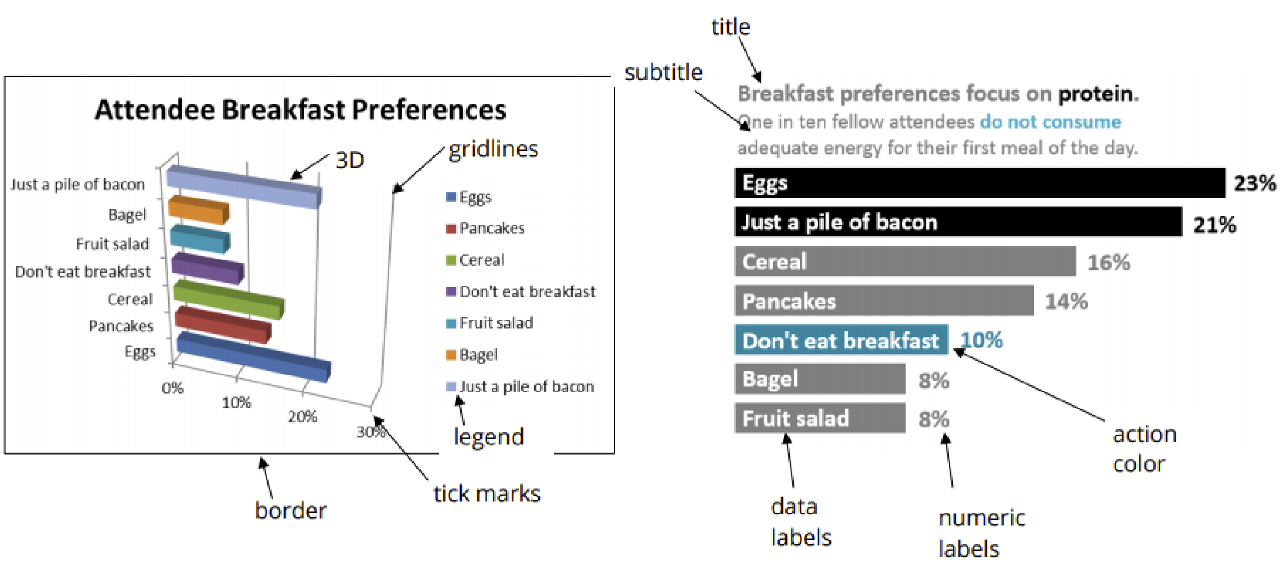
week 4: Representing Data Visually

recap:



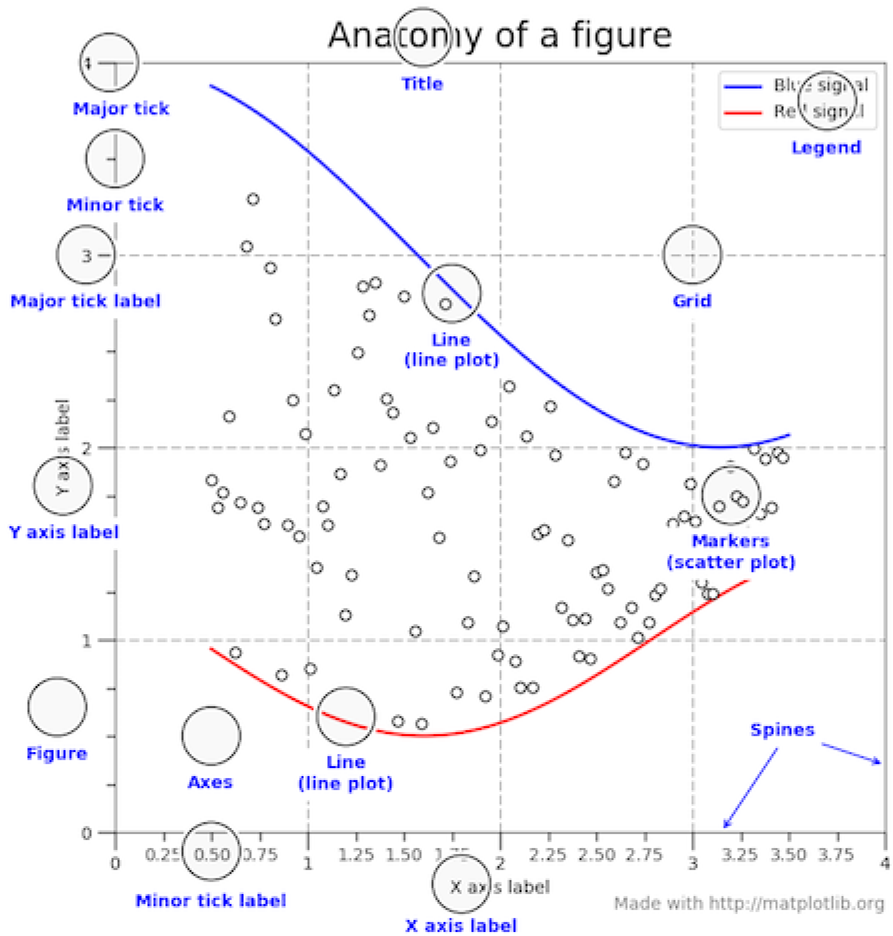


shows the same data graphed in two different ways (we'll talk later about which is a more effective visualisation!) and labels the main components.



3D, gridlines, border, tick marks, legend, title, subtitle, action color/colour, data labels, numeric labels? Notice that the x- and y- axis are not explicitly labelled here. Do they exist in the second graph?

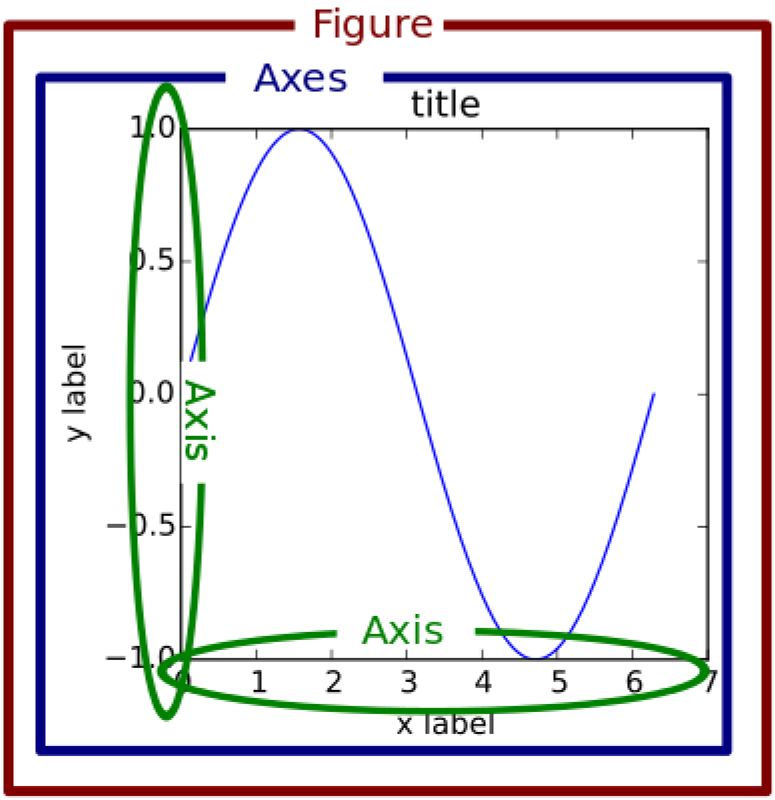
The image below labels a figure with the terms used in these libraries. This is especially useful when looking at the documentation to try and change a characteristic (colour, position, value, etc.) for one of these components. Most of these terms are the same for other libraries in python or R.



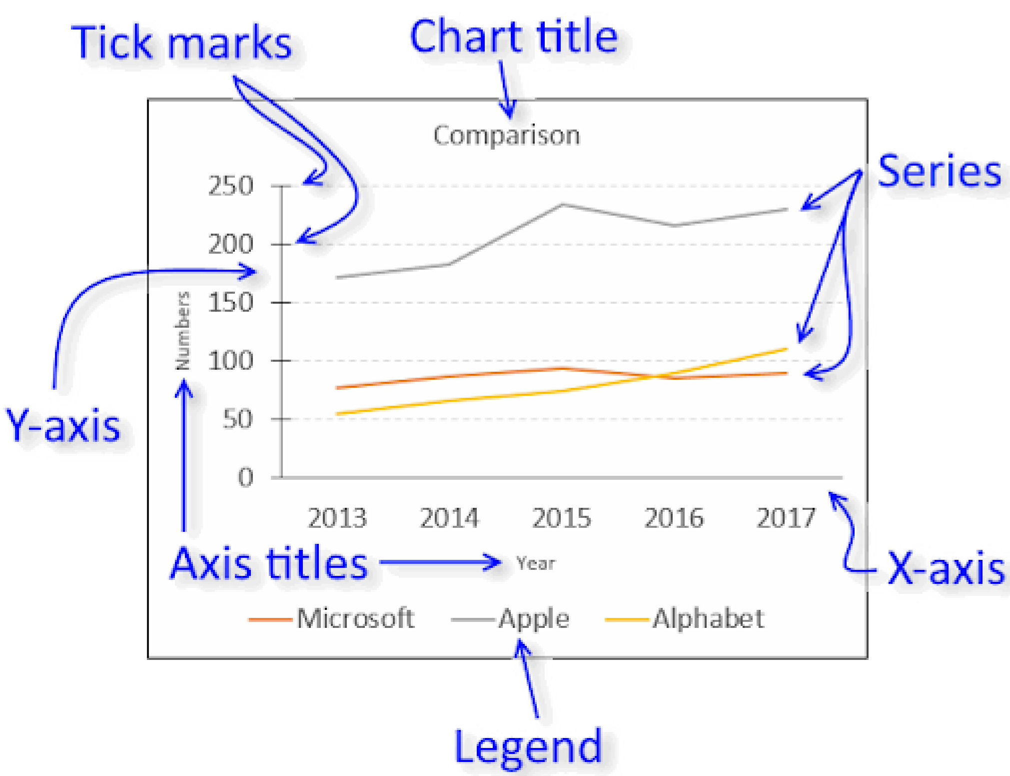
difference between an axes and an axis.

The ax**e**s is the area that your plot appears in, not the plural of ax**i**s (as commonly used in other documentation).

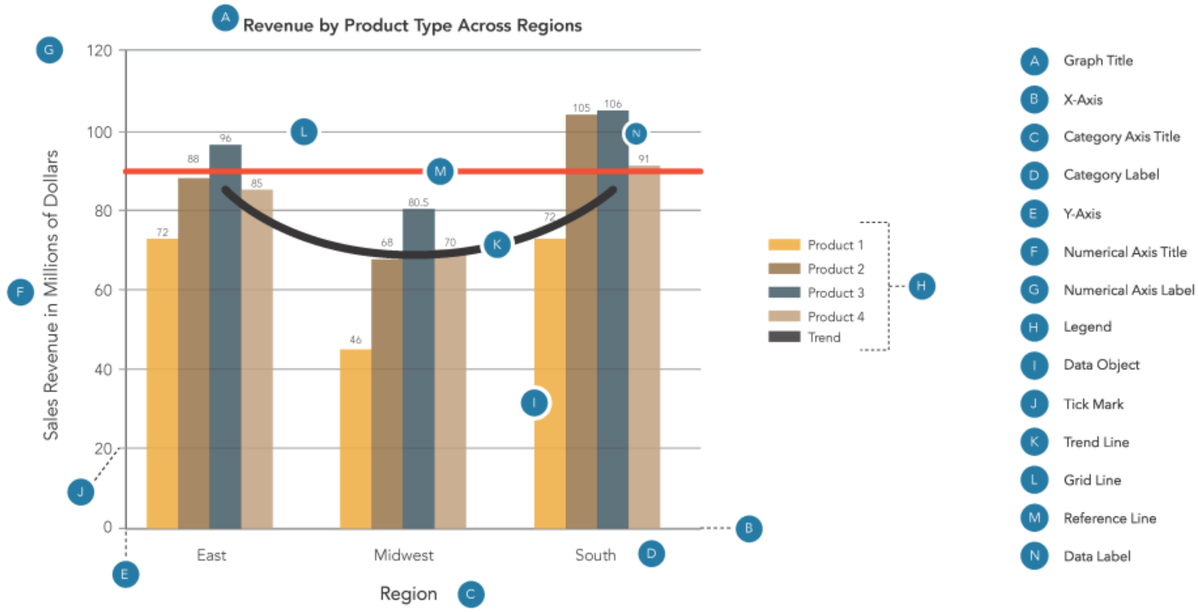
You can have multiple axes in a figure to get a multi-graph figure that may share the same axis/axes.



Excel chart components can be found here: http://www.mit.edu/~mbarker/formula1/f1help/10-ch-c2.htm



Finally, this article (https://chaione.com/blog/building-blocks-graphs/) provides a tool neutral discussion of graph components and some descriptions of each of the elements:



Encoding data

“The idea is to go from numbers to information to understanding.” (Hans Rosling)

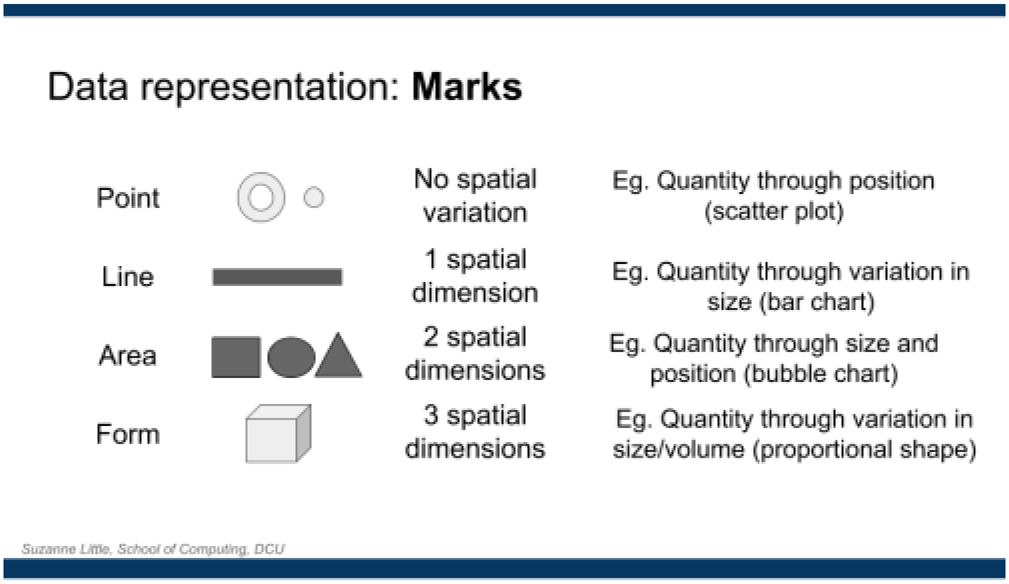
A **graph encodes data**. That is, it represents the information in a different way.

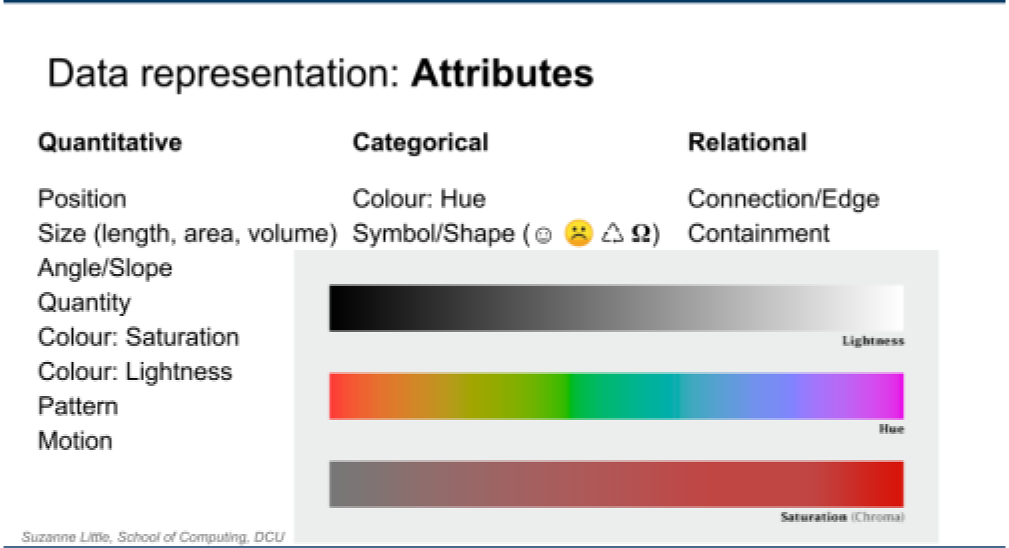
graphs use **marks and attributes** to represent info. Will look at the different types of *marks and their attributes* and match them to the data that they can be used to encode.

summary: Marks & Attributes

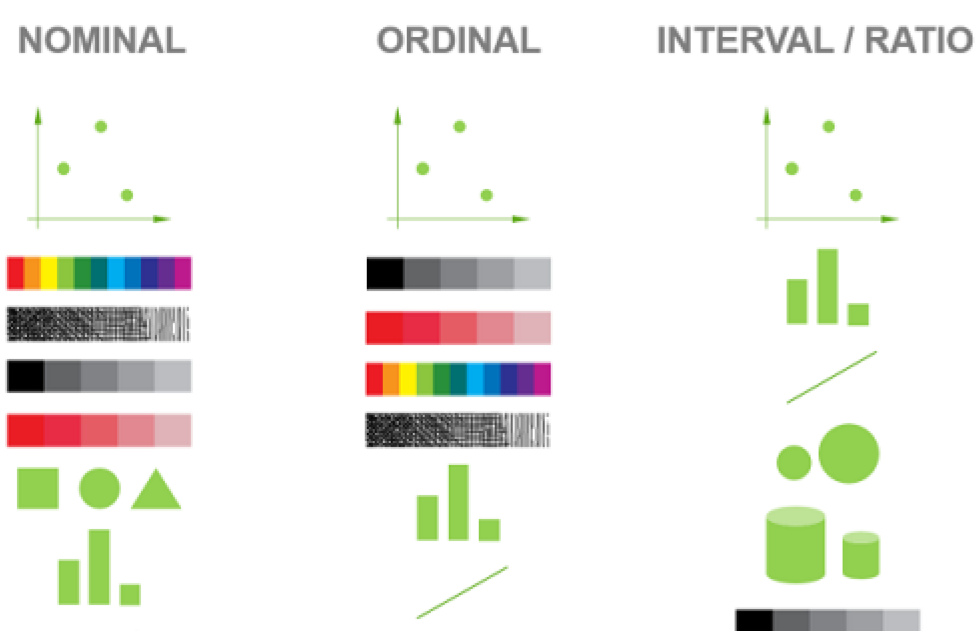
Marks and Attributes

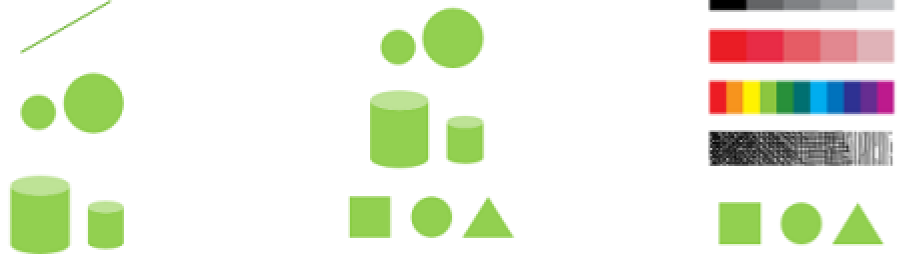
The figures below summarise the main points about marks and attributes and might be useful as a reference or reminder.





If you prefer a visual representation of the perceptual ranking, the graphic below was created by Patrik Lundblad.



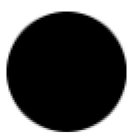


Marks

There are four variations of marks: point, line, area and form each of which can capture data by variations of different attributes.

Point

mostly used to represent quantitative data values either through position on a scale or the quantity of points used.



Line

commonly used to represent a quantitative value through variation in size or to show trends through variation in angle.

Area

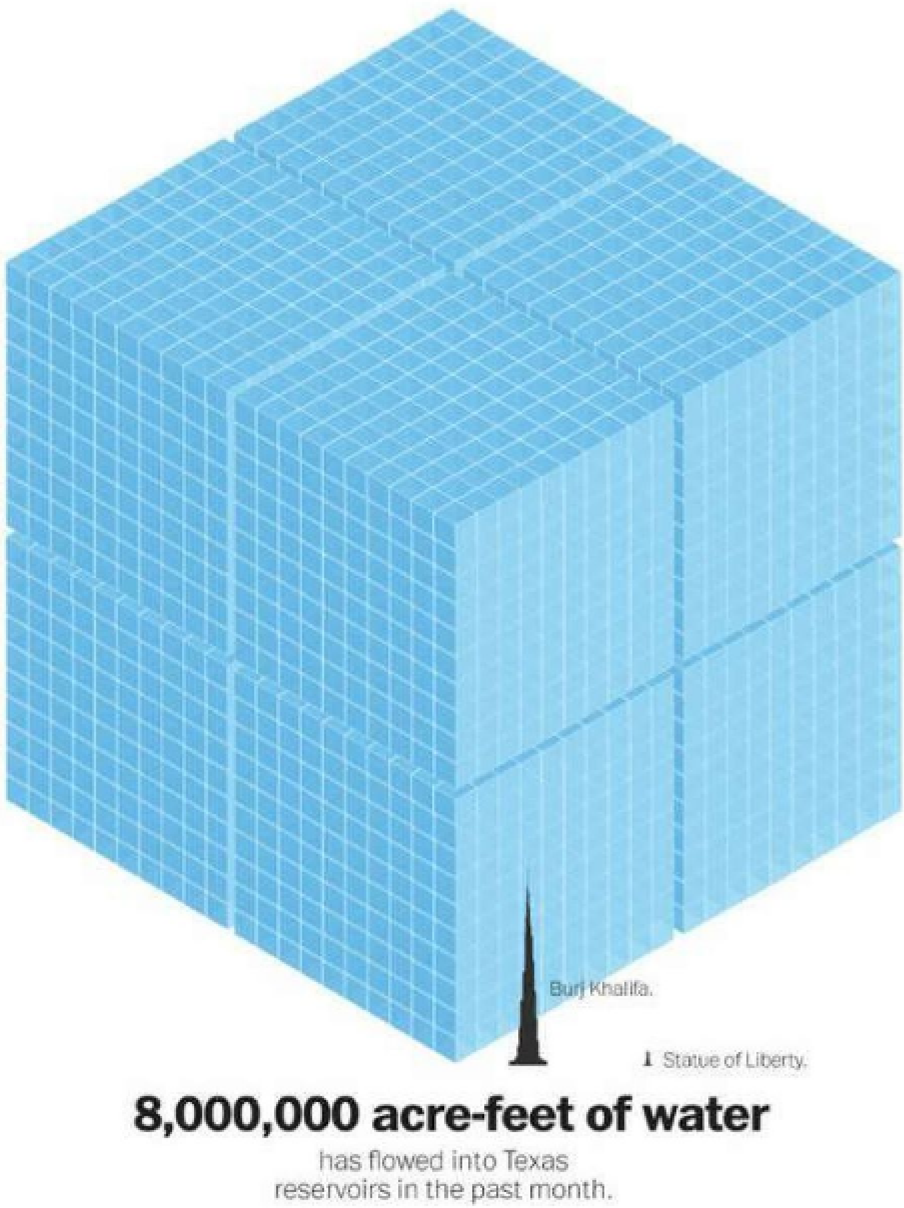
The area mark has two spatial dimensions (typically width and height or radius) and can represent quantitative values by variations in both position and size. The texture, colour or shape can also be varied to represent categories.

eg: https://ugc.futurelearn.com/uploads/assets/8b/47/large_hero_8b47bf17-203e-494f-aaf0-7eb13f8874c4.png

Form

represent quantitative values through variations in size (volume).

eg:



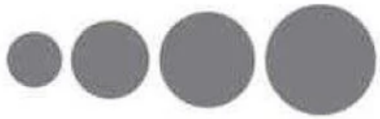
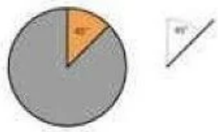
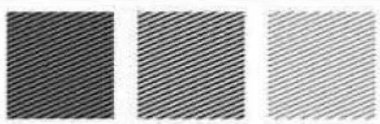
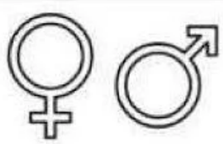
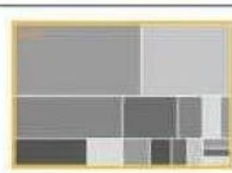
Attributes:

Now that we have the raw material to encode our data including points, lines and area (and occasionally forms), let's look at what attributes we can change to capture the data values

divided into three classes:

1. quantitative
2. categorical
3. relational

The table below from **Kirk (2016:161)**

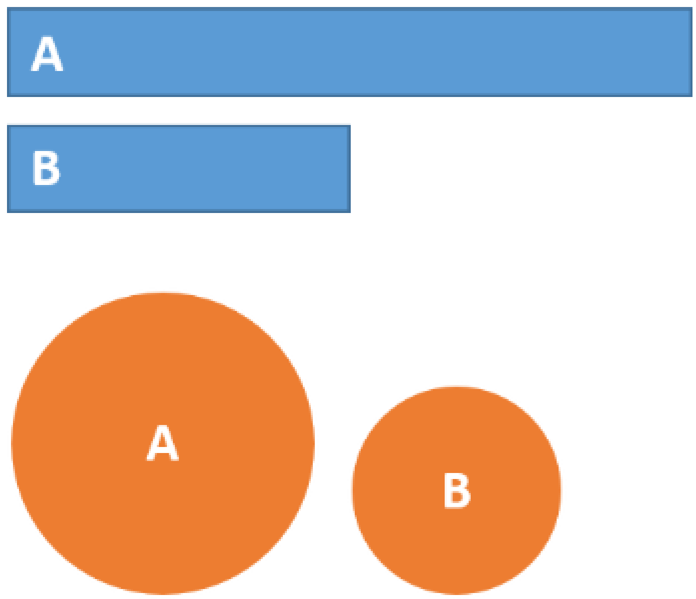
ATTRIBUTE	EXAMPLE	DESCRIPTION
QUANTITATIVE ATTRIBUTES		
Position		Position along a scale is used to indicate a quantitative value.
Size		Size (length, area, volume) is used to represent quantitative values based on proportional scales where the larger the size of the mark, the larger the quantity.
Angle/Slope		Variation in the size of angle forms the basis of pie chart sectors representing parts-of-a-whole quantitative values; the larger the angle, the larger the proportion. The slope of an incline formed by angle variation can also be used to encode values.
Quantity		The quantity of a repeated set of point marks can be used to represent a one-to-one or a one-to-many unit count.
Colour: Saturation		Colour saturation can be used (often in conjunction with other colour properties) to represent quantitative scales; typically, the greater the saturation, the higher the quantity.
Colour: Lightness		Colour lightness can be used (often in conjunction with other colour properties) to represent quantitative scales; typically, the darker the colour, the higher the quantity.
Pattern		Variation in pattern density or difference in pattern texture can be used to represent quantitative scales or distinguish between categorical ordinal states.
Motion		Motion is more rarely seen but it could be used as a binary indicator to draw focus (motion vs no motion) or by incorporating movement through speed and direction to represent a quantitative scale ramp.
CATEGORICAL ATTRIBUTES		
Symbol/shape		Symbols or shapes are generally used with point markers to indicate categorical association.
Colour: Hue		Colour hue is typically used for distinguishing different categorical data values but can also be used in conjunction with other colour properties to represent certain quantitative scales.
RELATIONAL ATTRIBUTES		
Connection/Edge		A connection or edge indicates a relationship between two nodes. Sometimes arrows may be added to indicate direction of relationship, but largely it is just about the presence or absence of a connection.
Containment		Containment is a way of indicating a grouping relationship between categories that belong to a related hierarchical 'parent' category.

Perceptual ranking

recall **NOIR** FROM PREV (nominal, ordinal, interval, ratio)

As an illustration of the difference that choice of encoding attribute can make in the understandability of your graph, look at the figure below (based on Figure 6.58 in Kirk 2016). If A is 10 then can you tell me the value of B in the respective bar (line) and circle (area) displays?

As an illustration of the difference that choice of encoding attribute can make in the understandability of your graph, look at the figure below (based on Figure 6.58 in Kirk 2016). If A is 10 then can you tell me the value of B in the respective bar (line) and circle (area) displays?

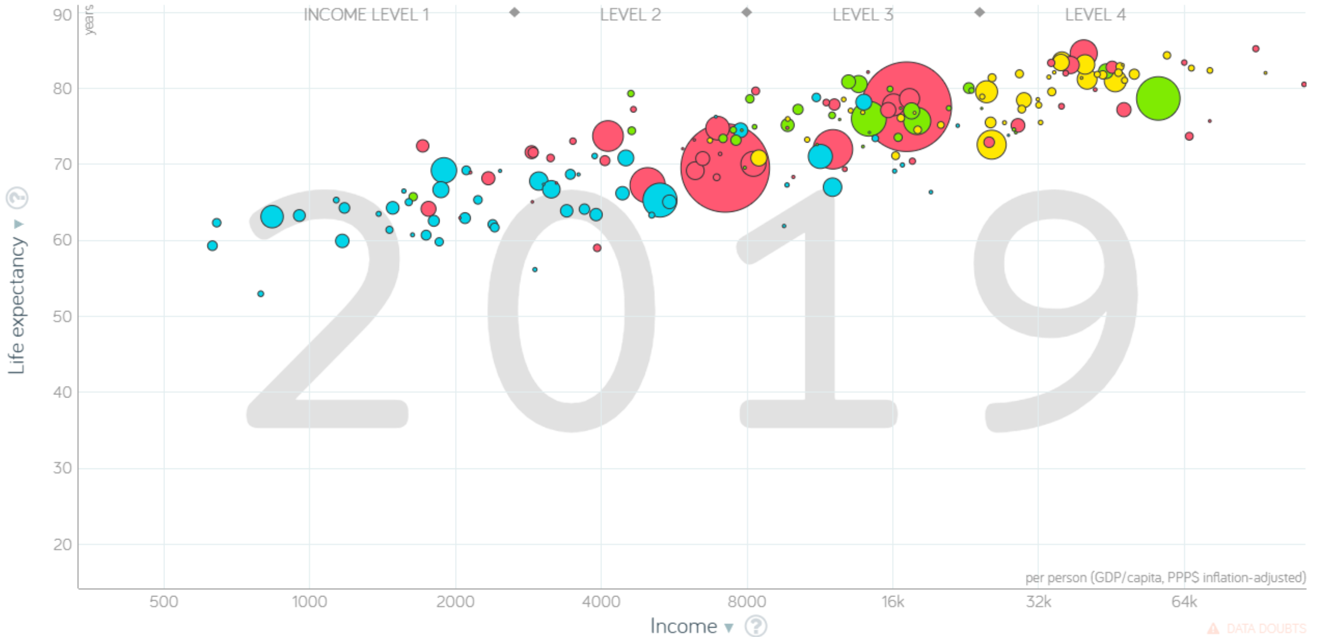


answer: The answer is 5 in both cases. It’s a lot easier to judge this in the line though!

Example of data encoding

https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600208

Let’s identify the marks and attributes used in this graph and consider the density of data encoded.



Data: Link to Google Sheets https://docs.google.com/spreadsheets/d/1uxf1q3lDkVKkEf-ov9Je7Y5wUctl_L0a8jkT7DcHqPI/edit?usp=sharing

Task: create three different charts using Google Sheets or Excel where the data is represented using a point, a line and a shape. What attributes can you use in each of these cases?

more here on creating viz: https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600211

wk4 quiz: https://www.futurelearn.com/courses/data-management-and-visualisation-data-driven-visualisations/3/steps/1600212/quiz/introduction

In [ ]: