

Data Visualization

Riccardo Caprile

September 2022

1 Introduction to Data Visualization

1.1 Defining Data Visualization(vis)

Visual representation of facts , events , elements and their attributes that are encoded by data. Not necessarily connected with computer science : data visualization exists since centuries.

The understanding of phenomena occurs through the interaction of methods and tools that facilitate the brain's ability to process information.

Computer based visualization system provide visual representations of datasets designed to help people carry out task more effectively. Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods. Viz allows people to analyze data when they don't know exactly what questions they need to ask in advance.

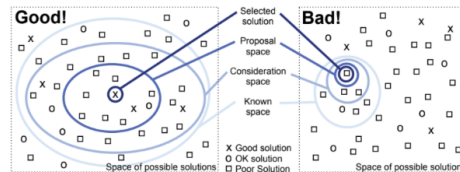
1.2 Possible uses of viz

- Long-term use for end users
- Exploratory analysis of scientific data
- Presentation of known results.
- Stepping stone to better understanding of requirements before developing models
- Help developers of automatic solution refine , determine parameters
- Help end users of automatic solutions verify, build trust

1.3 Information design principles

Idiom : Distinct approach to create or manipulate visual representation.

Idiom design space ; The design space of possible vis idioms is huge , and includes the considerations of both how to create and how to interact with visual representations.



The goal is to measure effectiveness , "It's not just pretty pictures".

1.4 Analysis : What , why , how

What is shown? : Data abstraction

Why is the user looking at it? : Task abstraction

How is it shown? : Idiom (visual encoding and interaction)

1.5 Vis Techniques



2 Visual Perception

2.1 The human visual system

The world around us can be understood as a vast display of visual data , from which we can pick out useful pieces of information. The shape of objects give us clues about their use , the textures of surfaces about properties of materials and the body language of humans and animals about their intentions.

At any moment , we can only accurately see a very small part of the entire visual field : **the fovea**, the small central area of the retina that provides accurate vision. All the rest is badly blurred and vaguely colored If we fix our

gaze on a point somewhere within our field of vision and try to detect objects in the remaining field without moving your eye it's difficult.

Where do we fixate our gaze??

Where there's something attractive and where we know there's something we want to look at.

The following three factors determine how easily we can find objects through visual searches :

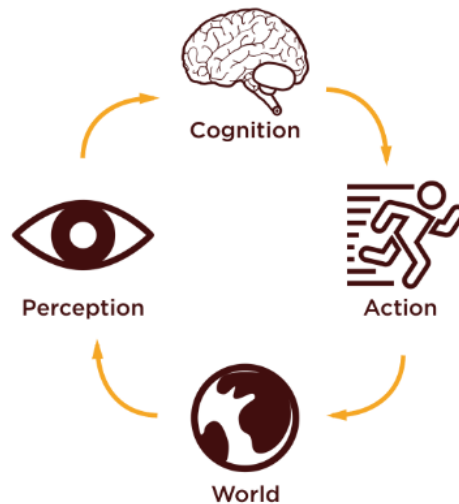
- **A priori salience** : features that our visual system is specialized to recognize
- **Top-down salience modification** : our sensitivity for different visual features is re-tuned based on what we are looking for.
- **Scene gist** : it is easier to interpret a familiar landscape or a figure than an unfamiliar one , as prior knowledge will help us select the appropriate search strategies.

2.2 Perception and Cognition

Visual perception does not occur in the eye , but in the brain.

Seeing is a multi-stage process , in which the sensory receptors in the eye transmit information via the optic nerve to the visual cortex for processing.

The image the brain receives has already been processed by the retina.



- **First Stage** : The neurons in the retina and in the visual cortex work in parallel searching for "low level" features in the visual field. Unconscious process

- Second Stage : The brain divides the visual field into regions and detects simple patterns , such as continuous contours.
- Third stage : The basic features identified in the previous stages are combined into more complex visual objects that are stored in the visual working memory that are compared with pre-existing items in memory.

Perception : Identification and interpretation of sensory information. From the physical stimulus to recognising information. Shaped by learning memory.

Cognition : The processing of information , applying knowledge

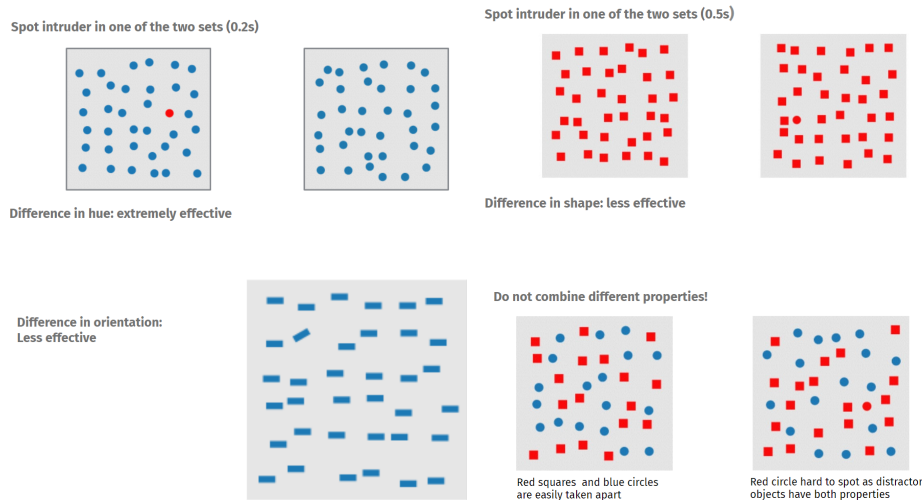
2.2.1 Pre-attentive processing

The distinctness of the features of an object in relation to its surroundings makes the object "pop out" from the background , catching our attention.

These features can be roughly divided into three channels :

- **Shape features** : include such things as position,size, elongation,orientation and texture
- **Color features** : include the hue,lightness and intensity of color, w
- **Motion** : include factors such as speed and direction of motion

We process these kind of features **pre-attentively**, which means that perceptual processing of such features happens faster than we can consciously direct our gaze.



Finding a red circle among red crosses or blue circles happens very quickly , but the search will slow down considerably when the task is to find a red circle among a mix of red crosses and blue circles. It is a good idea to make use of differences in these features sparingly and to use them only to highlight the most important details.

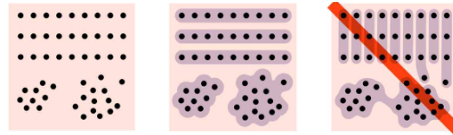
2.3 Gestalt Laws

A Gestalt is a shape that forms as a combination of tunable (pop up) features , the contours of elements , the areas enclosed by groups of elements. The perception of Gestalt relies on the focusing of attention. These laws describe the principles of our tendency to perceive individual visual features as **groups** and **entities**.

Now we see the 7 laws.

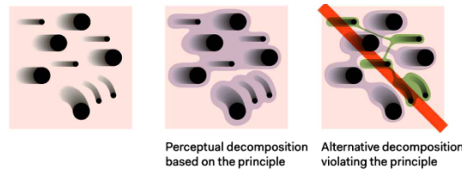
2.3.1 Law of proximity

Elements that are close to each other tend to be perceived as belonging together.



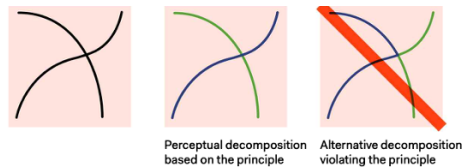
2.3.2 Law of common fate

Elements that move together tend to be perceived as belonging together



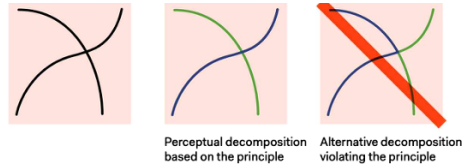
2.3.3 Law of Similarity

Elements that are similar to each other , for example with regard to color , size , or shape , tend to be perceived as belonging together



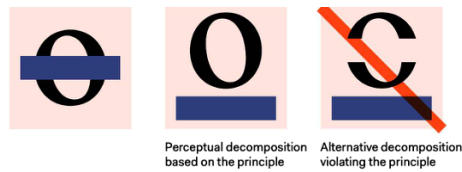
2.3.4 Law of continuity

With intersecting shapes, parts that form a continuous line are perceived as belonging together



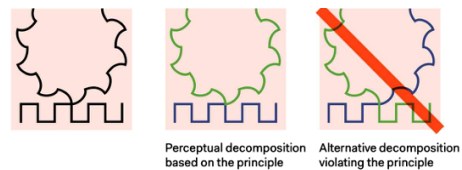
2.3.5 Law of closure

Elements that form a closed shape are perceived as belonging together. This is true even when the figure is missing elements.



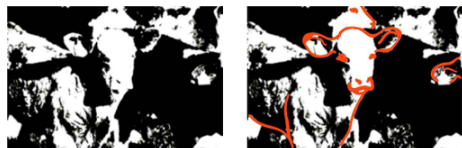
2.3.6 Law of good Gestalt

Elements that combine to form figures that have as simple a shape as possible tend to be perceived as belonging together or as being distinct.



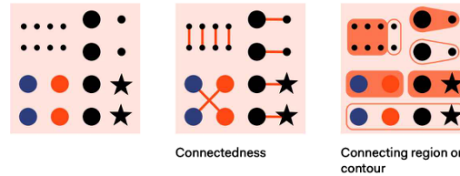
2.3.7 Past experience principle

Elements that have, according to past experience, been seen together are perceived as belonging together when seen again or in a new context



2.3.8 Connectedness and connecting regions

Cause elements to appear to belong together, even if the resulting combinations are in conflict with other Gestal principles.



3 Color and Perception

3.1 Physical and biological principles on vision

Colors have two distinct roles in information design.

On one hand, they are directly related to encoding and organizing of information. On the other hand, colors are also esthetically and culturally significant.

Colors are powerful visual codes, and readers always attach meanings to them.

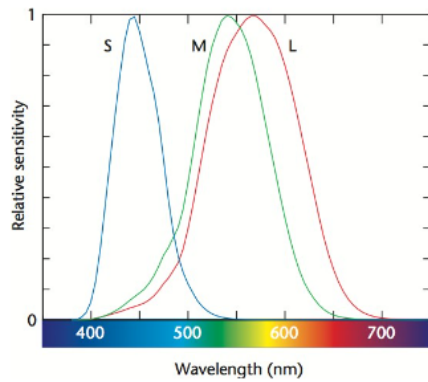
Humans have **trichromatic color vision**. The light sensitive cone cells in our eyes fall into three groups, each of which responds to different wavelength of light: red, green, blue. All other colors we see are mixtures of these primary colors.

Metamerism: is a perceived matching of colors with different spectral power distributions. Colors that match this way are called metamers.

3.1.1 Human eye

Our retina is a layer of tissue at the back of the eye. It contains cells called **photoreceptors**.

They convert light rays into nerve signals, which are then processed by nerve cells in the inner retina, sent to the brain, and translated as images. The two types of photoreceptors are known as **rods** and **cones**. Rods are responsible for peripheral and night vision. They detect brightness and shades of gray. Cones are responsible for day vision and color perception. Each type of cone is sensitive to a different range of frequencies in the visible spectrum. Called L, M, S from the zones of spectrum at which they have highest response.



3.2 Color Encoding

3.2.1 How do we encode colors?

Adjust the levels of three different light sources (R,G,B) to match a color in the spectrum. The levels of the three sources encode the given color with respect to such sources.

3.2.2 Grassman's Law

If tri-stimulus RGB1 matches S1 and tri-stimulus RGB2 matches spectrum S2, then tri-stimulus RGB1 + RGB2 matches spectrum S1 + S2.

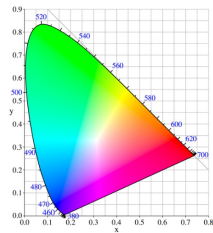
We can encode an infinite set of colors.

In practice :

1. Select three monochrome light sources (R 645 nm , G 526 nm , B 444 nm)
2. Sample the visible spectrum and match tri-stimuli for each sample
3. Interpolate color matching functions on the three channels.

CIE chromaticity diagram

- factor out luminance
 $x = X/(X+Y+Z)$ $y = Y/(X+Y+Z)$
- any color can be encoded as xyY
- chromaticity encoded as (x,y)
- pure saturated colors on the horseshoe shaped path
- the straight line below is called the *purple line*
- the more internal, the less saturated
- greyscale lies at the central point



3.2.3 RGB Space

Three sources of light R,G,B adjustable in the range [0,1].

Maps RGB space to XYZ by determining the tri-stimulus for each base color.

$$\begin{bmatrix} R & G & B \end{bmatrix} M = \begin{bmatrix} X & Y & Z \end{bmatrix}$$

$$M = \begin{bmatrix} X_R & Y_R & Z_R \\ X_G & Y_G & Z_G \\ X_B & Y_B & Z_B \end{bmatrix}$$

3.2.4 Color Composition

Three primary colors can be combined to obtain all others.

Additive composition : add colors by combining lights of a certain wavelength.

Subtractive composition: subtract colors from a reflective surface using pigments that absorb light of a certain wavelength.

3.2.5 The elements of color perception

- **Hue**: Refers to whether the color perceived is , for example red or yellow. The perceived hue depends on the wavelength composition of light reflected or emitted into the eye from the object.
- **Lightness** : is the visual system's interpretation of the brightness of a light source. Most significant
- **Saturation** : Refers to the intensity of color.



3.2.6 Transparency

Colors can be defined to have a certain degree of transparency/opacity. Transparency is encoded in a separate channel , called the alpha channel. Alpha in range [0,1] - [transparent,opaque]. Useful when colors overlap. Result depends on order of overlap $C = \text{afront} * C_{\text{front}} + (1 - \text{afront}) * C_{\text{back}}$.

3.3 Perceptual Illusion

Our eyes do not precisely measure the wavelength of the amount of light reflected from the object being looked at. Context and prior knowledge considerably affect our perception of color. We perceive an object to the same color , whether

lit by a lamp or sun.. This phenomenon is called **color constancy**. Color constancy creates problems when color is used for communicating abstract information , as the closely phenomenon of simultaneous contrast causes adjacent colors to affect each other. A color will appear lighter whether surrounded by dark colors.

3.3.1 Color anomalies in vision

- Color Blindness : Protanopia in case of red insensitivity and protanomaly in case of poor red sensitivity. Deuteranopia in case of insensitivity to green and deuteranomaly in case of poor sensitivity to green. Tritanopia in case of insensitivity to blue and tritanomaly in case of poor sensitivity to blue
- Low vision
- Trauma

3.4 The use of colors in data visualization

It is recommended to use saturated but light pastel colors for large surfaces and dark and intense colors for small details. In this way , the interference caused by simultaneous contrast can be avoided and the most important thing stand out

Using a color palette based on natural color has another advantage : a separate key is not necessarily required. Watery area blue , forest dark green etc.

3.5 Quantitative and Qualitative color scales

On a **Qualitative** color scale , different colors represent objects that belong in different groups or categories on a nominal scale. For example , election results can be presented in such a way that each party has its own distinctive color in the graphics.

The goal is to create a color palette in which different colors are as distinct from each other as possible.

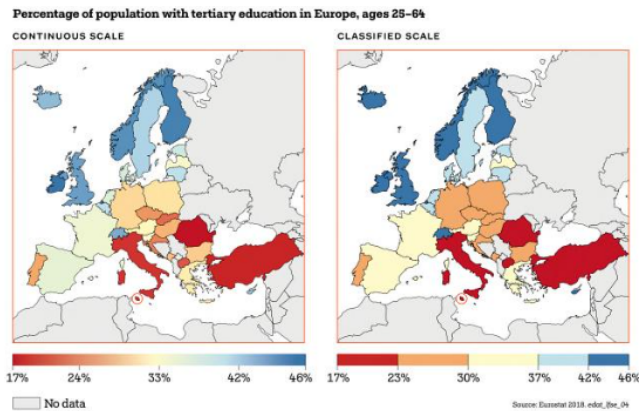
Using the highest number of colors is not practical and could be counterproductive.

Most often, when discussing color scales, what is meant is **quantitative**, rather than qualitative scales. Visual encoding of quantitative data should be based on variation of lightness of color.

In general , a color scale should be designed so that when the value of a variable increases , so does the contrast between the color and its background.

Quantitative color scales can be divided into **classified** and **unclassified** scales. Classified color scales are used to represent an ordinal scale of measurement , and unclassified ones for ratio and interval scales.

Classified colors scales are far easier to read precisely.



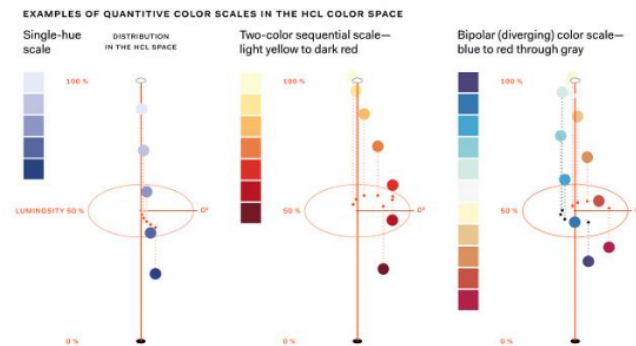
3.5.1 Types of color scales

The most common type of quantitative color scale is the **sequential scale**, in which the variation in the lightness of the color is used to encode values

Single hue scale : a sequential scale based solely on lightness of color. Individual colors are more reliably distinguished from each other in sequential multi-hue scales, where the variation in lightness is combined with variation in hue and saturation

Two-color sequential scale : is a multi-hue scale where a number of intermediate colors are located between two colors that mark the extremes of the scale in a uniform color space.

Bipolar scale : is multi-hue scale where an intermediate , neutral color is added halfway between the two extremes.



Scientific visualization , in particular , often use **spectral scale** , which is solely based on difference in hue. It has the advantage that reading precise values in a figure like this is easier than in visualization using an unclassified sequential color scale. Perceptually unordered and non linear.

