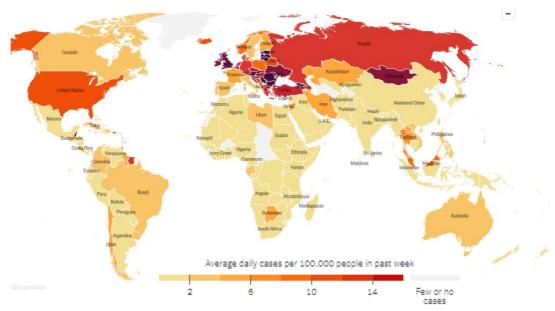
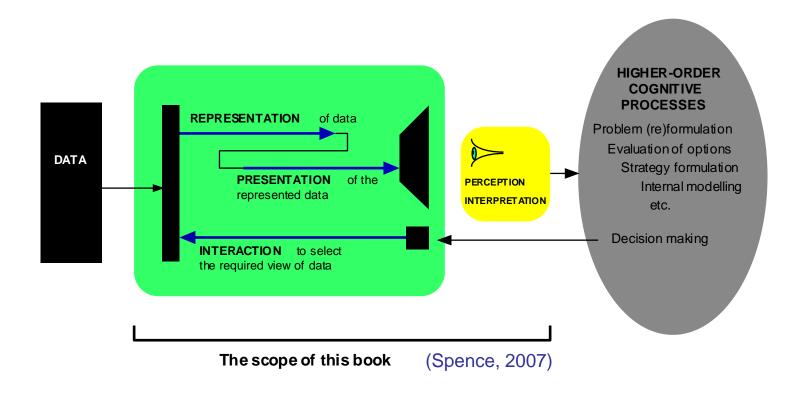


Representation Visual encoding of value



Covid-19 World Map: Cases, Deaths and Global Trends - The New York Times

Beatriz Sousa Santos, University of Aveiro, 2025

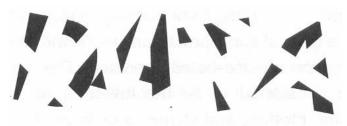


Interaction with data governed by high-order cognitive processes:

- Representation (how to code visually the data)
- Presentation (what/when/where to show on the screen)
- Interaction (how to let users explore the data)

Remember:

 The Human Visual system is the product of millions of years of evolution



 Although very flexible, it is tuned to data represented in specific ways

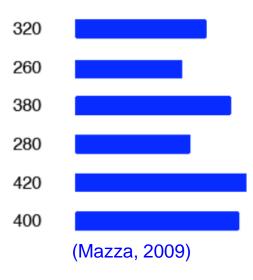


 If we understand how its mechanisms work we will be able to produce better results Pre-attentive attributes can help observers to see before though

Example: Count the number of 7s

https://www.youtube.com/watch?time _continue=121&v=AiD6etOB6qI Visual attributes as size, proximity are quickly processed by visual perception, before the cognitive processes come into play

Example: mapping numerical values to the length of bars:





Some more examples on how humans see...

And their limitations...

Example of **change blindness** (Spence, 2007)

Example of change blindness (Spence, 2007)

What is missing now?





Inattentional blindness

The Monkey Business Illusion

Change blindness

<u>Derren Brown - Person Swap</u>



Designing a Visualization

- The process must be preceded by good design
- The main problem in designing a visual representation is the choice of mapping, as to:
 - help the user to attain their goals
 - faithfully reproducing the information codified in the data

- The visual representation suitable depends on:
 - the nature of the data and phenomenon
 - the users' tasks and needs (the questions)
 - the user profile and context of use

- ...

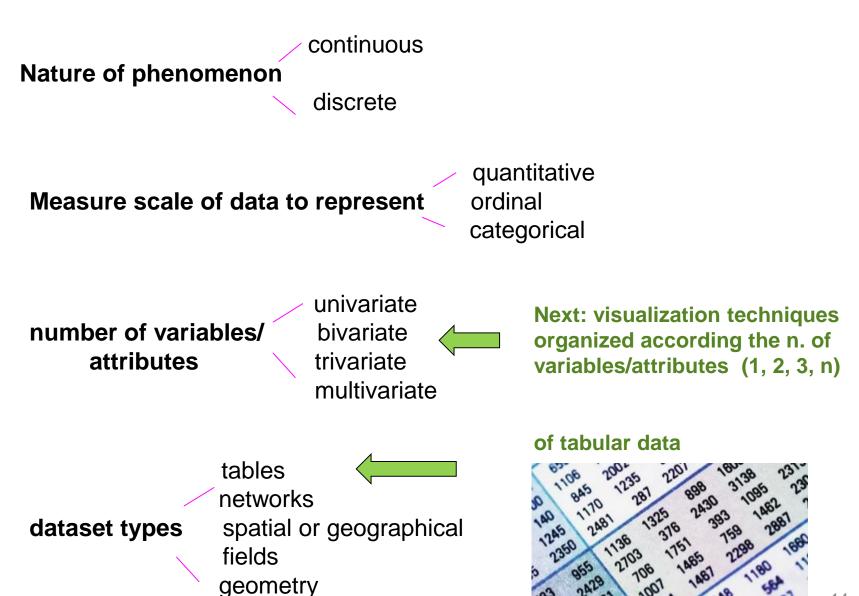
Procedure to follow to create visual representations of abstract data

Taking into consideration the users' tasks, profile and context of use:

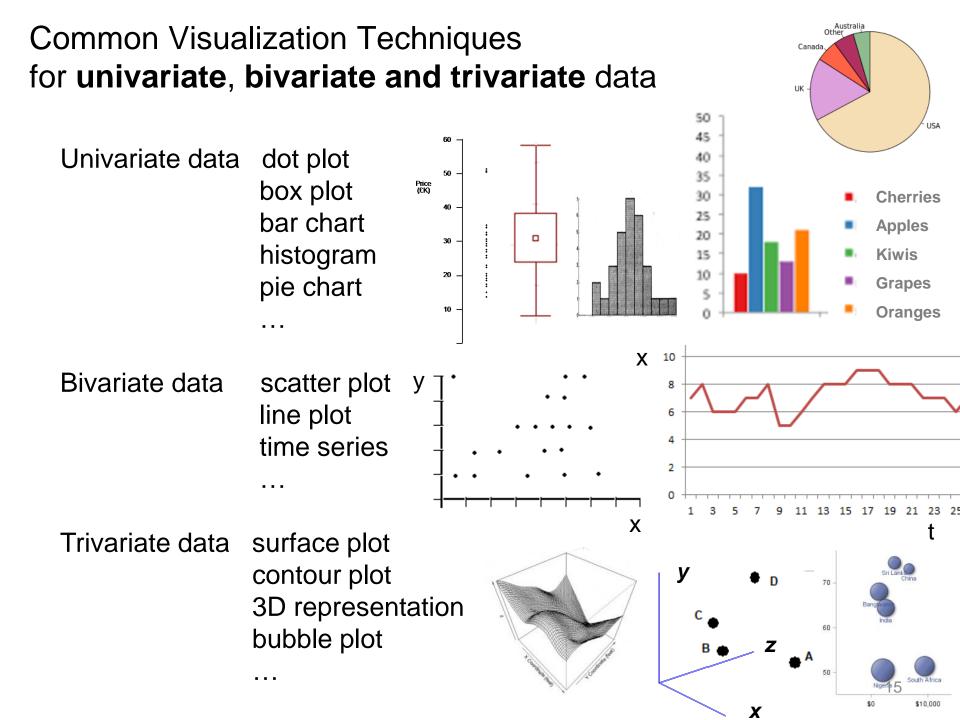
- 1. Define the problem and the users' questions, profile and context of use
- 2. Examine the nature of the phenomenon and data
- 3. **Pre-process** the data
- 4. Determine the **number of attributes**
- 5. Choose the **visual structures** (how to represent the data)
- 6. Establish **how and when** to present to the user

and the type of interaction

Important aspects to consider to select a Visualization technique:



14



Representing univariate data

A common situation consists in representing a set of values

Well established techniques exist

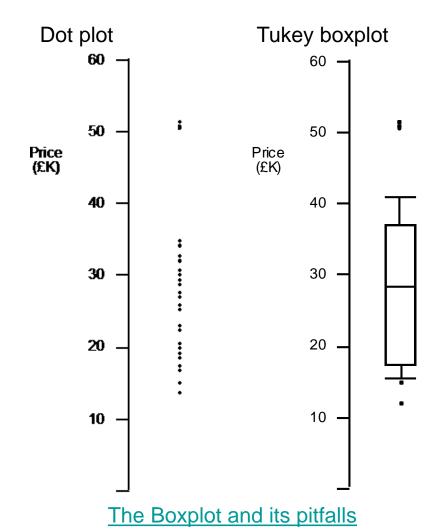
But new ones can be invented!

Example:

Price for a number of cars:

- dots on a linear scale
- box plot (that will answer many questions: median value, outliers,...)

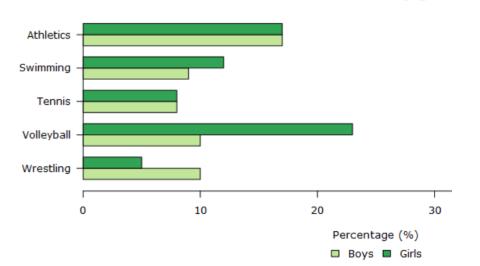
(Spence, 2007)

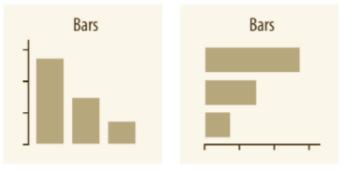


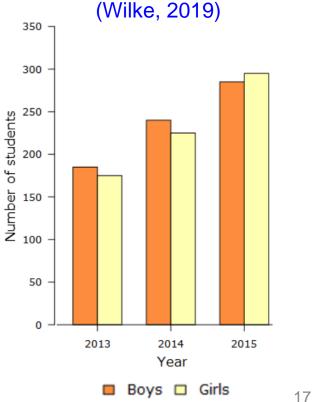
A common way to representing univariate data

- precise detail is often not needed
- A bar chart is a common way to represent one attribute,
- but we may combine to represent more attributes

Sports practiced by 15-year-old students by gender







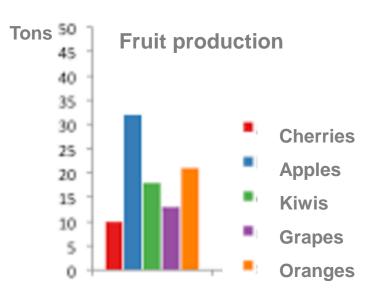
Simple (and common) representations of one attribute data

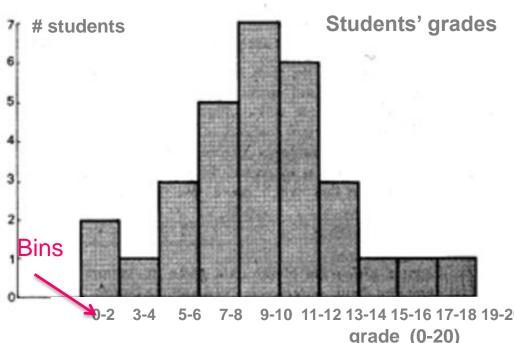
Two common techniques not to be confused!

Histogram represents a distribution of numerical data

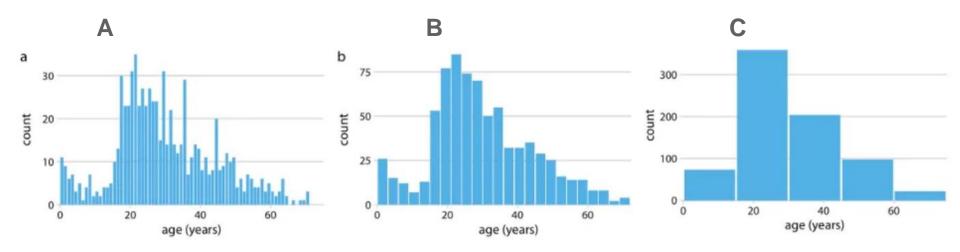
Bar chart represents number of occurrences of categorical/ordinal data

Both represent data by rectangular bars(vertical or horizontal) with length proportional to the values they represent





- Selecting an adequate bin width is important
- Which seems better?



(Wilke, 2019)

Another simple (and too common) representation

Pie Chart

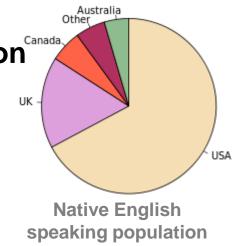
Represents numerical proportion, parts of an whole

The arc length of each slice (its central angle and area), is proportional to the quantity it represents

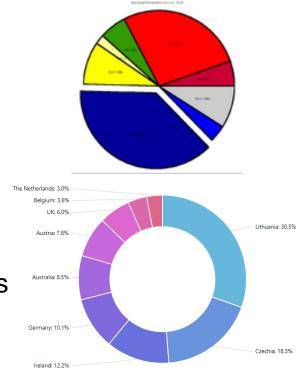
Are much controversial: many experts recommend avoiding them

Save the Pies for Dessert

It is difficult to compare different sections of a pie chart, or to compare data across different pie charts



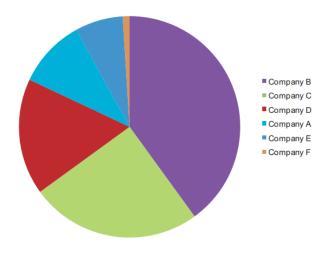
Variations of pie charts:



Exercise

What are the issues with this pie chart?

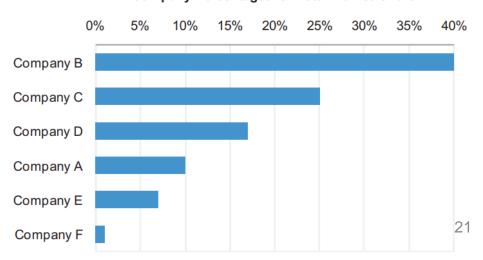
Companies	Percentage
Company B	40%
Company C	25%
Company D	17%
Company A	10%
Company E	7%
Company F	1%
Total	100%



Save the Pies for Dessert

 What if the user needs to assess and compare the companies market share?

Company Percentages of Total Market Share

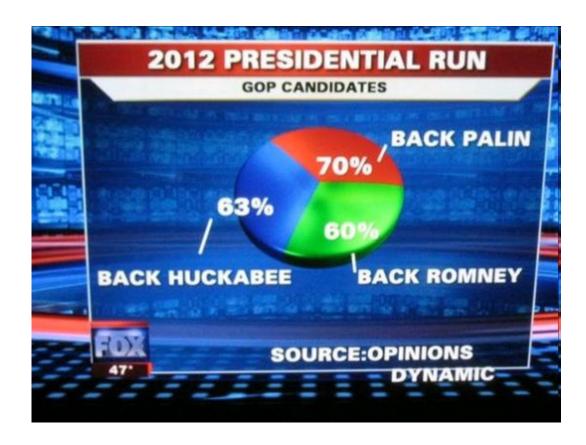


- Simple criteria to determine whether a pie chart is acceptable
- Consider it only if:
- The parts make up a meaningful whole
- The parts are mutually exclusive
- There are <6 parts and slices have not very different sizes

If the main purpose is to compare between the parts, use a different chart!

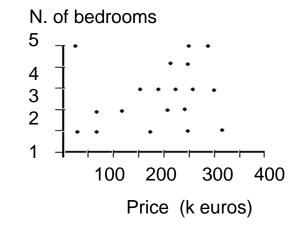
<u>Understanding Pie Charts - eagereyes.org</u>

Prize winning Pie chart!



Representing bivariate data

The scatterplot is a conventional representation



Each observation is represented by a point on a two dimensional space.

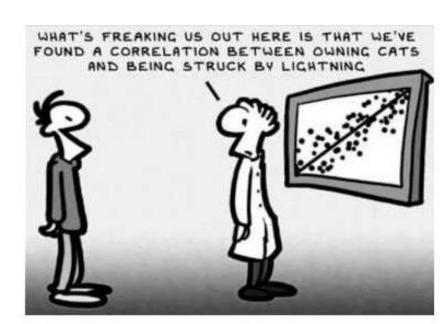
The axes are associated with these two attributes

This representation affords awareness of:

- general trends
- local trade-offs
- outliers



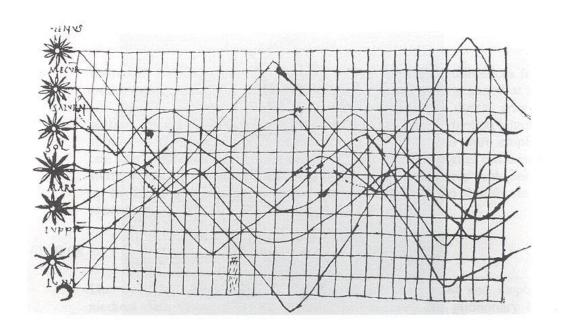
Correlation is not causation



Representing bivariate data

The line chart

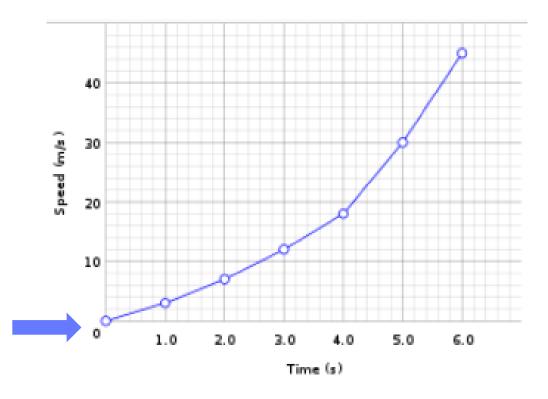
One of the oldest known and ubiquitous Visualizations



Inclination of orbits along the time - Xth century (Tufte, 1983)

 A line chart or line plot or line graph or curve chart displays a quantitative variable as a series of data points called 'markers' connected by straight line segments

- Basic type of chart common in many fields
- Often used to visualize a trend in data over intervals of time
- Is it reasonable to connect the data points by a solid line?

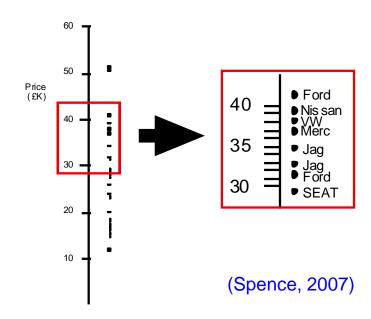


- If one attribute is more important than the other or must be examined first,
- it may be appropriate to employ logical or semantic zoom

Example:

Analyzing a list of cars:

- price is the first attribute to examine
- semantic zoom reveals data about a second attribute

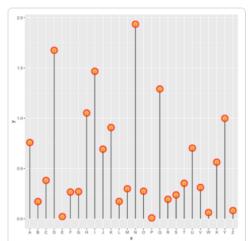


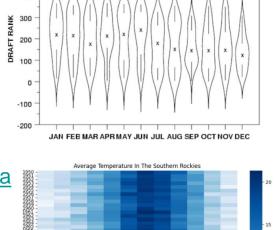
 This technique is quite general: it can encompass many attributes and many levels of progressive zoom

Some more visualization techniques:

- These are only some of the visualization techniques to represent value
- There are a lot more ...
- Here are some that you may have seen already
 - violin plots
 - lollipop charts
 - Heat-maps

<u>Heat map – Wikipedia</u>



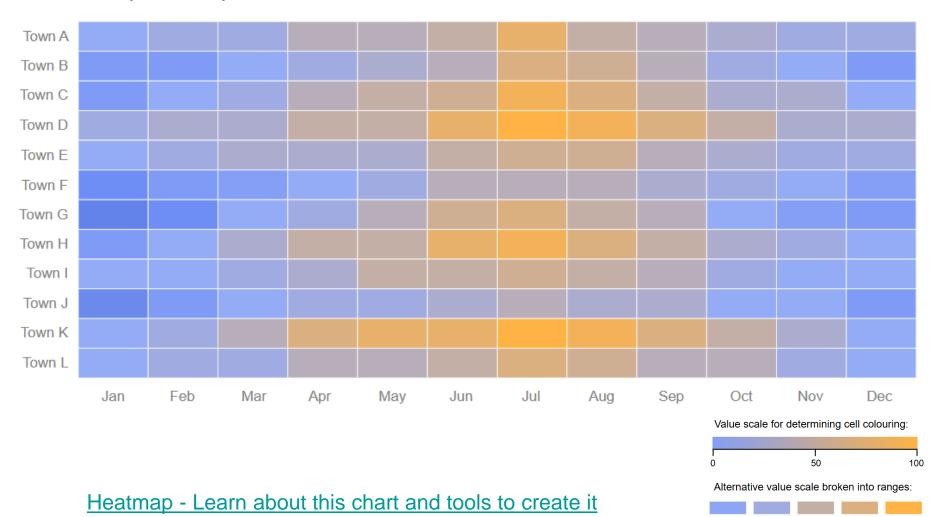


Violin plot - Wikipedia

Lollipop plot | the R Graph Gallery

Heat map

May represent quantitative or ordinal data also in tabular format through colour and help detect patterns

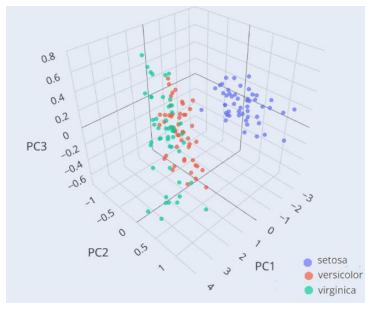


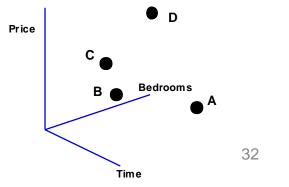
Representing trivariate data

 Since we live in a 3D world, representing trivariate data as points in a 3D space and displaying a 2D view seems natural

- However, these representations of abstract data can be ambiguous
- This can be mitigated by interaction, allowing the user to reorient the representation

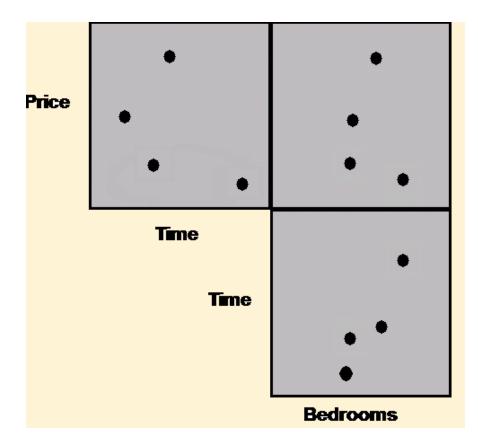
"for 3D to be useful, you' ve got to be able to move it" (Spence, 2007)

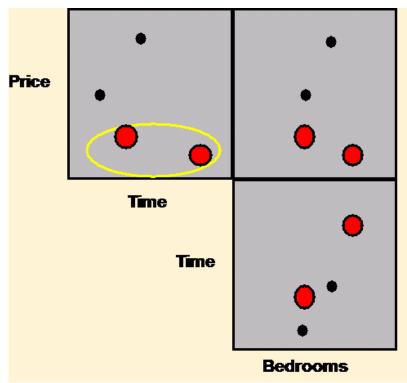




- Interaction (brushing) can help objects identified in one view are highlighted in the other two planes
- change blindness must be taken into account and ensure that the user notices the highlight in the other two planes

(Spence, 2007)





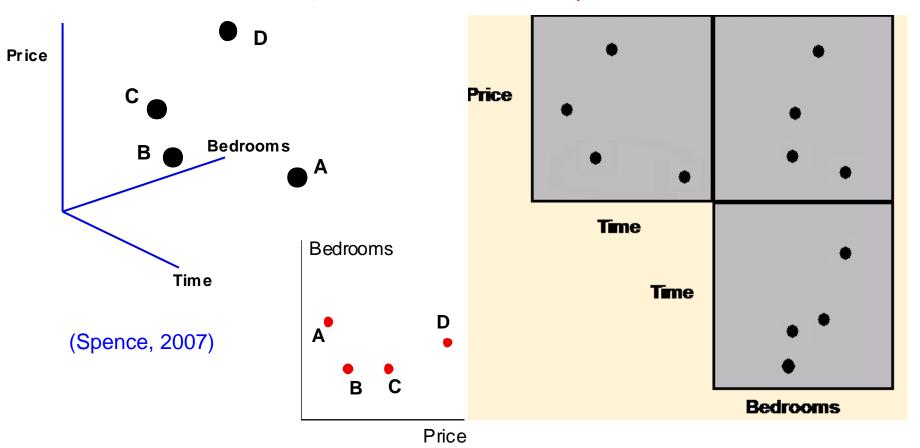
The highlighting of houses in one plane is brushed into the remaining planes.

33

 An alternative representation for trivariate (and hypervariate) data is a structure formed from the three possible 2D views of the data

Example: houses (price, number of bedrooms, time of journey to work)

Scatterplot matrix

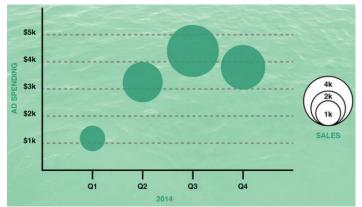


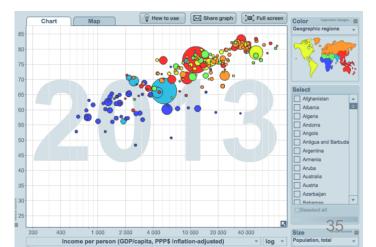
Other Simple (and common) representation of 3D data

In a bubble chart data are represented as a disk that expresses two
of the values through the disk's xy location and the third (less
accurately) through the size of the disk (radius or area?)

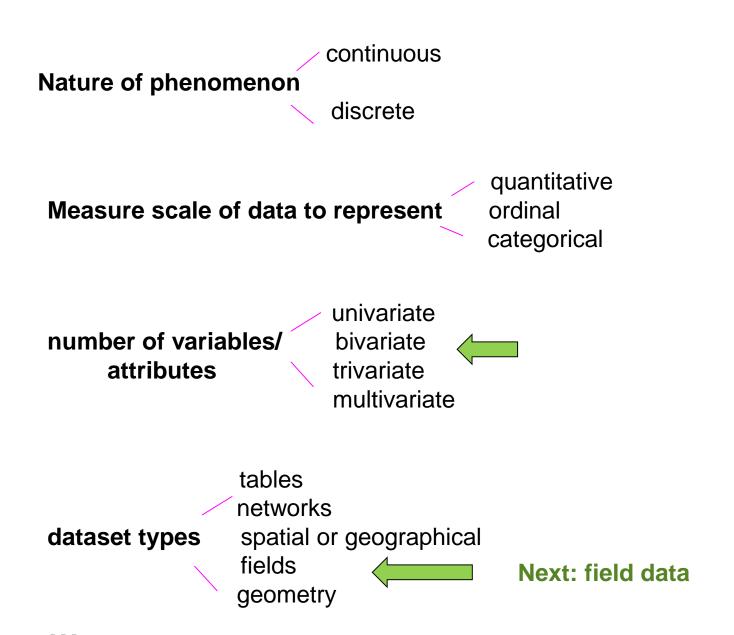
 Mapping the atribute/variable to disk size must be done carefully. The interpretation of may be ambiguous

 Representing one more dimension through color





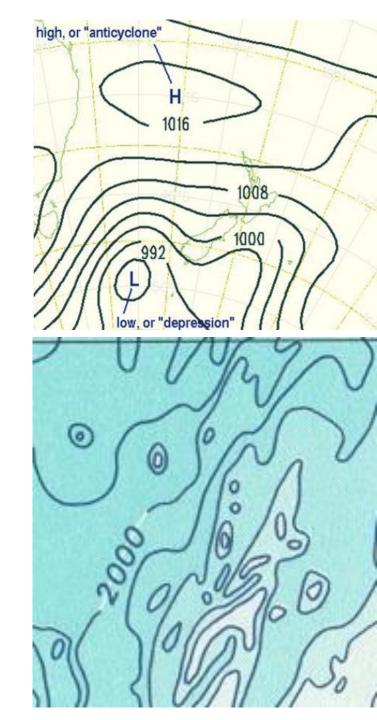
Important aspects to consider to select a Visualization technique:



36

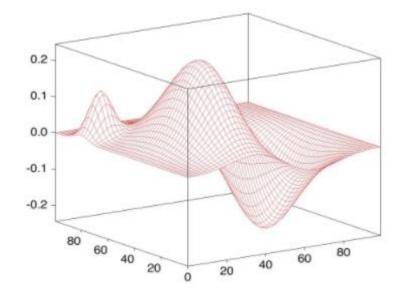
Simple representations of a function (field) of two variables

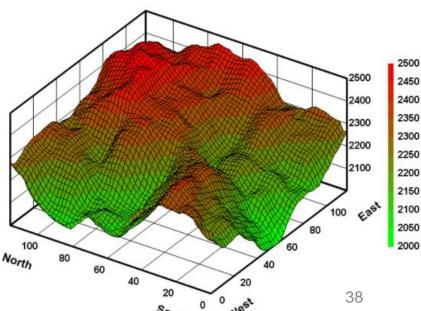
- Contour plots
- contour line (also isoline, isopleth, or equipotential curve) of a function of two variables is a curve along which the function has a constant value, so that the curve joins points of equal value.
- Often used in SciVis
- Typical in meteorological charts (isobars and isothermal curves)
- and maps (to represent altitude or depth)



- Surface plots (also often used in SciVis)
- May be combined with color

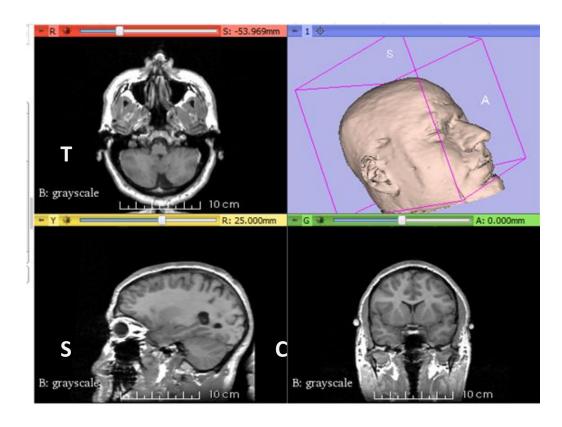
(preferably in a redundant way and carefully selecting the scale)

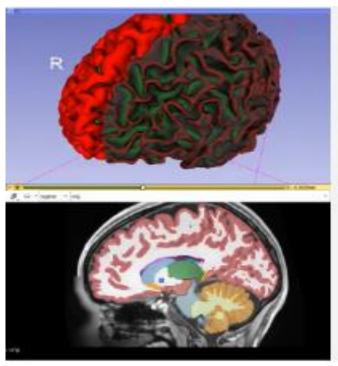




Representing a 3D scalar function volumetric data

Volumetric data

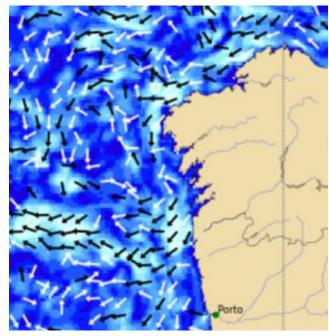


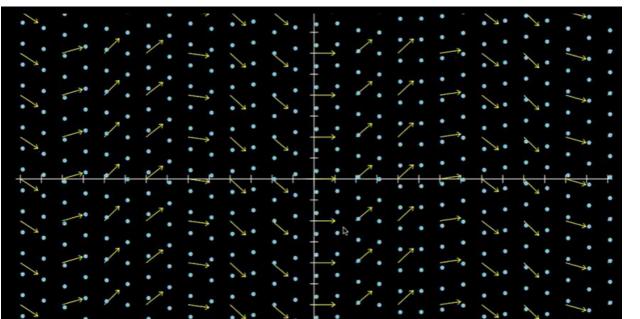


3D Slicer image computing platform | 3D Slicer

Representing vectors in a 2D field



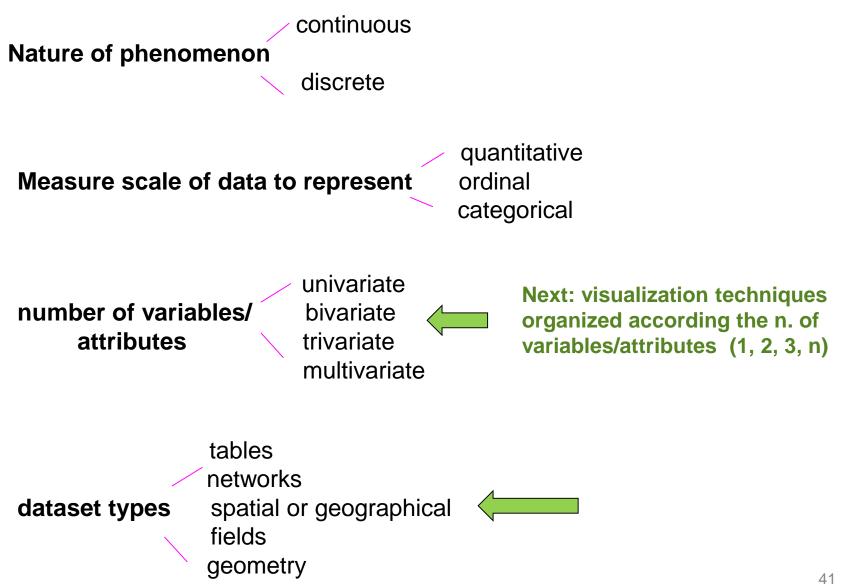




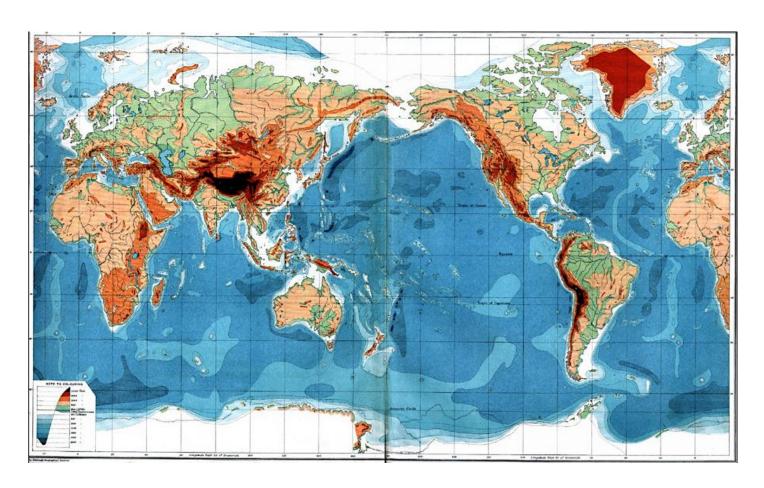
IPMA - Corrente à superfície

Fluid flow and vector fields (video) | Khan Academy 40

Important aspects to consider to select a Visualization technique:

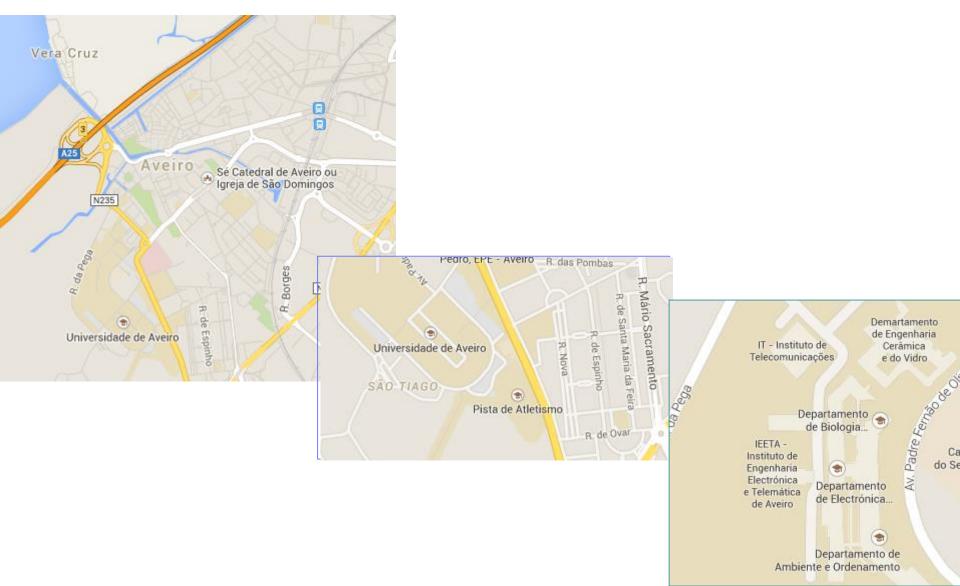


A special category of trivariate data: Maps (latitude and longitude + a value)



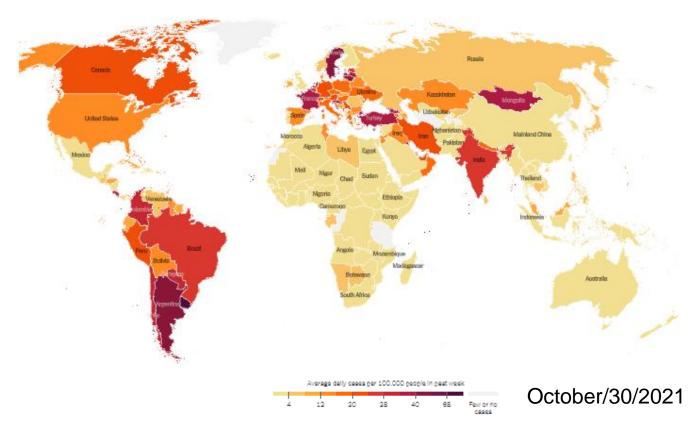
1915 - Orographic Chart of the World

Semantic zoom in Google Maps



Choropleth maps - A standard approach to communicating aggregated data by geographical areas using color encoding of the geographic area

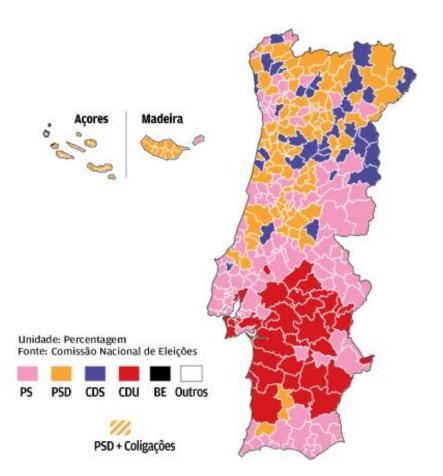
They require some care: what are the possible issues?



Covid-19 World Map: Cases, Deaths and Global Trends - The New York Times

They require some care: what are the possible issues?

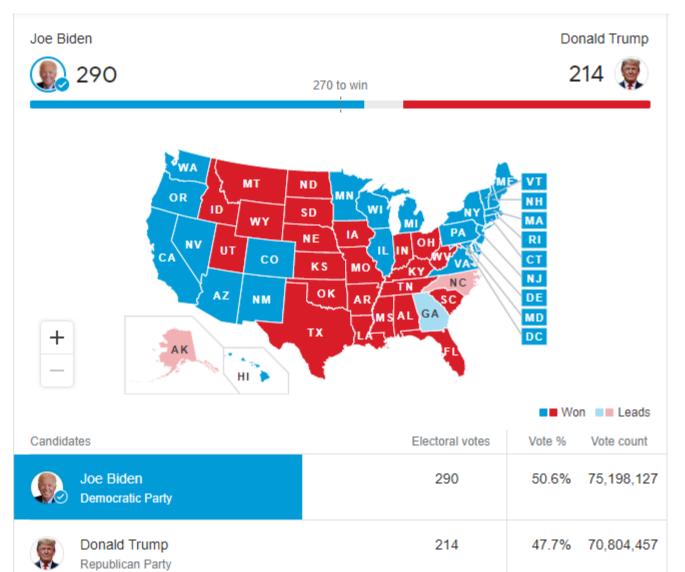
Portuguese elections in 1976





How can these issues be mitigated?

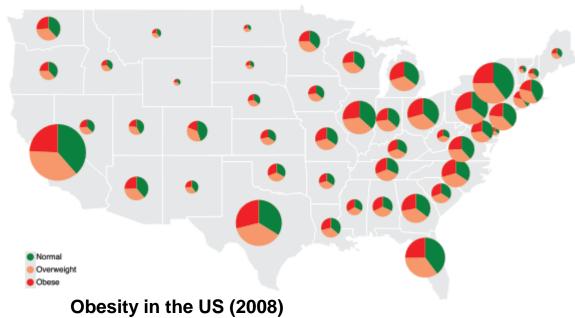
Visualizations of the US 2020 Election (choropleth + bar)



- Graduated Symbol Maps are an alternative to the choropleth map;
- Symbols are placed over an underlying map; may show more dimensions

(Heer et al., 2010)

Avoid confounding geographic area with data values



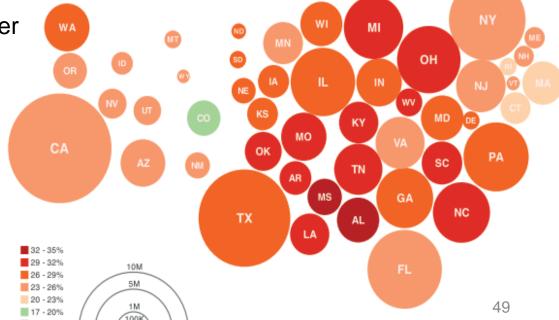
48

- Cartograms distort the shape of geographic regions so that the area directly encodes a data variable.
- There are several types
- Dorling cartograms represent each geographic region with a sized circle placed so as to resemble the true geographic configuration

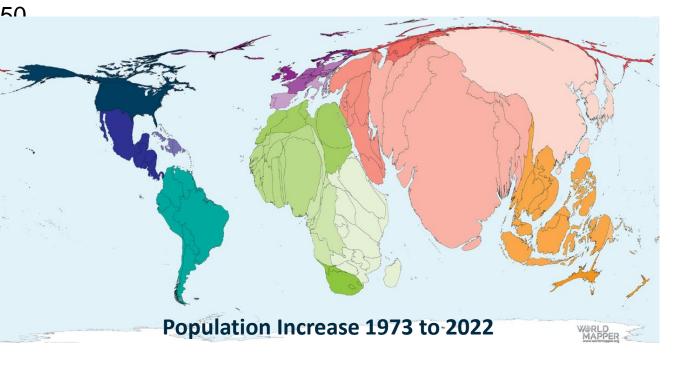
In this example:

- area encodes the total number of obese people per state

- color encodes percentage of obese population



Obesity in the US (2008) (Heer et al., 2010)



Cartograms:
Using distortion to represent a variable





Representing Hypervariate (multivariate) data

Many real problems are of high dimensionality
 (even after reducing dimensionality...)

 The challenge of representing hypervariate data is substantial and continues to stimulate invention

- Some of the mentioned representation techniques can be scaled to represent hypervariate data (to a limited extent),
- Reducing complexity can be mitigated through different ways (e.g. data preparation, interaction, ...)

Reducing complexity in Visualization

- Consider dimensionality reduction of the dataset
- Several methods can be used to reduce dimensionality (e.g.):
 - Principal Components Analysis
 - t-SNE (t-distributed stochastic neighbor embedding, etc.
- But there are other alternatives along the visualization pipeline (eg. filter)

Other Bibliography and interesting links .

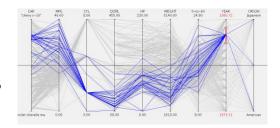
Available on Moodle



Techniques for Hypervariate (multivariate) data Visualization

Coordinate plots

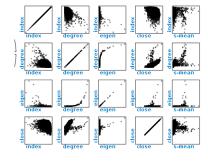
parallel coordinate plots



star (radar, or spider) plots



Scatterplot Matrix



Maps



Icons/glyphs





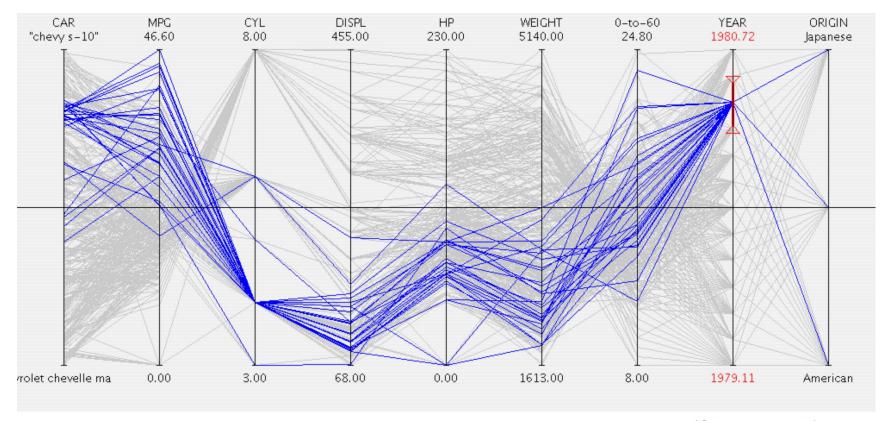
Parallel coordinates plots are one of the most popular techniques for

hypervariate data

They have a very simple basis

Make	Price (£)	MPG	Rating	Age (yrs)
Ford	15,450	31	****	3
Chevy	12,450	27	***	4

- - -





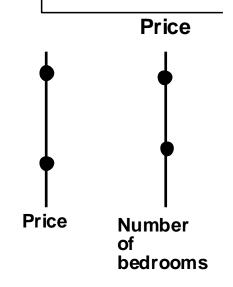
A

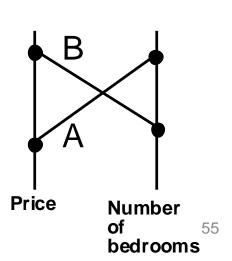
Consider a simple case of bivariate data:

1- A scatterplot represents the price and number of bedrooms associated with two houses

2- the axes are detached and made parallel; each house is represented by a point on each axis

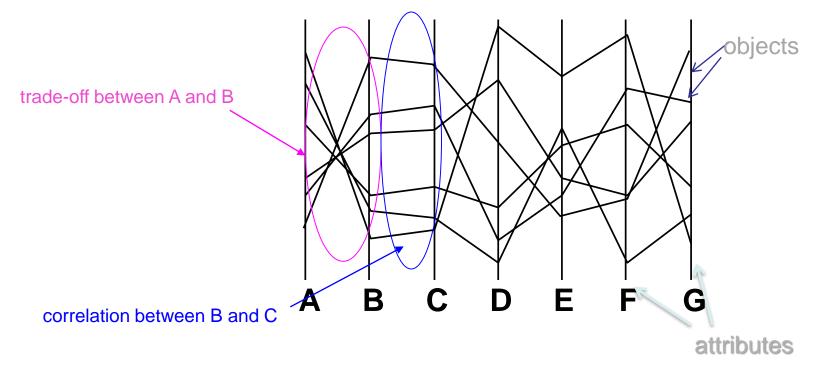
3- To avoid ambiguity the pair of points representing a house are joined and labeled



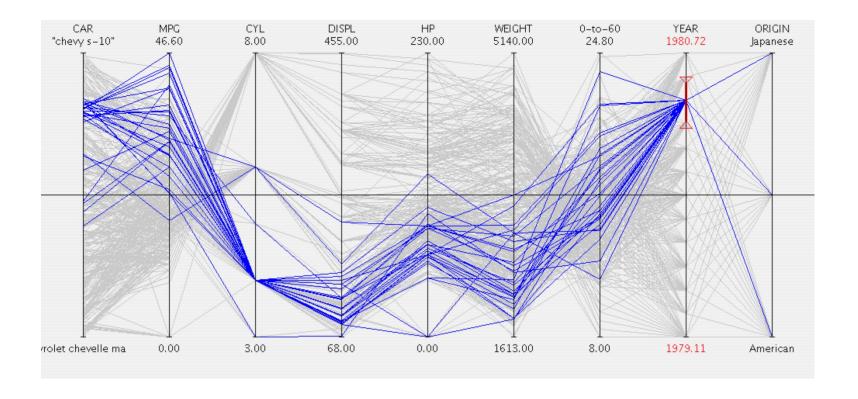


 For objects characterized by many attributes the parallel coordinate plots offer many advantages

A example for six objects, each characterized by seven attributes:

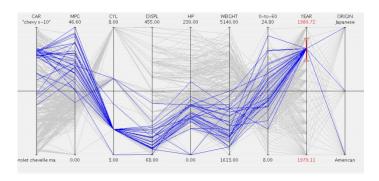


The trade-off between A and B, and the correlation between B and C, are immediately apparent. The trade-off between B and E, and the correlation between C and G, are not.



A parallel coordinate plot representation of a collection of cars, in which a range of the attribute *Year* has been selected to cause all those cars manufactured during that period to be highlighted.

Properties of parallel coordinate plots:

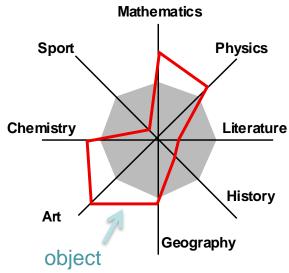


- Suitable to identify relations between attributes
- Objects are not easily discriminable; each object is represented by a polyline which intersects many others
- They offer attribute visibility (the characteristics of the separate attributes are particularly visible)
- The complexity of parallel coordinate plots (number of axes) is directly proportional to the number of attributes
- All attributes receive uniform treatment

 Star plots (aka radar or spider plots) have many features in common with parallel coordinate plots

 An attribute value is represented by a point on a coordinate axis

Attribute axes radiate from a common origin



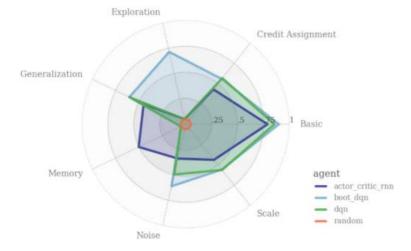
(Spence, 2007)

• For a given object, points are joined by straight lines

Other useful information such as average values or thresholds can be encoded

Properties of star plots:

 Their shape can provide a reasonably rapid appreciation of the attributes of the objects

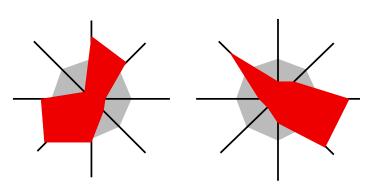


https://syncedreview.com/2019/08/16/deepmind-bsuite-evaluates-reinforcement-learning-agents/

They offer object visibility and are suitable to compare objects

(by visibility it is meant the ability to gain insight pre-attentively; without a great cognitive effort)





Bob's performance Tony's performance

Example: Use visualization techniques to help answer the following questions:

Is there a relation between wanted salary and experience? How many candidates ask for a salary in [30000, 50000] and in [55000, 75000]? How many candidates have an advanced level of English?

	Education	Age	Prof. Experience	English	Wanted salary
#	(MSc/PhD)	(years)	(years)	(Bas/Adv)	(\$\$/year)
1	MSc	22	0	Advanced	36000
2	MSc	23	0	Basic	36000
3	MSc	24	1	Advanced	36000
4	PhD	30	7	Advanced	72000
5	MSc	25	1	Basic	40000
6	PhD	29	5	Advanced	60000
7	MSc	31	7	Advanced	55000
8	MSc	23	0	Advanced	36000
9	MSc	26	2	Intermediate	40000
10	PhD	32	9	Intermediate	65000
11	BSc	30	7	Intermediate	30000
12	PhD	40	17	Advanced	80000
	MSc	28	4	Advanced	40000

the complete table has many more candidates and attributes, but you may test with these

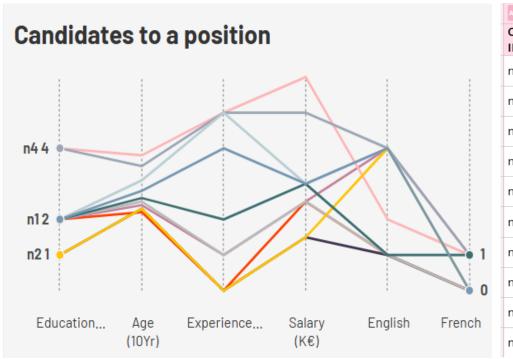
Example:

Parallel coordinates using **Examples | Flourish**

It is necessary to code the ordinal variables:

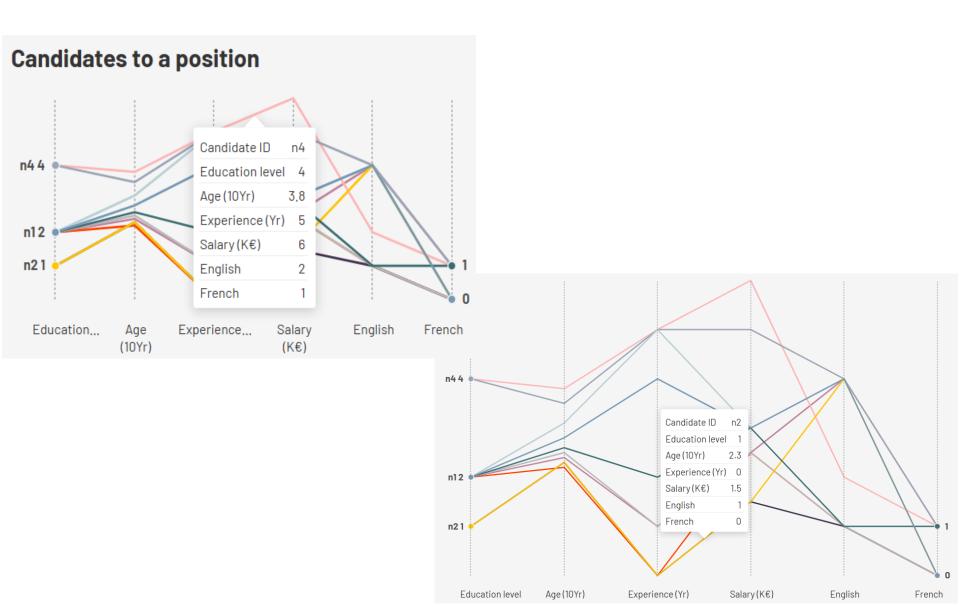
PhD-4 MSc-2 BSc-1 Advanced-4 Intermed-2 Basic-1 Y-1 N-0

Select the variables units to get similar values in all axes:



	p					-
авс А	123 B	123 C	123 D	123 E	123 F	123 G
Candidate ID	Education level	Age (10Yr)	Experience (Yr)	Salary (KEuros)	English	French
n1	2	2.2	0	2.5	1	0
n2	1	2.3	0	1.5	1	0
n3	2	2.4	1	2.5	4	1
n4	4	3.8	5	6	2	1
n5	2	2.5	1	2.5	1	0
n6	4	3.5	5	5	4	1
n7	2	3.1	5	3	4	0
n8	1	2.3	0	1.5	4	0
n9	2	2.6	2	3	1	1
n10	2	2.8	4	3	4	0

It is possible to have "details on demand" by user interaction



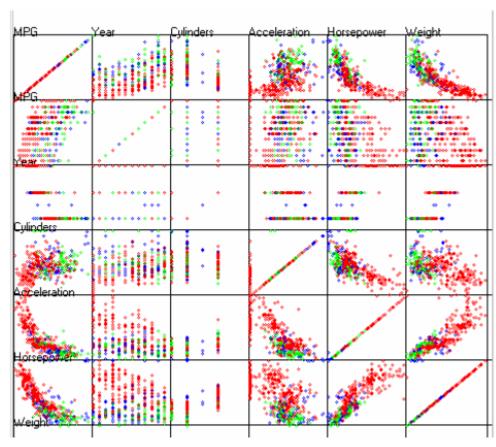
The scatterplot matrix (SPLOM) is applicable to higher dimensions

However, as the number of attributes increase, the number of different pairs

of attributes increases rapidly:

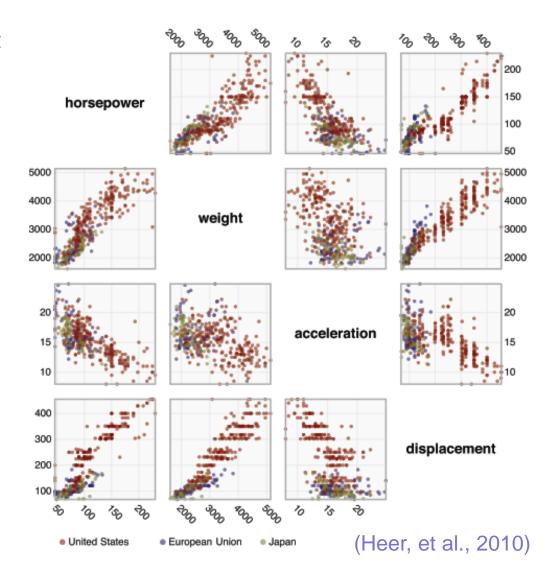
- 2 attributes -> 1 scatterplot
- 3 attributes -> 3 scatterplots
- 4 attributes -> 6 scatterplots

We may try to reduce the number of dimensions keeping the more relevant

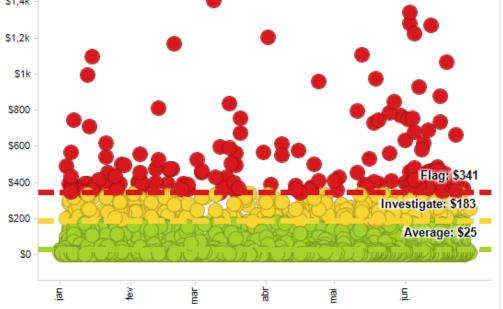


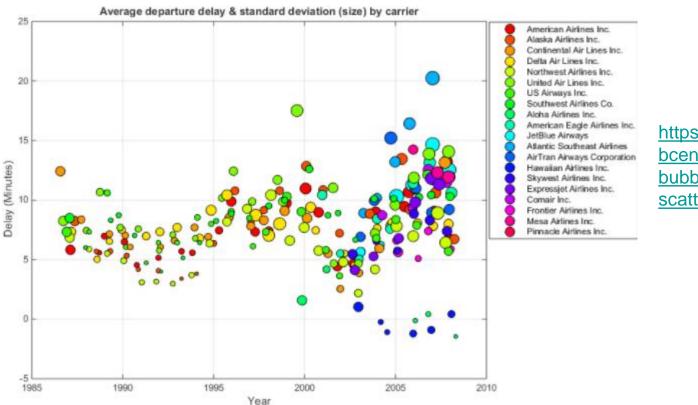
Scatterplot matrix for 6 attributes of a car dataset

 Another example of Scatterplot matrix for a car dataset



 A single scatterplot can be used together with other encoding techniques to represent data of higher dimension





https://www.mathworks.com/matla bcentral/fileexchange/48005bubbleplot-multidimensionalscatter-plots

A scatterplot representing 5 variables

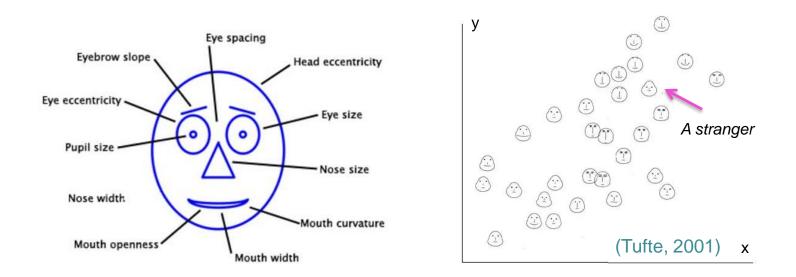
Hans Rosling's 200 Countries, 200 Years, 4 Minutes: 120 000 values

Income (x), Age expectancy (y), Time (t), Continent (color), Population (size of circle)



https://www.youtube.com/watch?v=jbkSRLYSojo

Icons (aka glyphs) represent a number of attributes qualitatively or quantitatively

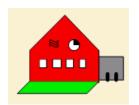


Chernoff Faces allow attribute values to be encoded in the features of cartoon faces

They were originally used to study geological samples, each characterized by 18 attributes

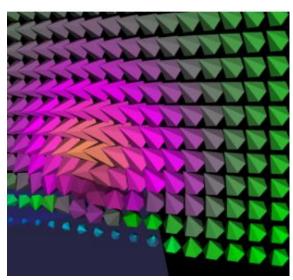
<u>Chernoff face - Wikipedia</u>

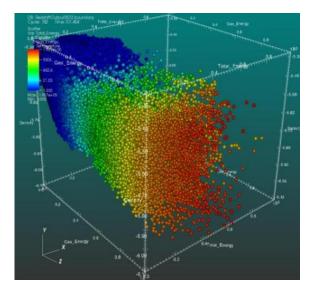
- Two examples of metaphorical icons:
 - with direct relation between icon and object (house icon)



- no direct relation between facial features and attributes they represent (Chernoff faces)

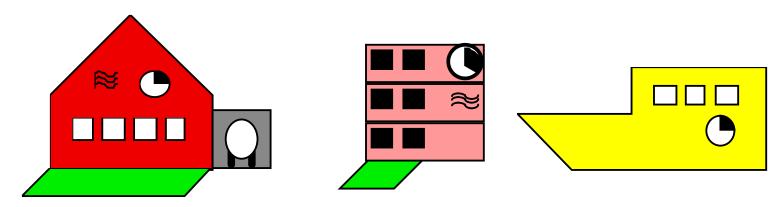
Examples in SciVis





Glyph (data visualization) -Wikipedia

Multidimensional icons representing eight attributes of a dwelling



house £400,000 garage central heating four bedrooms good repair large garden Victoria 15 mins flat £300,000 no garage central heating two bedrooms poor repair small garden Victoria 20 mins houseboat £200,000 no garage no central heating three bedrooms good repair no garden Victoria 15 mins

Textual descriptions of the dwellings represented by the multidimensional icons (Spence, 2007)

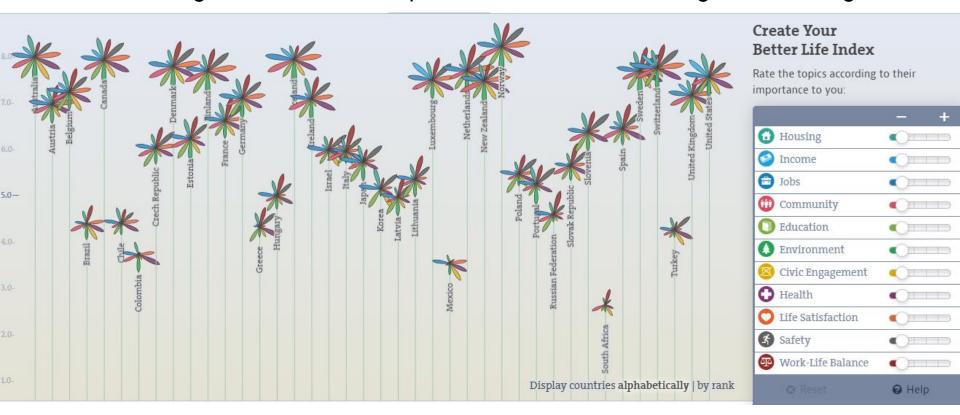
Lisbon

9.8 7.1 Access to 8.3 Environment 2.8 Civic 6.4 7.9 6.4 7.2

Glyph chart example:

Based on a shape being the main artifact of representation

The physical properties of the shape represent different categorical variables sized according to the associated quantitative value and distinguished through color



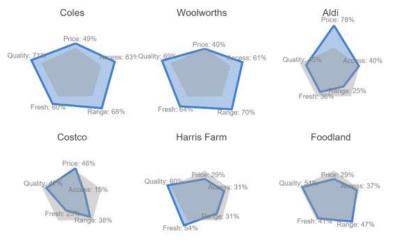
OECD Better Life Index | OECD (Kirk, 2012)

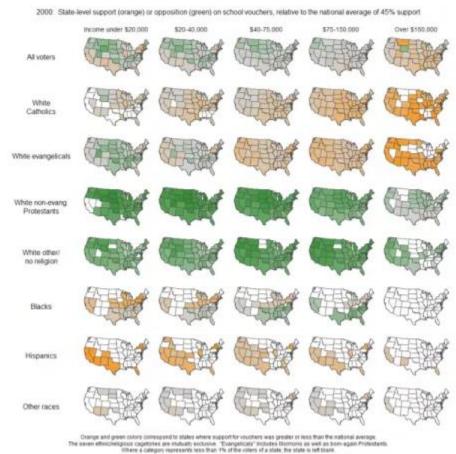
Useful arrangement of several charts

Small multiples:

arrangement approach that facilitates efficient and effective comparisons

(Kirk, 2012)





Dashboards

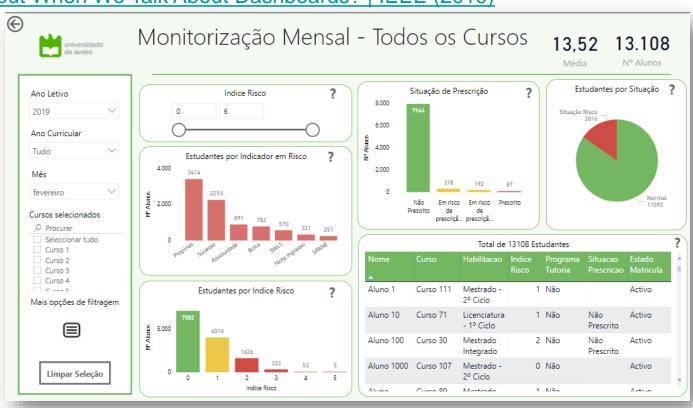
Visual display summarizing a dataset providing information at-a-glance (e.g. KPIs)

"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance" (Few, 2004)

https://www.nngroup.com/articles/dashboards-preattentive/

What Do We Talk About When We Talk About Dashboards? | IEEE (2019)

Prototype: "Portal dos indicadores, UA"

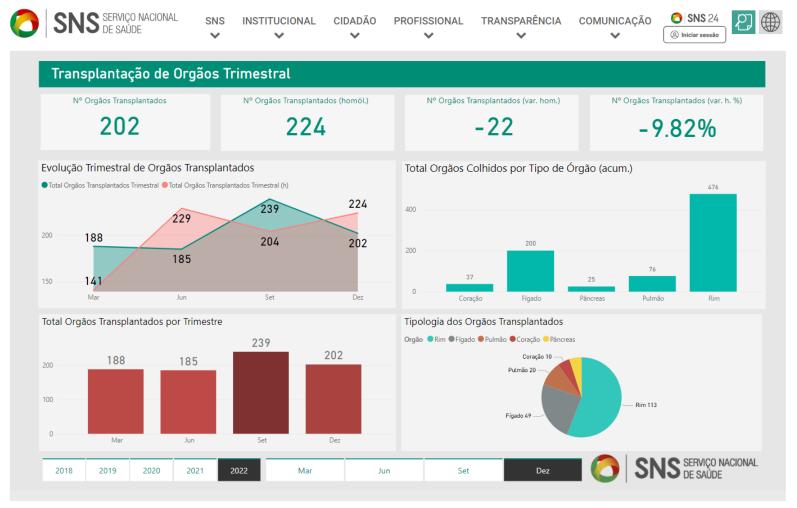




As seen by people with green-blind deuteranopia; does it work? https://www.color-blindness.com/coblis-color-blindness-simulator/

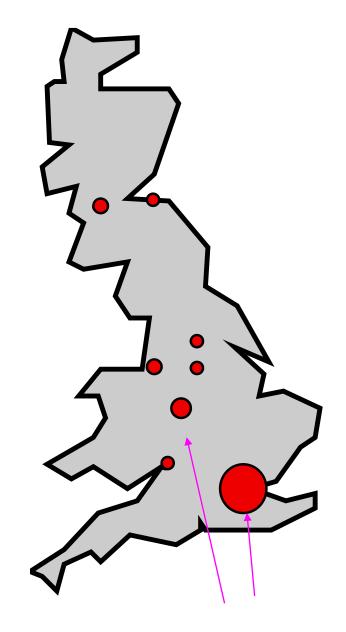
Dashboards

Visual display summarizing the Portuguese transplant situation



Some more examples on how humans see...

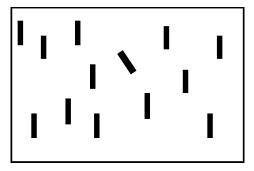
Population of major cities in England, Wales and Scotland. Circle area is proportional to population. (Spence, 2007)



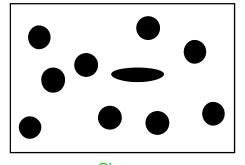
Things that "pop-out"

Pre-attentive processing: Things that "pop out"

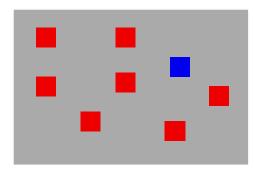
"We can do certain things to symbols to make it much more likely that they will be visually identified even after a very brief exposure" (Ware, 2004)



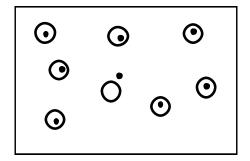
Orientation



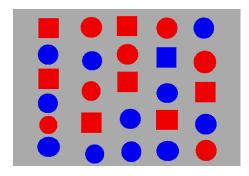
Shape



Colour



Enclosure



Where is the blue square?

(Spence, 2007)

But we should be careful...

Color is a strong visual cue

How many cherries?

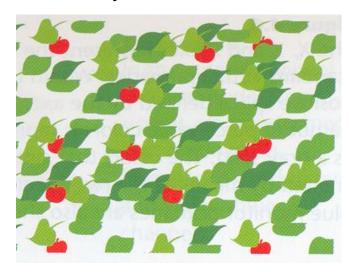


(Ware, 2004)

Color is a strong visual cue: it may help users perform their tasks

If correctly used

How many cherries?



Color may support users in many tasks!

Or not ...

Using color is complex as color perception is complex...

Color scales

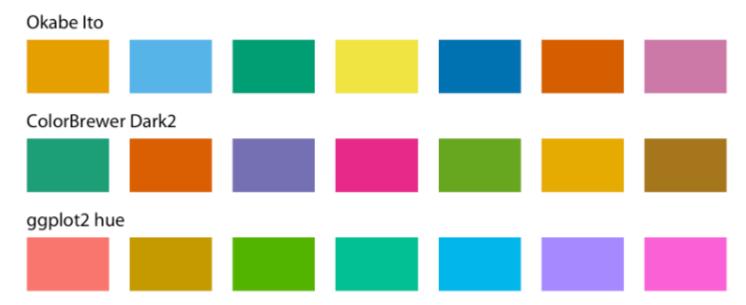
Fundamental use cases for color in visualization:

- distinguish groups of data (qualitative color scales)
- represent data values (sequential color scales)
- Highlight (accent color scales)

The types of colors and the way in which they should be used are quite different (some examples next)

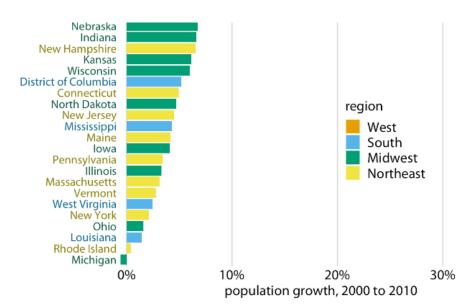
Fundamentals of Data Visualization

Qualitative color scales

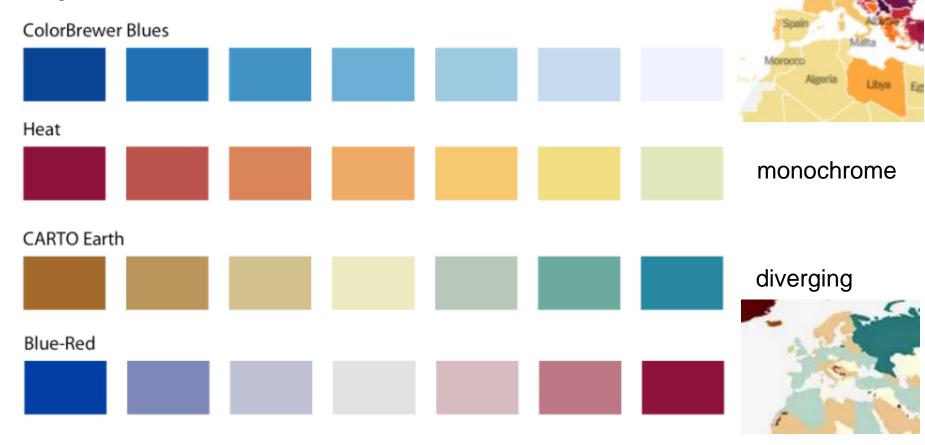


Colors are chosen to be clearly distinct and not stand out relative to others

Fundamentals of Data Visualization



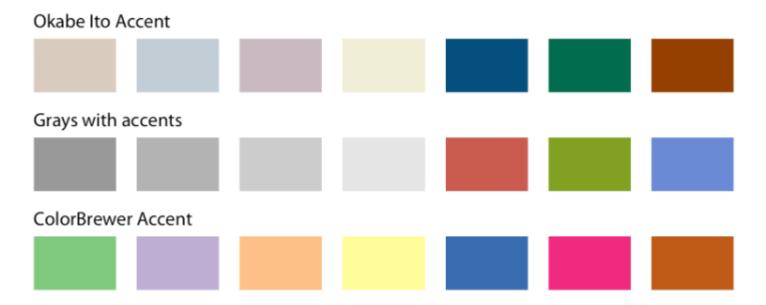
Sequential color scales



Colors should indicate which values are larger or smaller, and how distant two specific values are from each other, may be monochrome, diverging ...

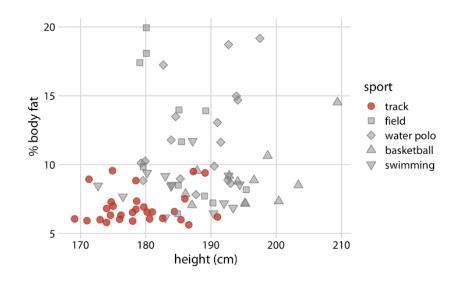
Fundamentals of Data Visualization

Accent color scales



These scales contain a set of subdued colors and a matching set of stronger, darker, more saturated colors

https://clauswilke.com/dataviz/



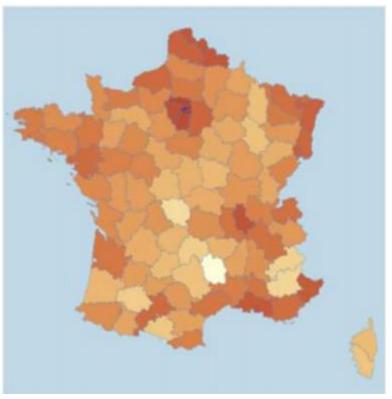


Color may not help or even make it more difficult!

Α

1968 100-150 90-99 80-89 70-79 60-69 50-59 40-49 30-39 0-29

В



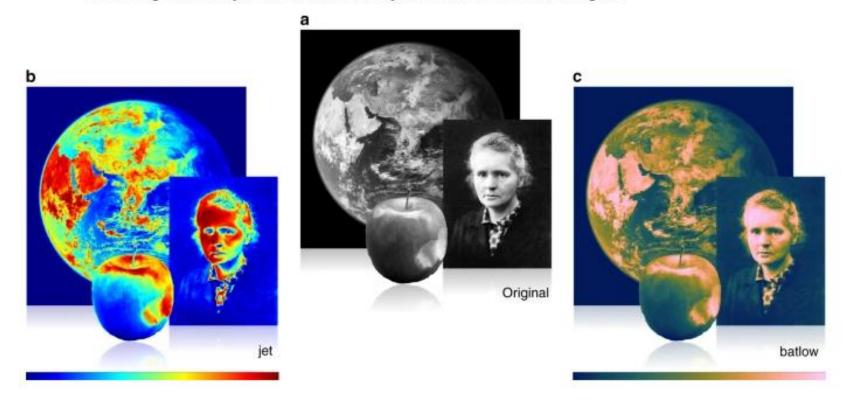
A- no preattentive association that allows efficiently determine the values (Kirk, 2012)

B- a single hue and a sequential color scheme representing values in an immediately understandable way

The misuse of colour in science communication



The superiority of scientifically derived colour maps.



The misuse of colour in science communication | Nature Communications

Remember:

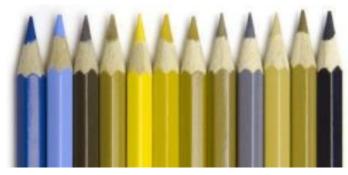
- Not everyone sees color:
- The most common form of color blindness is deuteranopia ("daltonism")
- There are color blindness simulators:
 Try this one:



<u>Coblis — Color Blindness Simulator —</u> Colblindor



Normal vision



Deuteranopia



Tritanopia

Home - Colour Blind Awareness

Some rules to use color in visualization

- Make it right is Black and White
- Less is more, or less is better
- Avoid using fully saturated colors in large numbers and in large areas
- Use fully saturated colors only when you want to highlight
- Use blue in larger areas and not in small areas
- Mind colorblindness and use simulators to test your designs
- When adopting color to distinguish, use colors that are easily distinct from each other

88

Color models

Are used to measure and produce color

The basic H/W oriented models are RGB (additive) and CMY (subtractive)



RGB: emitting/screens/projectors



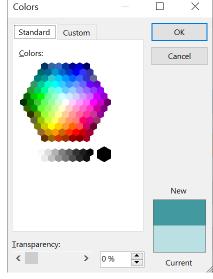
CMYK: reflecting/printers

Are not related to human perception, but to the physical process

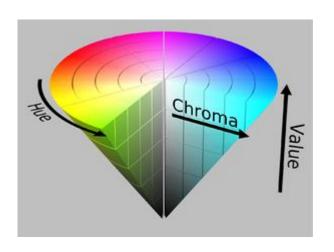
Should not be used directly to produce color scales

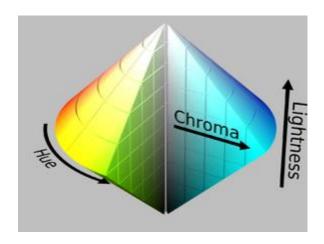
There are color models (<u>HSV and HSL</u>) based on perceptual variables:

hue (violet, blue, green, yellow, red ...) saturation (amount of white) value/brightness



used when we intuitively describe colors (e.g. light blue or dark green)





HSV HLS

Are more adequate for users to specify color

Other models (perceptually corrected) are better to specify quantitative color 90 scales (e.g., l.*, a*, b*, color model)

Bibliography

- Kirk, A., Data Visualization: a successful design process, Packt Publishing, 2012
- Kirk, A., Data Visualization: a handbook for data driven design *, Sage, 2016
- Mazza, R., Introduction to Information Visualization, Springer, 2009
- Munzner, T., Visualization Analysis and Design *, A K Peters/CRC Press, 2014
- Spence, R., Information Visualization, Design for Interaction, 2nd ed., Prentice Hall, 2007
- Spence, R., Information Visualization, an Introduction, 3rd ed., Springer, 2014
- Ware, C., Information Visualization *, 3rd ed., Morgan Kaufmann, 2013
- Wilke, C., Fundamental of Data Visualization, 2019
- Tufte, E., The Visual Display of Quantitative Data, 2nd ed, Graphics Press, 2001
- Books with * and other interesting books at:
 https://learning.oreilly.com/playlists/74bfec5e-4346-48ff-82b4-657fda6922b6