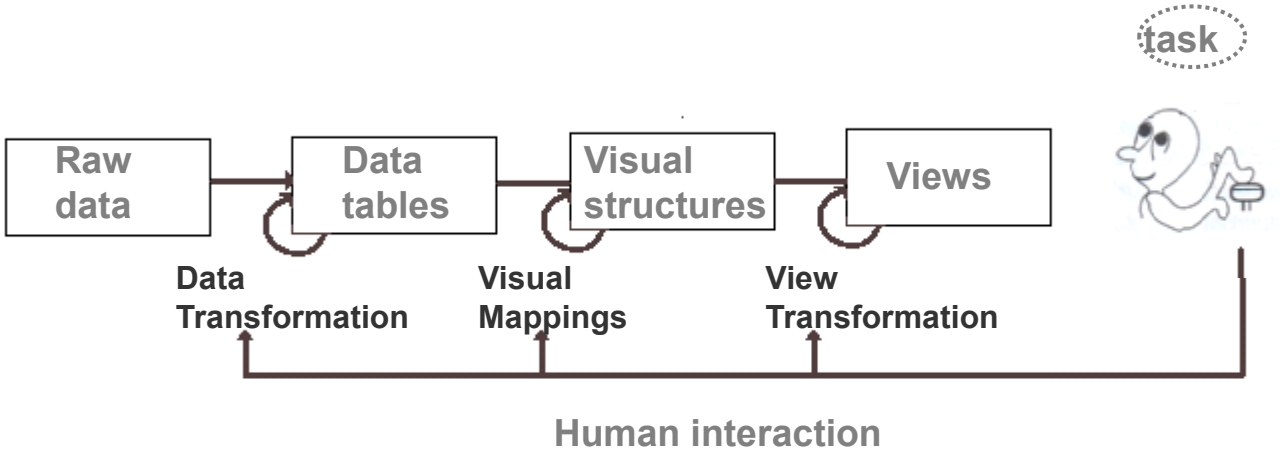
Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

An Introduction to Data and Information Visualization

Representation, Presentation & Interaction

Beatriz Sousa Santos
October/ 2021

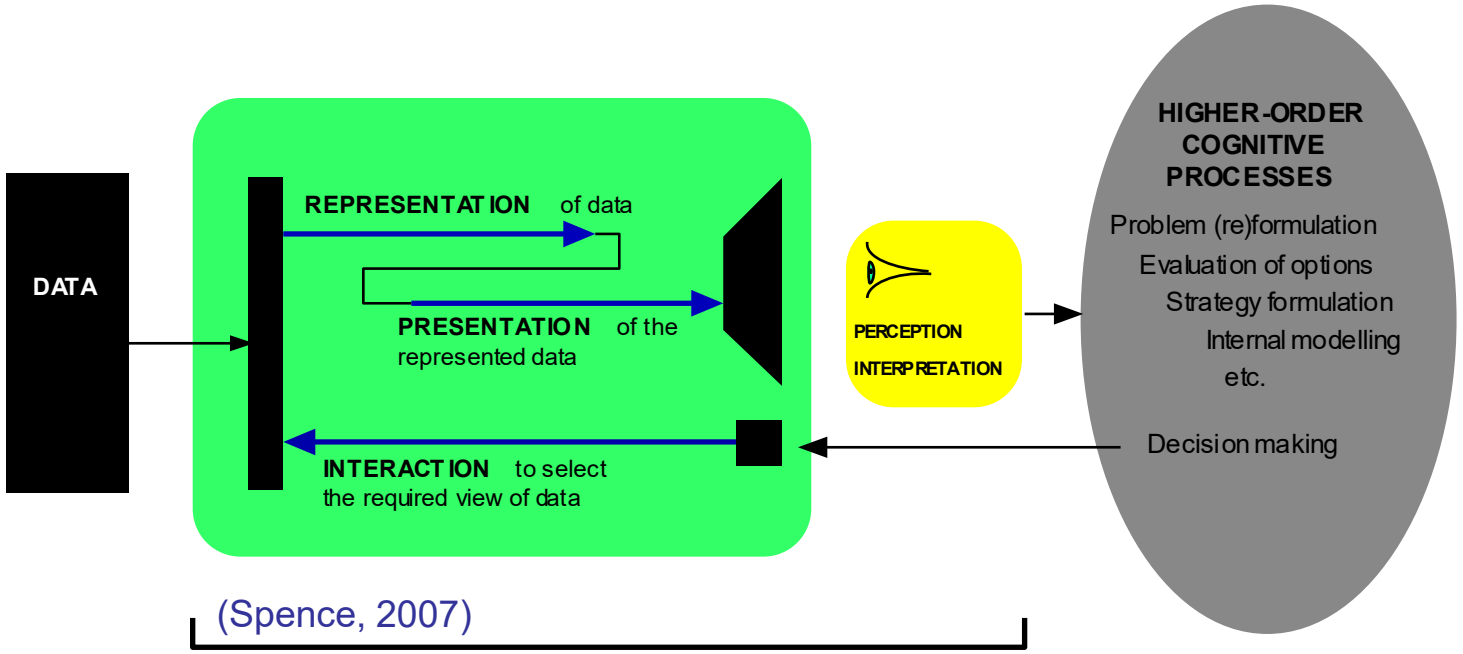
Information Visualization Reference Model



Visualization can be described as the mapping of data to visual form supporting human interaction for visual sense making (Card et al., 1999)

Visualization is a **Human in the loop process**! -> which calls for specific **methods**

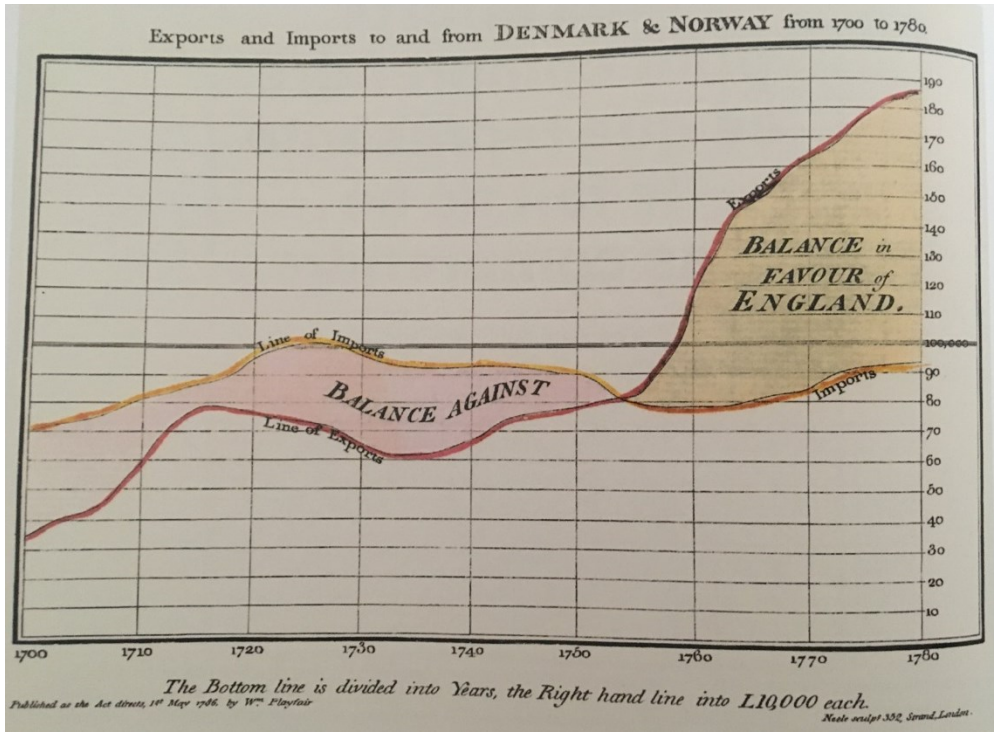
The process of visualization
a way to organize visualization techniques



Interaction with data governed by high-order cognitive processes involves:

- Representation
- Presentation
- Interaction

Representation – Encoding value
(some common Visualization techniques)



nature of the problem
- communicate
- explore
- confirm

nature of the data to represent
- categorical (nominal)
- ordinal
- quantitative

number of attributes
- univariate
- bivariate
- trivariate
- multivariate

data structures
- linear
- temporal
- spatial or geographical
- hierarchical
- network

To increase the “consideration space” in the design process it is necessary to know a lot of representation techniques:

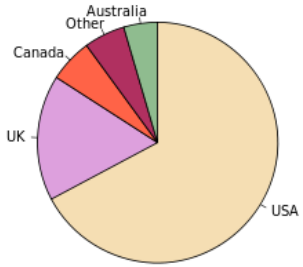
We study techniques to represent value organized according the number of attributes

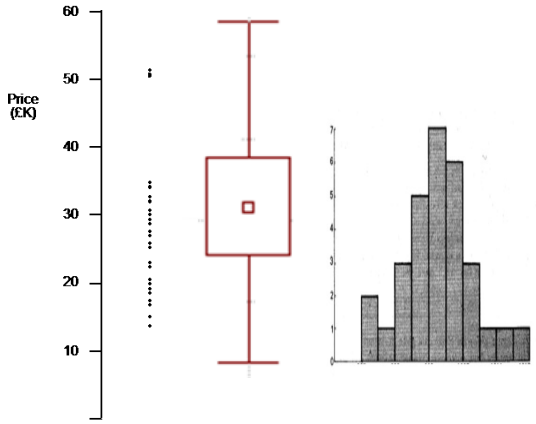
And techniques to represent relation (trees and networks)

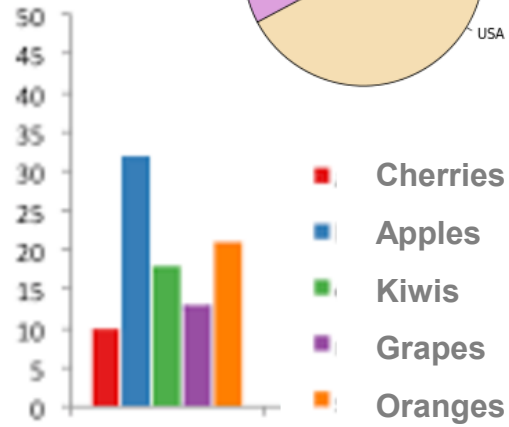
,

Common Visualization Techniques for univariate, bivariate and trivariate data

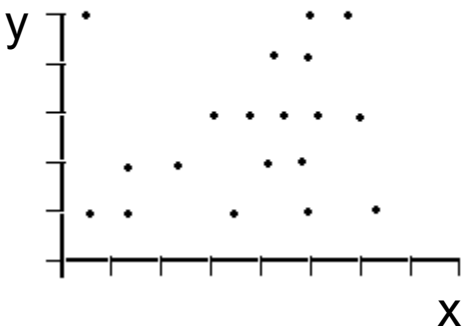
Univariate data
- dot plot
- box plot
- bar chart
- histogram
- pie chart
- …

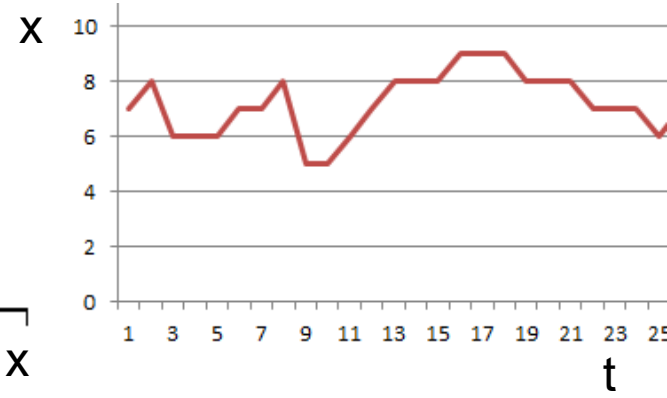




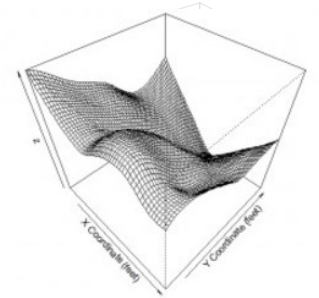


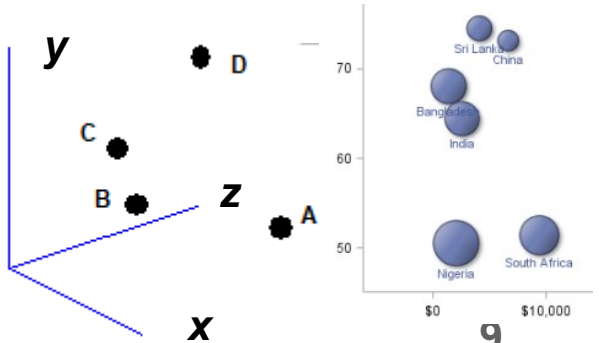
Bivariate data
- scatter plot
- line plot
- time series
- …





Trivariate data
- surface plot
- 3D representation
- bubble plot
- …



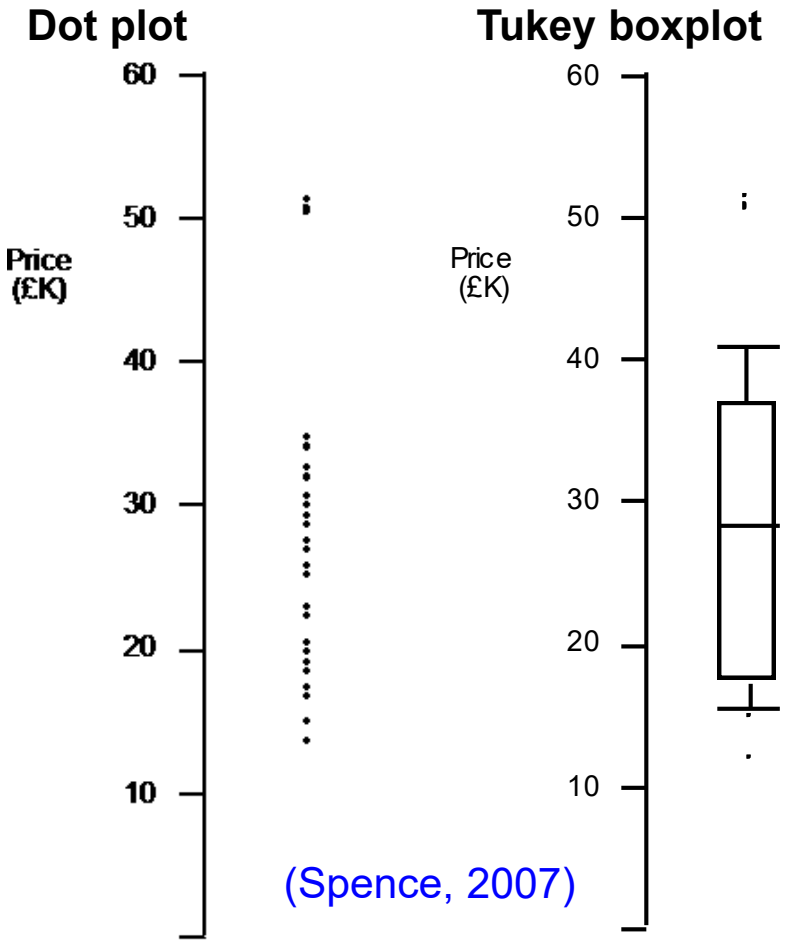


Representing univariate data

- A common situation consists in representing a **set of values**
- Well established techniques exist
- But new ones can be invented!

Example:

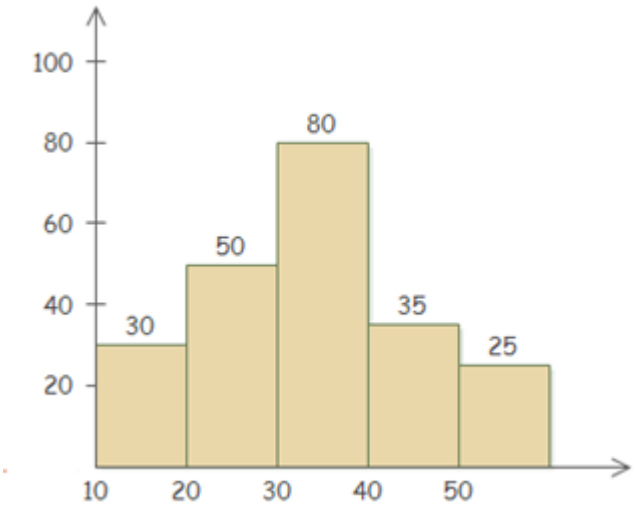
Price for a number of cars:
- dots on a linear scale
- box plot
(that will answer many questions:
median value, outliers,...)

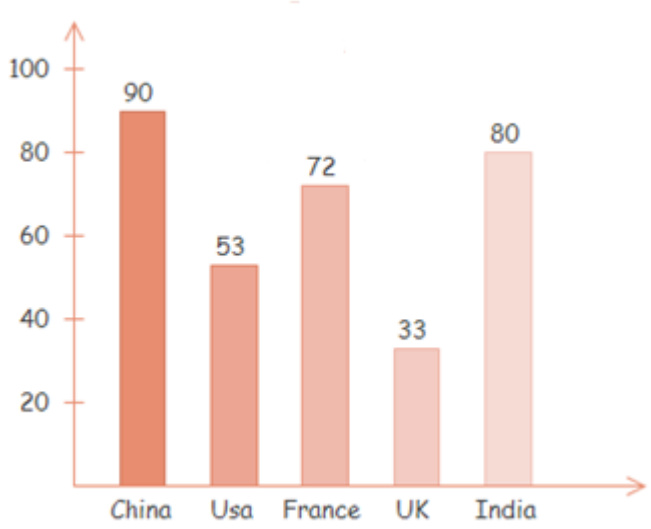


(Spence, 2007)

Histogram *vs* Bar chart

- Histogram or bar char?





Quantitative variable

Nominal variable

What is the difference?

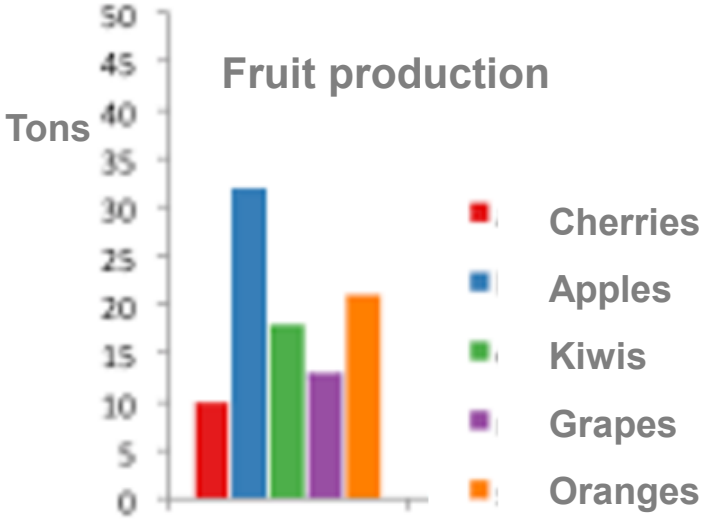
Simple (and common) representations of data

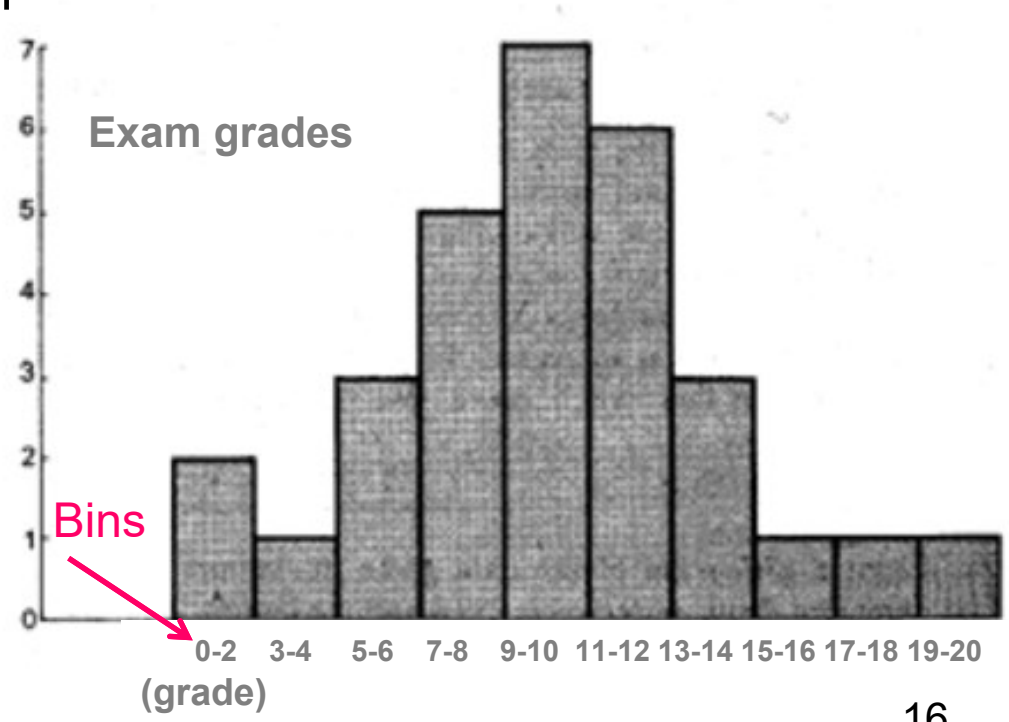
- Two common techniques not to be confused !

Histogram ⇒ represents a distribution of numerical data

Bar chart ⇒ represents the number of occurrences of a categorical/ordinal data

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

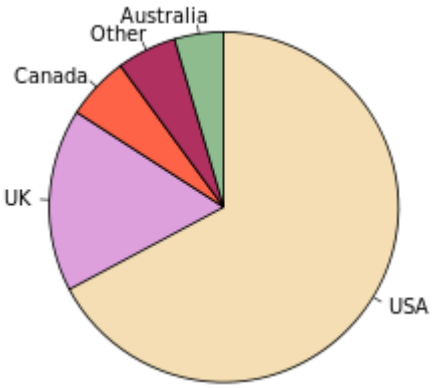




Another simple (and too common) representation

- Pie Chart

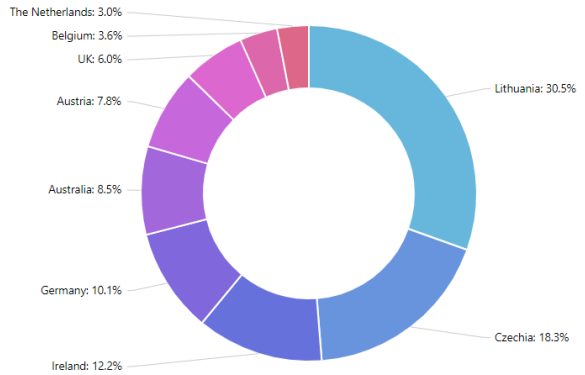
Represents numerical proportion, **parts of an whole**



Native English speaking population

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

Variations of pie charts:



Are much controversial:

many experts recommend avoiding them

http://www.perceptualedge.com/articles/08-21-07.pdf

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

- 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!

https://eagereyes.org/techniques/pie-charts

Representing bivariate data

- The **scatterplot** is the conventional representation

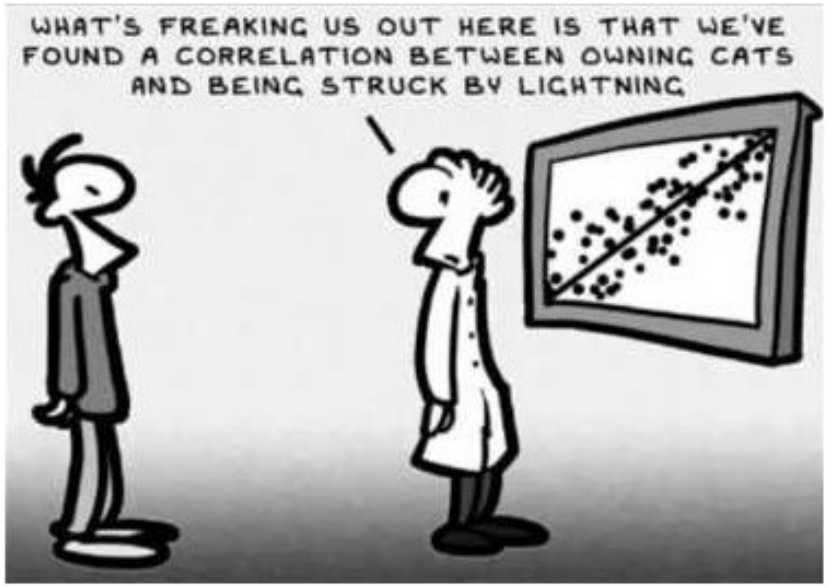
Each observation is represented by a point on a two dimensional space

The axes are associated with these two attributes

This representation affords awareness of:

- general trends
- local trade-offs
- outliers

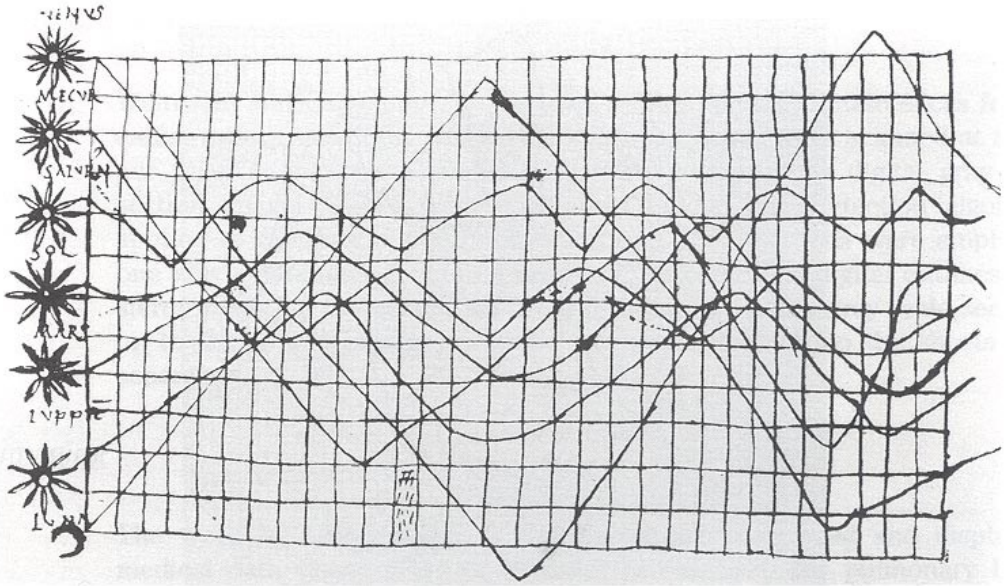
Correlation is not causation



Representing bivariate data

The **line chart**

One of the oldest known and ubiquitous Visualizations

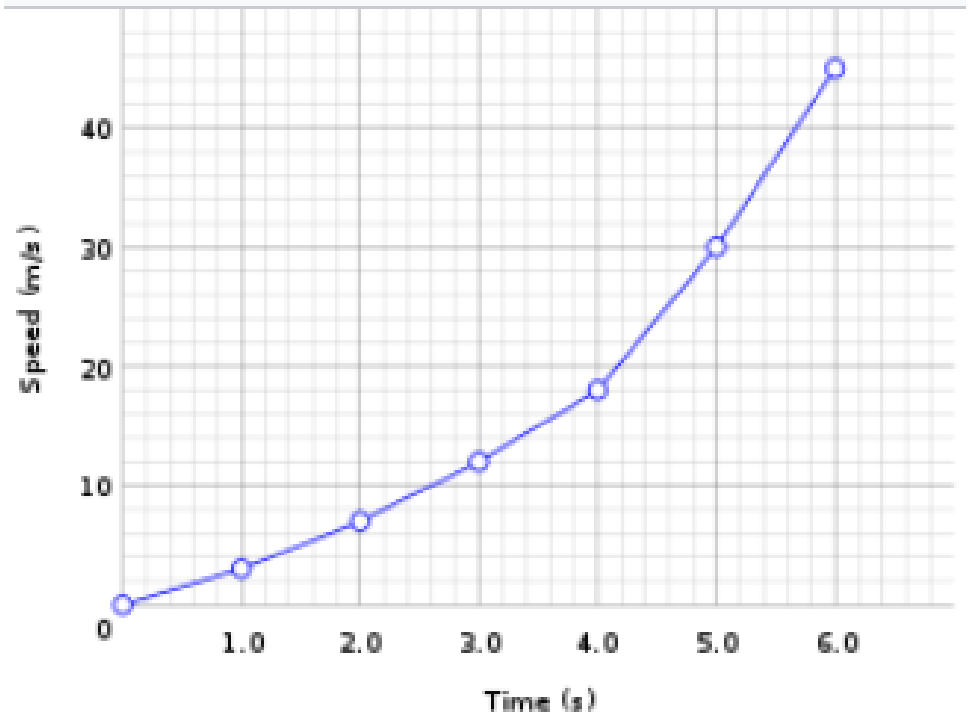


Inclination of orbits along the time - Xth century (Tufte,1983)

- A **line chart** or **line plot** or **line graph** or **curve chart** displays information 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



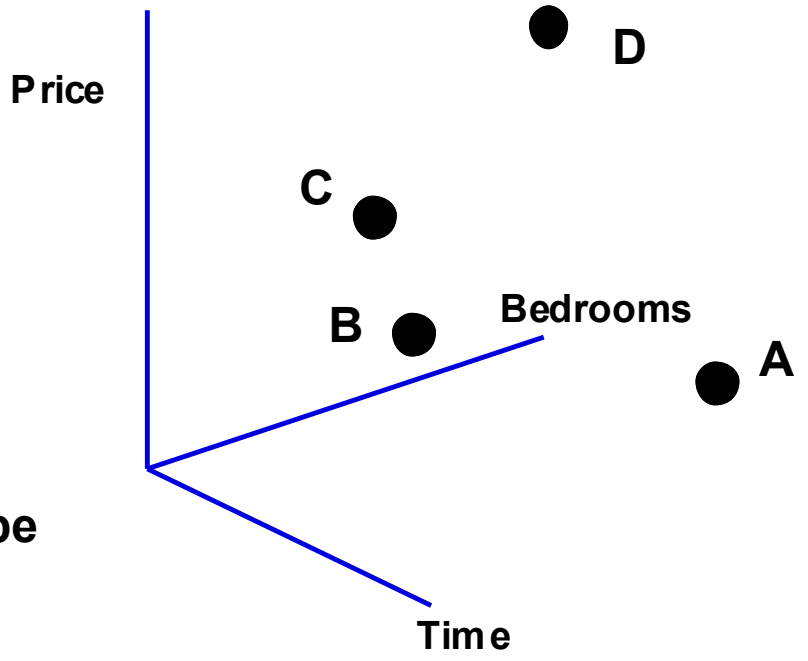
Representing trivariate data

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

- However, these representations can be ambiguous …

- How can we improve this issue?

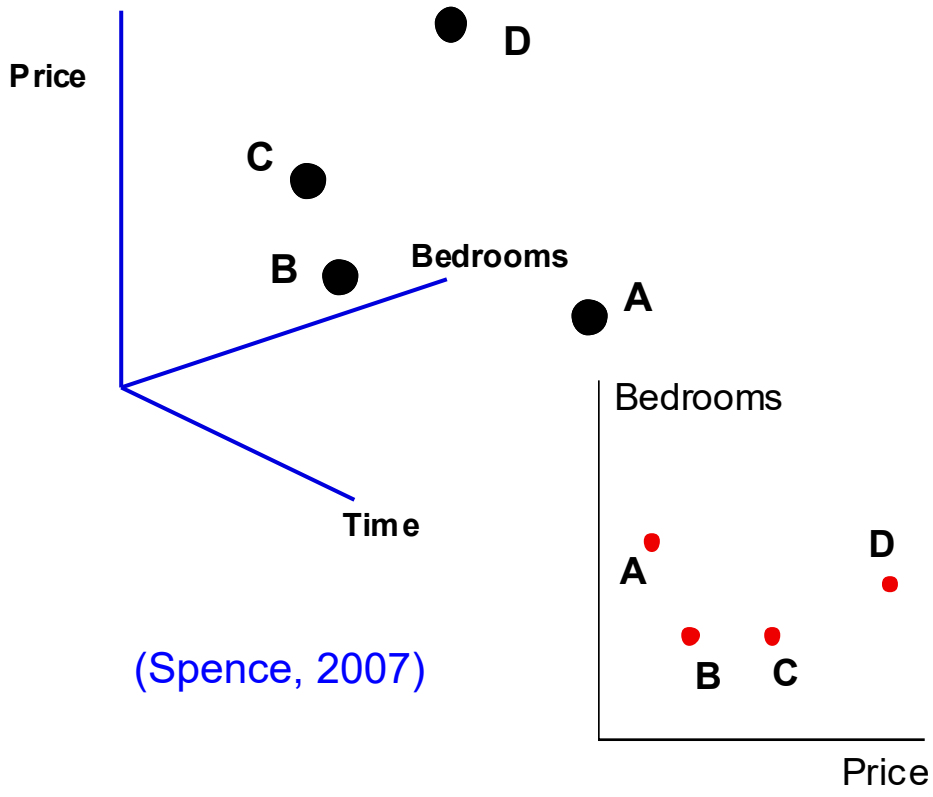
"for 3D to be useful, you' ve got to be able to move it" (Spence, 2007)



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

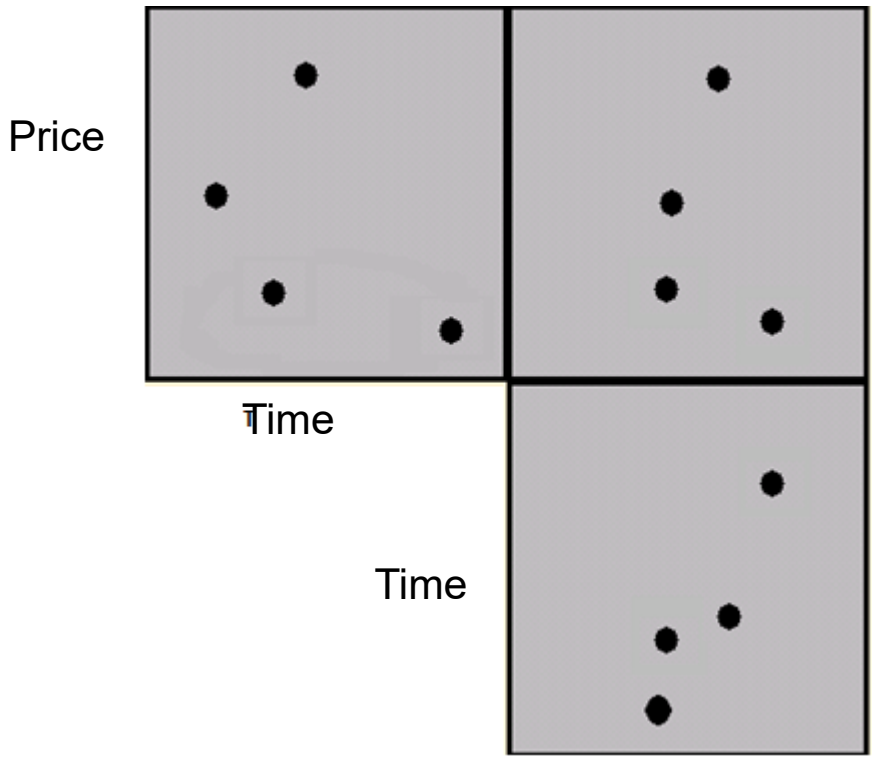
- An alternative representation for trivariate (and hypervariate) data is a structure formed from the **three possible 2D views** of the data

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



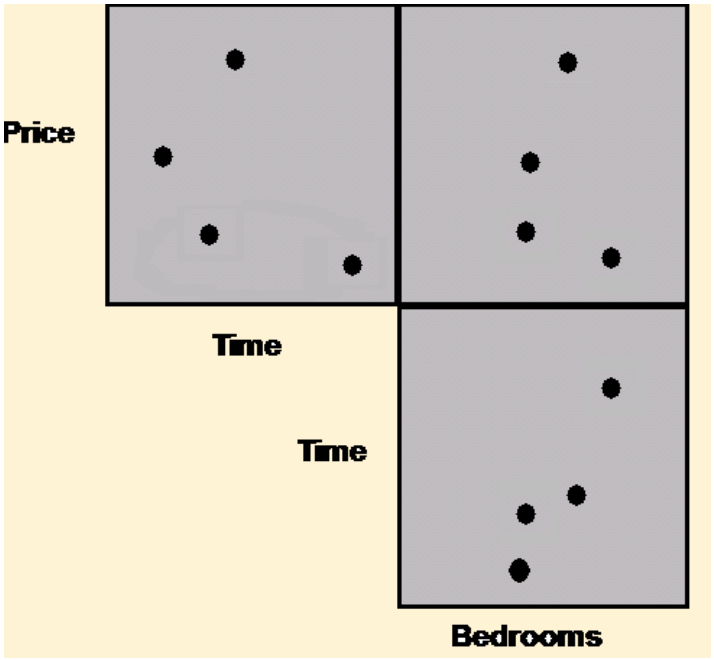
(Spence, 2007)

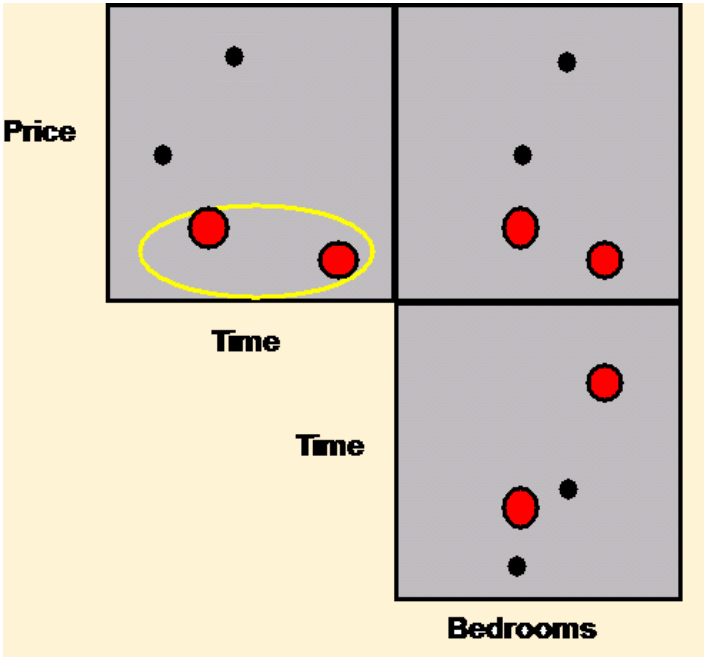
Scatterplot matrix



- 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

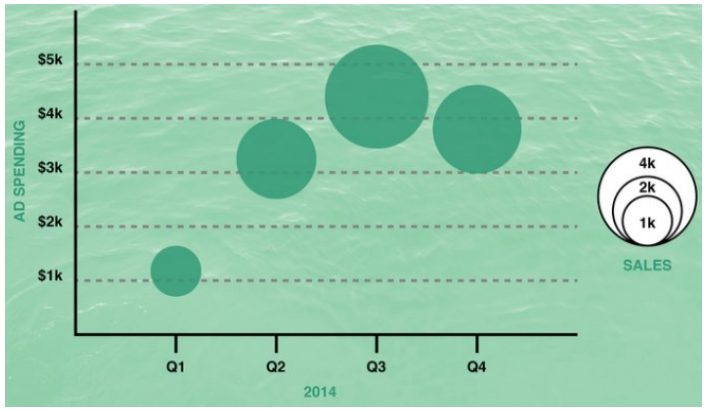




The highlighting of houses in one plane is brushed into the remaining planes (Spence, 2007)

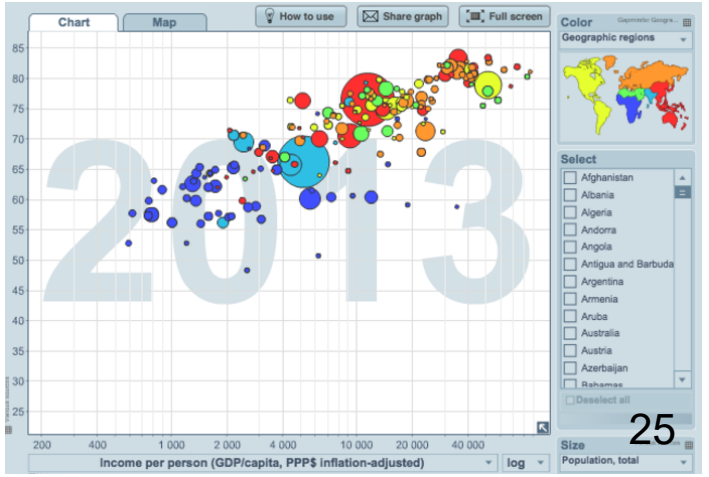
Other Simple (and common) representation of 3D data

- 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 through its size (radius or area?)



- Mapping the variable to size must be done carefully. The interpretation of size may be ambiguous

- Representing one more dimension through color

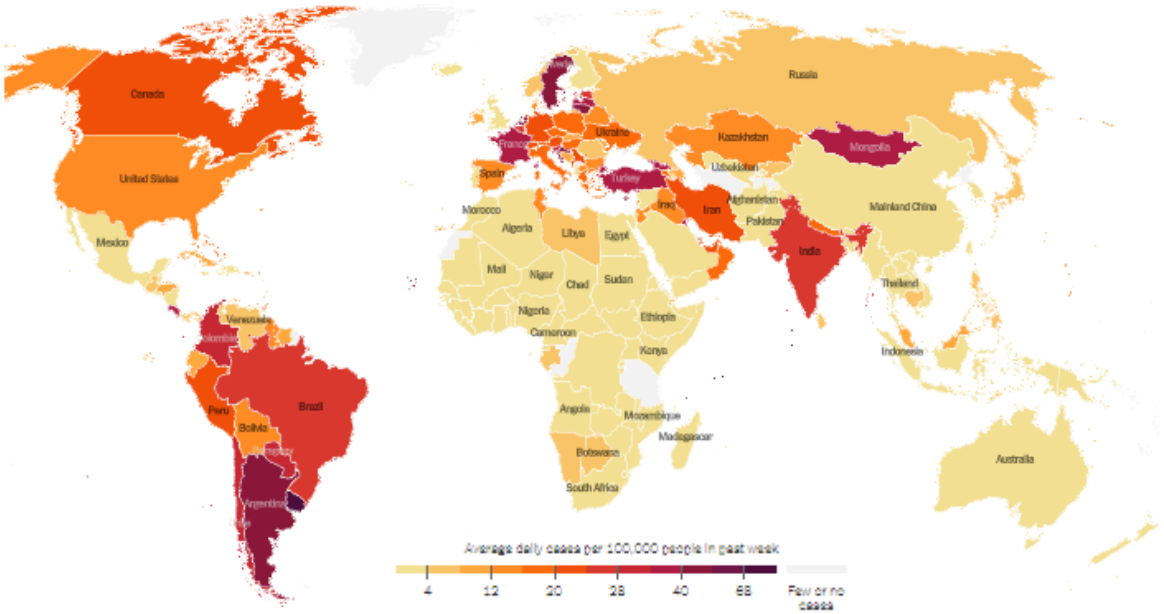


https://visage.co/data-visualization-101-bubble-charts/

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

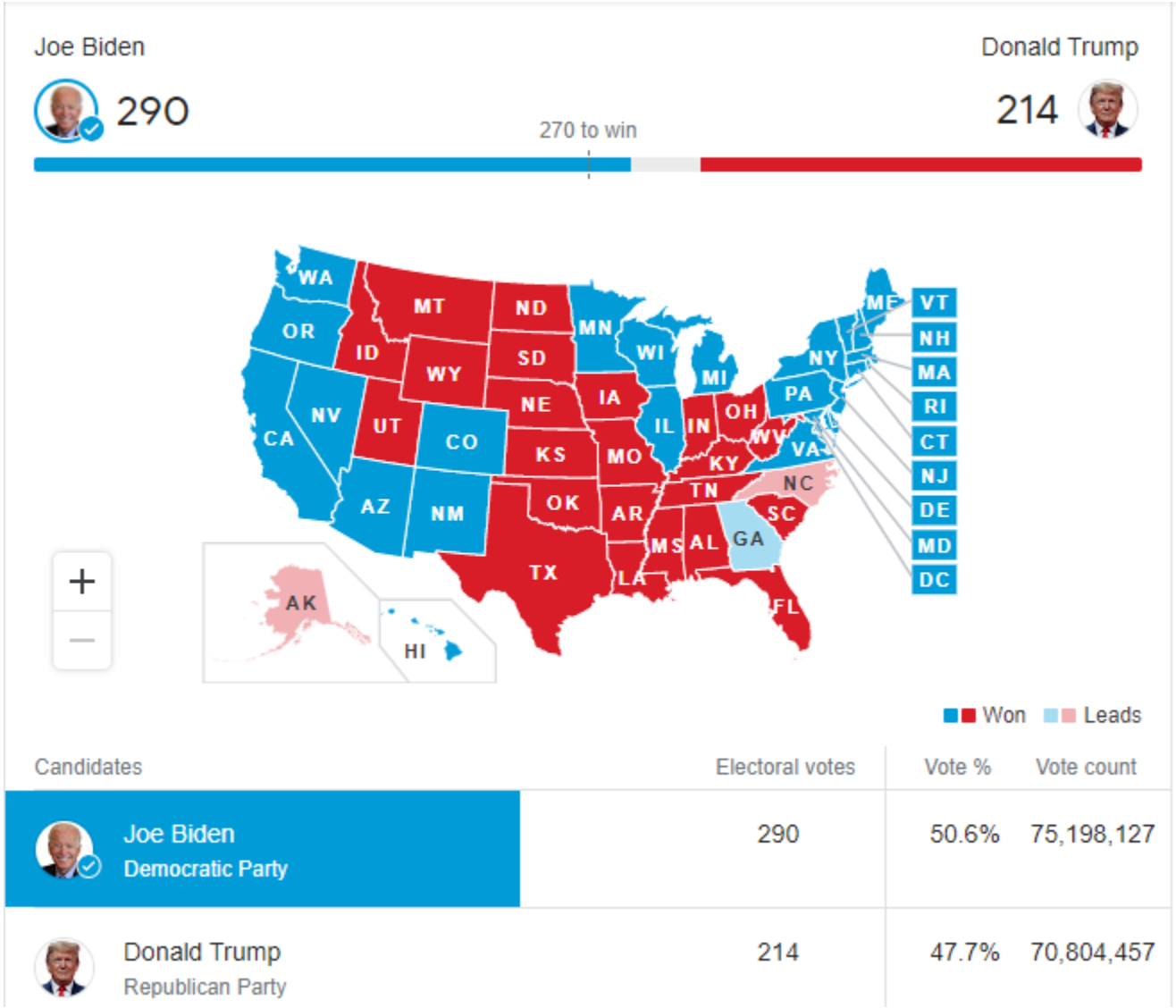
Choropleth maps - A standard approach to communicating aggregated data by geographical areas using color encoding of the geographic area

They require some care: what are the possible issues?



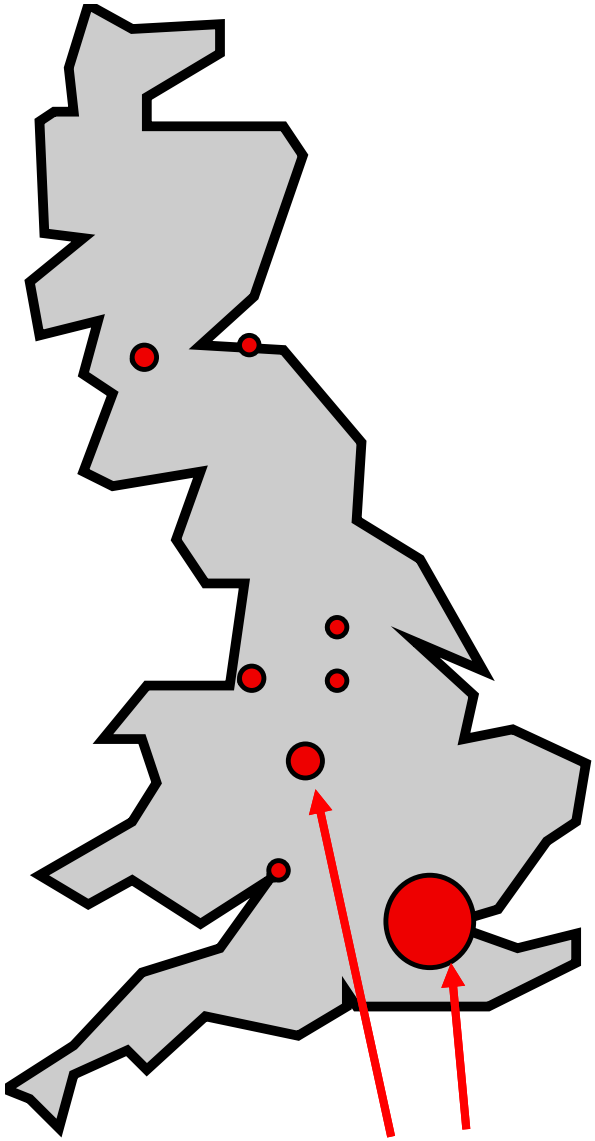
https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html

Visualizations of the US 2020 Election
(a more complete representation: choropleth + bar + text)



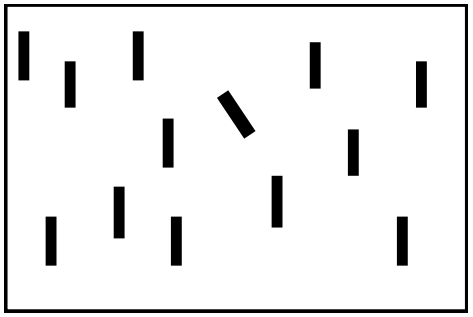
A special category of trivariate data: maps (latitude and longitude + a value)

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

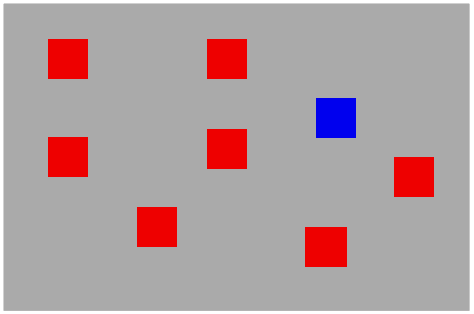


Pre-attentive processing: Things that “pop out”

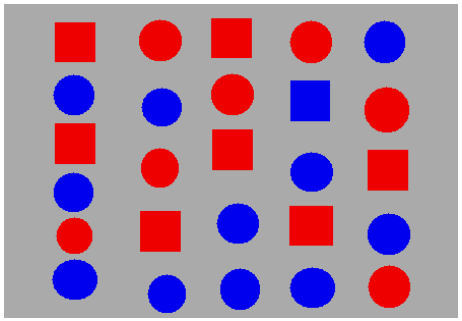
“We can do certain things to symbols to make it much more likely that they will be visually identified even after a very brief exposure” (Ware, 2004)



Orientation

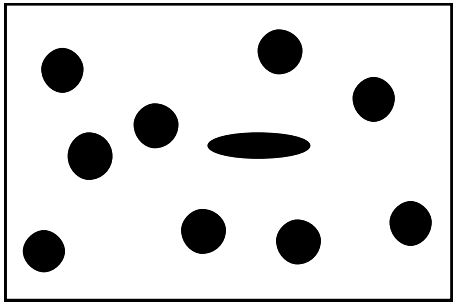


Colour

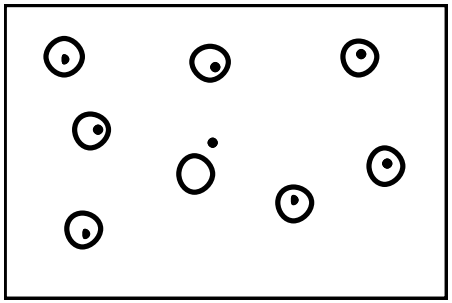


Where is the blue square?

(Spence, 2007)



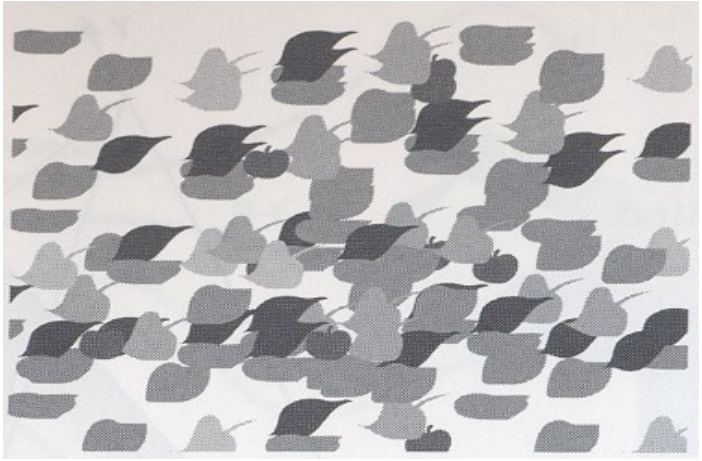
Shape



Enclosure

Color is a strong visual cue

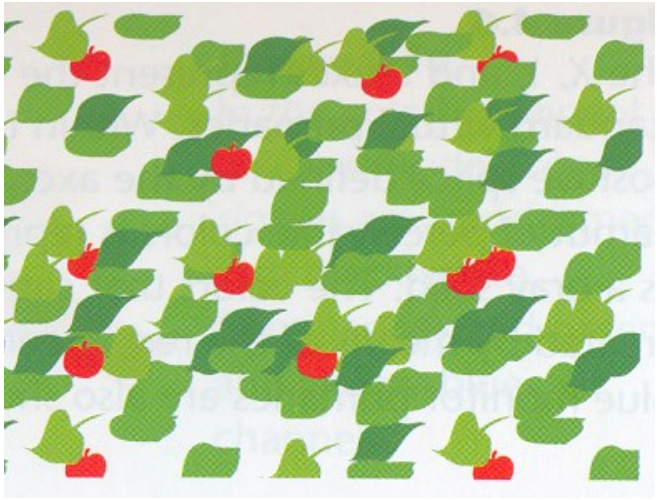
- How many cherries?



(Ware, 2004)

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

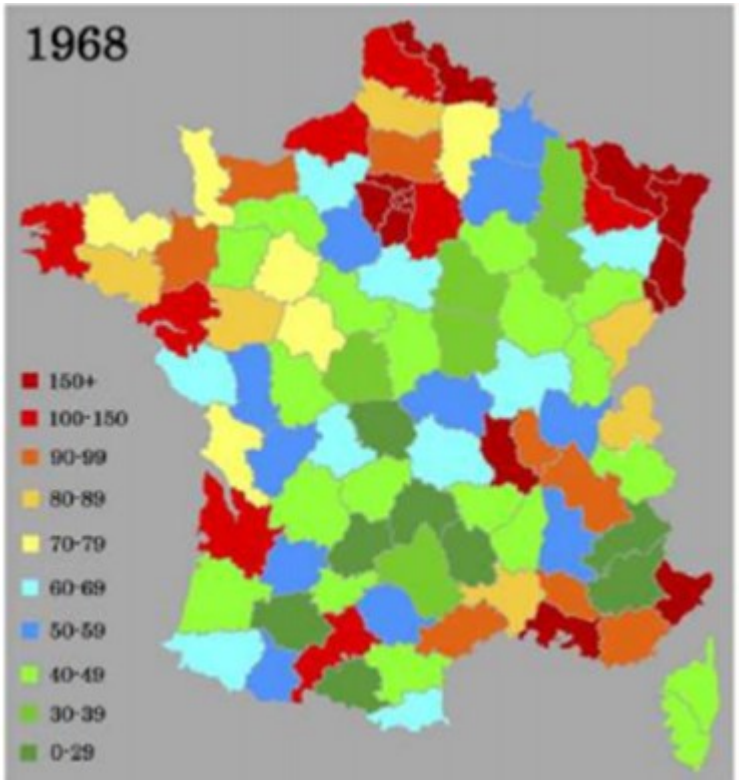
How many cherries?



Color may support users in many tasks!
Or not ...

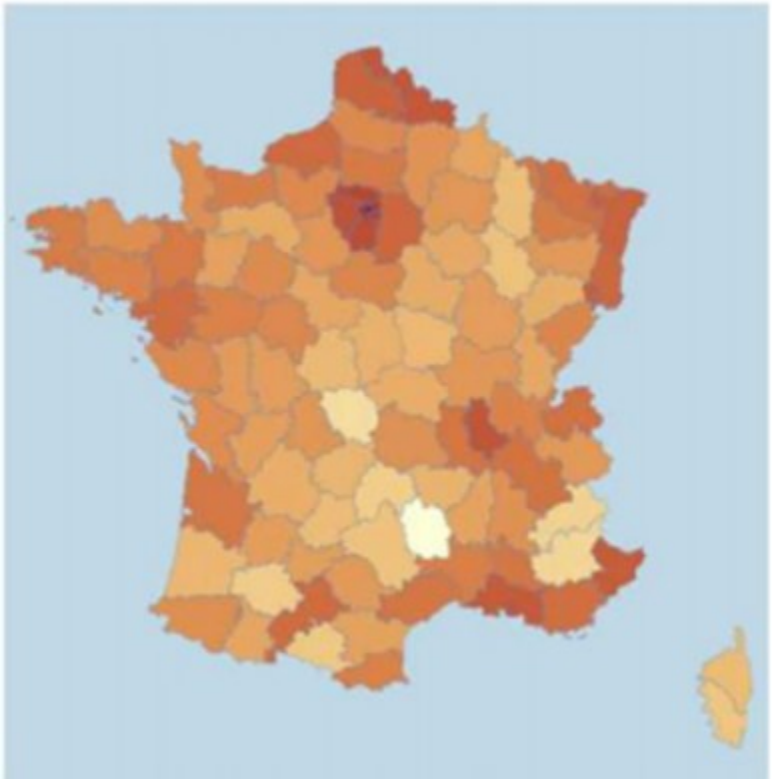
Color may not help or even make it more difficult!

A



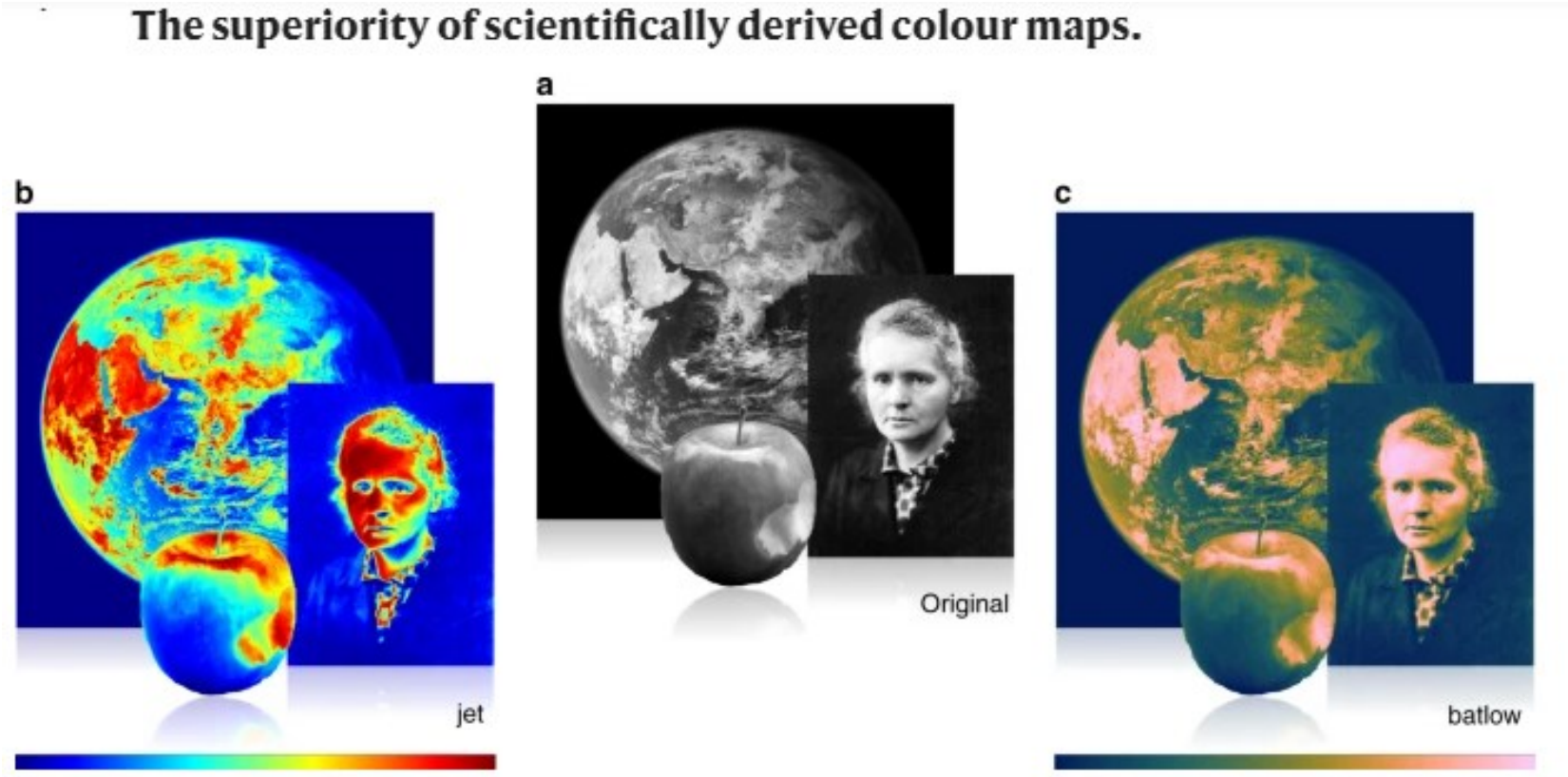
A- no preattentive association that allows efficiently determine the values (Kirk, 2012)

B



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

The misuse of colour in science communication



https://www.nature.com/articles/s41467-020-19160-7

Representing Hypervariate (or multivariate) data

- Many real problems are of high 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)

Example: Hans Rosling's 200 Countries,
200 Years, 4 Minutes
4 variables + time

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

Representing multivariate data using a simple scatterplot

Example: Hans Rosling's 200 Countries, 200 Years, 4 Minutes

- Which variables? How are they visually encoded?

- How is time represented?

Advantages?
Disadvantages?

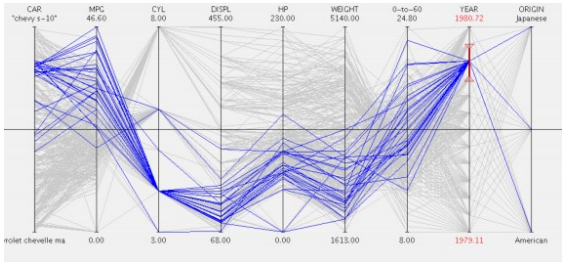
When is it interesting?

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

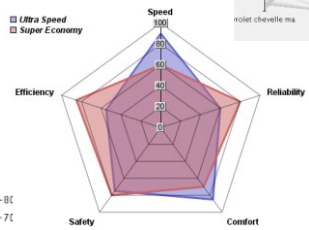
Representing hypervariate data

Coordinate plots

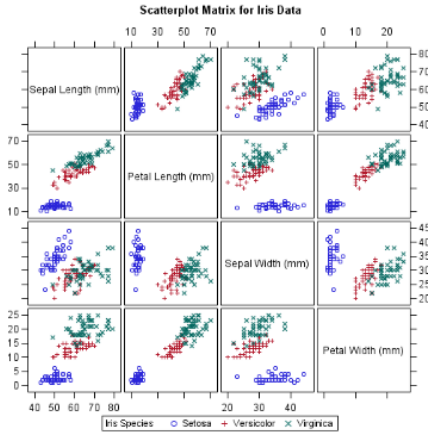
parallel coordinate plots



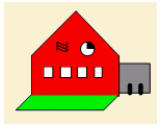
star/radar plots

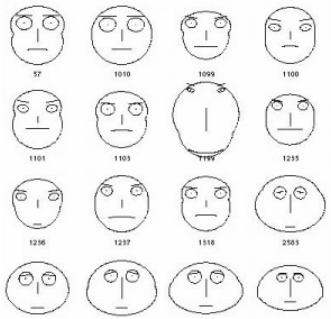


Scatterplot Matrix



Icons

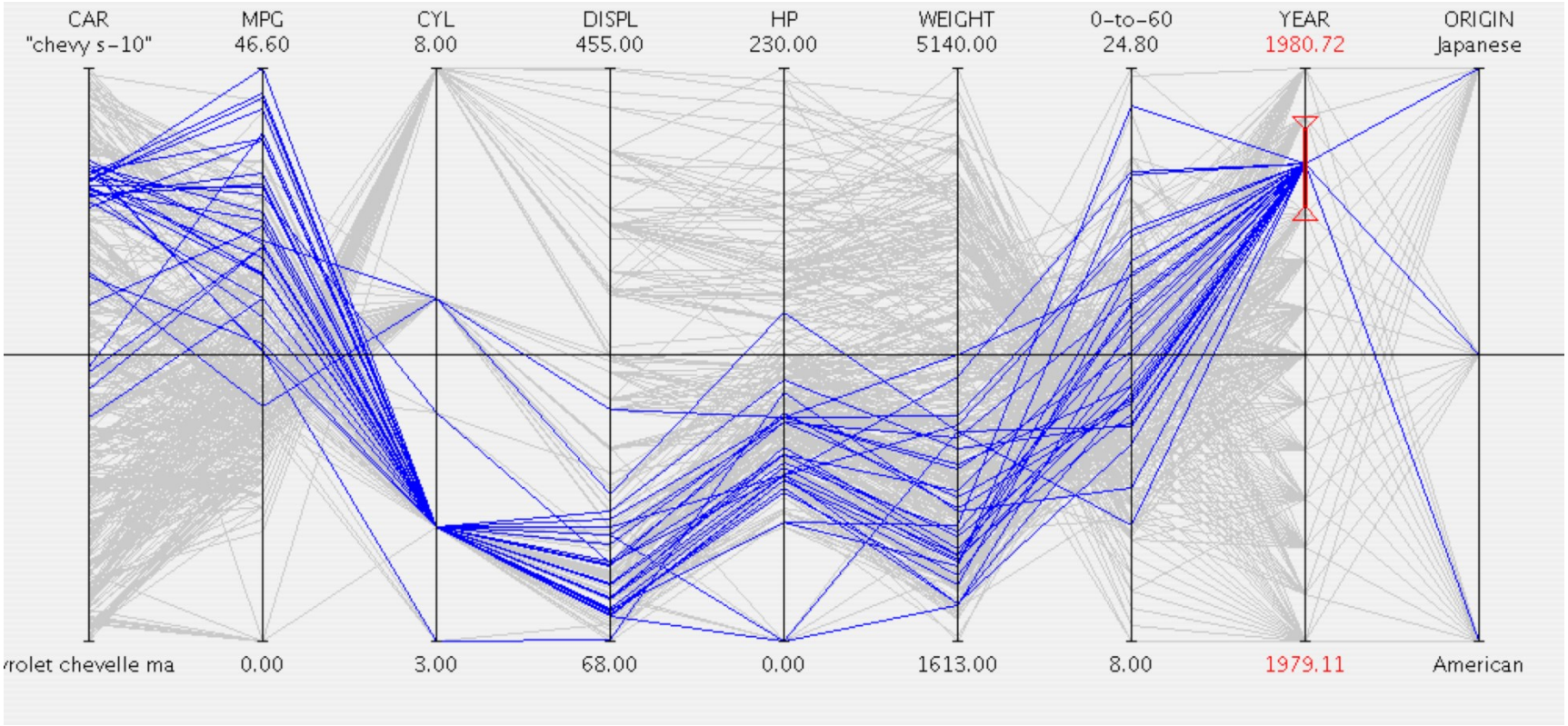




...

- **Parallel coordinates plots** are one of the most popular techniques for hypervariate data

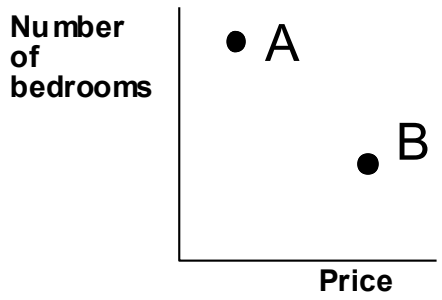
- They have a very simple basis



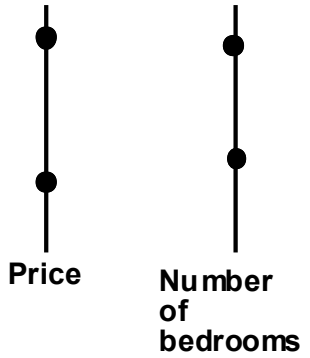
(Spence, 2007)

Consider a simple case of bivariate data:

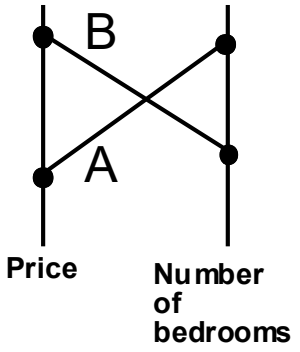
1- A scatterplot represents the price and number of bedrooms associated with two houses



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

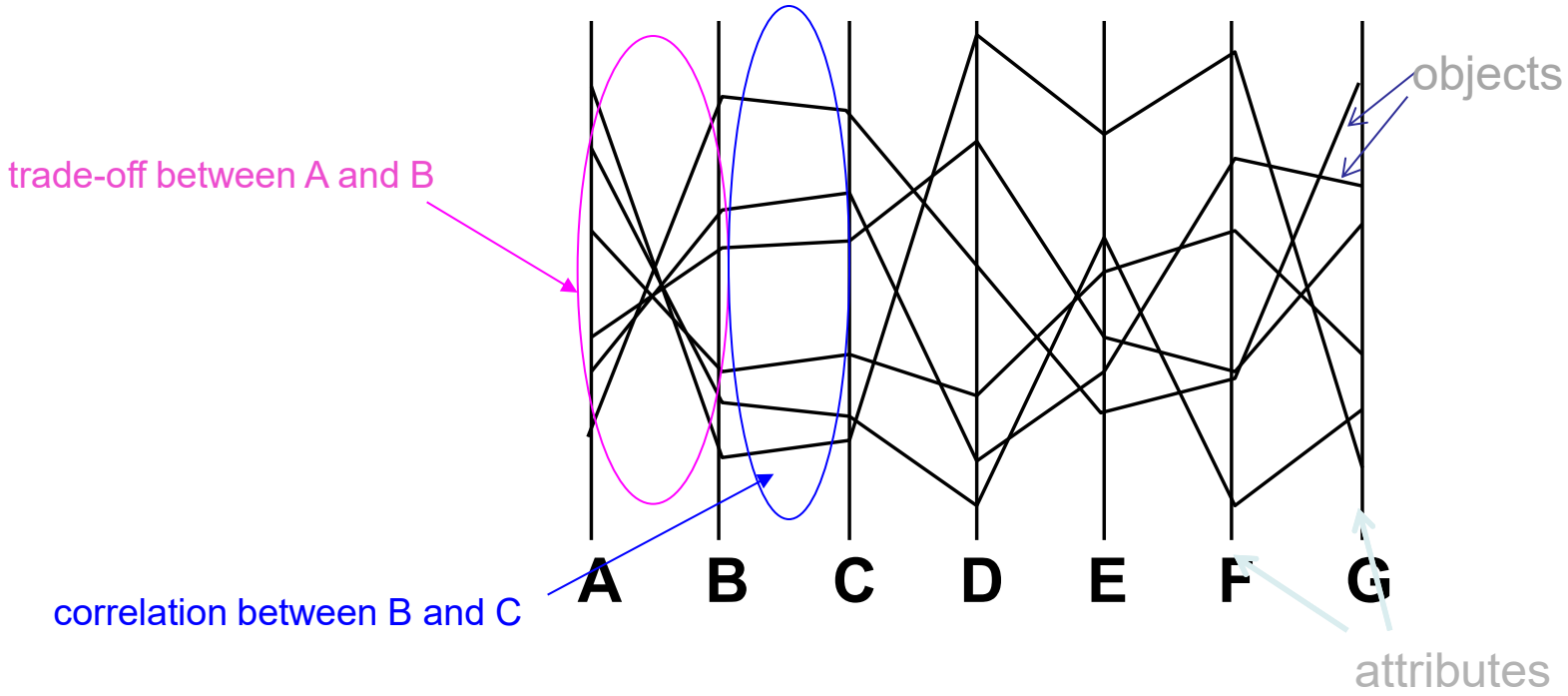


3- To avoid ambiguity the pair of points representing a house are joined and labeled

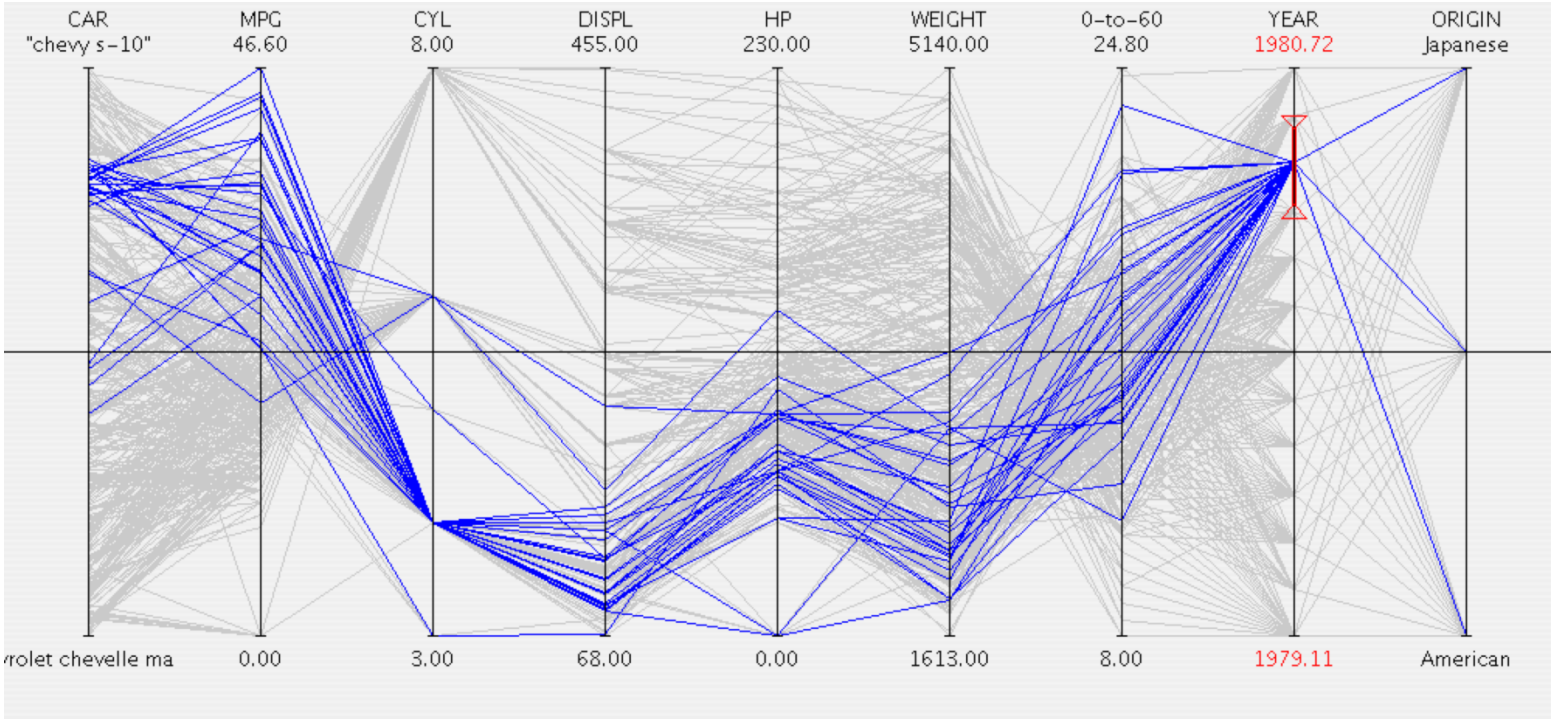


- For objects characterized by many attributes the parallel coordinate plots offer many advantages

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



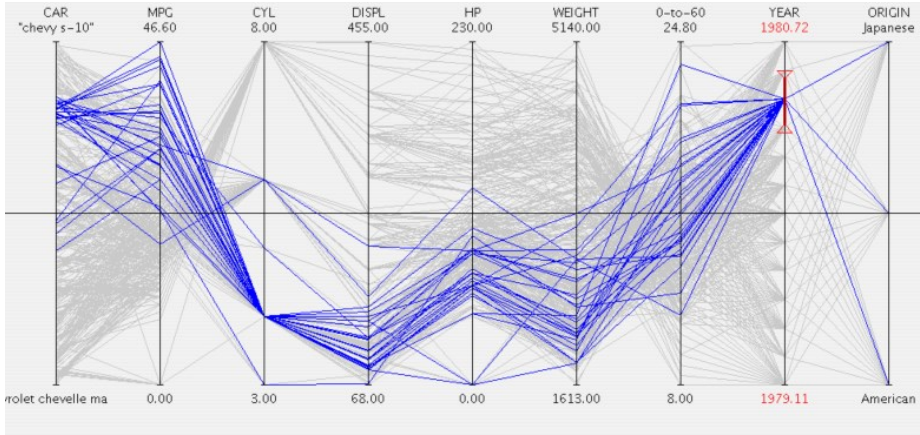
The trade-off between A and B, and the correlation between B and C, are immediately apparent. The trade-off between B and E, and the correlation between C and G, are not.



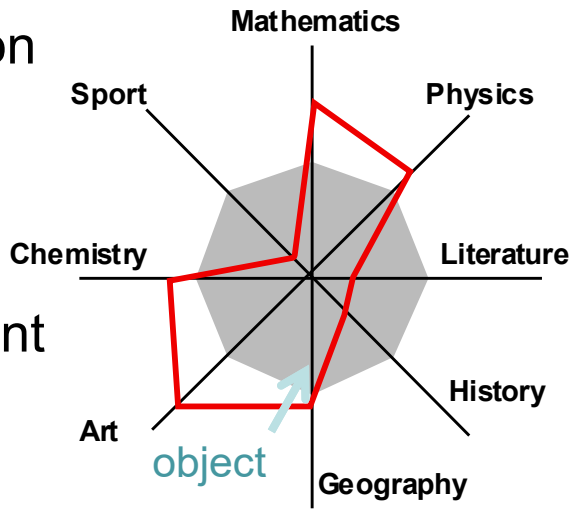
A parallel coordinate plot representation of a collection of cars, in which a range of the attribute *Year* has been selected to cause all those cars manufactured during that period to be highlighted.

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**
- Attributes have **uniform treatment**



- **Star plots** have many features in common with parallel coordinate plots

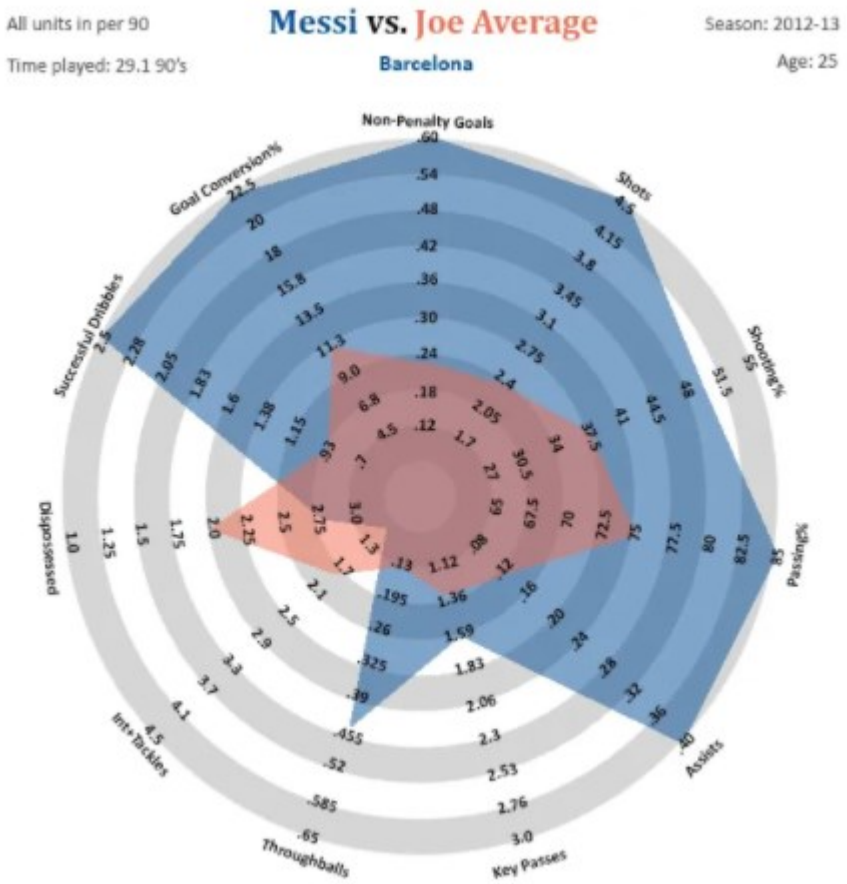


(Spence, 2007)

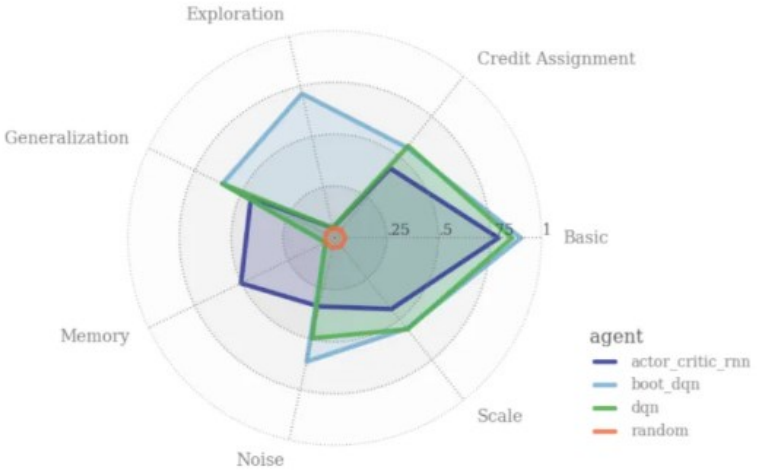
- An attribute value is represented by a point on a coordinate axis
- Attribute axes radiate from a common origin
- For a given object, points are joined by straight lines
- Other useful information such as average values or thresholds can be encoded

Examples of Star plot usage

Players



AI agents



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

Exercise

1- Draw a parallel coordinates representation of the following data: (candidates to a position)

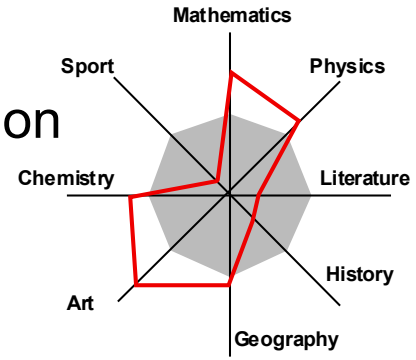
2- Draw a star plot

3- Ask a question better supported by the star plot

#	Education	Age	Experience	Gender	Salary	English	French
#	(MSc/PhD)	(years)	(years)	(F/M)	(euros)	(Bas/Ada)	(Y/N)
1	MSc	22	0	M	2500	Advanced	N
2	MSc	23	0	M	2500	Basic	N
3	MSc	24	1	F	2500	Advanced	Y
4	PhD	38	7	F	6000	Advanced	Y
5	MSc	25	1	M	2500	Basic	N
6	PhD	35	5	M	5000	Advanced	Y
7	MSc	31	7	M	3000	Advanced	N
8	MSc	23	0	F	2500	Advanced	N
9	MSc	26	2	F	3000	Basic	Y
10	MSc	28	4	M	3000	Advanced	N

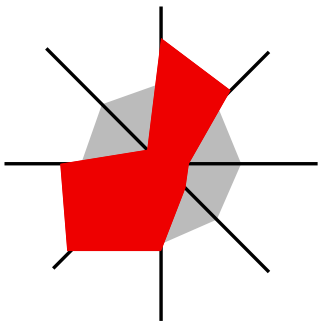
Properties of star plots:

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

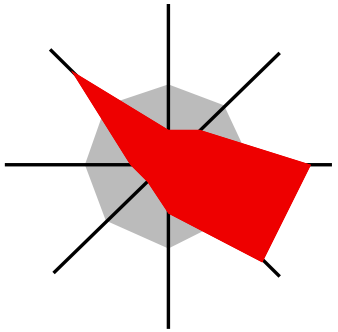


- They offer **object visibility** and are suitable to compare objects

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



Bob's performance



Tony's performance

(Spence, 2007)

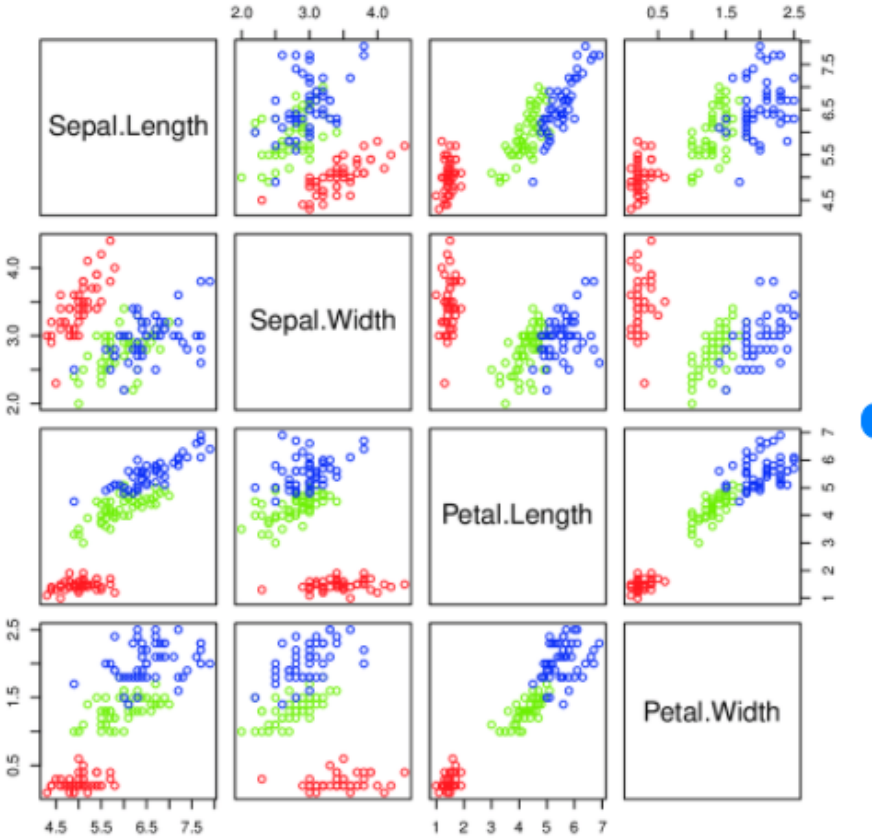
- The **scatterplot matrix** is applicable to higher dimensions

- However, as the number of attributes increases, the number of different pairs of attributes increases rapidly:

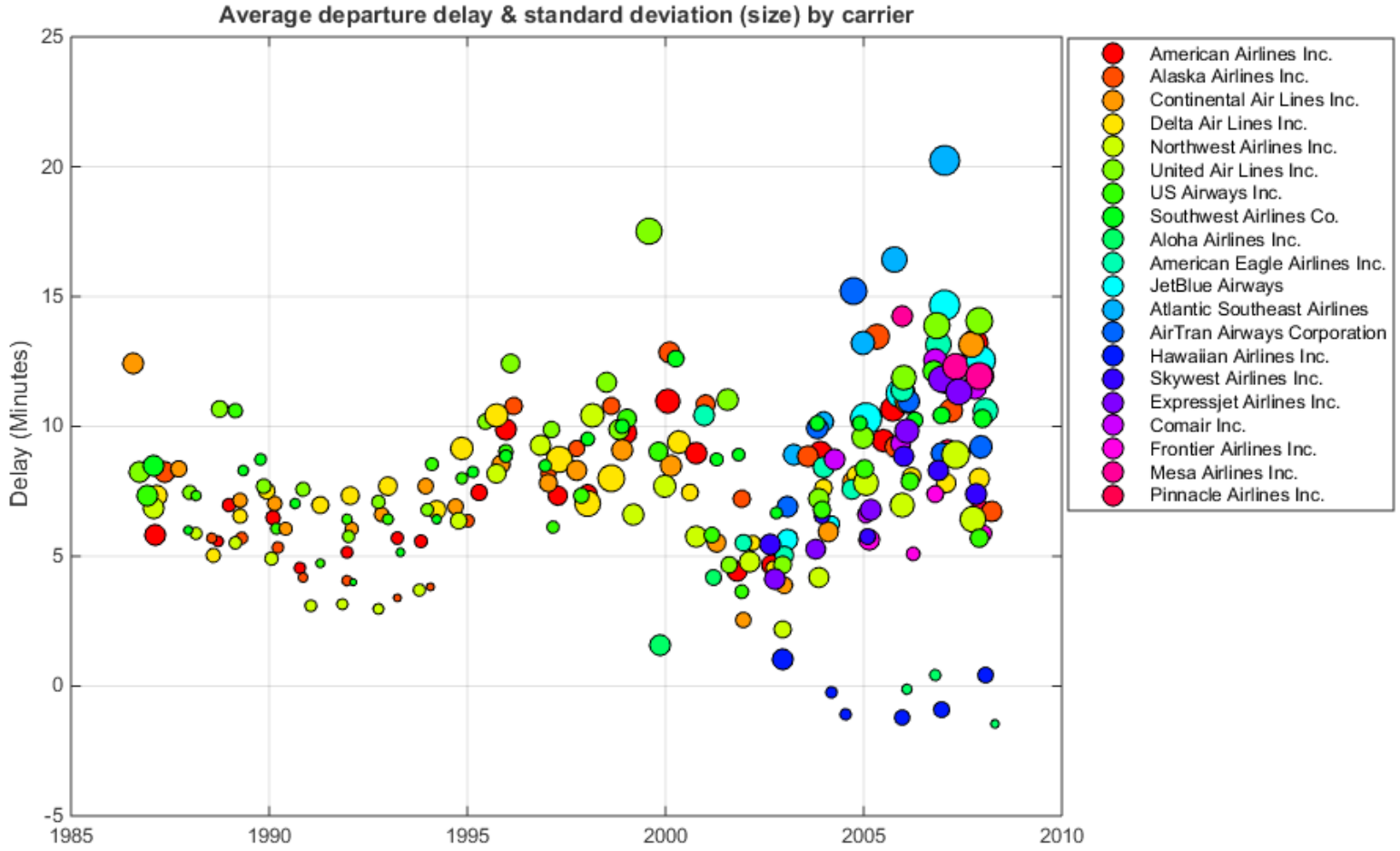
  – 2 attributes -> 1 scatterplot

  – 3 attributes -> 3 scatterplots

  – 4 attributes -> 6 scatterplots



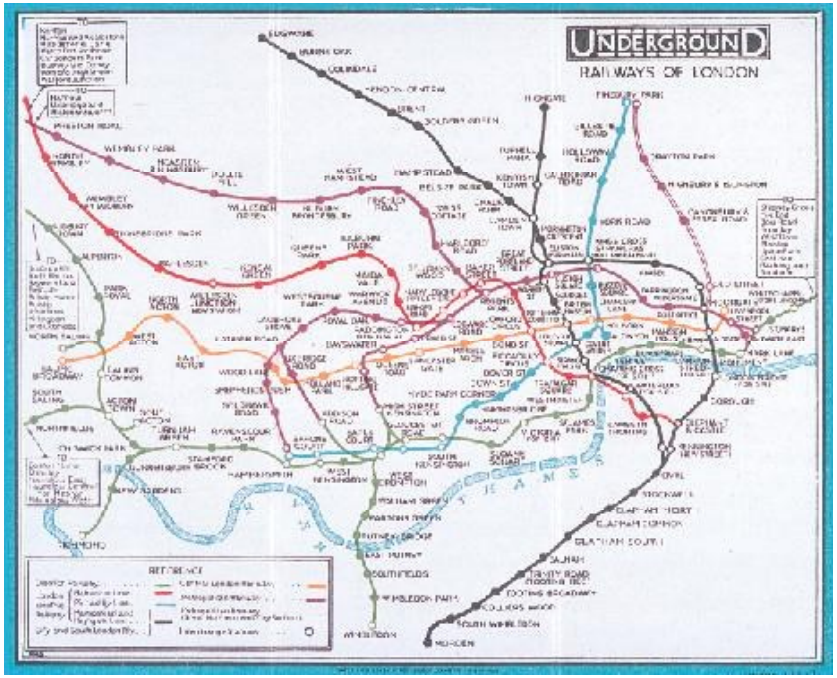
A single scatterplot can be used together with other encoding techniques to represent data of higher dimension



Which encoding techniques can be used?

Representation- II Encoding relation

(some common techniques)



- Types of data:
  - Value (already addressed)
  - Relation
  - …
- What is relation?

  A logical or natural association between two or more things; relevance of one to another; connection (in dictionary)
- How to visually represent relation?

Techniques for Visualization of Relation

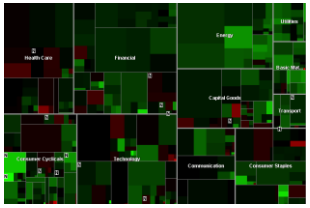
Lines

Venn diagrams

Trees

Hyperbolic browser

Treemap



...

- The nature of relations is different from values

- However, the **prime considerations** in the choice of a representation **are similar**:

  - an understanding of the task

  - the insight that is sought

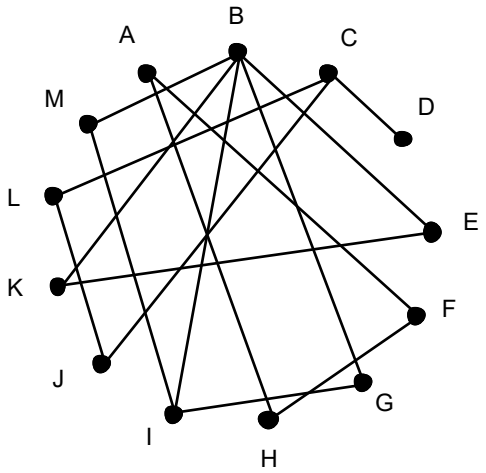
  - what questions might be asked of the relation

- Some representations of relation are **very simple, yet powerful**

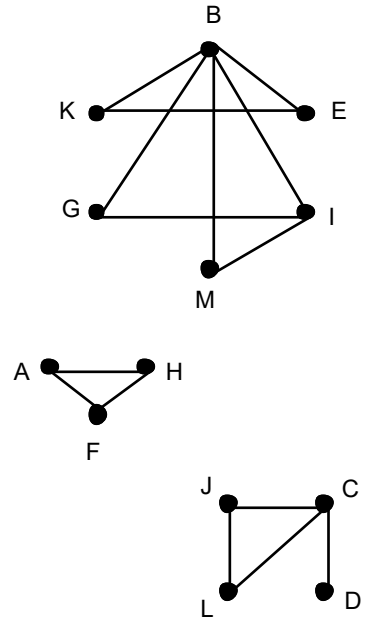
- **Lines** are perhaps the simpler way to represent a relation between entities

Originator	Receiver
A	H
C	L
I	M
B	E
F	H
G	I
I	B
B	M
K	B
G	B
K	E
C	J
D	C

(a)



(b)

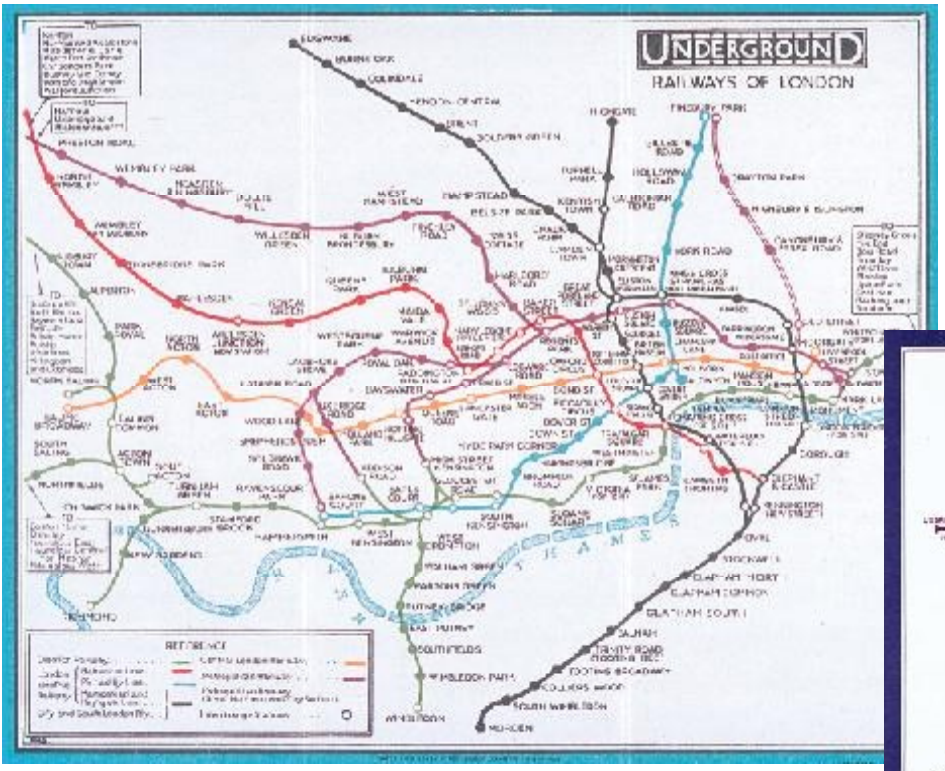


(c)

Insight into even a short list of telephone calls (a) is enhanced by their node-link representation (b), especially if disconnected subsets can be identified (c).

(Spence, 2007)

Perhaps the most familiar use of lines to represent relations is in transportation maps



Before Harry Back

London Underground

Harry Beck´s map (1933)

https://www.english-heritage.org.uk/visit/blue-plaques/harry-beck/

Topological transportation maps:

- Exploit the shapes into which lines connecting stations are arranged

- Use color to denote different lines

- Use symbols to denote different types of stations

- Have general (but not necessarily accurate) veracity

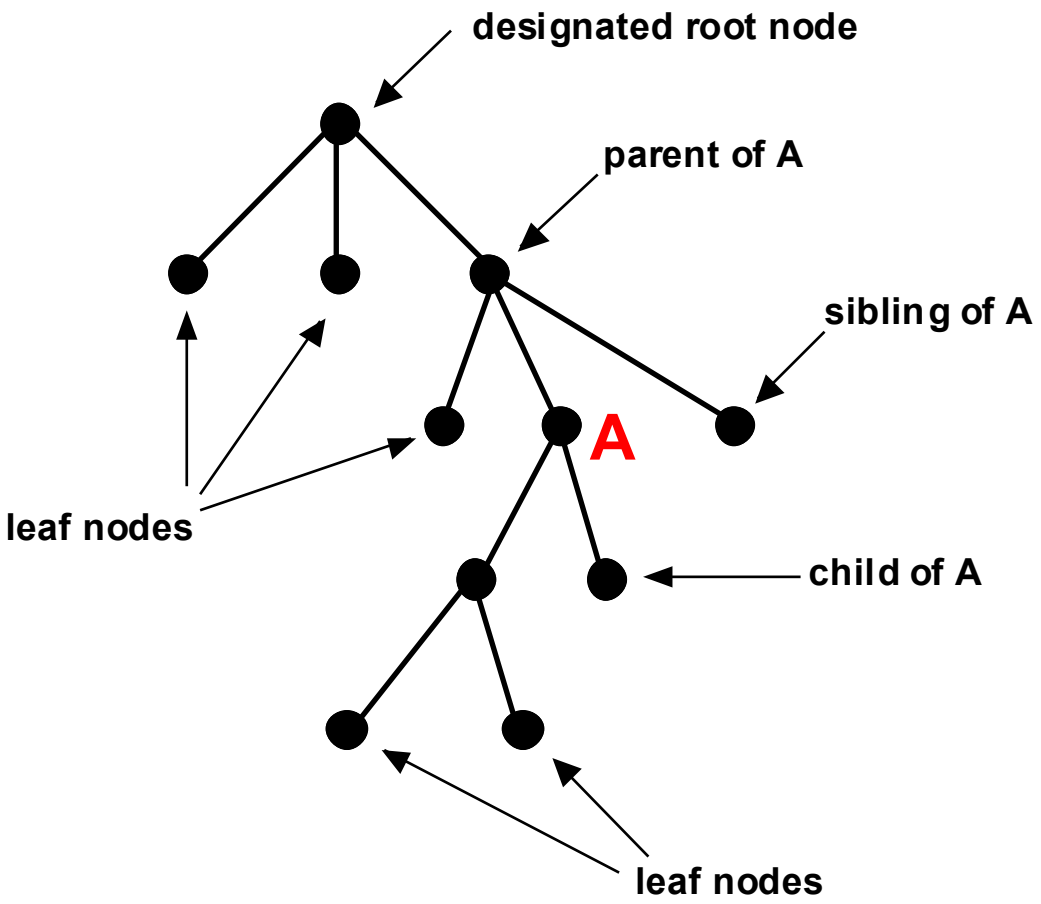
- The previous representations show relations with no restriction upon what is connected **(networks)**

- **Trees** are a class of networks which have no loops

- There are several common representations:

  - treemaps

  - hyperbolic browser, …



(Spence, 2007)

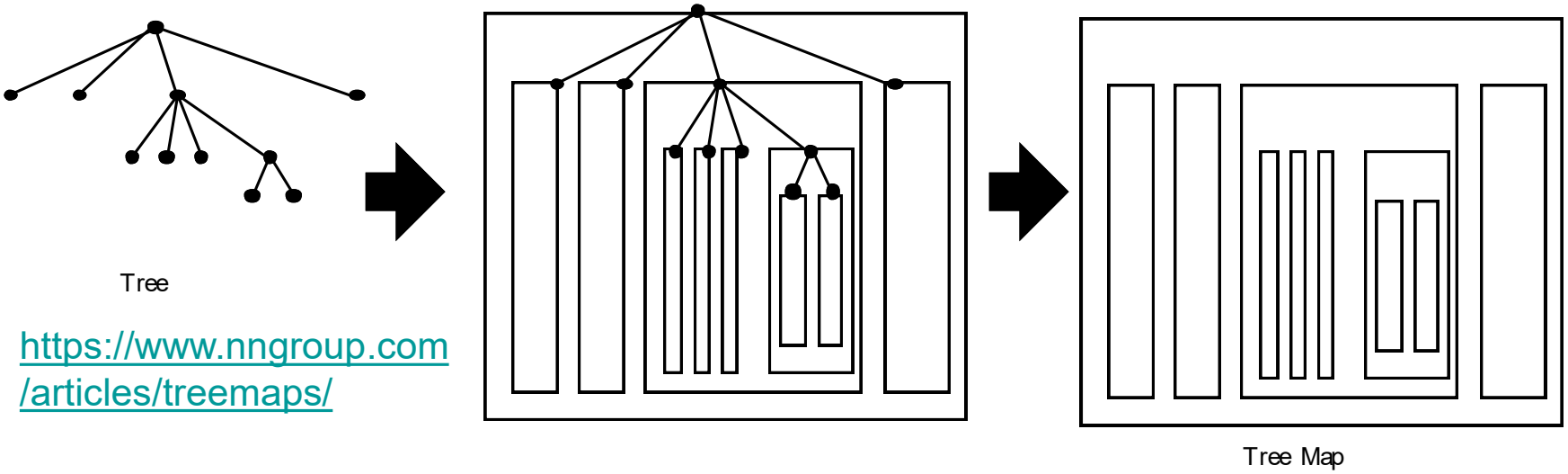
- The **Tree map** (Jonhson and Shneiderman, 1991)

  is a representation of hierarchical data:

  - the root node is represented by a rectangle (usually as large as possible)

  - within this rectangle there are smaller rectangles (one for each child node)

  - this is repeated until all the nodes are represented



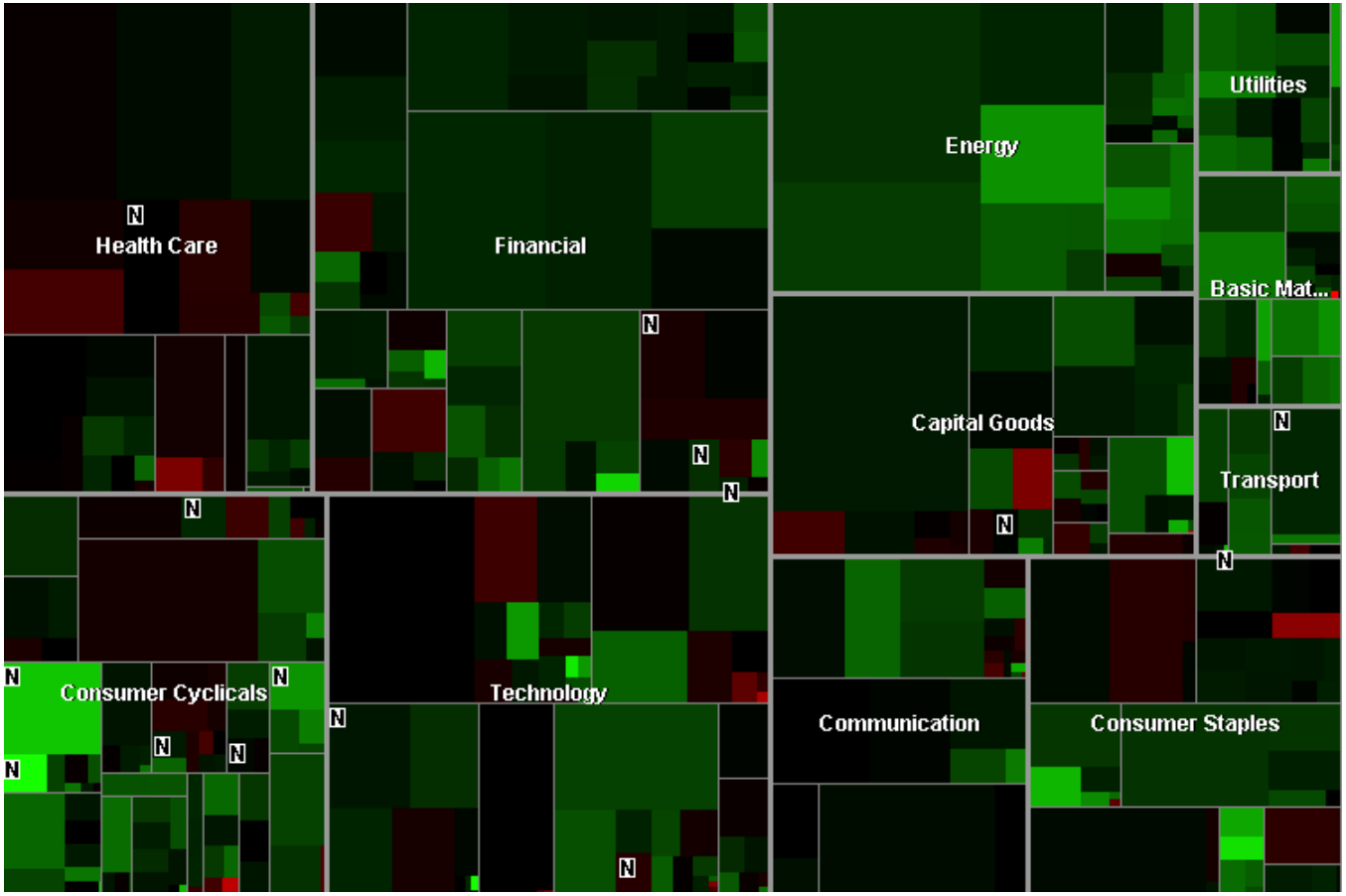
https://www.nngroup.com/articles/treemaps/

The construction of a Tree Map (Spence, 2007)

- The treemap offers a lot of opportunities for
  - encoding by color and area
  - interaction by mouse-over and selection of further detail by mouse click

https://finviz.com/map.ashx

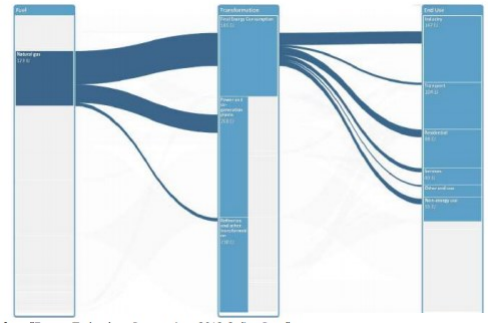
Not for balanced tree

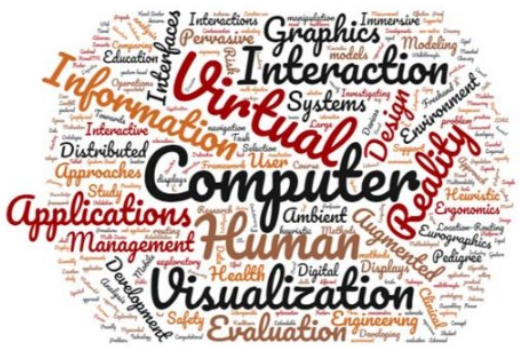


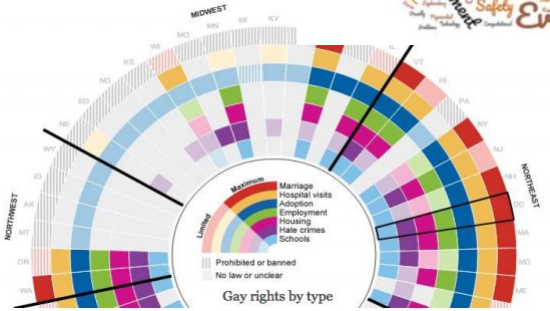
Status of companies within a number of sectors

Representation III – a few other

(less common) Visualization techniques …

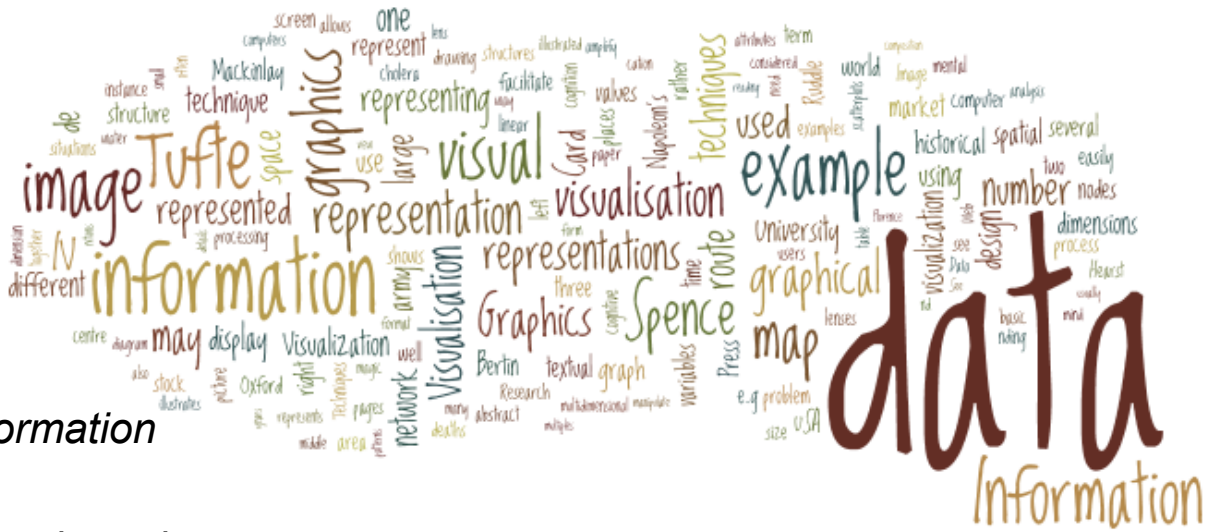






Word Cloud: simple representation of text

- Used to visualize free form text or tags
- After preprocessing the text the number of occurrences of each word is shown with font size or color



R. Mazza, *Introduction to Information Visualization*, 2004 (chap.1):

"Data" was the most often used word

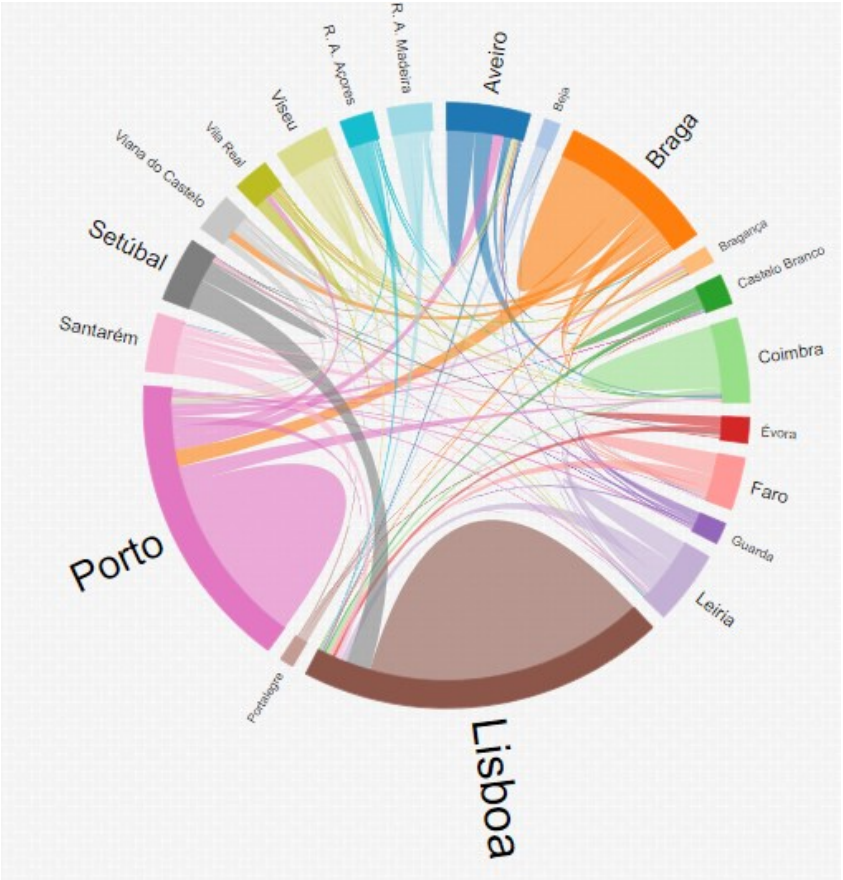
Small multiples

- An arrangement approach that facilitates efficient and effective comparisons

  (Kirk, 2012)

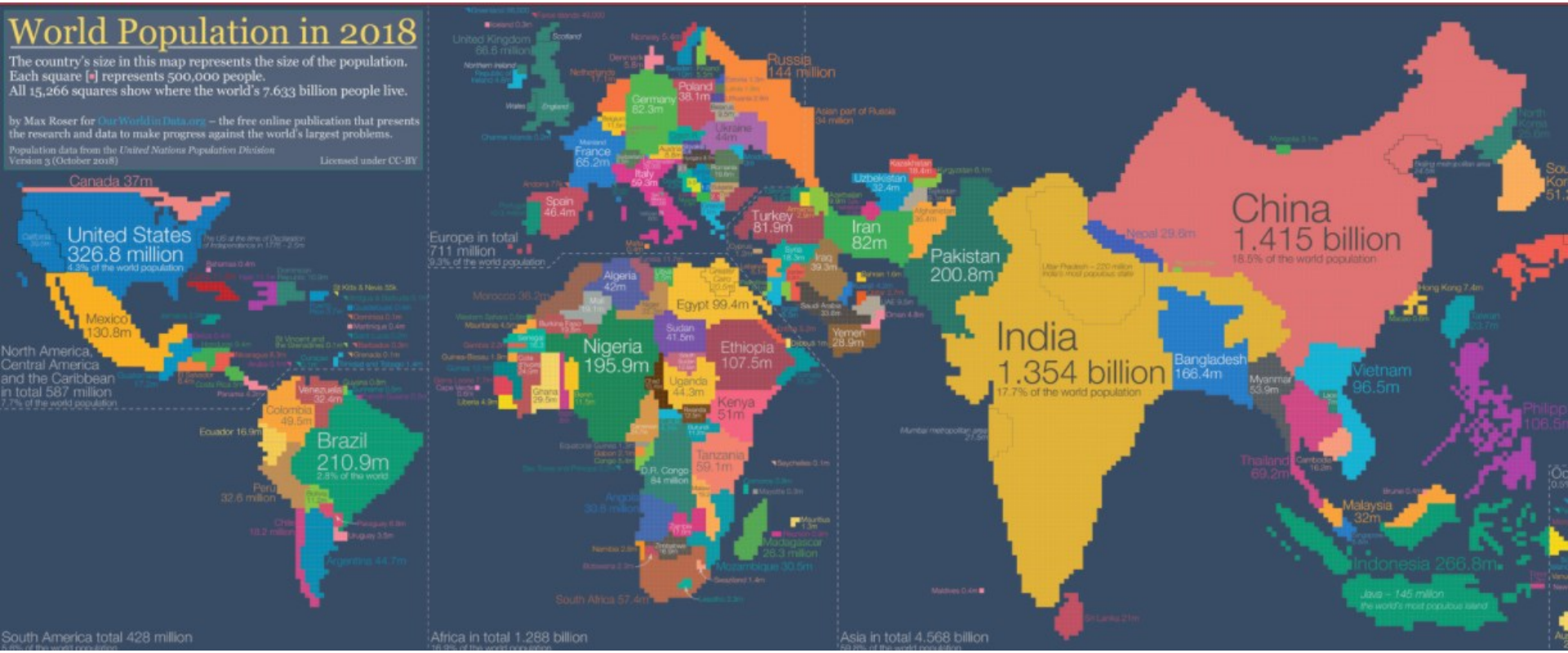
Sankey diagrams

- Useful to convey the idea of flow



Cartograms

Thematic map of a set of features, in which their geographic size is altered to be directly proportional to a selected ratio-level variable



https://en.wikipedia.org/wiki/Cartogram

Want to know a lot more?

See for instance:

A. Kirk, Data Visualization: A Successful Design Process, Packt Publishing, 2012

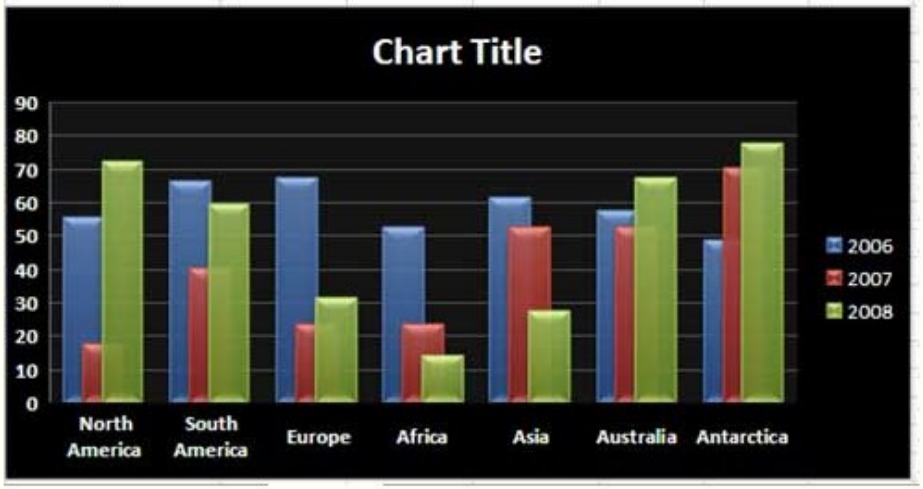
https://books.google.pt/books/about/Data_Visualization.html?id=I4qBVLfD3t4C&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false

J. Heer, M. Bostock, and V. Ogievetsky, “A tour through the visualization zoo,” *Communications of the ACM*, vol. 8, no. 1, pp. 6–22, 2010

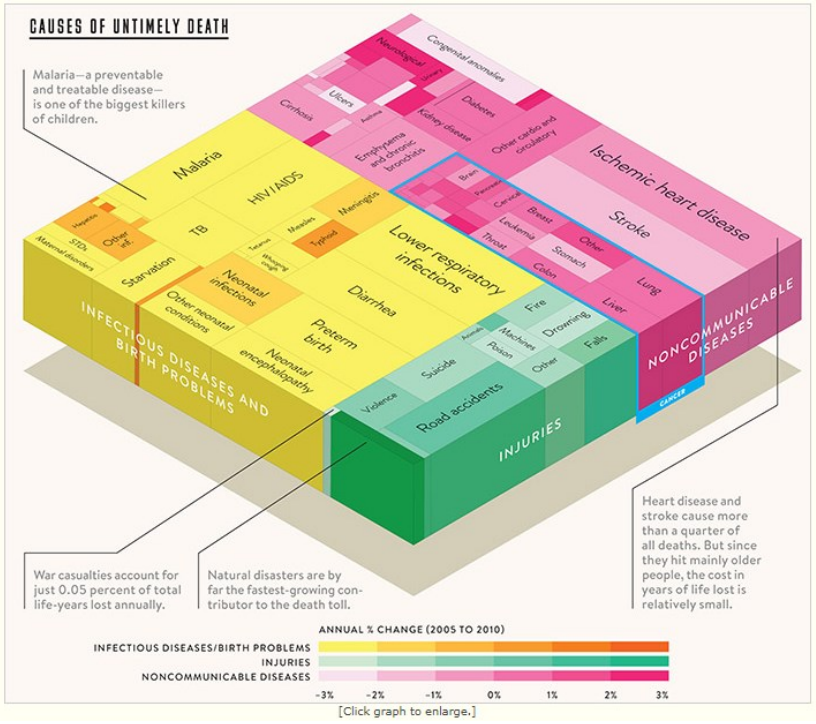
https://dl.acm.org/doi/pdf/10.1145/1743546.1743567

There are a lot of different ways to visually represent data
But often we see not adequate ways to do it…

Examples that could be more effective explained by S. Few:



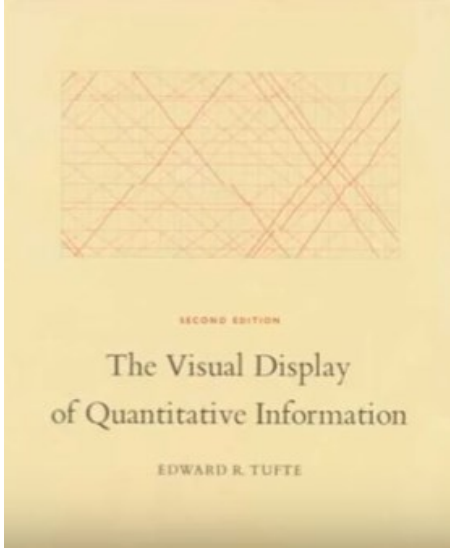
The bars are overlaid on one another, partially obscuring the first two sets and giving different visual salience …



3-D effect doesn't provide information and makes the treemap harder to decode …

http://www.perceptualedge.com/examples.php

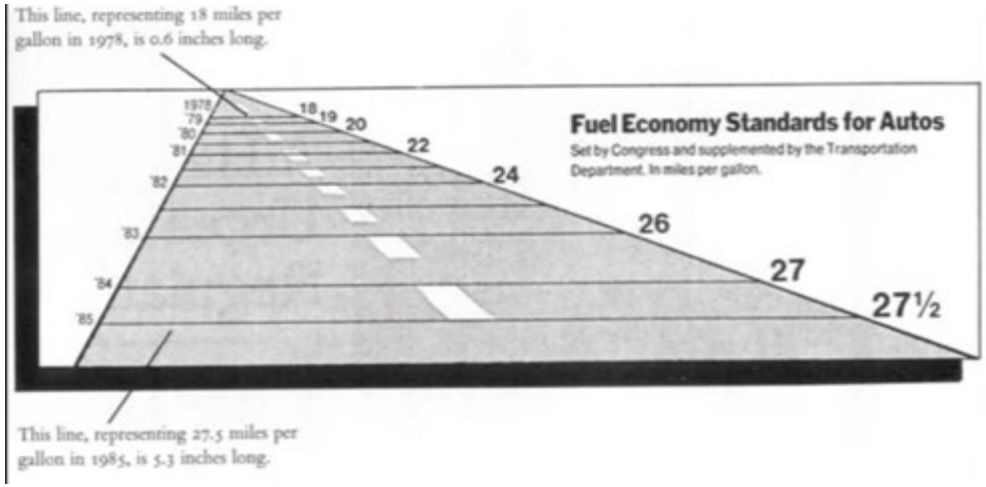
The issue of Effective Visualization



It implies saying the **truth** about the data

Tufte presents a lot of commented examples in his book:

Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983

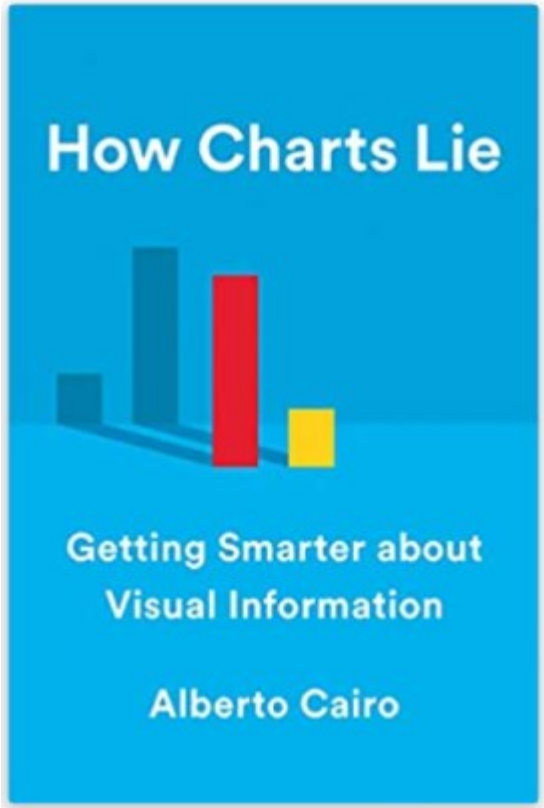


There are methods to evaluate visualizations that should be used along the process of creating a visualization

https://infovis-wiki.net/wiki/Lie_Factor

Can visualizations lie?

(And statistics? …)



An expression of concern:

Alberto Cairo, *How Charts Lie: Getting Smarter about Visual Information*, W. W. Norton & Company, 2019

https://books.google.pt/books/about/How_Charts_Lie_Getting_Smarter_about_Vis.html?id=qP2KDwAAQBAJ&printsec=frontcover&source=kp_read_button&hl=en&redir_esc=y#v=onepage&q&f=false

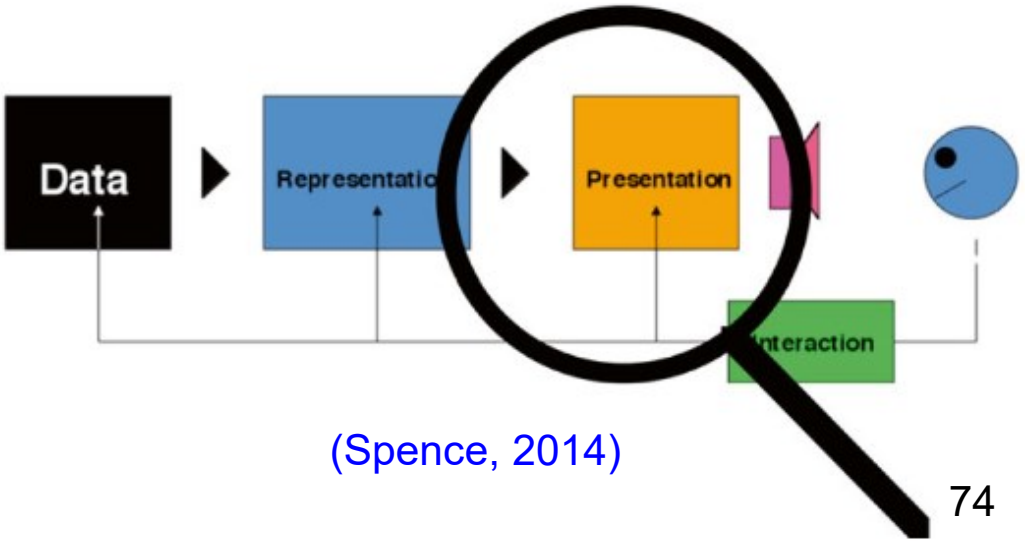
https://core.ac.uk/download/pdf/322505481.pdf

Presentation and Interaction

(some common techniques)

The presentation issue

- The issue of layout is important due to the limited screen real estate
- Irrespective of how data may be represented decisions have to be made:
  - how the representation is to be displayed
  - whether it is to be displayed
- Links to representation and interaction are important



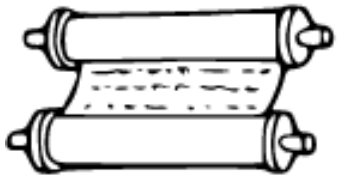
(Spence, 2014)

Space limitations

- Techniques often used to overcome space limitations (a more serious issue in mobile platforms)
  - Scrolling
  - Zoom and pan
  - Overview + detail
  - Distortion
  - Suppression
  - …

Simple often used techniques for presentation

- **Scrolling** is an obvious solution when a document is larger than the display

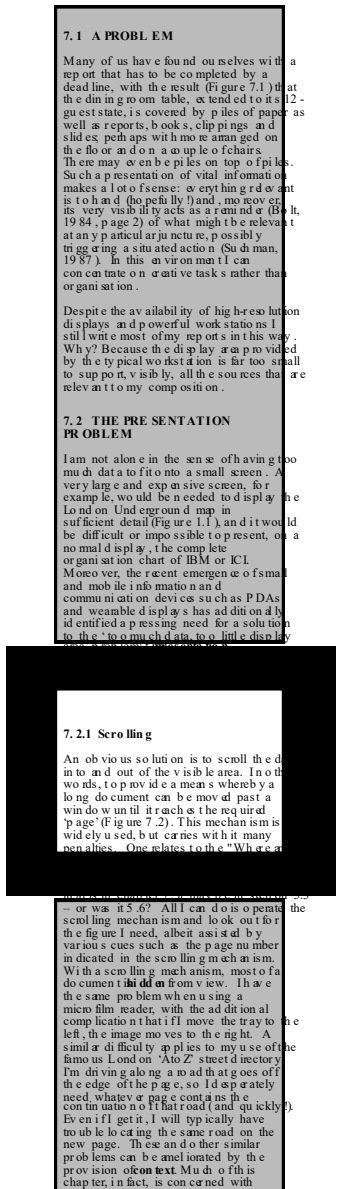


- A long document can be moved past a “window”

- Often it is not a satisfactory solution

- Scrolling hides most of a document:

  there is not a view of context as well as detail



7.1 A PROBLEM

Many of us have found ourselves with a report that has to be completed by a deadline, with the result (Figure 7.1) that the dining room table, extended to its 12-guest state, is covered by piles of paper as well as reports, books, clippings and slides, perhaps with more arranged on the floor and on a couple of chairs. There may even be piles on top of piles. Such a presentation of vital information makes a lot of sense: everything relevant is to hand (hopefully!) and, moreover, its very visibility acts as a reminder (Bolt, 1984, page 2) of what might be relevant at any particular juncture, possibly triggering a situated action (Suchman, 1987). In this environment I can concentrate on creative tasks rather than organisation.

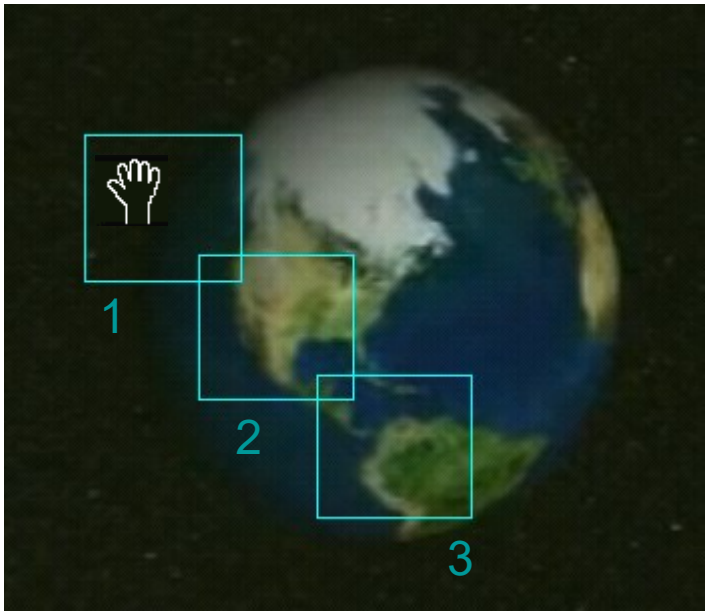
Despite the availability of high-resolution displays and powerful workstations I still write most of my reports in this way. Why? Because the display area provided by the typical workstation is far too small to support, visibly, all the sources that are relevant to my composition.

7.2 THE PRESENTATION PROBLEM

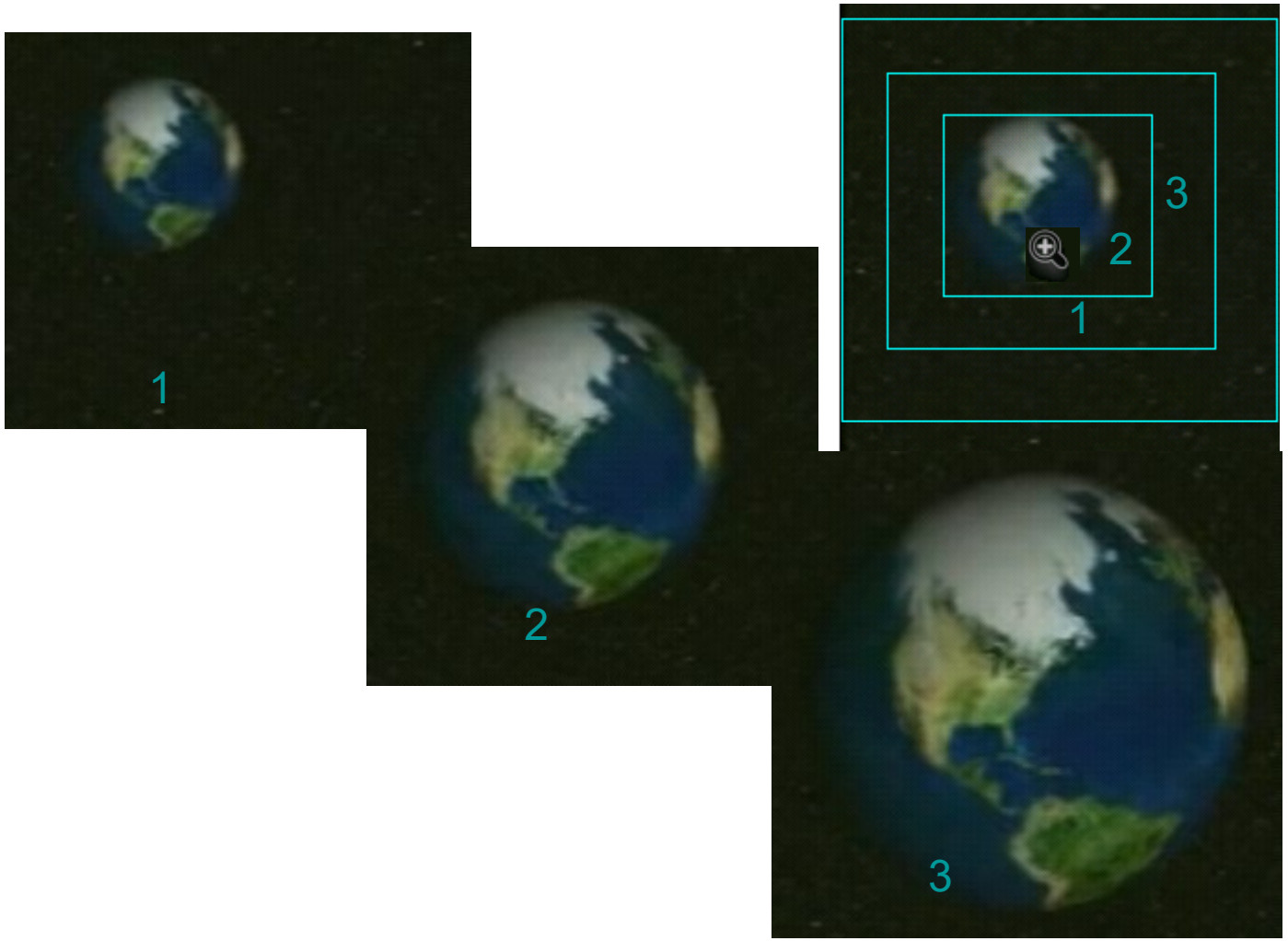
I am not alone in the sense of having too much data to fit onto a small screen. A very large and expensive screen, for example, would be needed to display the London Underground map in sufficient detail (Figure 1.1), and it would be difficult or impossible to present, on a normal display, the complete organisation chart of IBM or ICL. Moreover, the recent emergence of small and mobile information and communication devices such as PDAs and wearable displays has additionally identified a pressing need for a solution to the ‘too much data, too little display

7.2.1 Scrolling

An obvious solution is to scroll the document into and out of the visible area. In other words, to provide a means whereby a long document can be moved past a window until it reaches the required ‘page’ (Figure 7.2). This mechanism is widely used, but carries with it many penalties. One relates to the "Where am I?" problem. Am I looking at page 5 or 6 — or was it 5.6? All I can do is operate the scrolling mechanism and look out for the figure I need, albeit assisted by various cues such as the page number indicated in the scrolling mechanism. With a scrolling mechanism, most of a document is **hidden** from view. I have the same problem when using a microfilm reader, with the additional complication that if I move the tray to the left, the image moves to the right. A similar difficulty applies to my use of the famous London ‘AtoZ’ street directory. I'm driving along a road that goes off the edge of the page, so I desperately need whatever page contains the continuation of that road (and quickly!). Even if I get it, I will typically have trouble locating the same road on the new page. These and other similar problems can be ameliorated by the provision of **context**. Much of this chapter, in fact, is concerned with



Panning is the smooth movement of a viewing frame over a 2D image



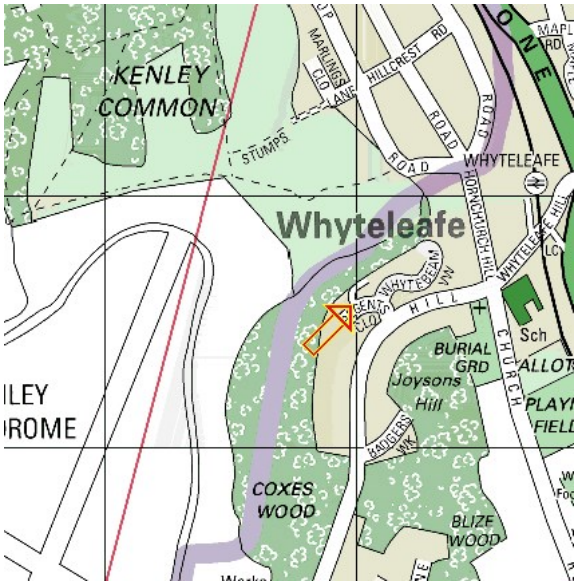
Zooming is the increasing magnification of a decreasing fraction of an image (or *vice versa*)

- Two separate views of detail and of context can be combined in a **overview + detail** view to alleviate the **focus + context** problem

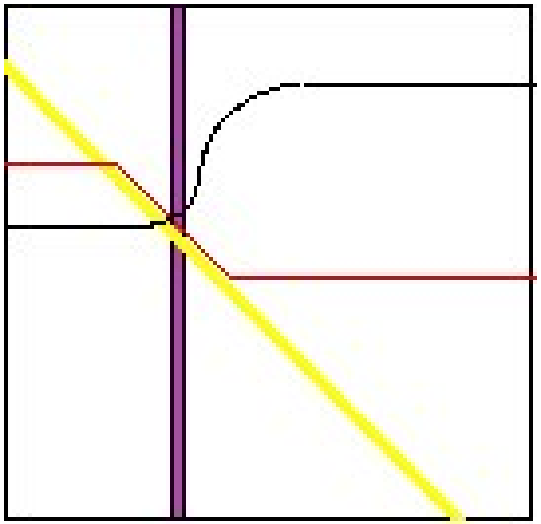
It is often necessary to provide a balance of local detail and global context

Suppression finds valuable application in the **Magic Lens** (Stone et al.,1994)

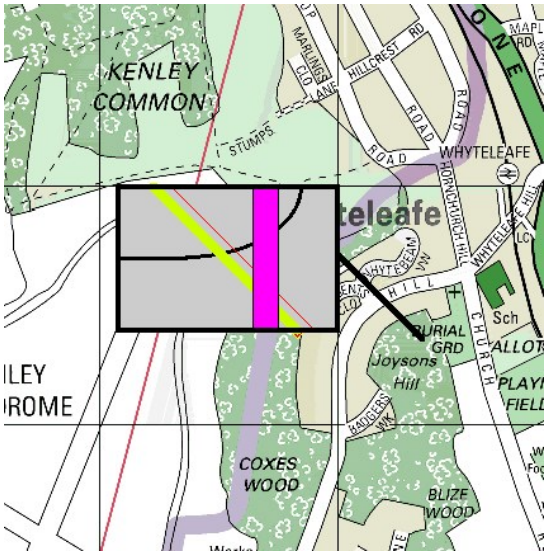
a)



b)



c)

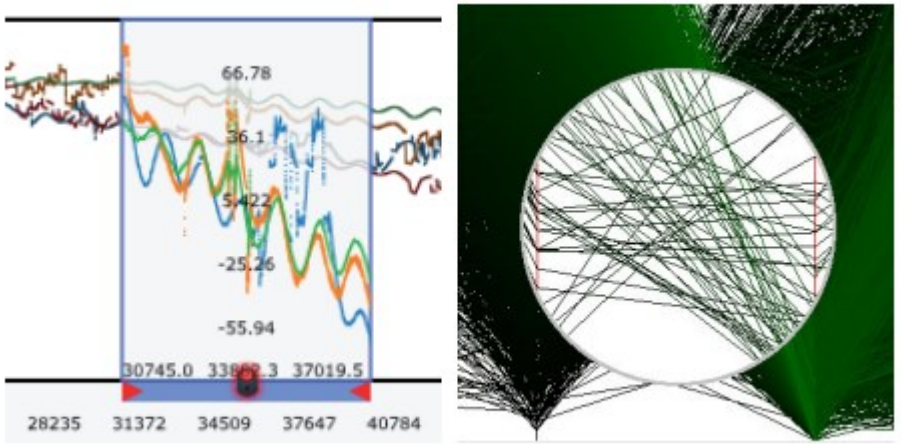


Magic Lens:
(a) shows a conventional map of an area,
(b) shows the location of services (gas, water and electricity pipes)
(c) a (movable) Magic Lens shows services in an area of interest, in context (Spence, 2007)

https://infovis-wiki.net/wiki/Magic_Lens

Magic lenses examples

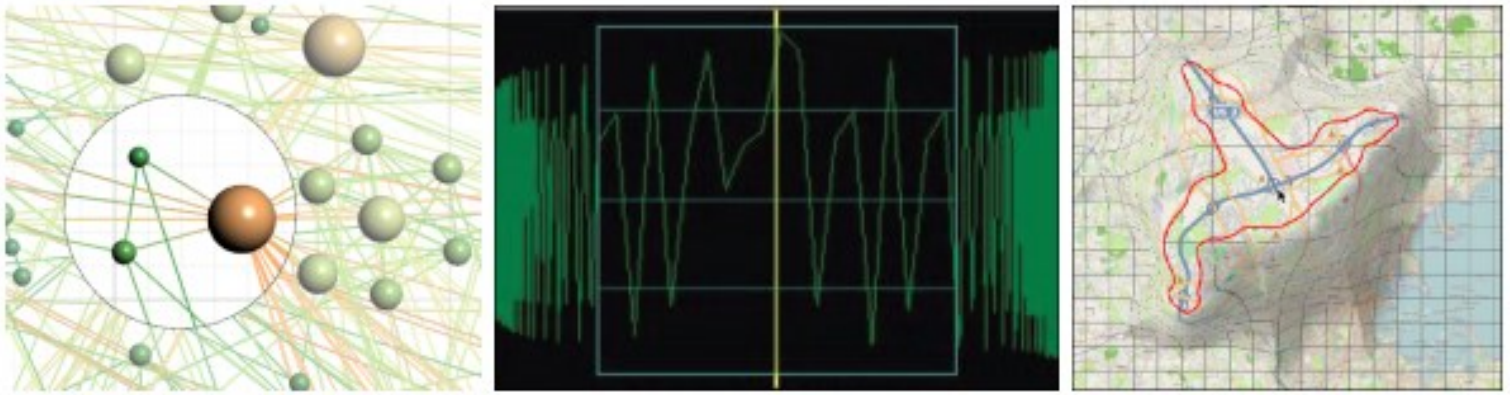
Different functions



(a) Alteration (b) Suppression

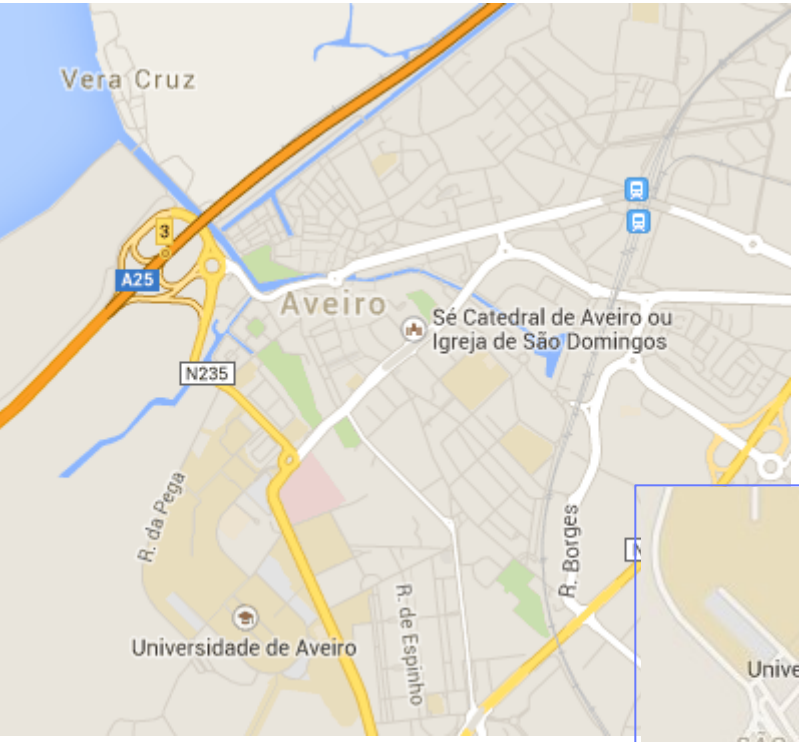
C. Tominski et al.
“A Survey on Interactive Lenses in Visualization”, *EuroVis 2014*

Different shapes



(a) Circular shape (b) Rectangular shape (c) Content-adaptive shape

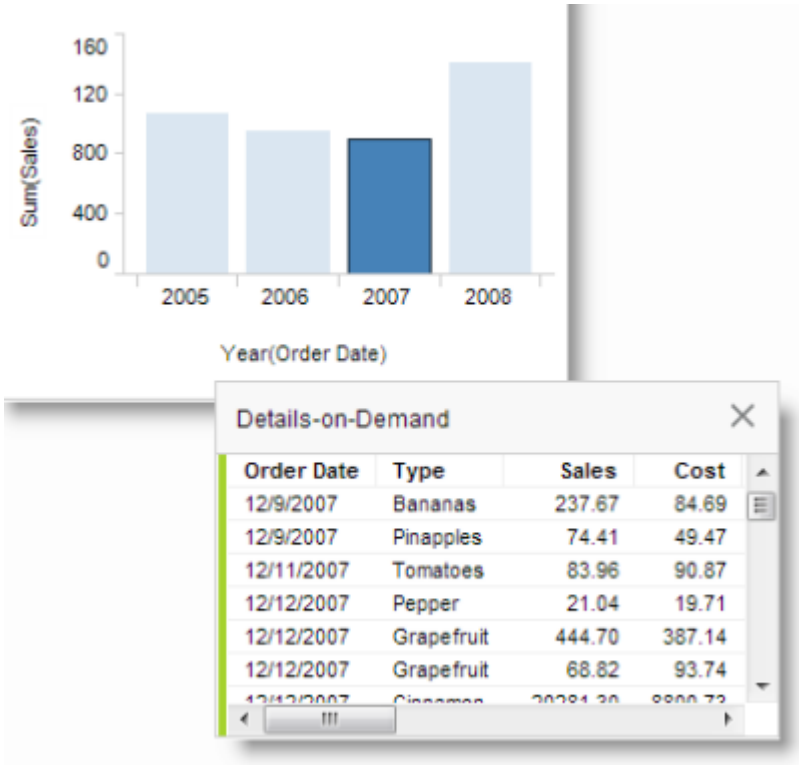
Example: (Semantic) zoom in Google Maps

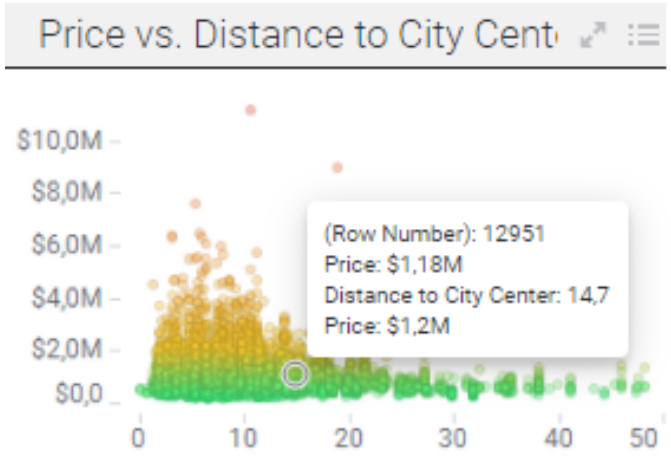


More (and different) information is displayed at different zoom levels

Details on demand

Displays the actual values of marked items in the active visualization

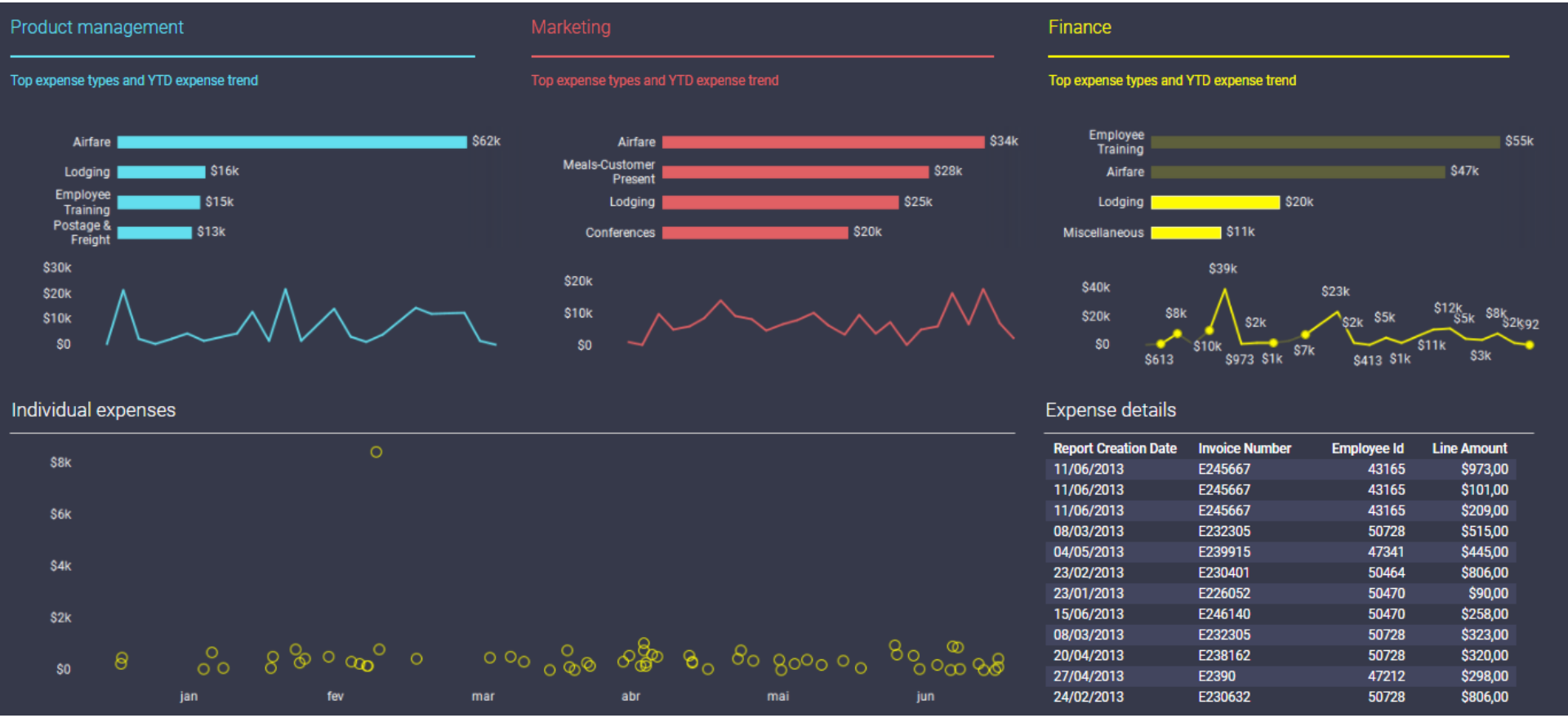




Linking and brushing

connecting different views of the same data,

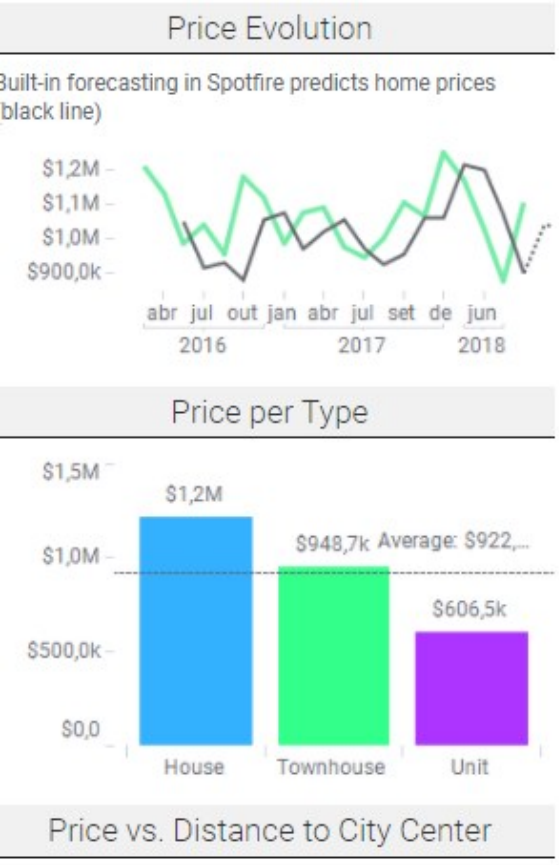
change in one view affects the representation in the other

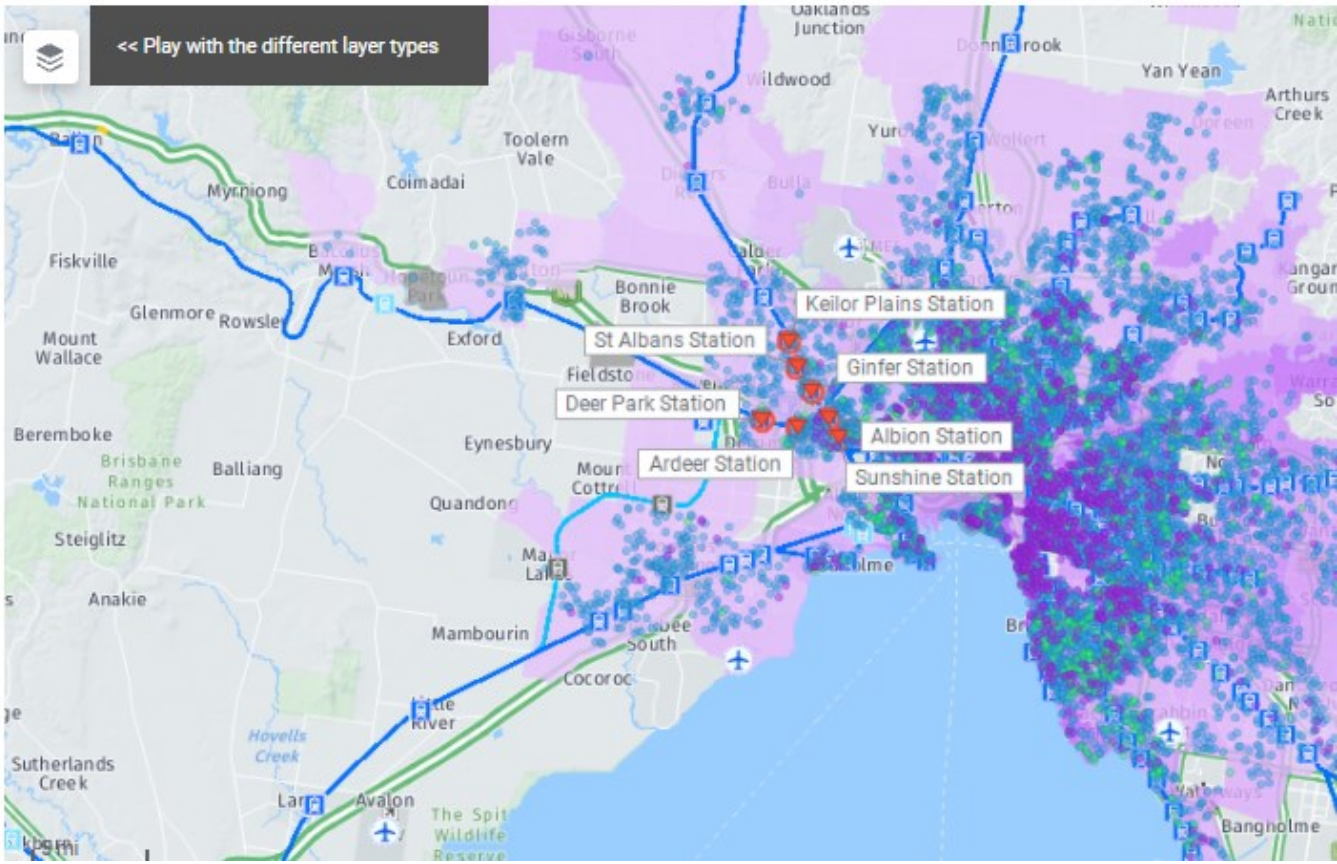


https://www.tibco.com/es/node/46276

Multiple coordinates views

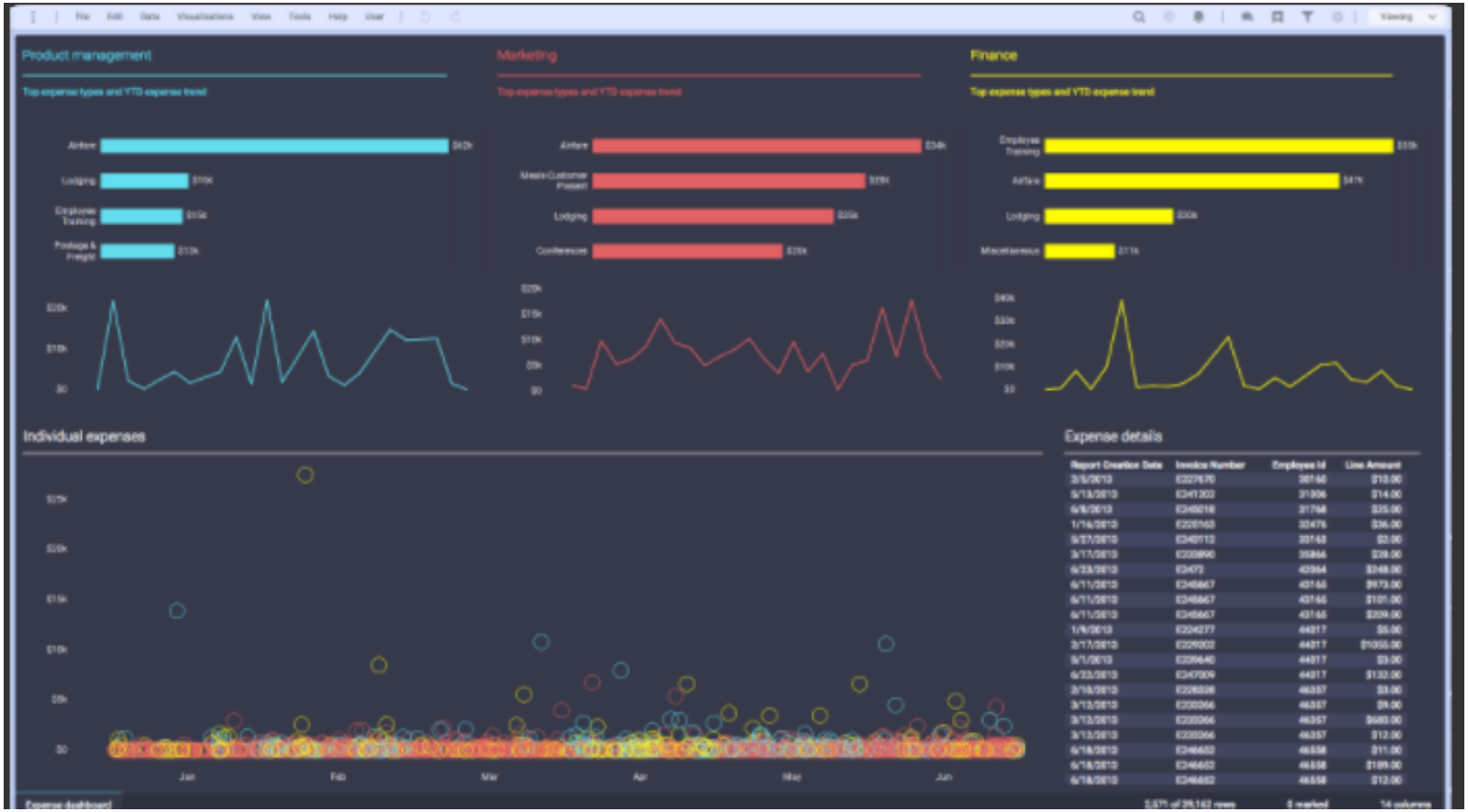
Real Estate Market Analysis





https://www.tibco.com/es/node/83641

Expense Analyzer Dashboard
Cross departmental expense analysis.

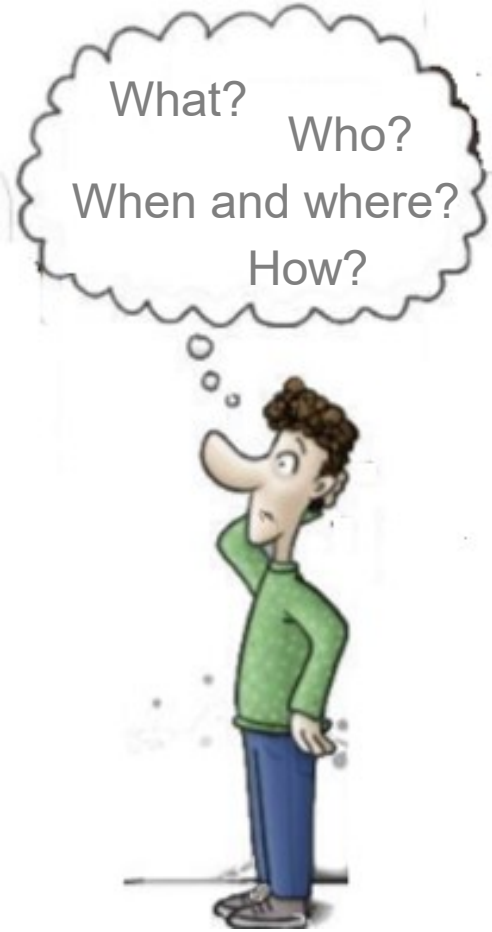


https://www.tibco.com/es/node/46276

In a nut shell:
Do you have a lot of data?

- Visualization may be the solution (or part of it)
- There are **a lot** of visualization techniques
- Should be selected according to the **phenomenon**, **data, users, tasks, and context of use**
- But,

  How to evaluate a solution? → next topic



Main Bibliography

- Spence, R., *Information Visualization, An Introduction*, Springer, 2014
- Munzner, T., *Visualization Analysis and Design*, A K Peters/CRC Press, 2014
- Kirk, A., Data VisualisationA Handbook for Data Driven Design, 2nd. Ed., Sage, 2019
- Mazza, R*., Introduction to Information Visualization*, Springer, 2009
- Ware, C., *Information Visualization, Perception to Design*, 4nd ed.,Morgan Kaufmann, 2020