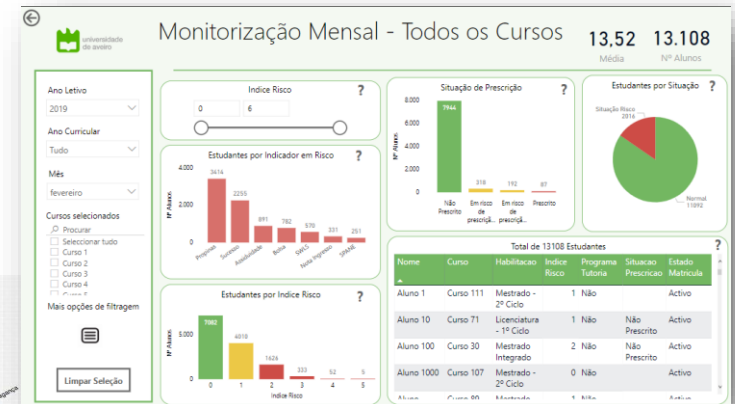
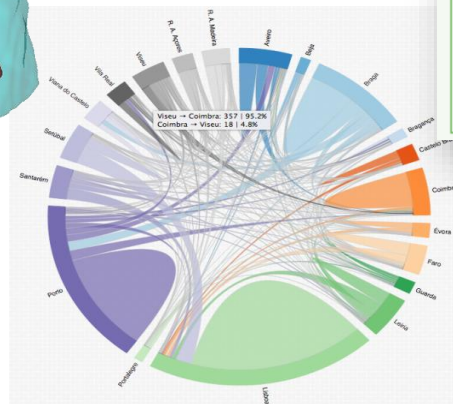
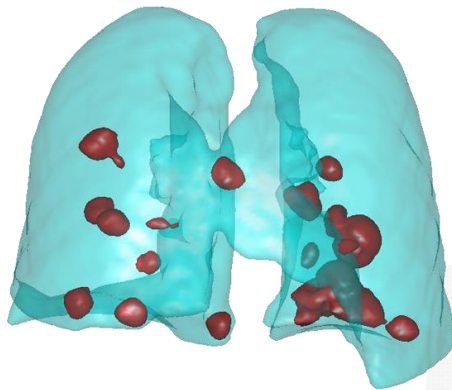
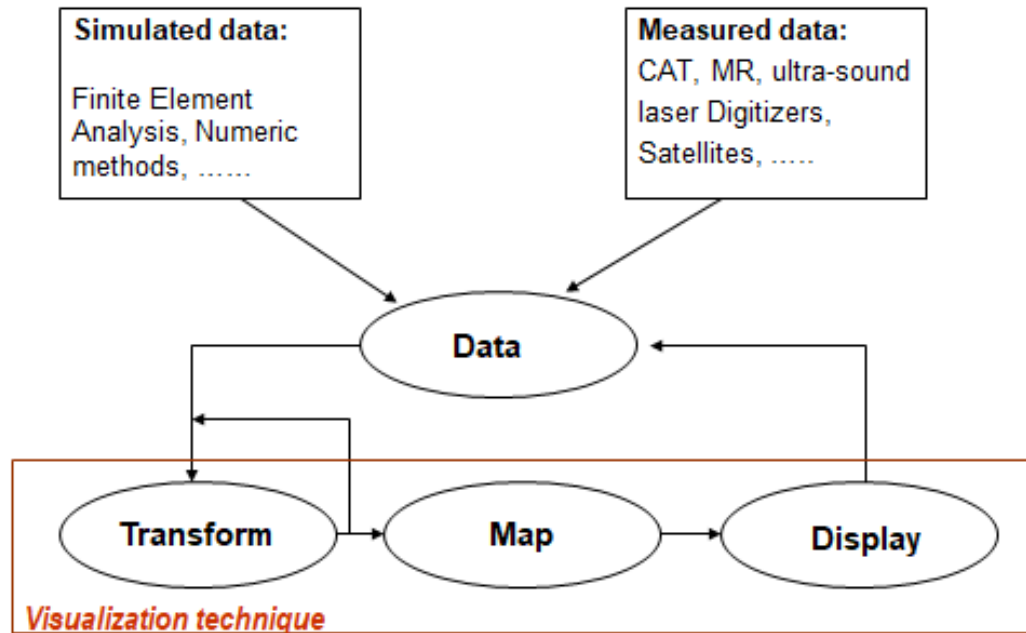




# Creating a Visualization



# Scientific Visualization reference model



(adapted from Schroeder et al., 2006)



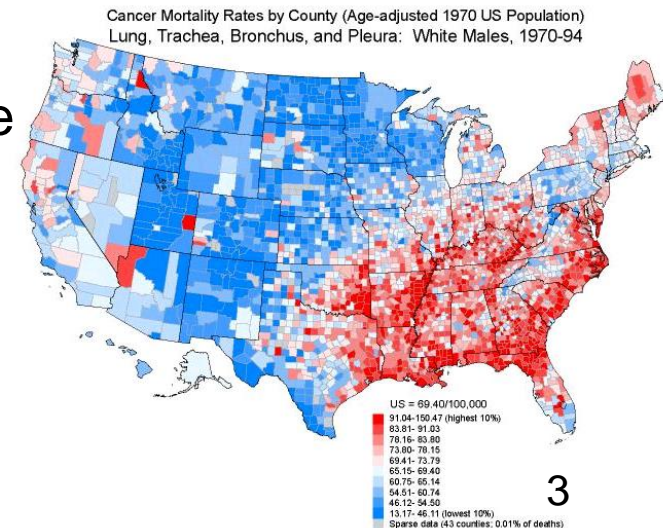
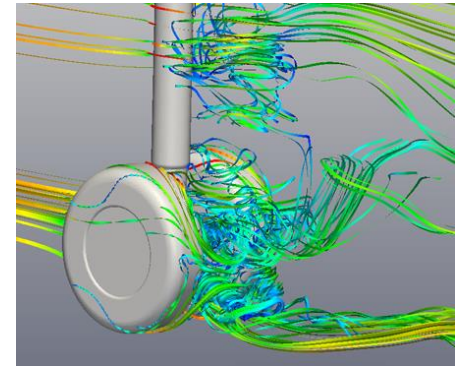
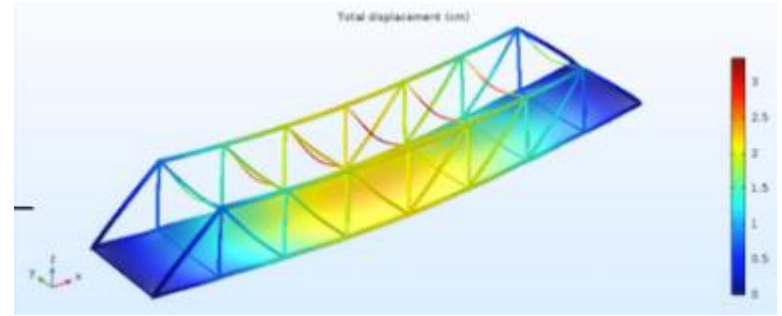
The visualization creator is involved in all phases

The user should get insights from the visualization



# Visual mapping

- It is necessary to decide:
  - which visual structures use to **represent** the data
  - their location in the display
- Some types of data can be easily mapped to a spatial location
- Examples:
  - . data with a topological or geographical structure
  - Abstract data don't have an easy correspondence with the dimensions of the physical space around us



Three **structures** must be defined in the **visual mapping/encoding**:

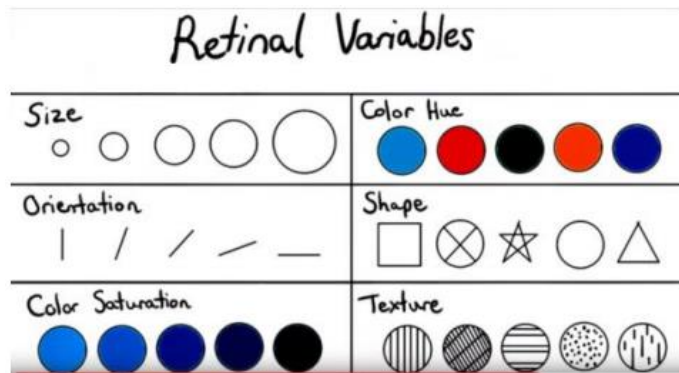
- spatial substrate
- graphical elements
- graphical properties

- **Spatial substrate** - dimensions in physical space where the visual representation is created (can be defined in terms of axes and type of data)
- **Graphical elements** - anything visible appearing in the space  
points, lines, surfaces, volumes
- **Graphical properties** – properties of the graphical elements to which the human retina is very sensitive - **retinal variables**:  
size, orientation, color, texture, and shape

- **Spatial substrate** axes (x, y, ...)  
type of data (quantitative, ordinal, categorical)

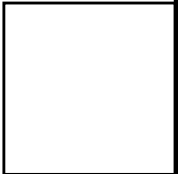
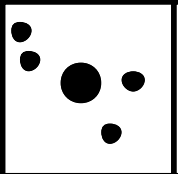
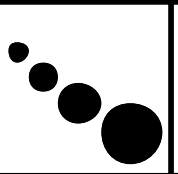
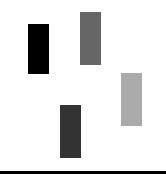

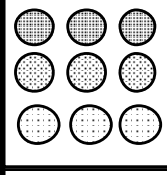
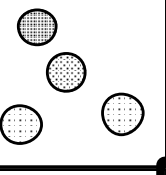
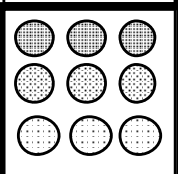
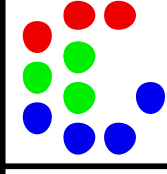
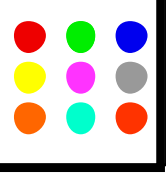
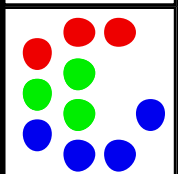
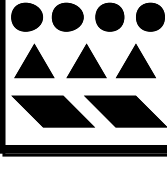
- **Graphical elements** points  
lines  
surfaces  
volumes

- **Graphical properties** retinal variables:



size,  
orientation  
color (depends on physiology and culture)  
texture  
shape

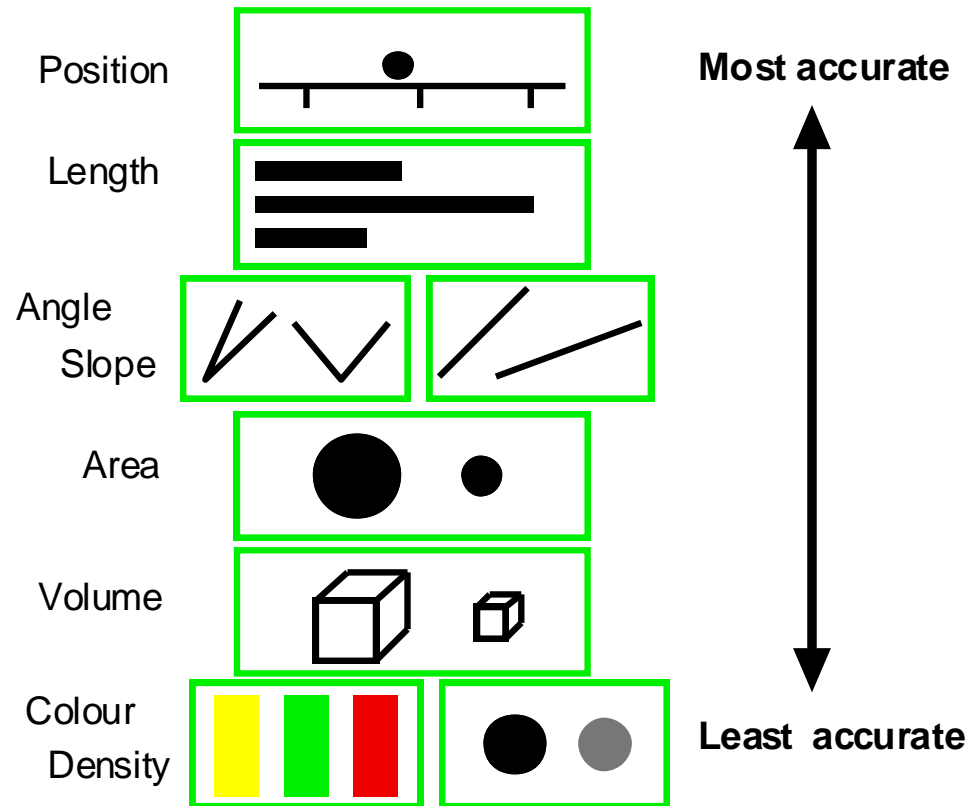
# How to select visual encodings?

	Association The marks can be perceived as SIMILAR	Selection The marks are perceived as DIFFERENT, forming families	Order The marks are perceived as ORDERED	Quantity The marks are perceived as PROPORTIONAL to each other
Size				
Value				
Texture				
Colour				
Orientation				
Shape				

Interpretation of Bertin's guidance regarding the suitability of various **encoding methods to support common tasks** (Spence, 2007)

Note that **only size is adequate to represent quantity accurately**

# How to select visual encodings to accurately represent quantity?



The relative difficulty of **assessing quantitative value** as a function of encoding mechanism, as established by Cleveland and McGill (Spence, 2007)

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

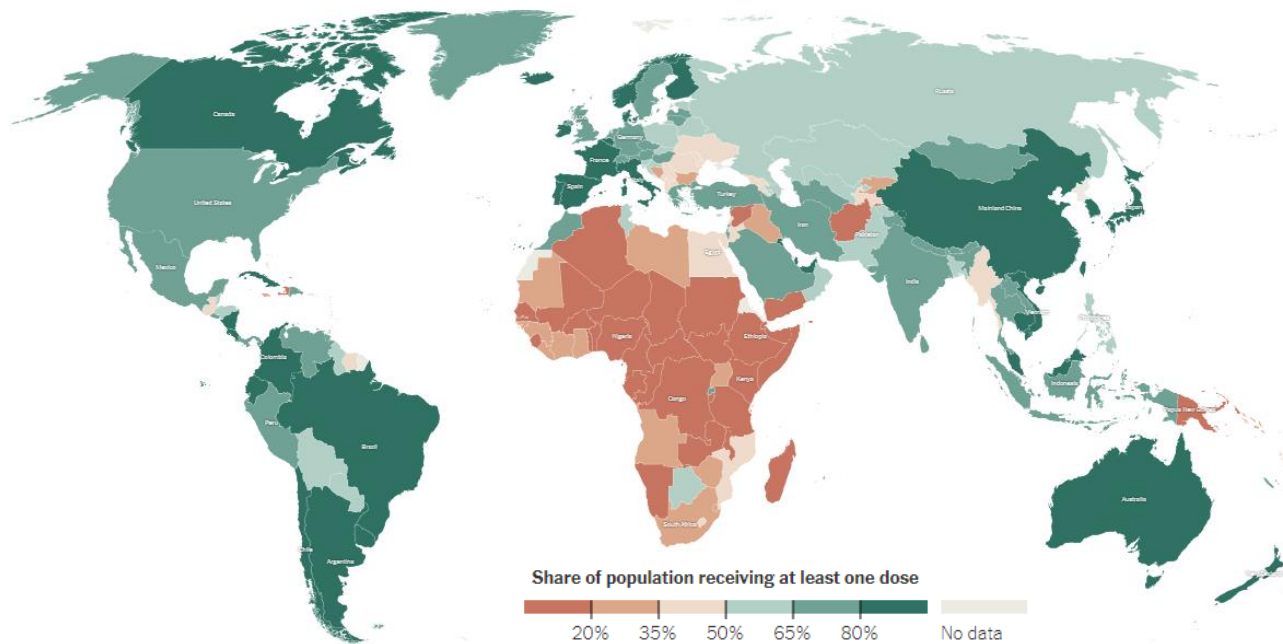
- Visualization may be the solution (or part of it)
- Creating a Visualization has several phases
- Visual mapping is core
- There are several possible visual encodings/  
visualization techniques
- But,

**How to select techniques ? → next topic**



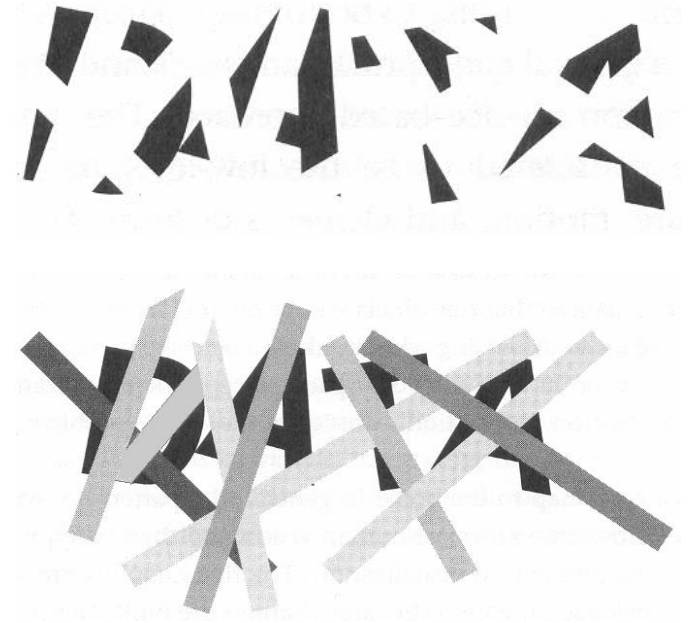


# Mapping - Visually encoding value



## 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

6970425934749  
3587282949546  
4244396854634  
2356658789376

[https://www.youtube.com/watch?time\\_continue=121&v=AiD6etOB6ql](https://www.youtube.com/watch?time_continue=121&v=AiD6etOB6ql)

Example: Count the number of 7s

69**7**0425934**7**49  
358**7**282949546  
4244396854634  
2356658**7**893**7**6

- Some visual attributes as **size**, **proximity** are also quickly processed by visual perception, **before the cognitive processes** come into play

Example:  
mapping numerical values  
to the length of bars:



(Mazza, 2009)

## Procedure to follow to create visual representations of data

1. Define the problem and the users' questions
2. Examine the nature of the data to represent and pre-process the data
3. Determine the number of attributes/variables/dimensions
4. Choose the visual structures to map

**test several ideas ...**

Important aspects to consider:

nature of the problem

- communicate
- explore
- confirm

nature of the data to represent

- quantitative
- ordinal
- categorical

number of attributes

- univariate
- bivariate
- trivariate
- multivariate

Next: visualization techniques organized according the n. of attributes

dataset types

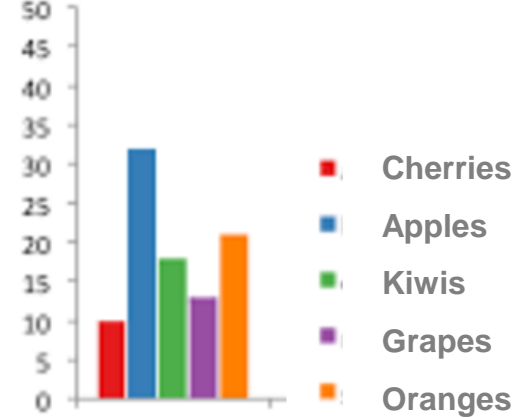
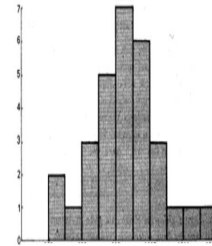
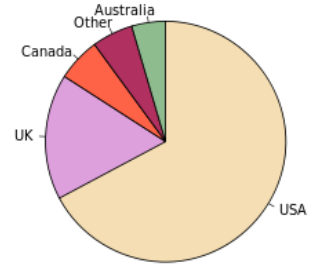
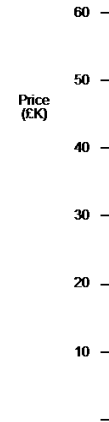
- tables
- networks
- spatial or geographical fields
- geometry

of tabular data

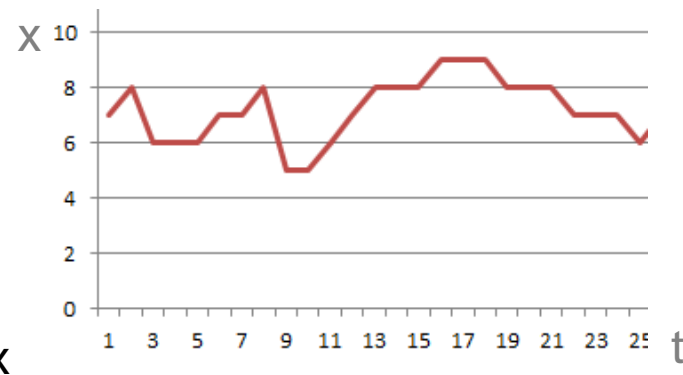
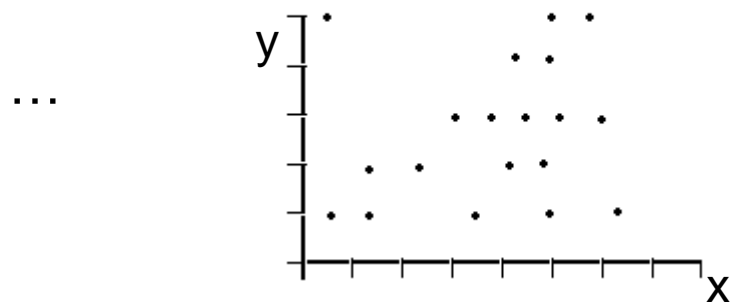


# Common Visualization Techniques to visually represent univariate, bivariate data

Univariate data   dot plot  
                         box plot  
                         bar chart  
                         histogram  
                         pie chart  
                         ...



Bivariate data   scatter plot  
                         line plot  
                         time series  
                         ...



# Representing univariate data

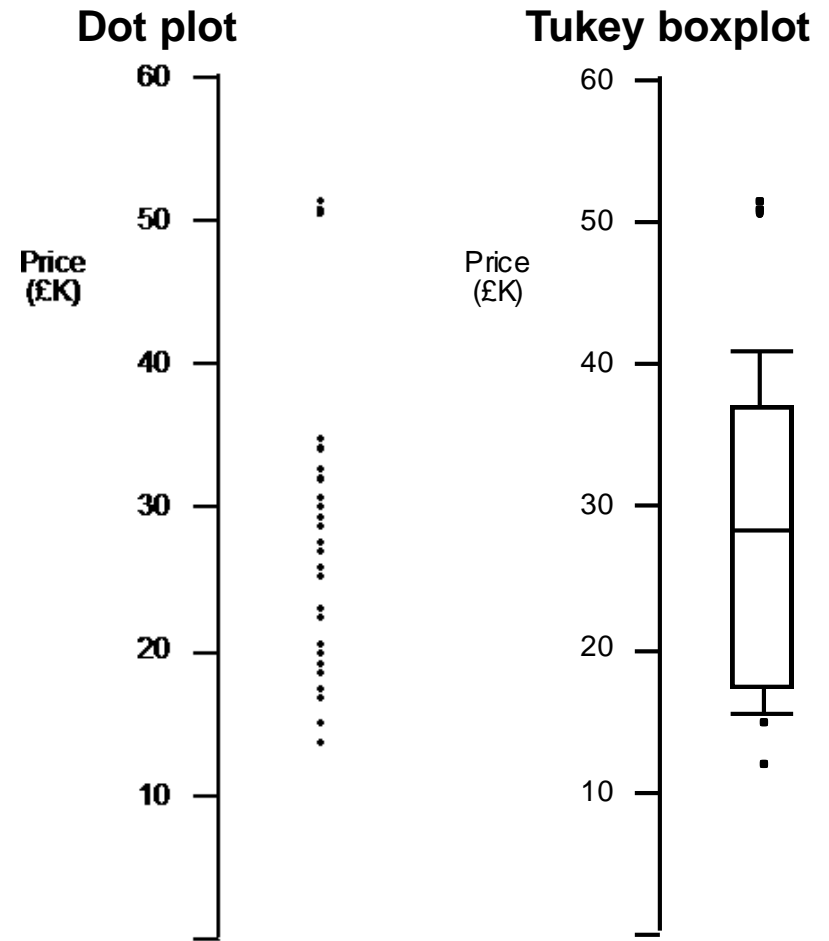
- A more common situation consists in representing a **set of values**
- Well established techniques exist
- But new ones can be invented!

Example:

Price for a number of cars:

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


(Spence, 2007)



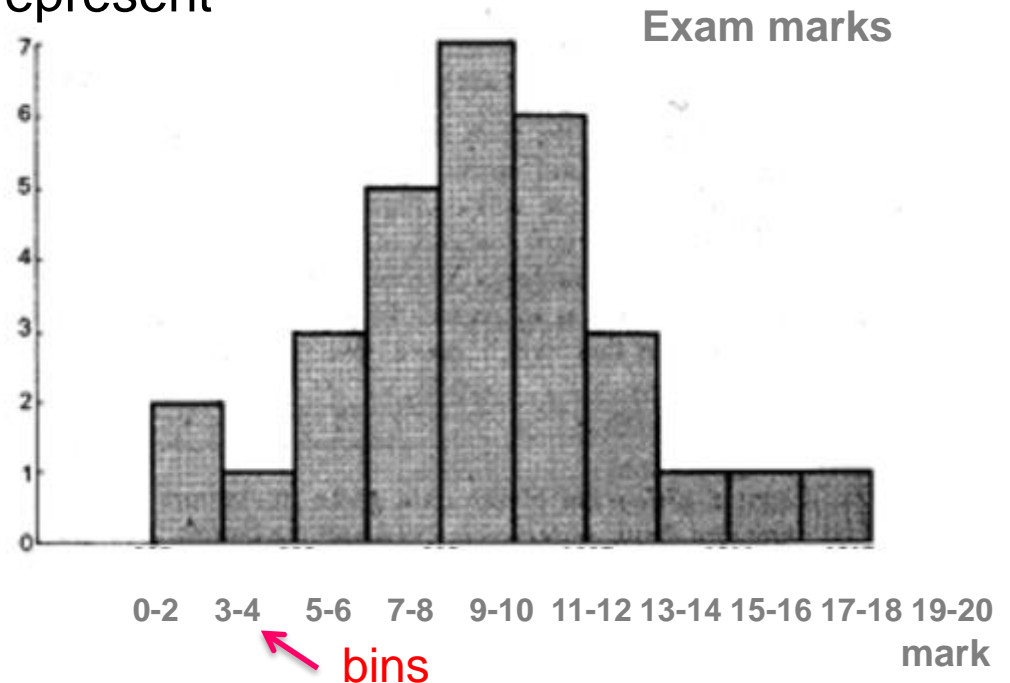
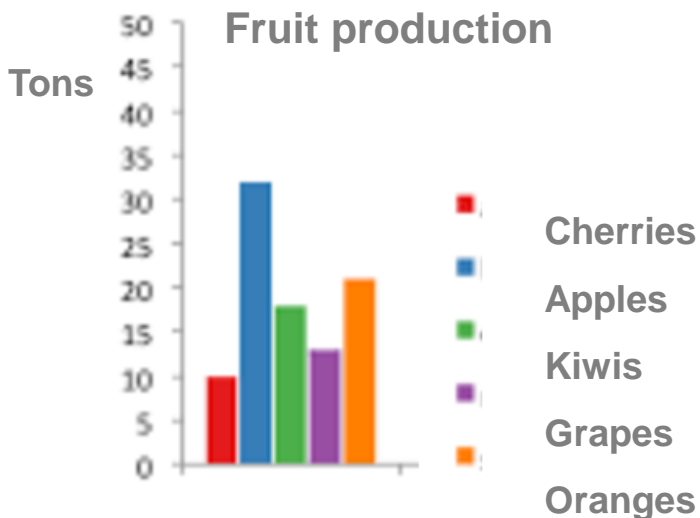


# Two common techniques not to be confounded !

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) technique

- 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

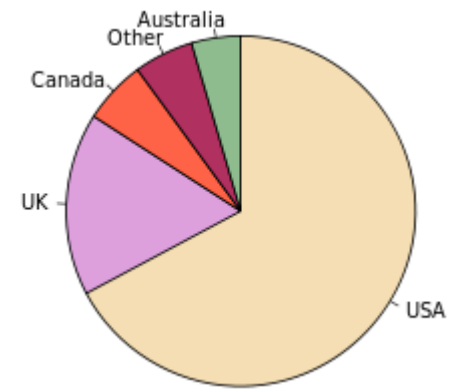
Are much controversial:

many experts recommend avoiding them

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

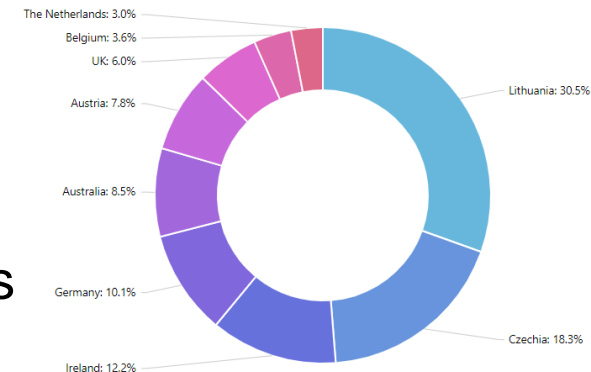
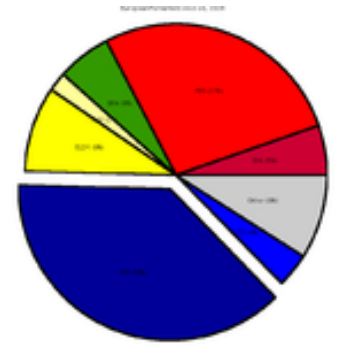


It is difficult to compare different sections of a pie chart, or to compare data across different pie charts



Native English speaking population

Variations of 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

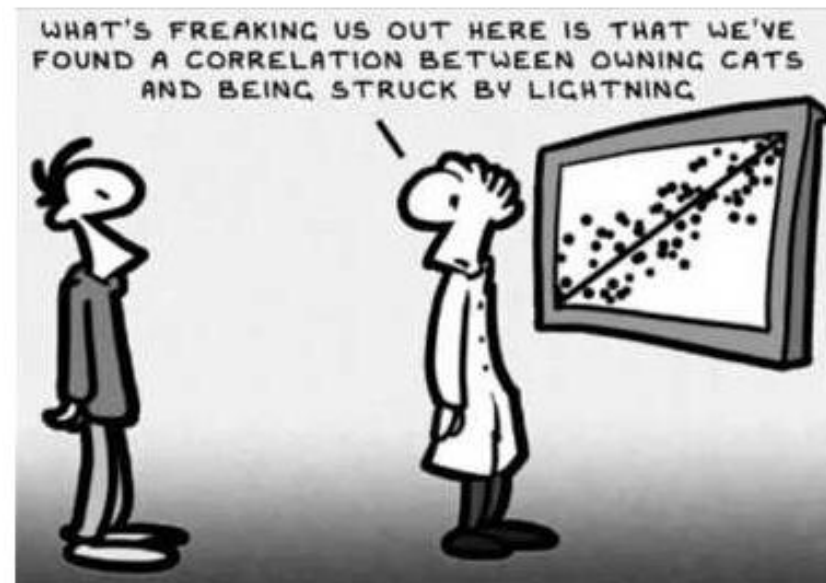
The axes are associated with these two attributes

This representation affords awareness of:

- general trends
- local trade-offs
- outliers



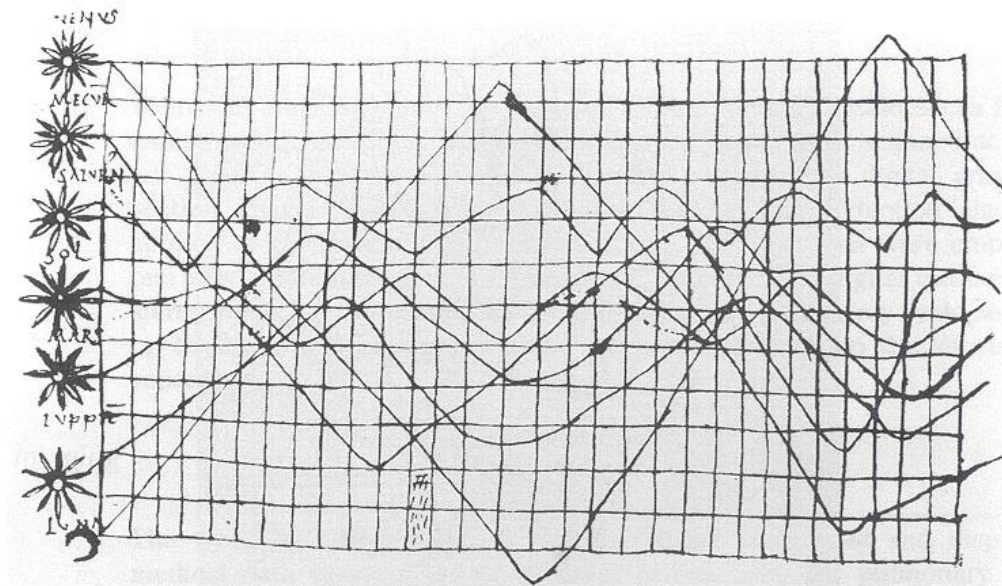
**Correlation is not causation**



# Representing bivariate data

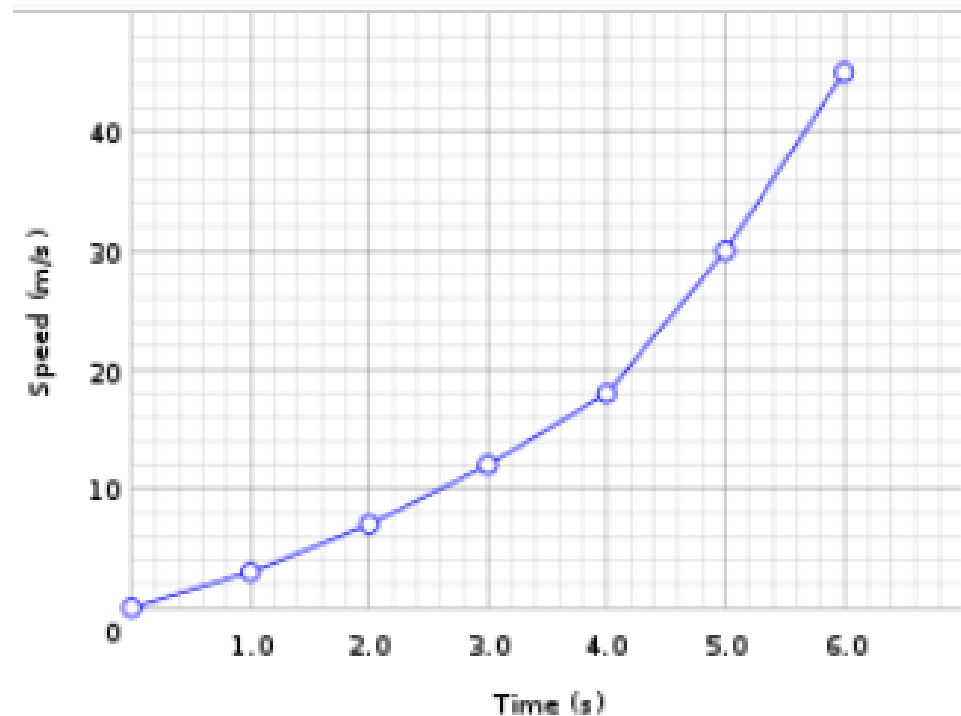
## The **line chart**

One of the oldest known and ubiquitous Visualizations



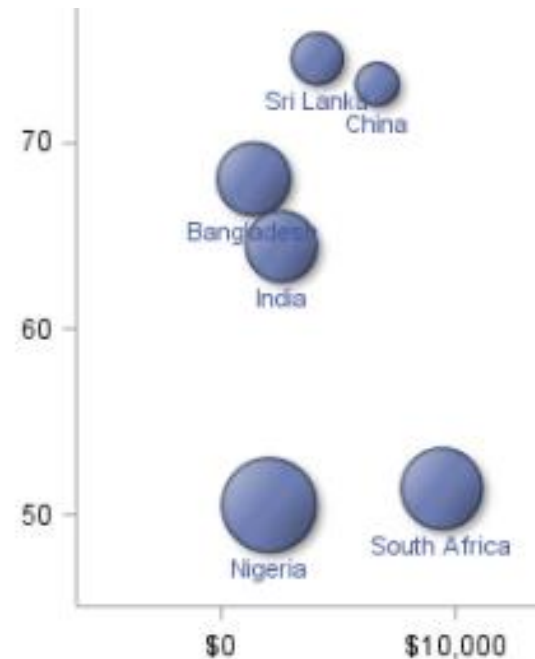
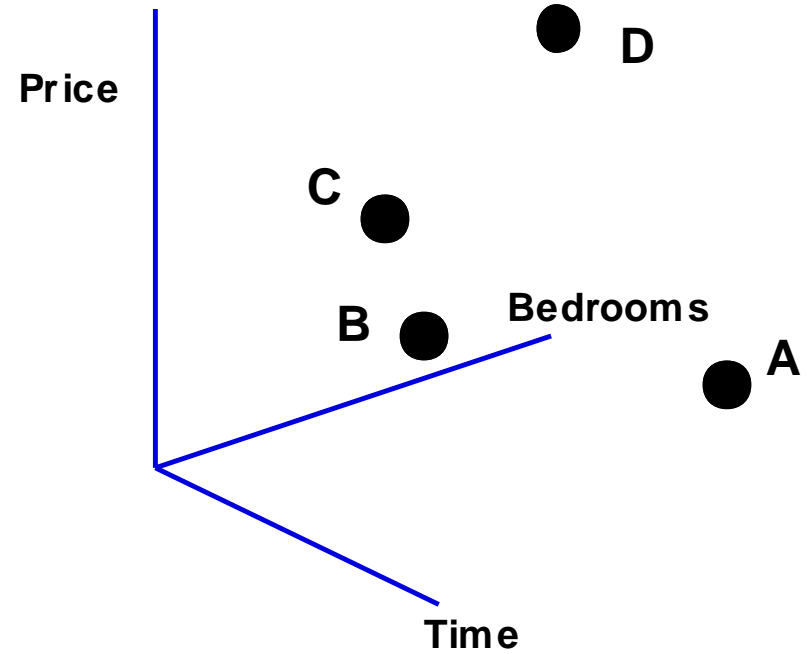
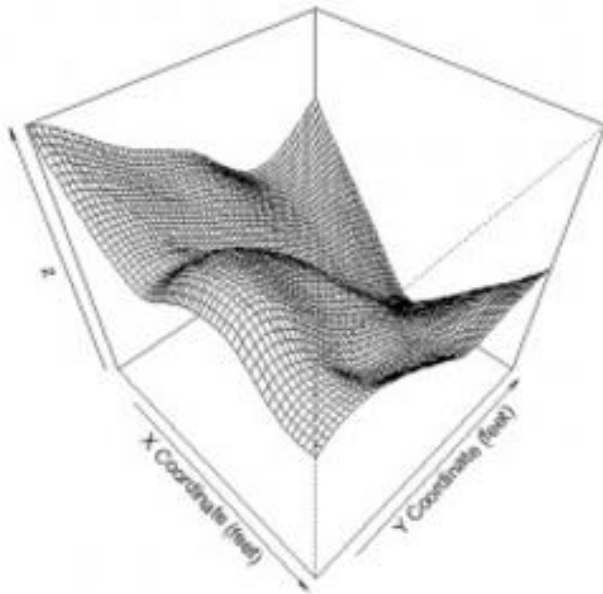
*Inclination of orbits along the time - Xth century (Tufte, 1983)*

- 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



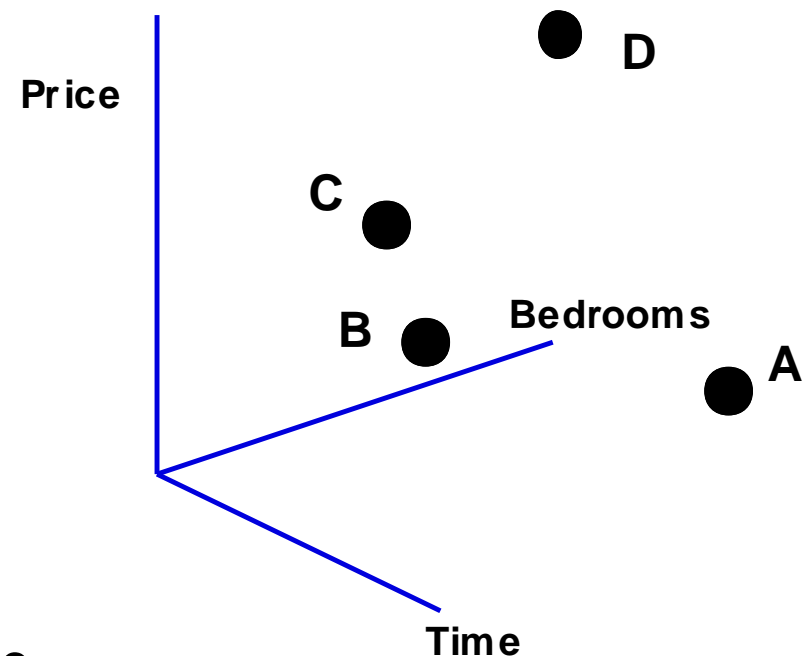
# Common Visualization Techniques to represent trivariate data

Trivariate data    surface plot  
                     contour plot  
                     3D representation  
                     bubble plot  
                     ...



# 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 of abstract data can be ambiguous
- This can be solved by interaction, allowing the user to reorient the representation

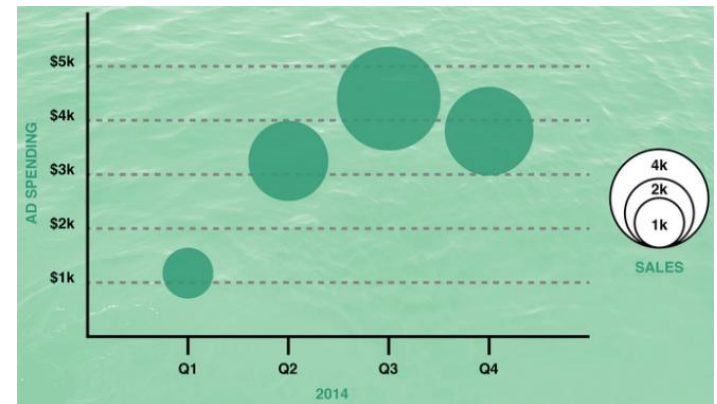


**“for 3D to be useful, you’ ve got to be able to move it”** (Spence, 2007)



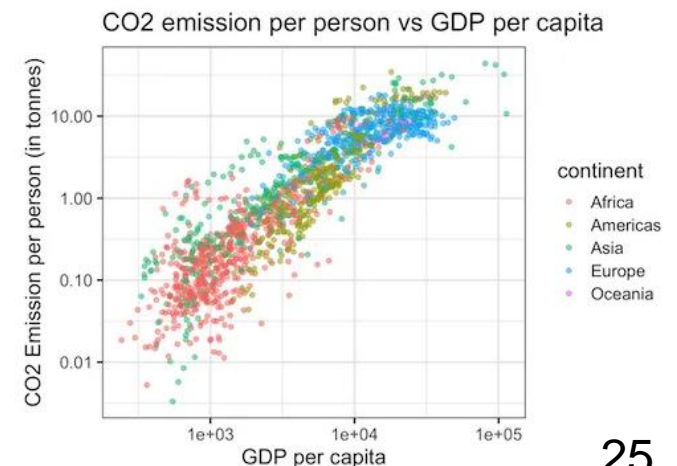
# Other Simple (and common) representations of trivariate data

- 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



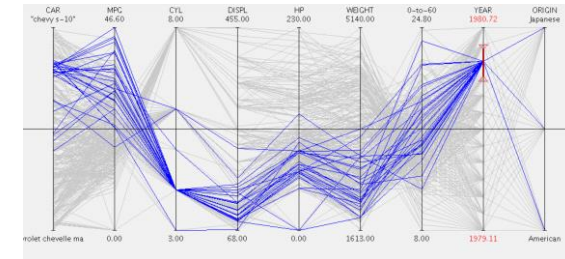
- Representing one more dimension through color

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

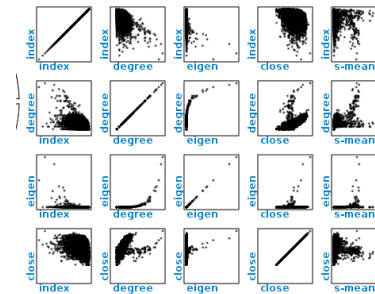


# Techniques for Multivariate (or Hypervariate) data

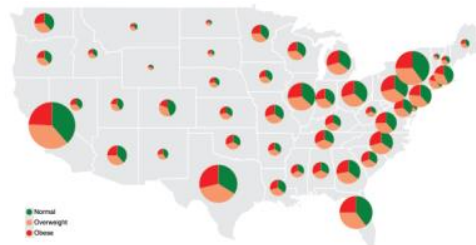
- Coordinate plots — parallel coordinate plots
- star (radar/spider) plots



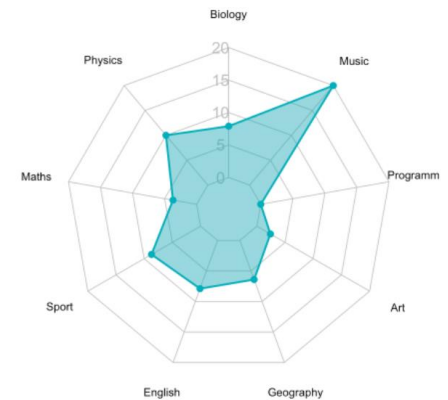
- Scatterplot Matrix



- Maps



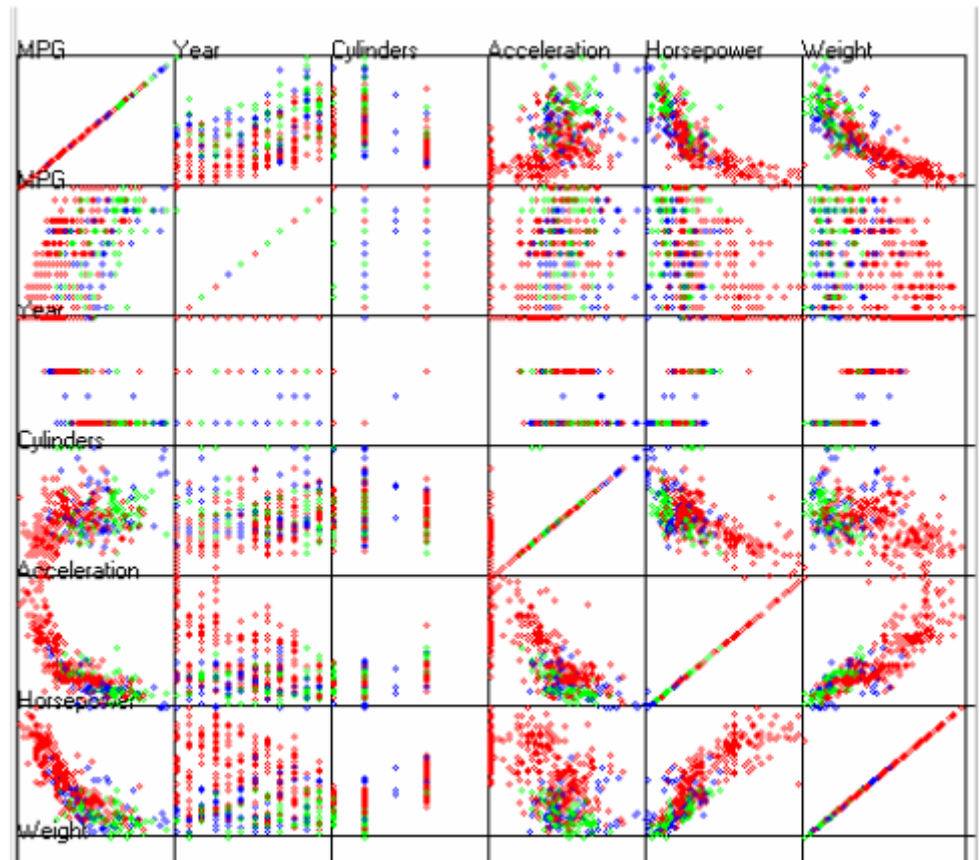
- Icons/glyphs



- The **scatterplot matrix** (SPLOM) is applicable to higher n. of variables
- However, as the number of attributes increase, the number of different pairs of attributes increases rapidly:

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

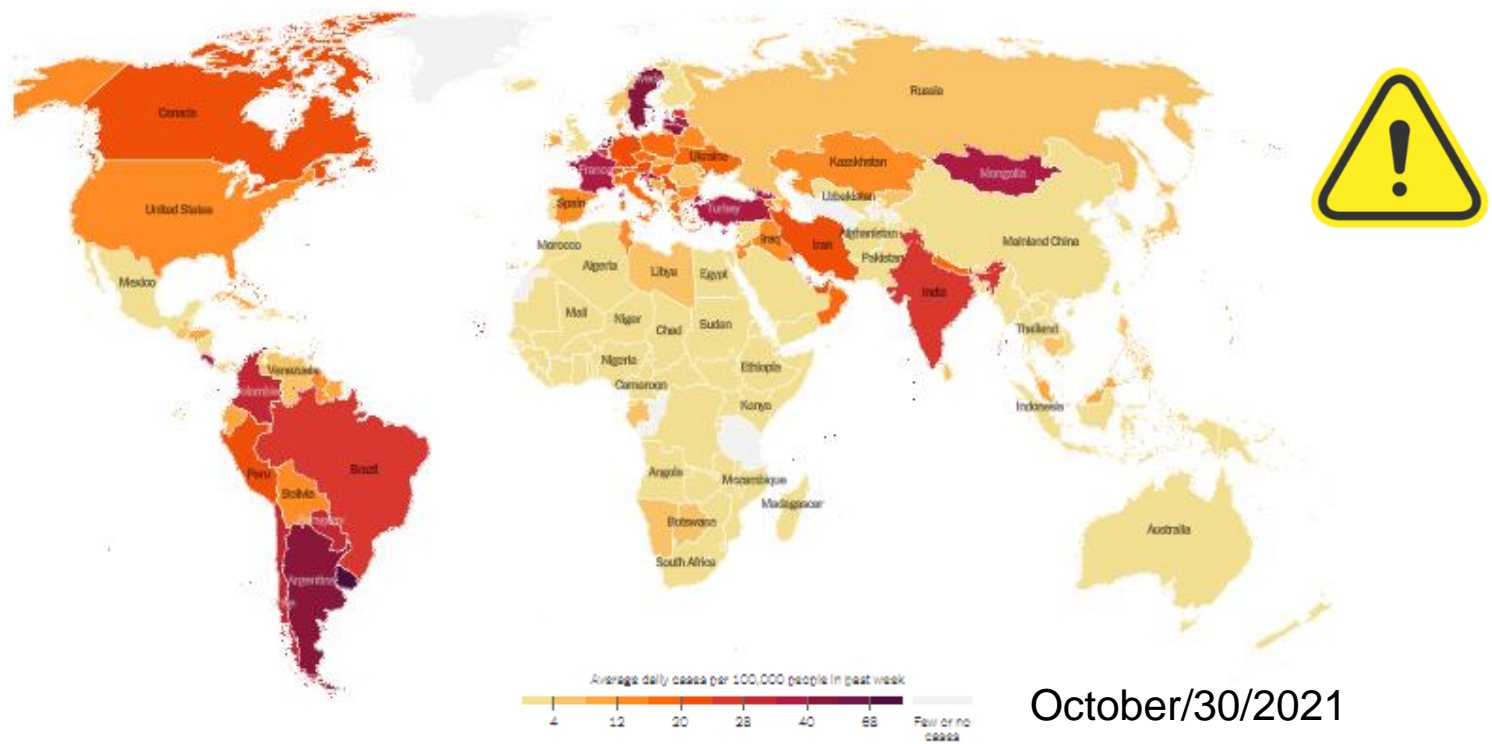
We may try to reduce the number of dimensions keeping the more relevant:  
**Dimensionality reduction!**



Scatterplot matrix for 6 attributes of a car dataset

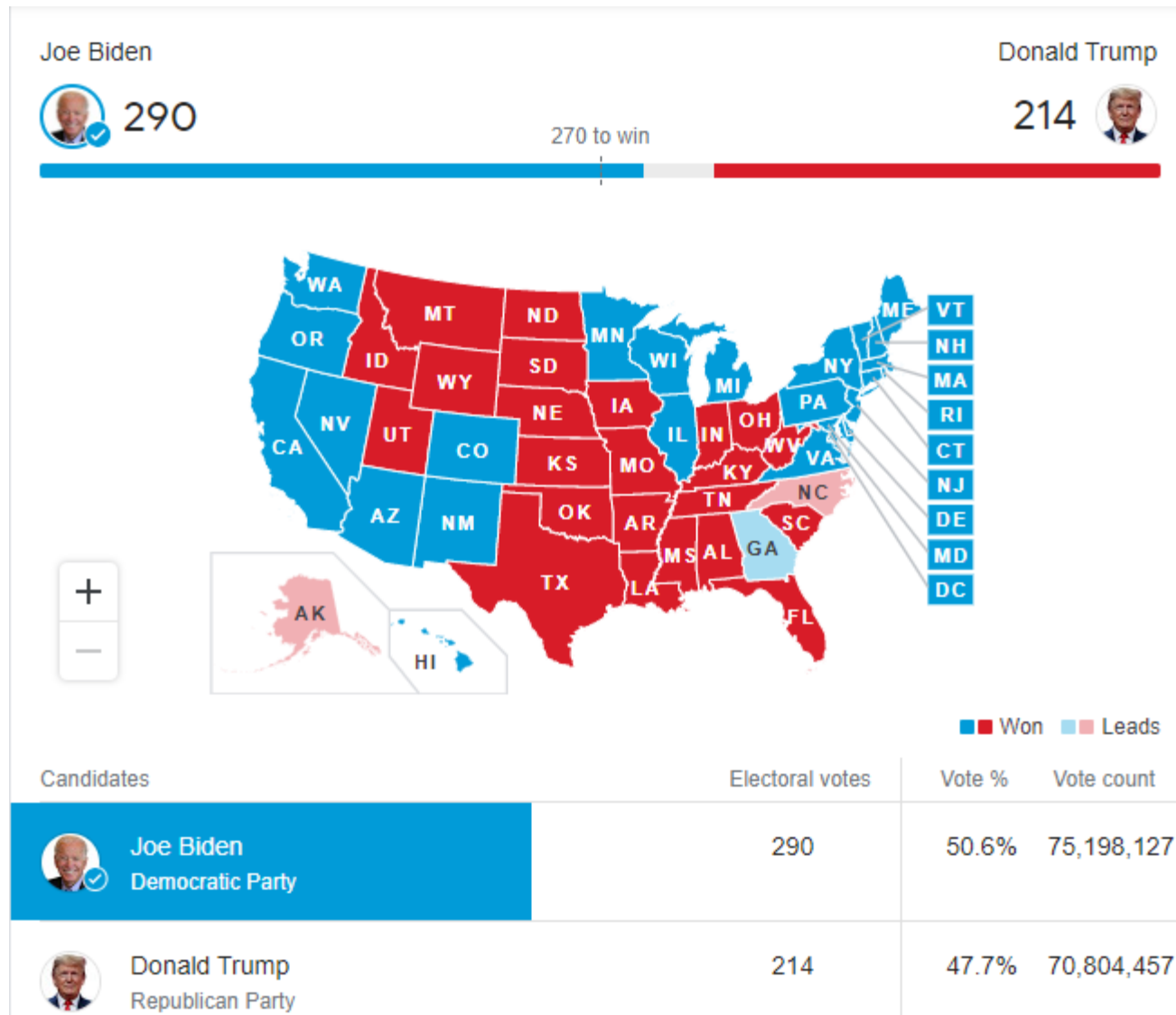
**Choropleth maps** - A standard approach to communicating aggregated data by geographical areas using color encoding of the geographic area

They require some care: what are the possible issues?



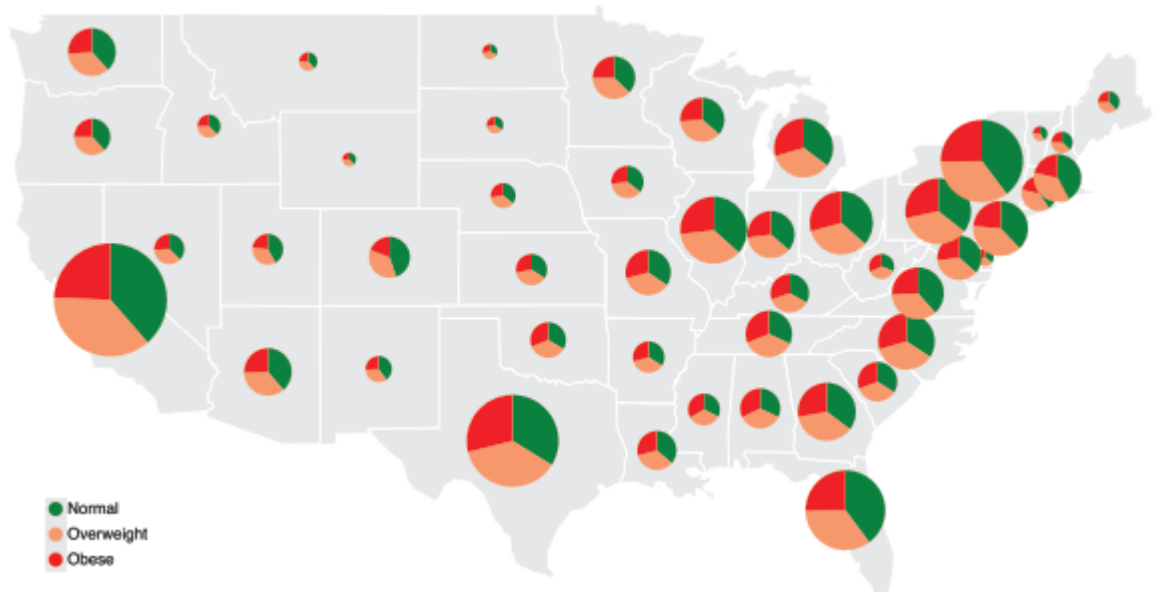
<https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html>

# Visualizations of the US 2020 Election (choropleth + bar) the bar helps better understand the ratio of votes



# Simple representations of attributes on a map

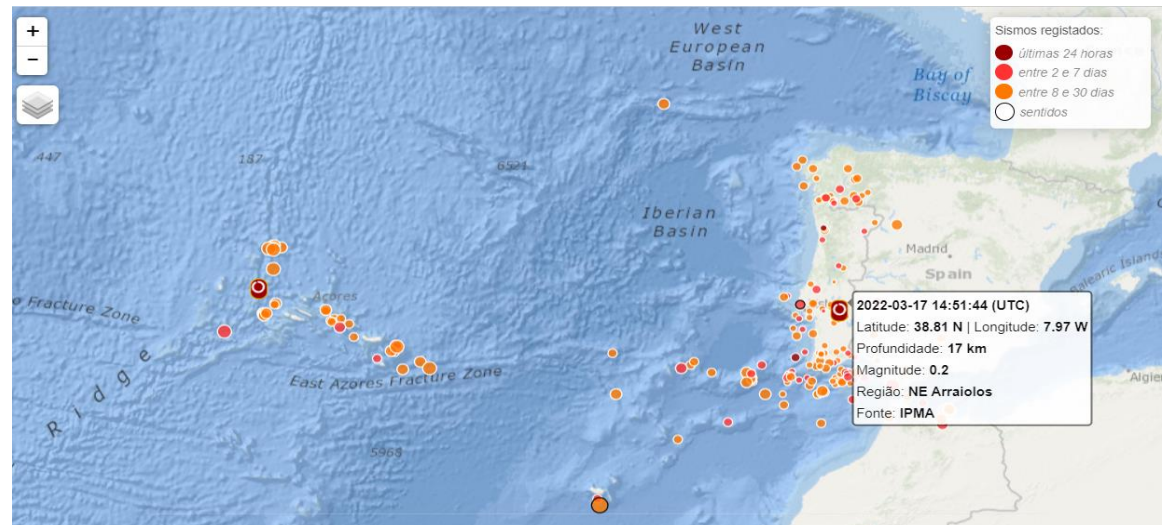
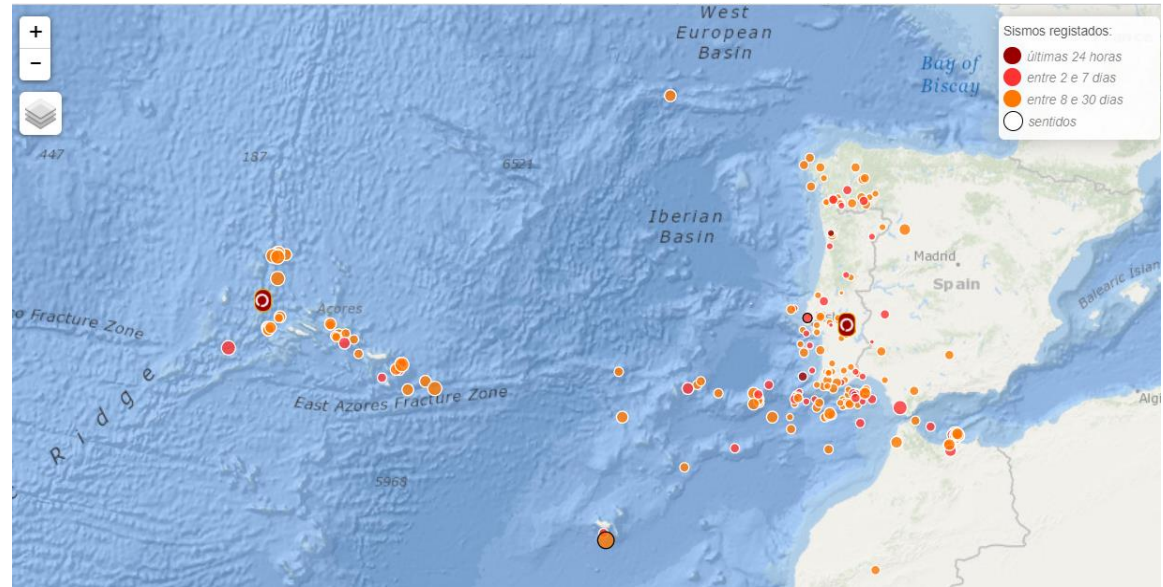
- **Graduated Symbol Maps** are an alternative to the choropleth map;
- Symbols are placed over an underlying map; may show more dimensions
- Avoid confounding geographic area with data values



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



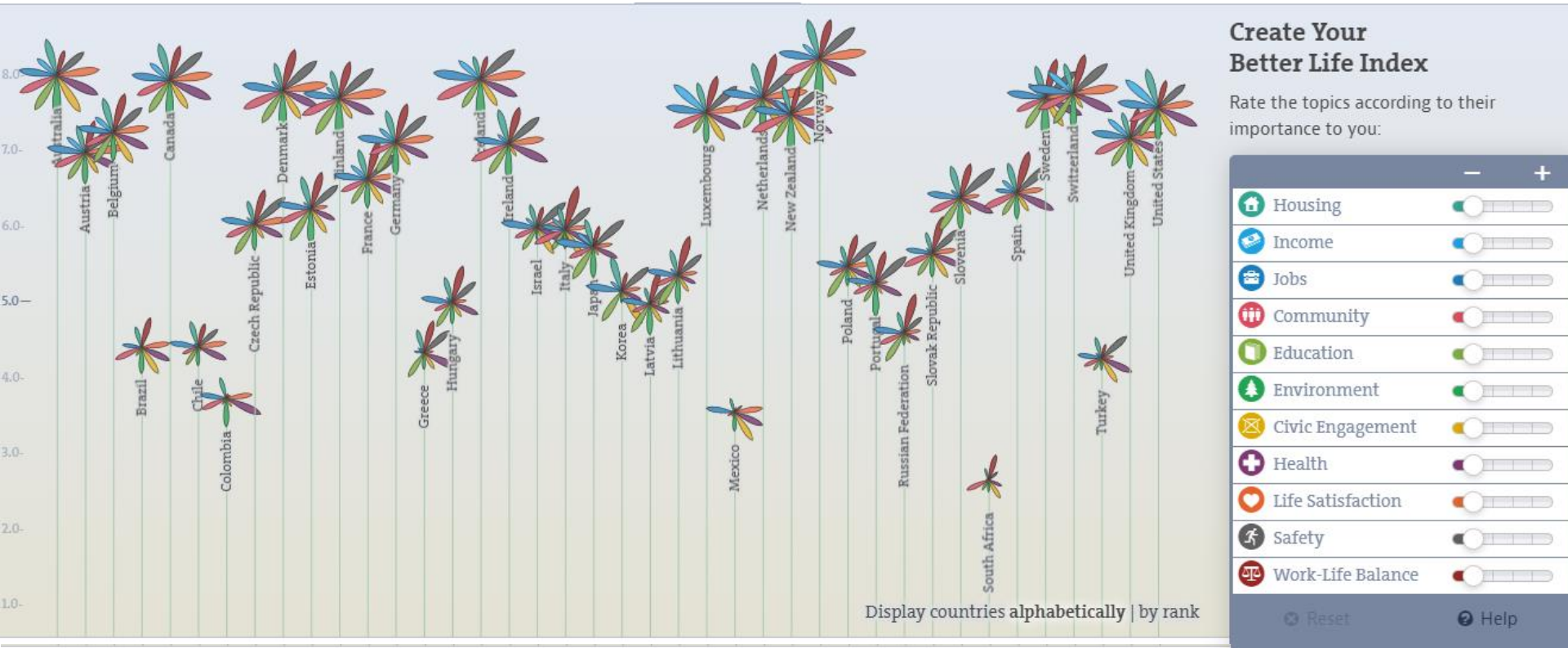
Seismic activity:  
Is something missing in  
this visualization?



<https://www.ipma.pt/pt/geofisica/sismicidade/index.jsp>

# Glyph chart example

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



(Kirk, 2012)

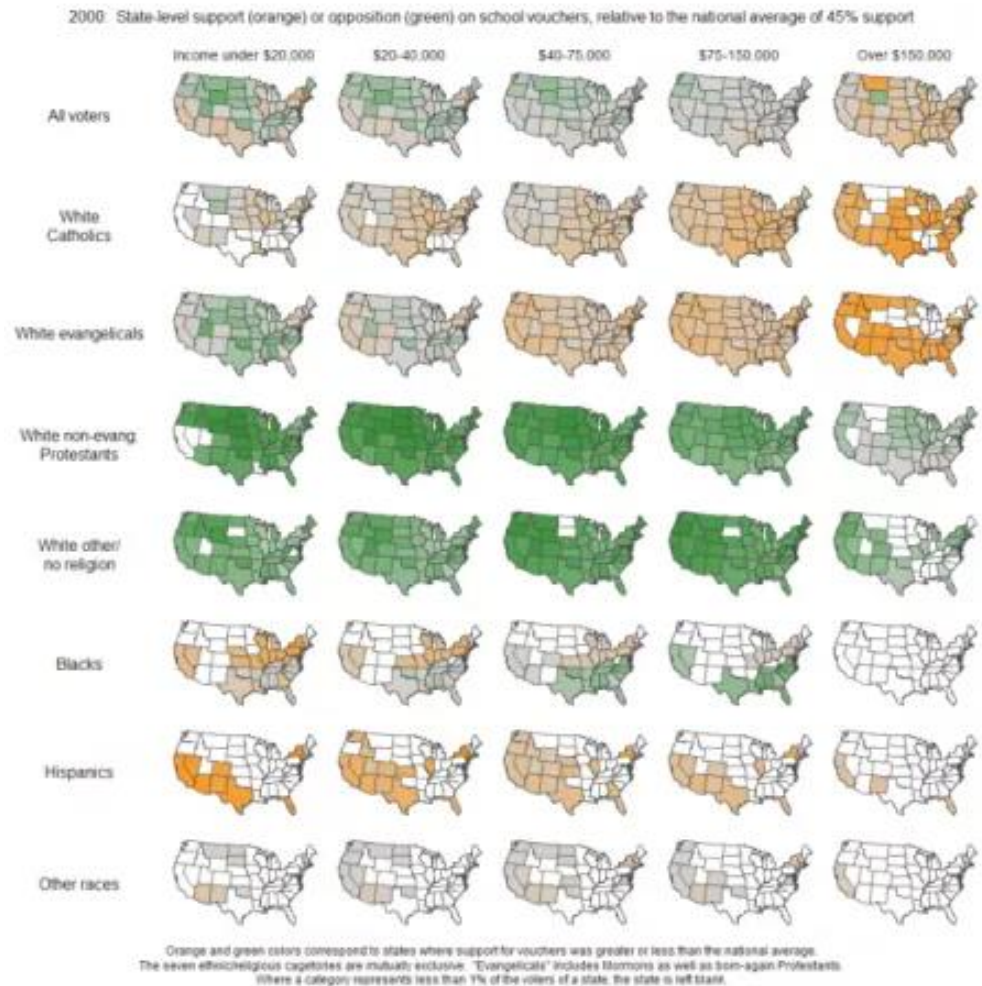
<http://oecdbetterlifeindex.org>



# Small multiples:

arrangement approach  
that facilitates efficient  
and effective  
comparisons

(Kirk, 2012)



**nature of the problem** / communicate  
/ explore  
/ confirm

**nature of the data to represent** / quantitative  
/ ordinal  
/ categorical

**number of attributes** / univariate  
/ bivariate  
/ trivariate  
/ multivariate

**Next: visualization techniques  
organized according the n. of  
attributes**

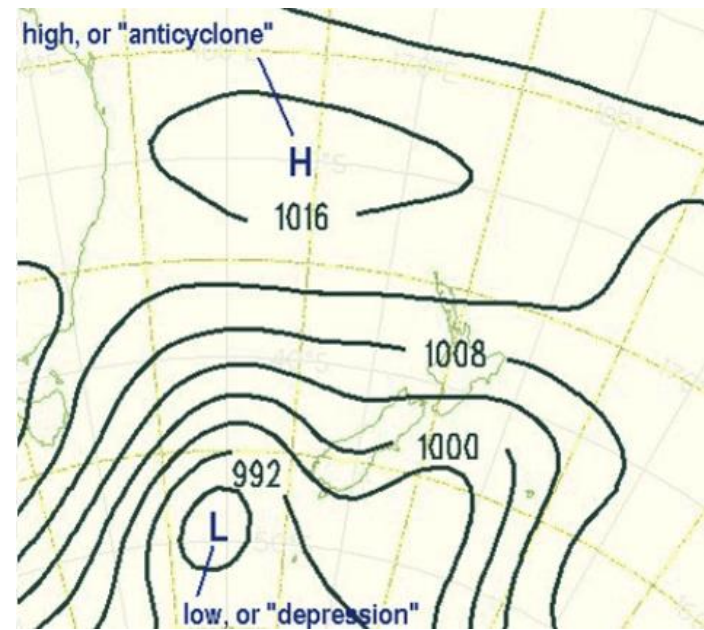
**dataset types** / tables  
/ networks  
/ spatial or geographical  
/ fields  
/ geometry

**of field and geometry data**

...

# Representations of a scalar in a 2D field

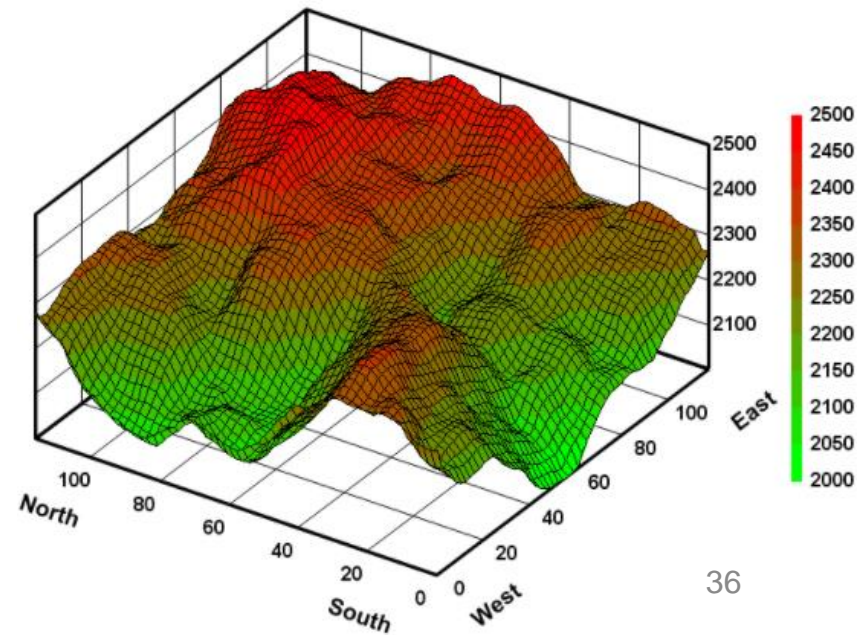
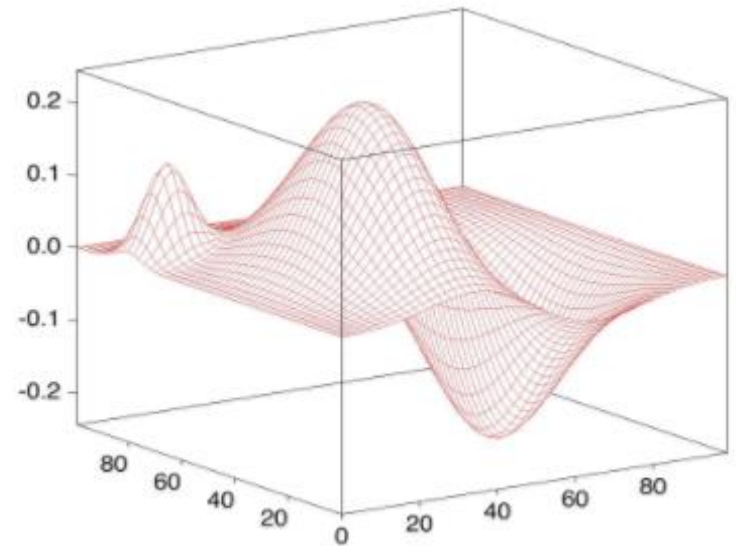
- Contour plots
- **contour line** (also **isoline**, **isopleth**, or **equipotential curve**) of a function of two variables is a curve along which the function has a constant value, so that the curve joins points of equal value.
- Typical in meteorological charts (isobars and isothermal curves)
-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



# Representations a 3D scalar function

- Surface plots
- May be combined with color

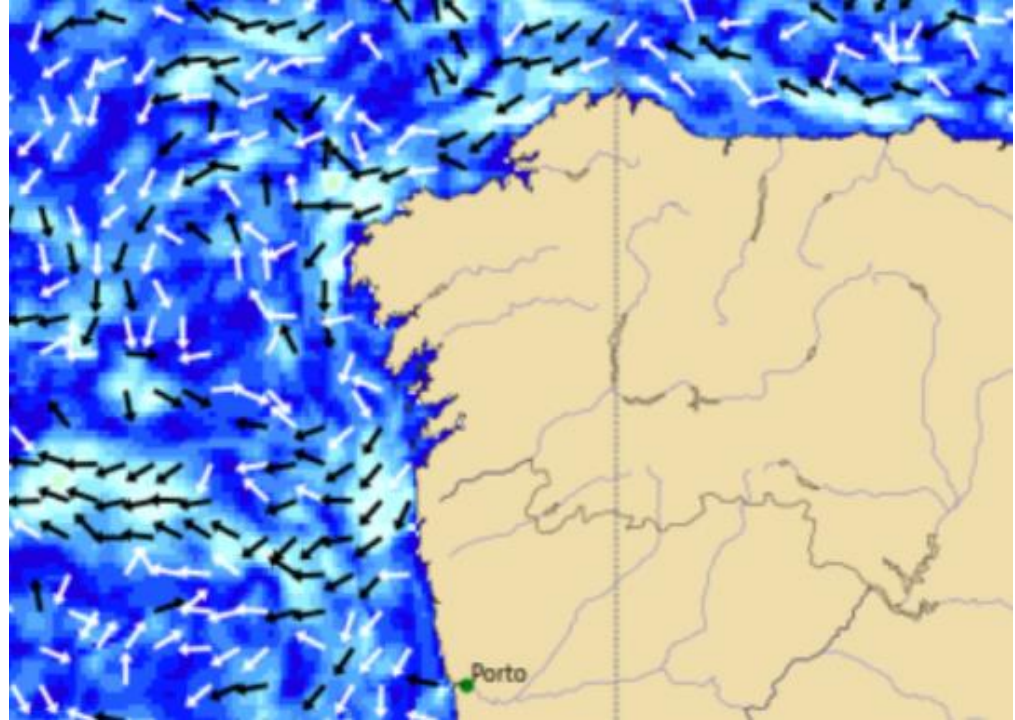
(preferably in a redundant way and carefully selecting the scale)



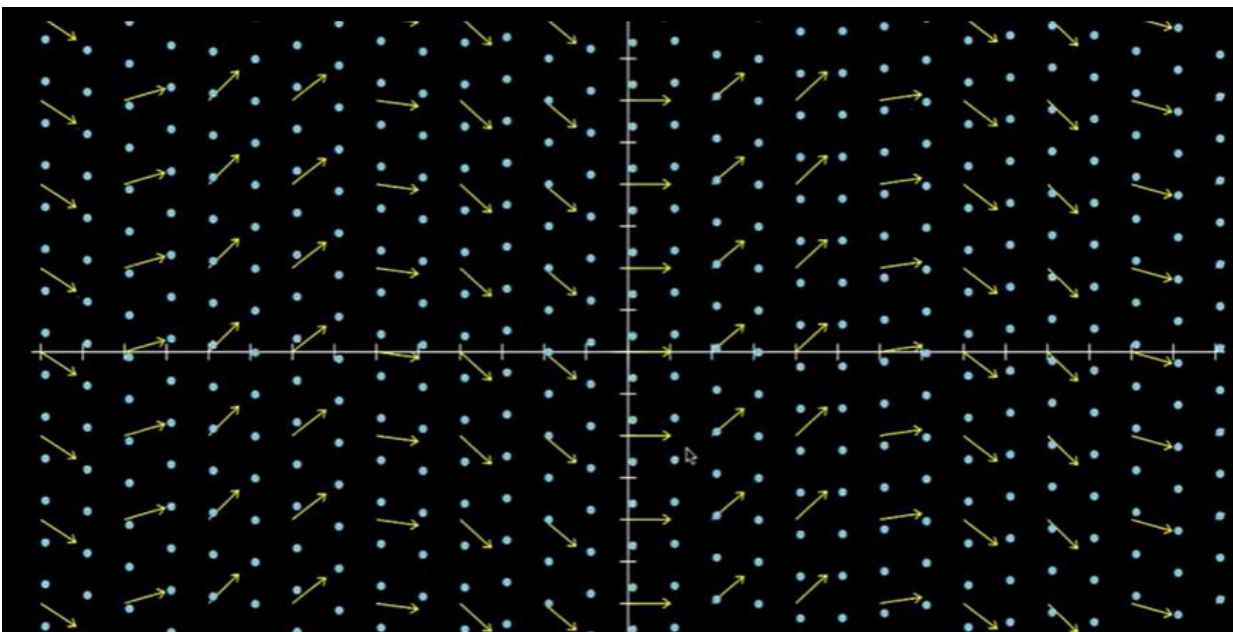


# Representations of vectors in a 2D field

- Representing sea currents at the surface →



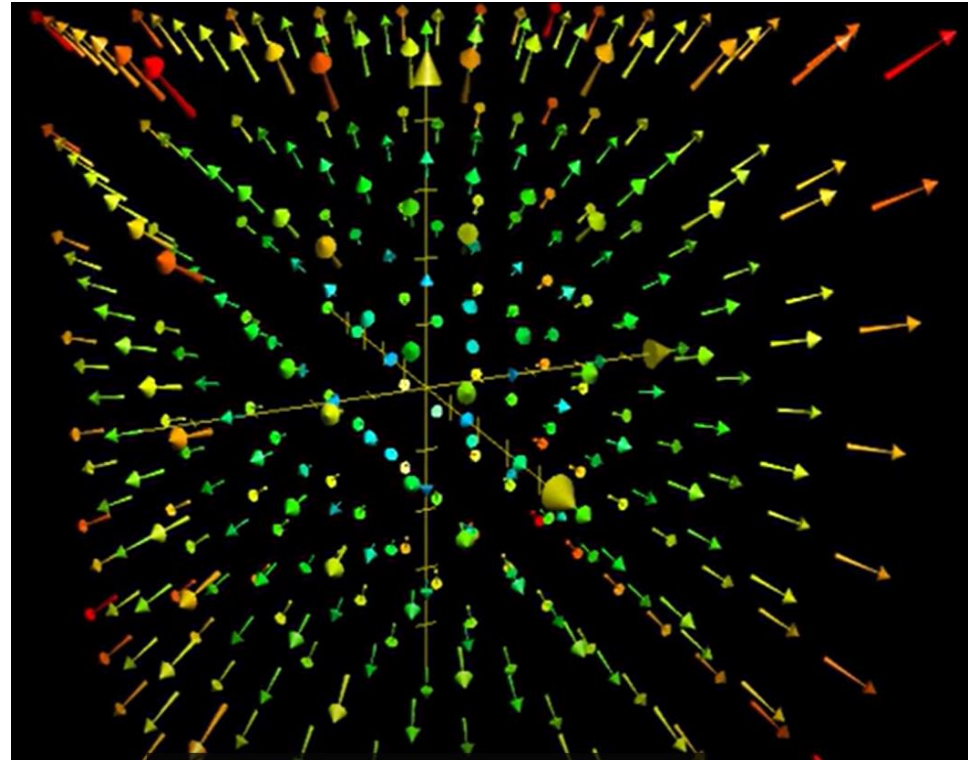
<https://www.ipma.pt/pt/maritima/currents/>



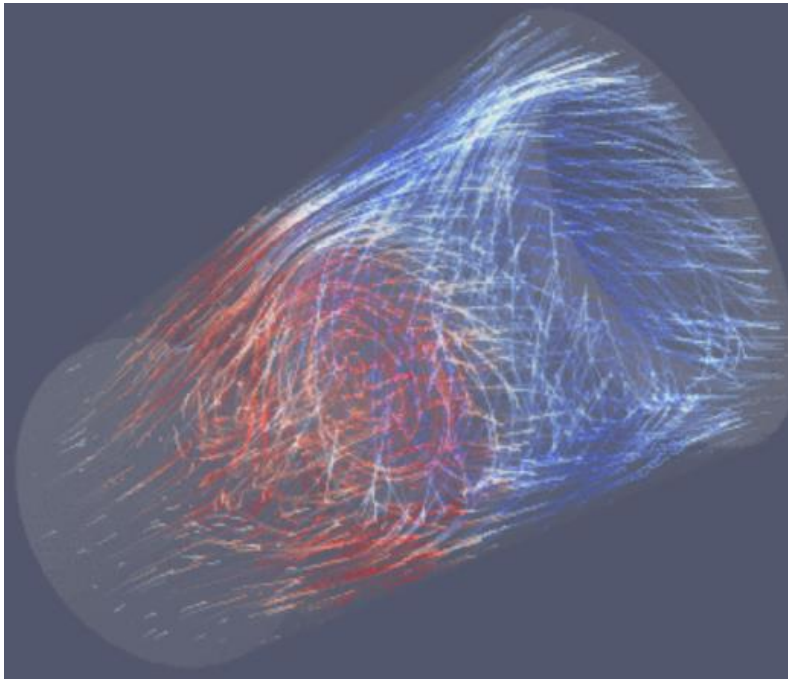
<https://www.khanacademy.org/math/multivariable-calculus/thinking-about-multivariable-function/visualizing-vector-valued-functions/v/fluid-flow-and-vector-fields>

# Representations of vectors in a 3D field

**Vector data visualization:** in 3D space a vector variable is visualized ➡



Stream Lines are another interesting representation



<https://www.khanacademy.org/math/multivariable-calculus/thinking-about-multivariable-function/visualizing-vector-valued-functions/v/3d-vector-field-example>

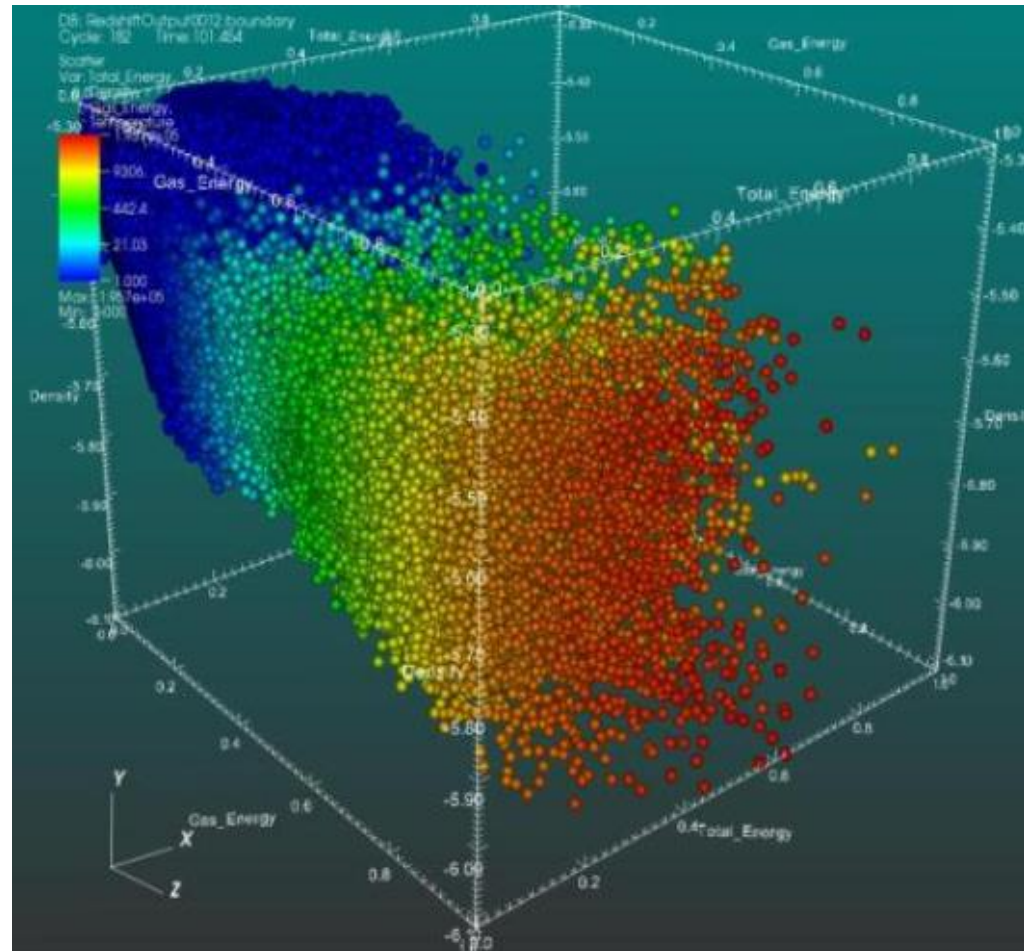
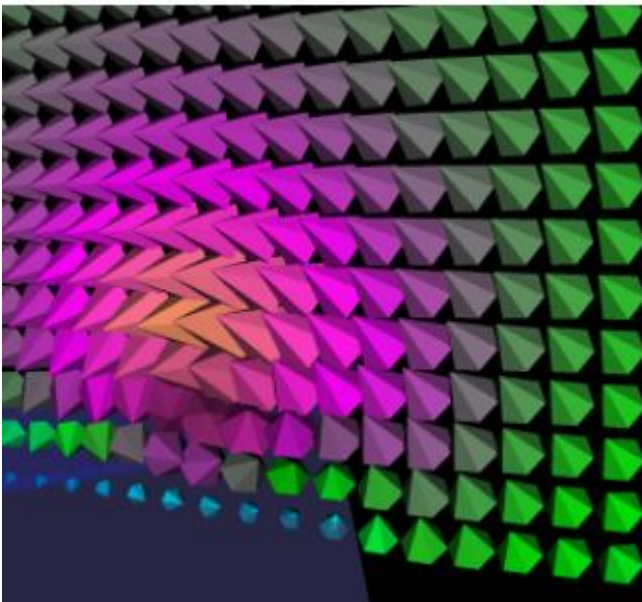
<https://www.kitware.com/new-animated-streamlines-representation-for-paraview-5-3/>



# Glyphs

**Four-dimensional data visualization:** in 3D space a fourth scalar variable is visualized using colored glyphs →

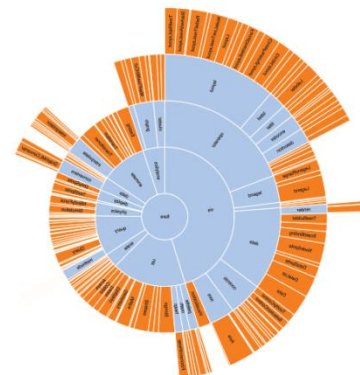
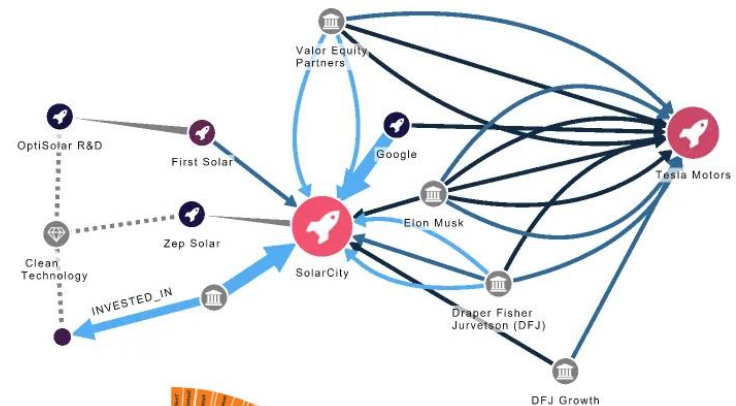
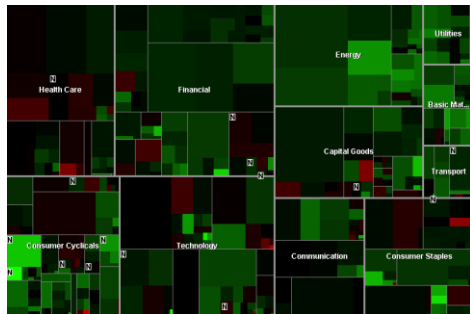
Glyphs for Visualizing a  
3D Vector Field



(Wikipedia/Glyph)

- These are only some of the visualization techniques to represent a value
- There are a lot more ...
- And we may visually want to visually represent beyond value: relation

- Networks
- Hierarchical data



<https://plotly.com/python/sunburst-charts/>



# Bibliography

- Spence, R., *Information Visualization, Design for Interaction*, 2nd ed., Prentice Hall, 2007
- Kirk, A., *Data Visualisation: A Handbook for Data Driven Design*, 2nd. Ed., Sage, 2019
- Mazza, R., *Introduction to Information Visualization*, Springer, 2009
- Ware, C., *Information Visualization, Perception to Design*, 2nd ed., Morgan Kaufmann, 2004
- Tufte, E., *The Visual Display of Quantitative Information*, 2nd ed., Graphics Press, 2001
- Heer, J., Bostock, M., & Ogievetsky, V. A tour through the visualization zoo. *Communications of the ACM*, vol. 8, n.1, 2010  
<https://doi.org/10.1145/1743546.1743567>