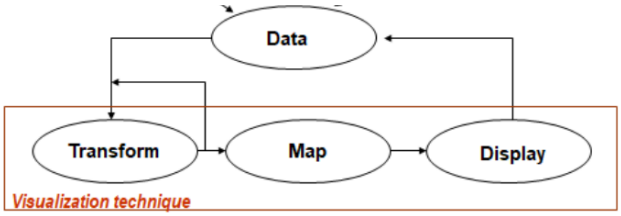
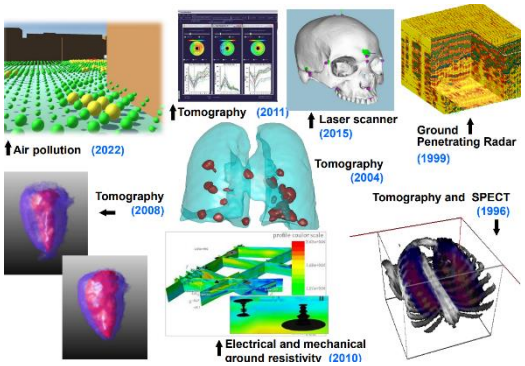
Visualização de Dados

Data(Scientific) Visualization (DV) -> dados com uma estrututa com espaço inerente.





The visualization creator is involved in all the phases after obtaining the data.

"**human-in-the-loop**" problems involve the user as a part of the system. They are very complex due to: humans are very complex systems; not well known; in general we cannot change them.

Data can be:

-**simulated** (finite element analysis, numeric methods, …);

-**measured** from real phenomena (CAT, MR, sensors, laser, …).

Then a **visualization technique** is applied, involving:

- data transformation through several methods;

- mapping to an adequate form to represent data visually;

- producing an image or sequence of images (rendering).

Different mappings or visualization techniques can be used:

- three-dimensional surface;

- pseudo-color;

-contours (isolines).

A visualization:

- Should allow **offload internal cognition and memory** usage to the **perceptual system**, using **carefully designed images** as a form of external representations (external memory).

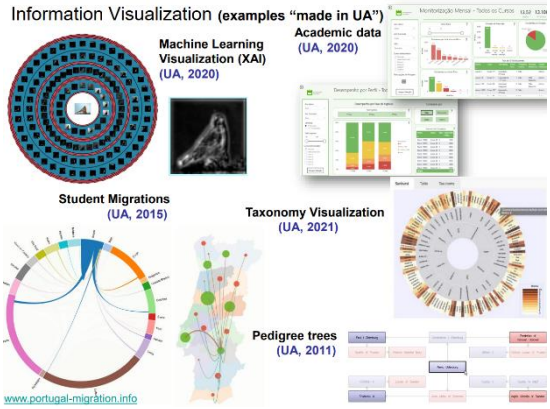
- To support users' tasks.

To design simple/complex visualizations:

- Need to find what are the questions users will ask!

Information Visualization (IV) -> dados tabelados "abstratos" sem espaço inerente.

Ambas(DV e IV) começam com (raw) data e permitem extrair informação.



Raw data -> is a set of information that was delivered from a certain data entity to the data provider and hasn't been processed yet by machine nor human.

Data

A data classification is important to:

- predict what visualization techniques are adequate;
- make easier the communication about the data;
- allow a more systematic approach to Visualization.

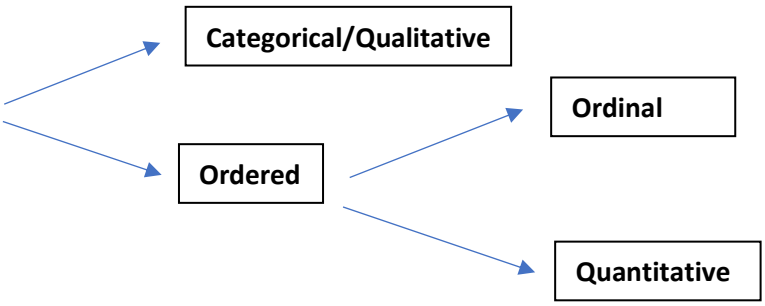
Data Abstraction

Four basic dataset types:

- tables;
- networks;
- fields;
- geometry.

Five basic datatypes:

- Items;
- **Attributes**;
- Links;
- Positions;
- Grids.



Measuring Scale:

- nominal (car brands, gender, animal species, …);
- ordinal (week days, preferences, …);
- interval (date, IQ, temperatures in ºC, …);
- ratio (temperatures in ºK, weight, height, …).

The **ratio scale** represents the **highest level or representation**, has a nonarbitrary zero (unlike the interval scale).

It is necessary to know the phenomenon behind the data as well as knowing the needs (questions) of the users!

Data preparation is very important and very time consuming.

Data integrity becomes more essential when the volume of data increases.

Cleansing Data

Data is dirty: it contains typos, inconsistencies, fails in some way to meet a standard.

Ex: birth date: fev/30/2000

Temperature: -300ºK

City: Lixboa

Transforming Data:

- **encoding**: answers to an open question need to be parsed and coded;
- **aggregation**: detail may be excessive (age: <18; 18-40; 41- 65; >65);
- **derived data**: add new relevant variables (T range = Tmax-Tmin);
- **removal**: remove data that are not needed;
- **standardization**: M/F; ºC or ºF.

Creating a Visualization

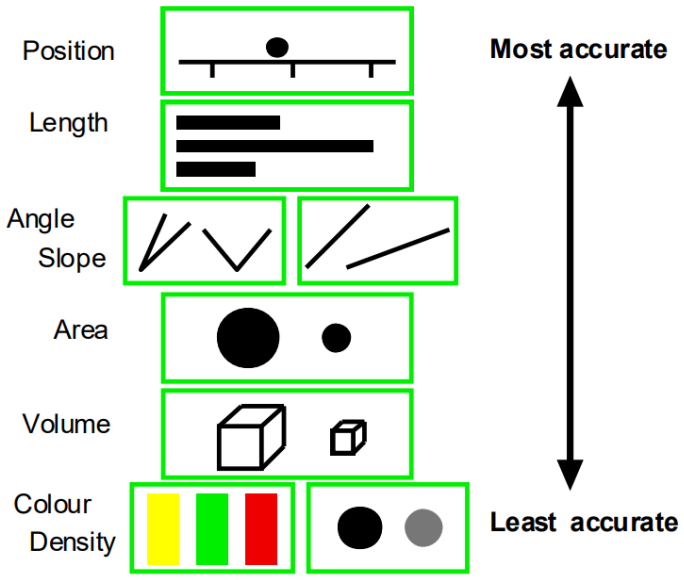
Three **structures** must be defined in the **visual mapping/encoding**:

- spatial substrate (dimensions in physical space where the visual representation is created); axes(x,y,z,…); type of data(quantitative, ordinal, categorical);

- 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 sensitivo - **retinal variables**: **size, orientation, color, texture and shape**)

PS: ONLY SIZE IS ADEQUATE TO REPRESENT QUANTITY ACCURATELY.



Nature of the Problem: communicate, explore, confirm.

Nature of the Data to Represent: quantitative, ordinal, categorical.

Number of Attributes: univariate, bivariate, trivariate,…

Dataset Types: tables, networks, spatial fields, geometry.

Univariate Data: dot plot, box plot, bar chart, histogram, pie chart,…

Bivariate Data: scatter plot, line plot, time series,…

Histogram: represents a distribution of numerical data.

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

Representing **bivariate** data

- The **scatterplot** is the convencional representation.

This affords awareness of:

- general trends;

- local trade-offs;

- outliers

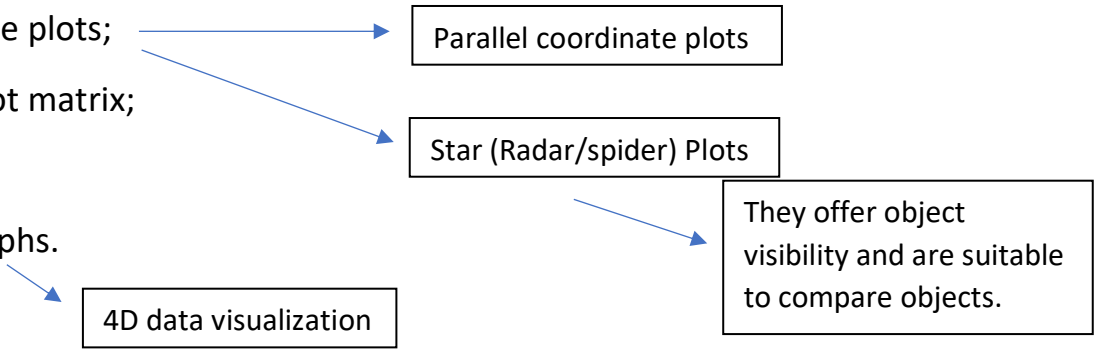
A **line chart/line plot/curve chart** displays information as a series of data points called “markers” connected by straight line segments.

Representing **trivariate** data

- “for 3D to be useful, you’ve got to be able to move it”

Representing **Multivariate** data

- coordinate plots;
- scatterplot matrix;
- maps;
- Icons/Glyphs.

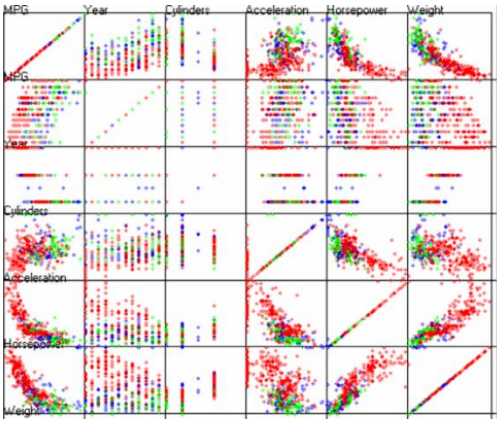


The **Scatterplot Matrix** is applicable to higher nº of variables.

2 variables -> 1 scatterplot

3 variables -> 3 scatterplots

4 variables -> 6 scatterplots.



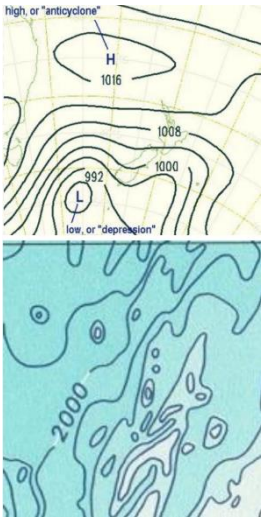
Dashboards -> visual display of the most important information needed to achieve 1 or + objectives.

Representations of a scalar in a 2D field

- contour plots;

- contour line (also isoline) of a function of 2 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 and maps.



Effective Visualization and Evaluation

Perception varies among people and with:

- context;

- experience.

2 Fundamental Principles to get graphical integrity:

- represent numbers, as physically measured on the surface of the graphic itself, directly proportional to the numerical quantities represented;

- clear and thoroughly label to defeat graphical distortion and ambiguity.

Guidelines to use color:

- Design first in black and white; then color apply sparingly and carefully;

- Use direct labeling instead of colors when you need to distinguish between more than about eight categorical items;

- Avoid large filled areas of overly saturated colors. They make it difficult for the reader to carefully inspect the figure;

- To make sure figures work for people with cvd, don't just rely on specific color scales. Instead, test figures in a cvd simulador;

- Do not use blue to color small objects (will be difficult to see).