Information Visualization

**Visualization** is the process of exploring, transforming and representing data as images (or other sensorial forms) to gain insight into phenomena. It includes not only image production from data, but also transformation and manipulation. Should allow offload interna cognition and memory usage to the perceptual system, using carefully designed images. Different areas:

* Scientific Visualization
* Data Visualization – data having an inherent spatial structure
* Information Visualization – data not having an inherent spatial structure

Contributes to Visualization:

* Computer Graphics
* Human-Computer Interaction
* Software Engineering
* Image Processing
* Signal Processing

Visualization deals mainly with multi-dimensional data, data transformation is fundamental and i tis is essentially interactie.

Benefits:

* Helps us think
* Reduces load on working memory
* Offloads cognition
* Uses the power of human perception

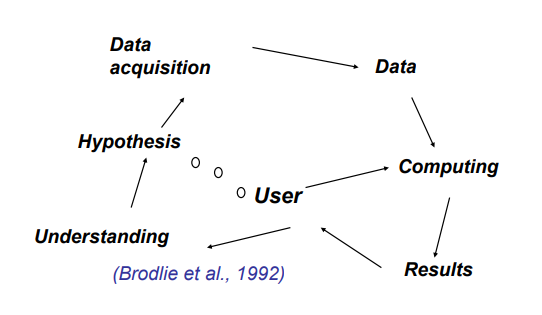
Applications of Data Visualization:

* Medicine
* Meteorology, climatology, oceanography
* Fluid Dynamics
* Cosmology

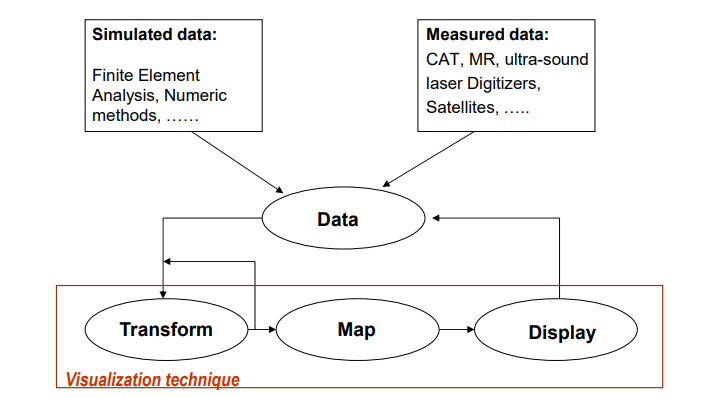
Purposes:

* Personal exploration
* Discussion with colleagues
* Presentation to other people
* Explorative analysis
* Confirmative analysis

**Human-In-The-Loop** problems involve the user as part of the system. Hey are complex because:

* Humans are very complex systems
* Not well known
* In general we cannot change them

**Ascombe quartet** are data sets with same simple statistical model.



**Data visualization process**

Data can be:

* Simulated
* Measured from real phenomena

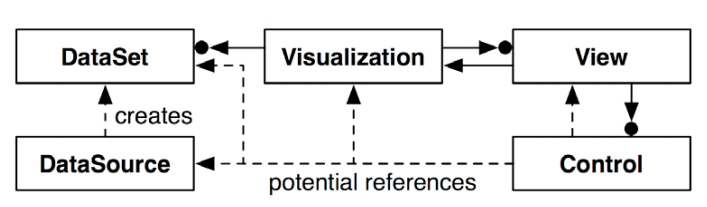
Must be then prepared

And then a visualization technique is applied, involving:

* Data transformation through several methods
* Mapping to an adequate form to representation
* Producing an image or sequence of images (rendering)

Repeat this process as needed.

**Visualization Applications**



Information visualization application development requires balancing:

* Data management
* Visual mappings
* Computer graphics/views
* Interaction

**How can we produce a Visualization?**

* There are principles derived from human perception and cognition
* There are methods to approach the process of Visualization
* A correct definition of the users and goals is fundamental to efficacy

**Why show the data in detail?**

Visualization helps in situations where seeing the dataset structure in detail is better than seeing only a brief summary of it (loosing information).

**Effective Visualization**

Implies saying the truth about the data.

**When are Visualization solutions appropriate?**

* To analyze data when people don’t know exactly what questions they need to ask in advance
* For long-term use, when a human intends to stay in the loop indefinitely
* For long-term use to monitor a system, so that people can take action if they spot unreasonable behavior
* For transitional use where the goal is to “work itself out of a job”, by helping the designers of future purely computational solutions

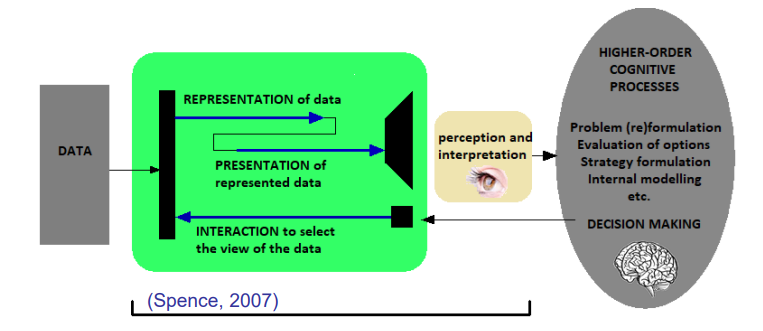
**How importante is Visualization in your job?**

* One of the most critical aspects of being a data scientist is to visualize what you are actually trying to make sense of
* It is impossible to build a model unless I understand what the data means
* You may do some boxplots, scatterplots, trend analysis
* Domain scientists play a very important role

**Visualization use framework**

* Why the user needs it
* How the idiom is design
* What data is hown

**The process of visualization**

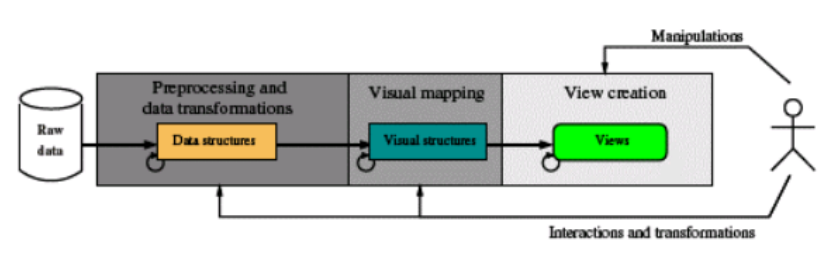


Interaction with data governed by high-order cognitive processes envolves:

* Representation
* Presentation
* Interaction

**Creating visualizations**

* Good design and evaluation is the key to success in producing a Visualization
* Visualization S/W can provide many visual templates
* In spite of variation, all S/W packages follow the same generation proces



Process:

* Preprocessing and transformation
  + Abstract data are rarely in a suitable format for automatic treatment and visualization
  + Raw data have to be given an organized logical structure to be processed by the Visualization S/W
* Visual mapping
  + Necessary to decide which structures to represente the data and their location in the display
  + Some types of abstract data can be easily mapped to a spatial location
  + Many types of data don’t have an easy correspondence with the dimensions of the physical space around us
* Creation of views
  + Views are the final result of the generation process
  + Producing them corresponds to the computer graphucs phase
  + Often the quantity of data to represent is too large for the available space
  + To overcome this problema there are presentation techniques as:
    - Zooming
    - Panning
    - Scrolling
    - Focus + Context
    - Magic lenses

**Data characteristics**

* Data may have a lot of different forms and there are many techniques and systems to visualize them
* A data classification is importante 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
  + Categorical
  + Ordered
    - Ordinal
    - Quantitative
* Links
* Poistions
* Grids

Data representation level:

* Qualitative (categorical)
* Quantitative (numeric)

Data nature:

* Continuous
* Discrete

Measuring scale:

* Nominal (car brands, gender, animal species, etc)
* Ordinal (week days, preferences, levels measured in a linkert-type scale)
* Interval (date, IQ, temperatures in ºC
* Ratio (temperatures in ºK, weight, height) – represents the highest level of representation, has a non-arbitrary zero

In order to understand data, i tis necessary to know:

* Semantics – real world meaning
* Type – structural or mathematical interpretation