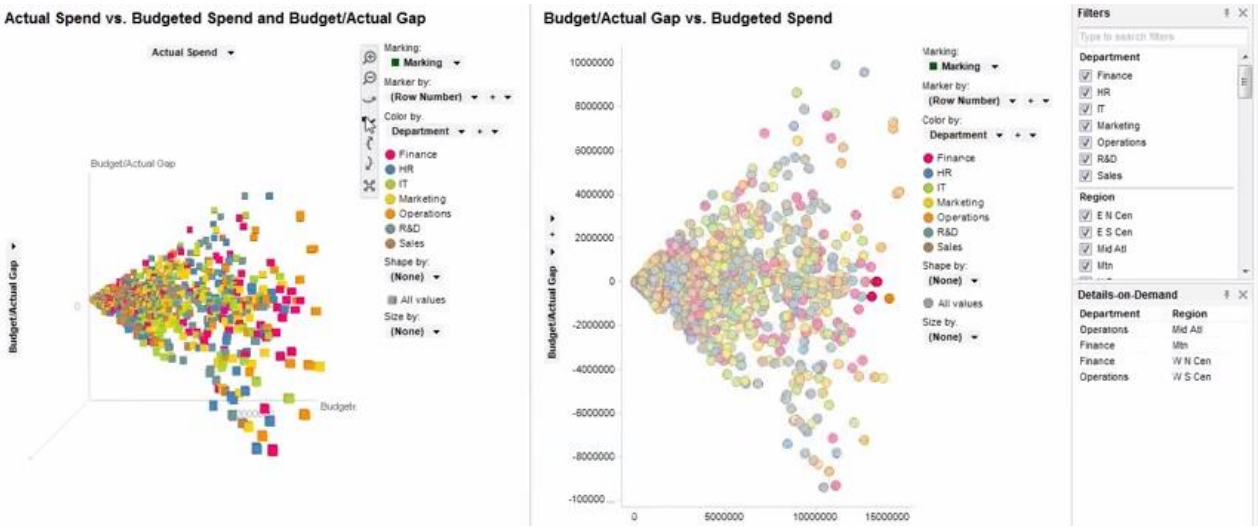
Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Issues in Visualization: Data and Design Process



Beatriz Sousa Santos, University of Aveiro, 2025

Interesting questions:

Why have a human in the decision-making loop?

Why have a computer in the loop?

Why use an external representation?

Why depend on vision?

Why show the data in detail?

Why use interactivity?

What is the design space of visualization idioms?

Why focus on tasks?

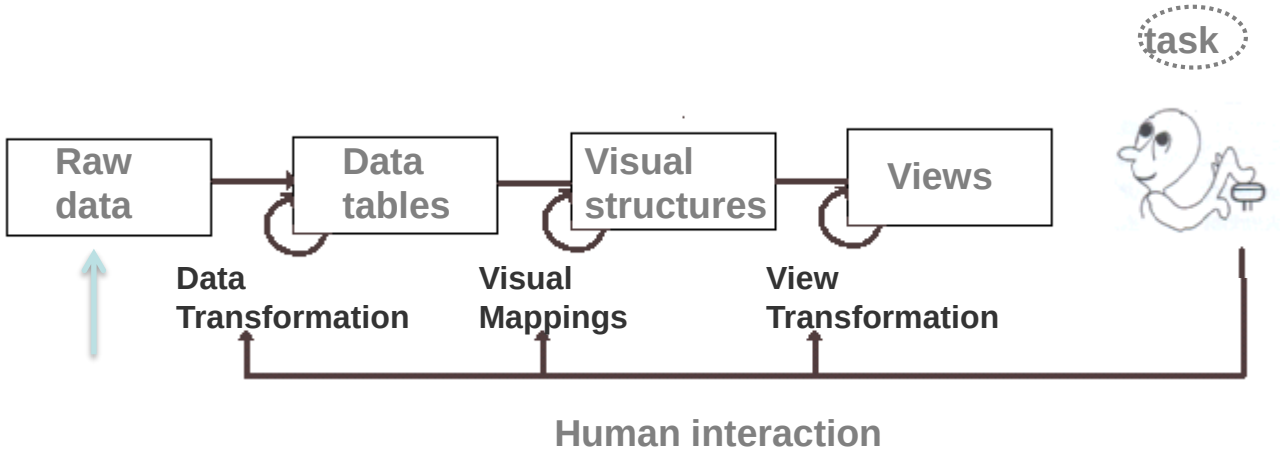
Why are most designs ineffective?

What resource limitations matter?

How can better be measured? (Munzner, 2014, chap. 1)

Information Visualization Reference Model

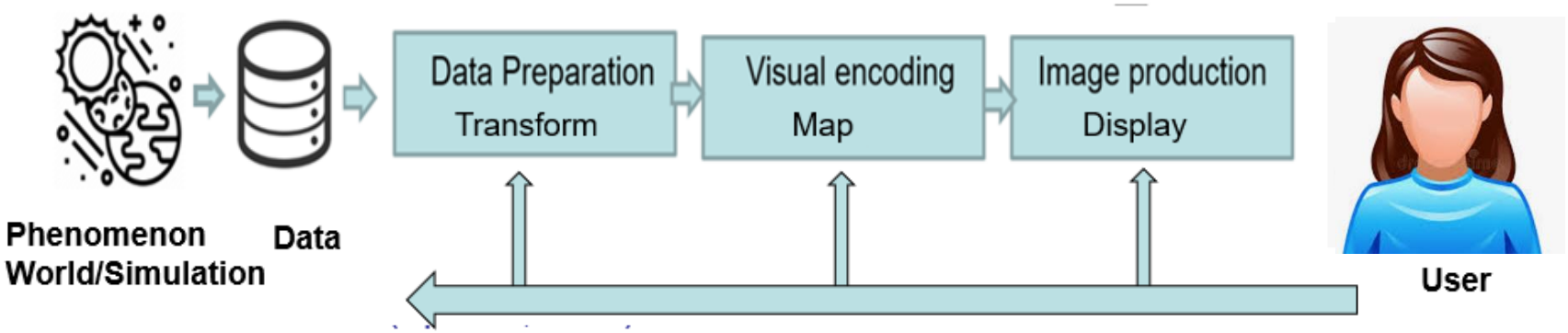
This course focus on Information Visualization



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 creating a Visualization/ a Visual Data Exploration application

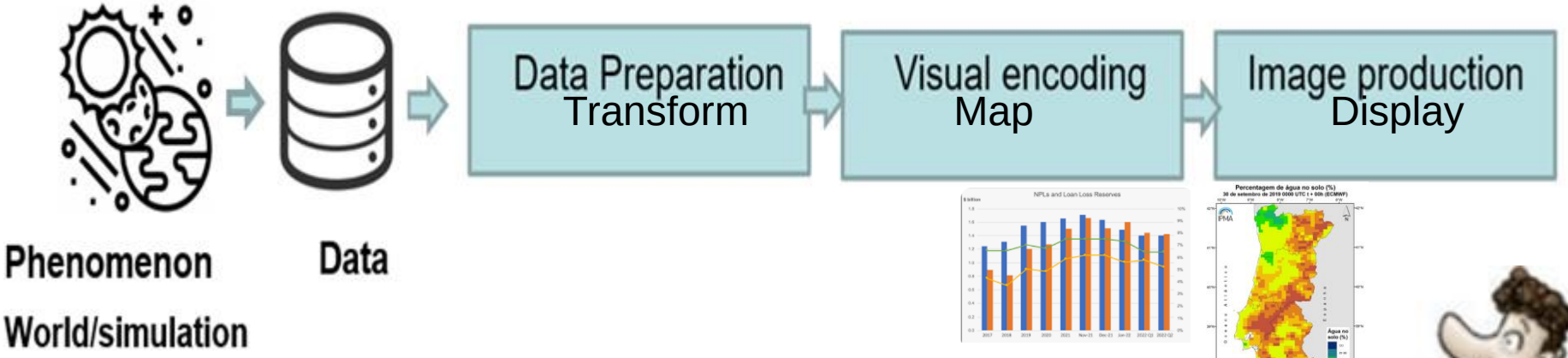


Human-in-the-loop process

The process of creating a (simple/static) Visualization

Data may be **acquired from the world** (e.g. sensors, questionnaires)

or **simulated** (e.g. Finite Element Analysis, weather models)



The users should be supported in their **tasks** (answer their questions) and get **insights** from analyzing the visualization

"***human-in-the-loop***" problems involve the user as a part

- They are **very complex** due to the facts that:

  - humans are very complex systems

  - not well known

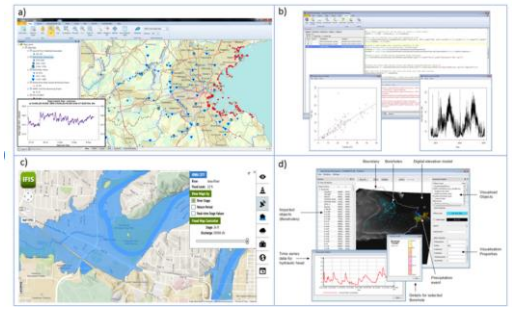
  - in general we cannot change them

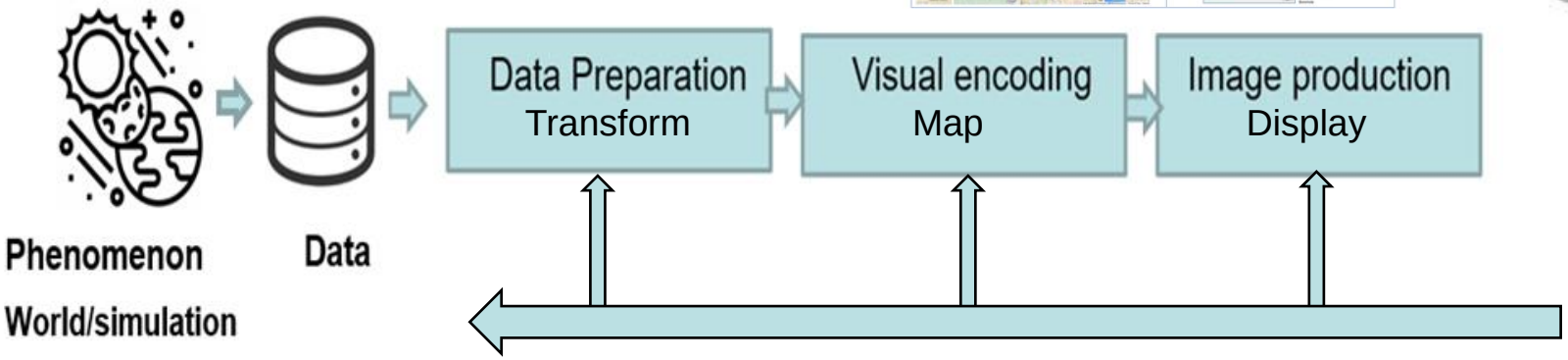
- Target users' **profile, needs, and context of use** must be carefully considered whenever designing a visualization

Visual Data Exploration applications

To let the users visually and interactively explore data it is necessary to provide ways to interact along the process

[An overview of visualization and visual analytics applications in water resources management - ScienceDirect]





Developing this type of application implies a specific approach:

Human-centred design

[What is Human-Centered Design (HCD)? | IxDF What Is Human-Centered Design? | Harvard BS]

- Visualization designers must take into account three very different kinds of **resource limitations**, those of:
  - Humans,
  - Computers,
  - Displays (more and more possibilities…)

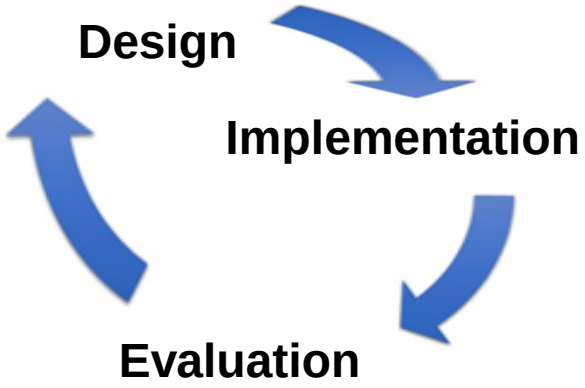
How can we produce a Visualization?

- Designing Visualizations/Vis applications is a "**wicked problem**"

What Are Wicked Problems and How Might We Solve Them? | IxDF

- There are **many possible** Visualization solutions, but how to select the more adequate to a specific case?

(Munzner, 2014)



- To obtain **efficacy** it is fundamental:
  - a correct definition of **goal** and **user tasks**
  - apply adequate methods and **evaluate**

  in **several iterations** until the goals are ~satisfied …

- I.e. using a **human-centered approach**

Framework for analyzing Visualization use

Questions to guide the Visualization creation process:

Who are the users (profile and context of usage)

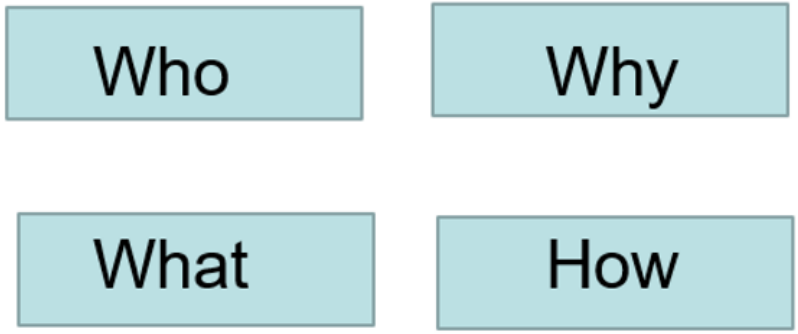
Why the user needs it (tasks and questions to support)

What data to visualize

How to design the visualizations (visualization techniques t)o use

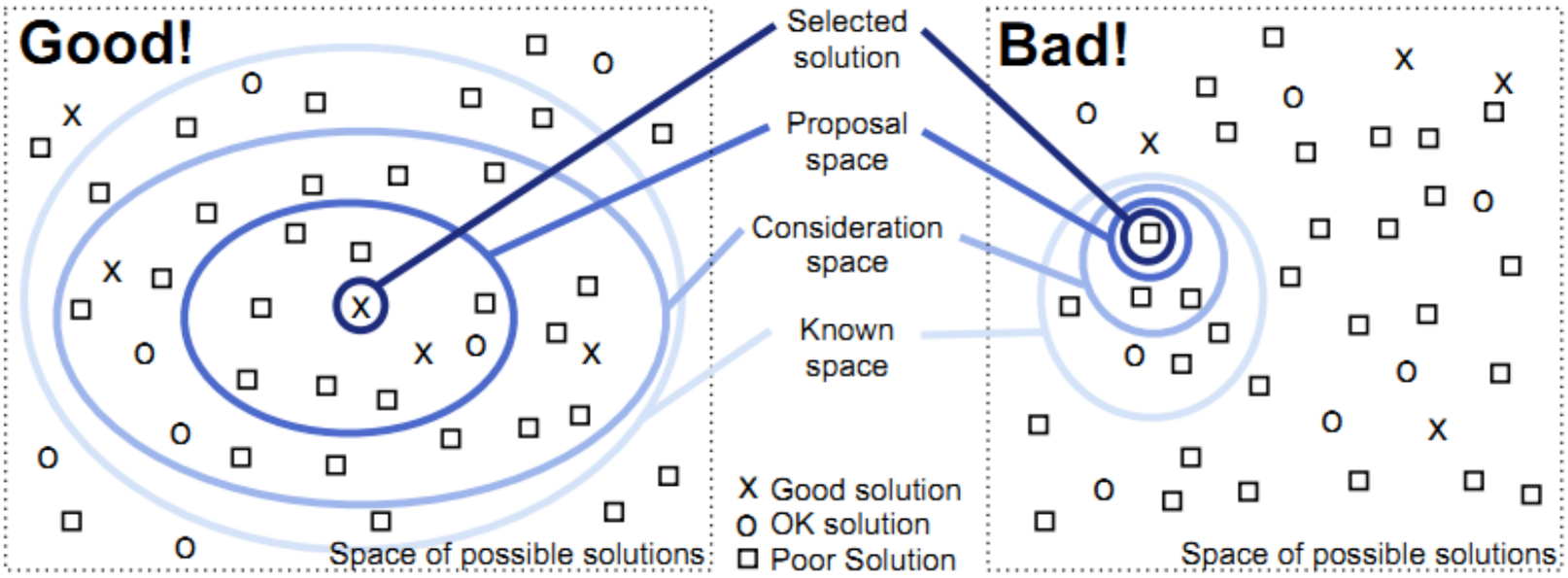
After answering the questions
it is easier to select the techniques

Test and iterate several ideas!



[Chapter 1 What's Vis, and Why Do It? | Visualization Analysis and Design]

The problem of Visualization design



(Munzner, 2014)

- Only a very small number of possibilities are reasonable …
  most are ineffective

Consider multiple alternatives that seem interesting and then select the best!

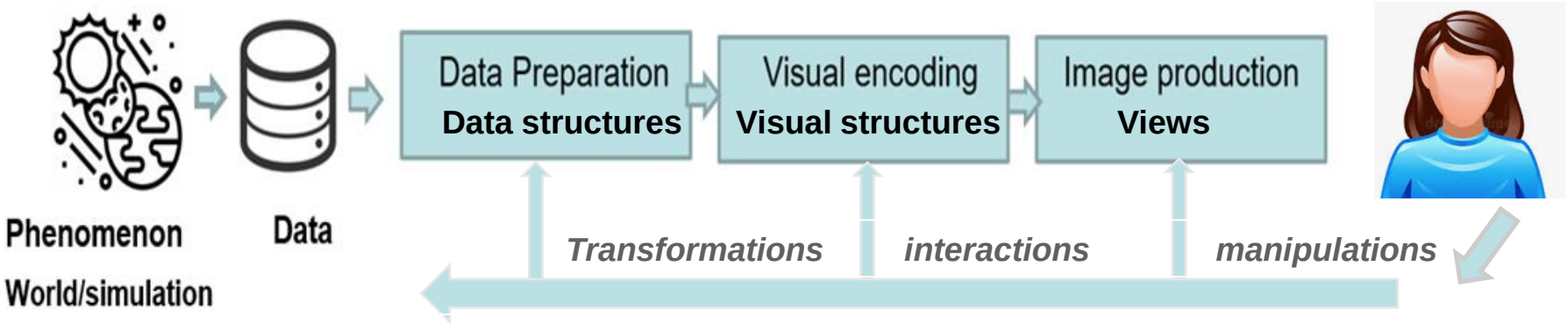
(based on evaluation …)

Creating Visualizations

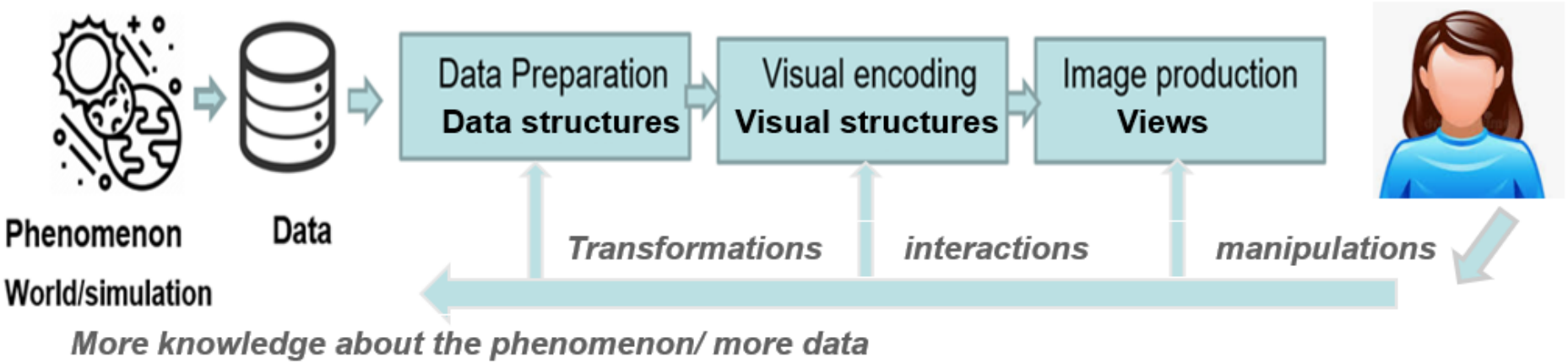
- **Good design and evaluation** is the key to success in producing a Visualization

- Visualization S/W (e.g. Tableau, Spotfire, PowerBi, Flourish Studio, etc.) can provide many visual templates;

- In spite of variation, all S/W packages follow the same generation process



(Mazza, 2009)



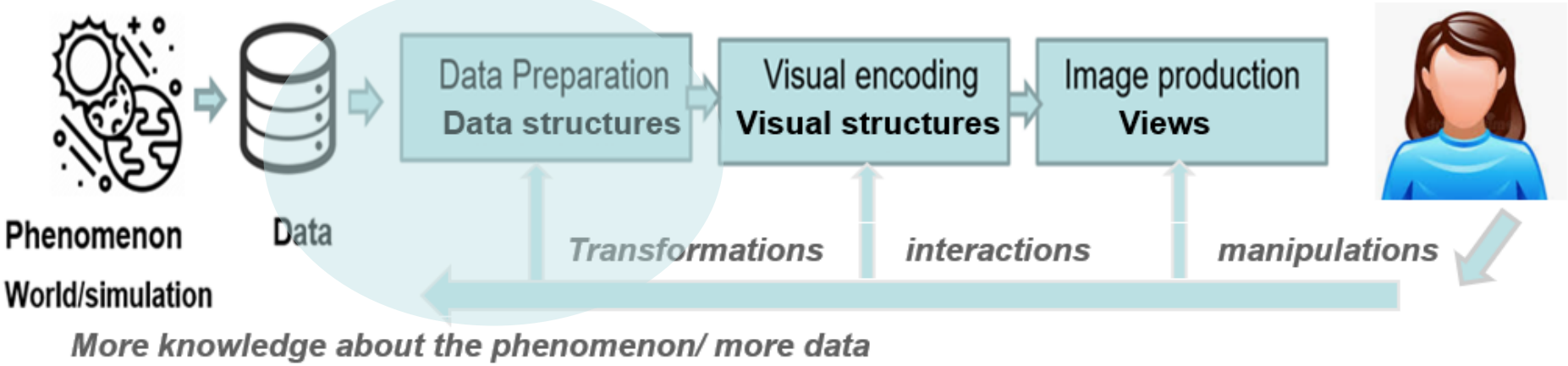
- Creating a visualization can be modeled as a process including several stages:

  (Mazza, 2009)

  - **data preprocessing and transformation**
  - **visual encoding/ mapping**
  - **creation of images/ views**

Note: this is similar to the visualization reference model

not forgetting evaluation in several iterations!



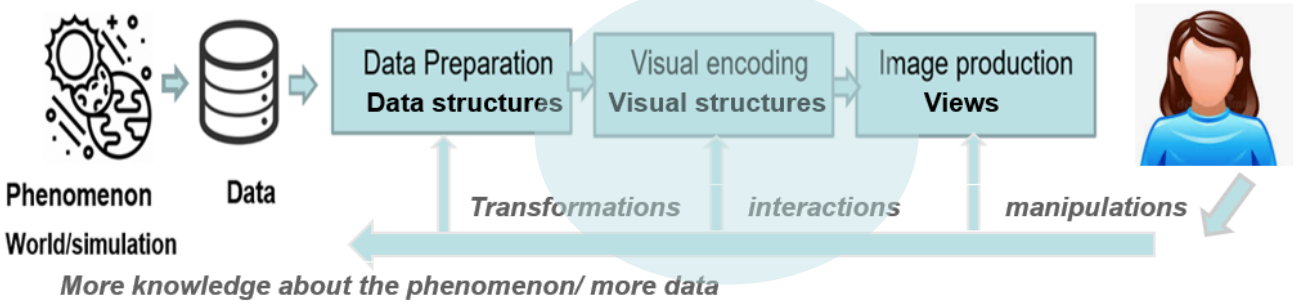
1. Preprocessing (Data transformation):

- Abstract data (which don´t have a specific connection with physical space) are rarely in a suitable format for automatic treatment and visualization

- Raw data (data supplied by the world around us, a.k.a. data sets) have to be given an organized logical structure to be processed by the Visualization S/W

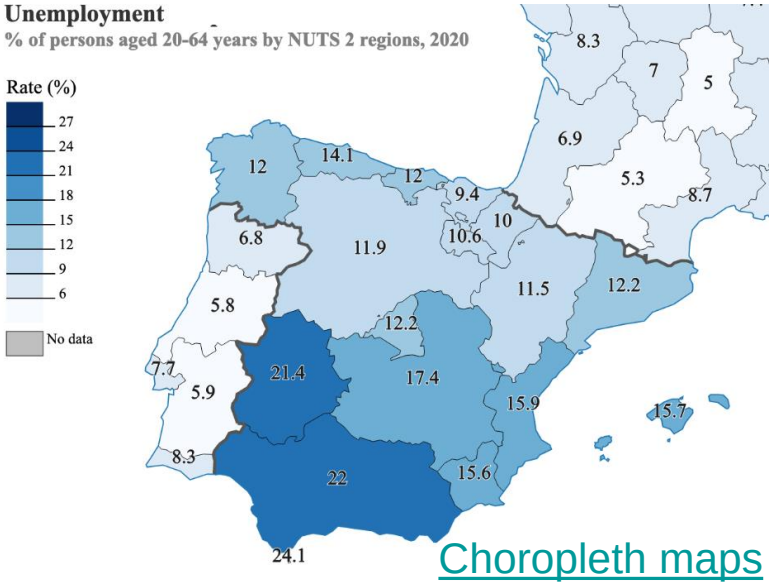
This is a fundamental step; it is assumed to be addressed in another course!

2. Visual encoding (mapping):



- It is necessary to decide:
  - which visual structures to use to **represent** the data
  - their location in the display

- Some types of abstract data can be easily mapped to a spatial location



Choropleth maps

- Examples:

.- data with a geographical or topological structure

- Yet, any types of data don't have an easy correspondence with the dimensions of the physical space around us

How to Display Complex Network Data with Information Visualization | IxDF

2. Visual mapping:

- It is necessary to decide:
  - which visual structures to use to **represent** the data
  - their location in the display

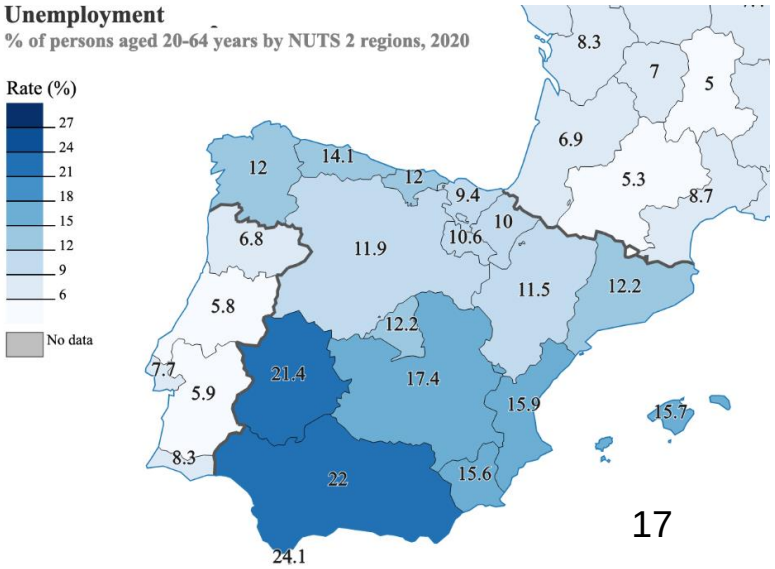
• Some types of abstract data can be easily mapped to a spatial location

• Examples:
.- data with a topological or geographical structure

• Yet, any types of data don't have an easy correspondence with the dimensions of the physical space around us

How to Display Complex Network Data with Information Visualization | IxDF

Choropleth maps



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

- 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

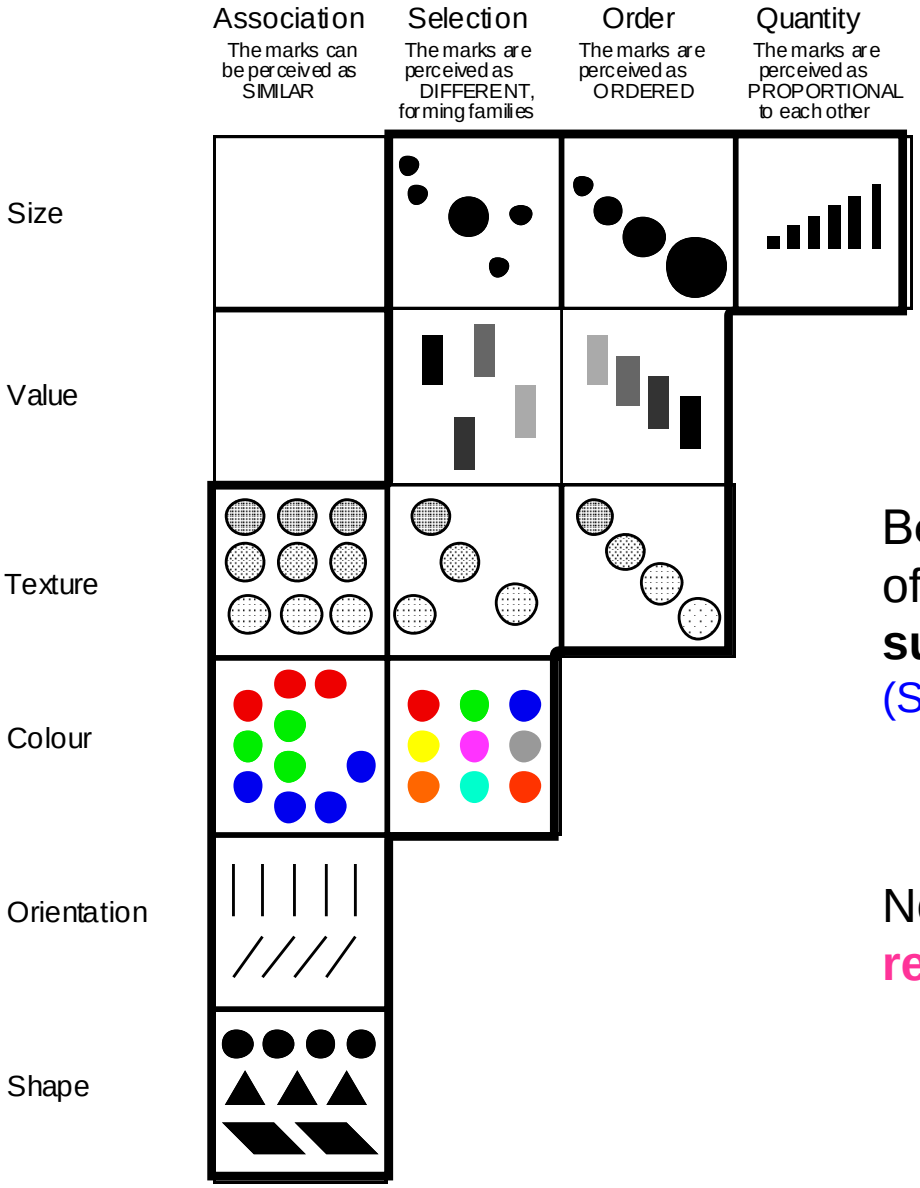
- **Graphical properties** – properties of the graphical elements to which the human retina is very sensitive - **retinal variables:** size, orientation, color, texture, and shape

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

- **Graphical elements** points
  lines
  surfaces
  volumes

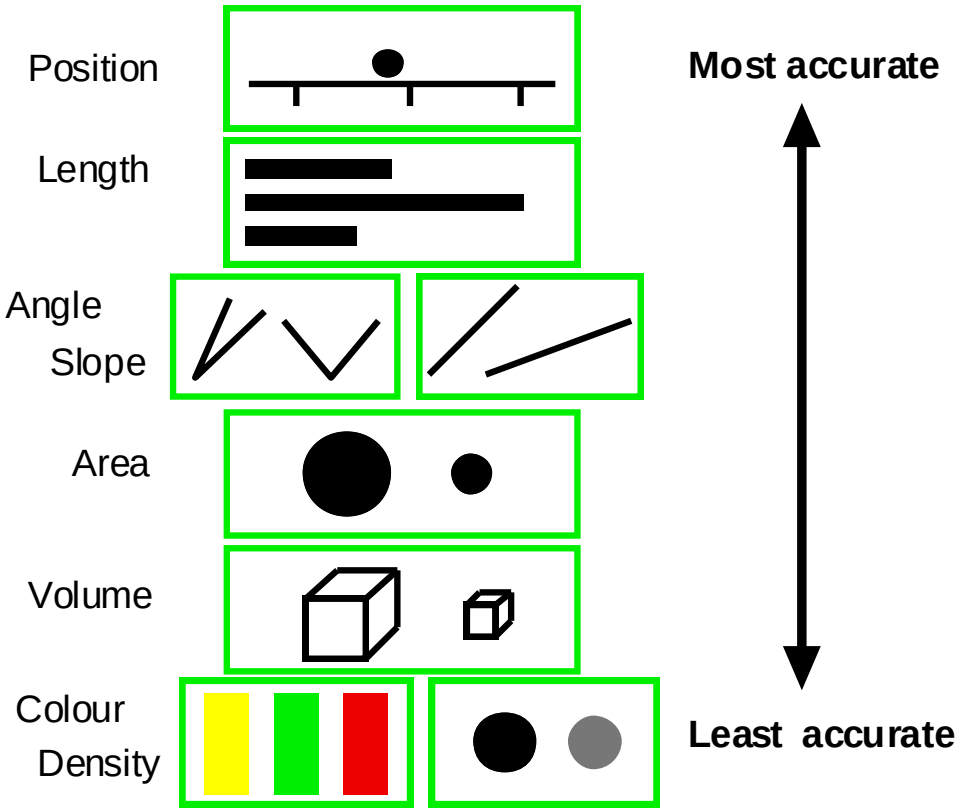
- **Graphical properties** retinal variables:
  size,
  orientation
  color (depends on physiology and culture)
  texture
  shape

•



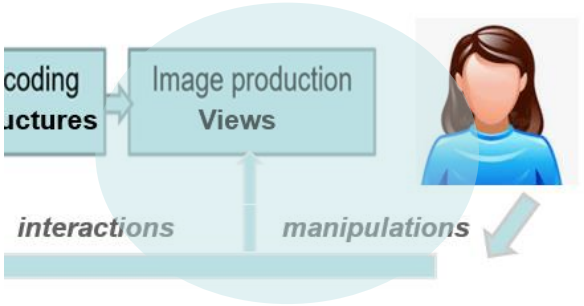
Bertin's guidance regarding the suitability of various visual **encoding methods to support common tasks**
(Spence, 2007)

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



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

3. Creation of views:



- Views are the final result of the generation process

- Producing them corresponds to the “computer graphics phase”:

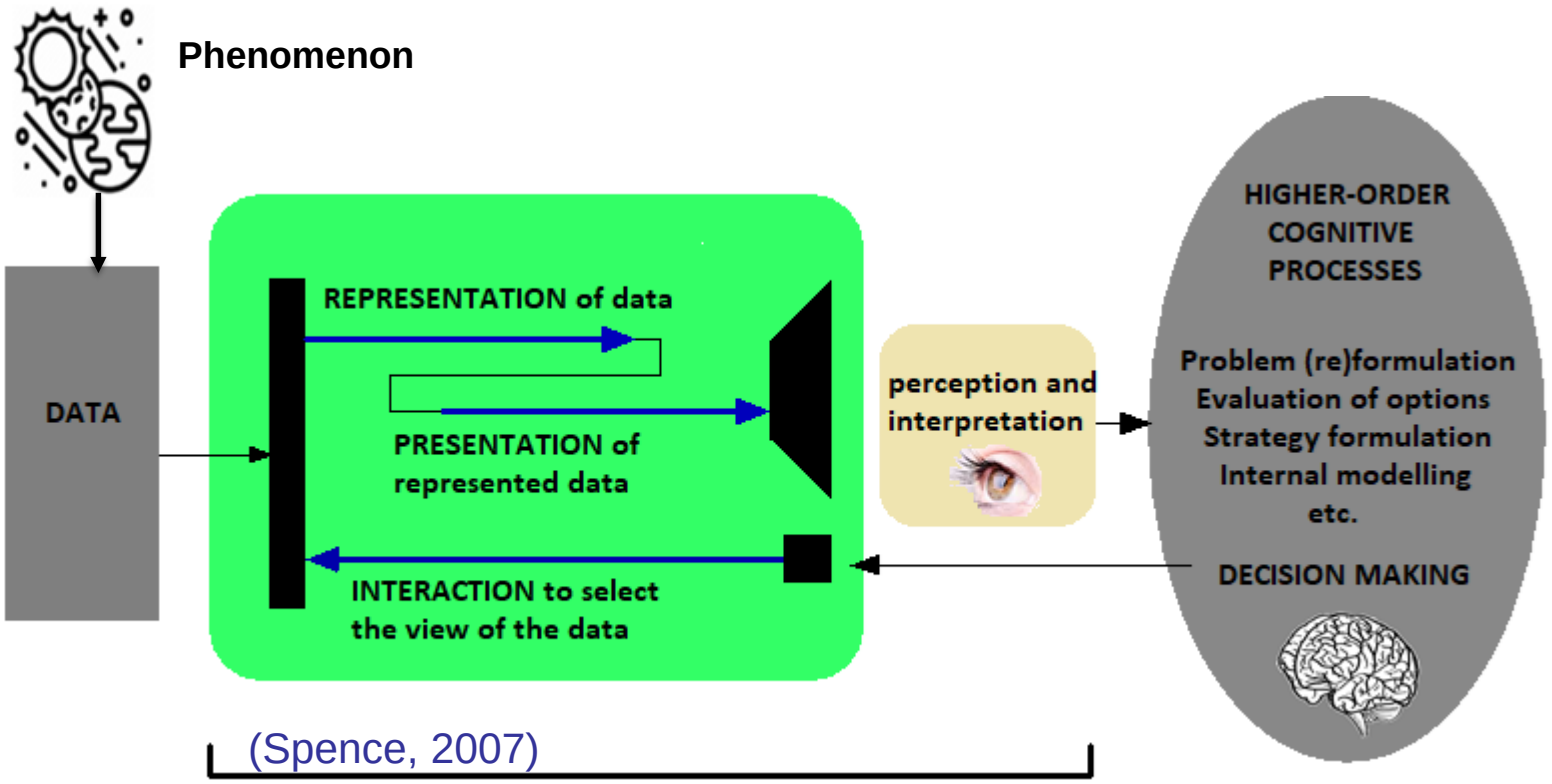
- Often the quantity of data to represent is too large for the available space

- To overcome this problem there are **presentation and interaction** techniques as:
  - Zooming
  - Panning
  - Scrolling
  - Focus + context
  - Magic lenses
  - ...

Bifocal Display | The Encyclopedia of Human-Computer Interaction, 2nd Ed.

The process of visualization

Let us increase the known solution space organizing the methods!



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

- Representation
- Presentation
- Interaction

Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Data Characteristics

- Data may have a lot of different forms and there are many techniques and systems to visualize them

- A data classification is important to:

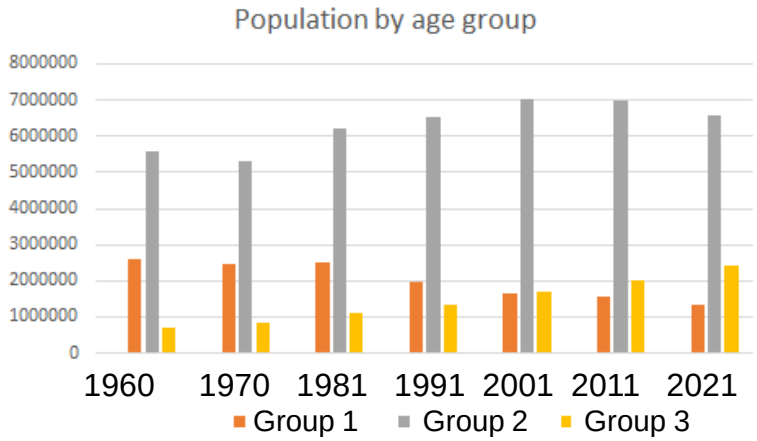
  - predict what visualization techniques are adequate

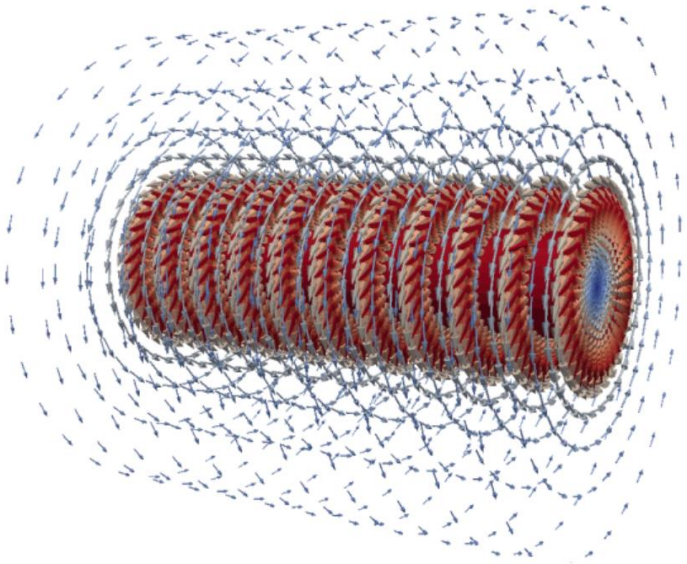
  - make easier the communication about the data

  - allow a more systematic approach to Visualization

  ….

3D/4D Spatial Data -> SciVis





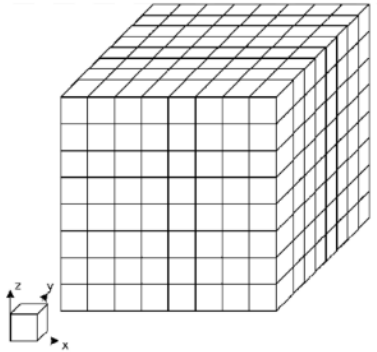
Attribute/variable

Census Year	Population by age group		
	Group 1: 0-14 years	Groups 2: 15-64 years	Group 3: 65 + years
1960	2591955	5588868	708569
1970	2451850	5326515	832760
1981	2508673	6198883	1125458
1991	1972403	6552000	1342744
2001	1656602	7006022	1693493
2011	1572329	6979785	2010064
2021	1331188	6588239	2423639

cell

Item/ object

https://www.paraview.org/ ParaView for Computational Fluid Dynamics



Tabular Data -> InfoVis

Data Abstraction

name	rank	gender	year
Jacob	1	boy	2010
Isabella	1	girl	2010
Ethan	2	boy	2010
Sophia	2	girl	2010
Michael	3	boy	2010

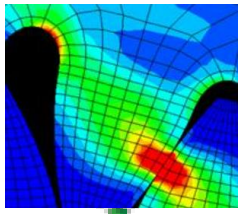
Item/object

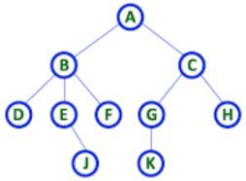
Attribute/variable

- Four basic dataset types:
  - Tables
  - Networks
  - Fields (continuous)
  - Geometry (spatial)

InfoVis

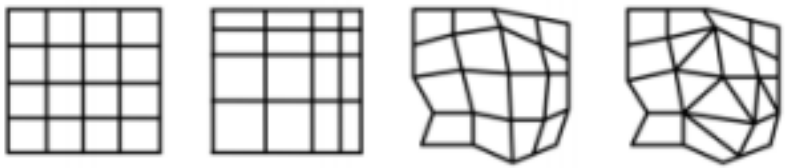
SciVis





- Five basic datatypes
  - Items
  - Attributes
  - Links
  - Positions
  - Grids

Attributes → Categorical
Attributes → Ordered → Ordinal
Ordered → Quantitative

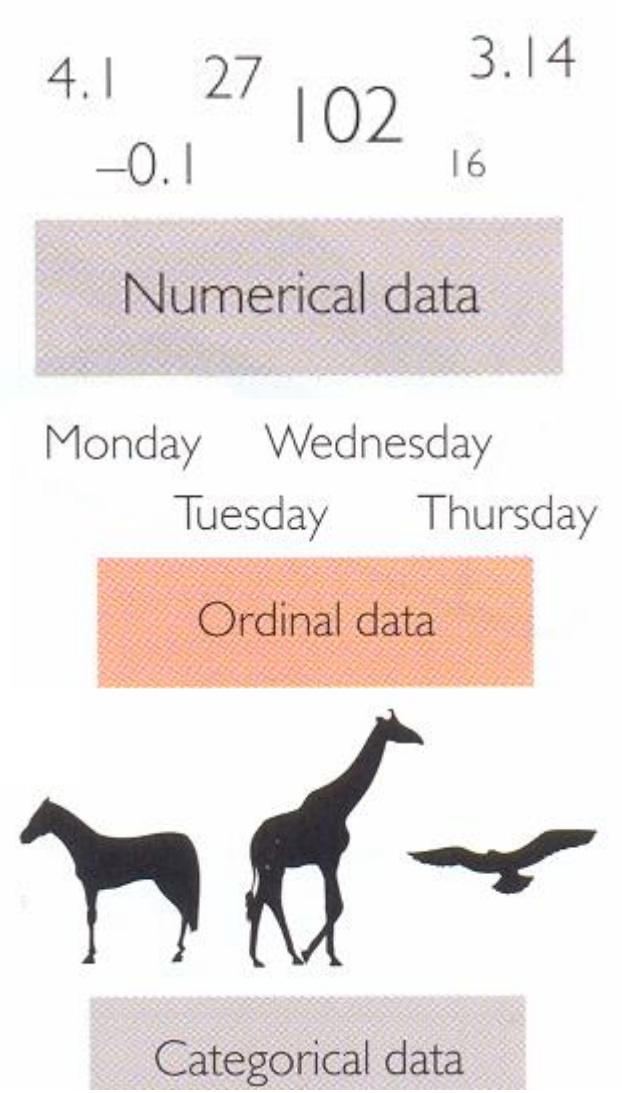


(Munzner, 2014)

- Data representation level*:*
  - Qualitative (or categorical)
  - Quantitative (or numeric)

- Data nature:
  - Continuous
  - Discrete

- Measuring scale:
  - Nominal
  - Ordinal
  - Interval
  - Ratio



(Spence, 2007)

- Examples of measuring scales and types of data:
  - nominal --> car brands, gender, animal species…
  - ordinal --> week days, preferences, levels measured in a Likert-type scale
  - Interval --> date, IQ, temperatures in ºC
  - Ratio --> temperatures in ºK, weight, height
- The ratio scale represents the highest level of representation, has a absolute, non-arbitrary zero (unlike the interval scale)
- This is a general classification and might be used to select the statistical methods to use with the data

Model, structure and format of the data to Visualize

- Consider a data set with three columns:

  latitude *longitude* *d*

- Which is the most adequate way to visualize these data?

- If *d* is depth, the selected visualization technique may involve interpolation (ex: contours)

- If data represent location and the number victims of traffic accidents, interpolation and contours do not make sense

Know the data structure is not enough
It is necessary to know the phenomenon behind the data!

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

  latitude *longitude* *d*

- Which is the most adequate way to visualize these data?

- If *d* is depth or altitude?

the selected visualization technique may involve interpolation

(e.g. isocontours, isosurfaces, 3D surface)

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

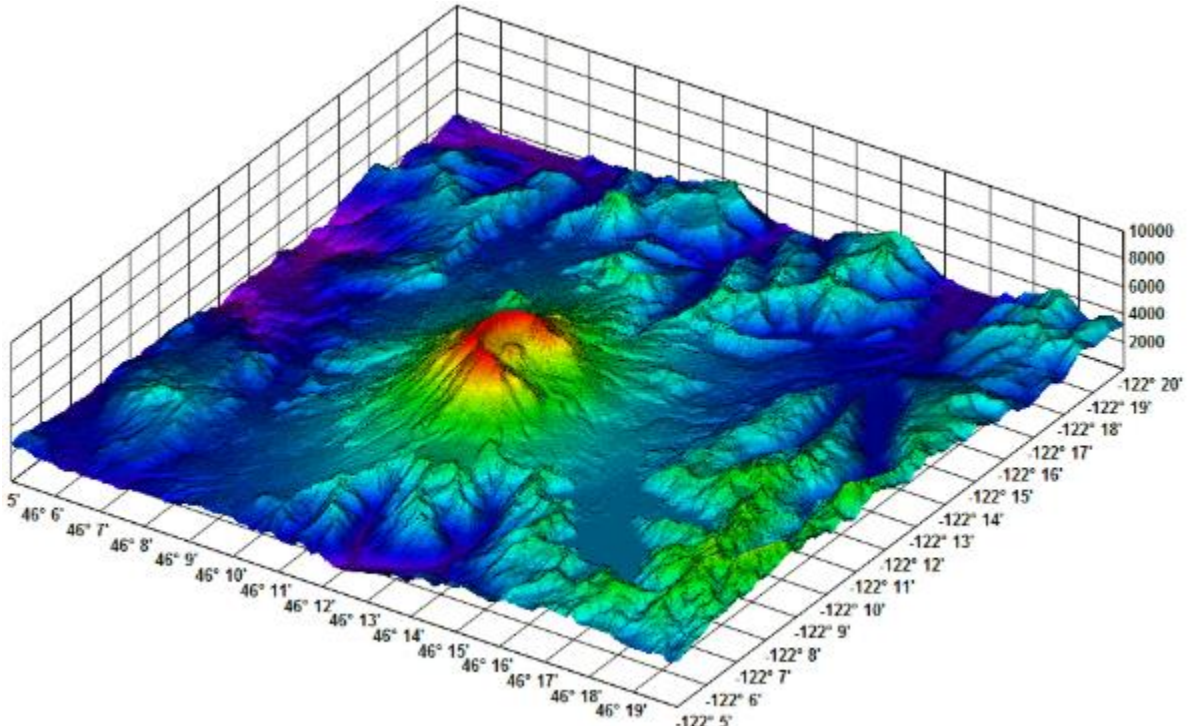
  latitude *longitude* *d*

- Which is the most adequate way to visualize these data?

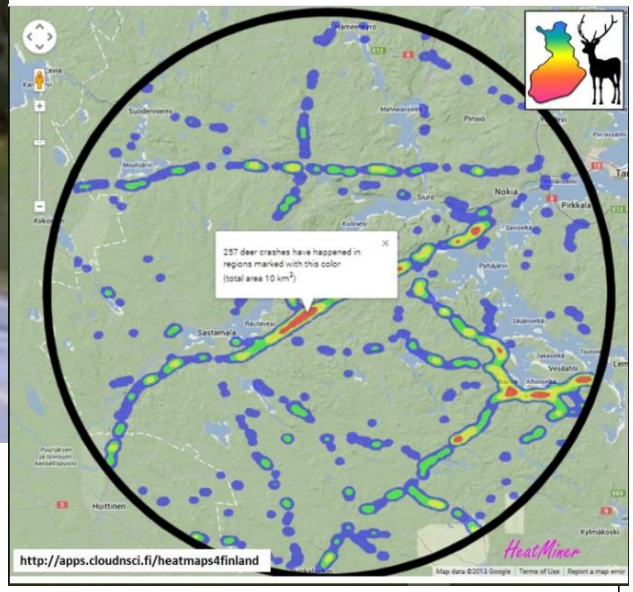
- If *d* is depth or altitude?

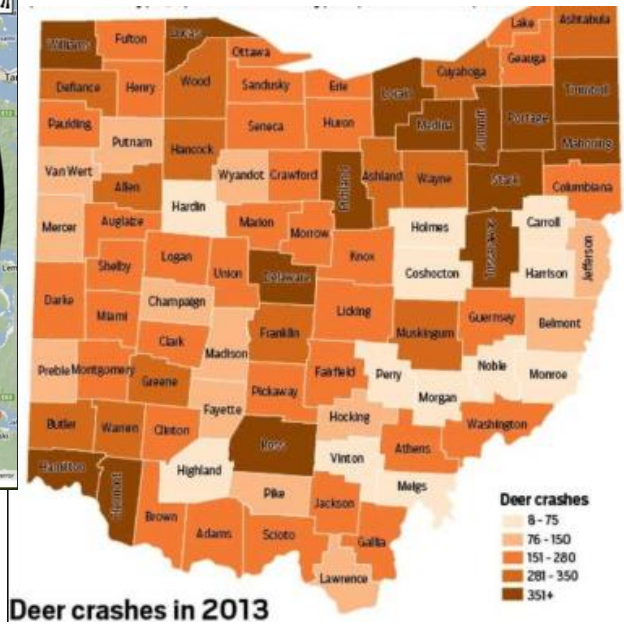
the selected visualization technique may involve interpolation

(e.g. isocontours, isosurfaces, 3D surface)



- What if the data represent location and the number of “deer crashes”?

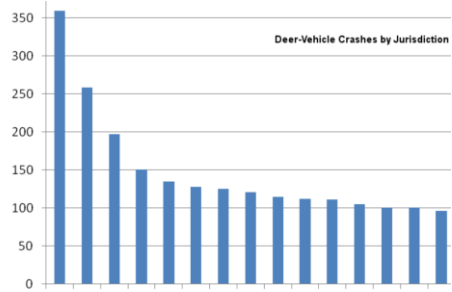




- Interpolation and contours don’t make sense!

Know the data structure is not enough

It is necessary to **know the phenomenon behind the data and the users’ tasks**



Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Data preparation

What is Data preparation? Steps in the Preparation Process

- Data preparation is very important and very time consuming

- Several phases and terms:
  - Data pre-processing
  - Data wrangling
  - Data cleaning, Data tiding …
  - Data transformation

Data integrity becomes more essential when the volume of data increases

4. Data Preparation | Data at Work: Best practices for creating effective charts and information graphics in Microsoft® Excel®

“Brilliant visualizations cannot redeem bad data!”

Or

“Garbage in garbage out …”

Cleansing Data

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

Transforming Data

(at the variable level)

- Encoding
- Aggregation
- Derived data
- Removal
- Standardization

Examples:

Cleansing Data

Birth date: Feb/30/2000

Temperature: -300 °K

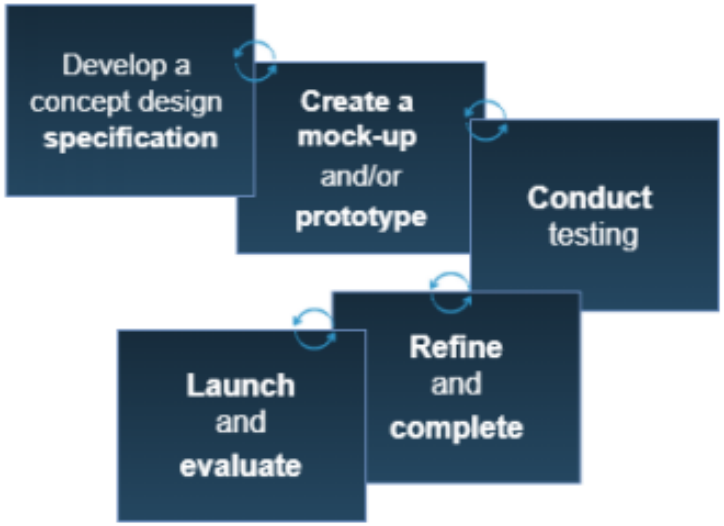
City: Lixboa

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 ($T_{range} = T_{max} - T_{min}$)
- Removal – remove data that are not needed
- Standardization – M/F; °C or °F

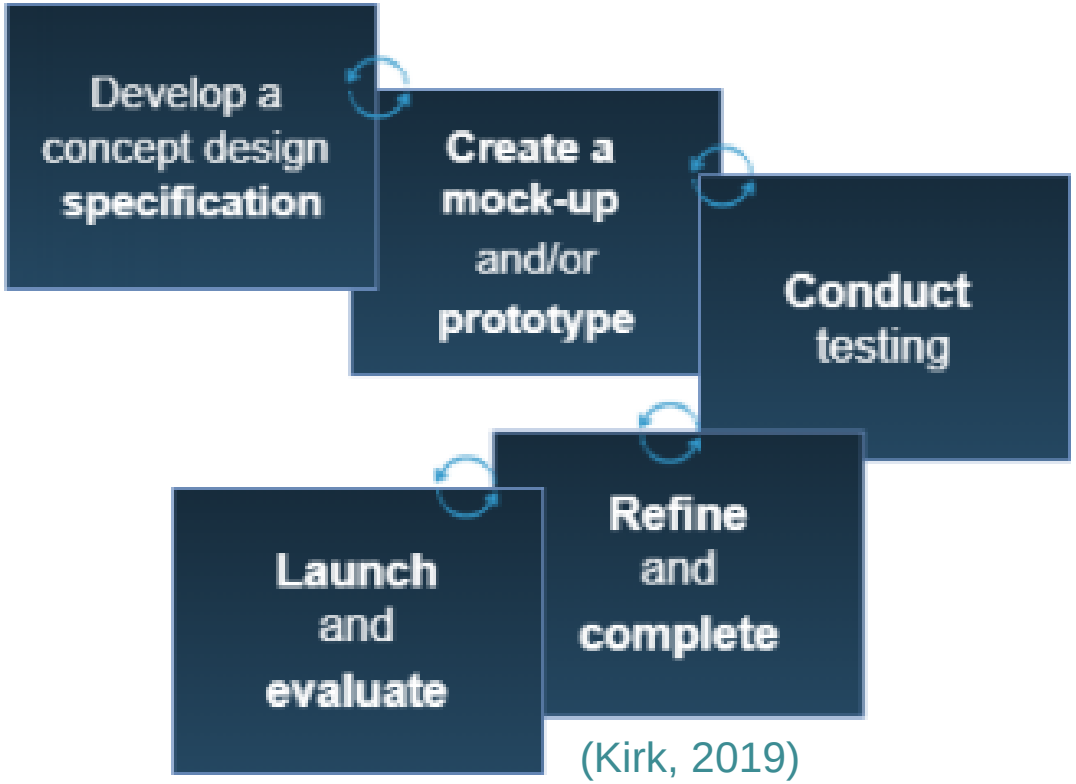
Developing a Visual data exploration application

a very brief introduction



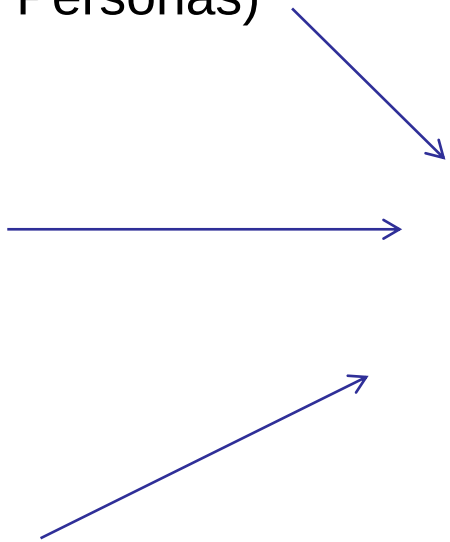
Phases of the development cycle
of a visual data exploration application

after understanding the users and their questions to elicit requirements:



(Kirk, 2019)

- There are several methods that can be used to help requirement elicitation of an interactive application, e.g.
  - User models (such as Personas)
  - Scenarios
  - Task analysis
  - Requirements
  - Low fidelity prototype

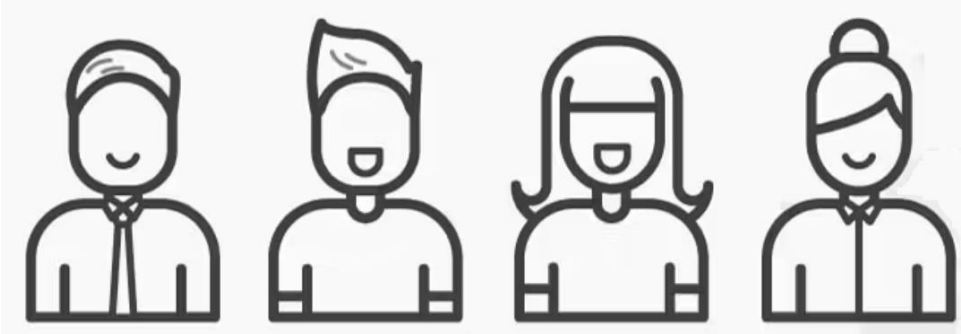


Typical in user-centered design approaches (UCD)

  - To apply in the Visualization mini-project

Personas

- Personas are **fictional characters** based on **user research** to help understand:
  - users' needs,
  - experiences,
  - behaviors
  - goals



- Make the design task at hand less complex

- Guide the ideation processes, and help to achieve the goal of creating a good user experience for the target user group

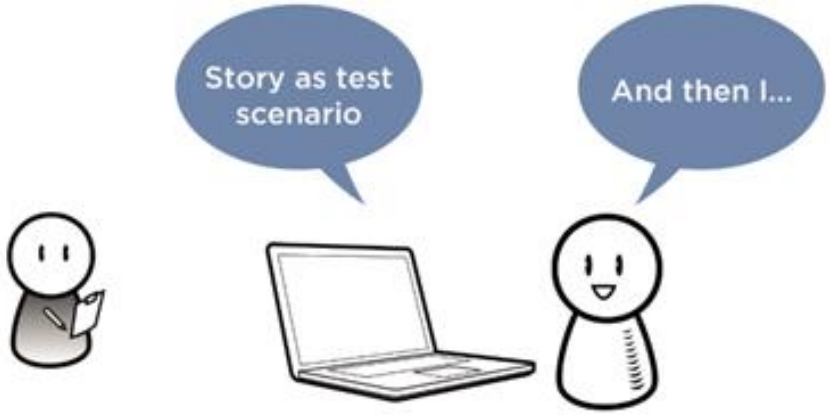
Personas – A Simple Introduction | IxDF

Scenarios

- Stories and contexts about how the user groups use a future product/service

- Note the goals and questions to be achieved and sometimes define the possibilities of how the user(s) can achieve them on the application

- Scenarios are critical for
  - designing
  - UX testing



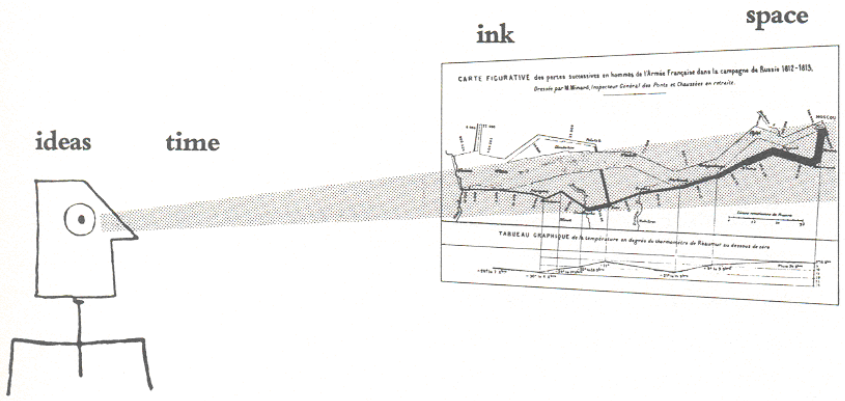
(Note: Different from User stories and Use cases)

Usability – Digital.gov

What are User Scenarios? | IxDF

Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Effective Visualization



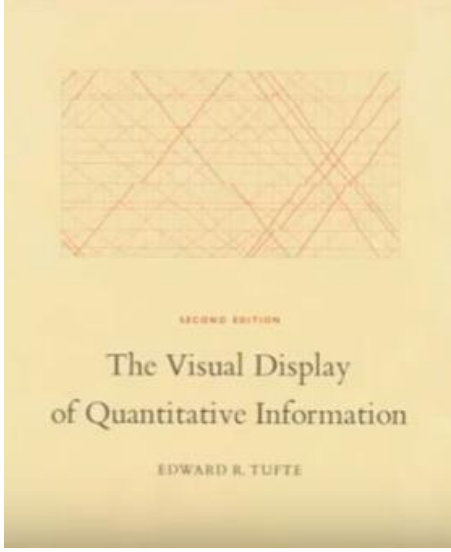
The Visual Display of Quantitative Information - Edward Tufte

Beatriz Sousa Santos, University of Aveiro

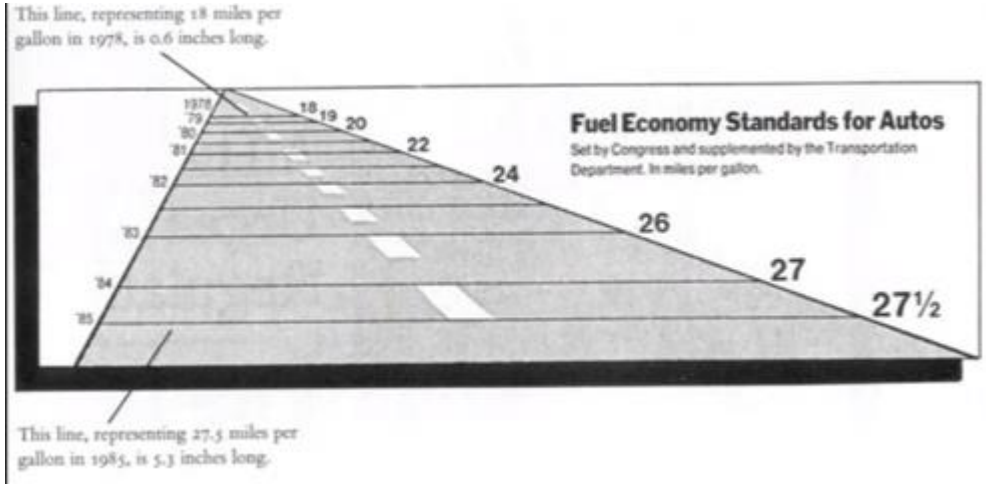
Effective visualization

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



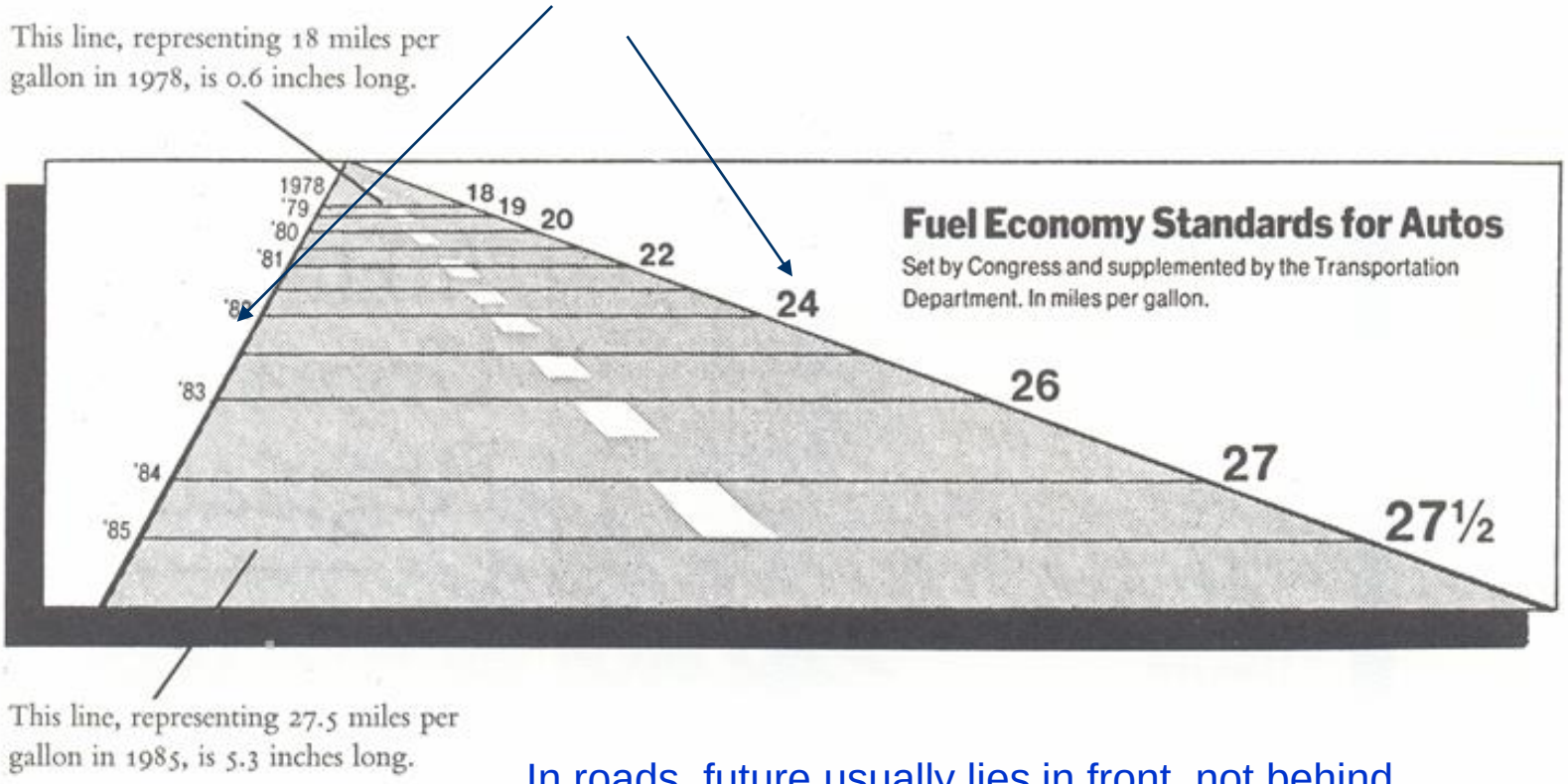
There are methods to evaluate visualizations that should be used along the process of creating a visualization

Lie Factor - InfoVis:Wiki

this example has several problems:

Lie Factor = 14.8

Legends have a constant size in one side and variable in the other



In roads, future usually lies in front, not behind

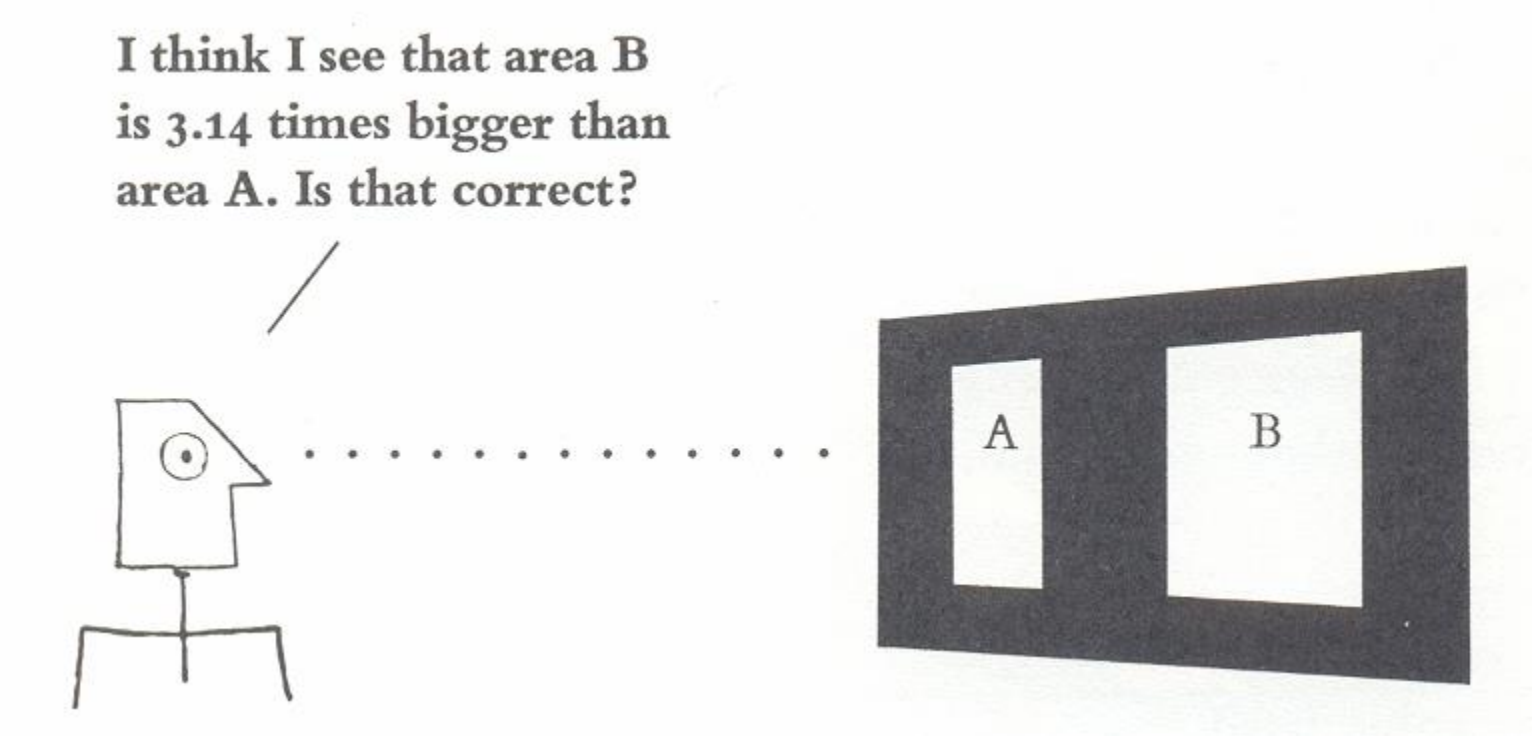
- Perception varies with

  - context

(Tufte, 1983)

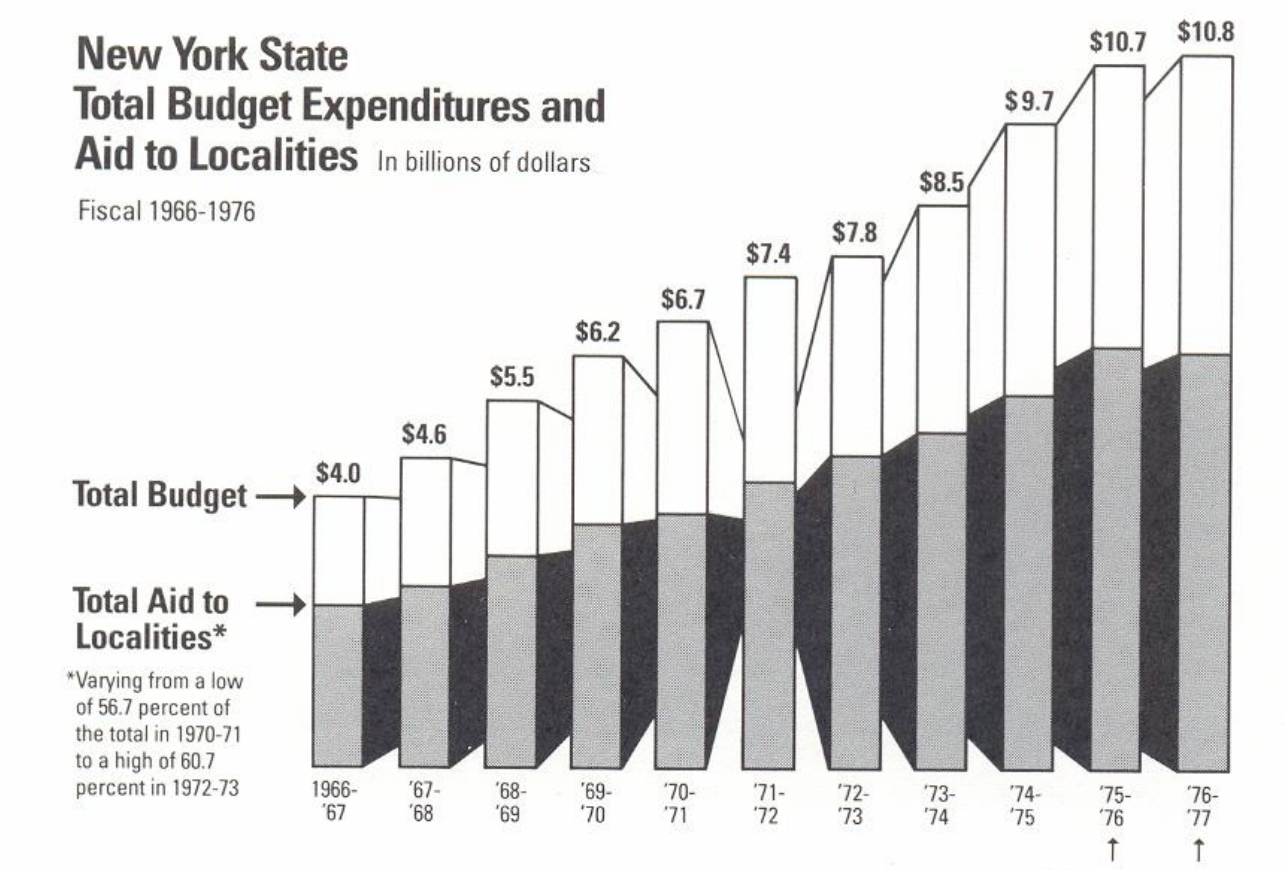
• How do we know that the visual image represents the underlying numbers?

• One way to try to answer these questions is to conduct experiments on the visual perception of graphics



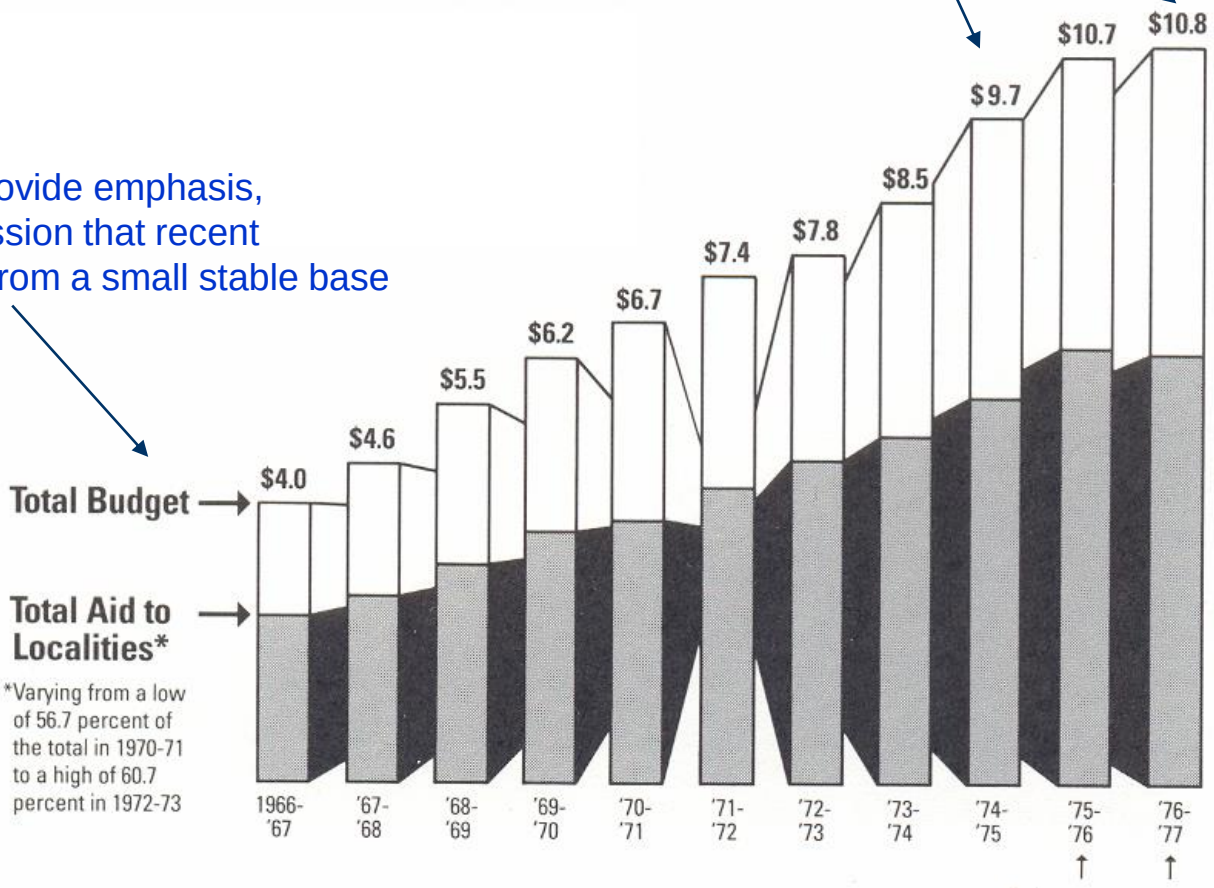
(Tufte, 1983)

A “classic” poor visualization example:



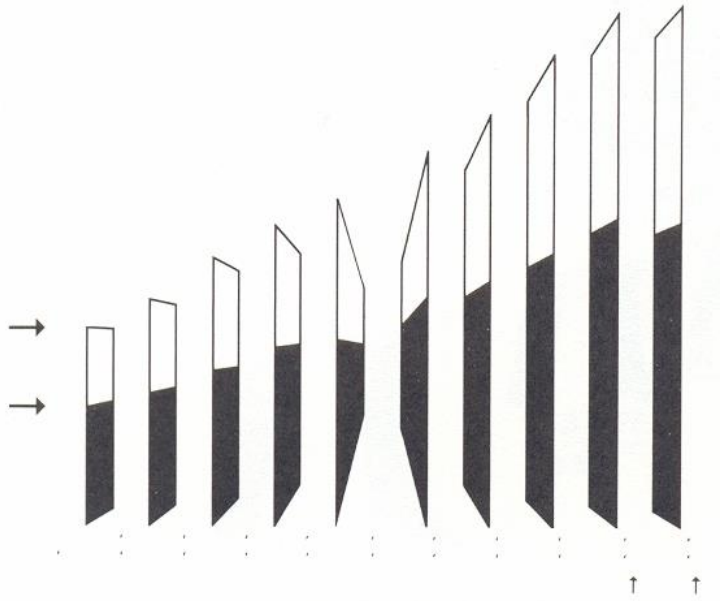
These three parallelipeds have been placed in na optical plane in front of the other eight, creating the image that the newer budgets tower over the older ones

Horizontal arrows provide emphasis, Encorage the impression that recent years have shot up from a small stable base

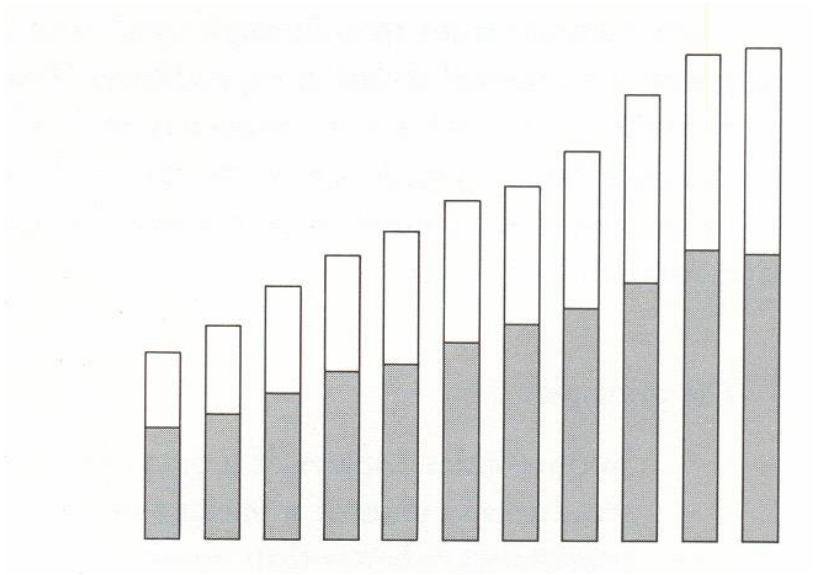


Arrows pointing straight up emphasize recent growth

Leaving behind the distortion



we have a calmer view:

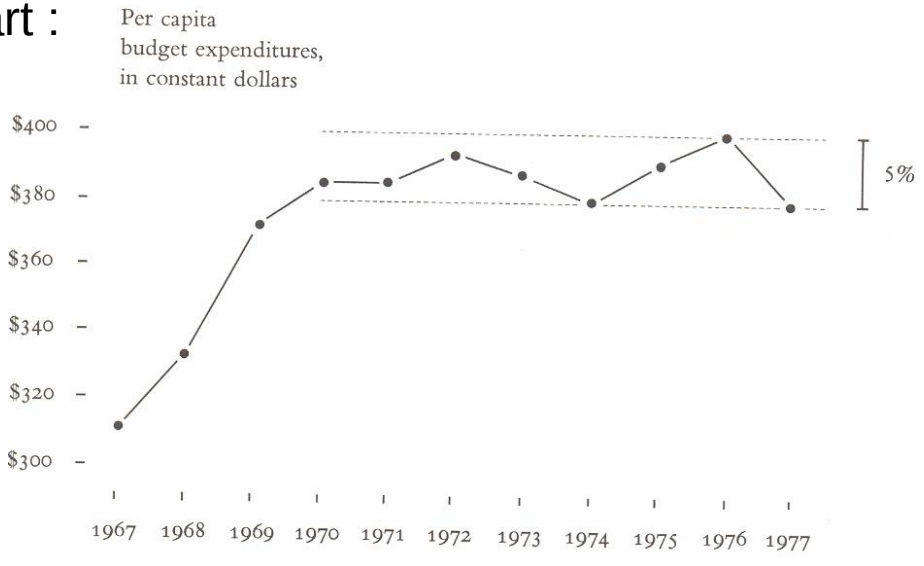


Two statistical lapses also bias the chart :

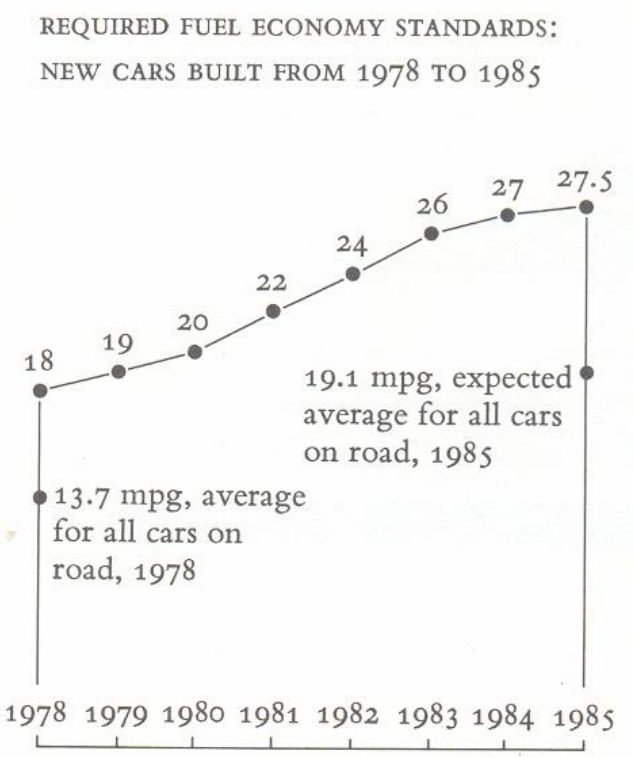
- Population increased10%

- there was substantial inflation

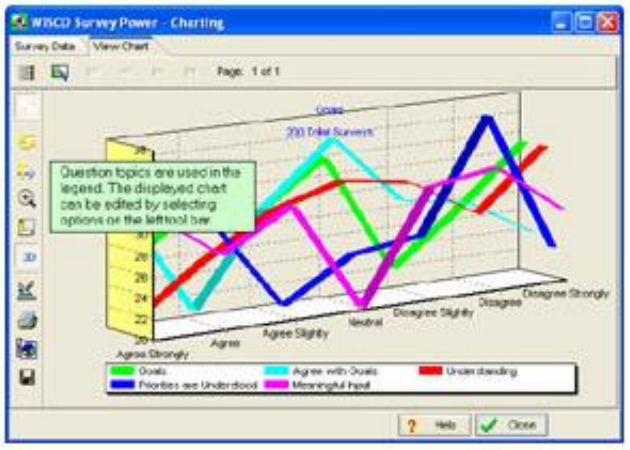
Final result →

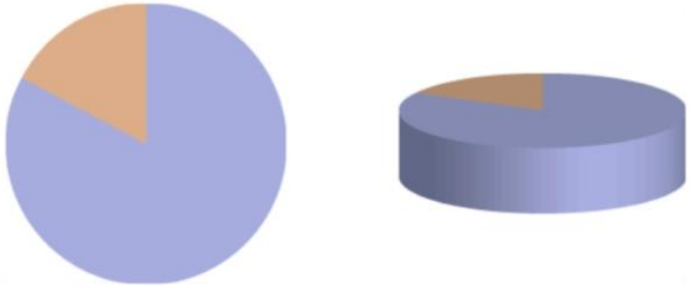


Decorations without lies:

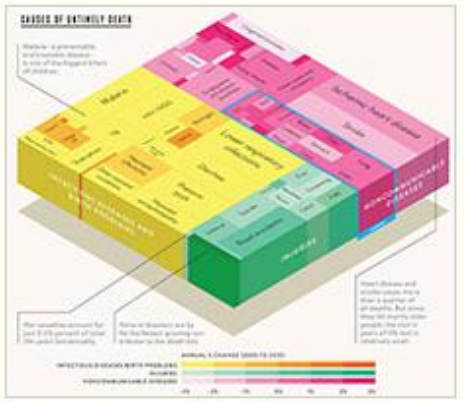


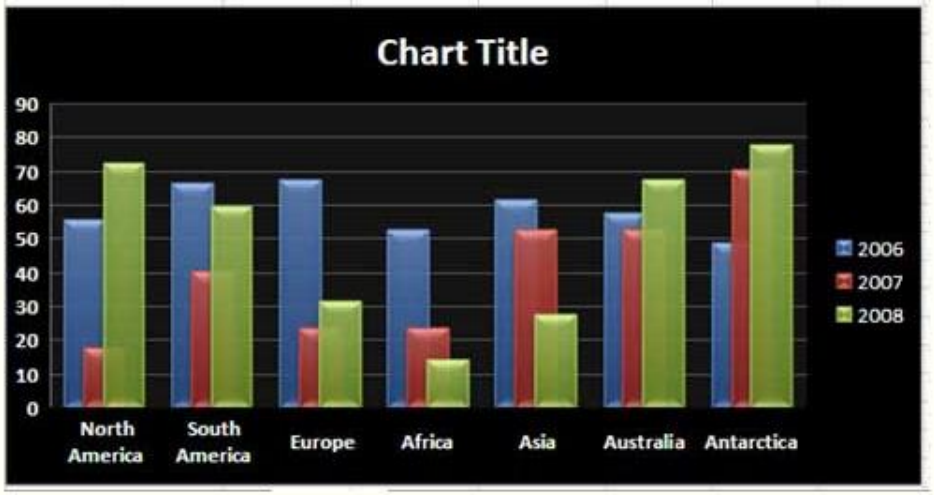
Effective visualization: more poor examples analyzed





[Visualization design objectives | Data Visualization: a successful design process]





[Perceptual Edge - Examples]

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

Main bibliography

- Spence, R., *Information Visualization, An Introduction*, Springer, 2014
- Munzner, T., *Visualization Analysis and Design* *, A K Peters, 2014
- Kirk, A., *Data Visualisation- A Handbook for Data Driven Design*, 2nd. Ed., Sage, 2019
- Fisher, D., Meyer, M.*, Making Data Visual. A practical Guide to using Visualization for insight* *, O'Reilly, 2017
- Camões, J., *Data at Work : Best practices for creating effective charts and information graphics in Microsoft Excel* *, Pearson Education, 2016
- Ware, C., *Information Visualization, Perception to Design* *, 3nd ed.,Morgan Kaufmann, 2012
- Alberto Cairo, *How Charts Lie: Getting Smarter about Visual Information*, W. W. Norton & Company; 1st edition, 2020
- Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983

- Explore books marked * and other books at the playlist:
  Playlist: CT- Data Visualization