Visualisation de données Massives

Introduction

2020

Marco Winckler

Université Nice Sophia (Polytech) | I3S | SPARKS team | bureau 446

winckler@unice.fr

http://www.i3s.unice.fr/~winckler/

M1 MSc DataScience
Information Visualization

Intro

Semestre 1, 2021-2021

Lecturer: Marco Winckler (UCA, I3S)
winckler@unice.fr

UNIVERSITÉ CÔTE D'AZUR

"Graphics *reveal* data"

- John Snow's map of water wells in London (1854)

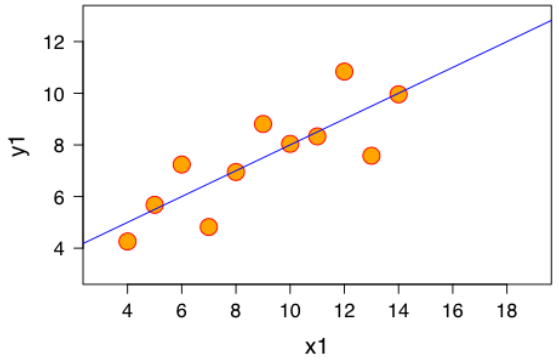
Tufte, Edward. *The Visual Display of Quantitative Information*. Cheshire, Graphics Press, 2001, 2nd edition

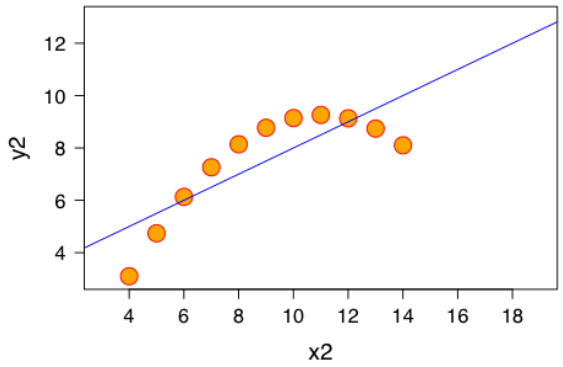
causality

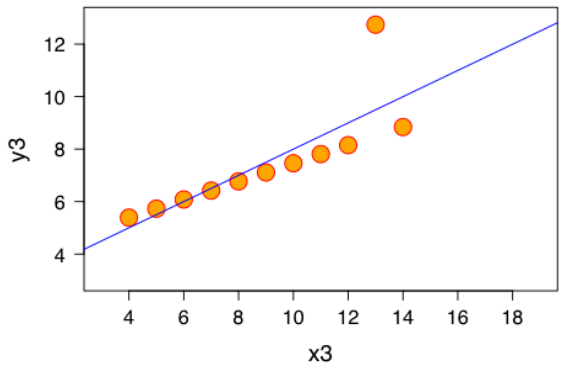
Basics

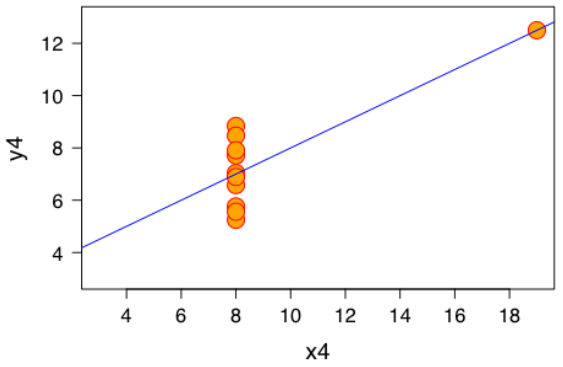
Anscombe's Quartet: Raw Data

	I		II		III		IV	
	x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
var.	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
corr.	0.816		0.816		0.816		0.816	





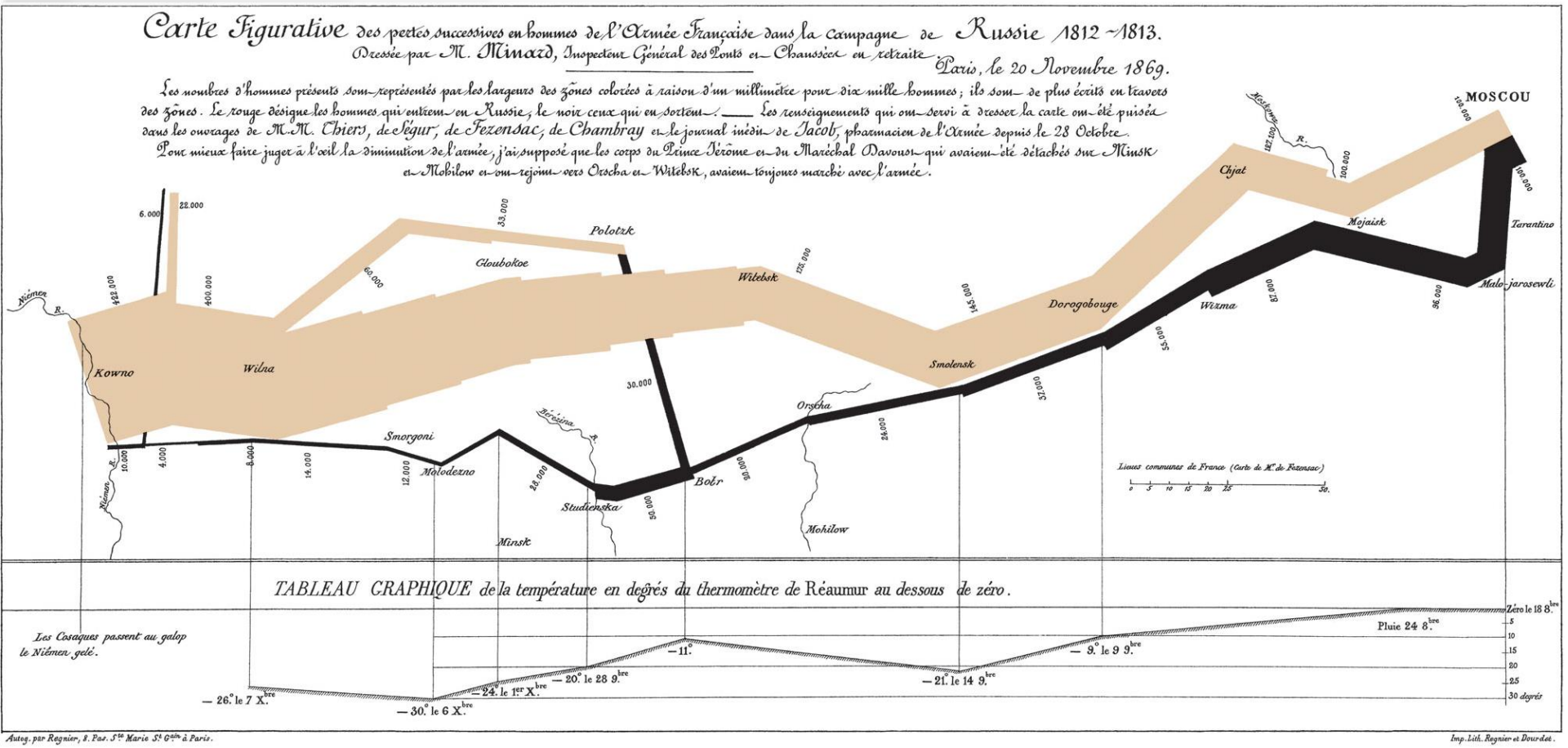




Anscombe, F.J. *Graphs in Statistical Analysis*. American Statistician 27 (1973), 17–21

Show differences, similarities and patterns

Adding complex data



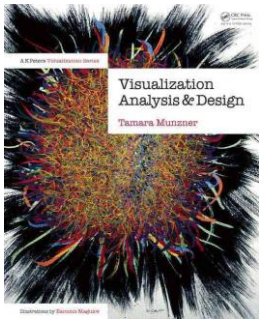
Charles Joseph Minard's map (1869)
1 mm = 6 mil people

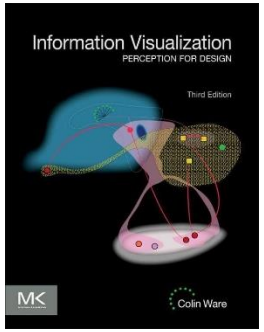
Communicating data

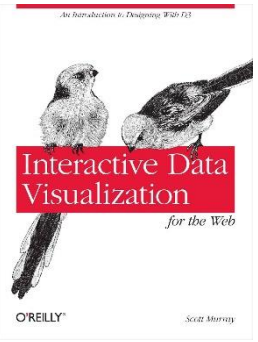
https://www.youtube.com/watch?v=jbkSRLYSojo

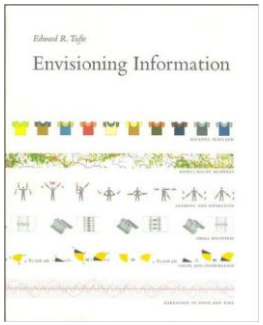
References

- Tamara Munzner. Visualization Analysis and Design. AK Peters Visualization Series, CRC Press (2014).
- Colin Ware. Information Visualization, Third Edition: Perception for Design (Interactive Technologies). Morgan Kaufmann. 536 pages (2012)
- Scott Murray. Interactive Data Visualization for the Web. O'Reilly Media. 273 pages (2013)
- Edward Tufte. The Visual Display of Quantitative Information. 1983









Objectives:

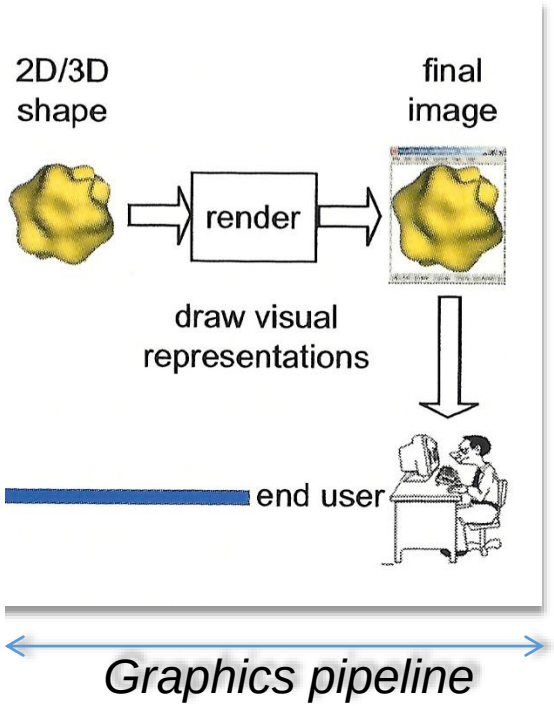
- The goal is to present information visualization techniques and apply them to solve problems related to the interaction with large datasets
- Understand the principles of information visualization
- Know the Schneiderman's mantra of information visualization
- Know the main information visualization techniques
- Know the tools for dealing with information visualization
- Know the data structures used to visualize data
- Implement a pipeline for information visualization
- Interact and use different information visualization techniques
- Be able to reuse information visualization techniques available
- Be able to program basic information visualization techniques

Planning

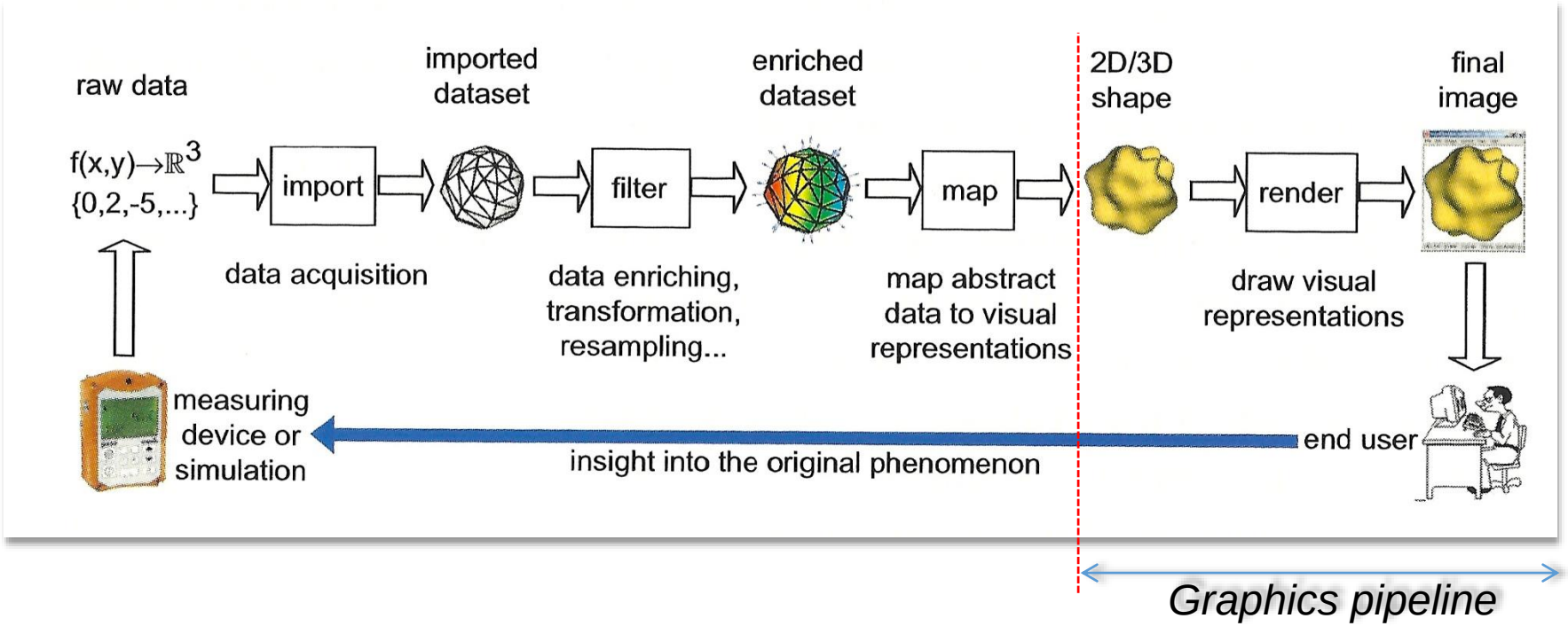
- Contents:
  - Introduction to information visualization
  - Information visualization pipeline
  - Information visualization techniques (ex. graphs, hierachies, multidimensional data, …)
  - Data processing
  - Programming of information visualization techniques

Information visualization
x
Computer Graphics

The domain


2D/3D
shape
final
image
render
draw visual
representations
end user
Graphics pipeline

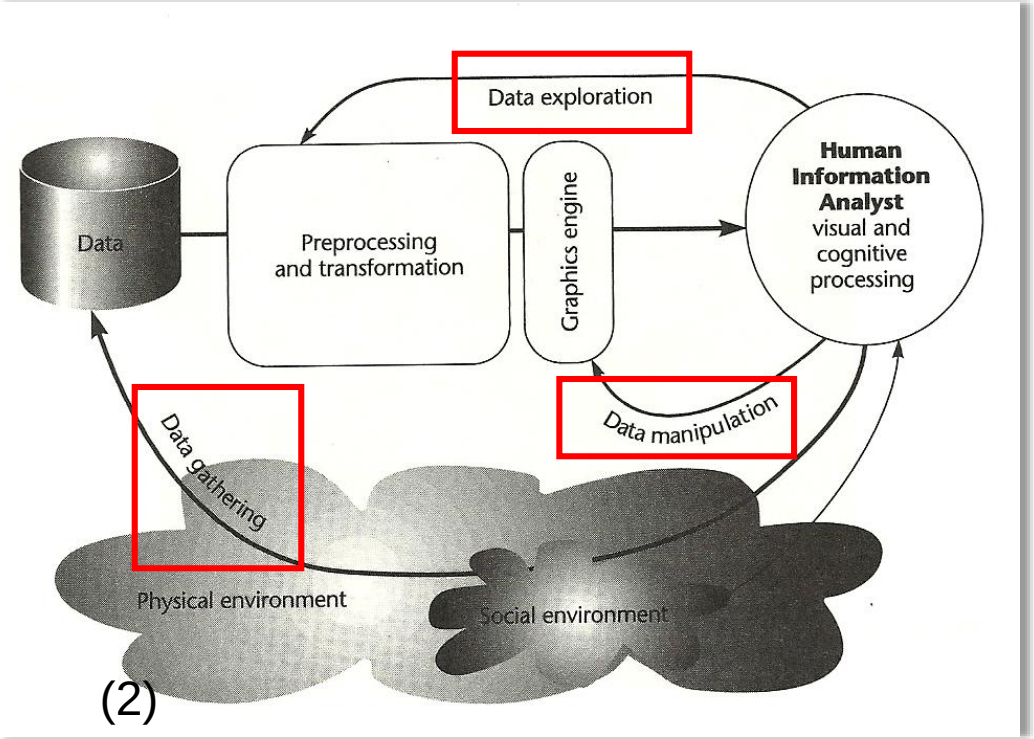
The domain



Telea, A.C. *Data Visualization – Principles and Practice*. Wellesley, MA: A K Peters, 2008

The domain

- Visualization is the communication of information using graphical representations [1].

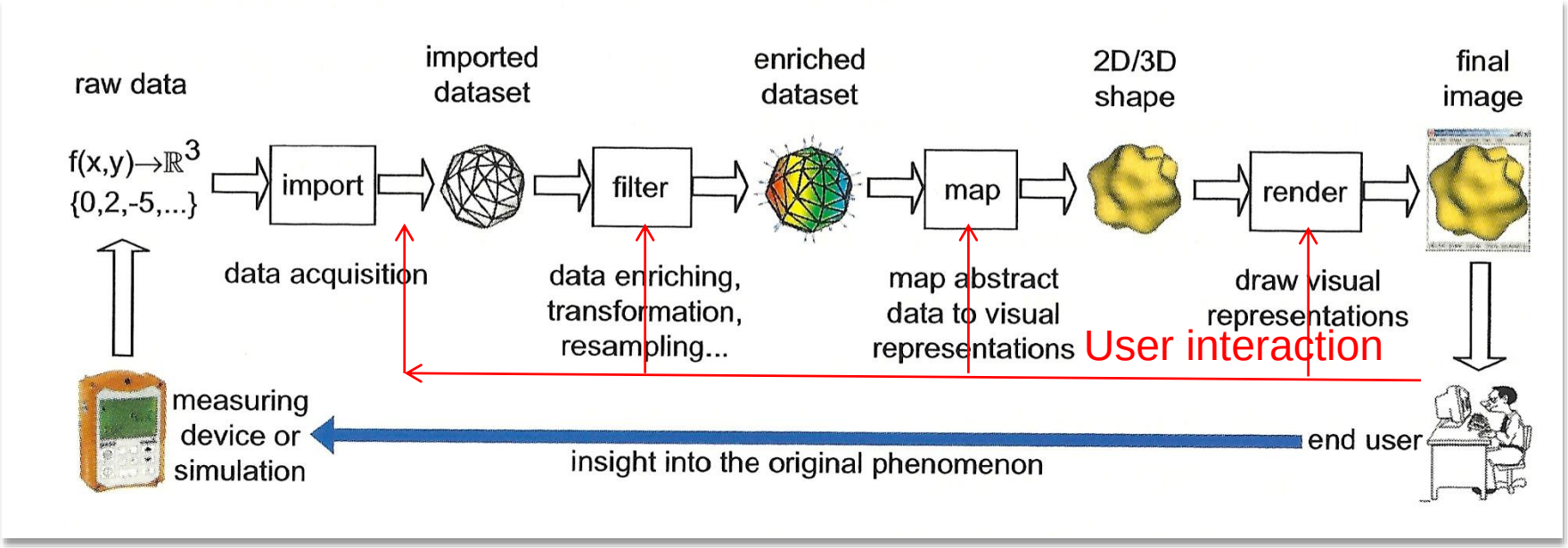


(2)

(1) Ward, M., Grinstein, G e Keim, D. *Interative Data Visualization – Foundations, Techniques and Applications*. Wellesley, MA: A K Peters, 2010

(2) Ware, Colin. *information Visualization – Perception for Design*. San Francisco, CA: Morgan Kaufmann, 2000

The domain



Telea, A.C. *Data Visualization – Principles and Practice*. Wellesley, MA: A K Peters, 2008

Principles of data visualization

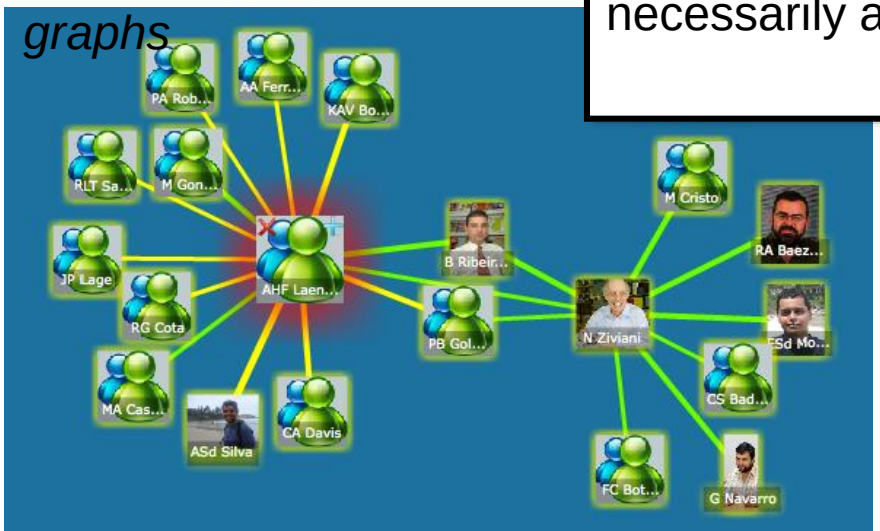
- Characterizing data
- User perception
- Users task and interaction

Scientific data: information associates with positions/regions of a space (implicit or explicit geometry)

Abstract data: information associated to an entity of an application domain, not necessarily a spatial one

Trajectories

Social graphs



Characterizing data

- Many classifications…

Keller e Keller (1994)

- Scalar (or scalar fields
- Nominal
- Direction
- Shape
- Position
- Region

Keller, P. e Keller, M. *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press, 1994

Shneiderman (1996)

- Unidimensional
- Bidirectional maps
- Tridimensional world
- Temporal
- Multidimensional
- Trees
- Networks

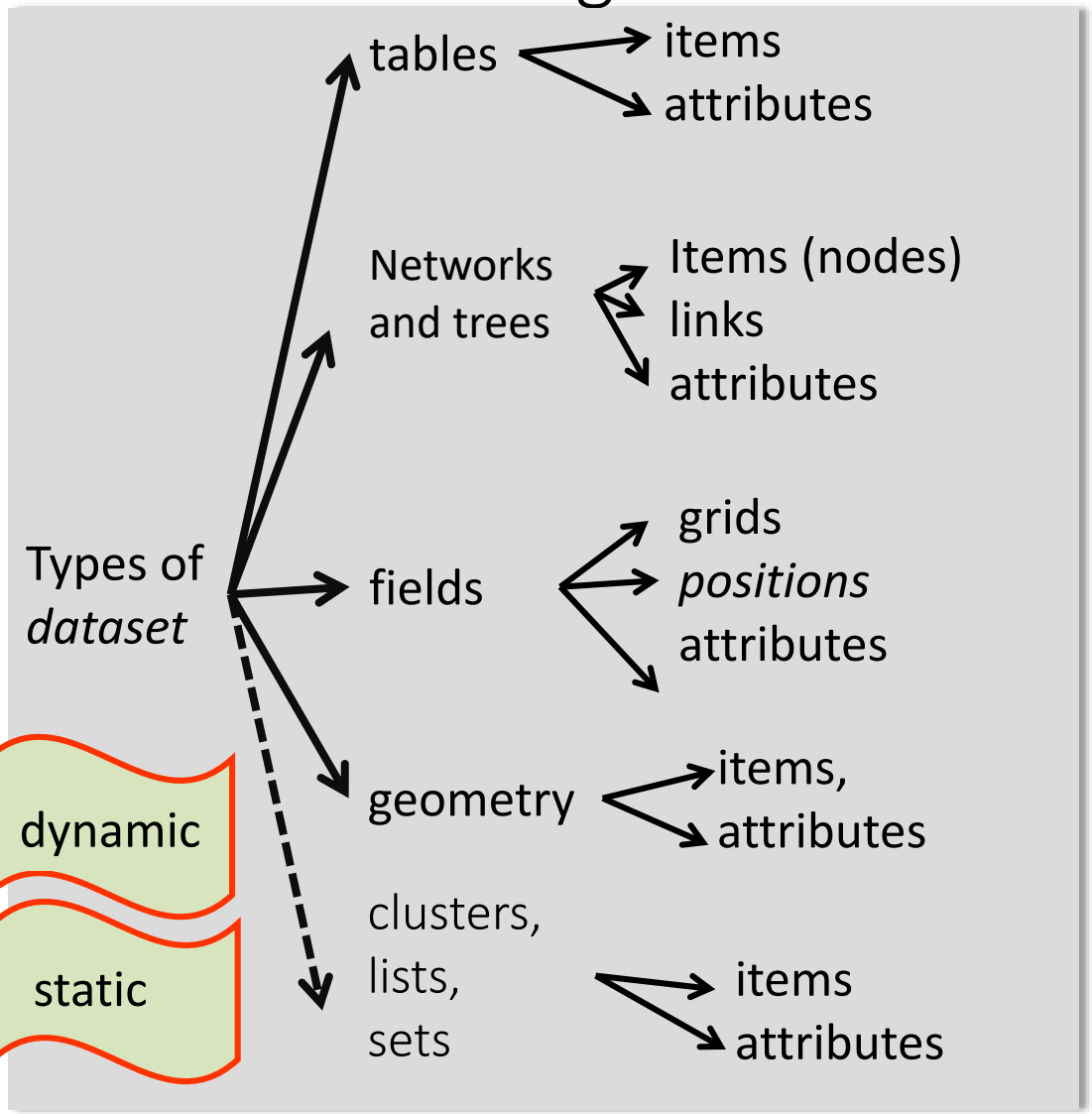
Shneiderman, Ben *The Eyes Have it: A Task by Data Type Taxonomy for Information Visualization*. 1996 IEEE Symposium on Visual Language, pp336-343

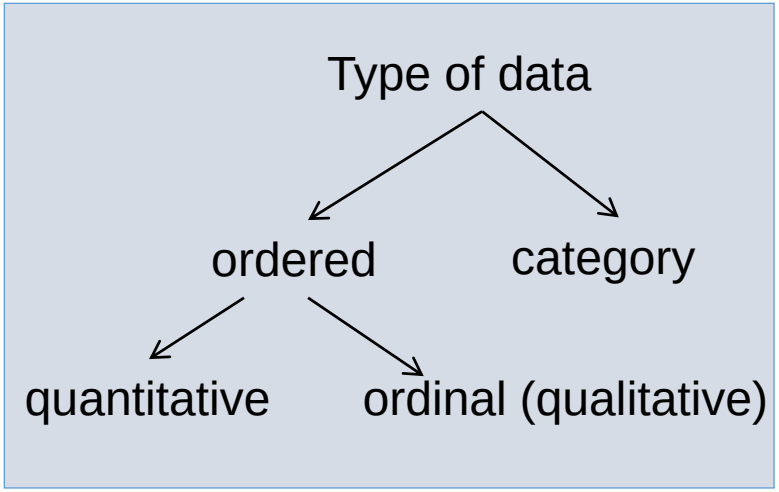
Keim (2002)

- Unidimensional
- Dimensional
- Multidimensional
- Text and hypertext
- Hierarchy and graphs
- Algorithms and software

Keim, Daniel *Information Visualization and Data Mining*. IEEE Transactions on Visualization and Graphics, 8:1(2002):1-8

Characterizing data

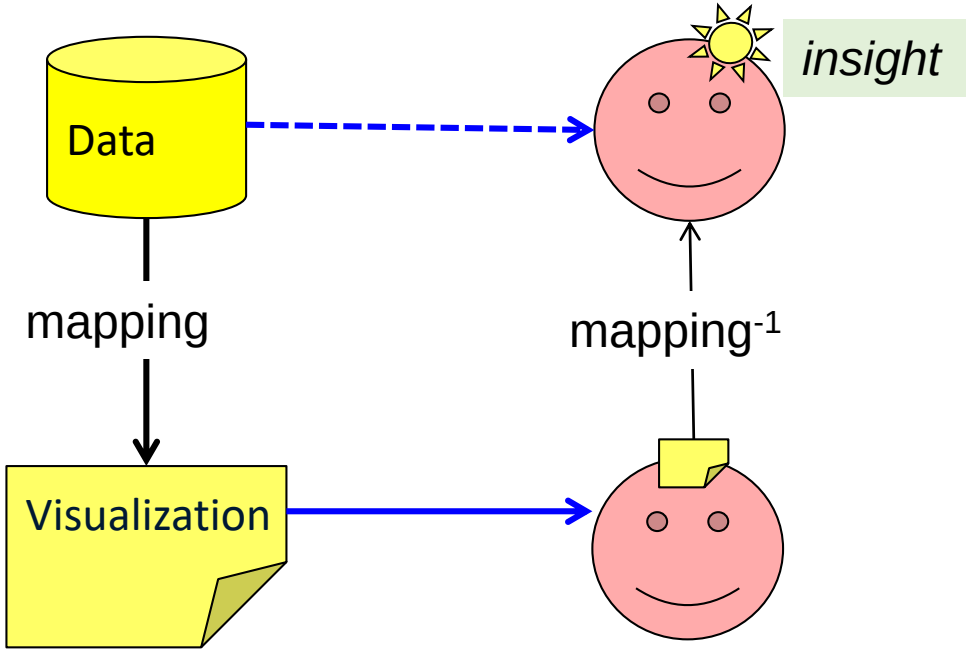




Munzner, Tamara *Visualization Analysis and Design*. Boca Raton, FL:Taylor & Francis, 2014

Principles of data visualization

- Characterizing data
- **User perception**
- Users task and interaction

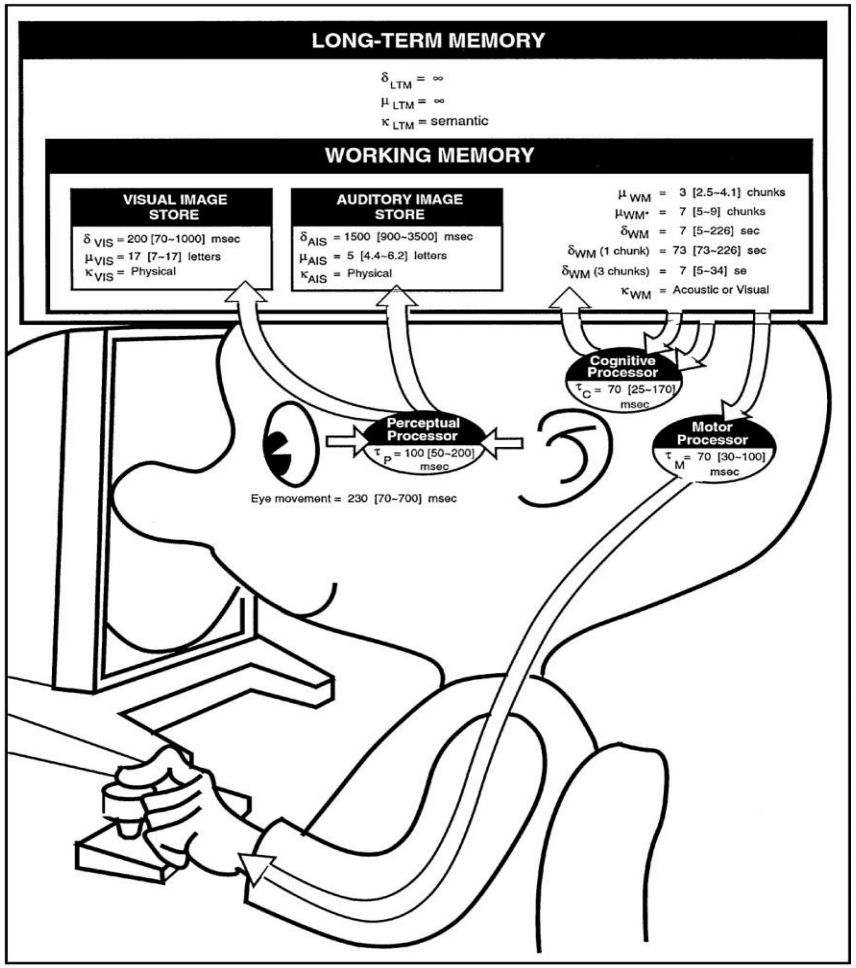


Process by which we interpret what is around us to create a mental representation of evidences

recognize, organiza and interpret sensorial information

Model of Human processing

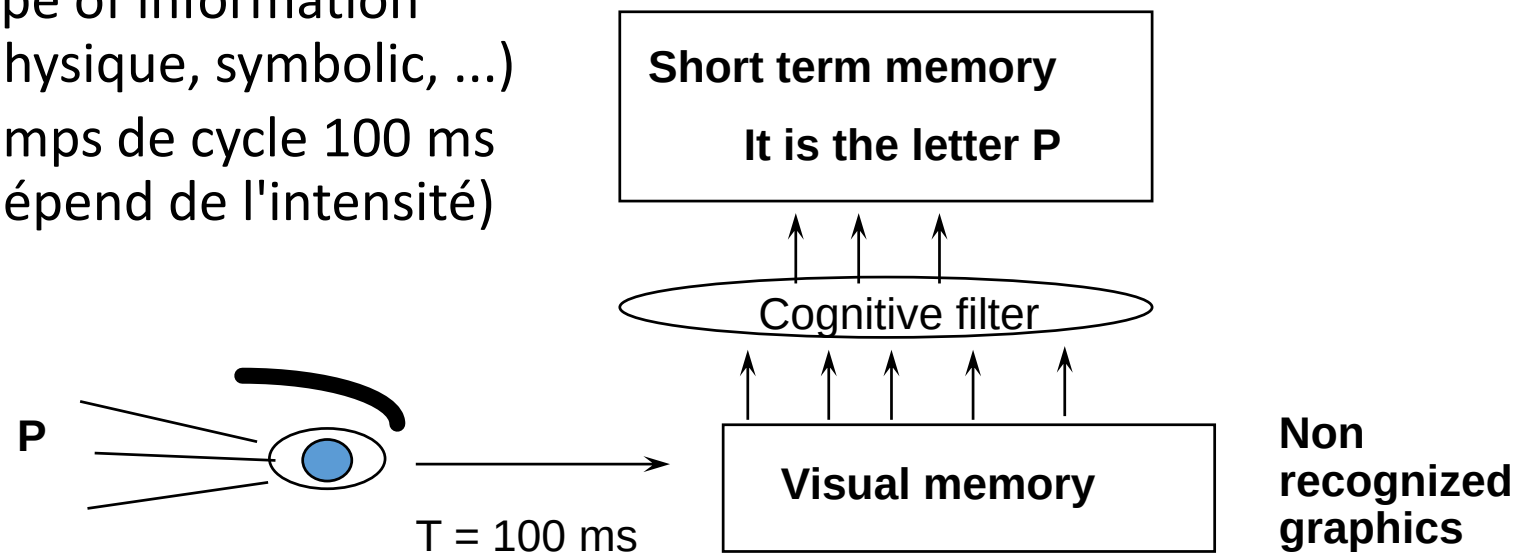
- Perceptive (every sense)
- Cognitive (memory + processing)
- Motor (movements, ex. Fitts' law)



Card, S.K; Moran, T. P; and Newell, A. *The Model Human Processor: An Engineering Model of Human Performance.* In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), **Handbook of Perception and Human Performance.** Vol. 2: Cognitive Processes and Performance, 1986, pages 1–35.

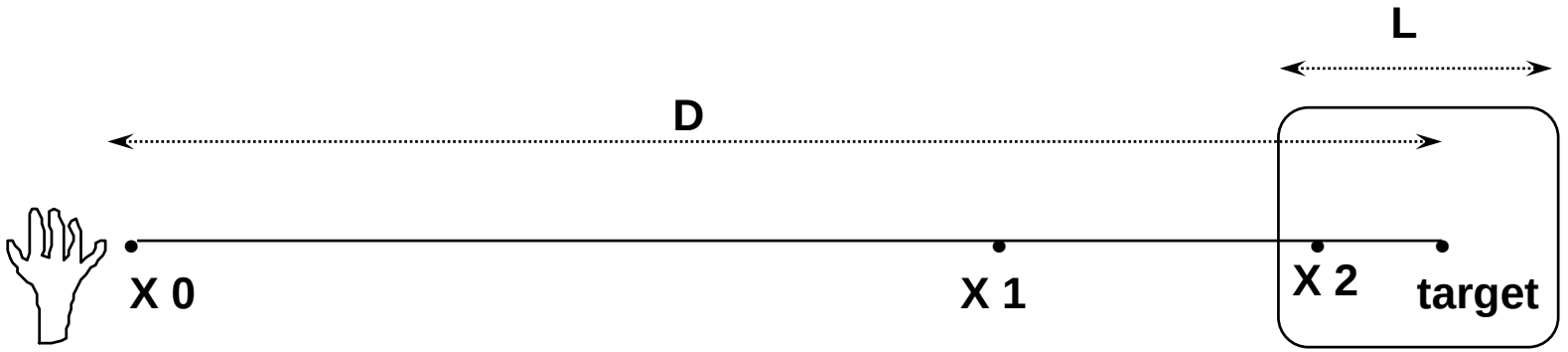
Perceptive system

- Representation non interpreted of inputs
- Information persistency = 200 ms for the visual memory, 1500 ms for the auditory memory
- Capacity of storage
- type of information (physique, symbolic, ...)
- temps de cycle 100 ms (dépend de l'intensité)



Motor system

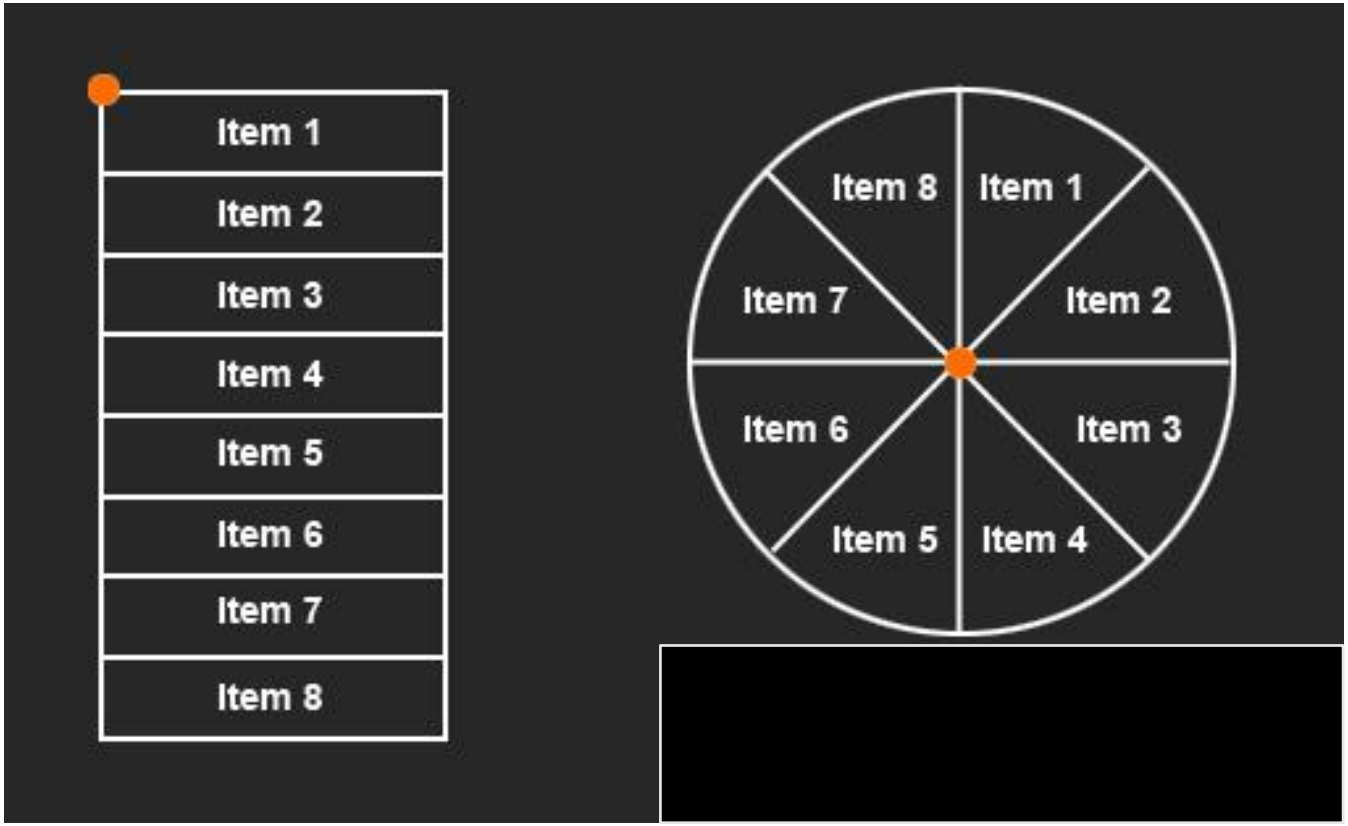
- Movements are not continuous pas a repetition of discreet micro-movements
- The movement corresponds to user interaction with physical devices
  - Time of micro-movement : 70 ms (cycle base of motor system)
  - Time of selection of a graphical element: T = I.log 2D/L with D : distance, L : size of the target, I = 0,1 sec. (Fitts' law 1954)



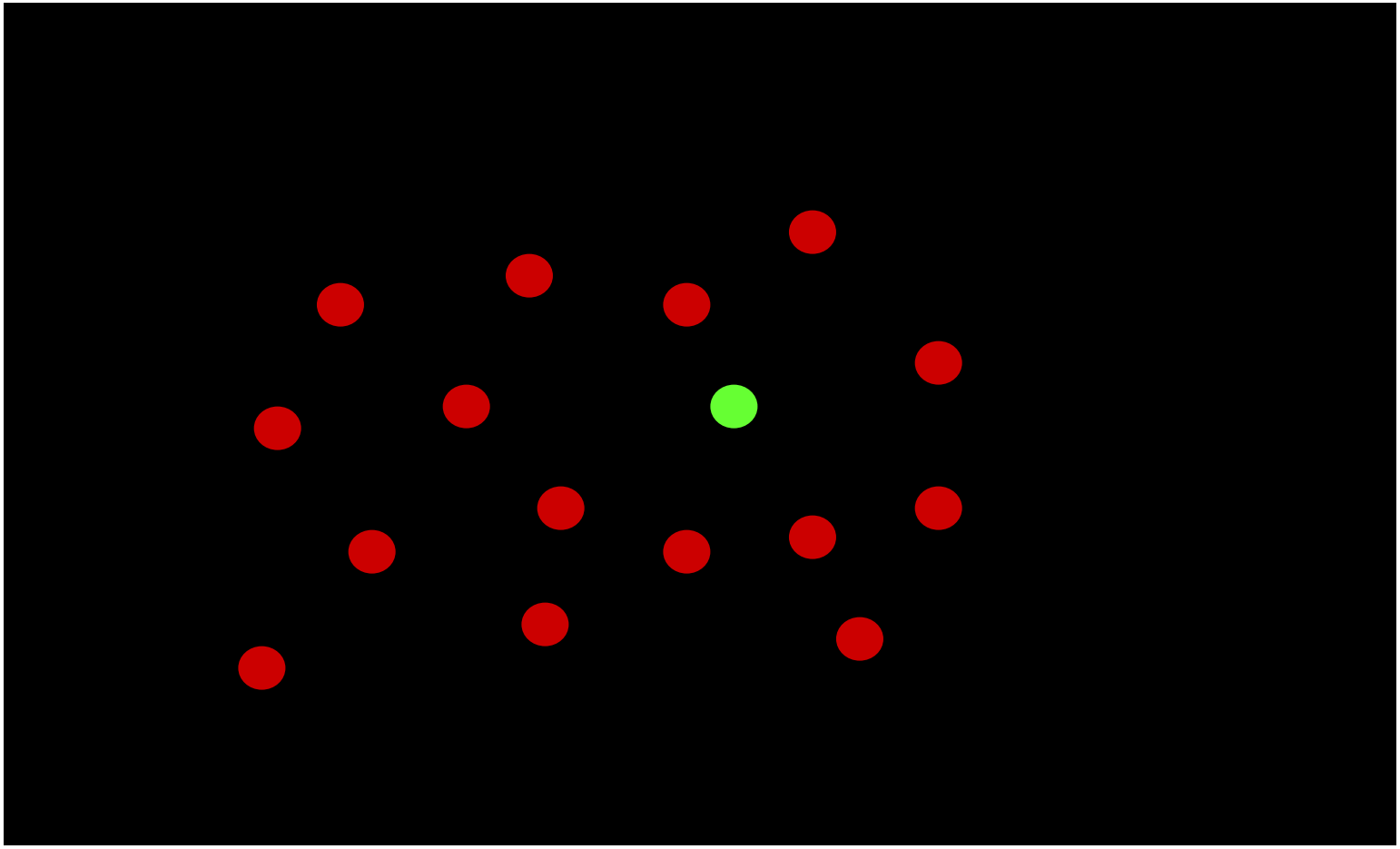
Paul M. Fitts (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, volume 47, number 6, June 1954, pp. 381–391. (Reprinted in *Journal of Experimental Psychology: General*, 121(3):262–269, 1992)

$$T = a + b \log_2 \left(1 + \frac{2D}{W}\right)$$

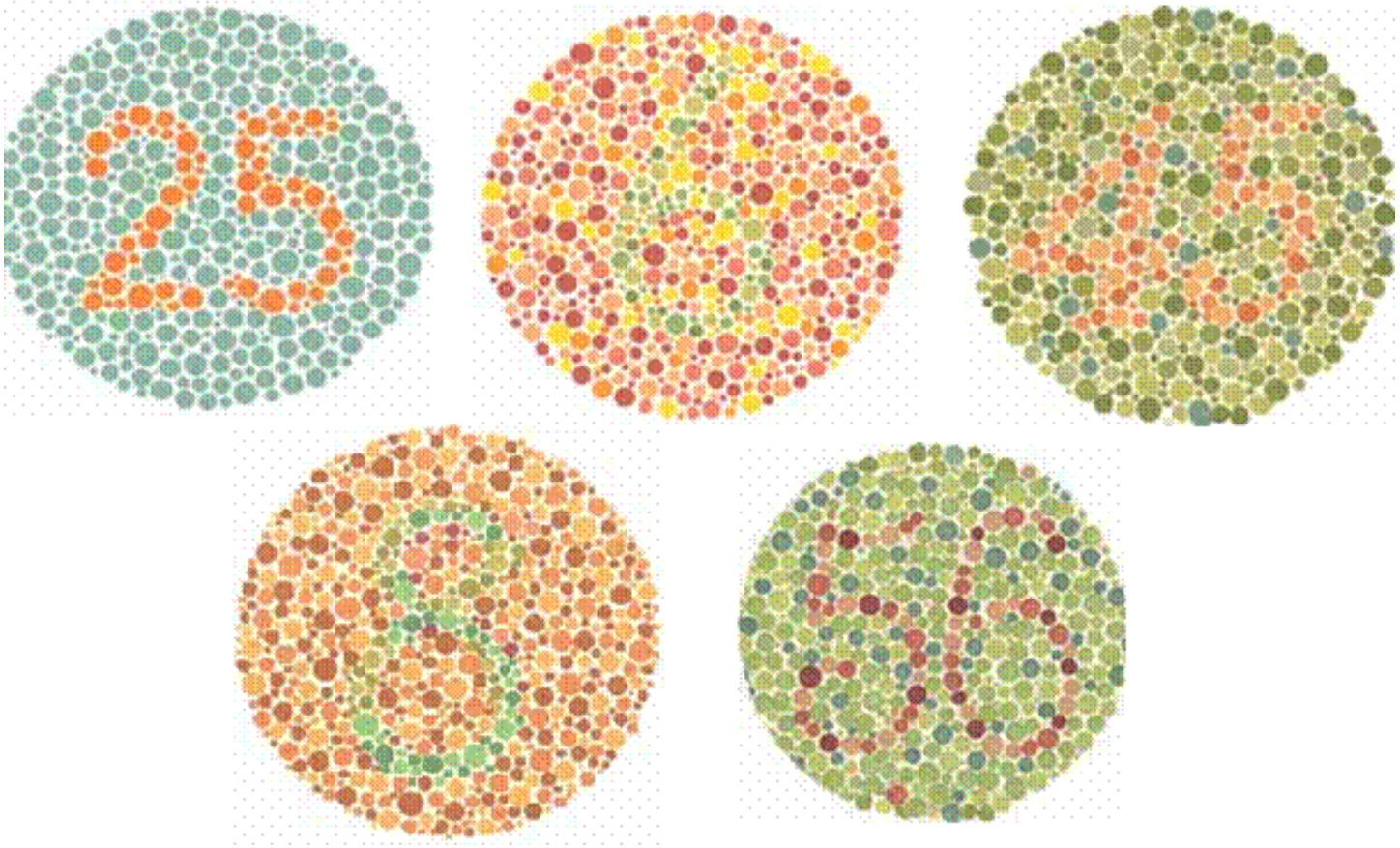
Example


Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7
Item 8
Item 8
Item 1
Item 7
Item 2
Item 6
Item 3
Item 5
Item 4

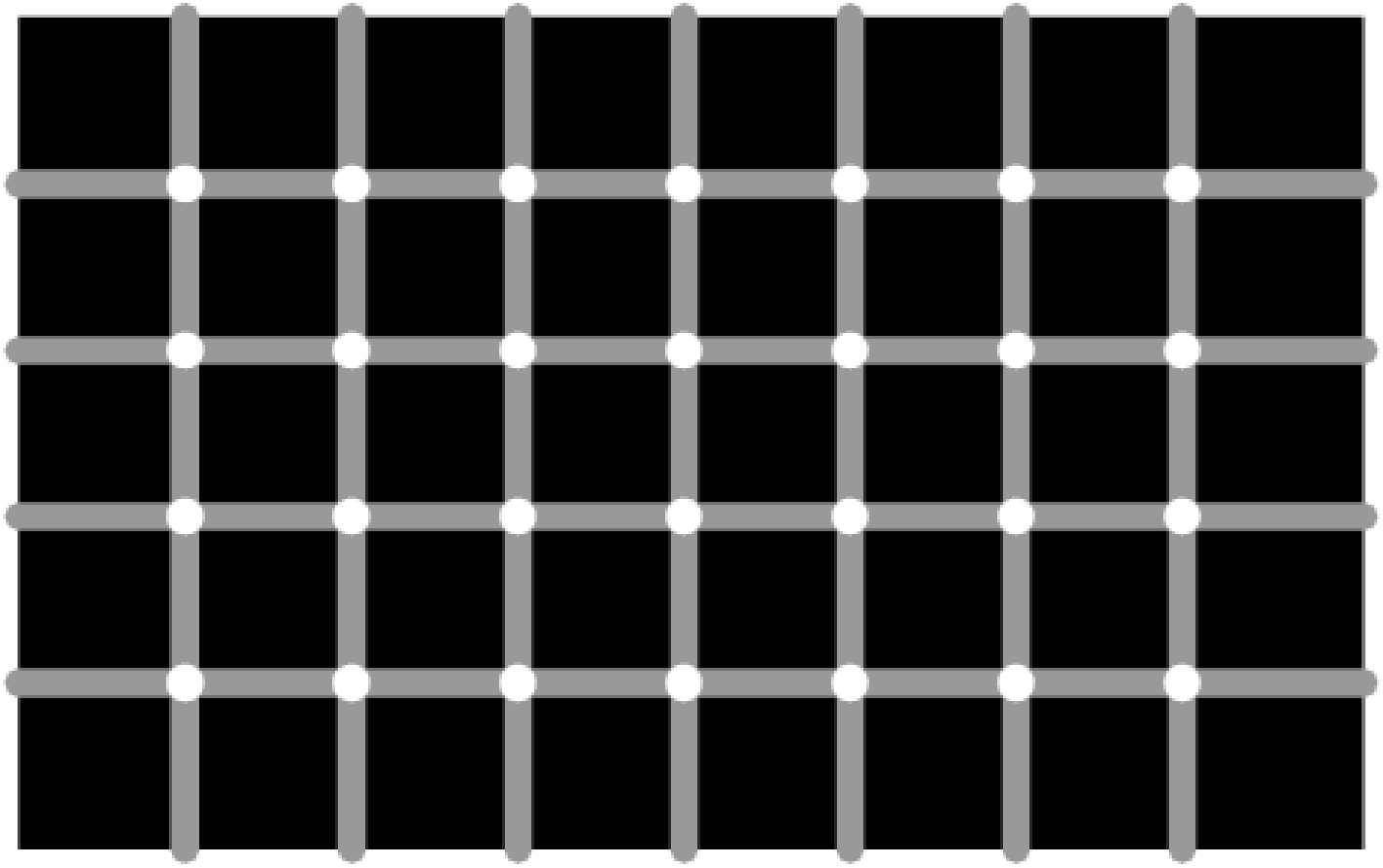
Perception



Perception



Perception

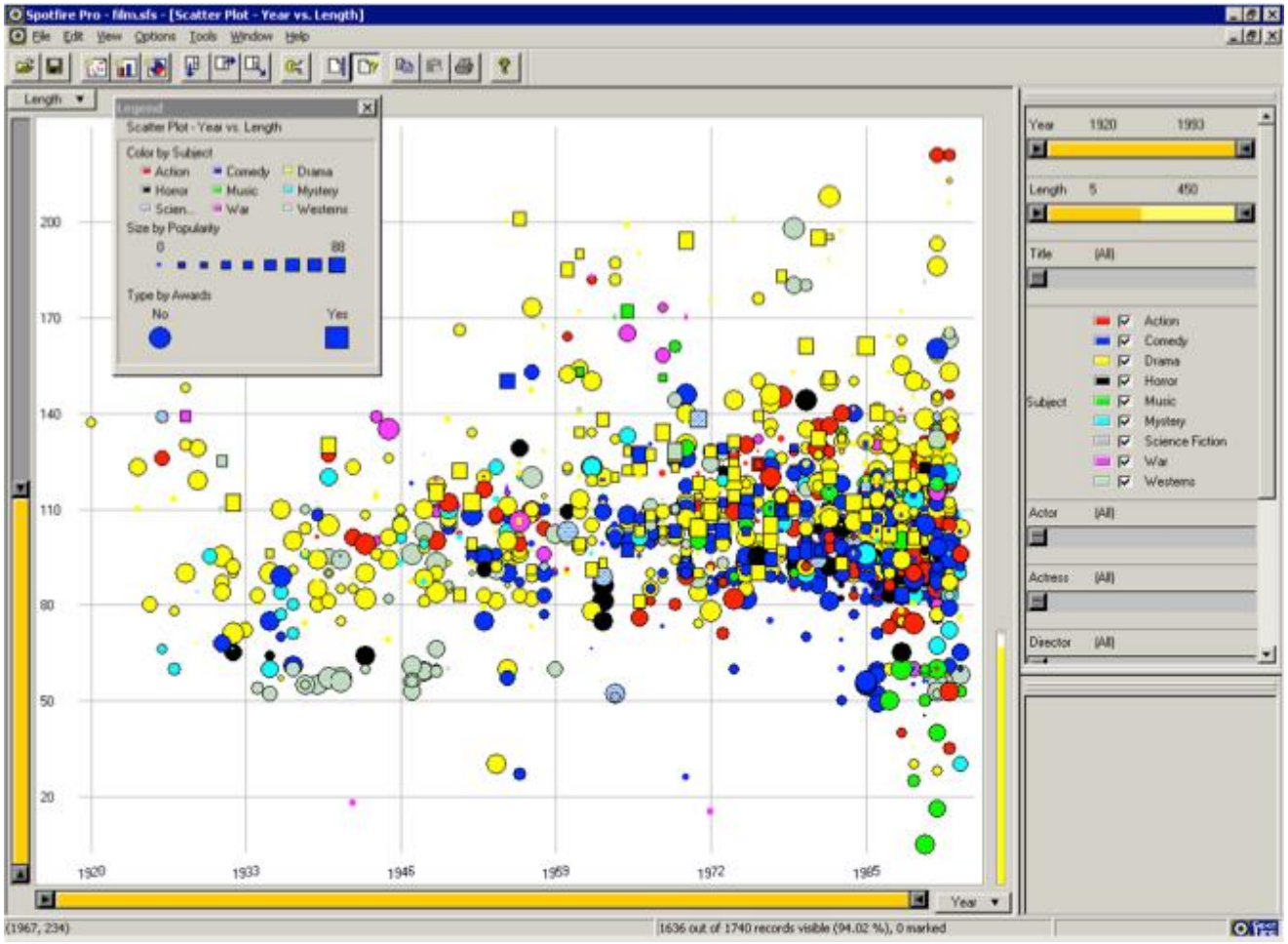


Perception: visual variables

- Position
- Shape
- Size
- Bright
- Color
- Orientation
- Texture
- Movement

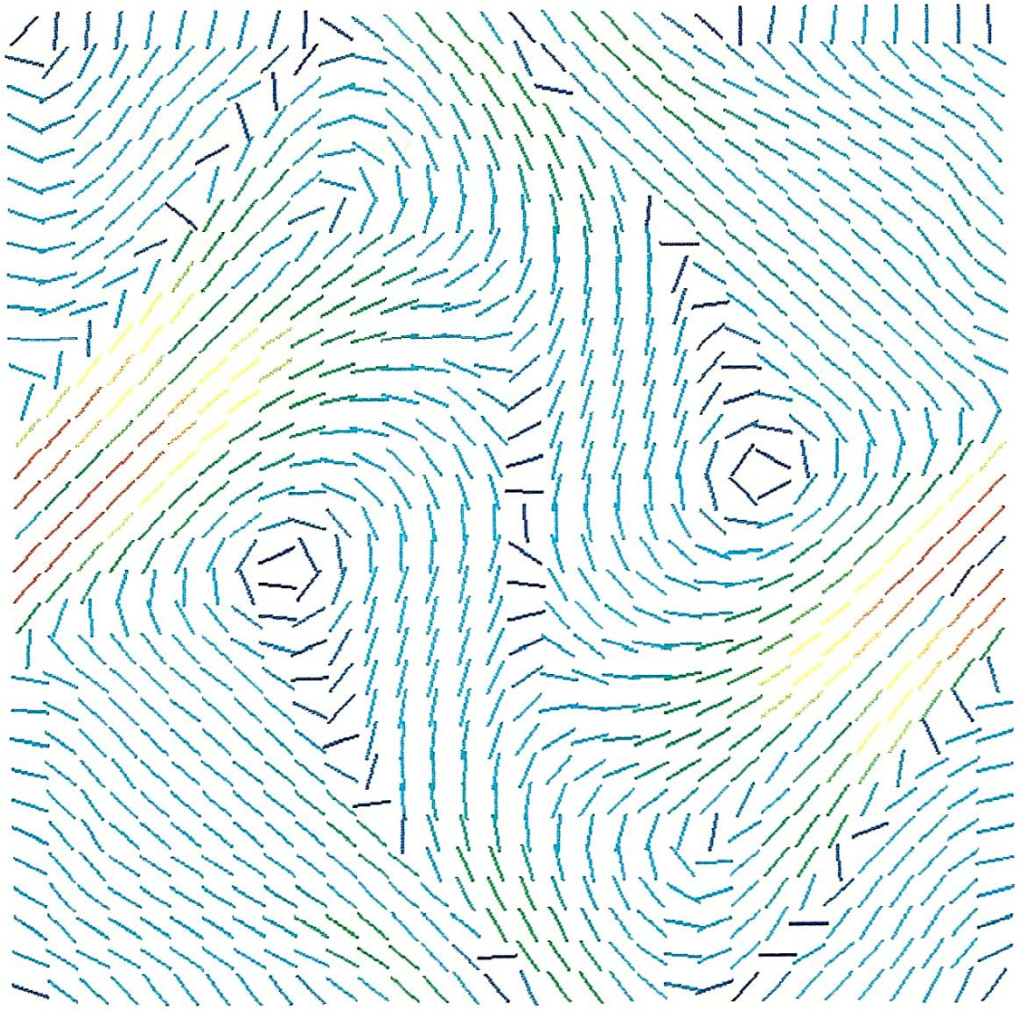
Perception: visual variables

- Position
- Shape
- Size
- Bright
- Color
- Orientation
- Texture
- Movement



Perception: visual variables

- Position
- Shape
- Size
- Bright
- Color
- Orientation
- Texture
- Movement



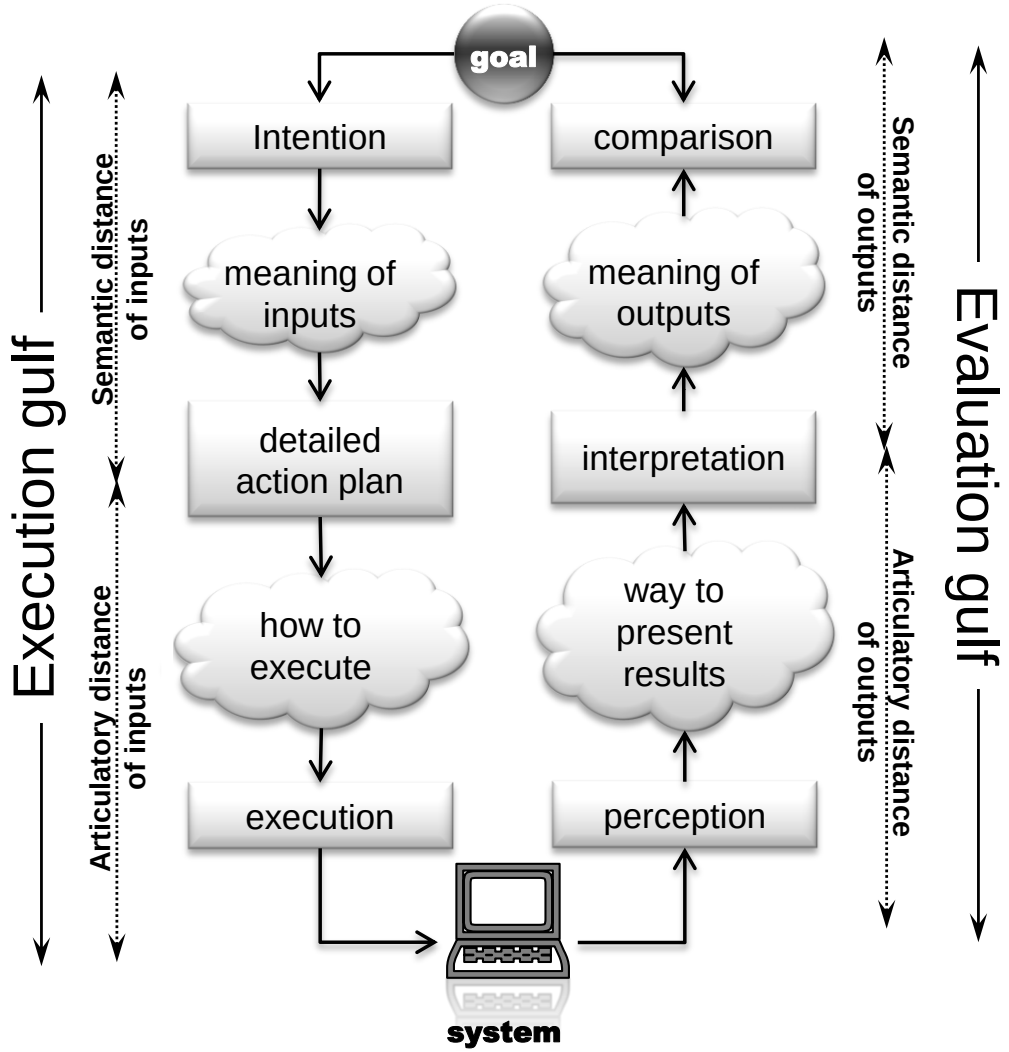
Principles of data visualization

- Characterizing data
- User perception
- Users task and interaction

User interaction

- Changing the scene
  - Selection
  - Navigation
  - Reordering/reorganizing
  - Changing visual coding
- Latency
- Feedback
- Costs
  - Time and user attention

Interaction gulfs *(according to D. Norman, 1986)*



Execution gulf is the effort required for a user to express an intention in terms of commands or instructions

Evaluation gulf refers to the way the results provided by the system are meaningful or understandable by the users, and in accordance with their goals

User tasks (for visualization)

Keller e Keller (1994)

- Identify
- Locate
- Distinguish
- Categorize
- *Cluster*
- Order
- Compare
- Associate
- Correlate

Keller, P. e Keller, M. *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press, 1994

Shneiderman (1996)

- **Overview**
- **“Zoom”**
- **Filter**
- **Details on demand**
- Relate
- History
- Export (data)

Shneiderman, Ben *The Eyes Have it: A Task by Data Type Taxonomy for Information Visualization*. 1996 IEEE Symposium on Visual Language, pp336-343

User tasks (for visualization)

- Wehrend and Lewis, 1990
- Springmeyer, 1990
- Shneiderman, 1996
- Zhou and Feiner, 1998
- Morse et al., 2000
- Amar and Stasko, 2004
- Amar et al., 2005
- Valiati et al., 2006

Low level analytical tasks

- Search value
- Filter dados
- Compute value
- Find limits
- Classify
- Determine range
- Characterize distribution
- Find annomalies
- *"Cluster"*
- Correlate

Exemple of techniques

Conclusion

- Many applications
  - Scientific ones (information processing, nature, social, …)
  - Application in industry (improving and monitoring processes)
  - Dedicated to specialized enterprise versus lay people
- Only recent recognition
- Recent expantion
  - 2005/2006 – Visual Analytics
    - Information visualization
    - Techniques for analyzing data

Exercise

- Find the data set and classify it by identifying
  - Variable perceptive (ex. color, bright, etc)
  - Type of dataset and type of data

Perception: visual variables

- Position
- Shape
- Size
- Bright
- Color
- Orientation
- Texture
- Movement

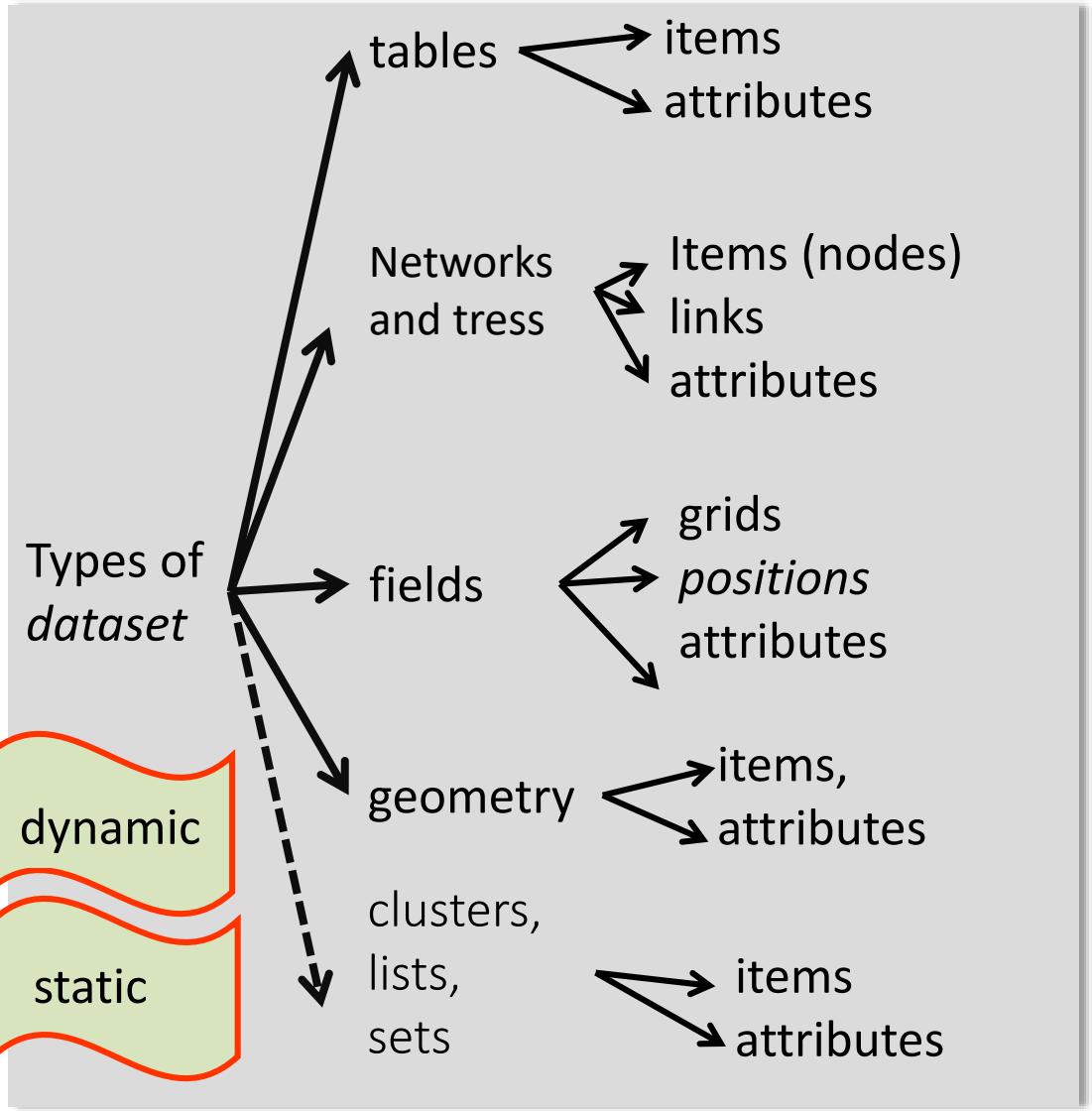
Classes of data

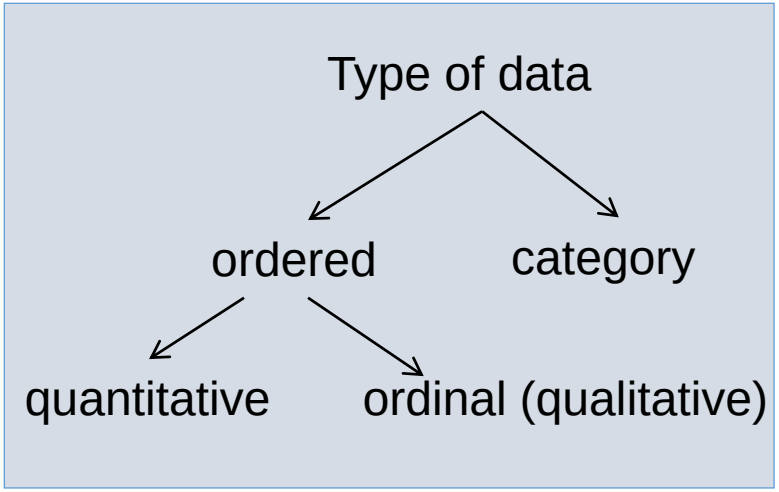
- Scalar (or scalar fields
- Nominal
- Direction
- Shape
- Position
- Region

Dimension of domain

- Unidimensional
- Bidirectional maps
- Tridimensional world
- Temporal
- Multidimensional
- Trees
- Networks

Type of data





Numeric, quantitative or ordered
Categories (items in a enumeration)

Munzner, Tamara *Visualization Analysis and Design*. Boca Raton, FL:Taylor & Francis, 2014