Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

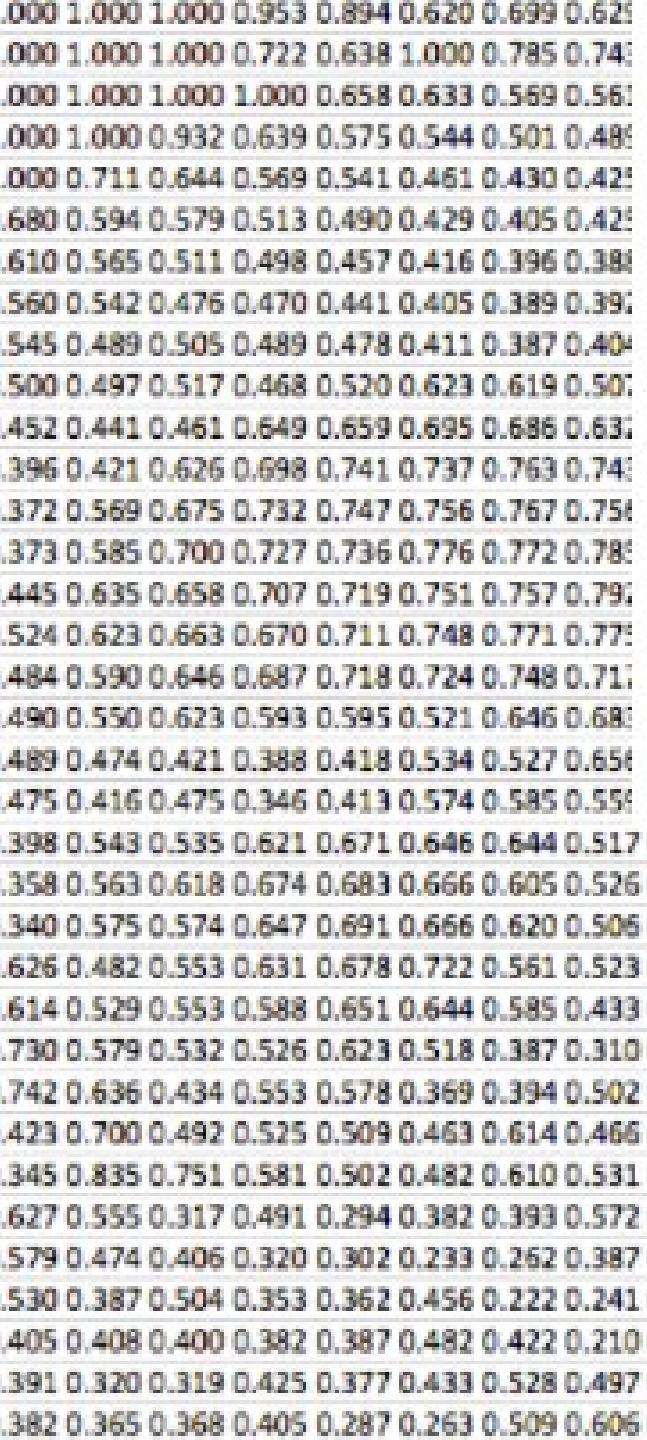
ECIU

An Introduction to Data and Information Visualization

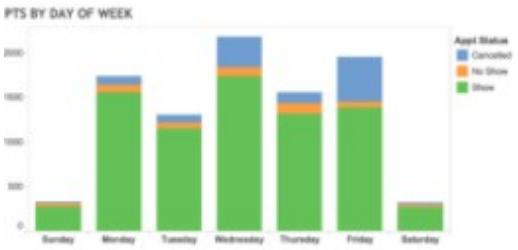
Beatriz Sousa Santos
October/ 2021

CLAD
Associação Portuguesa de Classificação e Análise de Dados

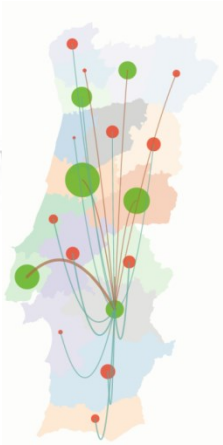
The problem …

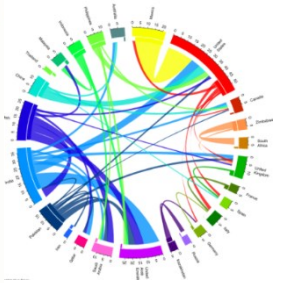


Machine learning?



Visualization?



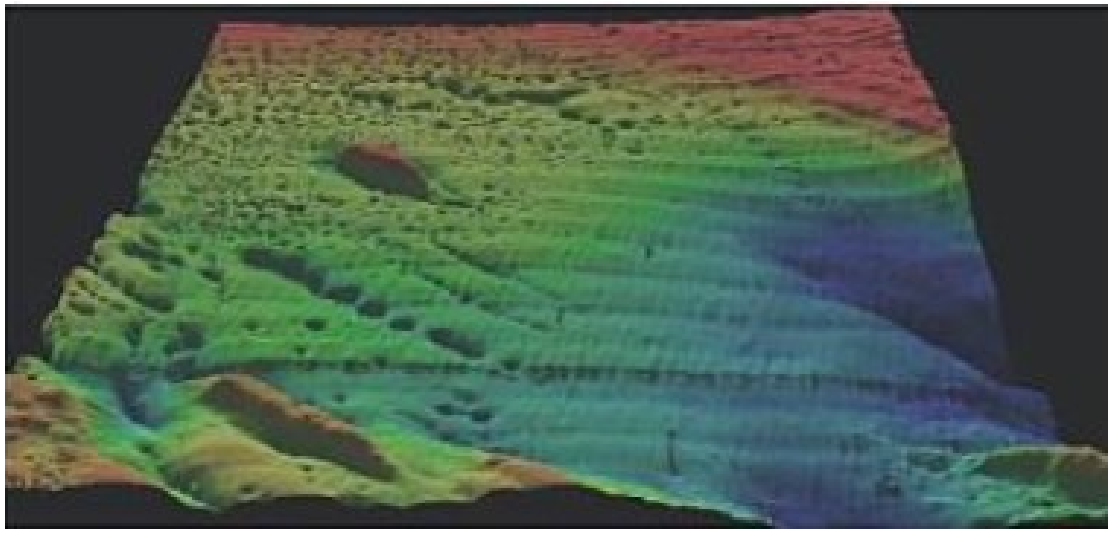


Statistics?

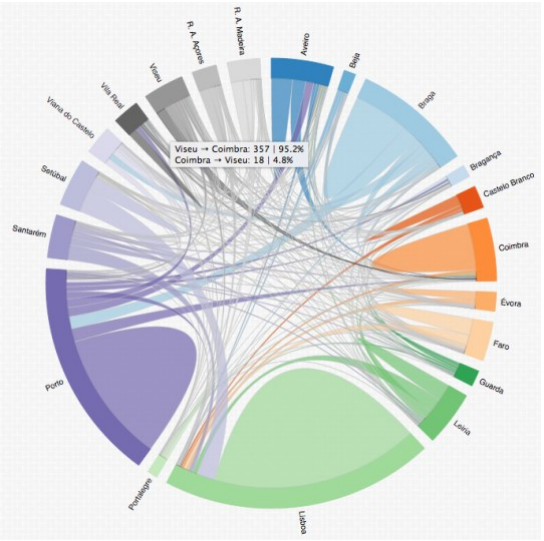
Motivation

What is Visualization?

Is the process of exploring, transforming and **representing data** as images to **gain insight into phenomena**



Passamoquoddy Bay
(10^6 measures)
(Ware 2019)



Portuguese Higher Education
(data from 120 000 candidates)

Why and how to represent data visually?

- The human visual system is a most powerful pattern seeker

  “seeing is understanding…”

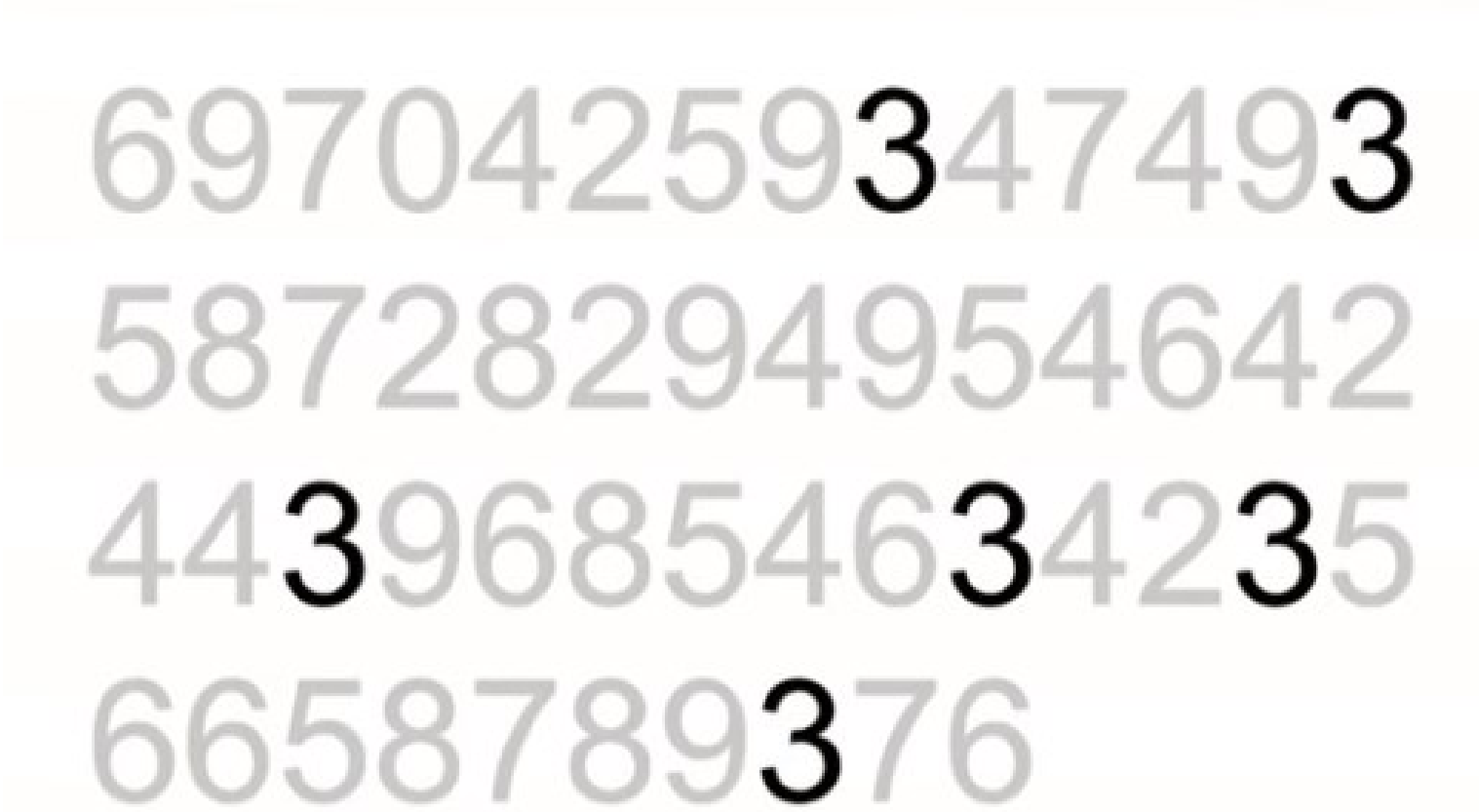
- We easily see patterns displayed in certain ways

  but not in others …

(Ware 2019)

An exercise in preattentive processing: how many “3”?

(Nussbaumer Knaflic, 2015)



C. Nussbaumer Knaflic, Storytelling with Data ,Talks at Google, 2015
https://www.youtube.com/watch?v=8EMW7io4rSI

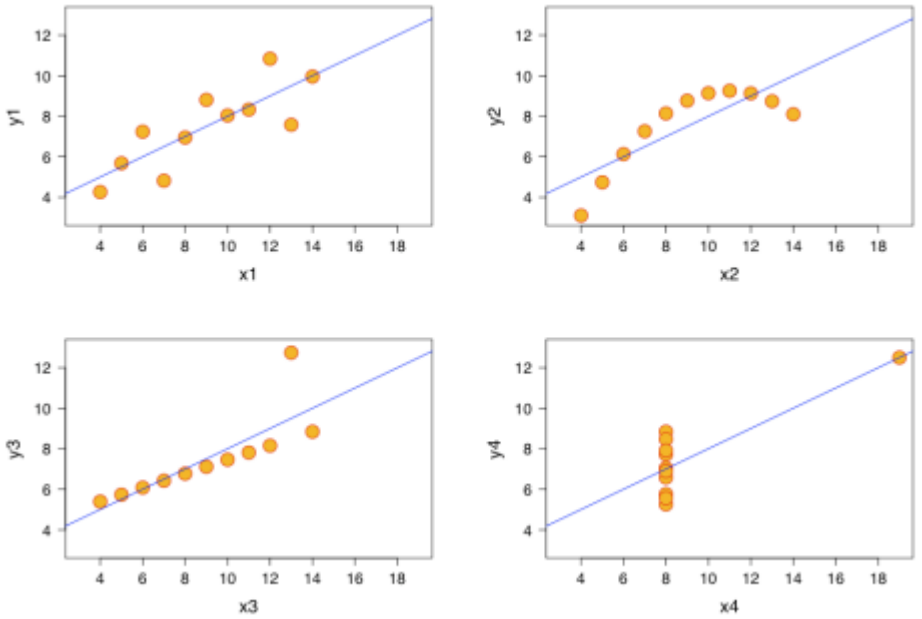
Why show the data in detail?
and not only aggregated parameters

Visualization helps in situations where seeing the dataset structure in detail is better than seeing only a brief summary of it (loosing information). (Munzner, 2014)

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	

Ascombe quartet: data sets with same simple statistical properties (Tufte, 1983)



Outline of the course:

- Introduction to Data and Information Visualization: definitions, historical perspective and relationship to other fields
- Reference model, phases of the visualization process and solution design
- Data: main aspects and visual encoding
- Representation, Presentation, Interaction – concepts, a taxonomy of techniques, and examples
    - representation of value, and relation – common techniques
    - presentation and interaction - common techniques
- Evaluation: main aspects and methods
- Tools, Bibliography and other learning resources

-1st session - morning

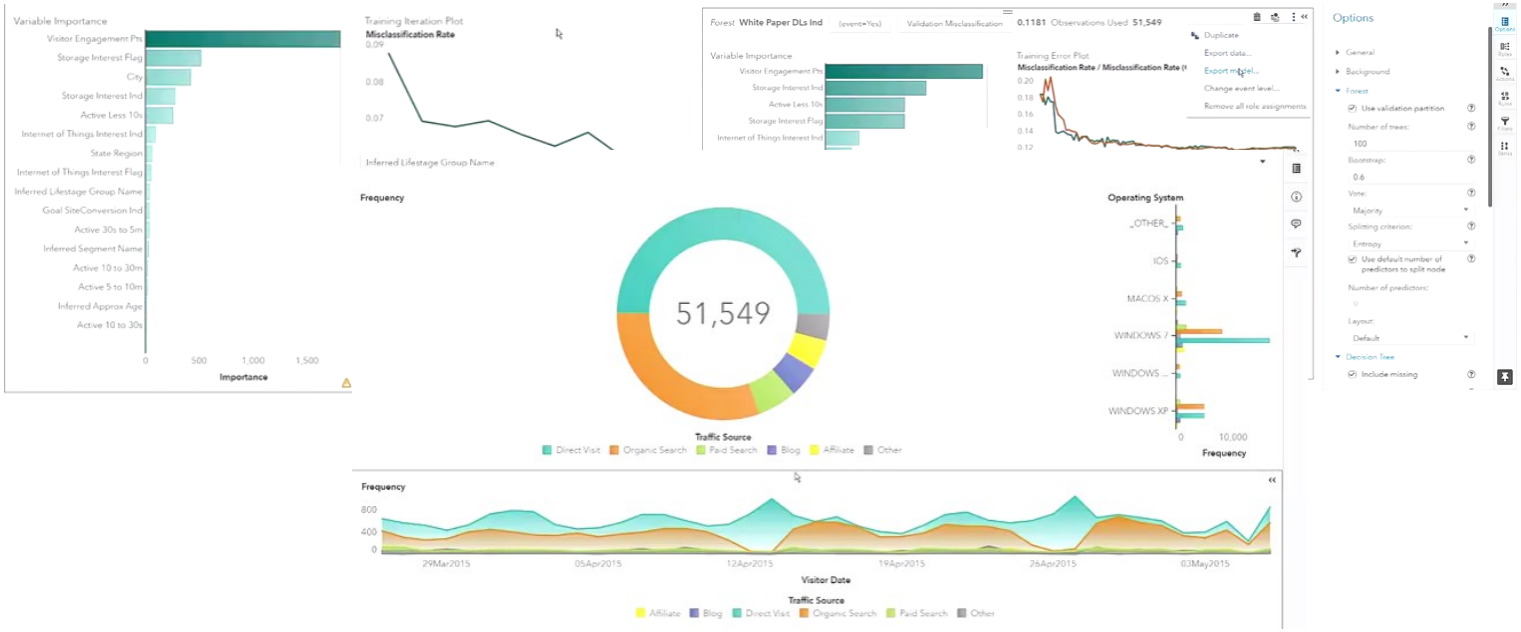
- Introduction to Data and Information Visualization: definitions, historical perspective and relationship to other fields
- Reference model, phases of the visualization process and solution design
- Data: main aspects and visual encoding
- Practical activity: critique of InfoVis examples
- Representation: concepts, a taxonomy of techniques, and examples

2nd session - afternoon

- Presentation and Interaction: concepts and methods
- Evaluation: main aspects and methods
- Visualization S/w and learning resources
- Practical activity: evaluation of a Visual Data Exploration application (group work)

Presentation of findings

Visualization, Visual Data Mining, and related fields



Visualization and Visual Data Mining

- Visualization is a field of Computing focused on how to **visually explore, transform and represent large amounts of data to gain insight into phenomena**

- **Visual representations take advantage of the human eye's broad bandwidth** pathway into the mind

- **Visual Data Mining uses visualization** in decision support to facilitate data exploration and understanding; it involves
  - Selecting data,
  - Transforming,
  - Representing visually

Visualization and Visual Data Mining

- Visualization as a scientific field is the process of **exploring, transforming and representing data as images** (or other sensorial forms) **to gain insight into phenomena**

- Visual data mining techniques are of high value in **exploratory data analysis** (Keim, 2002)

- Specially when **little is known about the data** and the exploration **goals are vague**

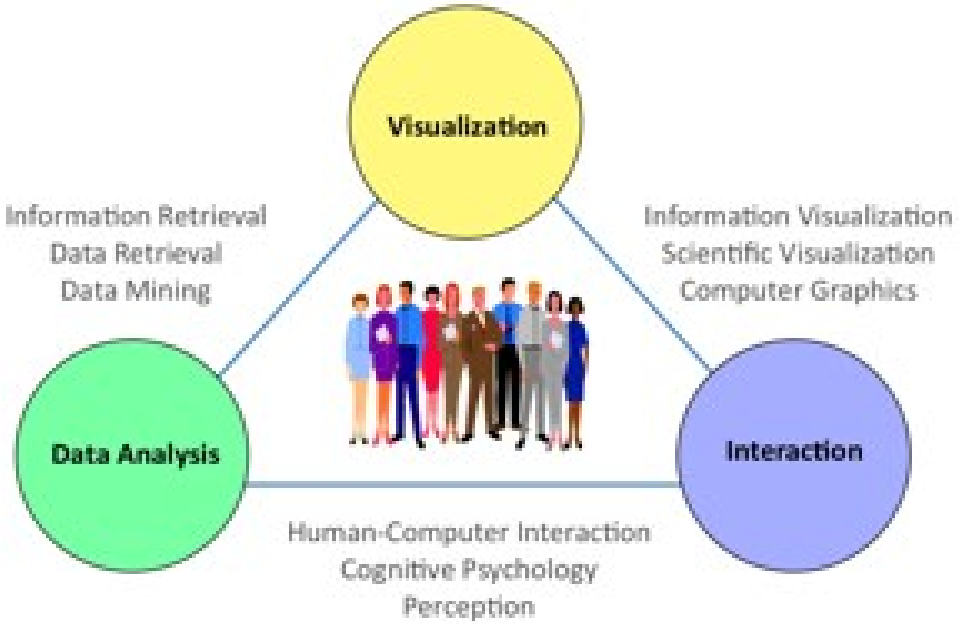
- Since the **user is directly involved**, shifting and adjusting the exploration goals is automatically done if necessary

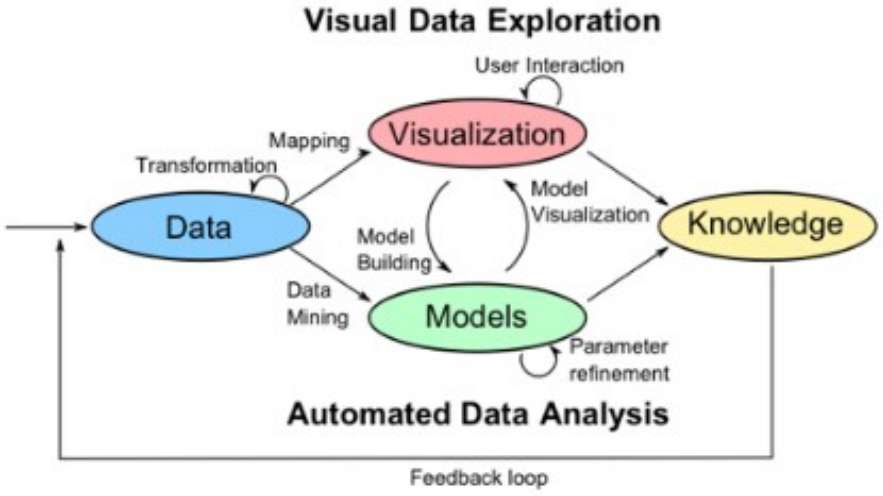
- **Main advantages** of visual over automatic data mining techniques (statistics or machine learning):
  - can easily deal with highly inhomogeneous and noisy data
  - is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters.
- Visual data exploration techniques provide a much higher degree of confidence in the findings of the exploration.
- This makes them indispensable in conjunction with automatic exploration techniques.

Visual Analytics

The **science of analytical reasoning supported by interactive visual interfaces**.

“Detect the expected and discover the unexpected”
Illuminating the Path - The Research and Development Agenda for Visual Analytics
(Thomas and Cook, 2006)





“The Visual Analytics Process **combines automatic and visual analysis methods** with a tight coupling through **human interaction** in order to gain knowledge from data.”

http://www.visual-analytics.eu/faq/

Visualization and Machine Learning

- Information Visualization and Visual Data Mining **leverage the human visual system to provide insight** and understanding of unorganized data

- Machine Leaning and Visualization **share a focus on data** and information

- **The main difference is the role of the user** in the data exploration and modeling:
  - Machine Learning -> has as ultimate goal to get read of the user
  - Information Visualization -> allows the user to discover patterns and adjust models (Keim et al., 2012)

http://drops.dagstuhl.de/opus/volltexte/2012/3506/pdf/dagrep_v002_i002_p058_s12081.pdf

When are Visualization solutions most appropriate?

- to analyze data when people **don't know exactly what questions** they need to ask in advance

- for long-term use, where a **human intends to stay in the loop indefinitely** (e.g. in scientific discovery, medical diagnosis)

- for long-term use to **monitor a system**, so that people can take action if they spot unreasonable behavior (e.g. in stock market)

- for transitional use where the goal is to "**work itself out of a job**", by helping the designers of future purely computational solutions, etc.

Interviews with Netflix Data Scientists:
How important 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

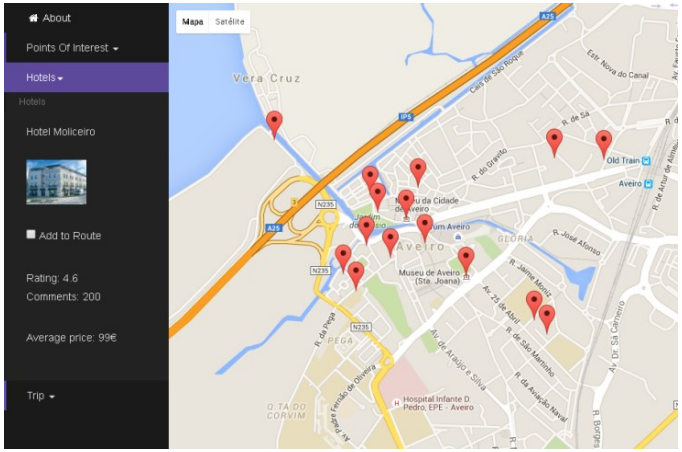
https://classroom.udacity.com/courses/ud404/lessons/9239573934/concepts/92387205290923

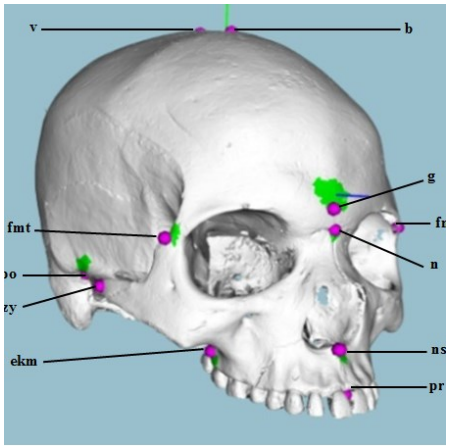
Visualization in the Data Science Process

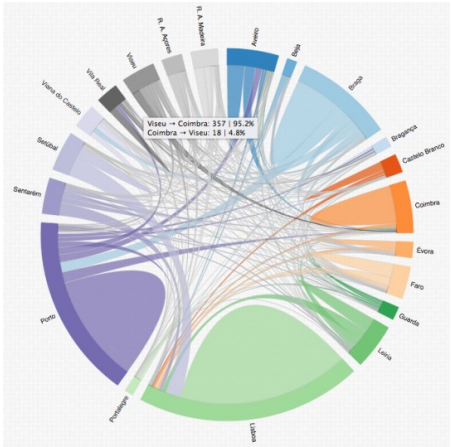
Information Visualization may be useful in several stages:

- Exploring the data
- Selecting the automatic models to use
- Monitoring the performance of the models
- Detecting when they need to be updated
- Explaining the models 
  XAI – Explainable AI:
  recent active trend in AI
- Analyzing the results …

An Introduction to Visualization (Data and Information)







Definition

Objectives

History

Model

How to create and evaluate a Visualization?

Different meanings of "**Visualization**"

Dictionary definition:

1- formation of mental visual images

2- the act or process of interpreting in visual terms or of putting into visible form (Merriam-Webster)

Visualization as a scientific field:

Is the process of exploring, transforming and representing data as images (or other sensorial forms) to gain insight into phenomena

May also be used to denote a specific visual representation of a data set

Objectives of Visualization

Is the process of exploring, transforming and representing data as images (or other sensorial forms) **to gain insight into phenomena**

Computer based visualization systems provide representations of data sets designed to **help people carry out tasks more effectively**
(Munzner, 2014)

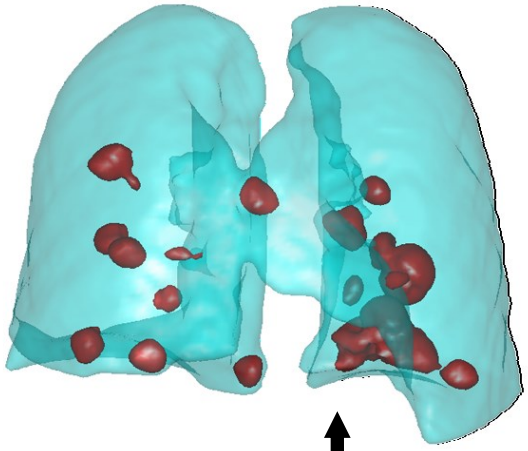
- Visualization as a scientific field:

- The differences among several areas are not completely clear, but
  - **Data/Scientific Vis** – data with 3D/4D physical structure (e.g. CAT, meteorological)
  - **InfoVis** – abstract data without a physical structure (e.g. business, text, S/W)

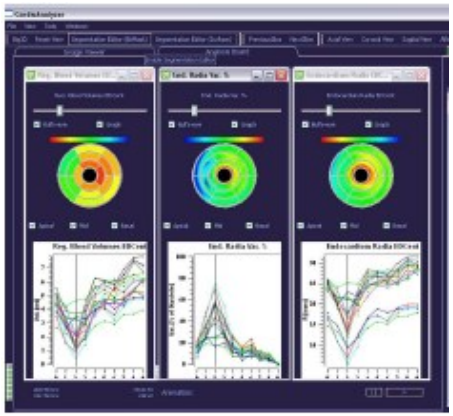
- These designations are misleading; **both start with (raw) data and allow to extract information**

Note: The Industry uses Data Visualization even for abstract data

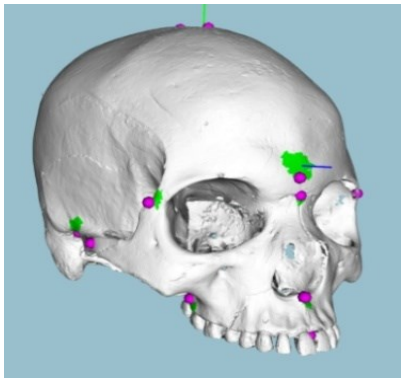
Data/Scientific Visualization **(examples "made in UA")**



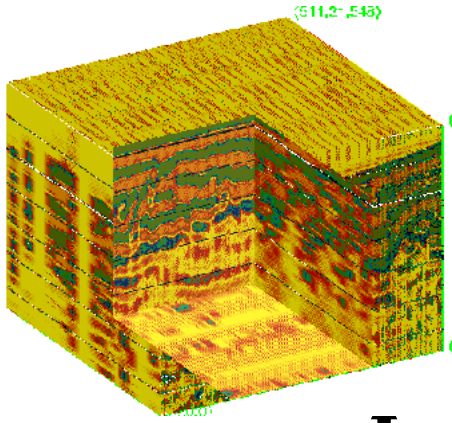
Tomography
(2004)



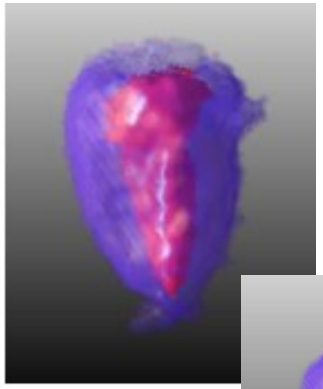
Tomography (2011)

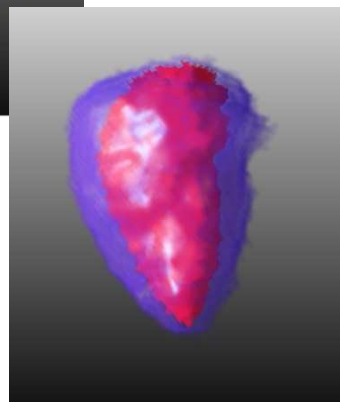


Laser scanner (2016)

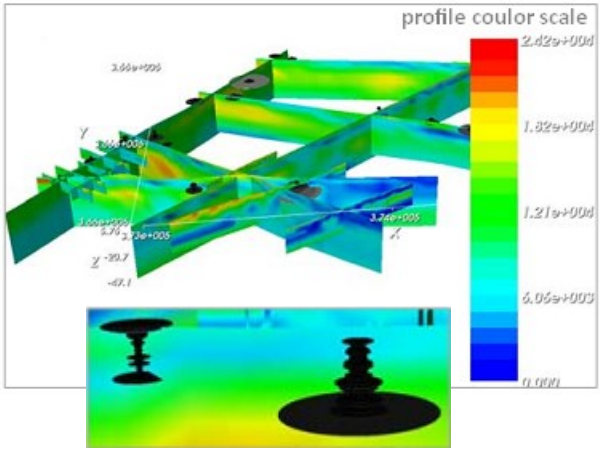


Ground Penetrating
Radar (1999)



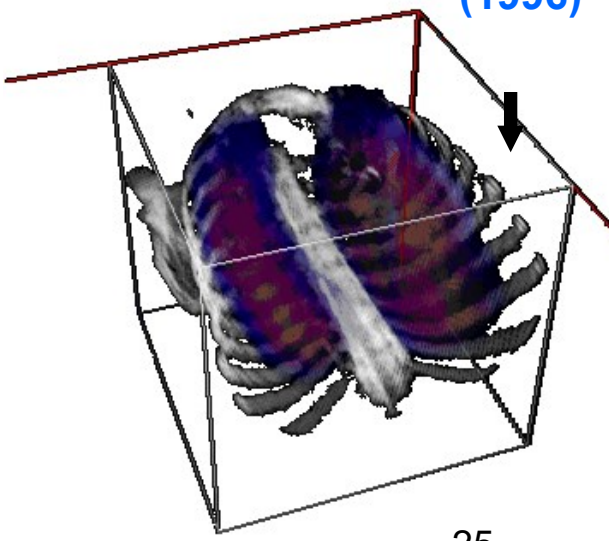


Tomography
(2008)



Electrical and mechanical
ground resistivity (2010)

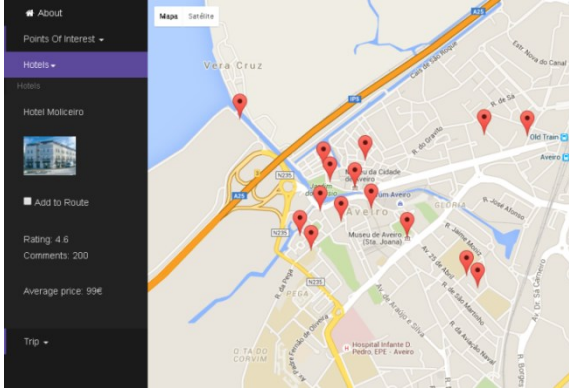
Tomography and SPECT
(1996)



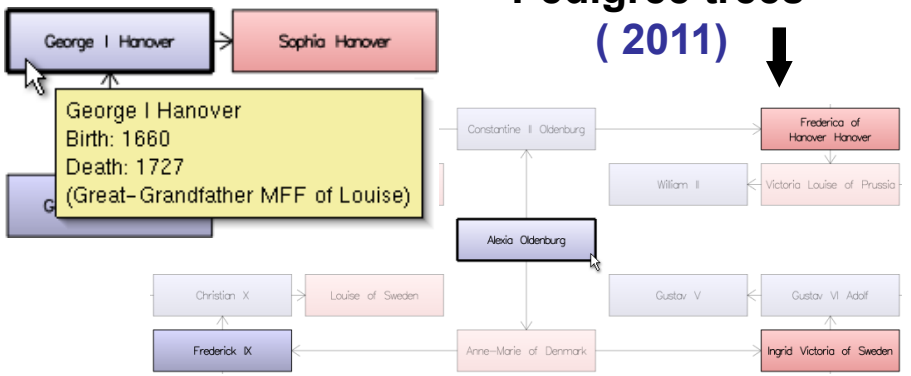
Information Visualization **(examples "made in UA")**

Academic data ⬅ **(2020)**

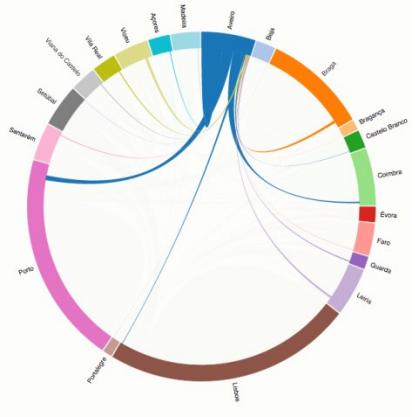
Location Routing in trip planning (2016) ➡

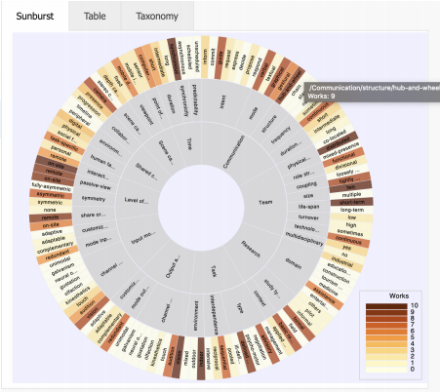


Pedigree trees (2011) ⬇



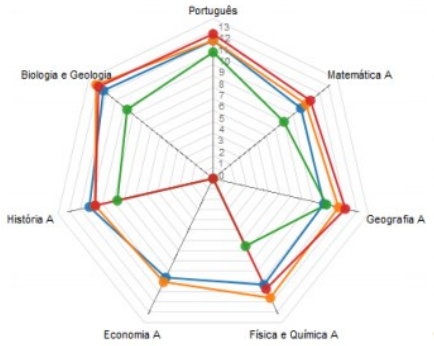
Human Migrations (2015) ⬇

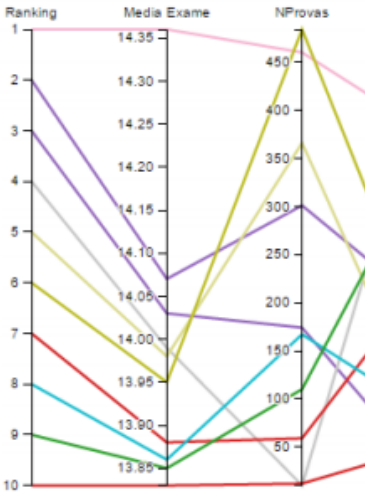


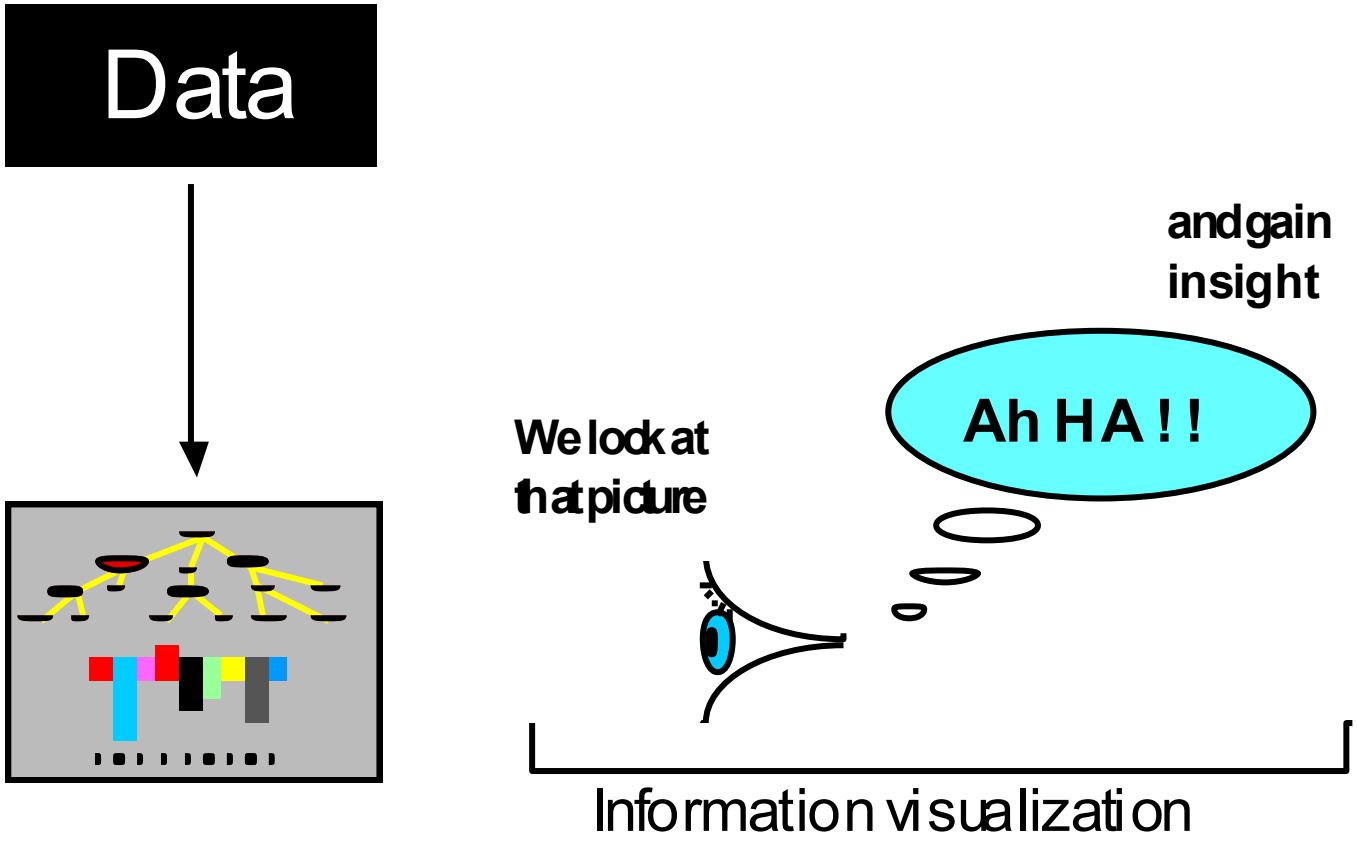


⬆ **Papers Taxonomy (2021)**

School Ranking (2016) ⬇







The process of information visualization: **graphically encoded data** is viewed in order to **form a mental model of that data and obtain insights**

(Spence, 2007)

Brief history

- The **usefulness** of graphical representations of large amounts of data has been recognized **long ago**; important examples are:

  XVIII e XIX centuries- use of graphics in statistics and science:

  W. Playfair, C. J. Minard

  XX century- principles and guidelines: J. Bertin, E. Tufte

- The use of the **computer made Visualization a more practicable** discipline:

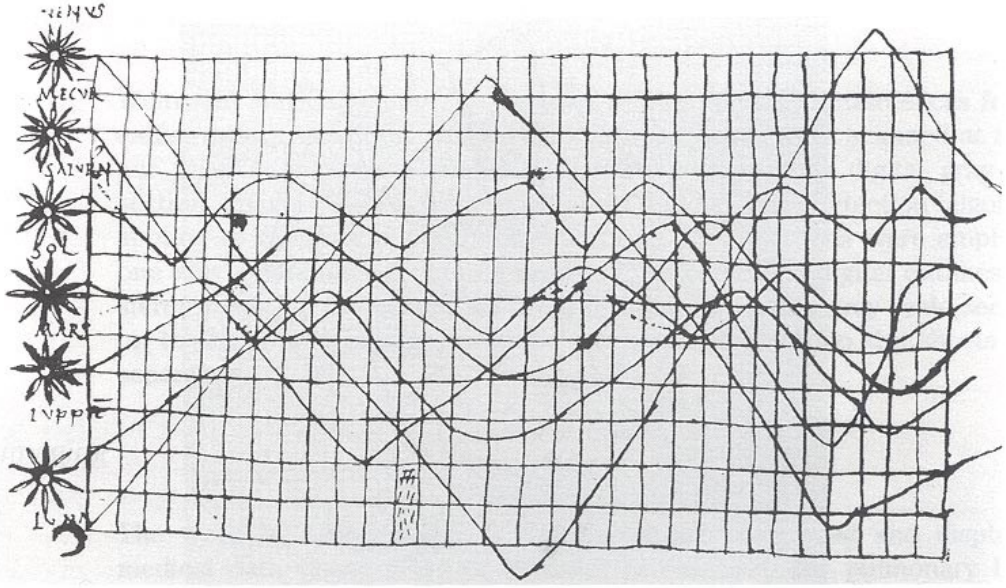
  1986 - Identification of Visualization as an autonomous discipline

  Visualization in Scientific Computing
  (McCormick, de Fanti and Brown – 1987)

Brief history

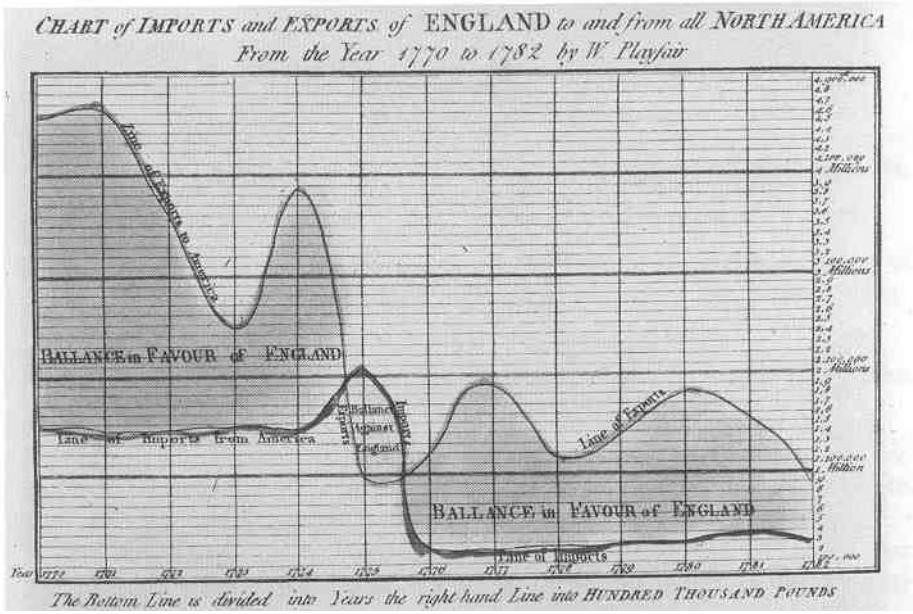
Examples of the pre-computer age

One of the oldest known Visualizations



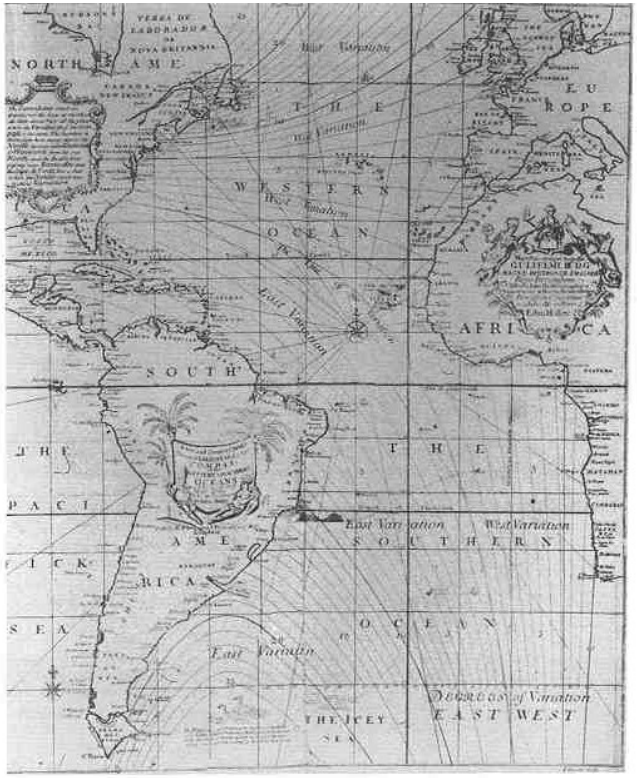
Inclination of orbits along the time - Xth century (Tufte,1983)

One of the first Visualizations used in “business”



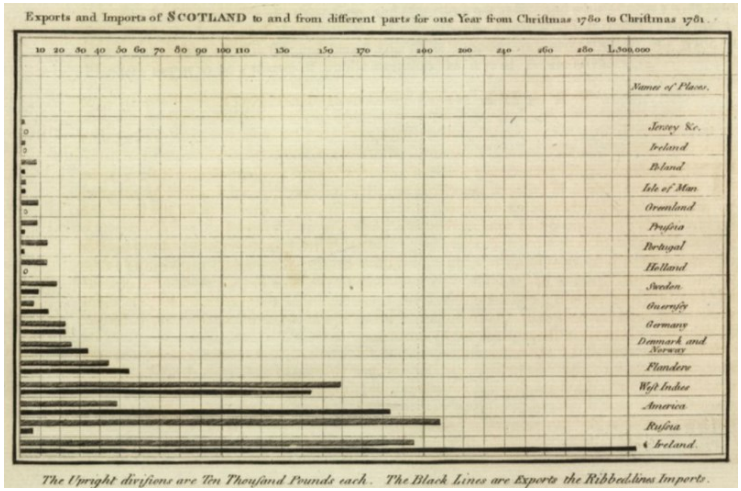
Import/export during the period from1770 to1782 by William Playfair (Tufte, 1983)

One of the first visualizations using contours (isolines)



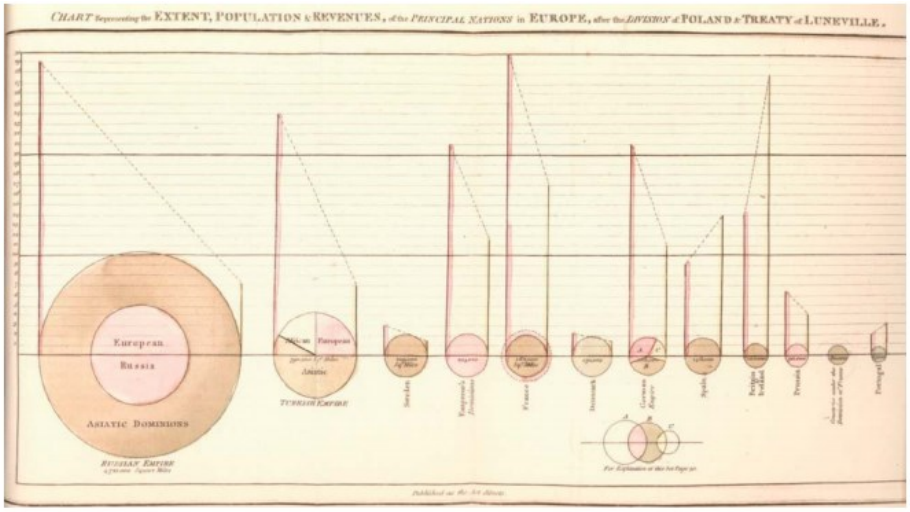
Magnetic declination 1701 Edmund Halley (Tufte,1983)

“Ancestors” of simple representations of univariate data

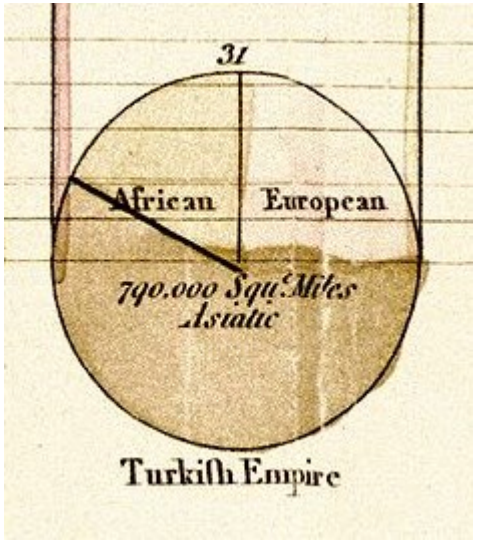


Exports and Imports of Scotland to and from different parts for one
W. Playfair’s *The Commercial and Political Atlas, 1871*

https://en.wikipedia.org/wiki/William_Playfair

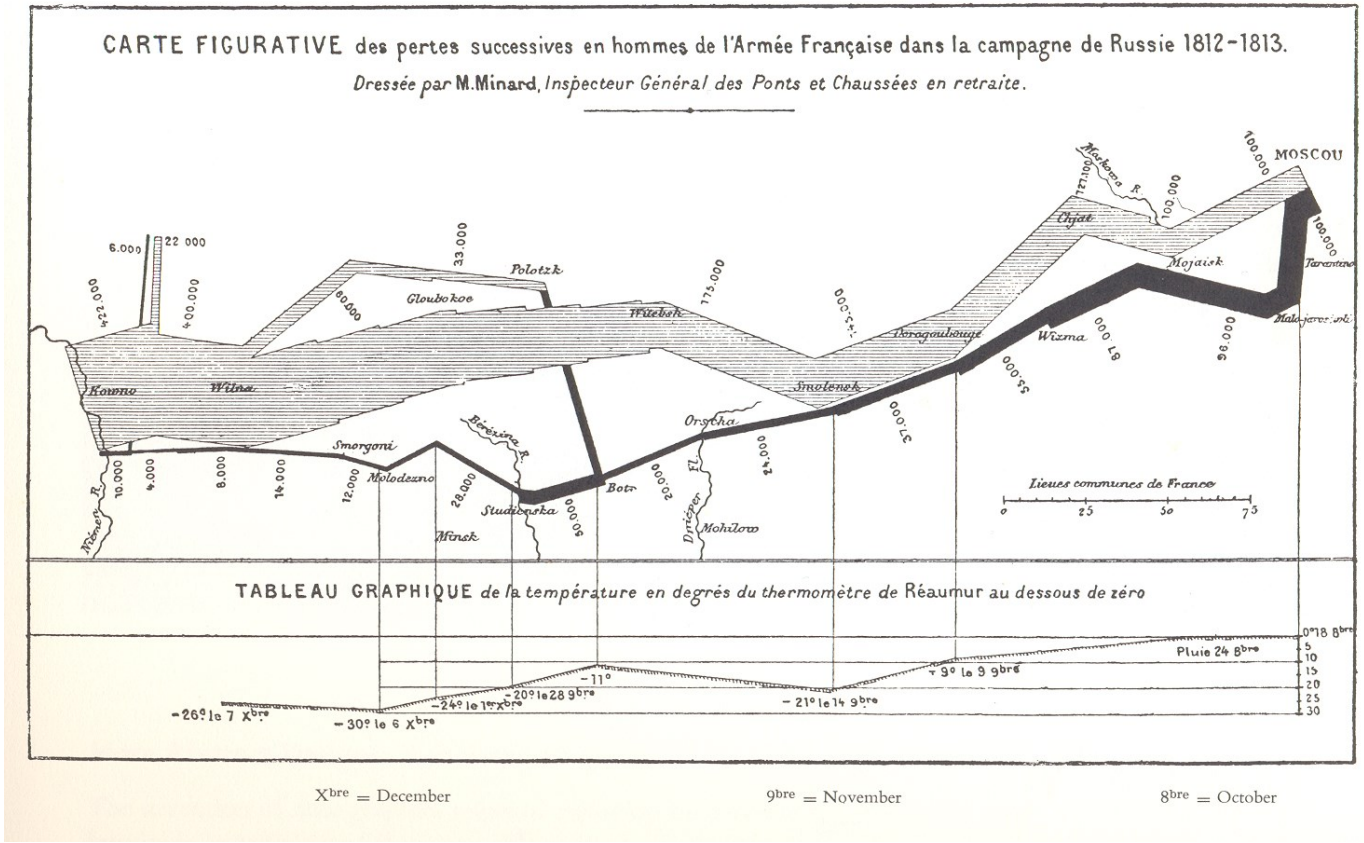


W. Playfair, *Statistical Breviary*, 1801



Multidimensional Visualization

6 dimensions: location (2), n. of men and direction of the army, date, temperature



Russia campaign of Napoleon 1861 by Charles Minard (Tufte, 1983)

Visualization in scientific discovery

Dr. John Snow

Discovering the cause of the London cholera out brake, 1853-54 (Wikipedia)

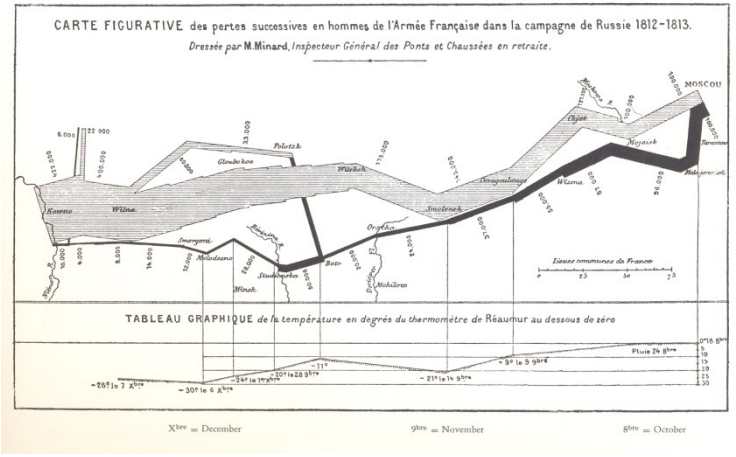
- Visualization may be used with different purposes:

- personal exploration	- explorative analysis
for	
- discussion with colleagues	- confirmative analysis
- presentation to other people	

Classical examples for:
a) exploration
b) presentation

a)

b)



Example of communication to a non-expert audience

Hans Rosling's 200 Countries, 200 Years, 4 Minutes: 120 000 values

Income (x), Age expectancy (y) , Time (t), Continent (colour), Population (size)

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

Example: how to select simple charts?

Max and Min temperatures along the month of February in a specific location (in ºC):

day	Max T	Min. T
1	15	7
2	14	8
3	13	6
4	13	6
5	12	6
6	13	7
7	13	7
8	14	8
9	15	5
10	12	5
11	13	6
12	12	7
13	11	8
14	11	8
15	12	8
16	12	9
17	13	9
18	14	9
19	14	8
20	13	8
21	13	8
22	12	7
23	12	7
24	11	7
25	11	6
26	11	7
27	13	6
28	14	6

Q1- What were the maximum and minimum values of MaxT?

Q2- What was the most frequent MaxT?

Q3- In how many days was that maximum MaxT value attained?

Q4- How were the daily temperature ranges?

Q5 – What was the maximum temperature range?

- What type of chart would you use to answer Q1?

- And the other questions?

Whatever the purpose, 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**

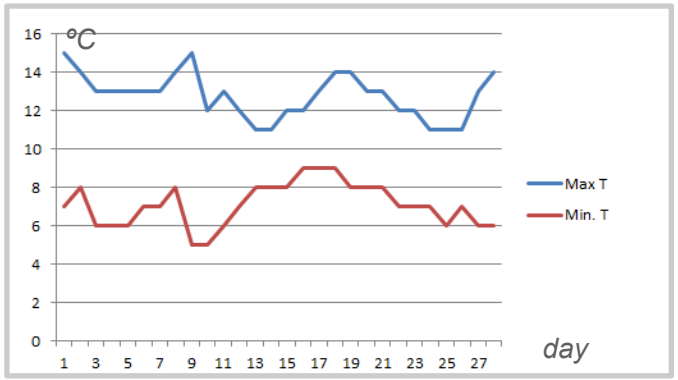
To design simple visualizations or visual data mining applications:

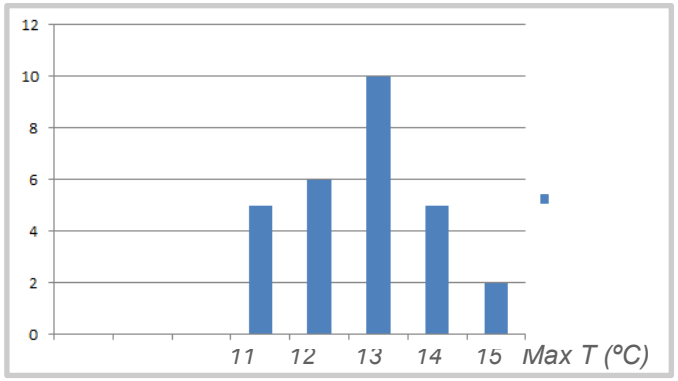
- Need to **find what are the questions** users will ask!

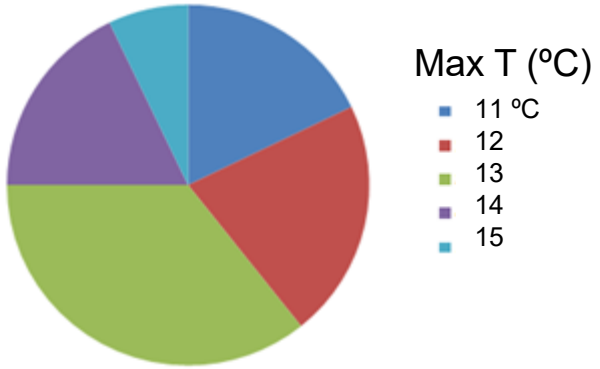
Example: how to select simple charts?

Temperatures along the month of February (in ºC): a few possible charts

day	Max T	Min. T
1	15	7
2	14	8
3	13	6
4	13	6
5	12	6
6	13	7
7	13	7
8	14	8
9	15	5
10	12	5
11	13	6
12	12	7
13	11	8
14	11	8
15	12	8
16	12	9
17	13	9
18	14	9
19	14	8
20	13	8
21	13	8
22	12	7
23	12	7
24	11	7
25	11	6
26	11	7
27	13	6
28	14	6

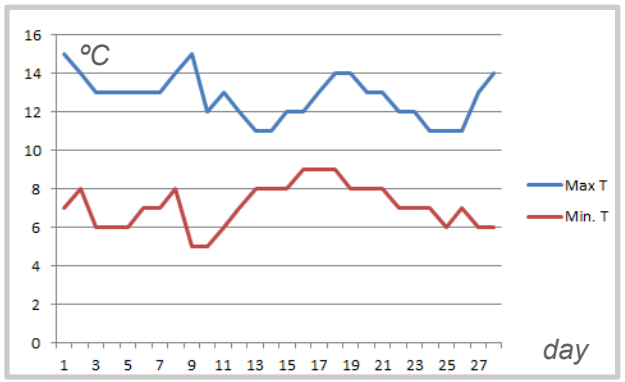






Simple example

Temperatures along the month of February (in ºC):

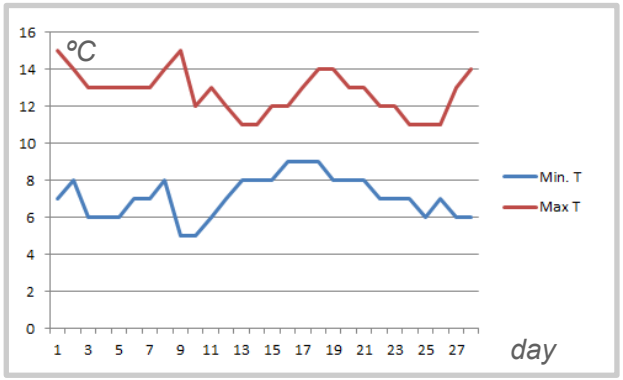


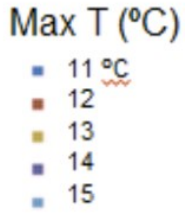
Anything “odd” about this chart?

What if the user is color-blind?

Test it using https://www.color-blindness.com/coblis-color-blindness-simulator/

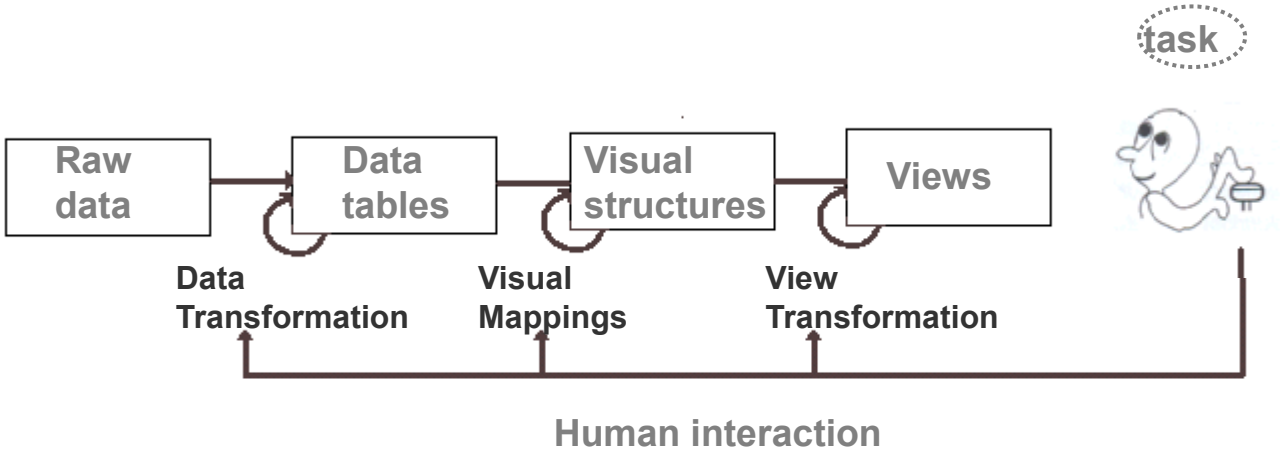
Would you prefer this one?





Do not forget “cultural” aspects,
nor individual differences!

Information Visualization Reference Model



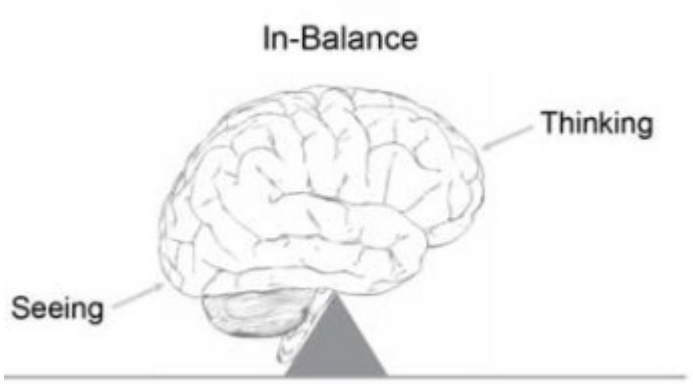
Visualization can be described as the mapping of data to visual form supporting human interaction for visual sense making (Card et al., 1999)

Visualization is a **Human in the loop process**! -> which calls for **specific methods**

When are Visualization solutions appropriate?

Vis systems are appropriate when the goal is to **augment human capabilities**, rather than completely replace the human in the loop

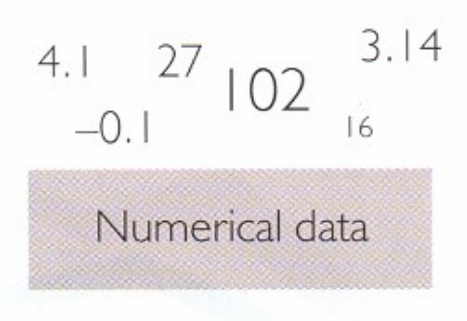
Should allow **offload internal cognition and memory** usage to the perceptual system, using carefully designed images as a form of external representations **(external memory**)



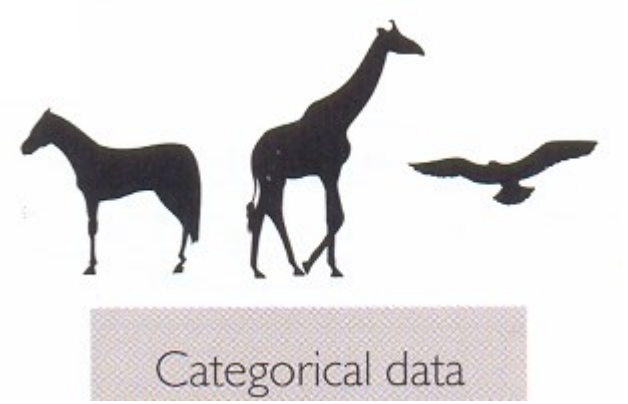
Data Characteristics

(wordcloud of chap.1, Mazza, 2009)

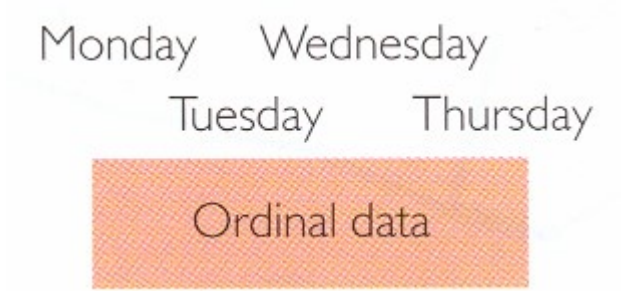
- Data representation level:
  - Qualitative (or categorical)
  - Quantitative (or numerical)



- Data nature:
  - Continuous
  - Discrete



- Measuring scale:
  - Nominal
  - Ordinal
  - Interval — quantitative
  - Ratio /



(Spence, 2007)

- Examples of measuring scales and types of data:
  - **Nominal/categorical -->** car brands, gender, animal species…
  - **ordinal -->** week days, preferences, levels measured in a Likert-type scale
  - **Interval -->** IQ, temperatures in ºC
  - **Ratio -->** weight, height

- The ratio scale represents the **highest level of representation,** has a non-arbitrary zero (unlike the interval scale)

- This is a general classification and may be used to select the methods to use with the data

http://lsc.cornell.edu/wp-content/uploads/2016/01/Intro-to-measurement-and-statistics.pdf
https://www.youtube.com/watch?v=KIBZUk39ncI

Example: beyond the structure of the data to Visualize

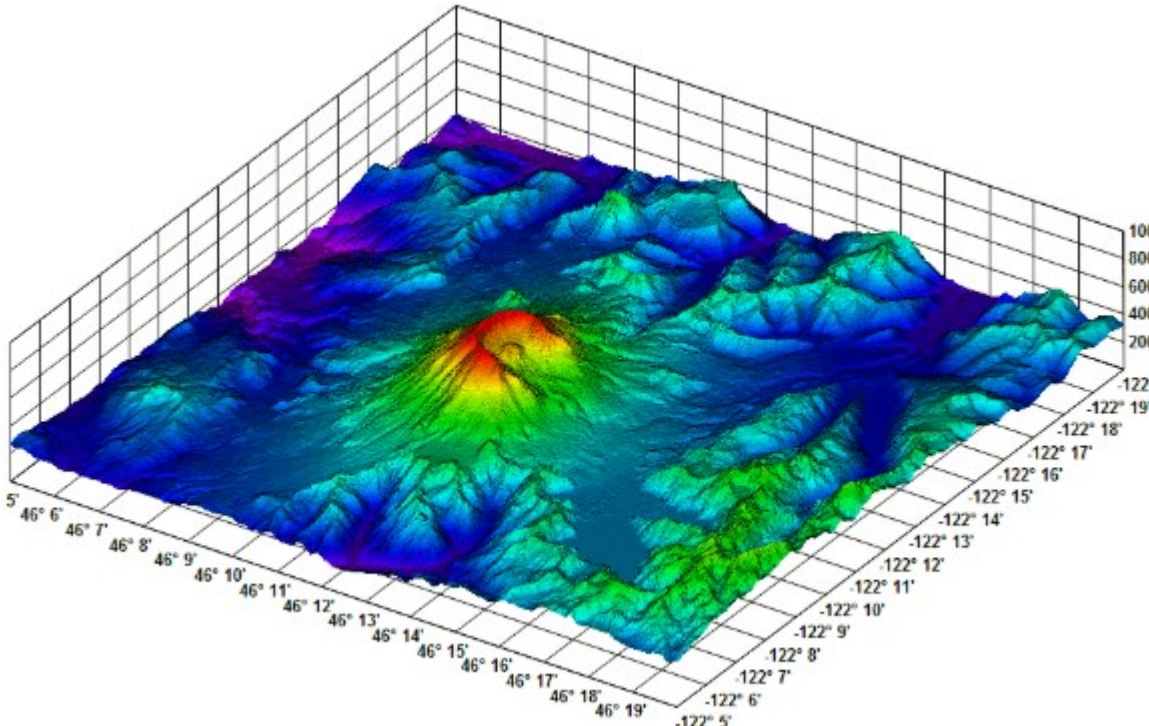
- Consider a data set with three columns:

  latitude *longitude* *d*

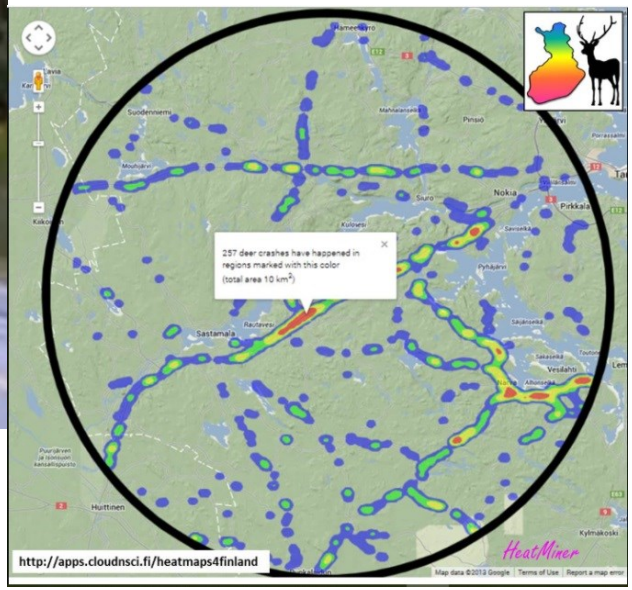
- Which is the most adequate way to visualize these data?

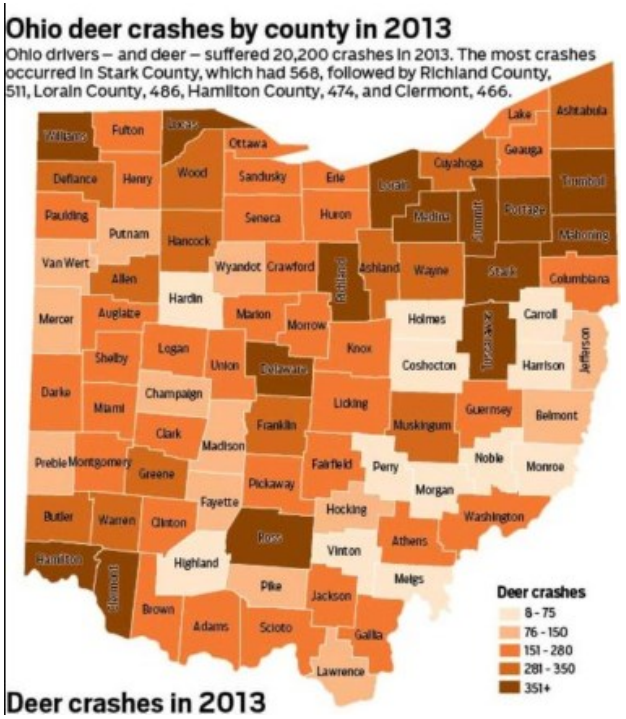
- If *d* is depth or altitude?

the selected visualization technique may involve interpolation (e.g. isocontours, isosurfaces)



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





http://cloudnsci.fi/wiki/index.php?n=Applications.Heatmaps4Finland

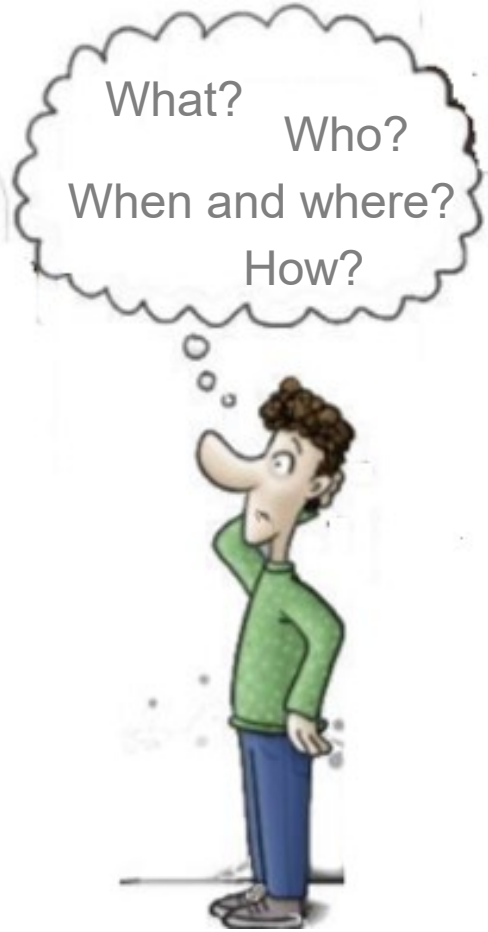
- Interpolation and contours don’t make sense!

Know the data structure is not enough

It is necessary to **know the phenomenon behind the data**

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

- Visualization may be the solution (or part of it)
- But:
  - **How to create a visualization?**



Acknowledgement
The author of these slides is grateful to all students and colleagues that have contributed in any way to these slides

Main bibliography for this topic

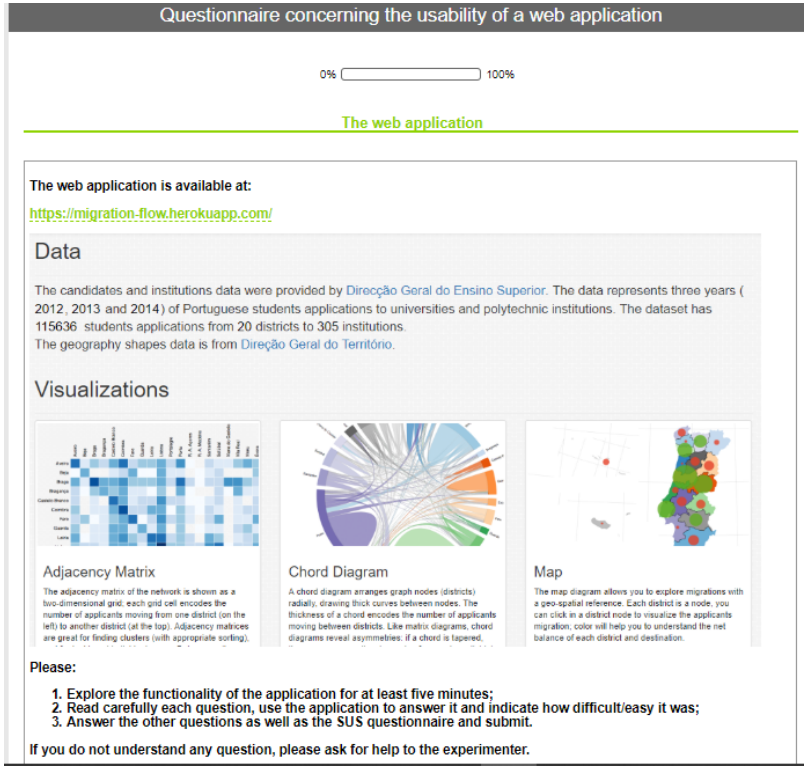
- Mazza, R., *Introduction to Information Visualization*, Springer, 2009
- Munzner, T., Visualization Analysis and Design, A K Peters/CRC Press, 2014
- Kirk, A., *Data Visualisation A Handbook for Data Driven Design*, 2nd. Ed., Sage, 2019
- Spence, R., *Information Visualization, An Introduction*, 3rd ed., Springer, 2014
- Spence, R., *Information Visualization, Design for Interaction*, 2nd ed., Prentice Hall, 2007
- Ware, C., *Information Visualization, Perception to Design*, 2nd ed., Morgan Kaufmann,2004
- http://www.wikiviz.org/wiki/Main_Page

Exercise (example of remote usability evaluation):

Explore the application:

https://migration-flow.herokuapp.com/

perform the tasks and answer the questions



https://forms.ua.pt/index.php?r=survey/index&sid=489227&lang=en