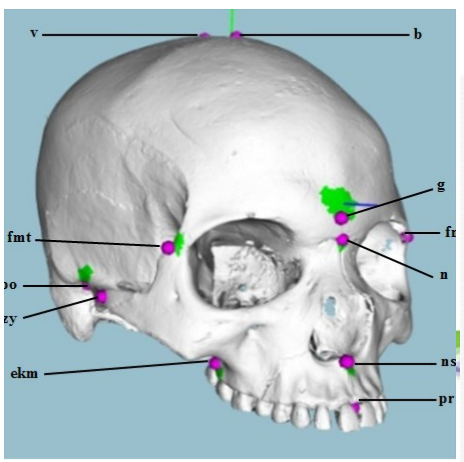
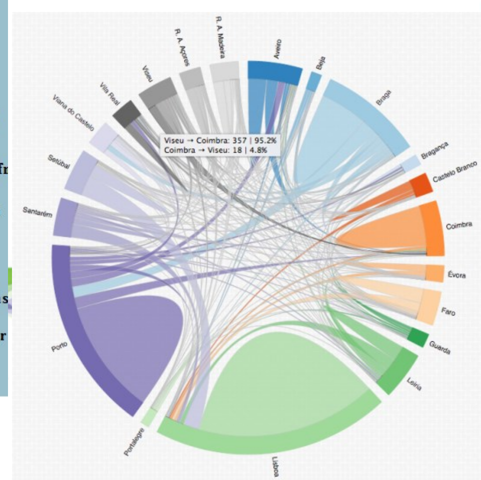
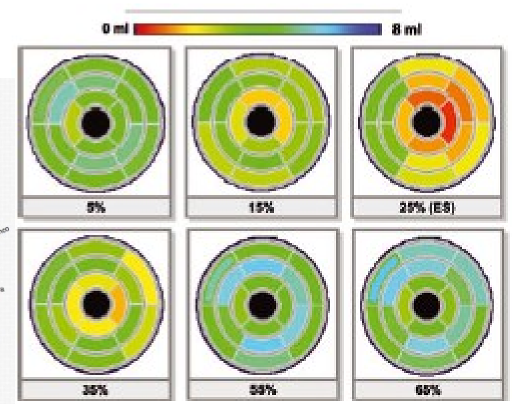
Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

An Introduction to Data and Information Visualization



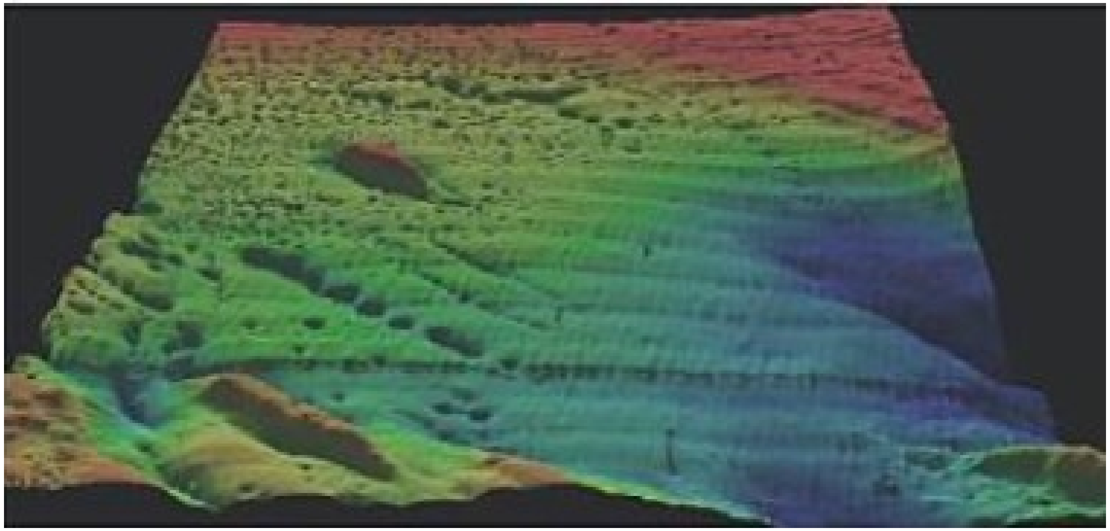




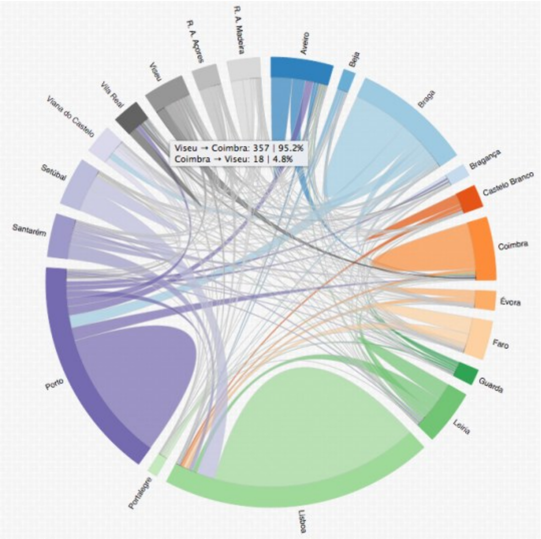
Beatriz Sousa Santos, University of Aveiro, 2021

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

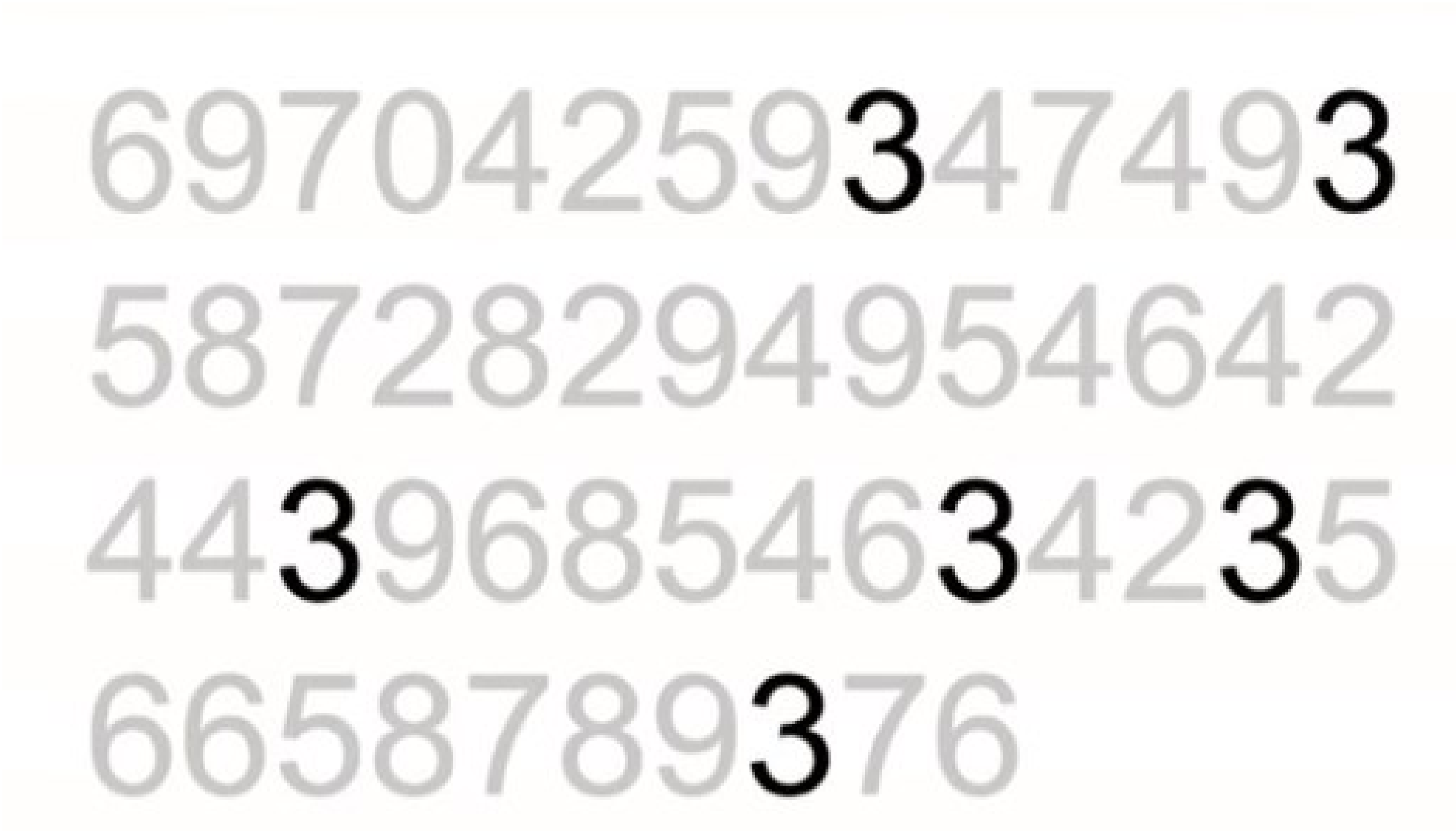
  but not in others …

(Ware 2019)

An exercise in preattentive processing: how many "3"?

69704259347493
58728294954642
44396854634235
6658789376

(Nussbaumer Knaflic, 2015)



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

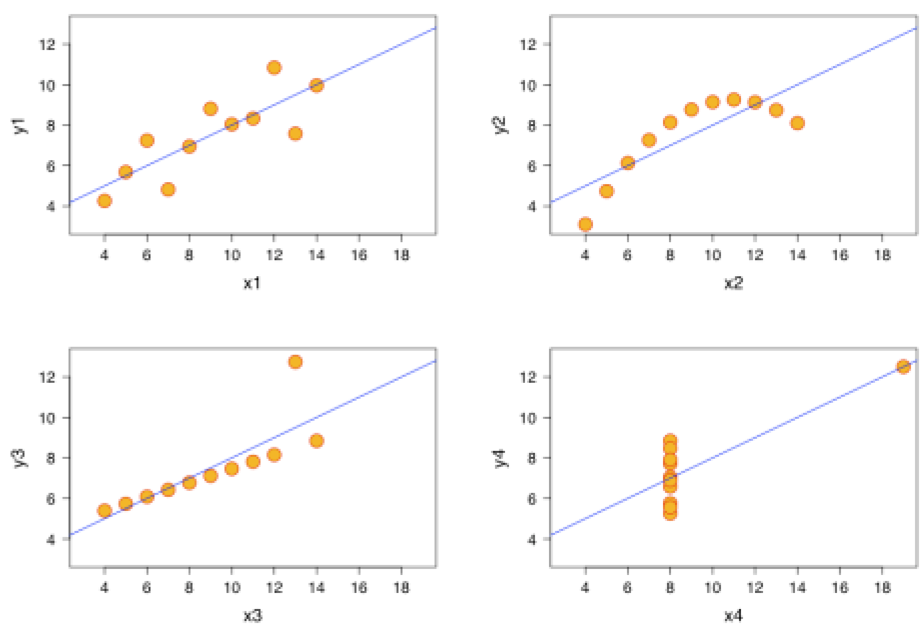
Why show the data in detail?
and not only aggregated parameters

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)

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

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	



Outline of the lectures:

Data and Information Visualization: Introduction
- Data Characteristics

Information Visualization:
- Main issues
- Representations
- Presentation
- Interaction
- Evaluation

Computer graphics:
- Interactive graphics systems; Primitives and attributes
- Geometric transformations and projections
- Visualization pipeline, visibility, illumination and surface rendering, geometric transformations and projections

Definition

Objectives

History

Applications

Model

How to obtain and evaluate a Visualization?

Definition

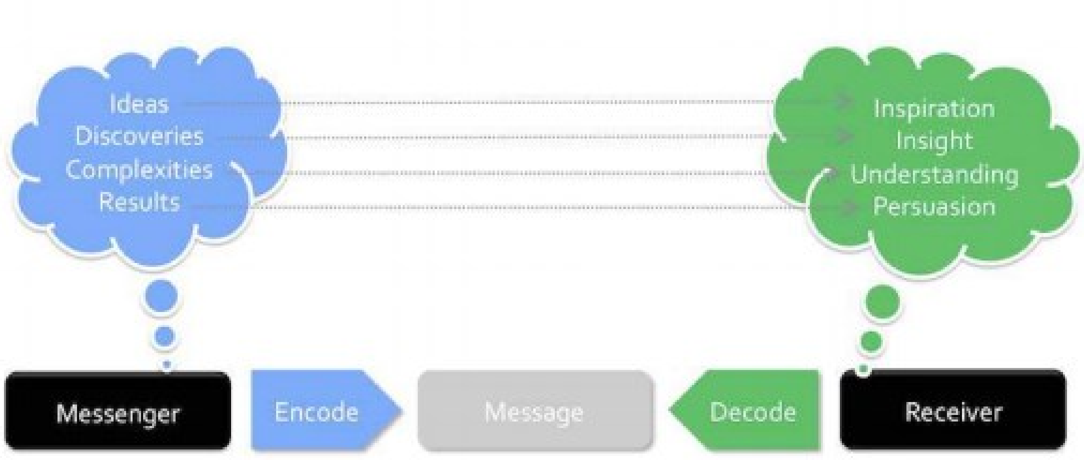
- Visualization is the process of exploring, transforming and representing data as images (or other sensorial forms) to gain insight into phenomena

- There are several expressions used to designate different areas of Visualization:
  - Scientific Visualization
  - Data Visualization
  - Information Visualization

- The differences among these areas are not completely clear
- Industry uses Data Visualization …

Definition (yet another)

- "The representation and presentation of data that exploits our visual perception abilities in order to amplify cognition" *(Kirk, 2012)*



(Kirk, 2012)

- *Intelligence Amplification* as opposed to *Artificial Intelligence* (Fred Brooks, 1999)

- Visualization may have a significant role in the amplification of human capacities

- Several scientific disciplines contribute to Visualization:
  - Computer Graphics
  - Human-Computer Interaction
  - Software Engineering
  - Image Processing
  - Signal Processing

- Visualization is different from Computer Graphics and Image Processing since:

    1- it deals mainly with multi-dimensional data

    (>= 3D, time varying)

    2- data transformation is fundamental

    (may be changed to increase insight)

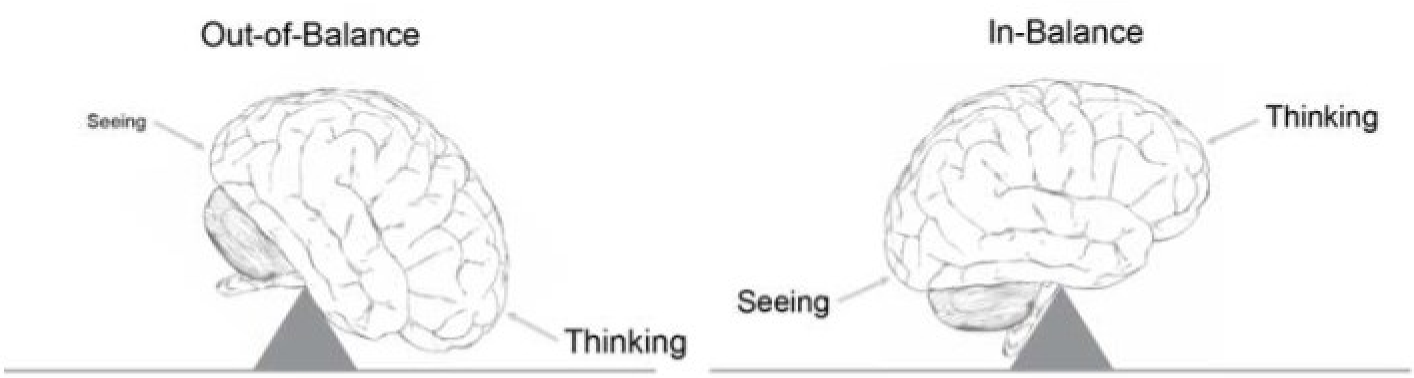
    3- it is essentially interactive,

    (including the user in the process of data creation, transformation and visualization)

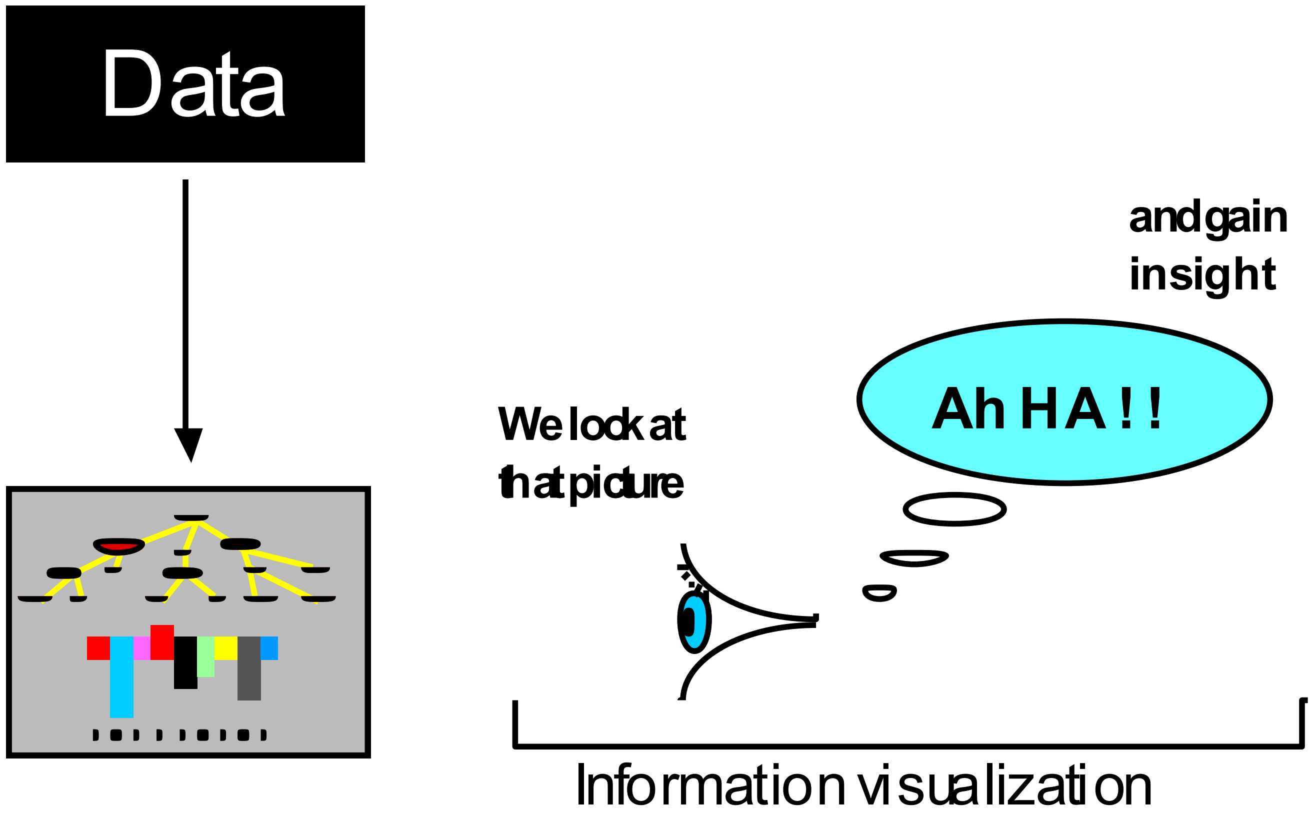
- However, there is some overlap:
  - The output of a visualization process is images
  - In general uses much CG

Visualization benefits

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

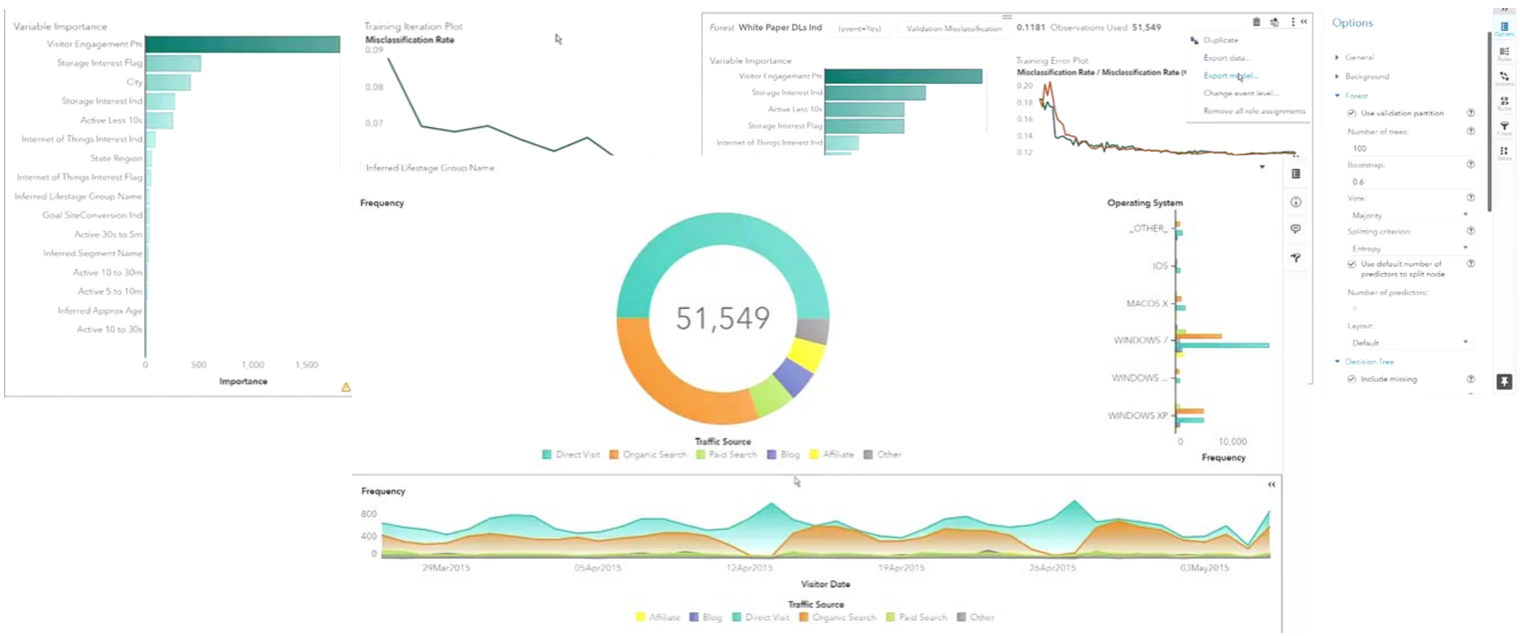


https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception



The process of information visualization: graphically encoded data is viewed in order to form a mental model of that data (Spence, 2007)

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)

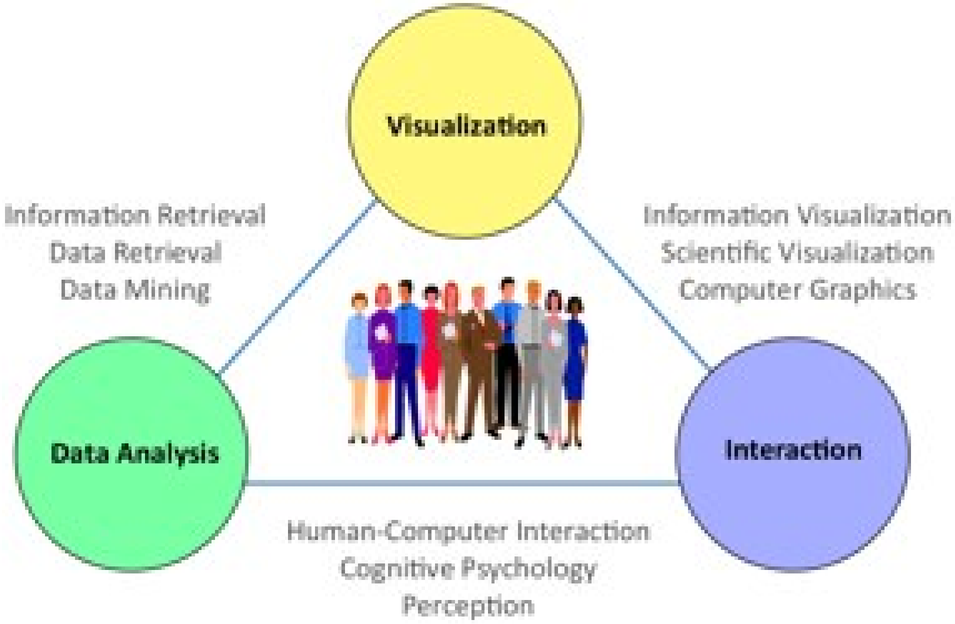
- Specially when **little is known about the data** and the exploration **goals are vague**

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

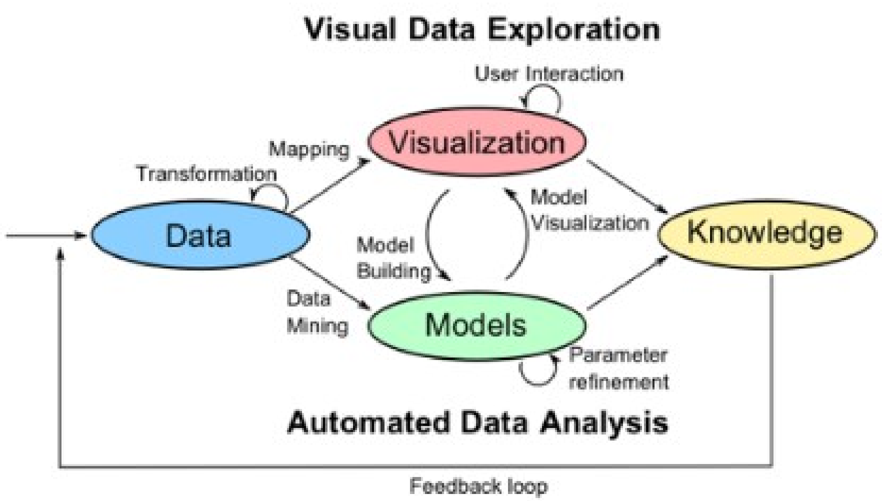
- **Main advantages** of visual over automatic data mining techniques (statistics or machine learning):
  - can easily deal with highly inhomogeneous and noisy data
  - is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters.
- Visual data exploration 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

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

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.

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 …

Brief history

- The usefulness of graphical representations of large amounts of data has been recognized long ago:

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

  W. Playfair, C. J. Minard

  XX century- J. Bertin, E. Tufte

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

  1987 - Identification of Visualization as an autonomous discipline

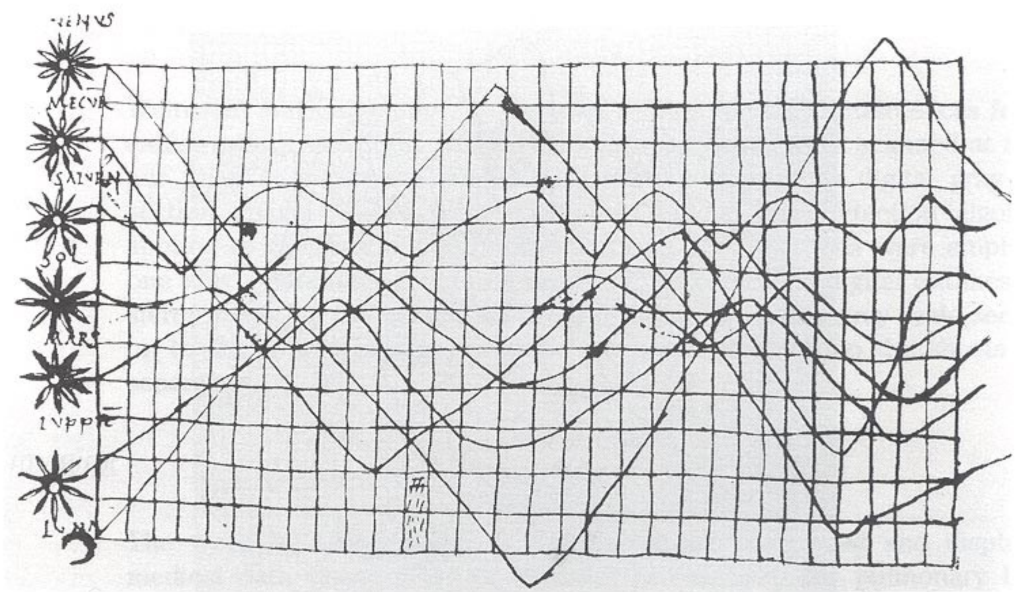
  Visualization in Scientific Computing
  (McCormick, de Fanti and Brown – 1987)

Brief history

- Plenty of Visualization examples of the “pre-computer age”:
    - Inclination of planetary orbits – Xth century
    - Import/ export (Playfair) – XVIIIth century
    - Magnetic declination (Halley) – XVIIIth century
    - Russia campaign of Napoleon (Minard) –XIXth century
    - Cholera out-brake in London (Dr. Snow) – XIXth century

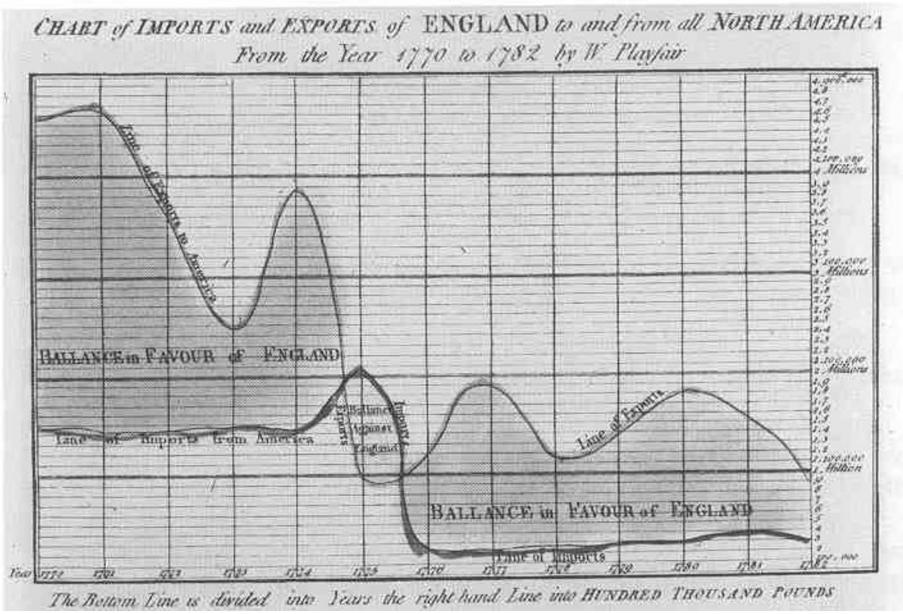
“Pre-computer” Visualization:

One of the oldest known Visualizations



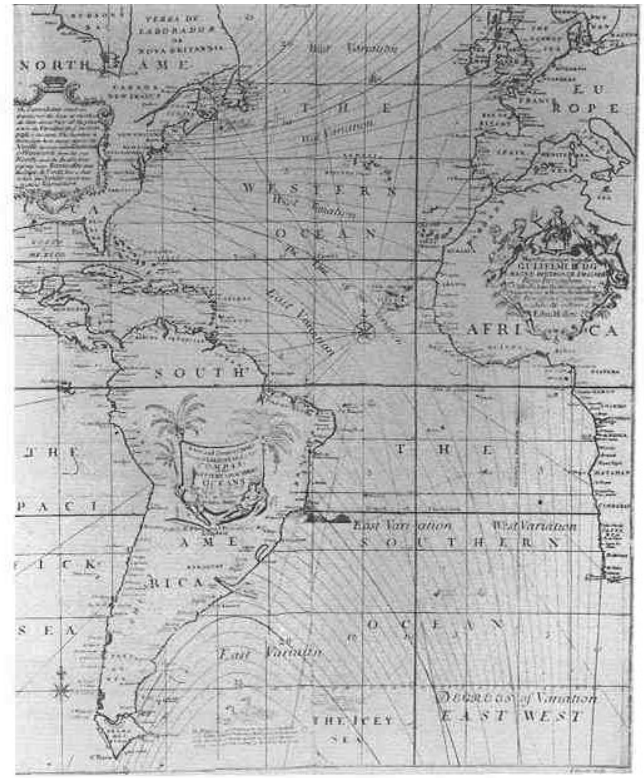
Inclination of orbits along the time - Xth century (Tufte,1983)

One of the first Visualizations used in “business”



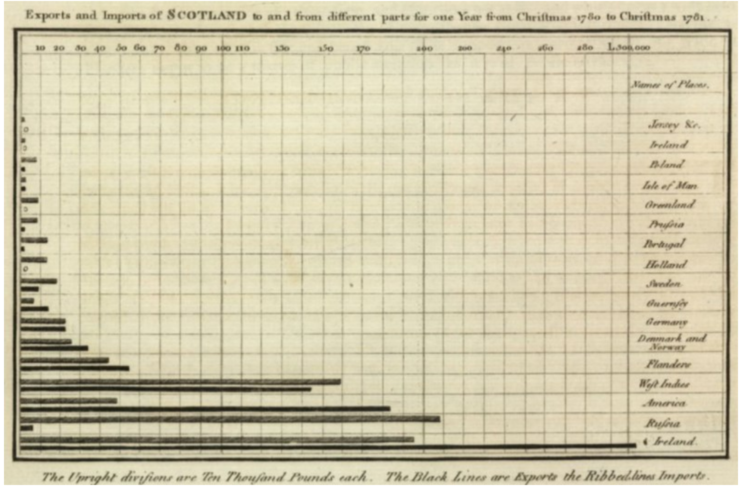
Import/export during the period from1770 to1782 by William Playfair (Tufte, 1983)

One of the first visualizations using contours (isolines)



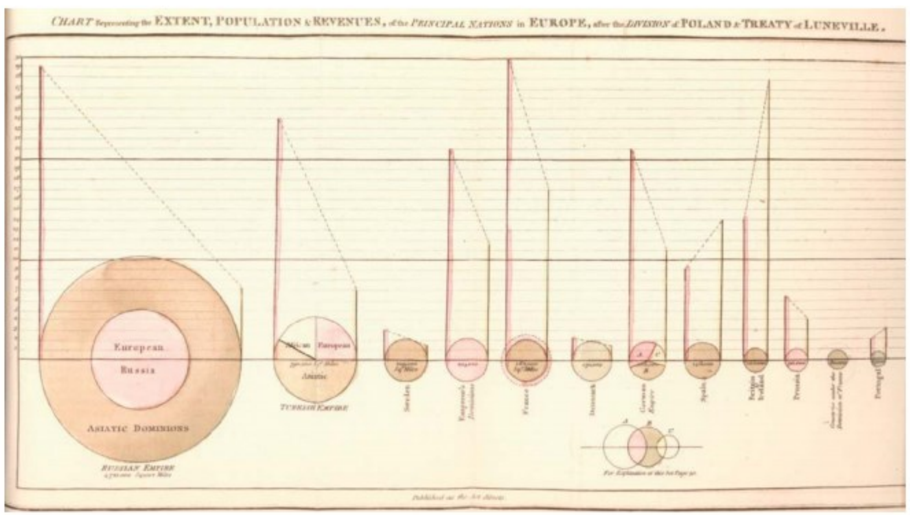
Magnetic declination 1701 Edmund Halley (Tufte,1983)

“Ancestors” of simple representations of univariate data

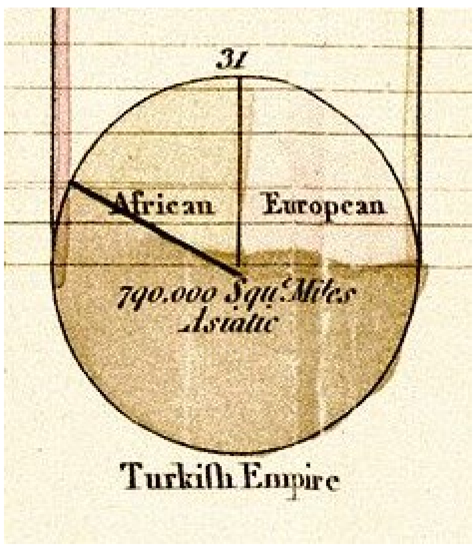


Exports and Imports of Scotland to and from different parts for one
W. Playfair’s *The Commercial and Political Atlas, 1871*

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

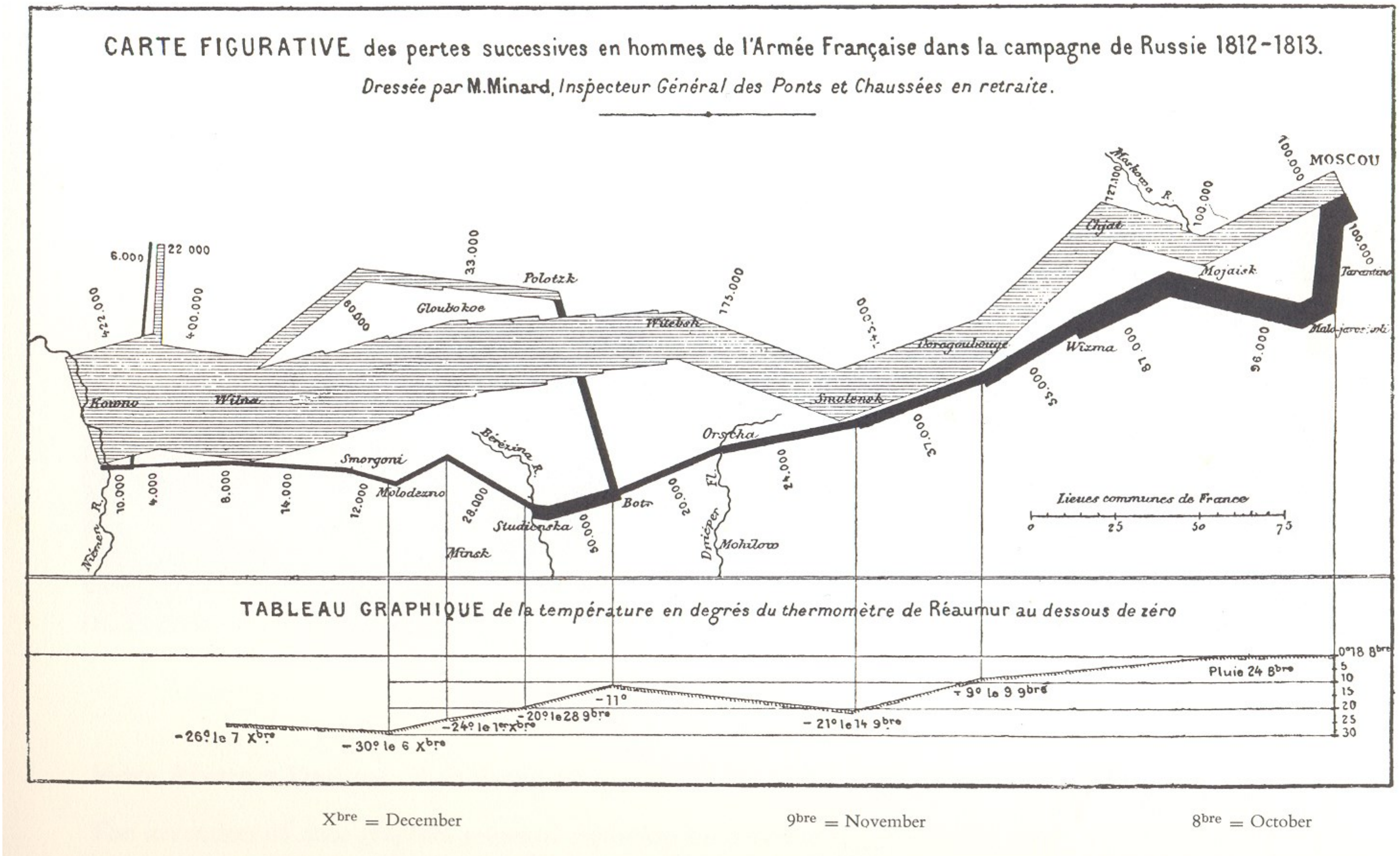


W. Playfair, *Statistical Breviary*, 1801



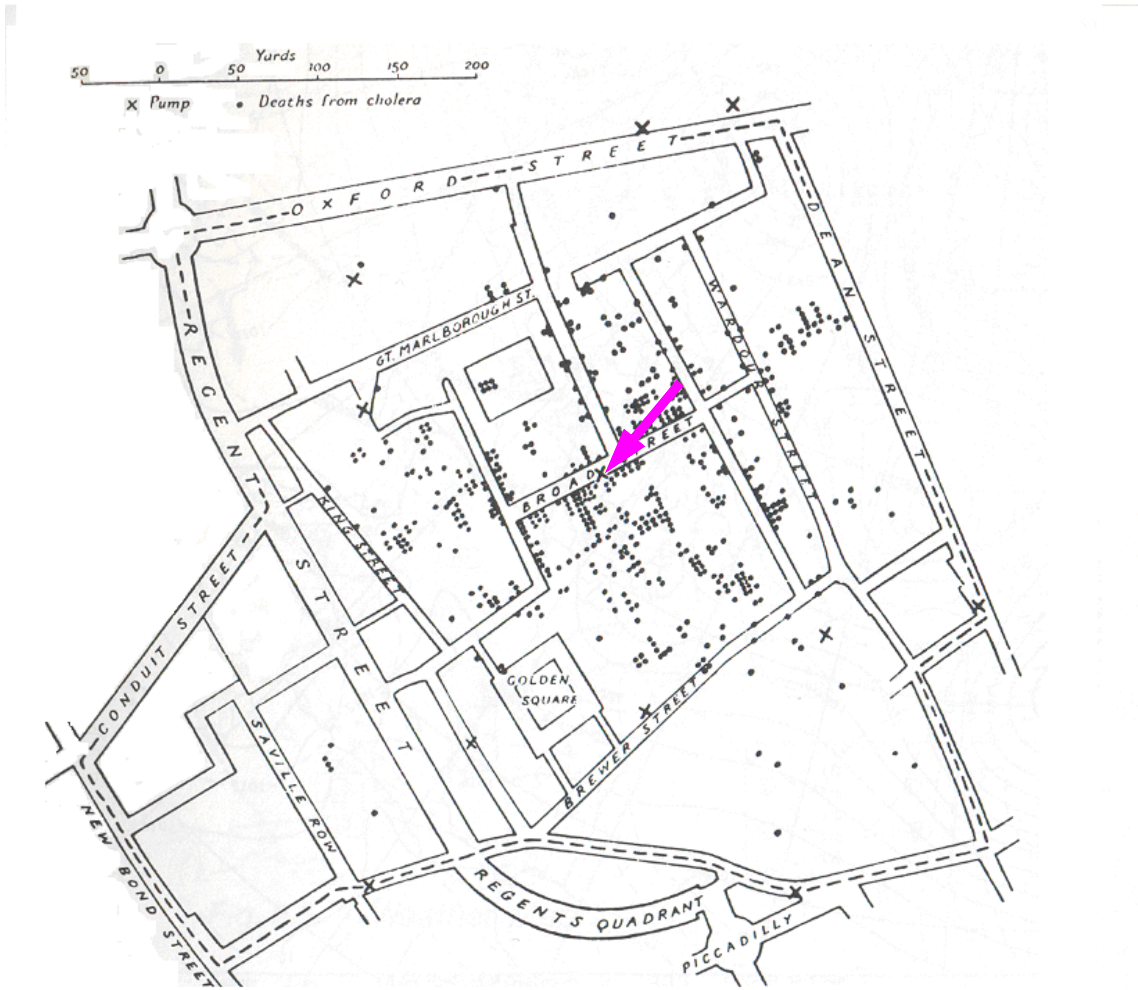
Multidimensional Visualization

6 dimensions: place (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)

Data and Information Visualization

- In general:

  Data (scientific) Visualization (DV) - Data having an inherent spatial structure

  (e.g., CAT, MR, geophysical, meteorological, fluid dynamics data)

  Information Visualization (IV) – “Abstract” data not having an inherent spatial structure (tabular data)

  (e.g., stock exchange , S/W, Web usage patterns, text)

- These designations may be misleading; both DV and IV start with (raw) data and allow to extract information

- Borders between these areas are not well defined …

IEEE Visualization: evolution of names in a major conference ….

- Vis -> SciVis ~2015

T. Dang , H.N. Nguyen, and V. Pham, WordStream: Interactive Visualization for Topic Evolution EUROVIS 2019

Applications of Scientific Vsualization

- **Scientific Visualization** is currently used in many scientific areas:
  - Medicine
  - Meteorology, climatology, oceanography
  - Fluid dynamics
  - Cosmology
  - etc., etc.

- Let us see some examples …
- Can you think of an area where data visualization cannot be applied?

Medicine
(education)

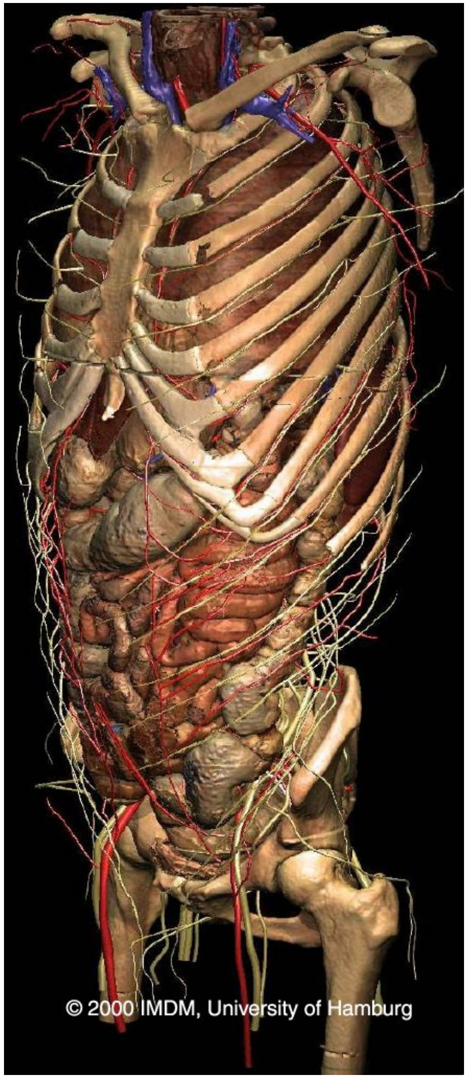
- Human anatomy
- using volume rendering
- VOXELman (University of Hamburg)
- Visible Human project (National Library of Medicine-USA)

https://www.visiblebody.com/

http://www.voxel-man.de/3d-navigator/inner_organs/

http://www.nlm.nih.gov/research/visible/visible_human.html

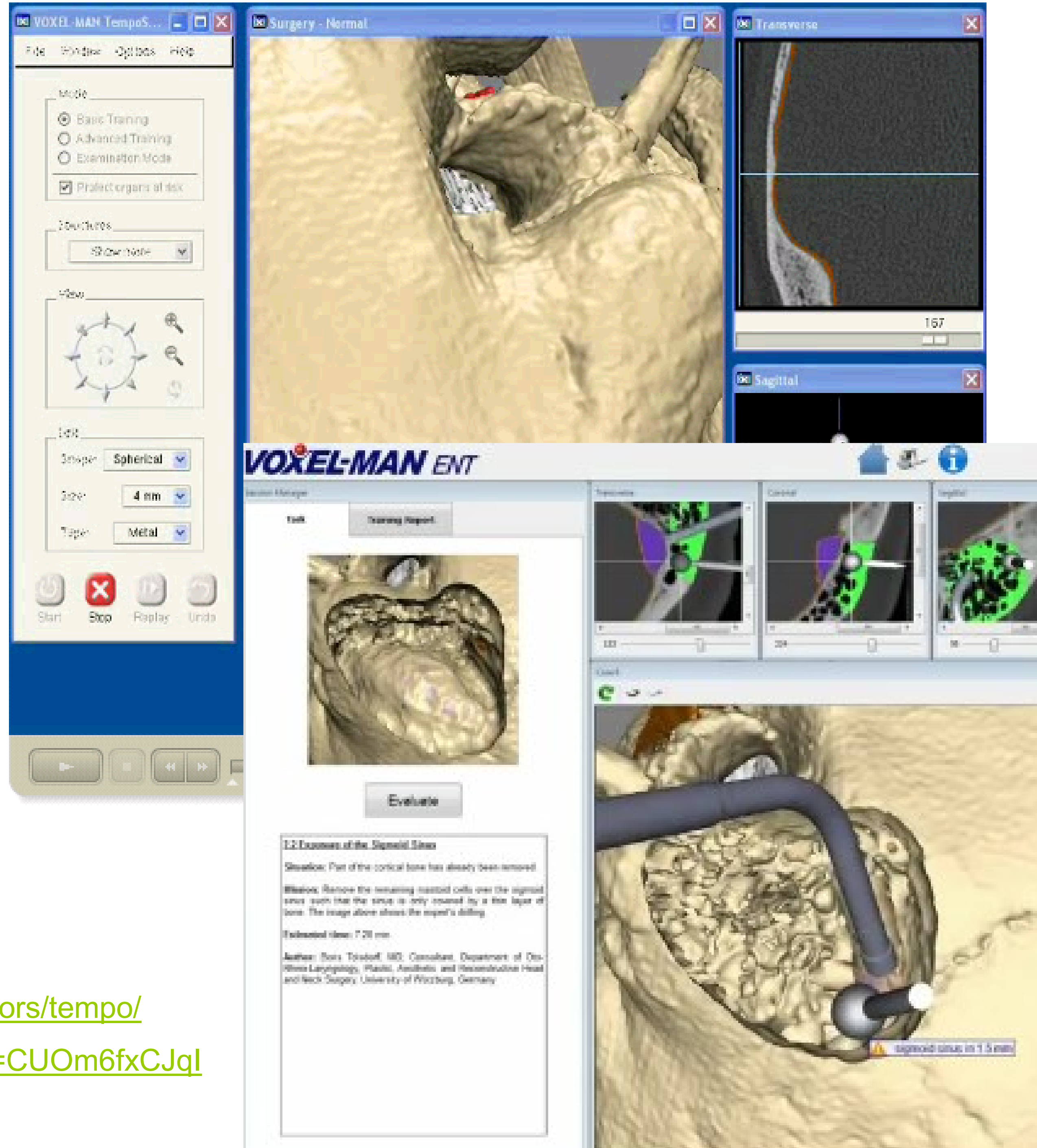
https://www.nlm.nih.gov/research/visible/applications.html



Medicine
(e.g. surgery training)

VOXELman,
University of Hamburg

- Temporal bone surgery
- Movement of the drill is controlled with a force feedback device



https://www.voxel-man.com/simulators/tempo/

https://www.youtube.com/watch?v=CUOm6fxCJqI

Dentistry (e.g. training)

Stereoscopic display + glasses

Interaction devices:

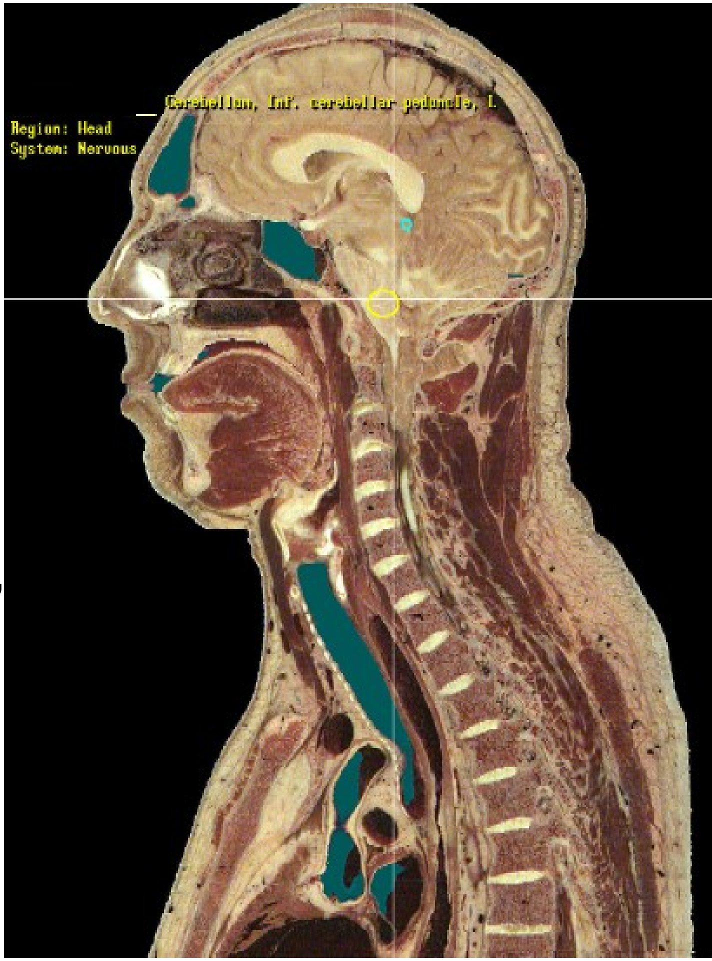
- two phanton (force feedback)
- foot pedal

https://www.voxel-man.com/simulators/dental/

https://www.youtube.com/watch?v=CB_vdW6K42o

An example of Scientific Visualization: The visible Human Project

(1994,1995)



The data sets were designed to serve as
(1) a reference for the study of human anatomy,
(2) public-domain data for testing medical imaging algorithms,
(3) a test bed and model for the construction of network-accessible image libraries.

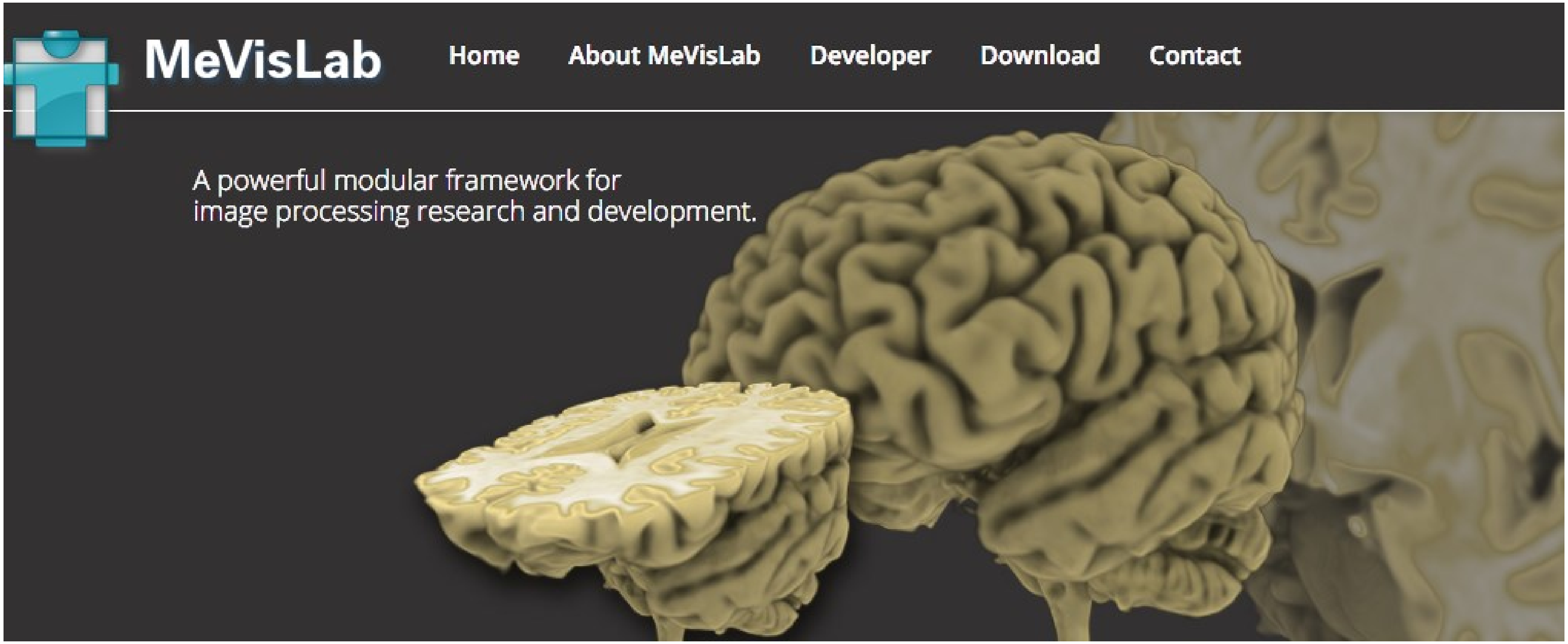
Have been applied to a wide range of educational, diagnostic, treatment planning, virtual reality, artistic, mathematical, and industrial uses.

About 4,000 licensees from 66 countries

As of 2019, a license is no longer required to access the VHP datasets.

https://www.nlm.nih.gov/research/visible/applications.html

Medical Imaging



https://www.mevislab.de/

http://www.nlm.nih.gov/research/visible/visible_human.html

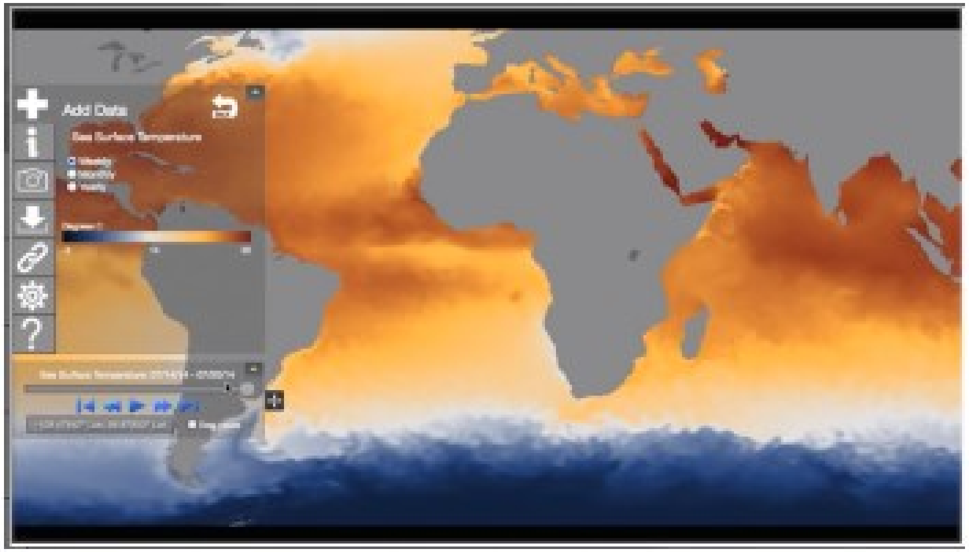
https://www.nlm.nih.gov/research/visible/applications.html

Example in Climate research (by NOAA)

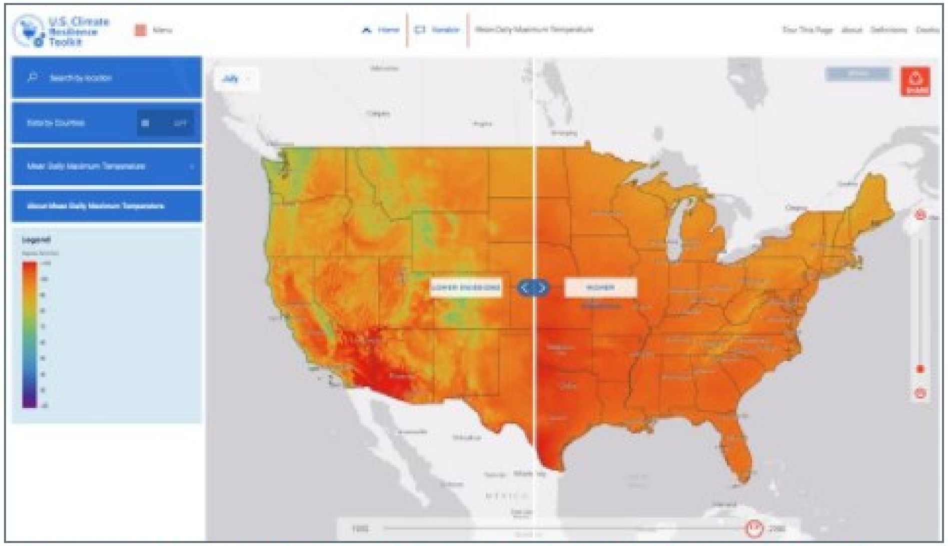
•The Climate Explorer offers graphs, maps, and data of observed and projected temperature, precipitation, and related climate variables for every county in the contiguous US

• The tool shows projected conditions for two possible futures:

– one in which humans make a moderate attempt to reduce global emissions of heat-trapping gases,
– one in which we go on conducting business as usual.



https://www.climate.gov/maps-data/primer/visualizing-climate-data

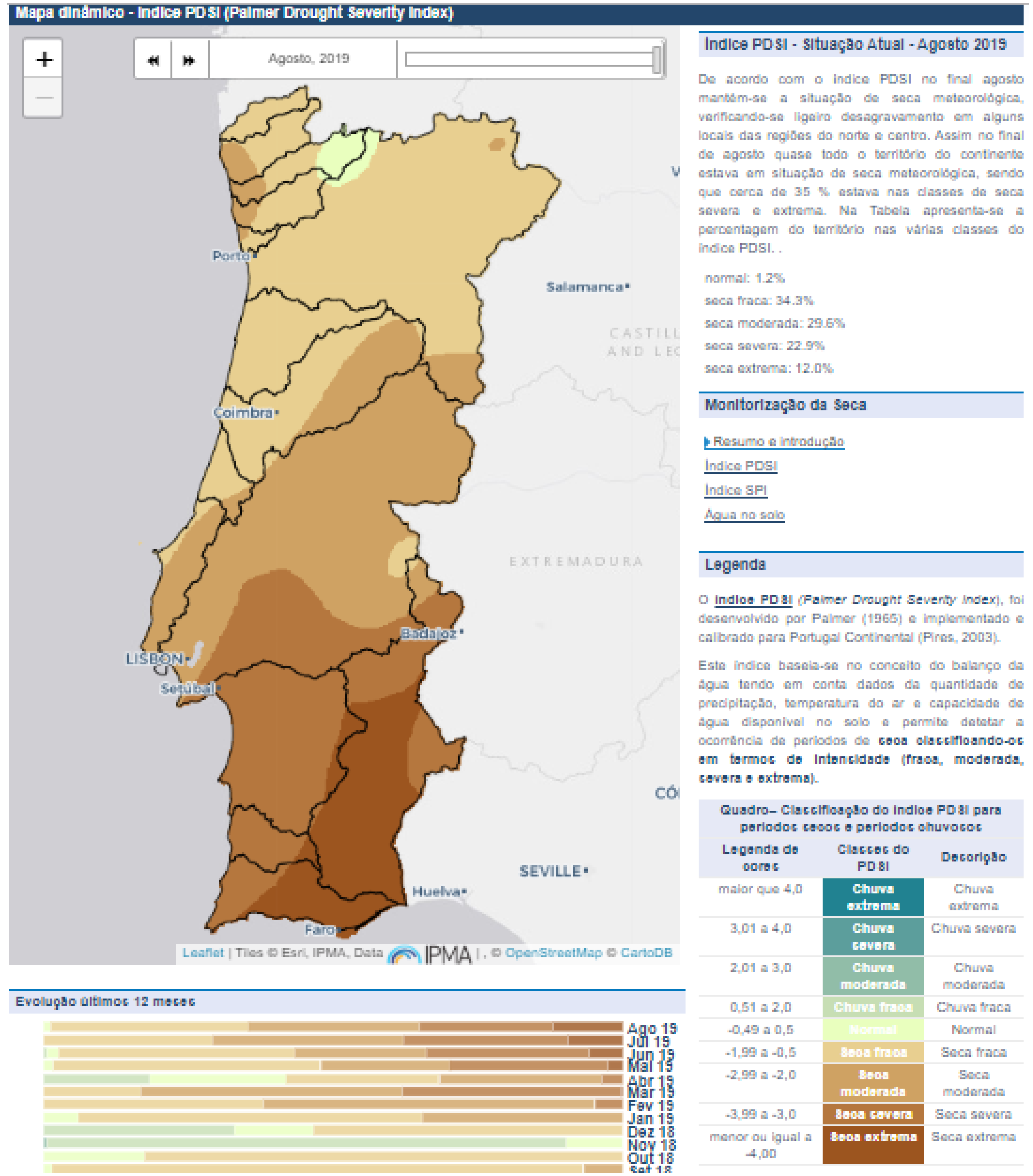


View by Variable interface. View Maximum Daily Temperature variable in Climate Explorer.

https://toolkit.climate.gov/tools/climate-explorer

Example in Climate monitoring:

Drought Severity Index (by IPMA)

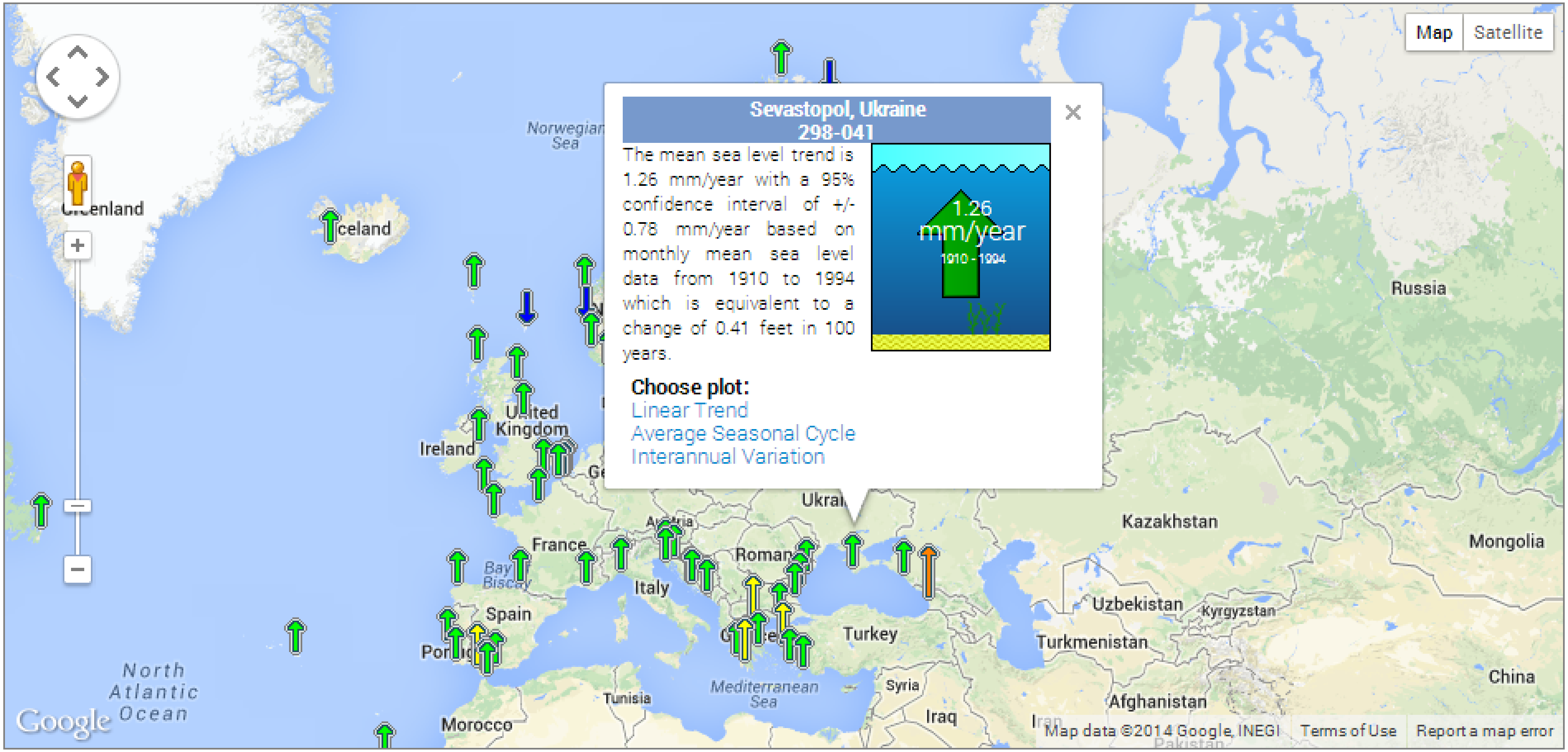


http://www.ipma.pt/pt/oclima/observatorio.secas/

Sea Level Trends

East Coast | West Coast | Gulf Coast | Alaska | Hawaii | Global

View in Google Earth



The map above illustrates regional trends in sea level, with arrows representing the direction and magnitude of change. Click on an arrow to access additional information about that station.

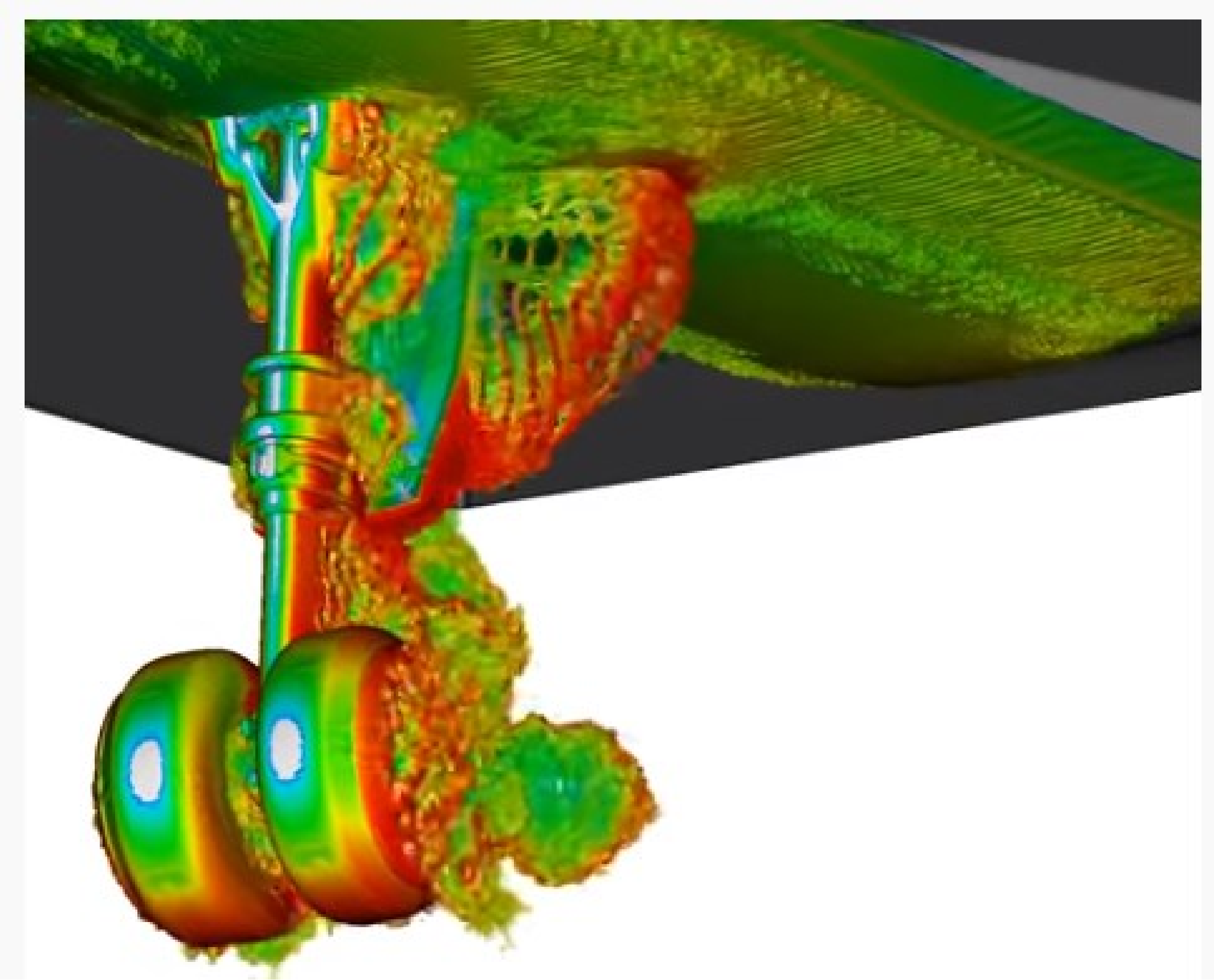
Sea Level Trends
mm/yr (feet/century)

15 to 21 (5 to 7)	6 to 9 (2 to 3)	-3 to 0 (-1 to 0)	-12 to -9 (-4 to -3)
12 to 15 (4 to 5)	3 to 6 (1 to 2)	-6 to -3 (-2 to -1)	-15 to -12 (-5 to -4)
9 to 12 (3 to 4)	0 to 3 (0 to 1)	-9 to -6 (-3 to -2)	-18 to -15 (-6 to -5)

https://tidesandcurrents.noaa.gov/sltrends/regionalcomparison.html?region=GNEATL

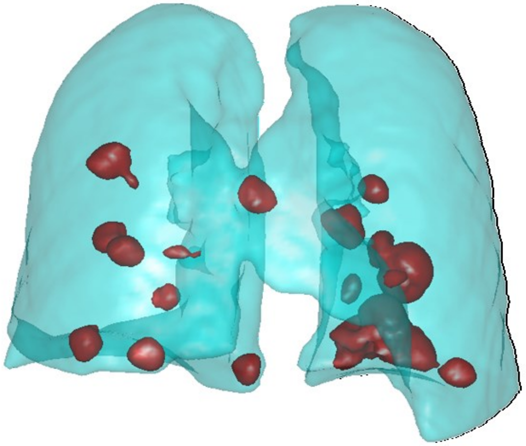
Example of fluid mechanics visualization

NASA/Boeing CFD visualization of vortices responsible for the noise created by the 777's noise landing

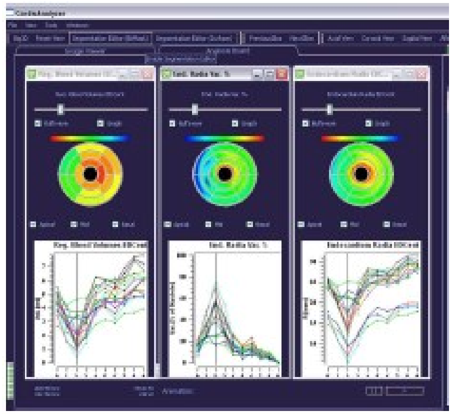


https://www.youtube.com/watch?v=F9EFx7aQuhw

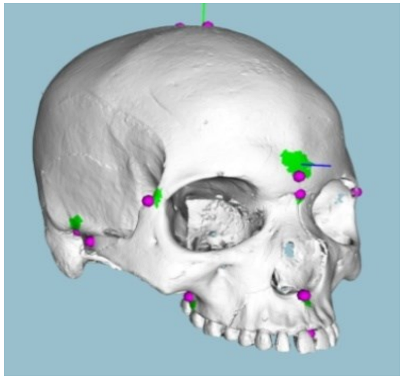
Scientific Visualization **(examples "made in UA")**



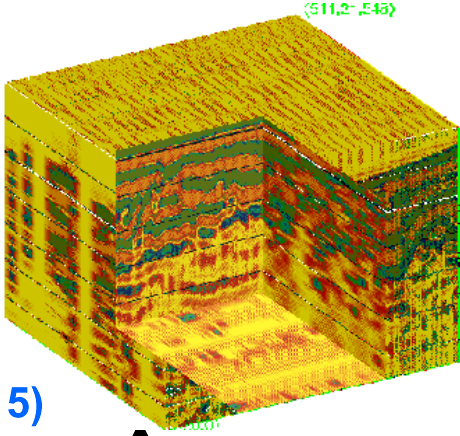
Tomography (2004)



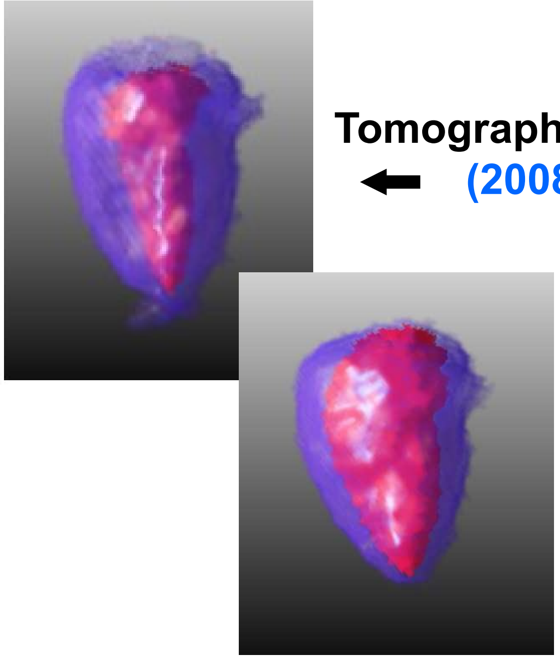
Tomography (2011)



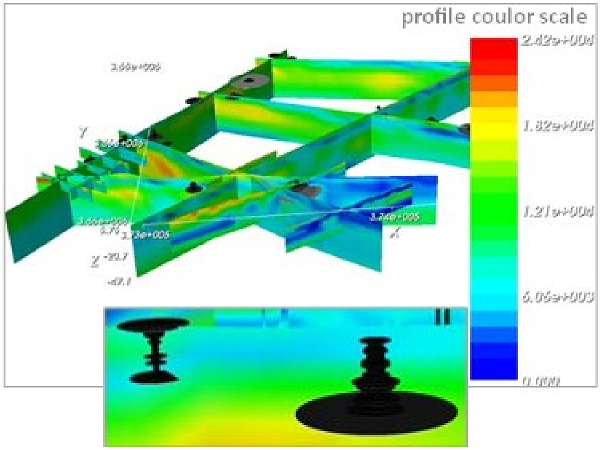
Laser scanner (2015)



Ground Penetrating Radar (1999)

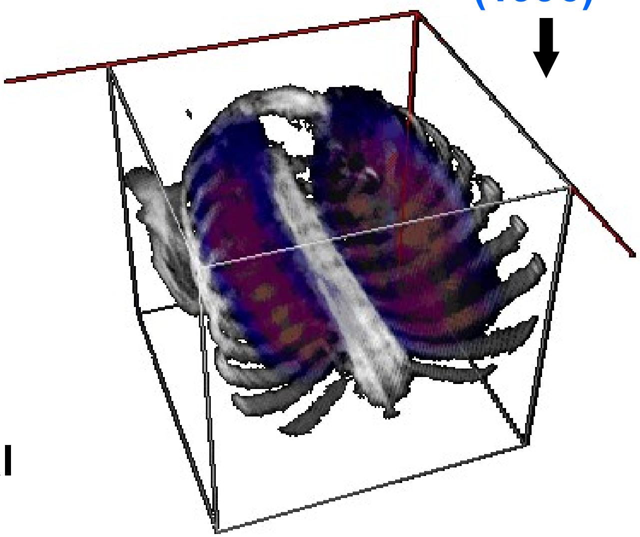


Tomography (2008)

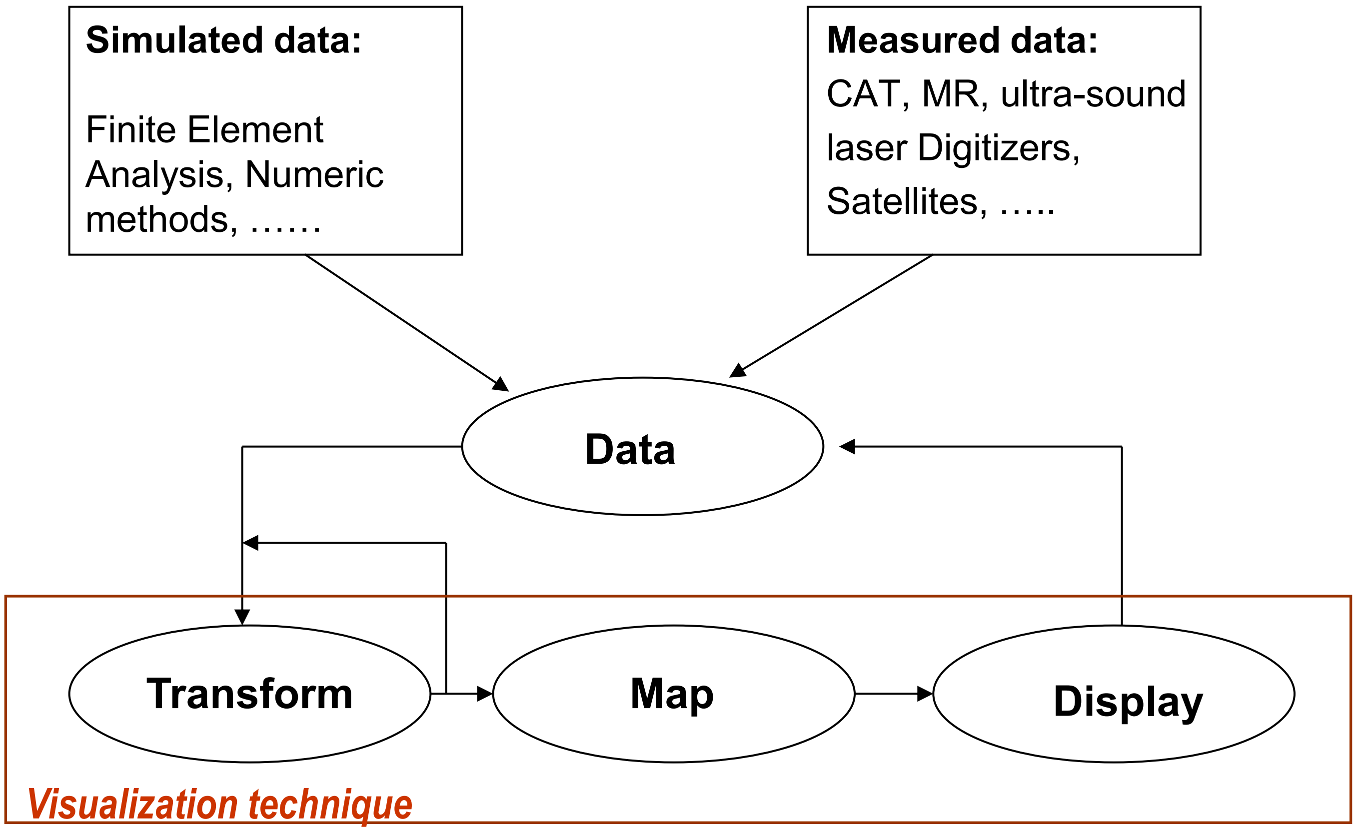


Electrical and mechanical ground resistivity (2010)

Tomography and SPECT (1996)



Scientific Visualization reference model



(adapted from Schroeder et al., 2006)

Adequate data pre-processing is vital!

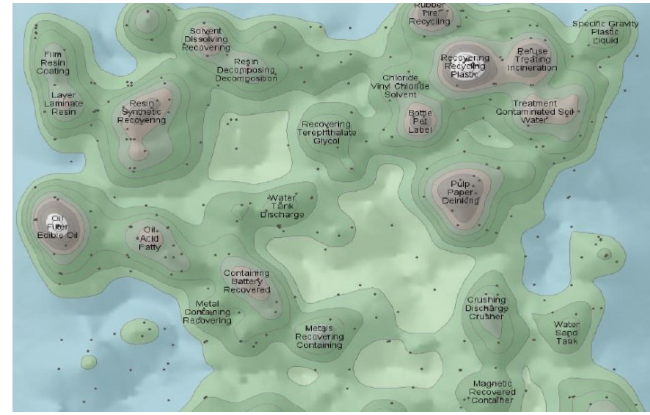
In this course we will not address in detail this phase;
we assume it is tackled in other courses

- The choice of the right mapping is fundamental, and is particularly difficult in InfoVis

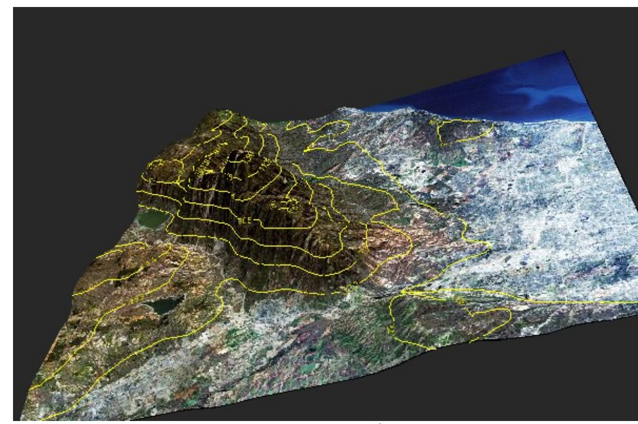
- It's generally easier in Data Visualization, since the data are inherently spatial

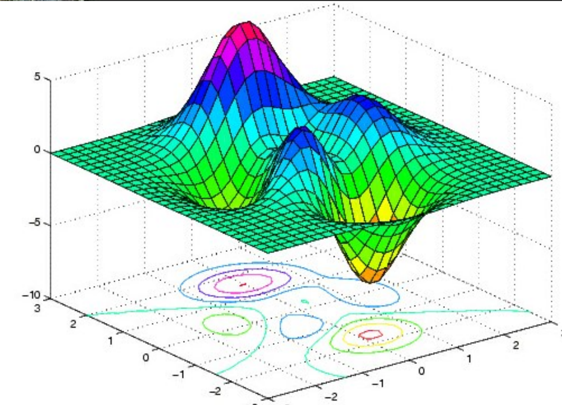
- Consider terrain altitude data, sea depth or values of a function:

  - different mappings or abstract visualization objects can be used,
  e.g.
    - contours (iso-lines)
    - pseudo-color
    - three-dimensional surface

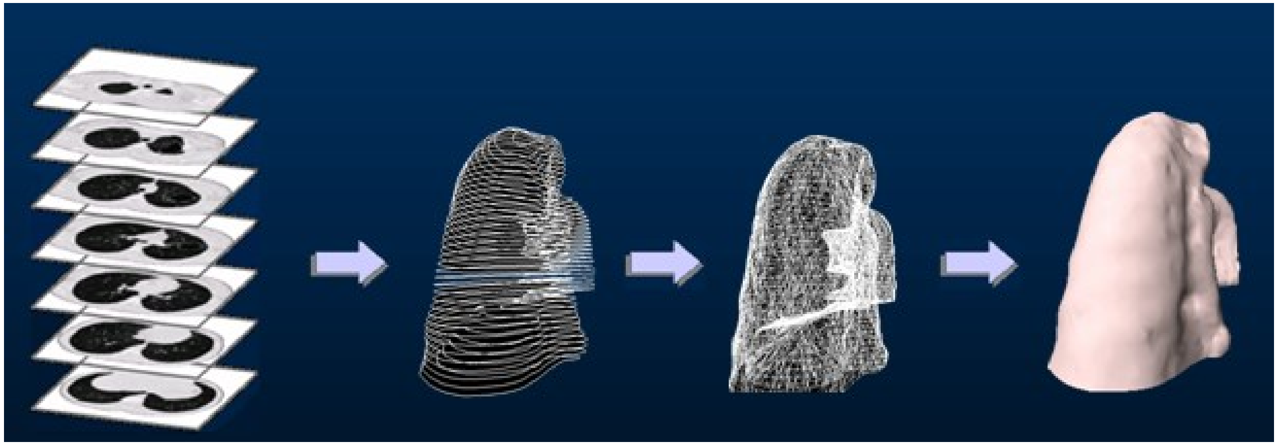


Patent landscape *(Cheng, 2003)*
http://www.ipo.gov.uk/informatic-recycleseparate.pdf

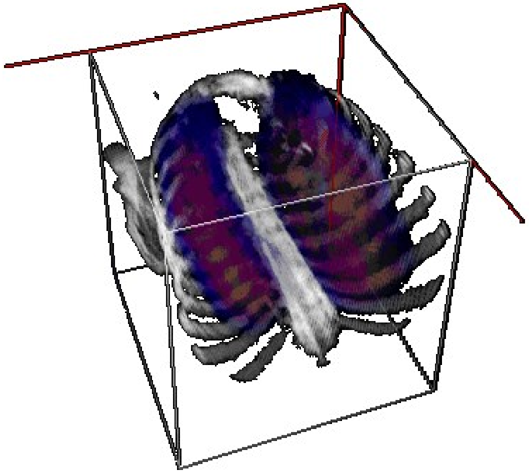


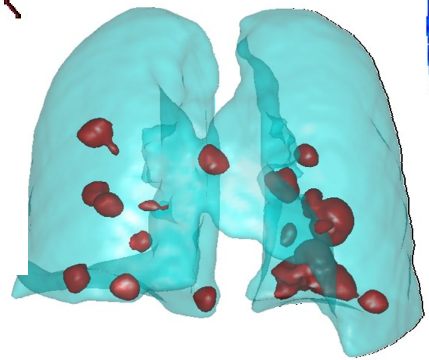


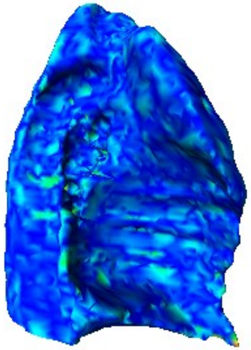
- E.g. medical data visualization, visualizations tend to be “literal” and thus the mapping phase does not vary as much as in other applications



- however, there are some possibility of variation







- In Visualization of abstract data there is no “natural metaphor”
- Making the visual mapping process more difficult

- Visualization may be used with different purposes:

  - personal exploration
  - discussion with colleagues
  - presentation to other people

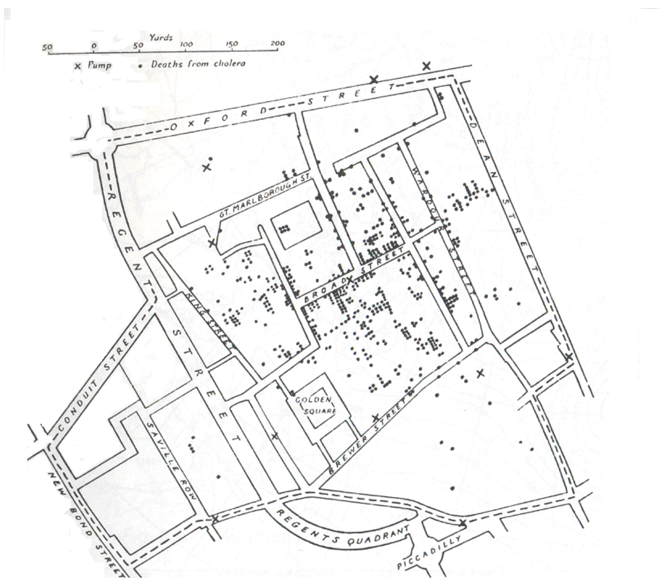
  for

  - explorative analysis
  - confirmative analysis

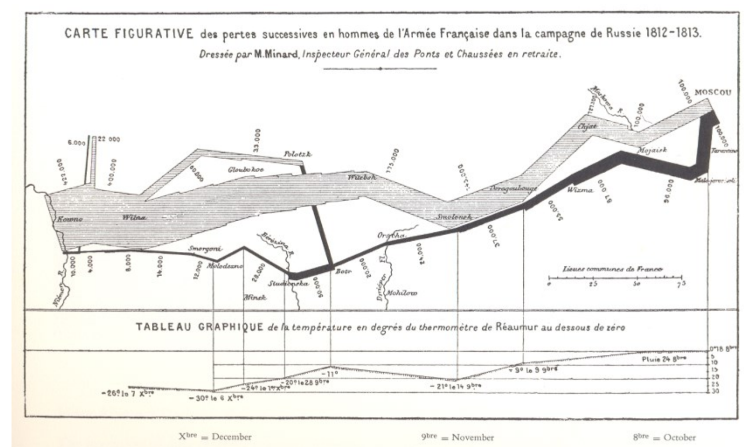
Classical examples for:

a) exploration
b) presentation

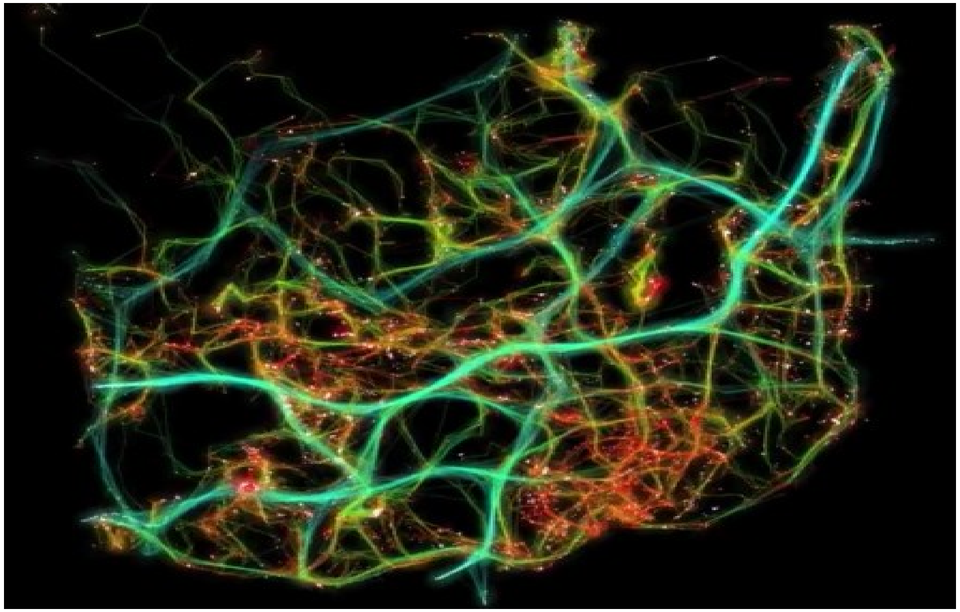
a)



b)



- Most often to promote insights and support users in work scenarios
- also in “Casual Visualization”, to depict personally meaningful information in visual ways that support everyday users also in non-work situations



https://vimeo.com/91325884

Pousman, Z., Stasko, J.T. and Mateas, M., 2007. Casual Information Visualization: Depictions of Data in Everyday Life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), pp. 1145-1152.

Presentation example: World health by Hans Rosling: 200 years of health/income – 120 000 values in 4 min

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

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?

Max and Min temperatures along the month of February (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 MaxT value attained?

Q4- How were the daily temperature ranges?

Q5 – What was the maximum temperature range?

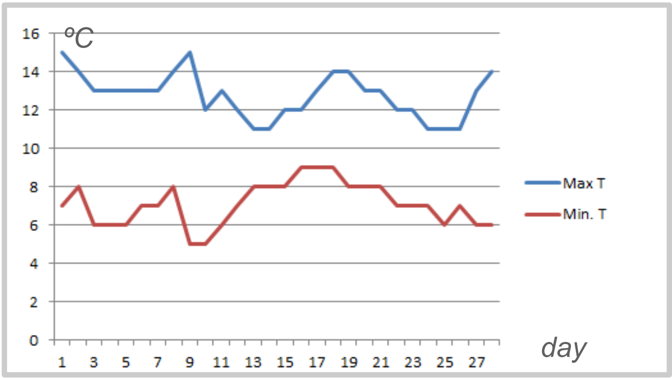
- What type of chart would you use to answer Q1?

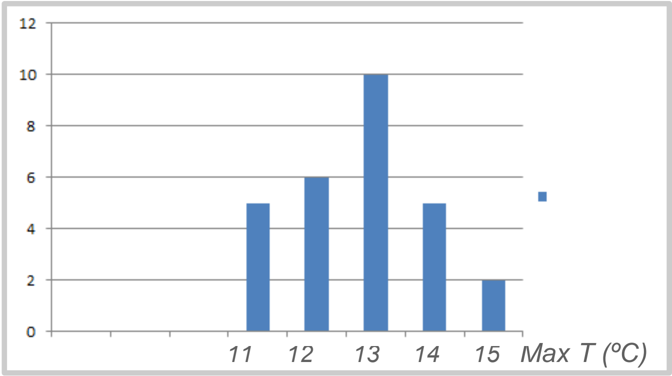
- And the other questions?

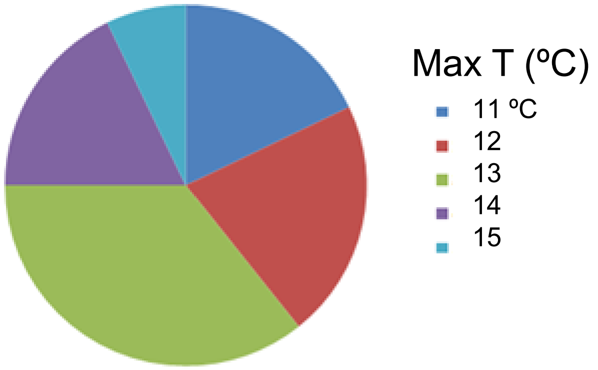
Example: how to select simple charts?

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



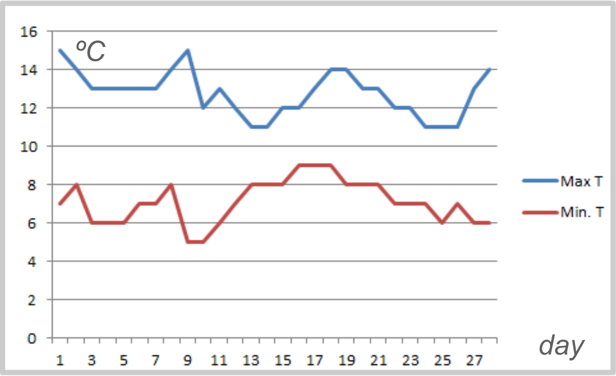




Simple example

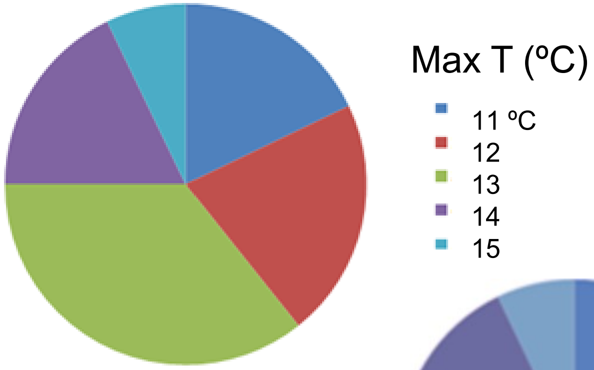
Temperatures along the month of February (in ºC):

Anything "odd" about this chart?

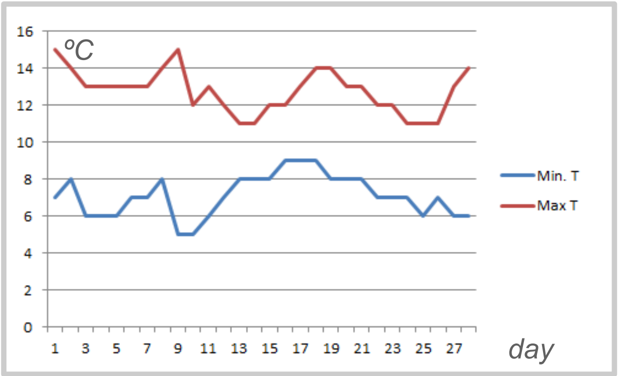


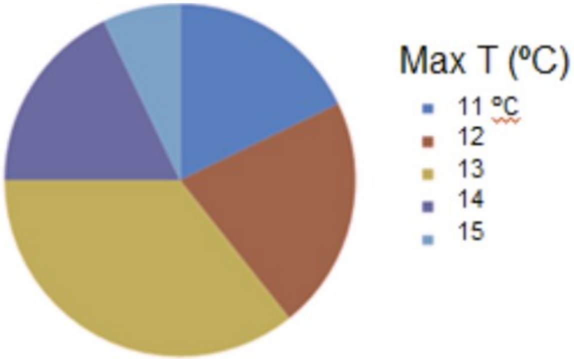
What if the user is color-blind?

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



Would you prefer this one?

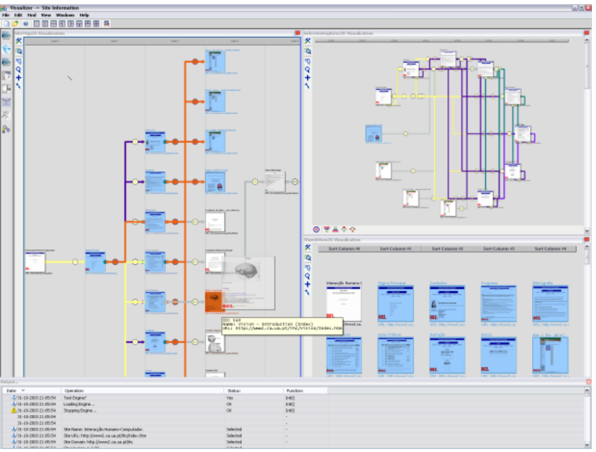


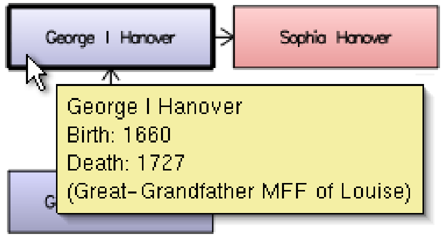


Do not forget "cultural" aspects,
nor individual differences!

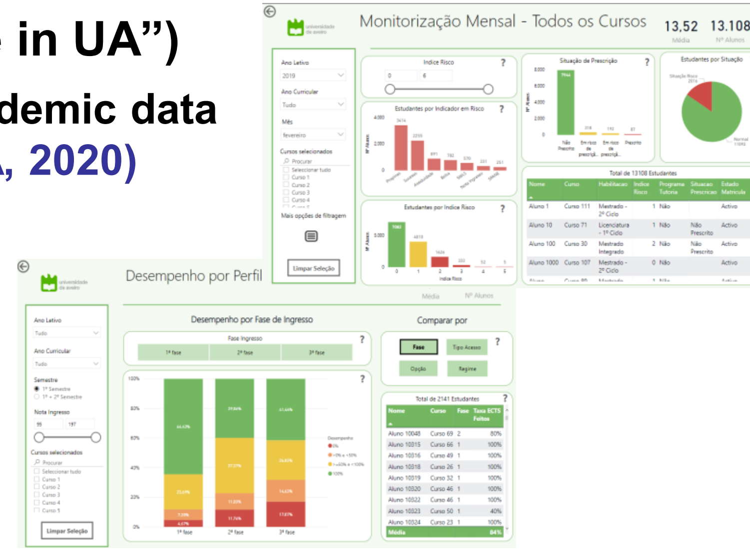
Information Visualization (examples “made in UA”)

Web site usage
(UA, 2004)

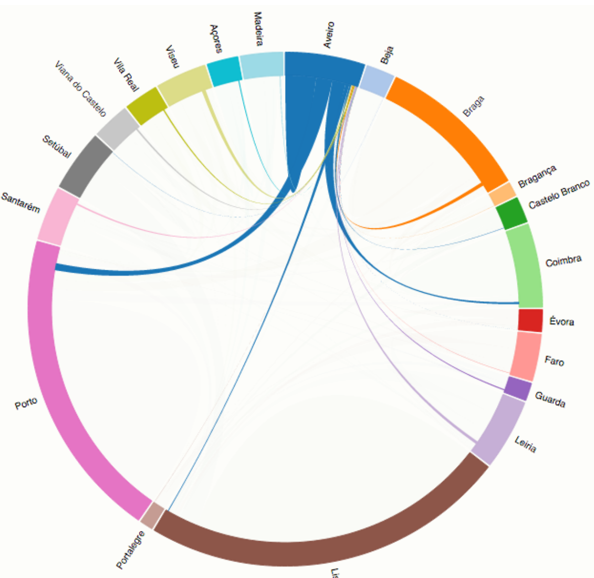


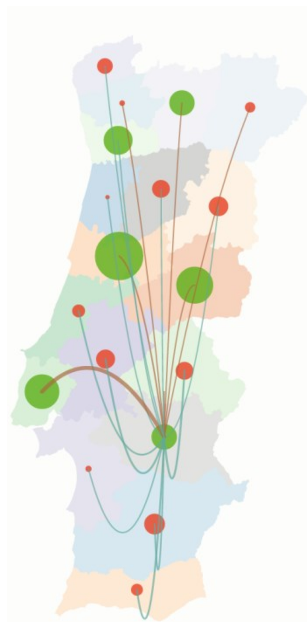


Academic data
(UA, 2020)



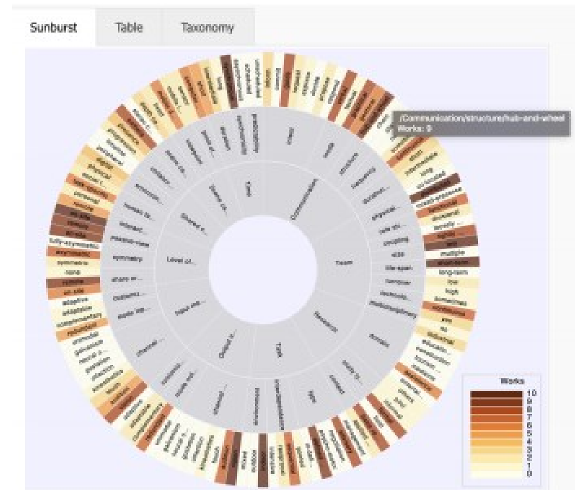
Student Migrations
(UA, 2015)



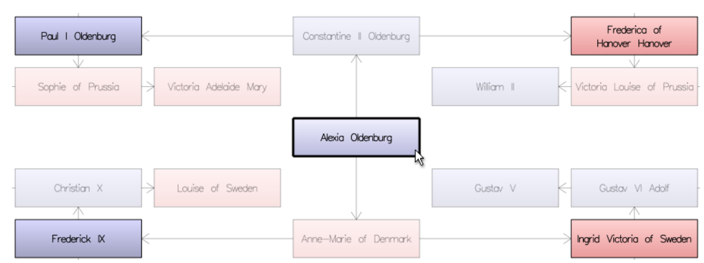


www.portugal-migration.info

Taxonomy Visualization
(UA, 2021)

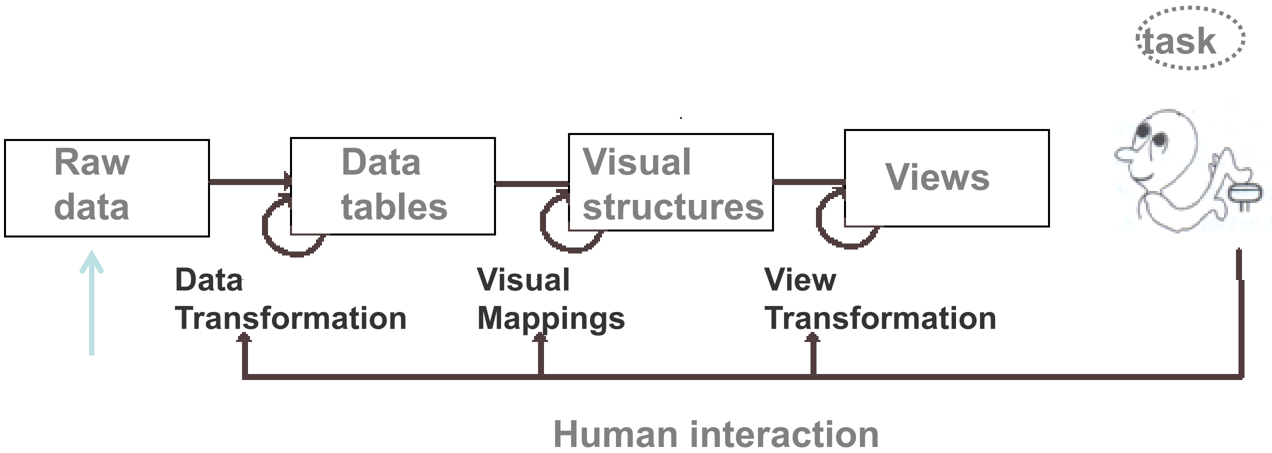


edigree trees
(UA, 2011)



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**

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

- Visualization may be the solution (or at least part of it)
- But:
  - How to produce a visualization?
  - And evaluate it?

We will address these topics in the next lectures …

- Data characteristics
- Representations, presentation and interaction
- The process of developing a Visualization application
- Human characteristics fundamental for Visualization
- Evaluation
    - Important skills in the profile of “**Data Scientist**”

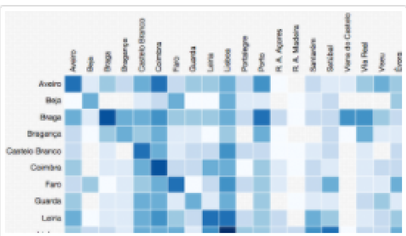
Example of a usability test of a visual exploration app

based on a web questionnaire to be answered by a user while observed by an experimenter

Data

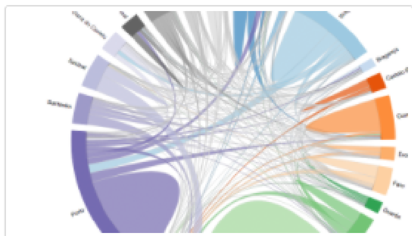
The candidates and institutions data were provided by Direcção Geral do Ensino Superior. The data represents three years (2012, 2013 and 2014) of Portuguese students applications to universities and polytechnic institutions. The dataset has 115636 students applications from 20 districts to 305 institutions.
The geography shapes data is from Direção Geral do Território.

Visualizations



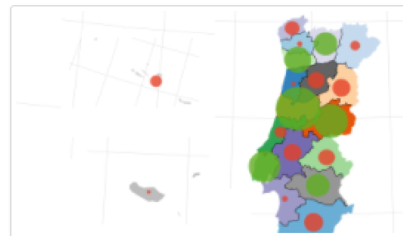
Adjacency Matrix

The adjacency matrix of the network is shown as a two-dimensional grid; each grid cell encodes the number of applicants moving from one district (on the left) to another district (at the top). Adjacency matrices are great for finding clusters (with appropriate sorting),



Chord Diagram

A chord diagram arranges graph nodes (districts) radially, drawing thick curves between nodes. The thickness of a chord encodes the number of applicants moving between districts. Like matrix diagrams, chord diagrams reveal asymmetries: if a chord is tapered,



Map

The map diagram allows you to explore migrations with a geo-spatial reference. Each district is a node, you can click in a district node to visualize the applicants migration; color will help you to understand the net balance of each district and destination.

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

InfoVis Books

- **Spence, R., *Information Visualization, An Introduction*, Springer, 2014**
- **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, Design for Interaction*, 2nd ed., Prentice Hall, 2007**
- **Mazza, R., *Introduction to Information Visualization*, Springer, 2009**
- Ware, C., *Information Visualization, Perception to Design*, 2nd ed.,Morgan Kaufmann, 2004
- Card, S., J. Mackinlay, B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, Academic Press, 1999
- Bederson, B. , B. Shneiderman, *The Craft of Information Visualization: Readings and Reflections*, Morgan Kaufmann, 2003
- *Tufte, E., The Visual Display of Quantitative Information*, Graphics Press, 1983
- Tufte, E., *Envisioning Information,* Graphics Press, 1990
- Friendly, M., "Milestones in the history of thematic cartography, statistical graphics, and data visualization“, 2008

Data Vis Books

- Ward, M. G. Grinstein and D. Keim, *Interactive Data Visualization: Foundations, Techniques, and Applications*, A K Peters/CRC Press , 2010
- Schroeder, W., K. Martin, B. Lorensen, *The Visulization Toolkit- An Object Oriented Approach to 3D Graphics*, 4th ed., Prentice Hall, 2006