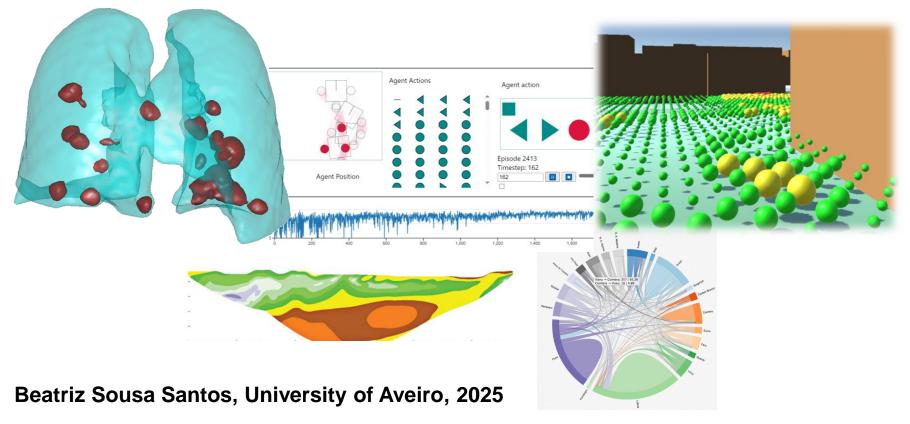
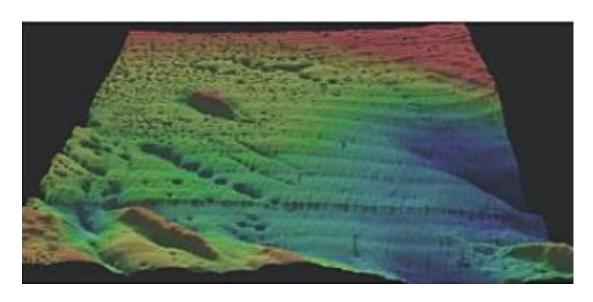


# An Introduction to Data and Information Visualization

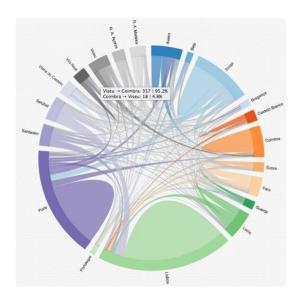


#### What is Visualization?

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



Passamoquoddy Bay (10<sup>6</sup> 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..."

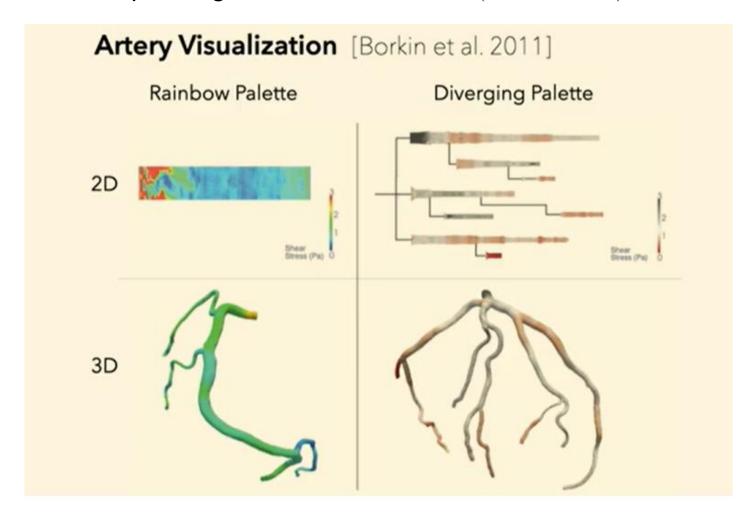
We see patterns displayed in certain ways more easily

than in others ...

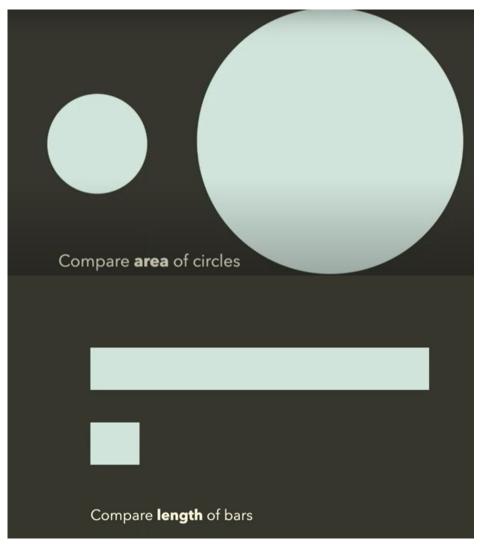


#### Some visual encodings are better than others...

Visualization is not just "intellectual curiosity, but potentially saves lives, depending on the decisions..." (Heer, 2023)



#### Some visual encodings are better than others...



How many times is the largest circle larger?

How many times is the largest rectangle larger?

(Heer, 2023)

<u>Jeff Heer on Augmenting Data Scientists: The Promise and Peril of Al-Assisted Analysis</u> (youtube.com)

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

(Nussbaumer Knaflic, 2015)

The best encodings depend on the user's tasks...

Storytelling with Data | Cole Nussbaumer Knaflic | Talks at Google

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

Anscombe's Quartet: Raw Data								
	I		II		III		IV	
	x	$\mathbf{y}$	x	$\mathbf{y}$	x	$\mathbf{y}$	x	$\mathbf{y}$
	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's quartet - Wikipedia

#### Outline of the lectures:

#### **Data and Information Visualization: Introduction**



#### Information Visualization:

- Creating a visusalization
- Data characteristics
- Representation
- 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

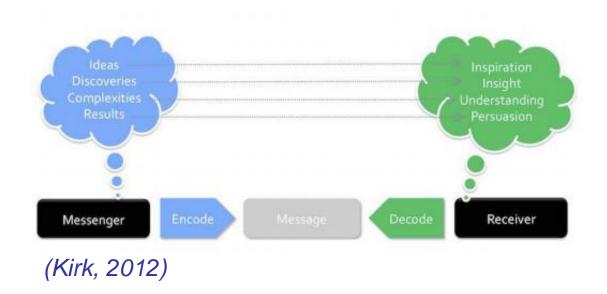
How to obtain and evaluate a Visualization?

#### **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 mainly Data Visualization ...

#### Definition (yet another)

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



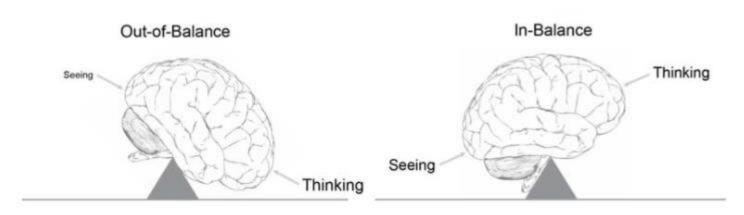
 Visualization is different from Computer Graphics and Image Processing since:

```
1- it deals mainly with multi-dimensional data (>= 3D, time varying)
```

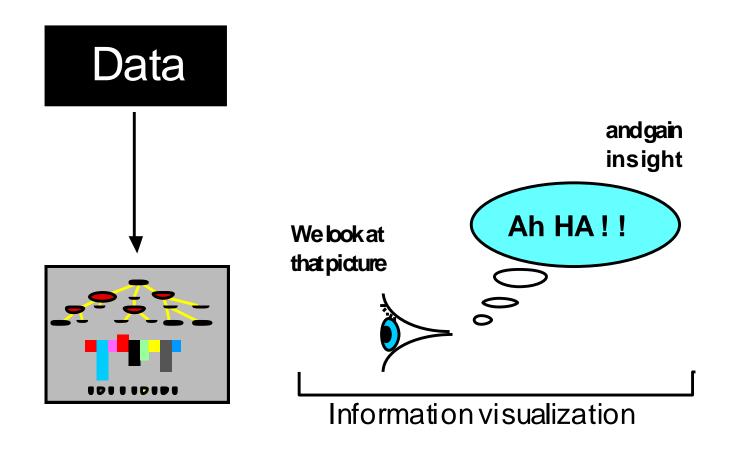
- 2- data transformation is fundamental (may be changed to increase insight)
- However, there is some overlap:
  - The output of a visualization process is images
  - In general uses much CG

#### Visualization benefits

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

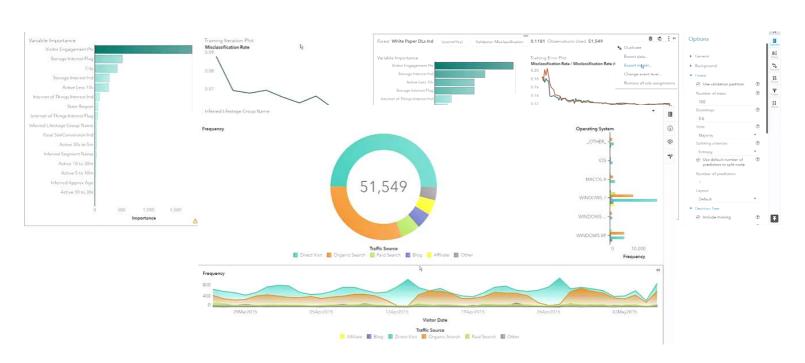


Data Visualization for Human Perception | The Encyclopedia of Human-Computer Interaction, 2nd Ed.



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.

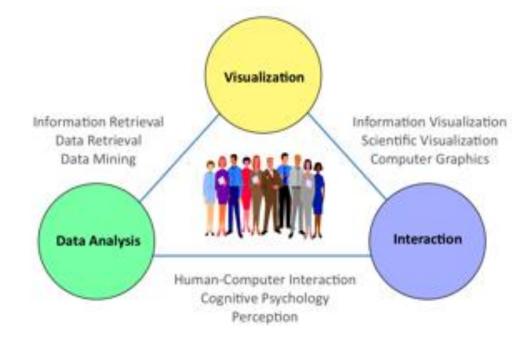
#### Visual Analytics

The science of analytical reasoning supported by interactive visual interfaces.

"Detect the expected and discover the unexpected"

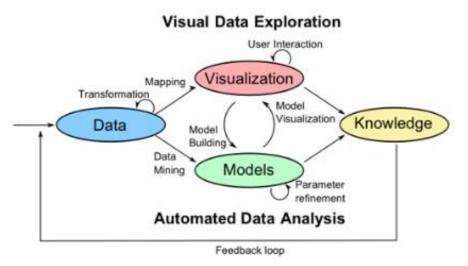
Illuminating the Path - The Research and Development Agenda for Visual Analytics

(Thomas and Cook, 2006)



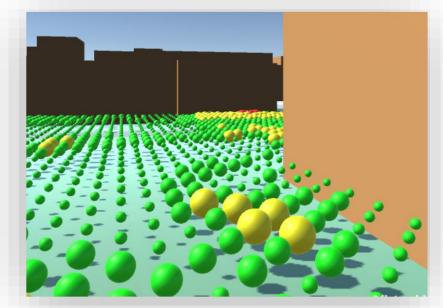
"The Visual Analytics Process combines automatic and visual analysis methods with a tight coupling through human interaction in order to gain knowledge from data."

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

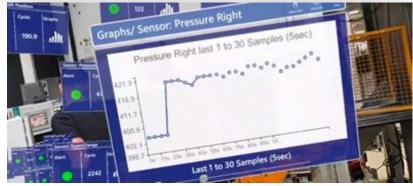


#### Situated Visualization

Examples in industry and pollution awareness combining Visualization & Augmented Reality









#### Immersive Visual Analytics



#### 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 rid 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



#### 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

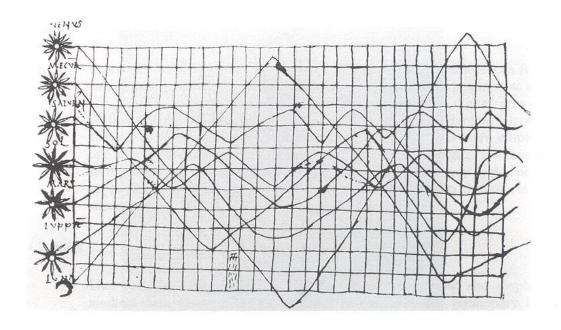
Visualization in Scientific Computing (McCormick, de Fanti and Brown – 1987)

#### 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

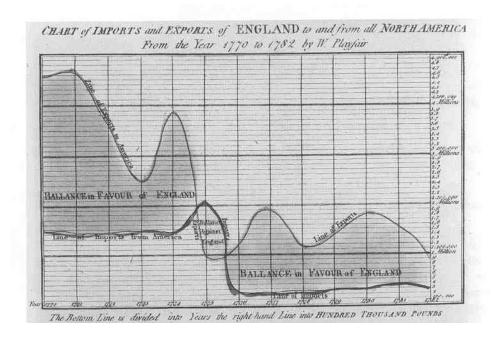
#### "Pre-computer" Visualization:

#### One of the oldest known Visualizations



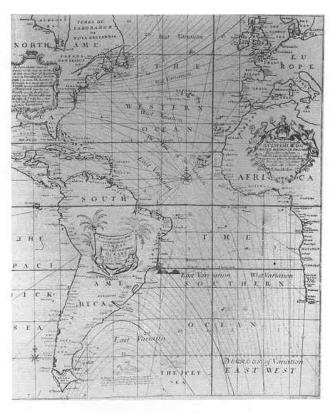
Inclination of orbits along the time - Xth century (Tufte, 1983)

#### One of the first Visualizations used in "business"



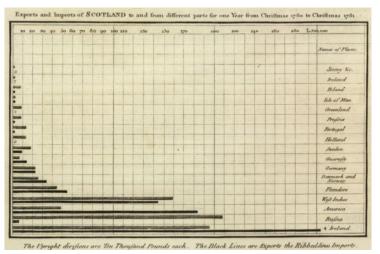
Import/export during the period from 1770 to 1782 by William Playfair (Tufte, 1983)

### One of the first visualizations using contours (isolines)



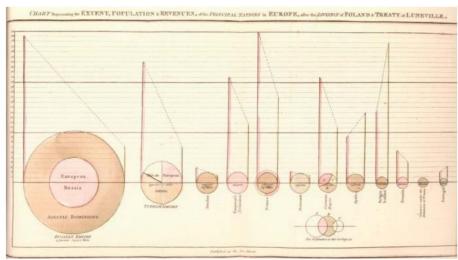
Magnetic declination 1701 Edmund Halley (Tufte, 1983)

#### "Ancestors" of simple representations of univariate data

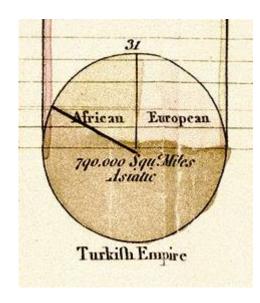


Exports and Imports of Scotland to and from different parts for one W. Playfair's *The Commercial and Political Atlas, 1871* 

https://en.wikipedia.org/wiki/William\_Playfair

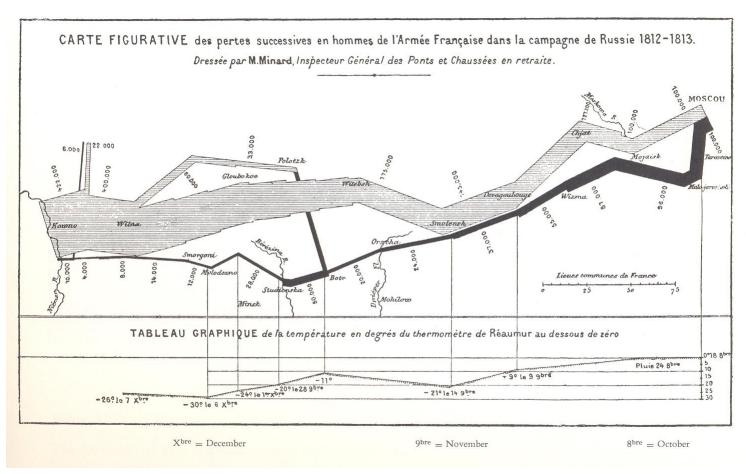


W. Playfair, Statistical Breviary, 1801



#### Multidimensional Visualization

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

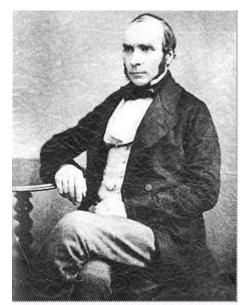


Russia campaign of Napoleon 1861 by Charles Minard (Tufte, 1983)

#### Visualization in scientific discovery



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



Dr. John Snow



#### 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 ...

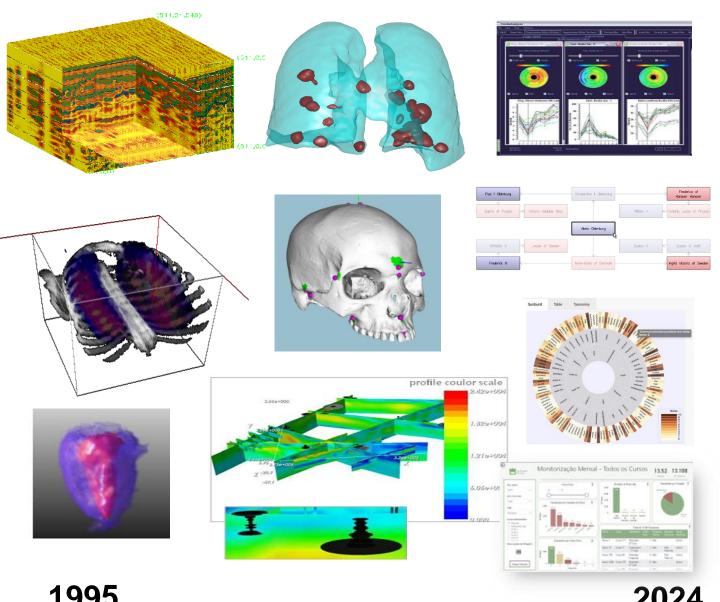
#### IEEE Visualization: evolution of names in a major conference ....

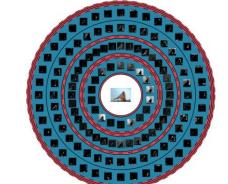
Vis -> SciVis ~2015

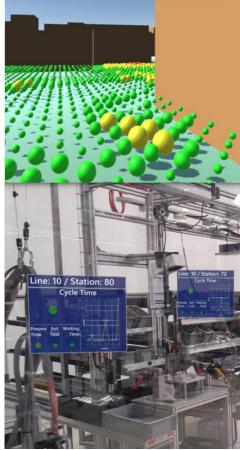


T. Dang, H.N. Nguyen, and V. Pham, WordStream: Interactive Visualization for Topic Evolution EUROVIS 2019

#### Our works: from Sci Vis to InfoVis to Sit Vis & XAI







## Applications of Scientific Visualization

- 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?

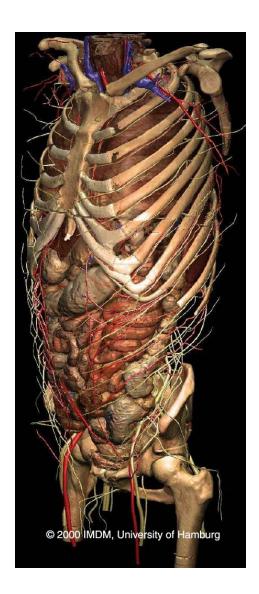
## Medicine (education)

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

Visible Body - Virtual Anatomy to See Inside the Human Body

The National Library of Medicine's Visible Human Project

The Visible Human Project - Applications



An example of Scientific Visualization:

The visible Human Project

(1994, 1995)

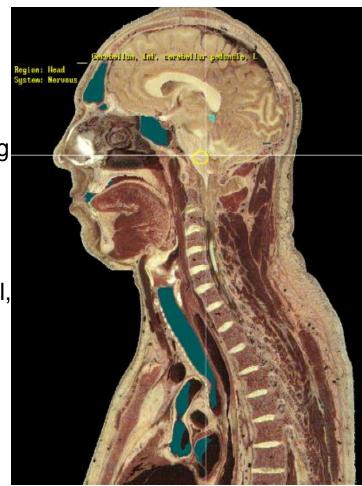
The data sets were designed to serve as

- (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.



## Medicine (e.g. surgery training)

VOXELman, University of Hamburg

 Movement of the drill is controlled with a force feedback device



<u>UKE - Faculty of Medicine - VOXEL-MAN</u>
Cavity Preparation with Voxel-Man Dental

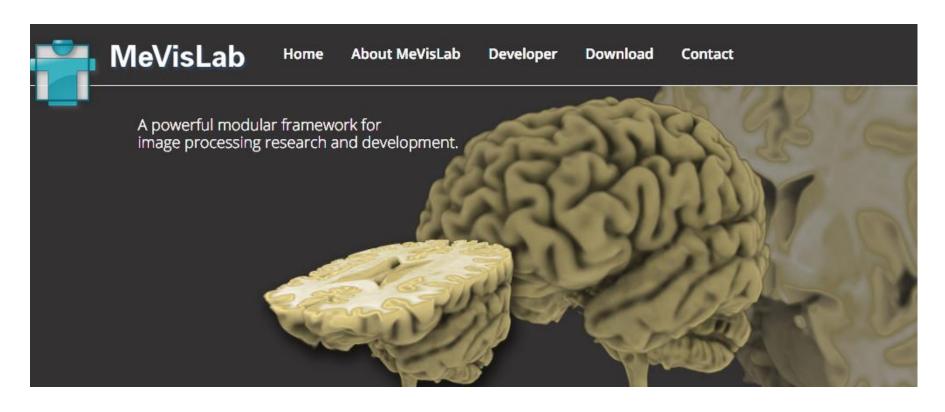
Combining imaging from MRIs, CT scans and angiograms to create a three-dimensional model that physicians and patients can see and manipulate — just like a virtual reality game — Stanford Medicine



<u>Virtual reality system helps surgeons, reassures patients | Medical Center Development | Stanford Medicine</u>

Virtual reality, augmented reality can improve surgeons' training | STAT

## **Medical Imaging**



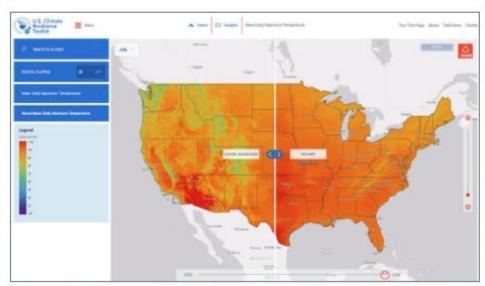
MeVisLab: MeVisLab

## 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.



Visualizing Climate Data | NOAA Climate.gov

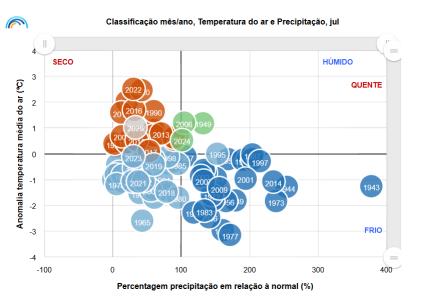


View by Variable interface. View Maximum Daily Temperature variable in Climate Explorer.

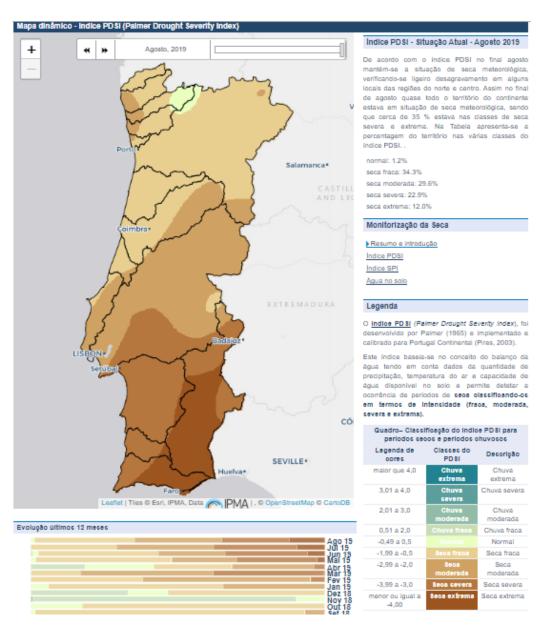
Climate Explorer | U.S. Climate Resilience Toolkit

## Climate monitoring:

# Drought Severity Index Precipitation (by IPMA)

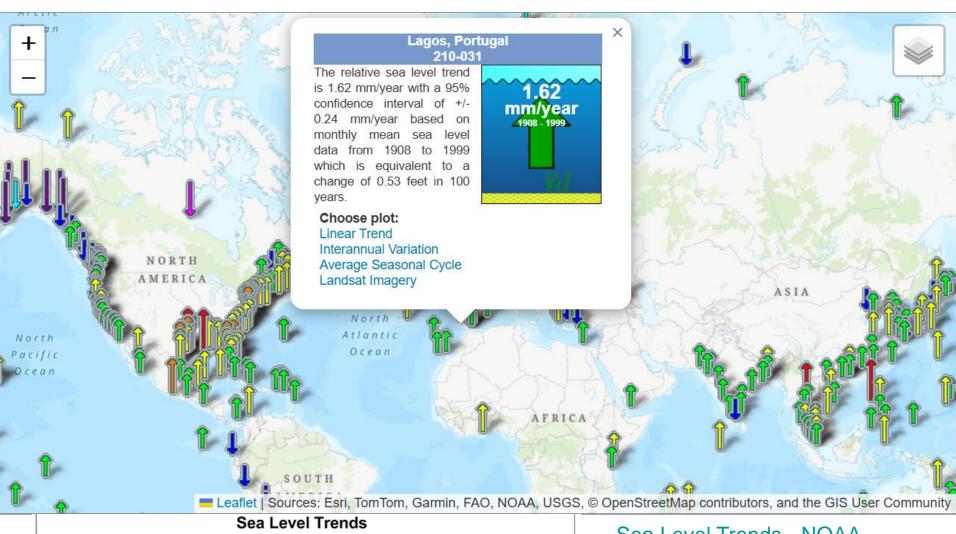


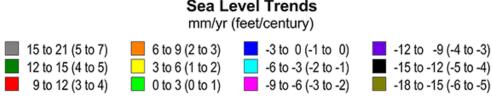
IPMA - Resumo boletim clima



IPMA - Monitorização da Seca Meteorológica

#### See Level Trends





Sea Level Trends - NOAA
Tides & Currents

## Visualization and Virtual Reality at the Automotive Industry

90% of the new Maserati M20 was digitally developed

Tested in a VR simulator, improving results, reducing time and cost of development

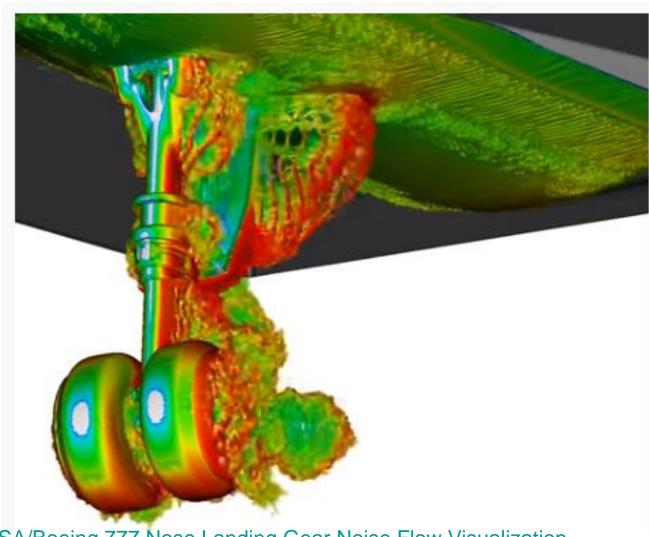


Maserati Innovation Lab LIVE TOUR | Motor Valley Fest Digital 2020

- YouTube

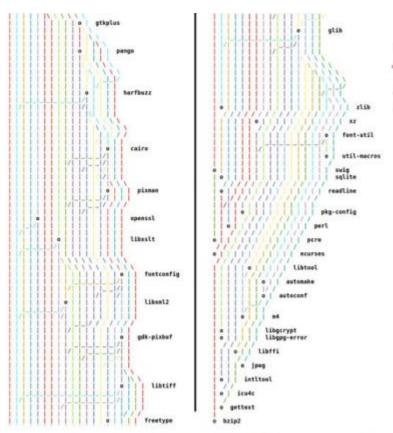
## Example of fluid mechanics visualization

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



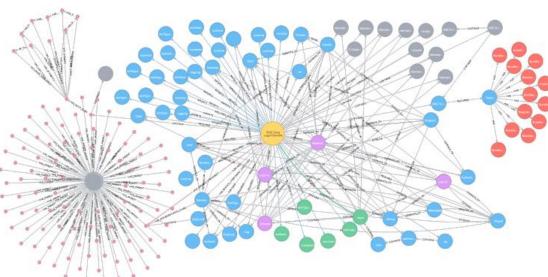
NASA/Boeing 777 Nose Landing Gear Noise Flow Visualization

### Software visualization

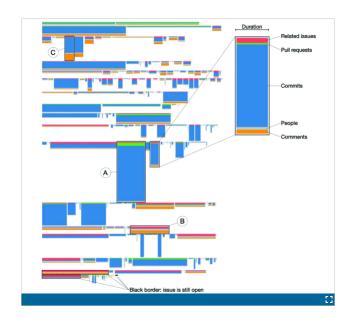


git-style package dependency graph of dia (also shown in Fig. 1). The freetype node has been duplicated to show alignment between the two halves.

https://ieeexplore.ieee.org/document/8419271



https://ieeexplore.ieee.org/document/8742198



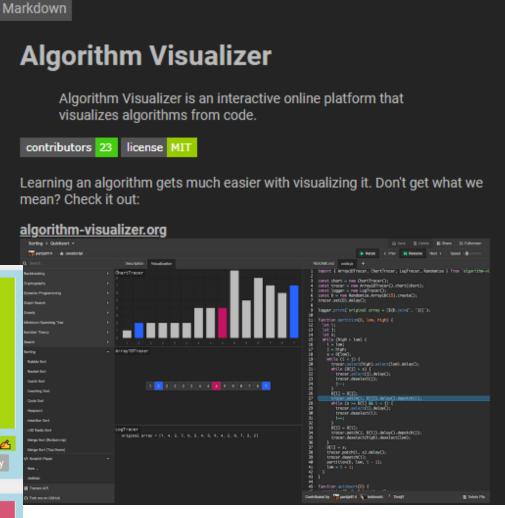
https://ieeexplore.ieee.org/document/9604892

## Algorithm visualization

 Beyond mathematical and empirical analyses of algorithms



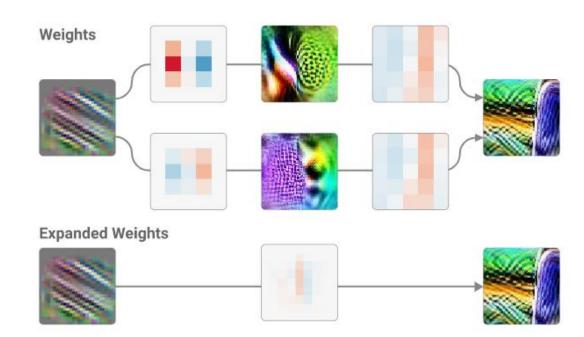


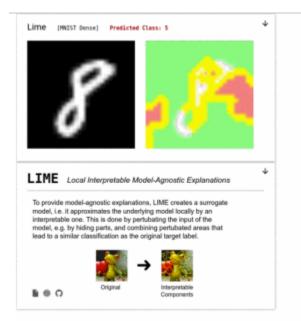


visualising data structures and algorithms through animation
- VisuAlgo

## Machine Learning visualization

 To help understand the "inner workings" of neural networks and other AI methods





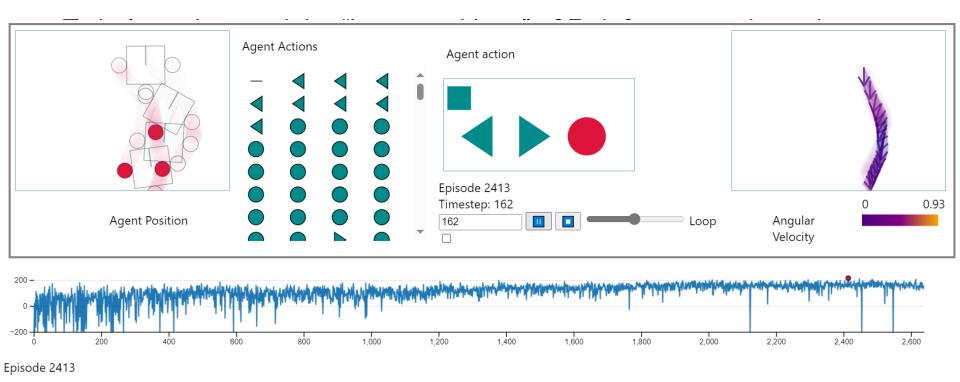


#### Visualizing Weights

Visualization Tools for Machine Learning Pipelines: A Review -2025 | IEEE Xplore

A survey of surveys on the use of visualization for interpreting machine learning models - 2020

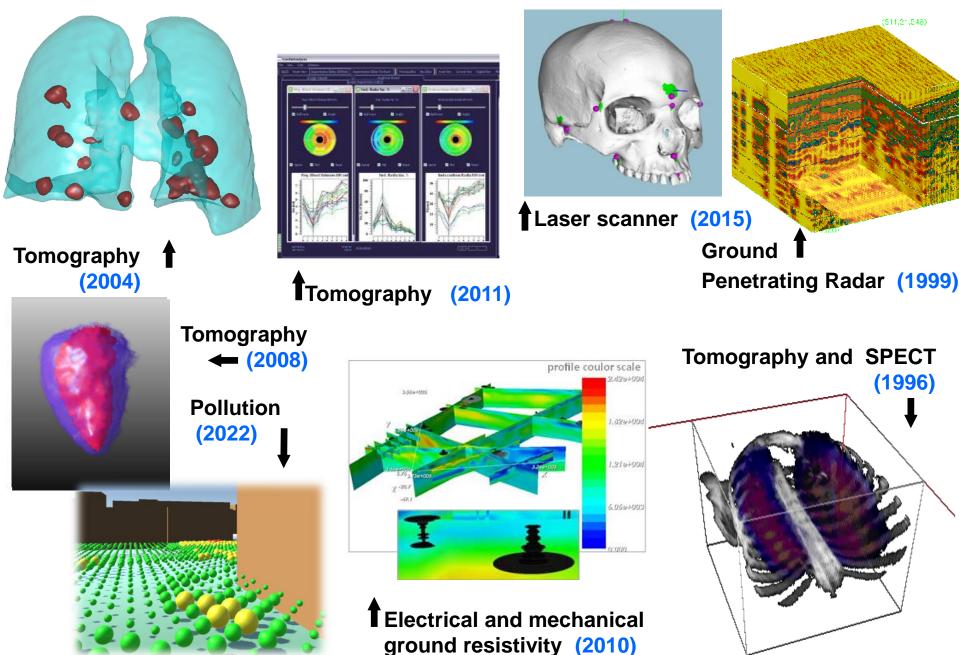
## Machine Learning visualization



Experiment with this Reinforcement Learning simple example:

Main (tiagodavi70.github.io)

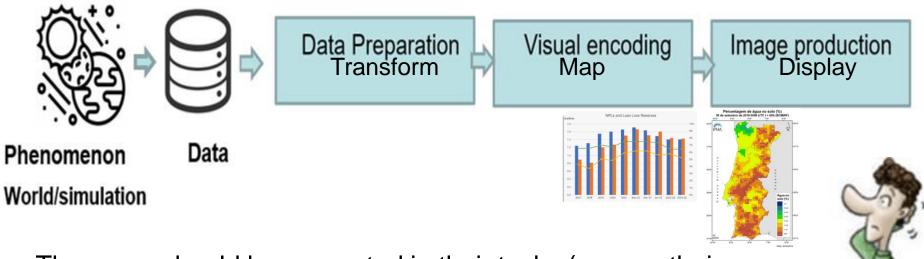
## Scientific Visualization (examples "made in UA")



## The process of creating a (simple) 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

It is a human-in-the-loop process

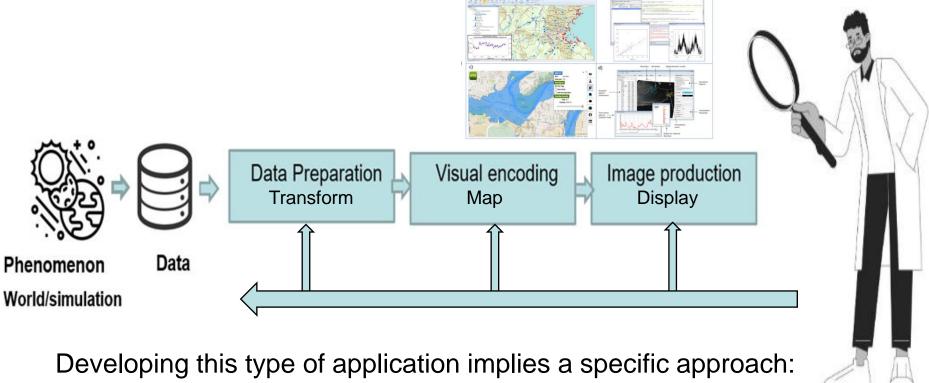
"human-in-the-loop" problems involve the user as a part

- 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



**Human-centred design** 

#### Data can be

- simulated

```
(e.g. stress of a mechanical part, phantom of the human body, etc.)
```

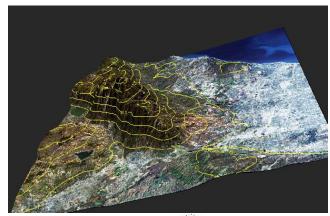
- acquired from the real world

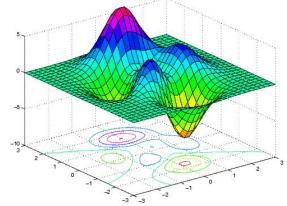
- Then a visualization technique is applied, involving:
  - data **transformation** through several methods (e.g. scale transformation, noise filtering, outlier elimination, changing resolution, etc.)
  - mapping to an adequate form to represent data visually (e.g. lines, points, color)
  - producing an image or sequence of images (display)
- This process is repeated as needed to provide insight

- 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

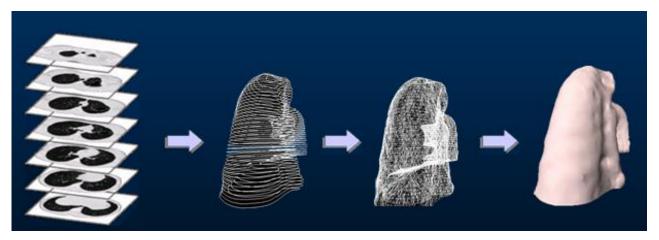


Patent landscape (Cheng, 2003) <a href="http://www.ipo.gov.uk/informatic-recycleseparate.pdf">http://www.ipo.gov.uk/informatic-recycleseparate.pdf</a>

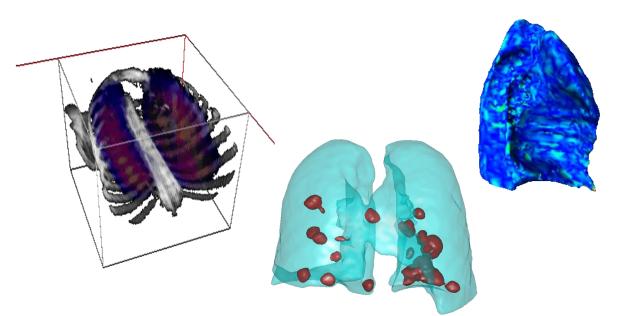




 E.g. medical data visualization, visualizations tend to be "literal" and thus the mapping phase does not vary as much as in other applications



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

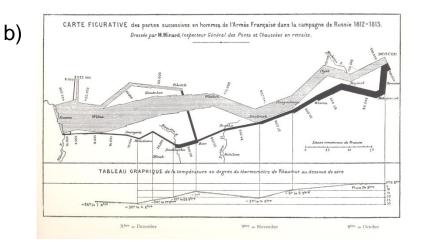
- explorative analysis
- confirmative analysis

Classical examples for:

a) exploration

for

b) presentation



## Presentation example: World health

by Hans Rosling: 200 years of health/income - 120 000 values in 4 min



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 °C):

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?

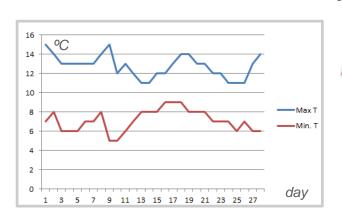
## Example: how to select simple charts?

## Temperatures along the month of February (in °C): a few possible charts

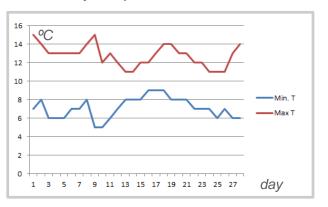
day	Max T	Min. T	Max and Min Temperatures		
1	15	7	16		
2	14	8	10	T	
3	13	6	14		
4	13	6	40	×	
5	12	6	12	I	
6	13	7	16 OC 10		
7	13	7	14	Т	
8	14	8	12 2 8		
9	15	5	16 0C 10 10 10 10 10 10 10 10 10 10 10 10 10		
10	12	5		$\perp$	
11	13	6	Min. T		
12	12	7			
13	11	8	4		
14	11	8	0		
15	12	8	o + dav		
16	12	9	1 3 5 7 9 11 13 15 17 19 21 23 25 27	Max temp Min temp	
17	13	9			
18	14	9	12		
19	14	8	10	Max T (%	
20	13	8		■ 11 °C	
21	13	8	8	12	
22	12	7	6	13	
23	12	7		14	
24	11	7	4	15	
25	11	6	2	. 19	
26	11	7			
27	13	6	0 44 40 45 44 45 T/00)		
28	14	6	11 12 13 14 15 Max T (°C)		

## Simple example

#### Temperatures along the month of February (in °C):



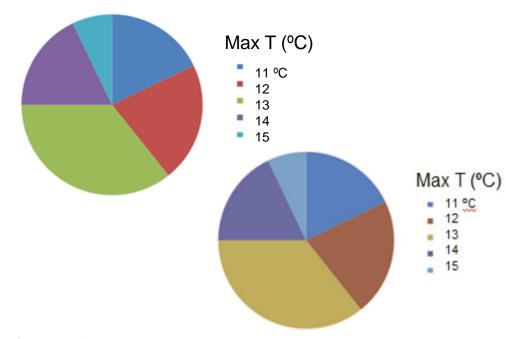
#### Would you prefer this one?



## Anything "odd" about this chart?



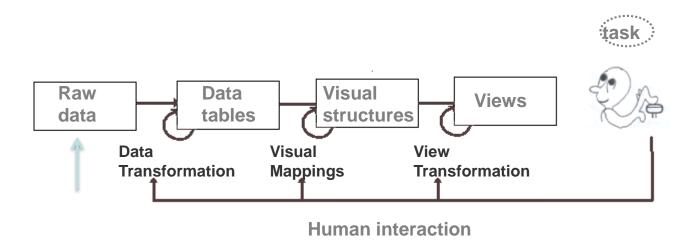
Test it using <u>Coblis — Color Blindness Simulator</u> <u>— Colblindor/</u>



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

### Information Visualization Reference Model (Card et al., 1999)

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

#### Remember:

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

## What future?

- Visualization seems to have becoming and continuing to be:
  - More real-time
  - More interactive
  - More intelligent
  - More immersive
  - More collaborative
  - More multisensorial
  - More accessible and ethical
  - **—** ...
  - Integrating more other technologies (edge and quantum computing...)

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 ...



## **Next sessions:**

- 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"

#### **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
- Fisher, D., Meyer, M., Making Data Visual: A Practical Guide to Using Visualization for Insight, O'Reilly, 2018
- Ware, C., Information Visualization, Perception to Design, 3rd ed., Morgan Kaufmann, 2012
- Ward, M., Grinstein, G., Keim, D., Interactive Data Visualization: Foundations, Techniques, and Applications, 2nd ed., CRC Press, 2015
- 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

## **Interesting Talks**

Michael Friendly: History of Visualizations (youtube.com)

<u>Jeff Heer on Augmenting Data Scientists: The Promise and Peril of Al-Assisted Analysis</u>

