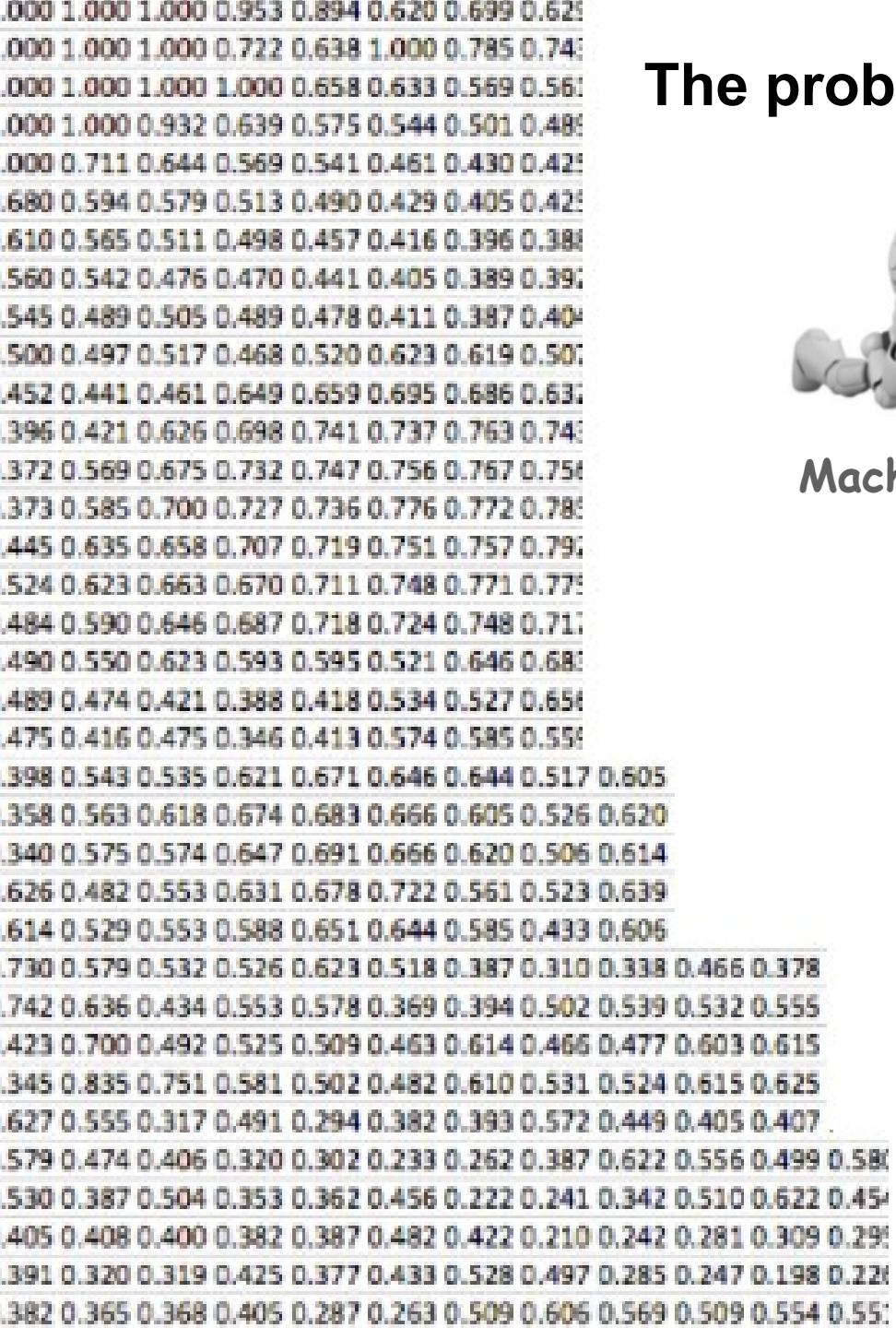
A Brief introduction to Visualization

Beatriz Sousa Santos

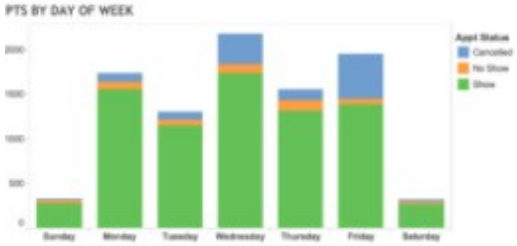
DETI/ University of Aveiro

Nokia, March, 2019

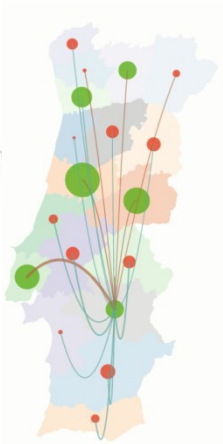
The problem …

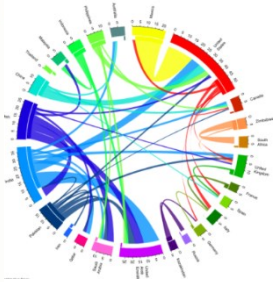


Machine learning?



Visualization?





Statistics?

Outline of the talk:

Visual data Mining/Visualization in the Data Science process

Brief historical overview

Data / Information Visualization

Information Visualization:

- Main issues
- Data and Design
- Representation, Presentation, Interaction

Main takeaways and some guidelines

To probe further: books, conferences, papers, sites, tools …

“Icebreaker”

- Who has already used visualization?
- Who uses visualization on a regular basis?

Visual Data Mining

- "The basic idea is to **present the data in some visual form**, allowing the human to **get insight into the data**."

  (Keim, 2002)

- Specially useful in **exploratory analysis** when **little is known about the data** and the exploration **goals are vague**

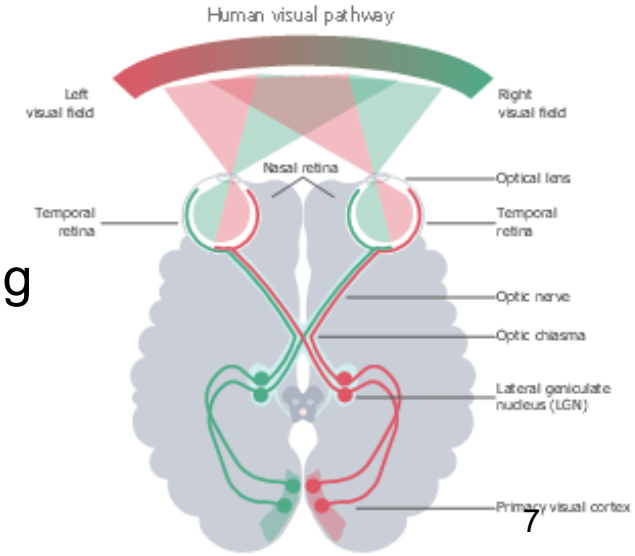
- Since the **user is directly involved**, shifting and **adjusting the exploration goals is automatically** done if necessary

Visual Data Mining and Visualization

- Visualization is a field of Computing focused on how to **visually represent and explore large amounts of data**

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

- **Visual Data Mining uses visualization** to facilitate data exploration and understanding



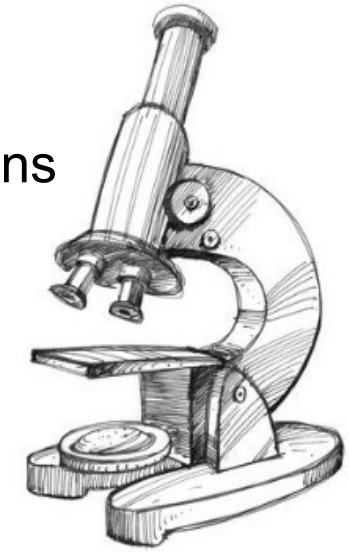
The **purpose of visualization is insight, not pictures**

- Like a telescope or microscope …

visualization amplifies cognitive abilities

to understand complex processes to support better decisions

(Ben Shneiderman)



https://medium.com/multiple-views-visualization-research-explained/the-purpose-of-visualization-is-insight-not-pictures-an-interview-with-visualization-pioneer-ben-beb15b2d8e9b

Visualization and Machine Learning

- **Share a focus on data** and information

- **The main difference is the role of the user** in the data exploration and modeling: (Keim et al., 2012)
  - Machine Learning → **get rid of the user**
  - Information Visualization → **allow the user to discover patterns and adjust models**

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

Main advantages of visual over automatic data mining

(Keim, 2002)

- deal **easily with highly inhomogeneous and noisy data**
- **intuitive** and **require 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

When are Visualization solutions 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.

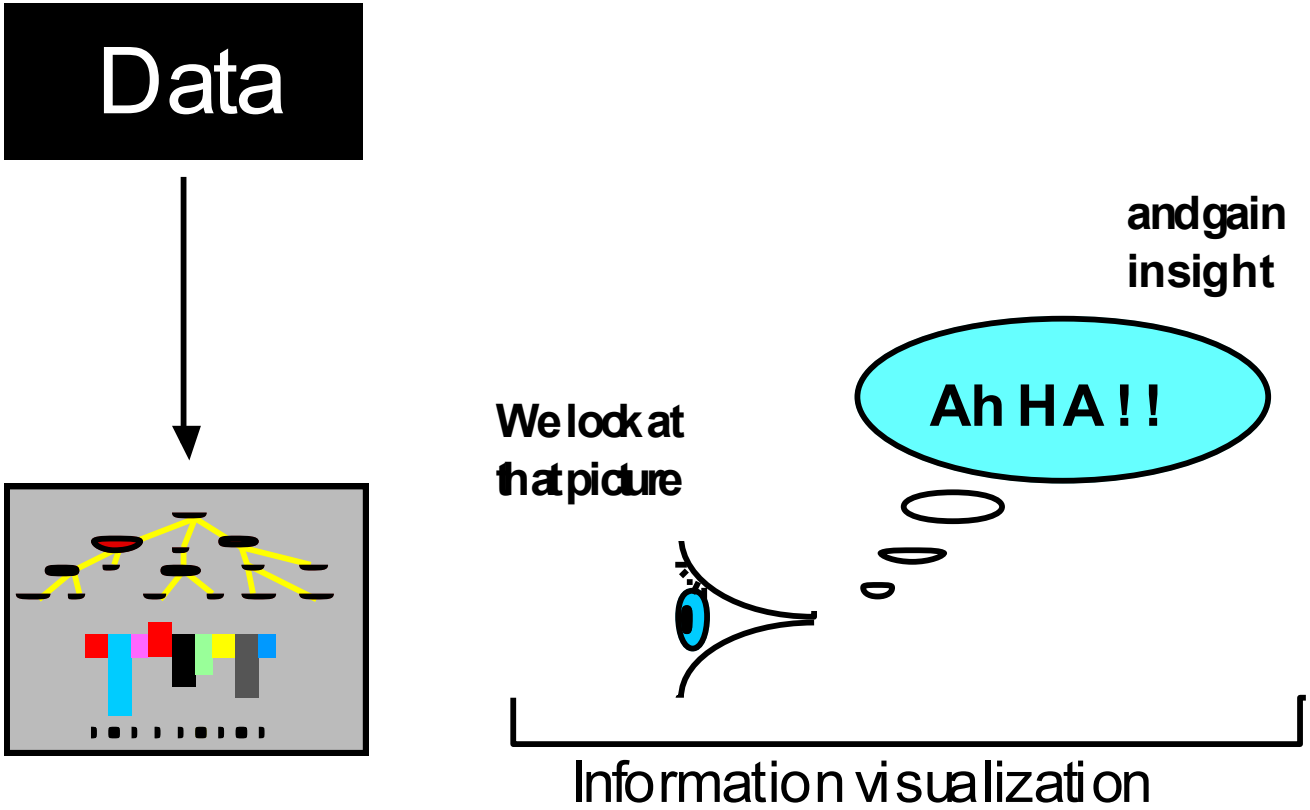
(Munzner, 2014, chap. 1)

Information 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
- Analyzing the results …

https://nips.cc/Conferences/2018/Schedule?showEvent=10986

The process of Visualization



graphically encoded data is viewed in order to form a mental model of that data and understand the phenomenon

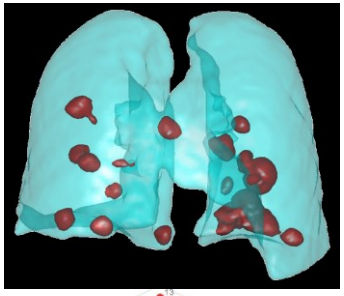
(Spence, 2007)

It is a **Human in the loop process**!

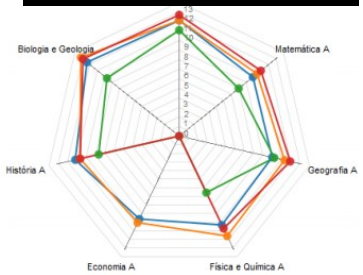
Visualization as a scientific field

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

- 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)



- B**oth start with (raw) data and allow to extract information …**

Brief history of Visualization

- 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:

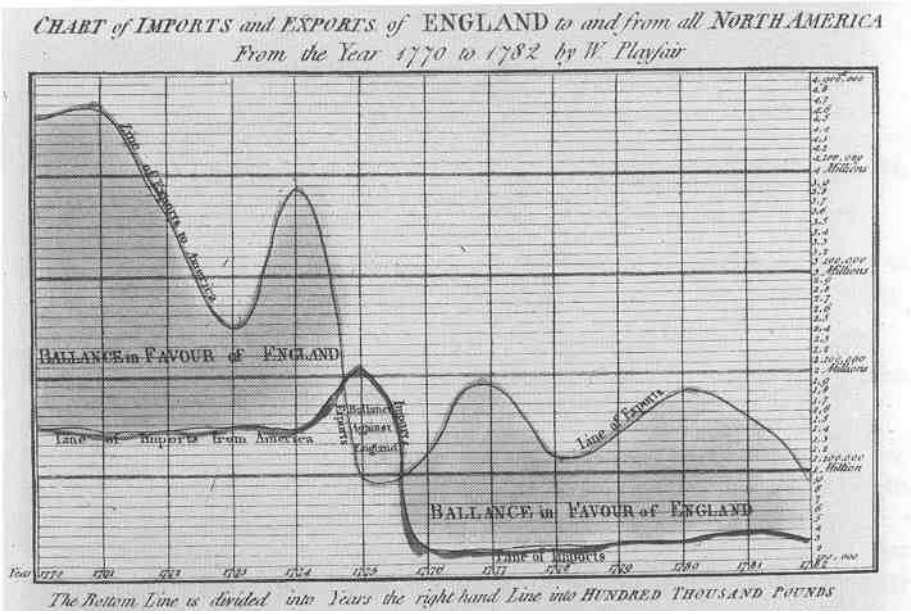
W. Playfair, C. J. Minard

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

- The use of the **computer made Visualization more practicable**

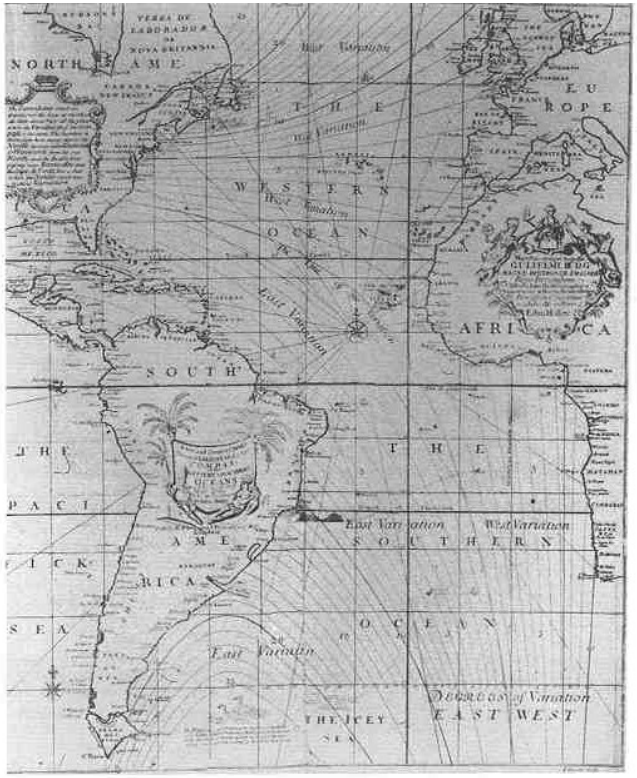
1987 - Identification of Visualization as an autonomous discipline

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)

Visualization in scientific discovery

Dr. John Snow

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

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**

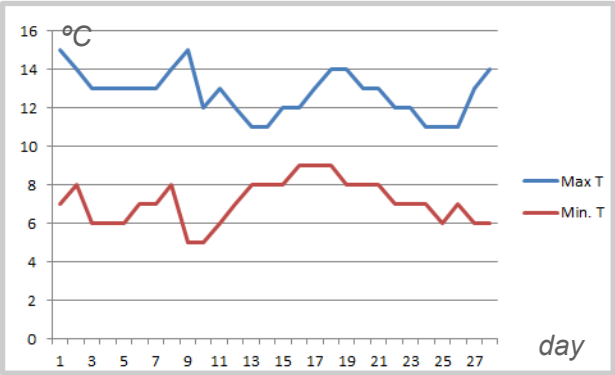
  (**questions** they ask)

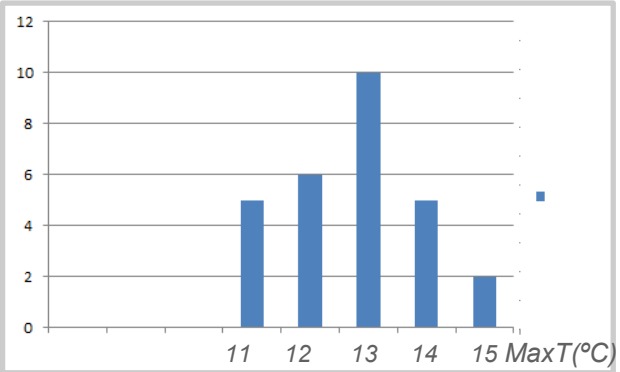
Example: how to select simple charts for a specific case? (the visual mapping problem)

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

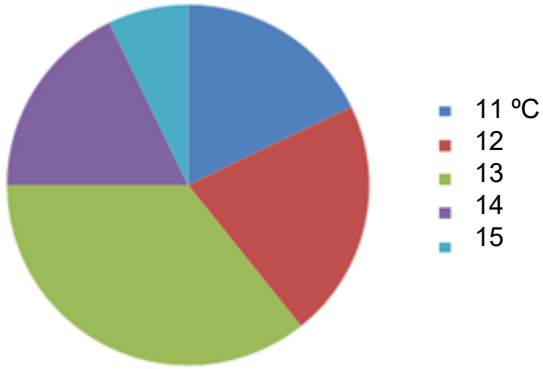
Q1- What were the maximum and minimum values of MaxT?
Q2- What was the most frequent Max temperature?
Q3- In how many days was that Max temperature attained?





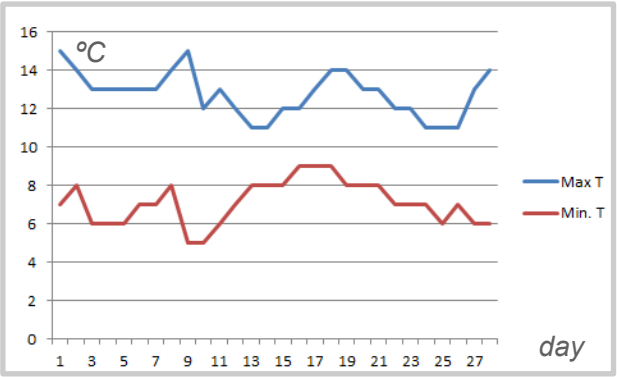
Which chart would you use to answer Q1?

Q2? And Q3?



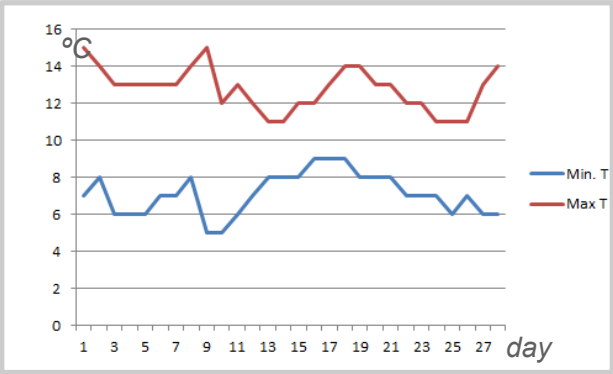
Temperatures along the month of February (in ºC):

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



Anything “odd” about this chart?

Would you prefer this one?



Don’t forget cultural issues!

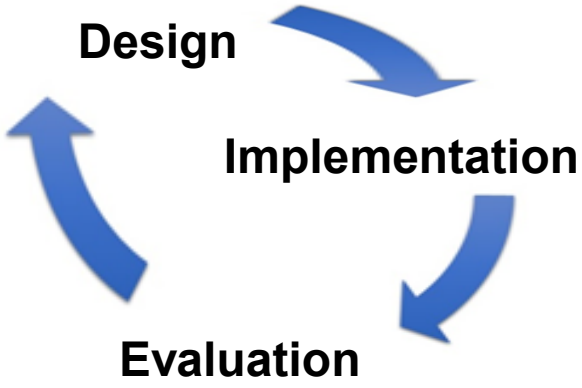
How can we produce a Visualization?

- There are **no "recipes"** to chose adequate Visualization techniques

- There are **principles** (derived form human perception and cognition)

  paradigms (examples resulting form past experience)

  and many **methods**

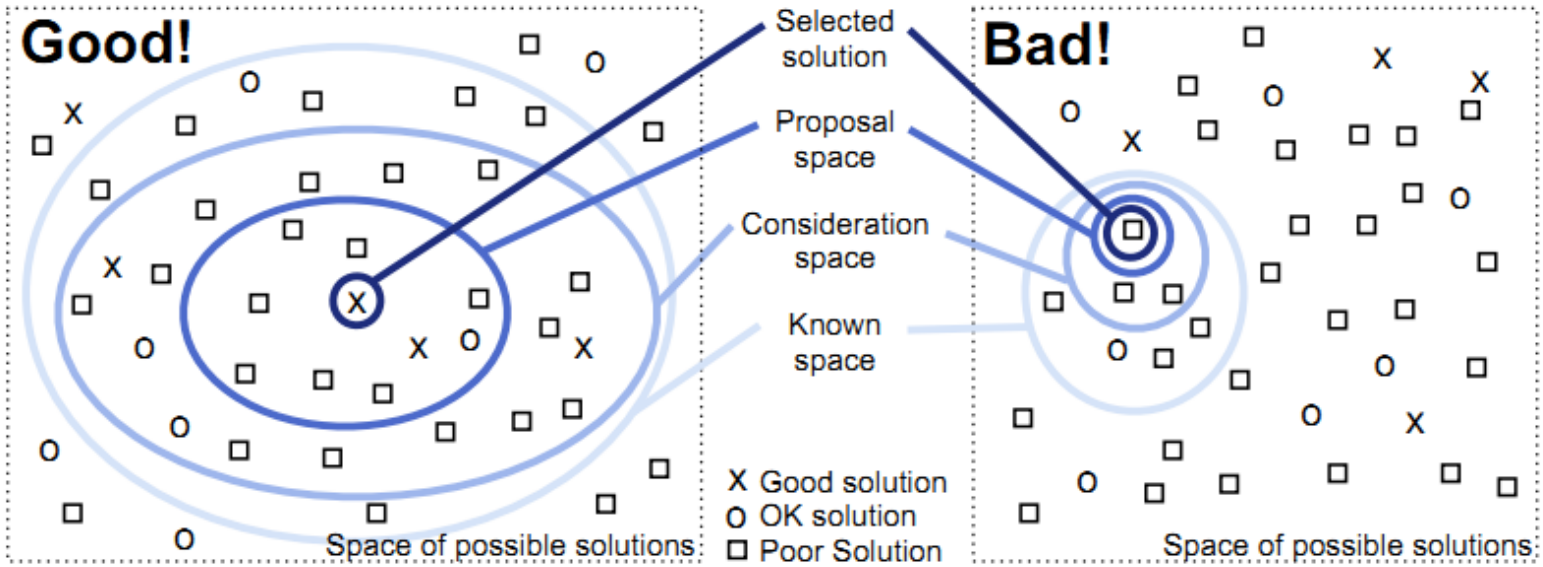


- To obtain **efficacy** it is fundamental:
  - a correct definition of **goal** and **user tasks**
  - apply adequate methods and **evaluate**

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

How to select an adequate visualization method for a specific case? (the problem of the design space of visualization idioms)

A search space metaphor for Visualization design:

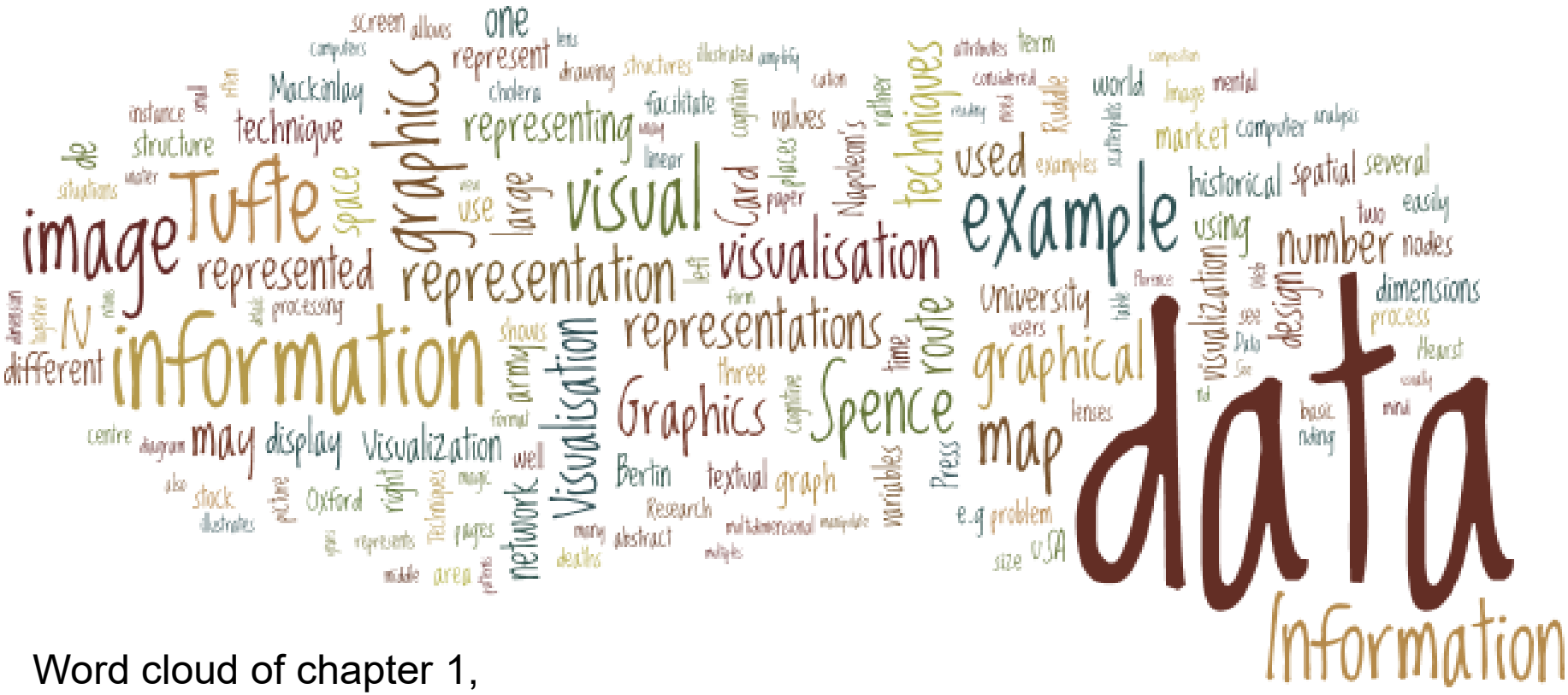


(Munzner, 2014)

- Only a few possibilities are reasonable; most are ineffective

Consider multiple alternatives and then select the best!

(based on evaluation …)



Word cloud of chapter 1,

Introduction to Information Visualization

(Mazza, 2004)

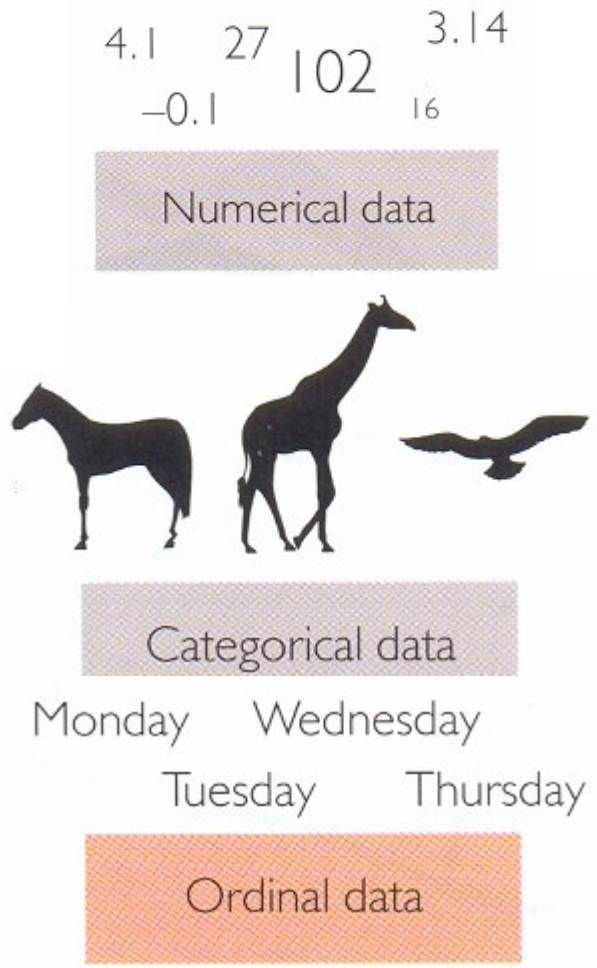
“Data” appears as the most “important” word!

Data characteristics

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

- Data/phenomenon nature:
  - Continuous
  - Discrete

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



(Spence, 2007)

Example: beyond the structure of the data to Visualize

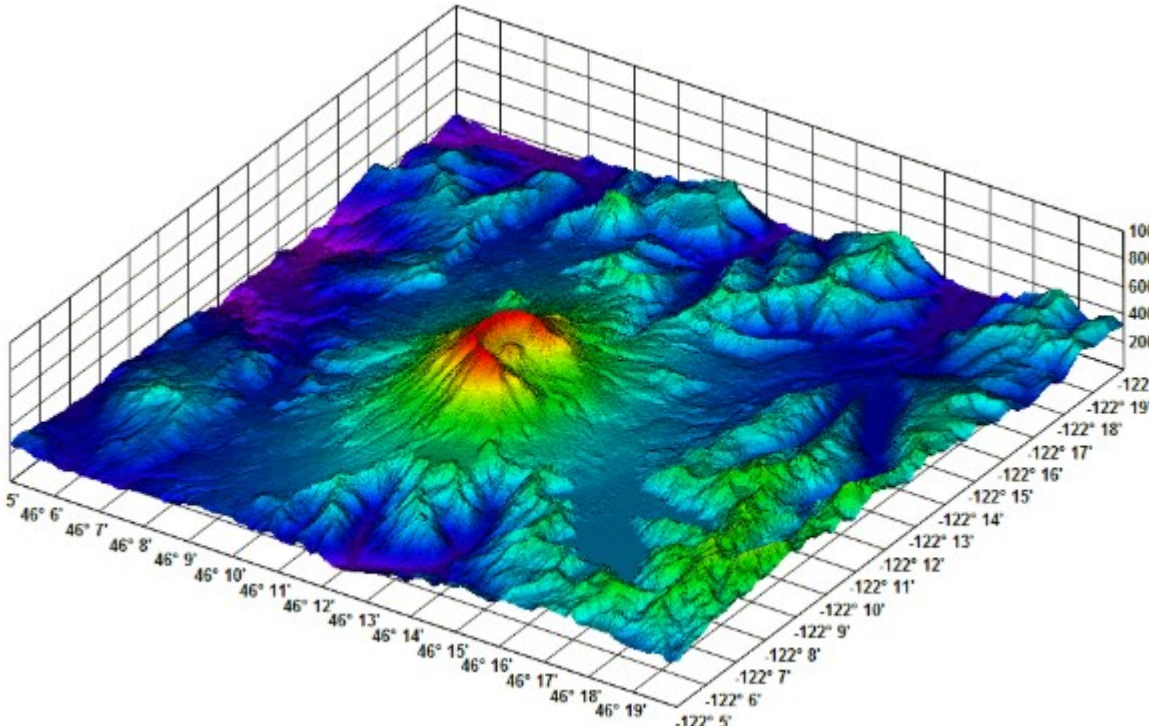
- Consider a data set with three columns:

  latitude *longitude* *d*

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

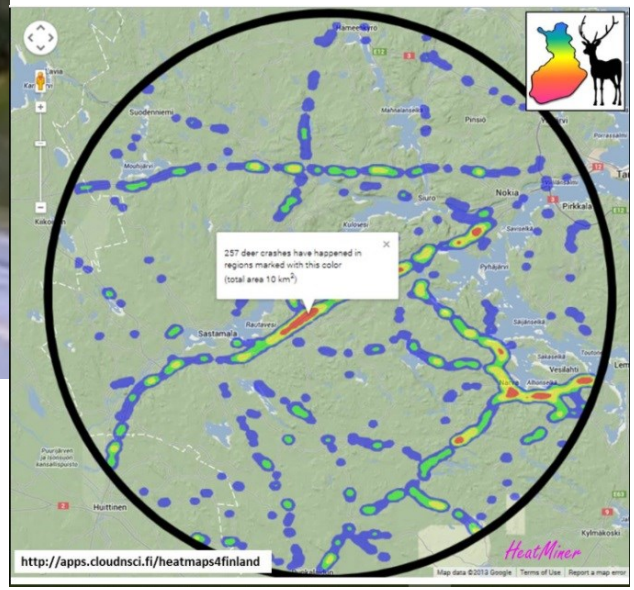
- If *d* is depth?

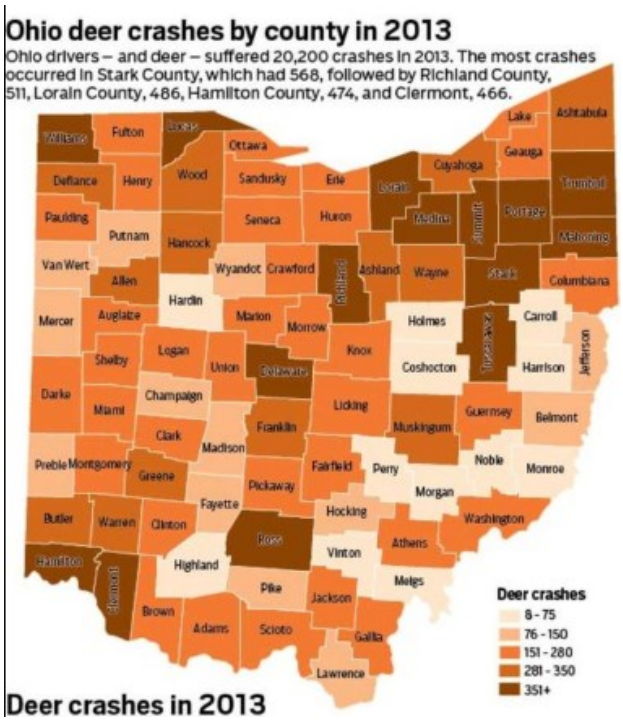
- the selected visualization technique may involve interpolation (e.g. isocontours, isosurfaces)



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

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



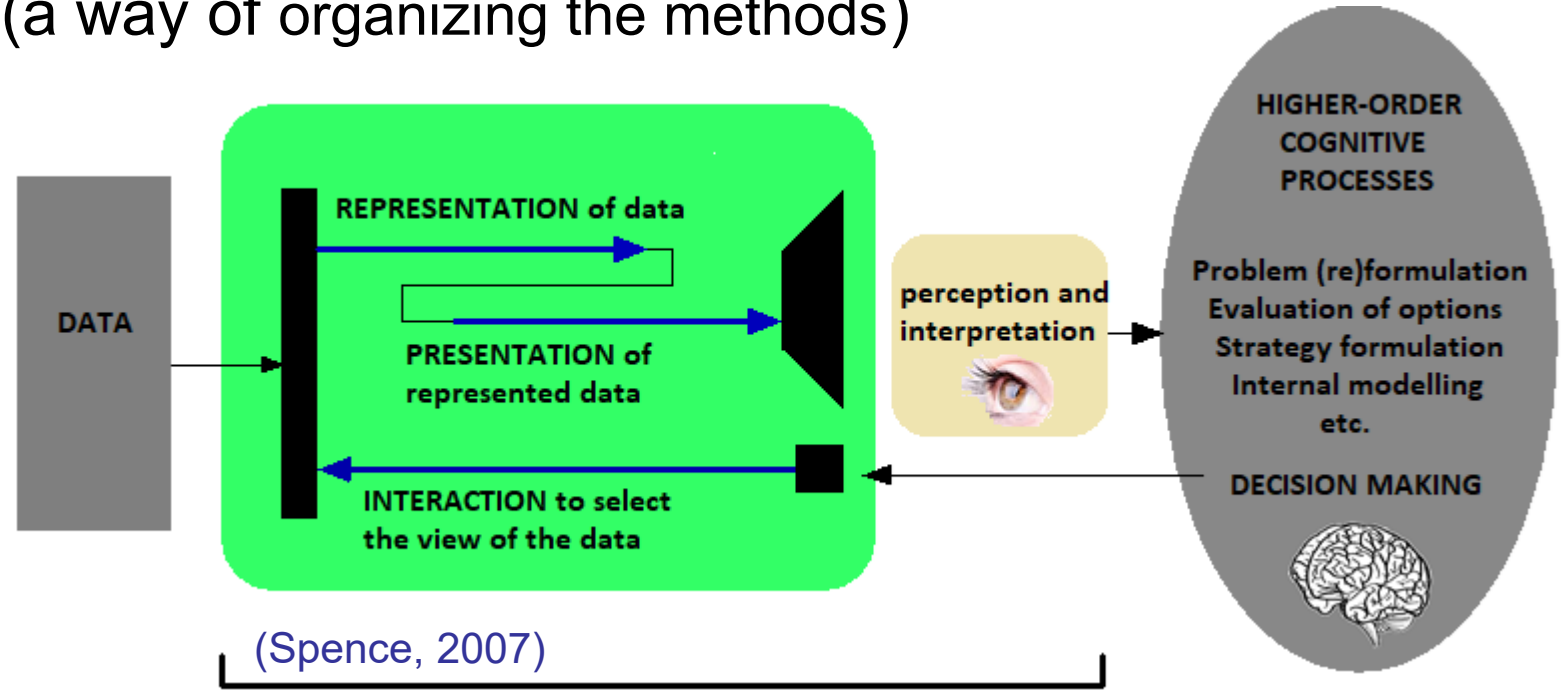


- Interpolation and contours don’t make sense!

  Know the data structure is not enough!

It is necessary to know the phenomenon behind the data

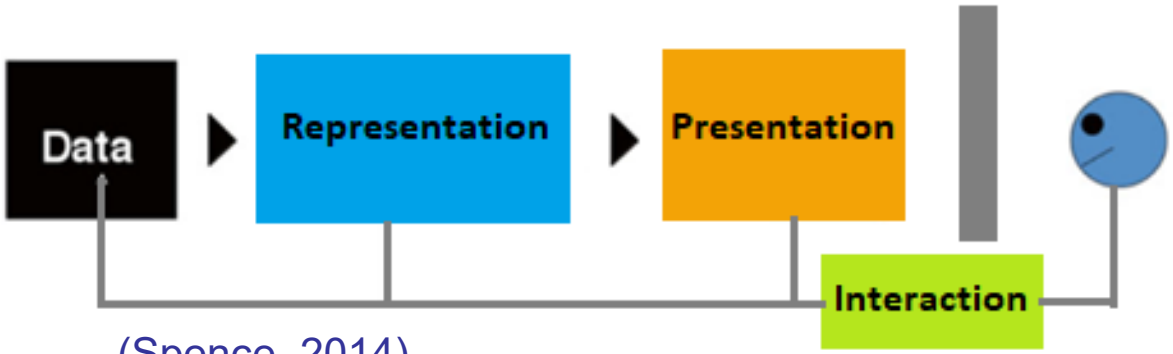
The process of visualization
(a way of organizing the methods)



(Spence, 2007)

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

- Representation
- Presentation
- Interaction



(Spence, 2014)

Representation techniques:

To increase the “consideration space” in the design process it is necessary to know a lot of representation techniques:

number of attributes

- univariate
- bivariate
- trivariate
- multivariate

techniques to represent value organized according the number of attributes

data structures

- linear
- temporal
- spatial or geographical
- hierarchical
- network

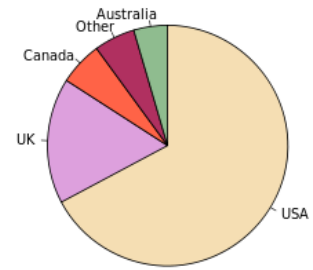
techniques to represent relation (trees and networks)
…

Presentation/interaction techniques:

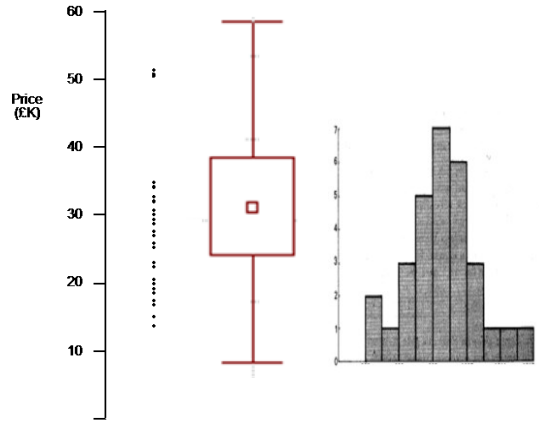
scroll
zoom
pan
suppression
distortion
…

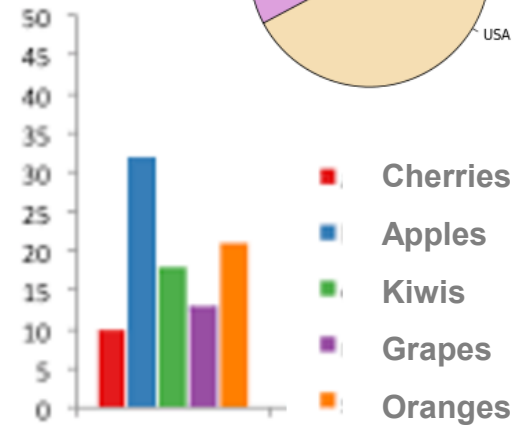
Techniques to present the representation selected (as the screen is limited)

Common Visualization techniques for univariate, bivariate and trivariate data

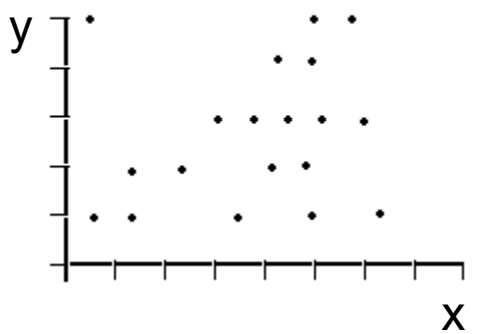


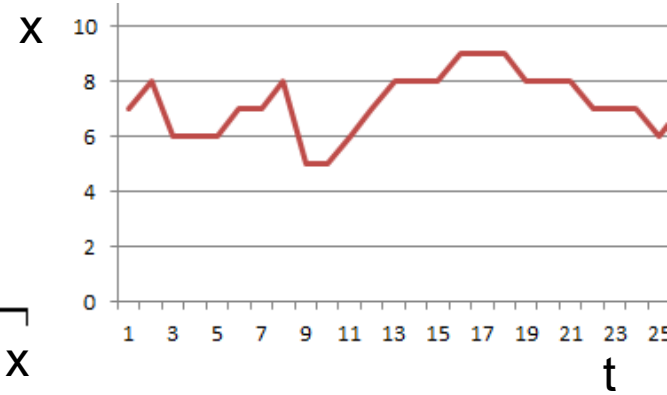
Univariate data
- dot plot
- box plot
- bar chart
- histogram
- pie chart
- …



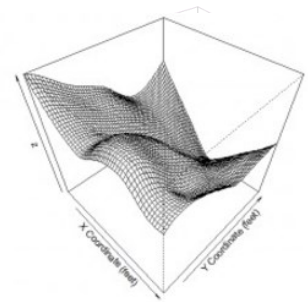


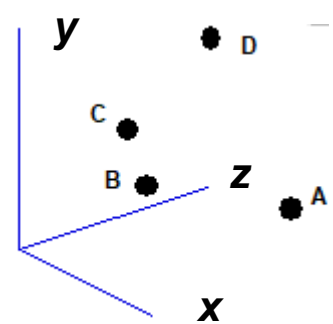
Bivariate data
- scatter plot
- line plot
- time series
- …

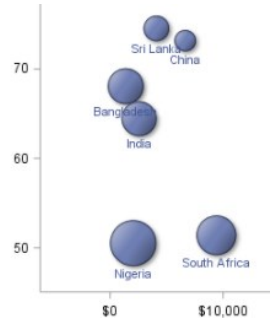




Trivariate data
- surface plot
- 3D representation
- bubble plot
- …

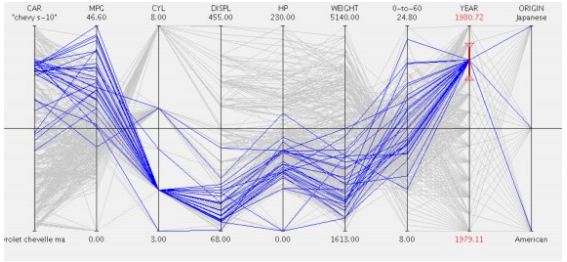




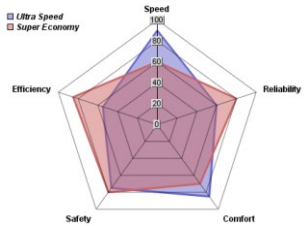


Techniques for Visualization of hypervariate data

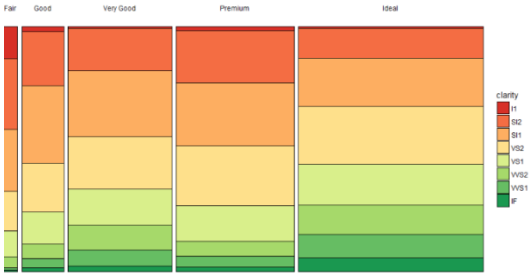
Coordinate plots parallel coordinate plots



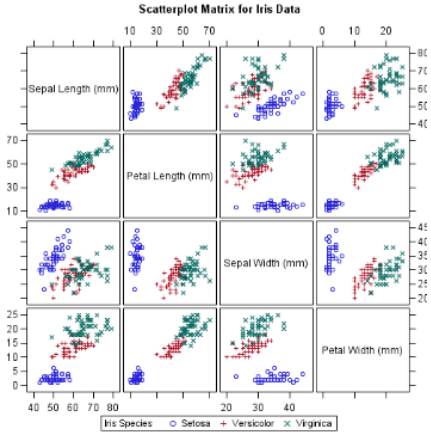
star plots



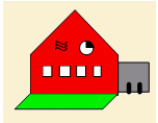
Mosaic Plots

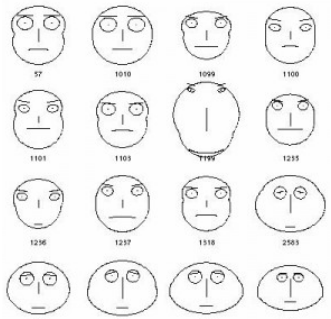


Scatterplot Matrix



Icons



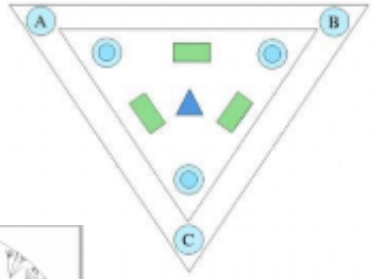


...

Techniques for Visualization of Relation

Lines

Venn diagrams



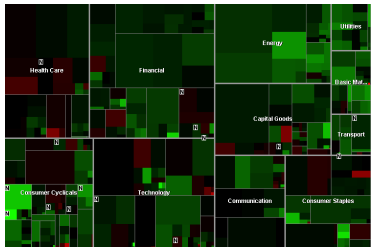
Maps and diagrams

InfoCrystal

- Trees

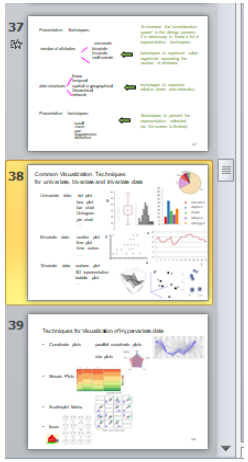
Hyperbolic browser

Treemap



…

Presentation and interaction techniques:



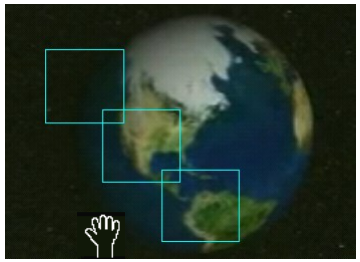
scroll

Scrolling is used when a document is larger than the display: a long document can be moved past a “window”

zoom

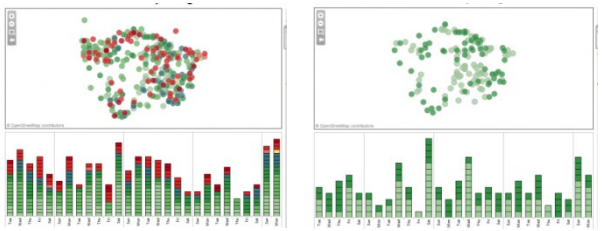
Zooming is the increasing magnification of a decreasing fraction of an image (or *vice versa*)

panning



Panning is the smooth movement of a viewing frame over a 2D image

suppression



Suppression helps focus on the important data at the moment

distortion

Distortion may help with the **focus + context** issue

…

- Visualization solutions currently are interactive

- Should be **usable** as any other interactive system

- Usability is, according to ISO 9241-11:

“the extent to which a product can be used by **specified users** to achieve **specified goals** with **effectiveness**, **efficiency** and **satisfaction** in a **specified context** of use”

- How to measure it??

Evaluation methods:

Some methods from Human-Computer Interaction have been adapted to evaluate **Visualization solutions**

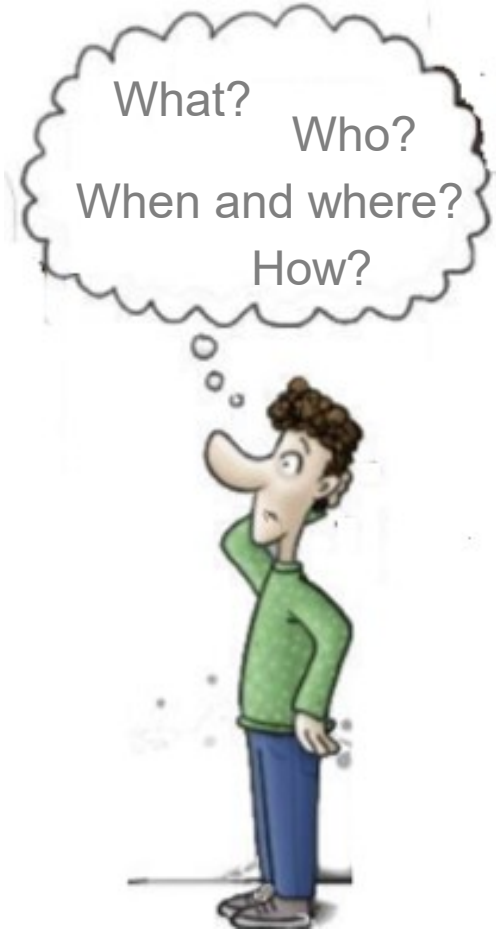
- **Analytical** (without users)
  - **Heuristic Evaluation**
  - Cognitive Walkthrough
  - Model based methods
  - Review methods
  - ...
- **Empirical** (involving users)
  - Observation
  - Query
  - Controlled Experiments
  - ...

usability tests

But there are methods specific to Visualization

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

- Visualization may be the solution (or part of it)
- There are **a lot** of visualization techniques
- Should be selected according to the **phenomenon**, **data, users, tasks, and context of use**
- And should **always be tested (with users)** before giving them to users (or use yourself!)



To probe further: Books

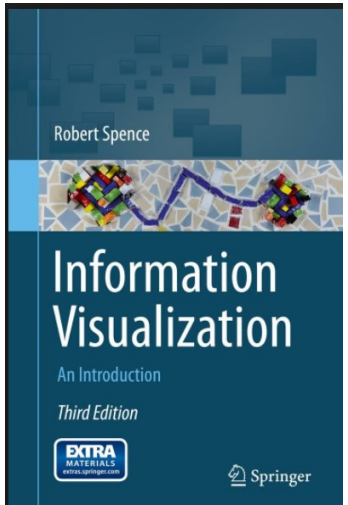
- Spence, R., *Information Visualization, An Introduction*, Springer, 2014

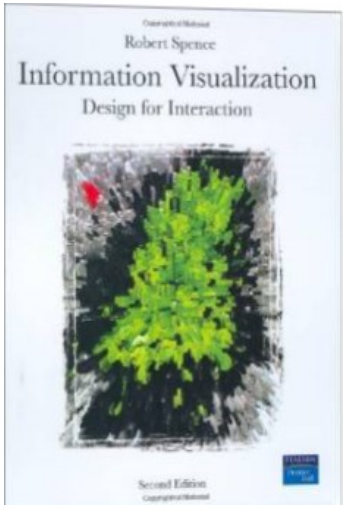
- Spence, R., *Information Visualization, Design for Interaction*, 2nd ed., Prentice Hall, 2007

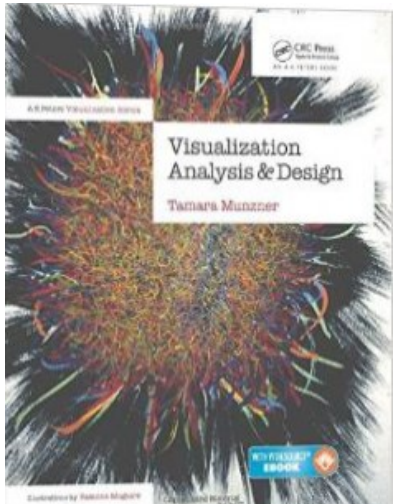
- Munzner, T., *Visualization Analysis and Design*, A K Peters/CRC Press, 2014

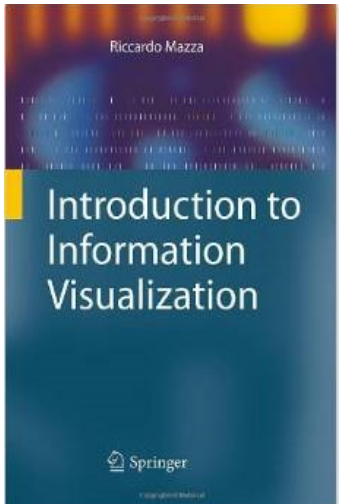
- Mazza, R., *Introduction to Information Visualization*, Springer, 2009

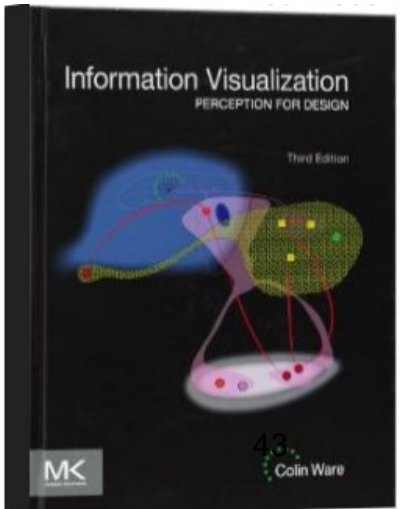
- Ware, C., *Information Visualization, Perception to Design*, 3nd ed.,Morgan Kaufmann, 2012











Scientific Journals/Conferences

- IEEE Transactions on Visualization and Computer Graphics
- IEEE Computer Graphics and Applications
- Computer Graphics Forum
- Computers and Graphics
- Information Visualization
- IEEE Vis (http://ieeevis.org/)
- Eurovis (http://eurovis2017.virvig.es/)
- Information Visualization (http://www.graphicslink.co.uk/IV2017/)

Videos

- Visualization for Machine Learning
  Fernanda Viégas · Martin Wattenberg
  https://nips.cc/Conferences/2018/Schedule?showEvent=10986

Other Bibliography

- Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983
- Tufte, E., *Envisioning Information,* Graphics Press, 1990
- Card, S., J. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think,* Morgan Kaufmann, 1999
- Bederson, B. , B. Shneiderman, *The Craft of Information Visualization: Readings and Reflections*, Morgan Kaufmann, 2003
- Few, S*.,* "Data Visualization for Human Perception". In: Soegaard, M. and Dam, R. (eds.). *The Encyclopedia of Human-Computer Interaction*, 2nd Ed. The Interaction Design Foundation https://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html
- Keim, D., Rossi, F., Seidl, T., Verleysen, M., & Wrobel, S. (2012). *Information Visualization, Visual Data Mining and Machine Learning* (Dagstuhl Seminar 12081). Dagstuhl Reports, 2(2), 58–83. http://doi.org/10.4230/DagRep.2.2.58
- Papers and sites …

Bibliography: papers

- Endert, A., Ribarsky, W., Turkay, C., Wong, B. L., Nabney, I. I.. Blanco, D. and Rossi, F., “The State of the Art in Integrating Machine Learning into Visual Analytics,” *Comput. Graph. Forum*, vol. 36, no. 8, pp. 458–486, 2017

- Heer, J., Bostock, M., & Ogievetsky, V, “A tour through the visualization zoo”. *Communications of the ACM*, *53*(6), 59, 2010

- Keim, D., “Information visualization and visual data mining,” *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 1–8, 2002

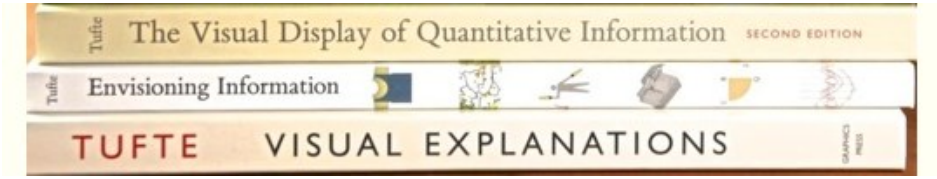
- Keim, D., Sips, M., and Ankerst, M. “Visual data-mining techniques1,” in *Visualization Handbook*, C. Johnson and C. Hansen, Eds., pp. 813–825, 2005

- Keim, D., Andrienko, G., Fekete, J., Carsten, G., Melan, G., “Visual Analytics : Definition , Process and Challenges” *Inf. Vis. - Human-Centered Issues Perspect*., Springer, pp. 154–175, 2008.

- Y. Lu, Y., Garcia, R., Hansen, B., Gleicher, M., and Maciejewski, R., “The State-of-the-Art in Predictive Visual Analytics,” *Comput. Graph. Forum*, vol. 36, no. 3, pp. 539–562, 2017.

Interesting links

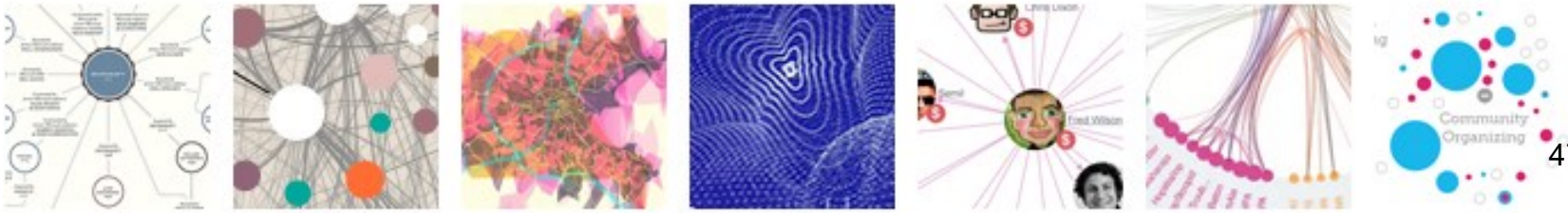
- http://www.infovis-wiki.net/
- https://eagereyes.org/
- https://www.edwardtufte.com/tufte/



- http://albertocairo.com/
- https://medium.com/multiple-views-visualization-research-explained

MULTIPLE VIEWS Visualization Research Explained

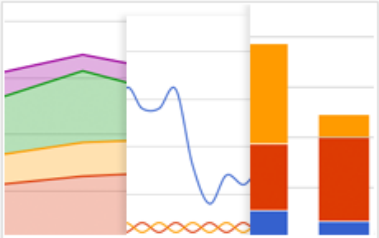
- http://www.visualcomplexity.com/vc/



Visualization Tools

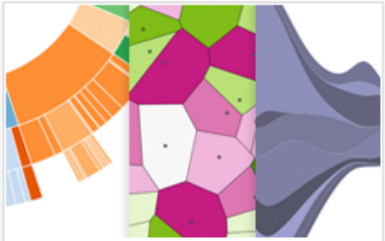
- There are a lot, of different types and with different purposes
  from very simple to very complex … free to very expensive…

- Some interesting tools for different types of users and uses



Google Chart Tools

A collection of simple to use, customizable and free to use interactive charts and data tools.



D3.js

A small, flexible and efficient library to create and manipulate interactive documents based on data.

R

A software environment for statistical computing and graphical techniques.

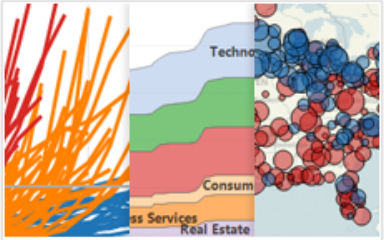


Tableau Public

A desktop application to build and post interactive graphs, dashboards, maps and tables to the web.

http://selection.datavisualization.ch/
https://www.springboard.com/blog/31-free-data-visualization-tools

2019 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms

Future trends in Visualization?

- IMO visualization should/may become more:
  - Accessible, Useful, Usual and Usable
  - Intelligent (+AI)
  - Interactive
  - Multimodal (gestures, haptic…)
  - Mobile
  - Qualitative
  - Situated (+AR)
  - Wearable
  - Etc.