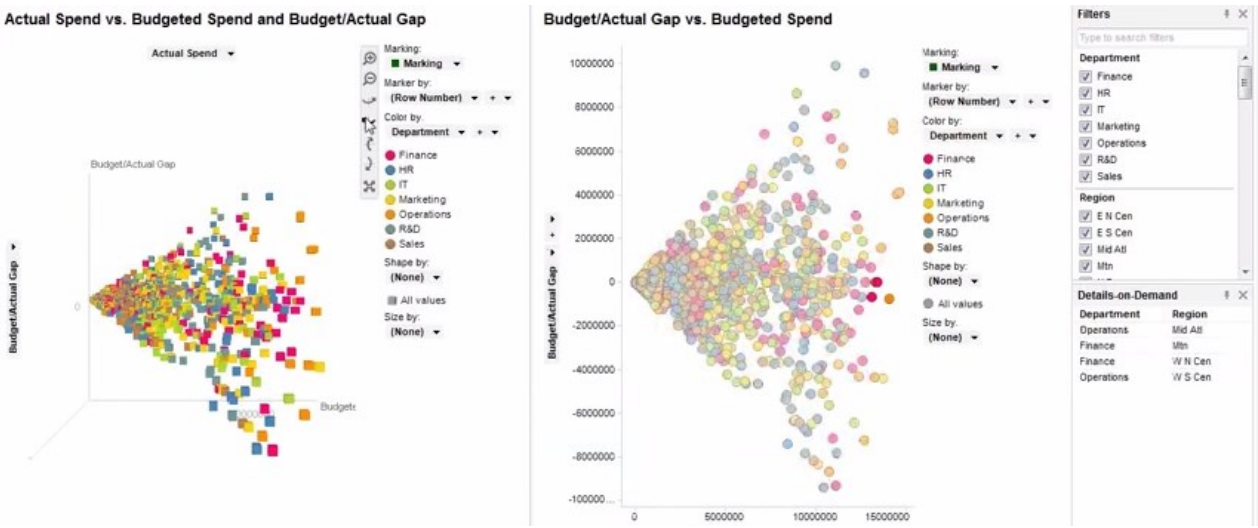
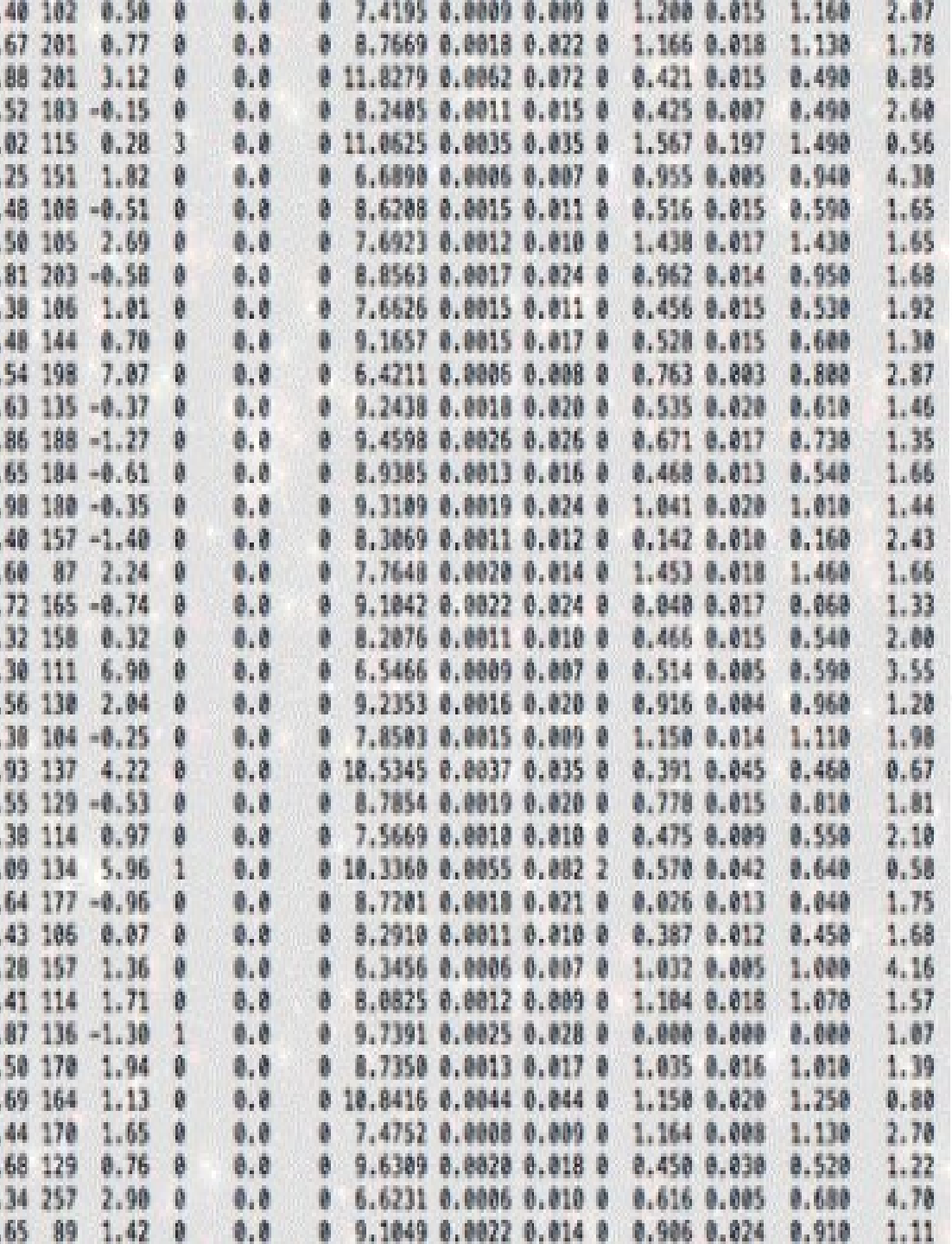
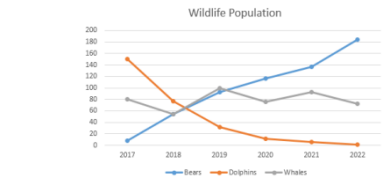
Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

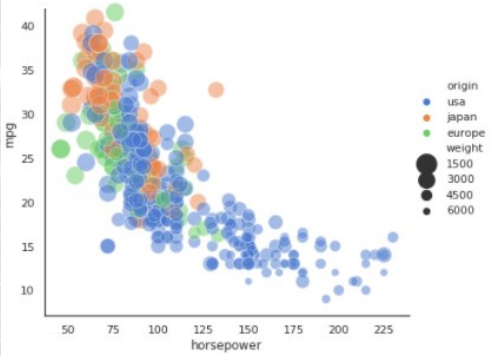
Data



Beatriz Sousa Santos, University of Aveiro, 2022





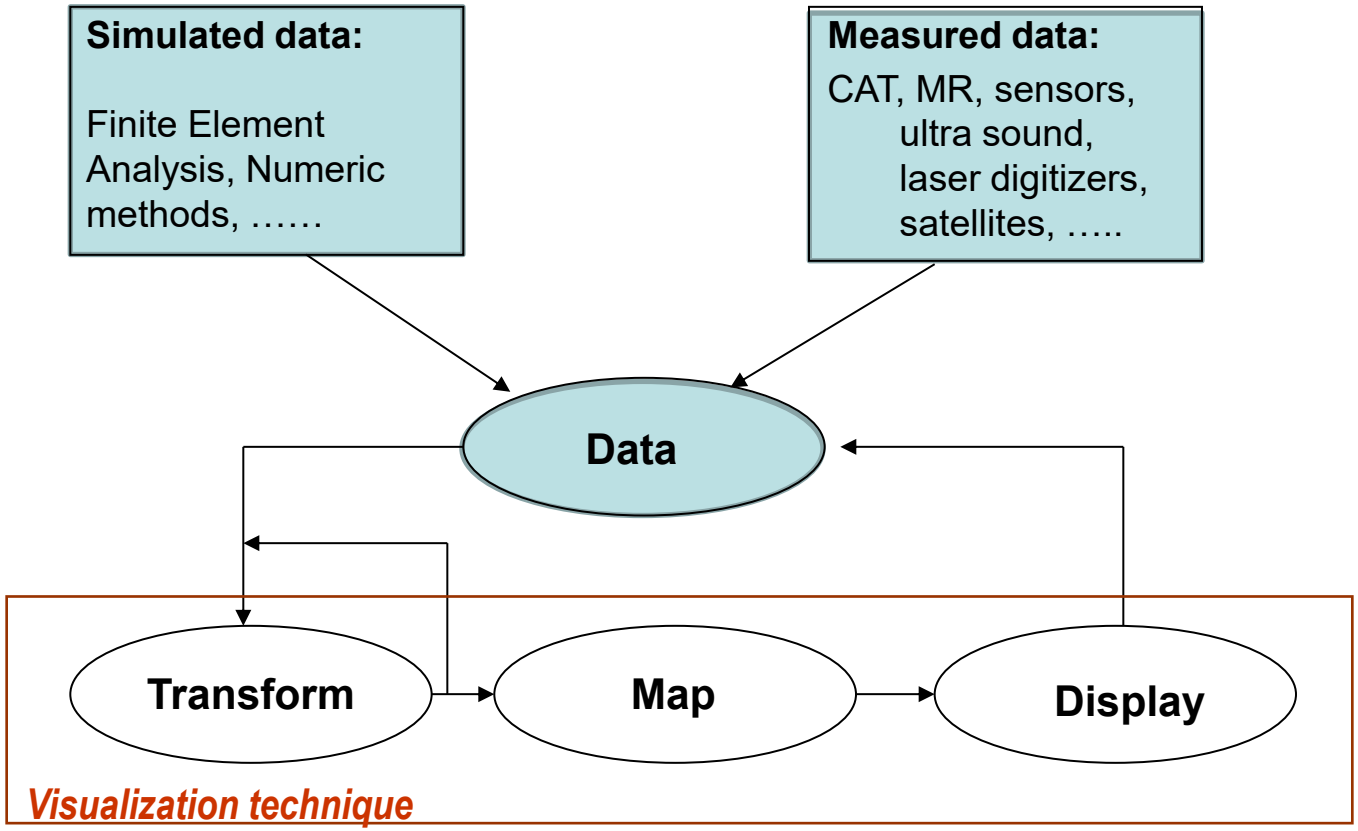


Data visualization?

Machine learning?

Statistics?

Scientific Visualization reference model



(adapted from Schroeder et al., 2006)

Adequate data pre-processing is vital!

Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Data Characteristics

- Data may have a lot of different forms and there are many techniques and systems to visualize them

- A data classification is important to:

  - predict what visualization techniques are adequate

  - make easier the communication about the data

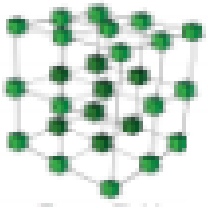
  - allow a more systematic approach to Visualization

  ….

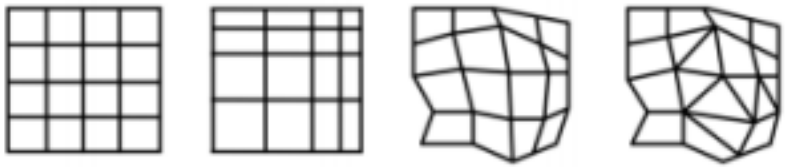
Data Abstraction

name	rank	gender	year
Jacob	1	boy	2010
Isabella	1	girl	2010
Ethan	2	boy	2010
Sophia	2	girl	2010
Michael	3	boy	2010

- Four basic dataset types:
  - Tables
  - Networks
  - Fields
  - Geometry



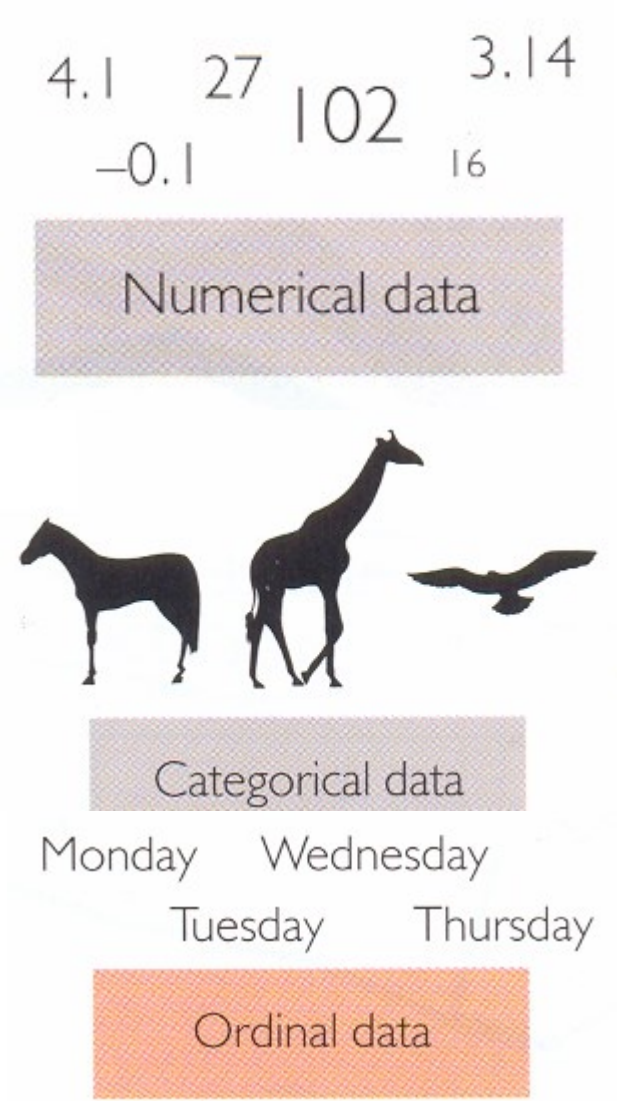
- Five basic datatypes
  - Items
  - Attributes
    - Categorical
    - Ordered
      - Ordinal
      - Quantitative
  - Links
  - Positions
  - Grids



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

- Data nature:
  - Continuous
  - Discrete

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



(Spence, 2007)

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

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

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

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

  latitude *longitude* *d*

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

- If *d* is depth or altitude?

the selected visualization technique may involve interpolation

(e.g. isocontours, isosurfaces, 3D surface)

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

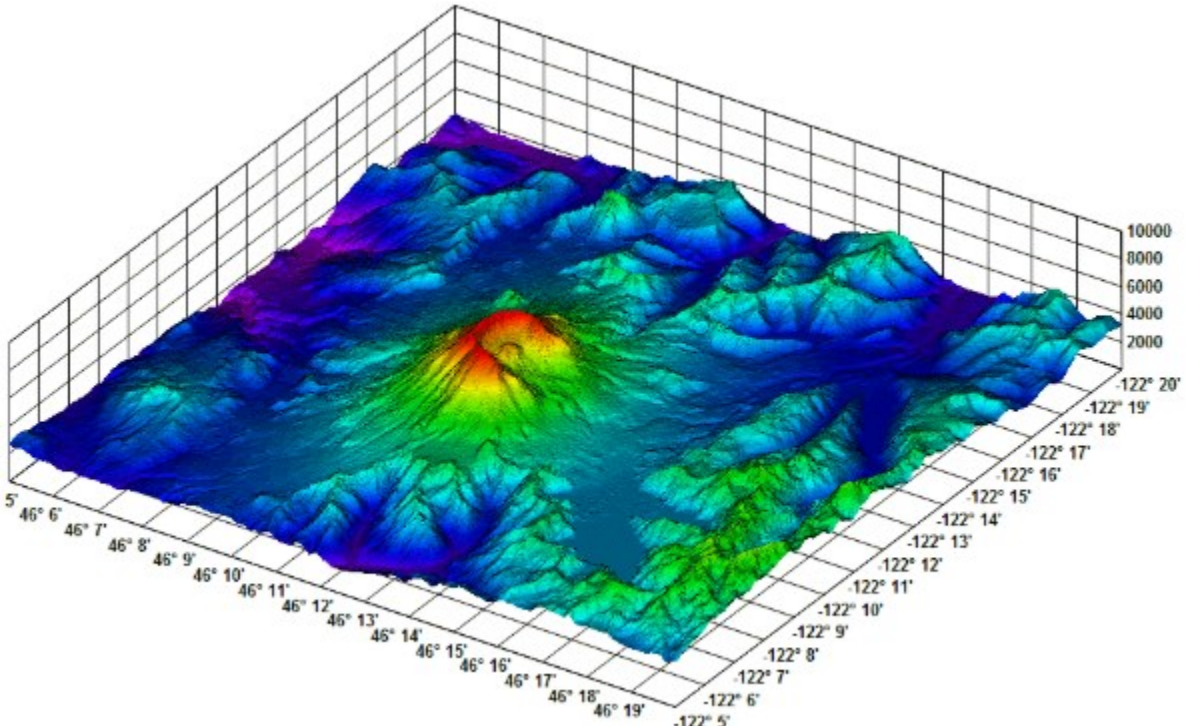
latitude *longitude* *d*

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

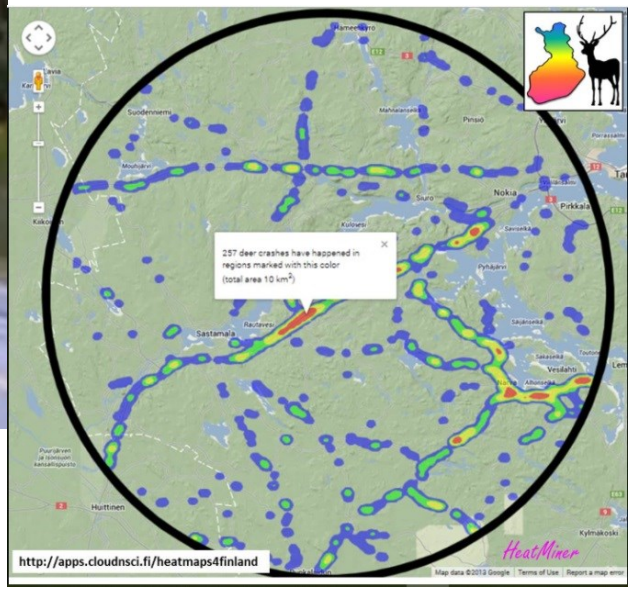
- If *d* is depth or altitude?

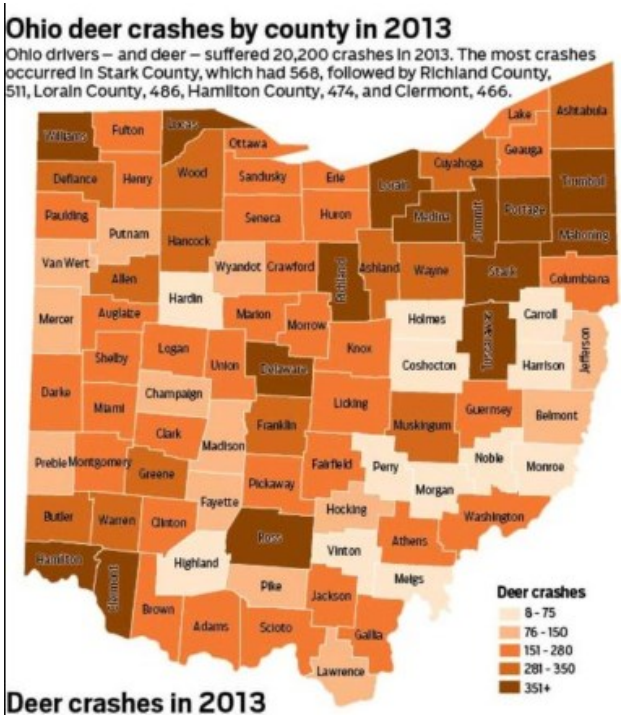
the selected visualization technique may involve interpolation

(e.g. isocontours, isosurfaces, 3D surface)



- 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**

Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Data preparation

- Data preparation is very important and very time consuming

- Several phases and terms:
  - Data pre-processing
  - Data wrangling
  - Data cleaning, Data tiding …
  - Data transformation

Data integrity becomes more essential when the volume of data increases

“Brilliant visualizations cannot redeem bad data!”

Or

“Garbage in garbage out …”

Cleansing Data

- Data is dirty: it contains typos, inconsistencies, fails in some way to meet a standard…

Transforming Data

(at the variable level)

- Encoding
- Aggregation
- Derived data
- Removal
- Standardization

Examples:

Cleansing Data

Birth date: Feb/30/2000

Temperature: -300 °K

City: Lixboa

Transforming Data

- Encoding – answers to an open question need to be parsed and coded
- Aggregation – detail may be excessive (age: <18; 18-40; 41-65; >65)
- Derived data – add new relevant variables ($T_{range} = T_{max} - T_{min}$)
- Removal – remove data that are not needed
- Standardization – M/F; °C or °F

Main bibliography

- Camões, J., *Data at Work : Best practices for creating effective charts and information graphics in Microsoft Excel*, Pearson Education, 2016 (Chap.4)
  https://learning.oreilly.com/library/view/data-at-work/9780134268798/title.html

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

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

- Kirk, A., *Data Visualization : a successful design process*. Packt Publishing., 2012 (Chap.3)
  https://books.google.pt/books/about/Data_Visualization.html?id=I4qBVLfD3t4C&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false