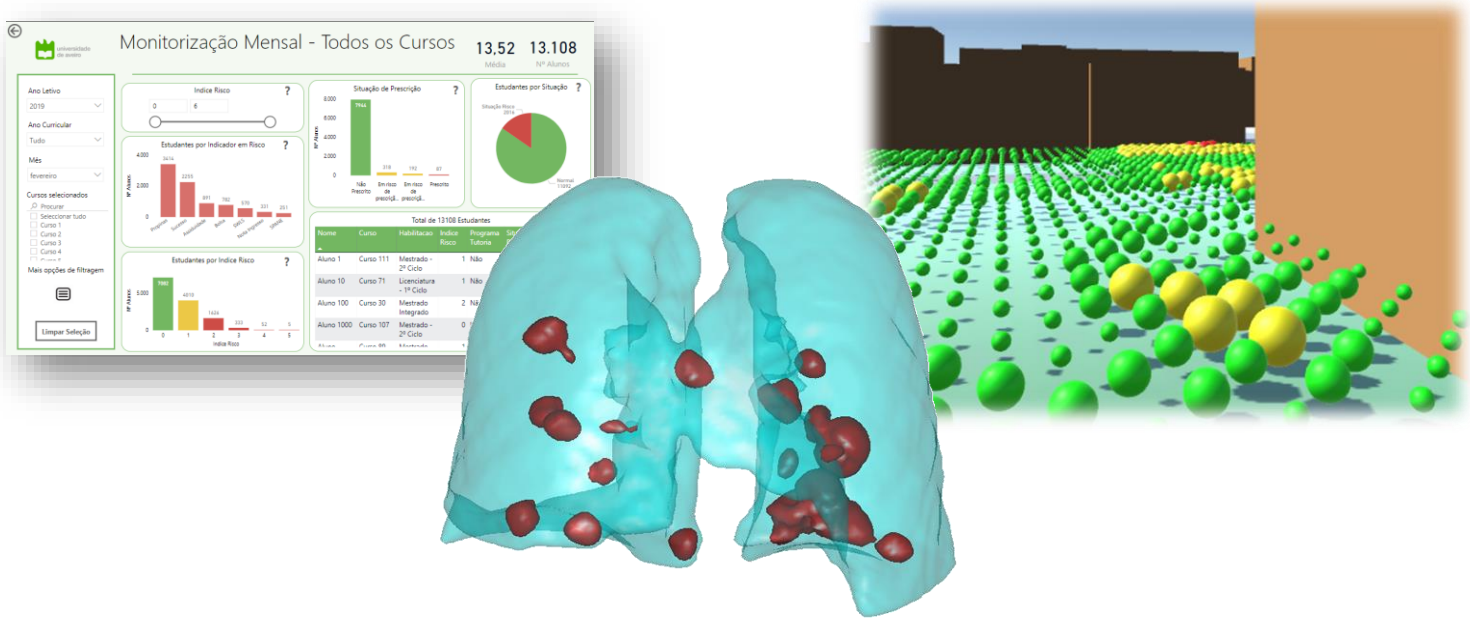
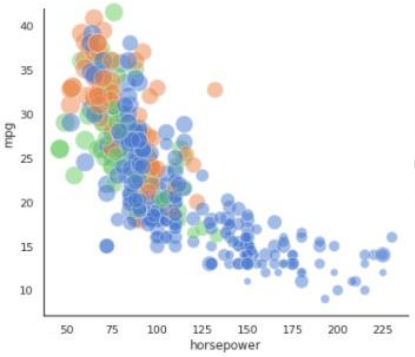
Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

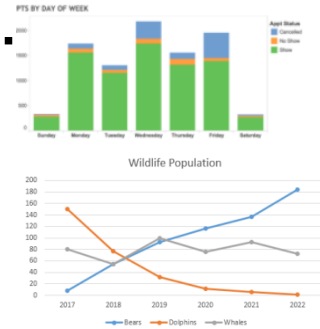
Data Visualization module 2023
Introduction



Beatriz Sousa Santos, Joaquim Madeira 2022/23

The problem…





Data visualization?

Machine learning?

Statistics/ Data analysis?

What is Data Visualization?

- Visualization is focused on how to **visually represent and explore large amounts of data**
- Taking advantage of the **human visual system capacities**
- Providing "**insights**" concerning the phenomenon behind the data

What it **is not**:

- just "pretty pictures"!

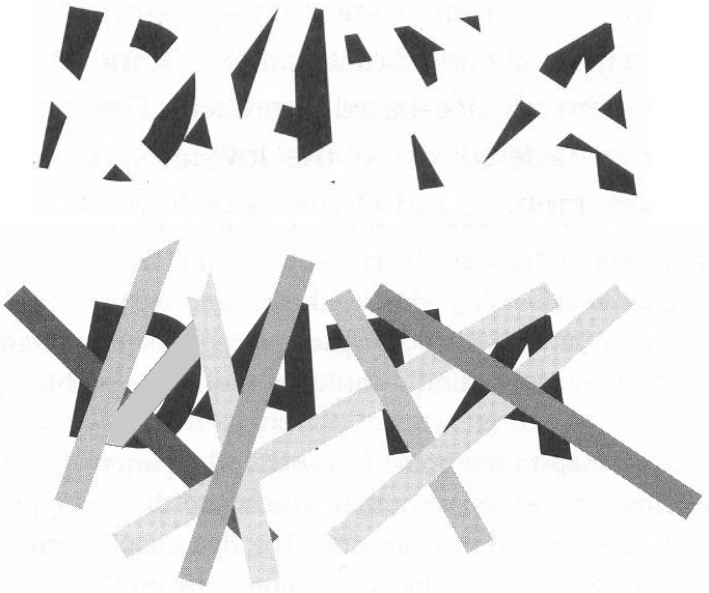
Why and how to represent data visually?

- The human visual system is a most powerful pattern seeker

  “seeing is understanding…”

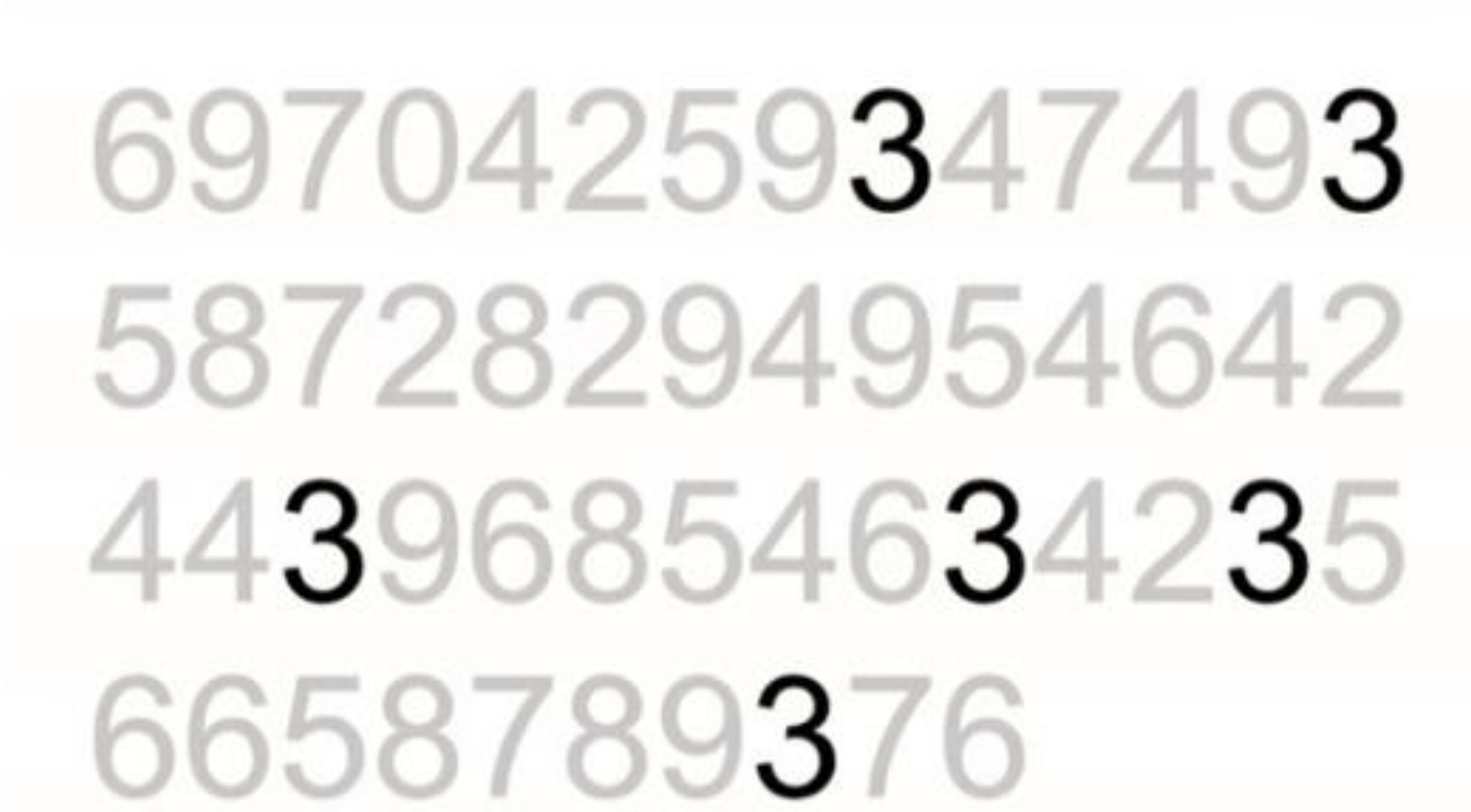
- We easily see patterns displayed in certain ways

  but not in others …



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

(Nussbaumer Knaflic, 2015)



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

Why represent data visually?

Vis helps in situations where seeing the dataset structure in detail is better than seeing only a brief summary of it.

(Munzner, 2014)

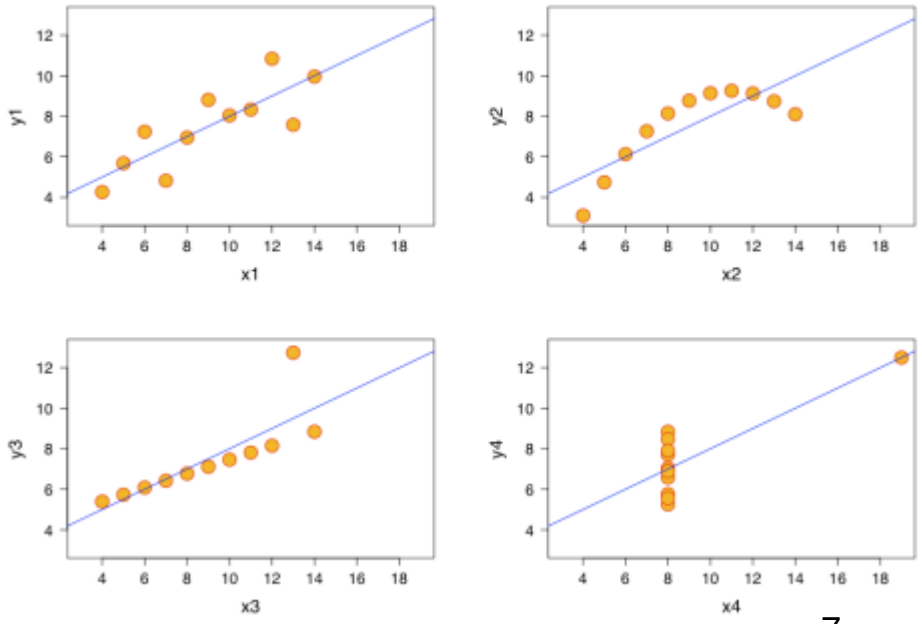
Anscombe's Quartet: Raw Data

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
var.	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
corr.	0.816		0.816		0.816		0.816	

https://www.sjsu.edu/faculty/gerstman/StatPrimer/anscombe1973.pdf

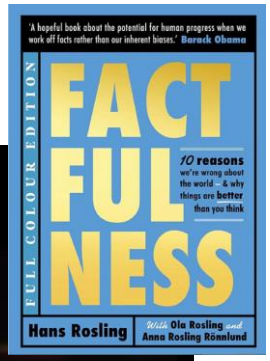
Ascombe quartet: data sets with the same simple statistical properties

(Tufte, 1983)



Example of Presentation based on a simple Visualization: World health by Hans Rosling

200 years of health/income – 120 000 values in 4 min



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

https://books.google.pt/books/about/Factfulness

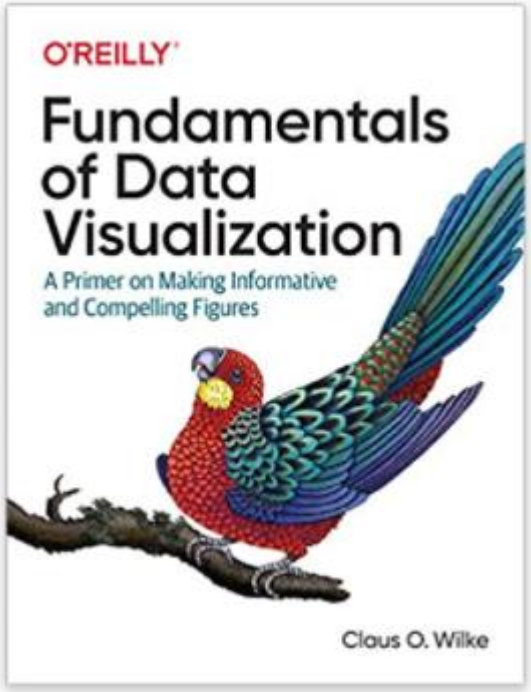
This module:

- Objective, bibliography and tools
- Introduction to Data Visualization
- Data and phenomena: types and pre-processing
- Human-in-the-loop process: perceptual and cognitive aspects
- Visualization of quantitative data:
  - 1D and 2D Data – main methods; exercises
  - 3D and nD data - main methods; exercises
- Visualization of other types of data
  - networks, hierarchical data, etc.
- Effective Visualization

Bibliography

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

- Wilke, C., *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*, O'Reilly, 2019
  https://learning.oreilly.com/library/view/fundamentals-of-data/9781492031079/

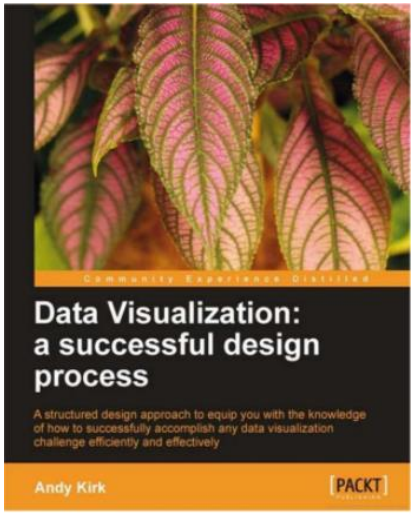


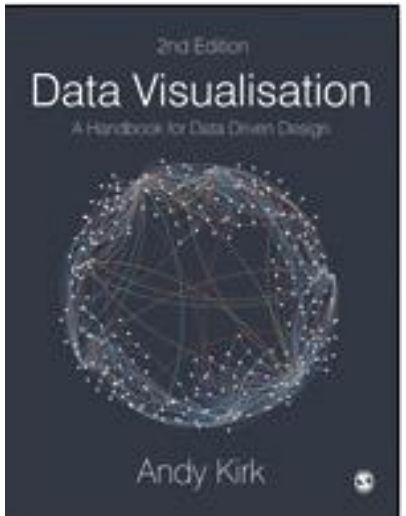
More advanced bibliography

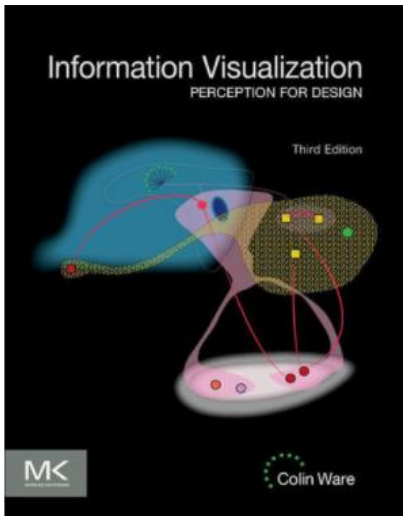
- Kirk, A., Data Visualisation A Handbook for Data Driven Design, 2nd. Ed., Sage, 2019
- Ware, C., Information Visualization, Perception for Design, 3rd ed., Morgan Kaufman, 2012
- Munzner, T., *Visualization Analysis and Design*, A K Peters/CRC Press, 2014
- Kirk, A., *Data Visualization: A successful design process*, Pack Publishing, 2012

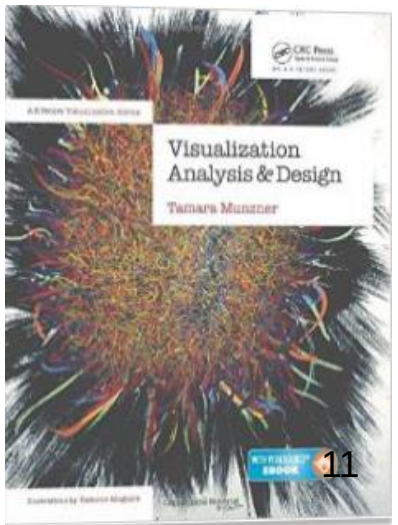
Explore other books available at the playlist:

https://learning.oreilly.com/playlists/74bfec5e-4346-48ff-82b4-657fda6922b6









Interesting links

- https://excelcharts.com/author/jorge-camoes/

- https://eagereyes.org/

eagereyes

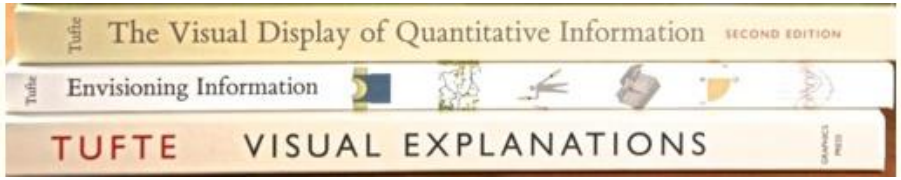
- http://www.perceptualedge.com/

perceptual edge

Visual Business Intelligence
for enlightening analysis and communication

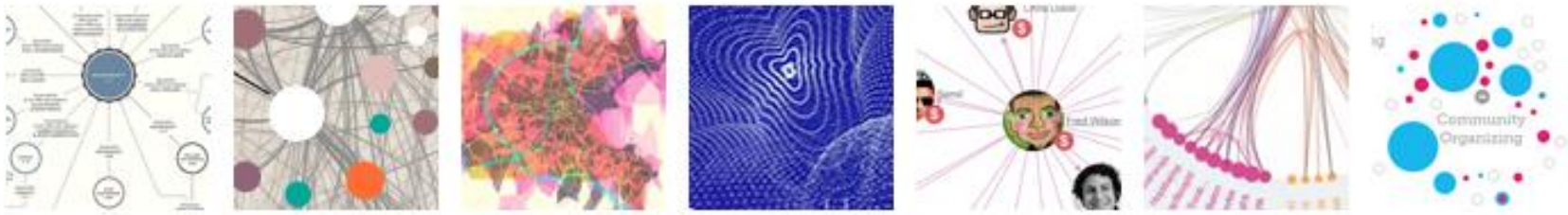
- http://www.thefunctionalart.com/

- https://www.edwardtufte.com/tufte



Interesting links

- https://medium.com/multiple-views-visualization-research-explained
- http://seeingdata.org/
- https://flowingdata.com/about
- http://www.visualcomplexity.com/vc/



Moodle

Slides

- Introduction to the module
- Introduction to Visualization
- Data
- Creating a Visualization
- Effective Visualization

Examples and Practical assignments

- Assignment - How Visualization was used in a specific research/development work
- Example: Using Visualization in the development of Multilayer radiating structures for mm Waves
- Template for the assignment - presenting how Visualization was used

Bibliography

- Camões, J., Data at Work : Best practices for creating effective charts and information graphics in Microsoft Excel, Pearson Education, 2016
- Wilke, C., Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures, O'Reilly, 2019
- Healy, K., Data Visualization a Practical Introduction, Princeton, 2019
- Kirk, A., Data Visualization: A successful design process, Pack Publishing, 2012
- Spence, R., Design for Interaction (2nd Edition), Pearson, 2007
- More Advanced Bibliography- playlist

Assessment

During the semester:
Test – 60%
Practical Assignment - 30%
Attendance and participation – 10%

Alternative at the end of semester:
Exam – 90%
Attendance and participation – 10%

Test – last session of the module – June, 5th

Assignment submission (via Moodle) – May, 28th

Assignment presentation and discussion – May, 29th

Practical activities

Next lecture bring your own laptop!

Assignment (team work)

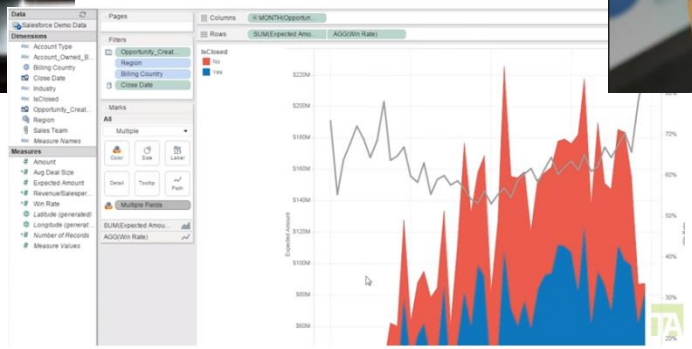
- Prepare and present an example illustrating the usage of visualization in a specific engineering field
(e.g. chemical, computer, electrical, mechanical, etc.)

or

- Select and present a Visualization paper

Questions regarding the module?

Visualization S/W



Which Visualization Tool?
It depends …

- Exploratory (discover pattern, multiple views)

  or Explanatory (View of the data presenting discovered highlights)?
- Type of data (Maps, Charts, Data,…)
- Developer or non-developer?
- Scientific or information Visualization (2D,3D, structured or not?)
- Interactive or Static?
- Web or local?
- Easy to use or Flexible?
- Protection of data?…

Visualization Tools – Some possible choices

- If you want to produce a few simple charts for a report or paper: Excel, …
- If you want to produce some charts and have some programming skills: MATLAB, R, python libraries, …
- If you use some statistics/analytics S/W: Statistica, SPSS, …
- If you are in a large company: Tableau or Qlickview may be adequate (very powerful and expensive business intelligence S/W)
- If you just want to make a few simple charts for your web page and have programming skills: google charts
- If you want to develop an interactive visualization Web application to visually explore data: D3.js

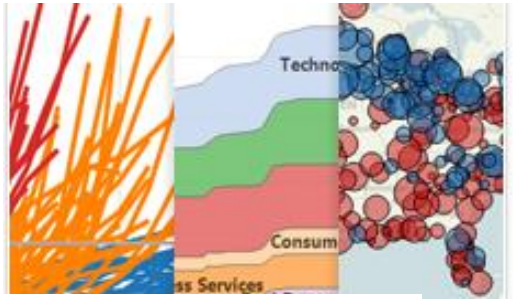
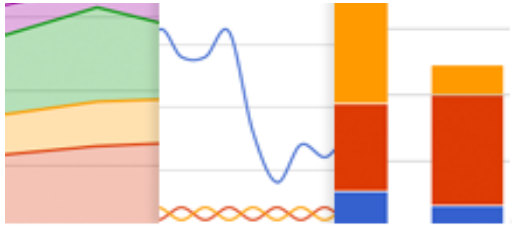


Tableau Public

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



Google Chart Tools

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

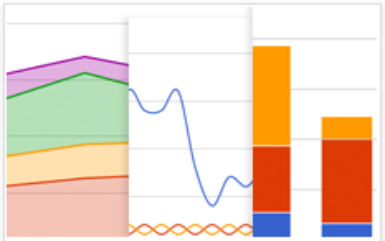
D3.js

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

Visualization S/W

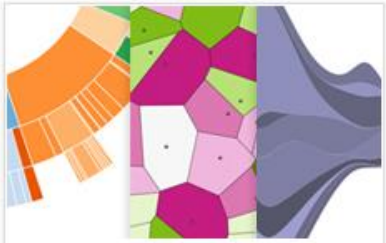
- There are a lot of tools, of different types and with different purposes from very simple to very complex …

- Some interesting and widely used tools:



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.

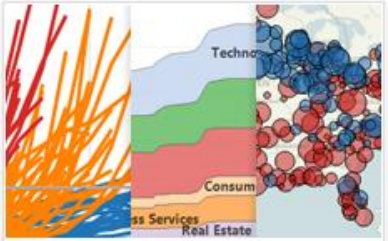
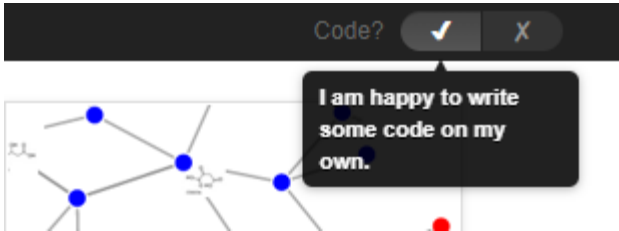


Tableau Public

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

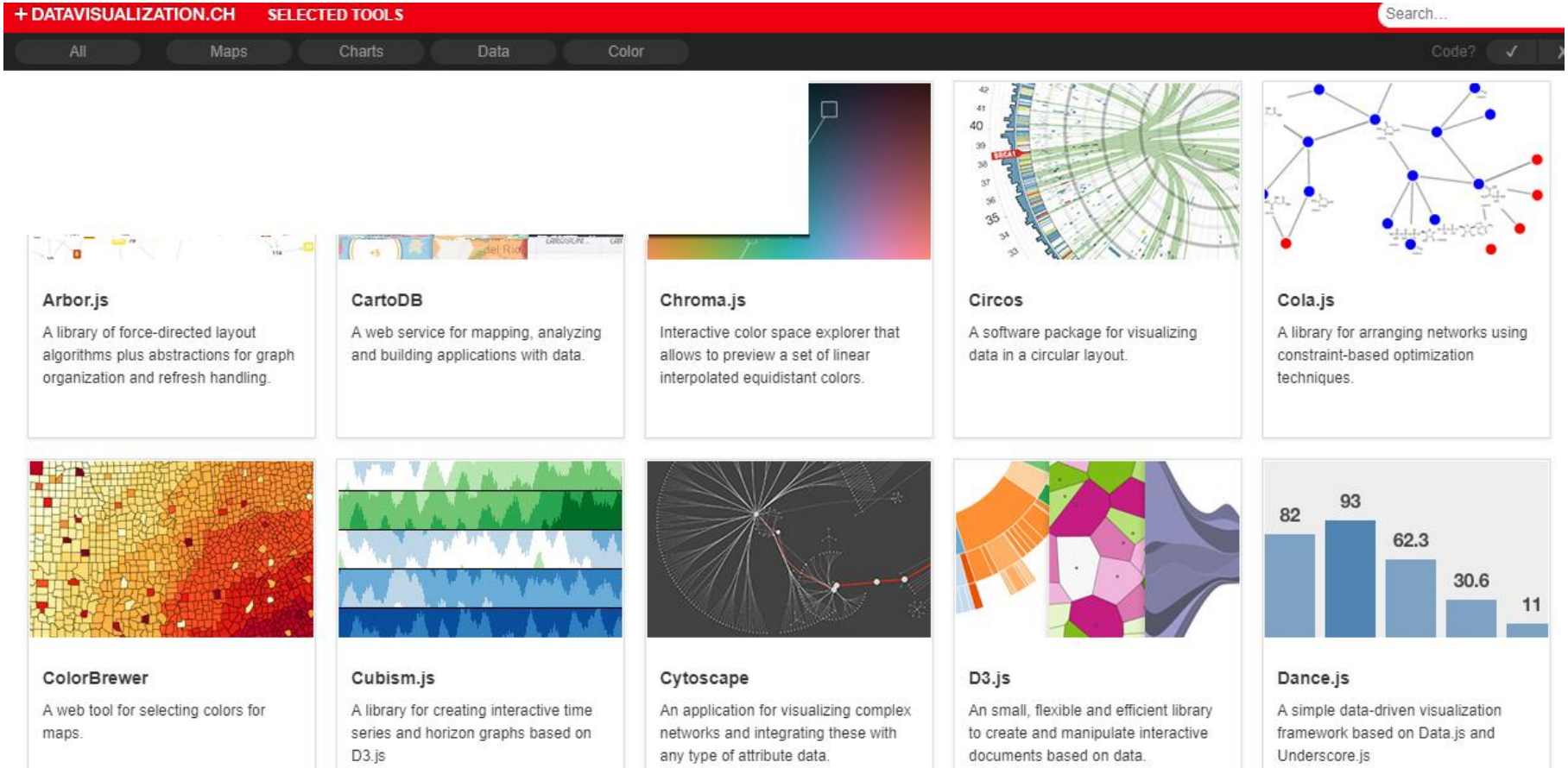
see e.g.
https://www.kdnuggets.com/2020/05/top-10-data-visualization-tools-every-data-scientist.html
https://www.kdnuggets.com/2019/04/7-qualities-big-data-visualization-tools.html
http://selection.datavisualization.ch/



Visualization Tools

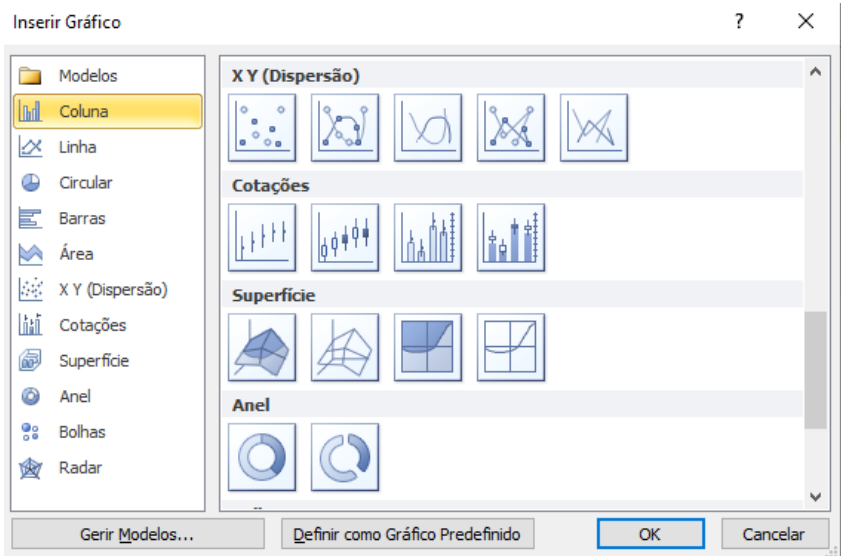
- There are a lot, of different types and with different purposes

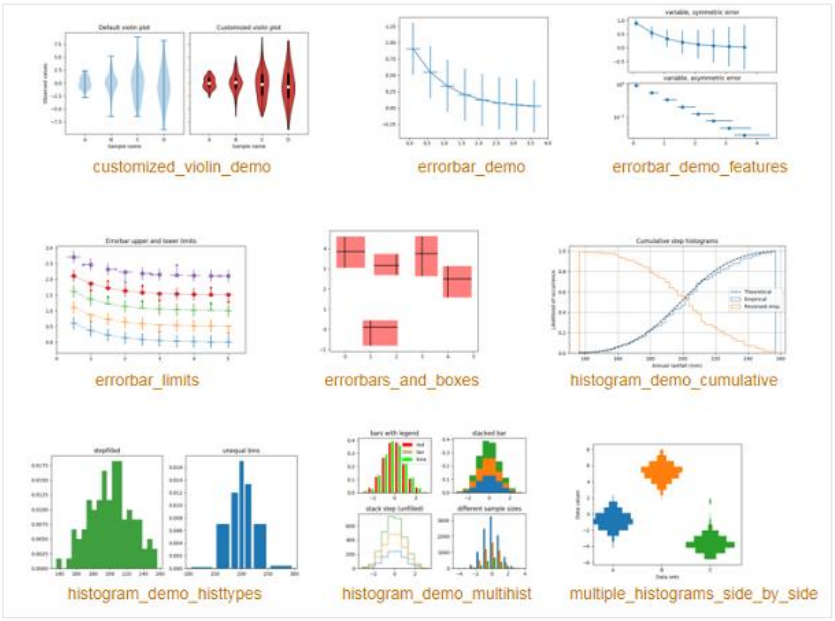
(see e.g. http://selection.datavisualization.ch/)

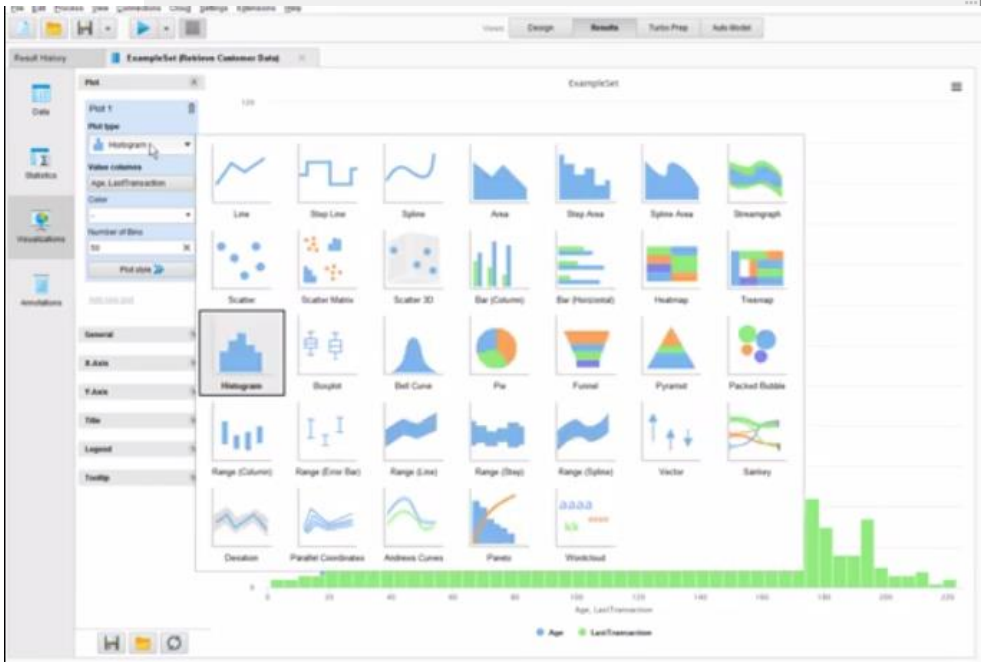


We may use the Visualization capabilities of other S/W or online tools

- Excel
- SPSS
- Rapidminer
- Weka
- ….
- Python graphics libraries (e.g. Bokeh)
- …







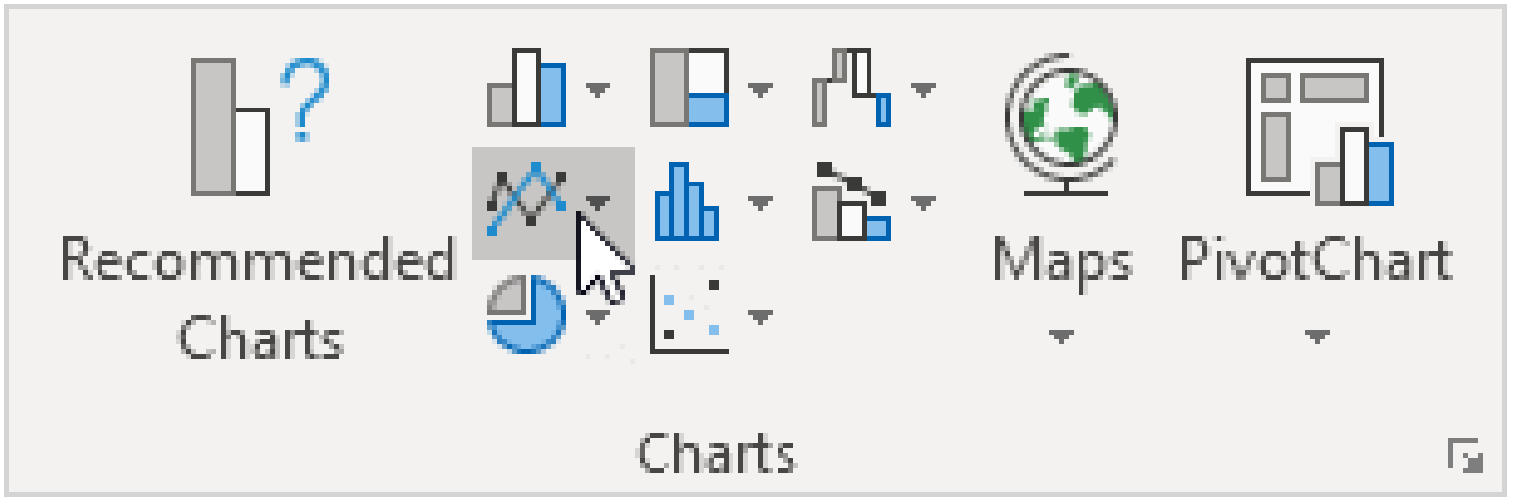
2022 Gartner Magic Quadrant
for Analytics and Business Intelligence Platforms

Note: the report is available upon request

We will use

- Excel
- Python + Matplotlib

(but you may use also other S/W if you master it)



Suggestion for more advanced programmers…

- Python + Bokeh (https://bokeh.org/)

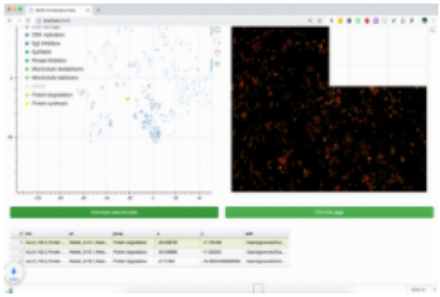
Bokeh (bokeh.org)

User Showcase

Dask

Dask is a tool for scaling out PyData projects like NumPy, Pandas, Scikit-Learn, and RAPIDS. It is supported by Nvidia, Quansight, and Anaconda.

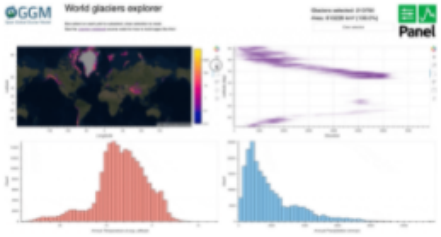
The Dask Dashboard is a diagnostic tool that helps you monitor and debug live cluster performance.



Microscopium

Microscopium is a project maintained by researchers at Monash University.

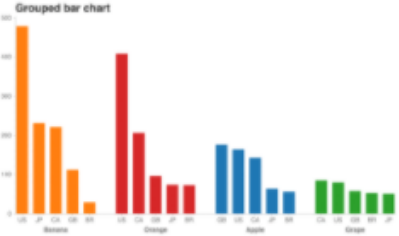
It allows researchers to discover new gene or drug functions by exploring large image datasets with Bokeh's interactive tools.



Panel

Panel is a tool for polished data presentation that utilizes the Bokeh server. It is created and supported by Anaconda.

Panel makes it simple to create custom interactive web apps and dashboards by connecting user-defined widgets to plots, images, tables, or text.



Chartify

Chartify is an opinionated high-level charting API built on top of Bokeh, created by Spotify.

With smart default styles, consistent tidy data format, and a simple API, it's easy for you to concentrate on your work.

For next session:

- Select assessment mode:
  - (along semester/ final)
- If along the semester:
  - Organize groups (5 students)
  - Select assignment topic