Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

ieeta

Visualization in Data Science: a human-centered perspective

(Data Science for Social Sciences Seminar)

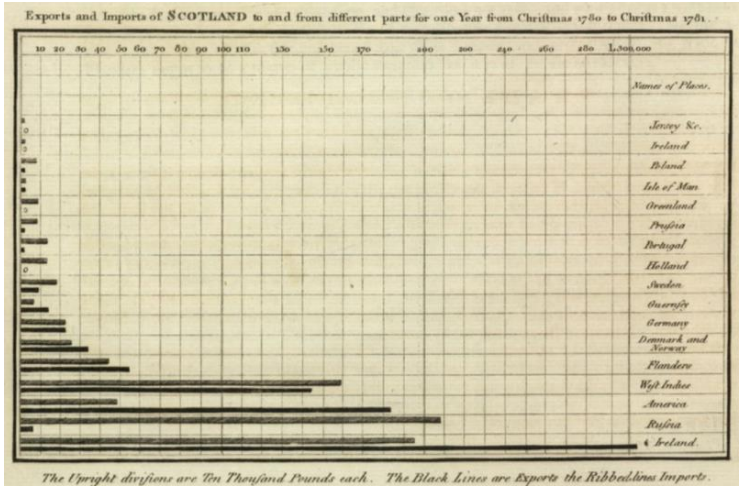
Beatriz Sousa Santos / 2023

The problem…

- What to do with ever more data?
  - Statistics and Data Analysis?
  - Machine Learning?
  - **Visualization**?
- These are not mutually exclusive approaches …

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
- It existed before computers …



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

What it **is not**:

- just "pretty pictures"!

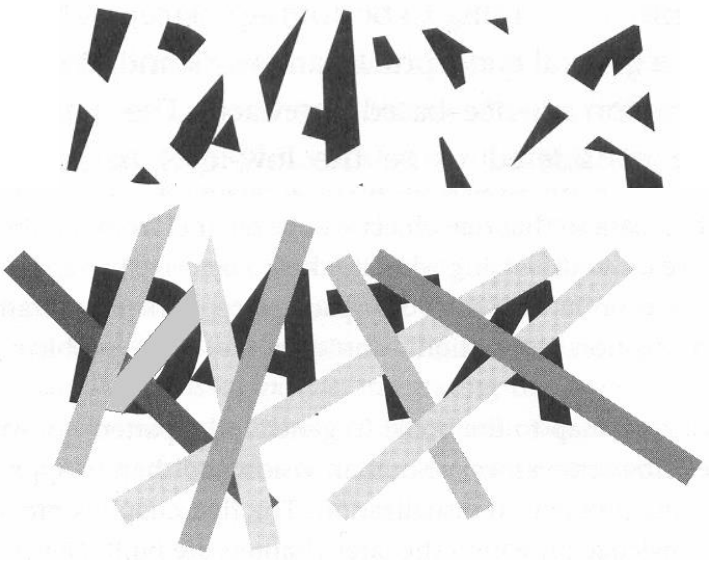
Why and how to represent data visually?

- The human visual system is a most powerful pattern seeker

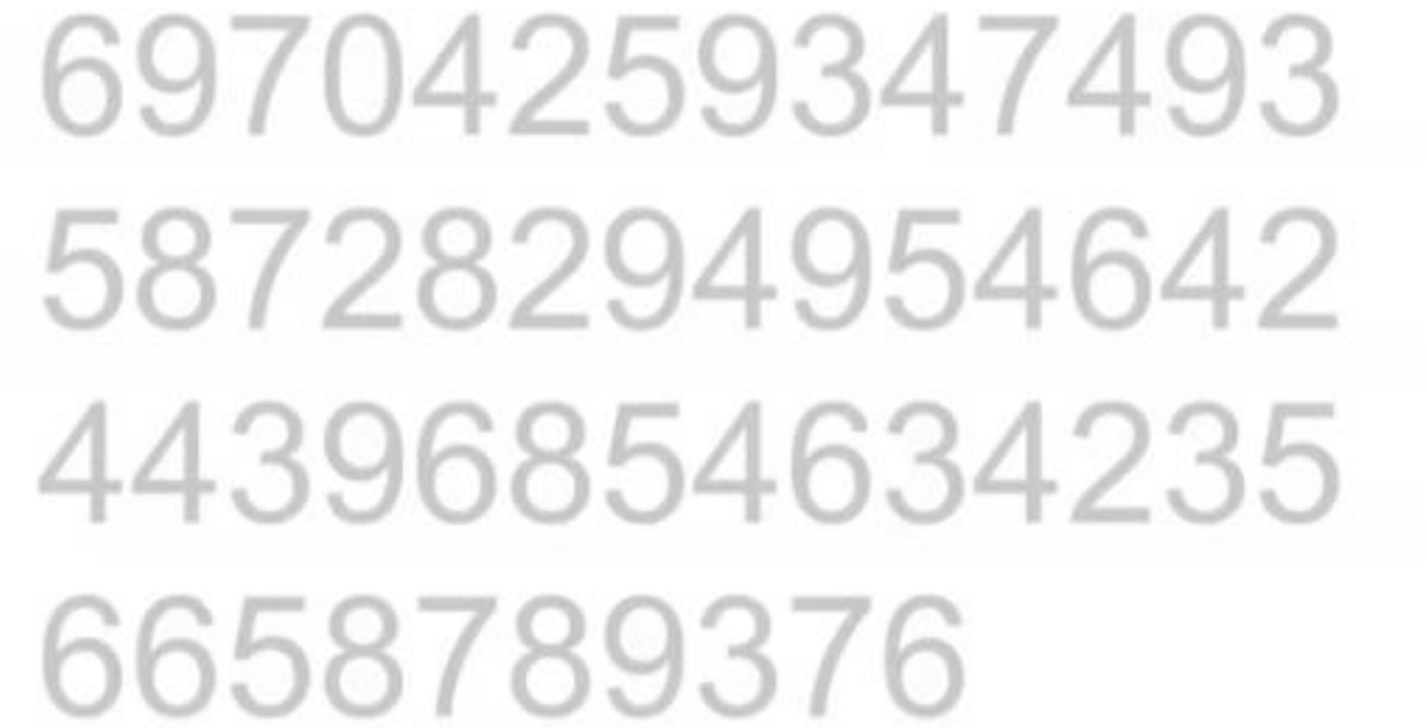
  “seeing is understanding…”

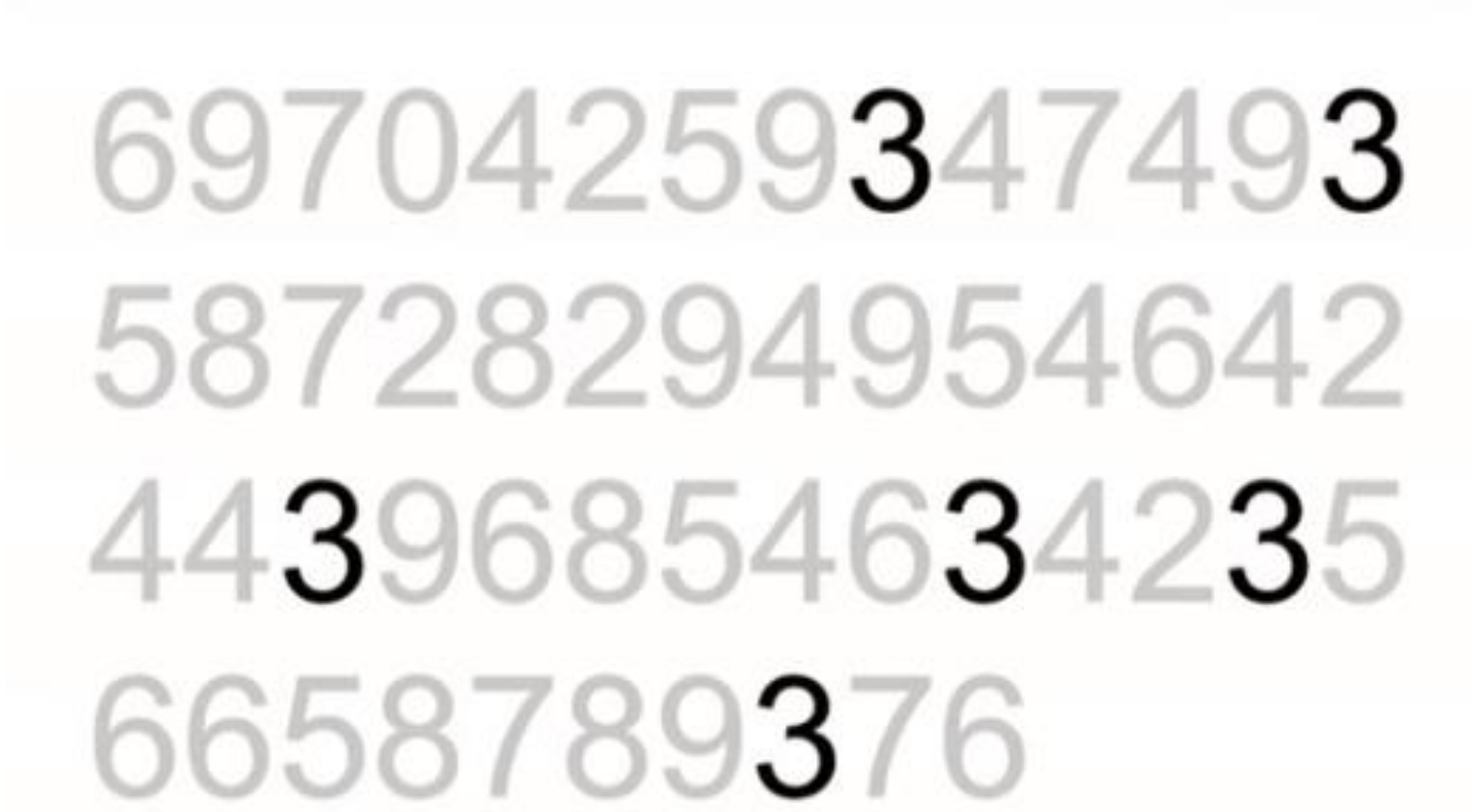
- We easily see patterns displayed in certain ways

  but not in others …



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





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.

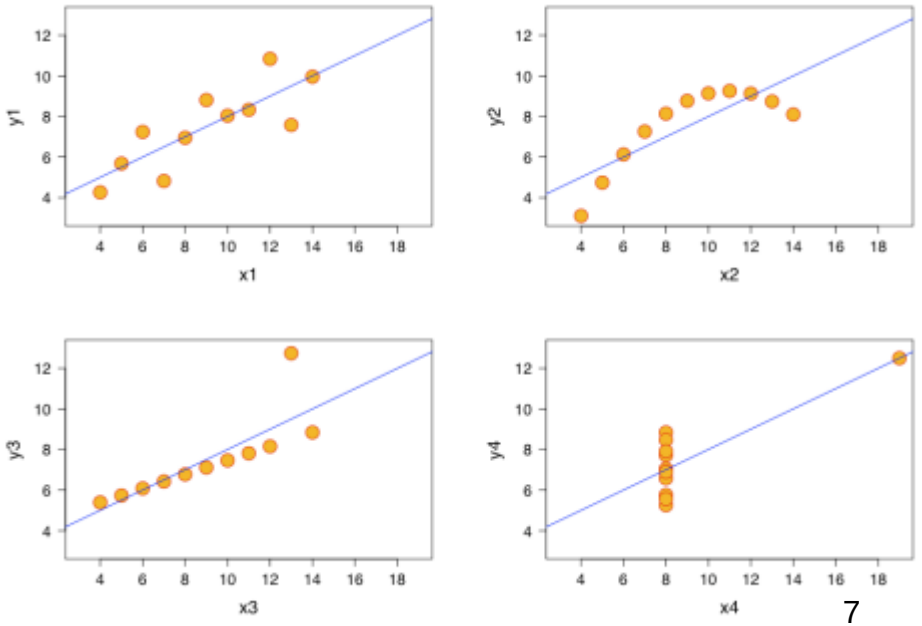
(Munzner, 2014)

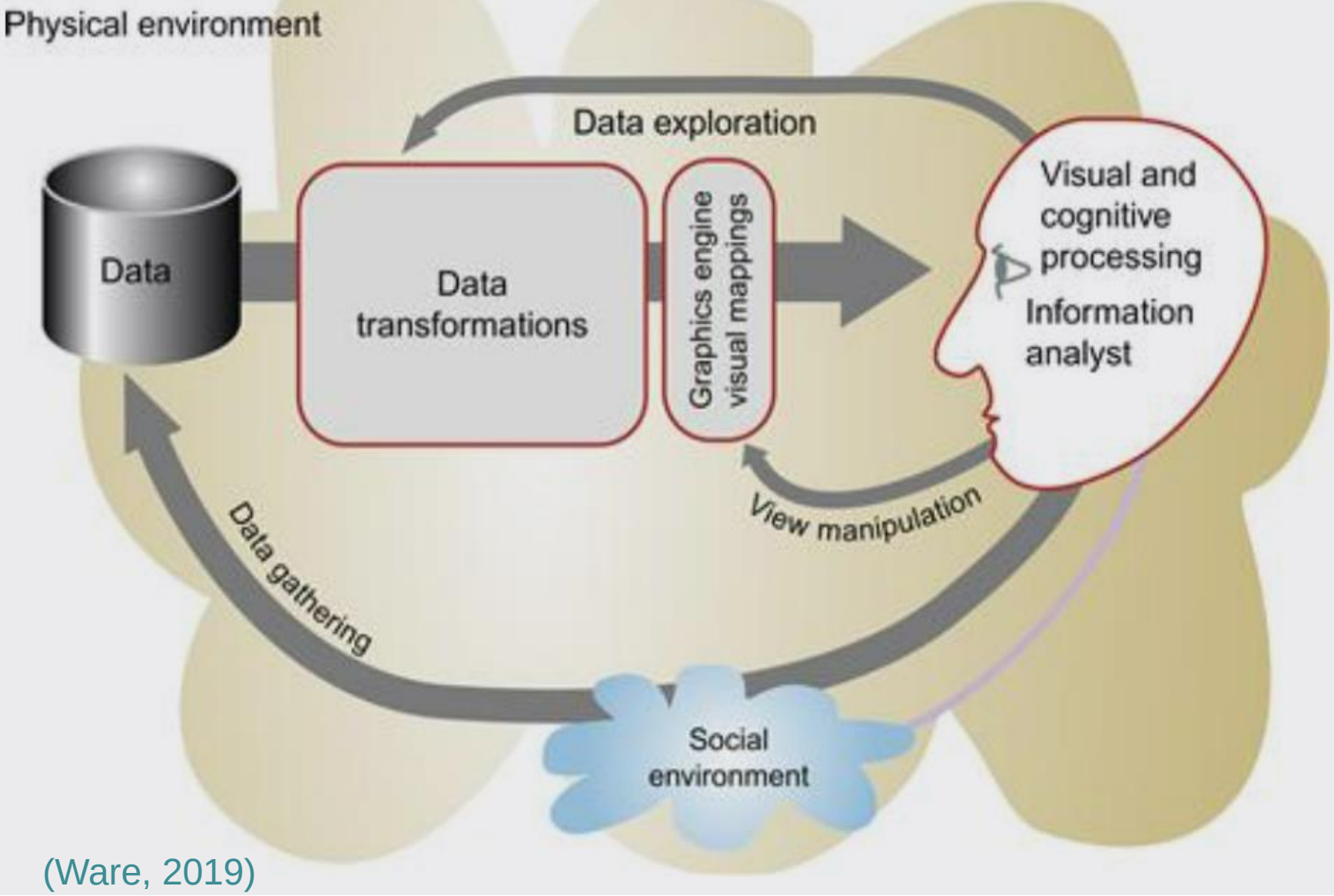
Anscombe's Quartet: Raw Data

	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	

Ascombe quartet: data sets with the same simple statistical properties

(Tufte, 1983)





(Ware, 2019)

Visualization is a Human in the loop process!
It is necessary to take the human into consideration

This talk:

- Introduction to Visualization
- Visualization, Visual Data Mining, Visual Analytics and Machine Learning
- Data and phenomena: types and pre-processing
- Human-in-the-loop process: perceptual and cognitive aspects
- Most used Visualization techniques of quantitative data:

  1D, 2D, 3D and nD Data
- Visualization of other types of data

  networks, hierarchical data, etc.
- Effective Visualization: some principles
- Visualization tools: hints on how to select a tool
- Bibliography and other learning resources

Visual Data Mining

- "The basic idea of visual data exploration is to **present the data in some visual form**, allowing the human to **get insight into the data**, draw conclusions, and directly interact with the data." (Keim, 2002)

- Visual data mining techniques are of high value in **exploratory data analysis**

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

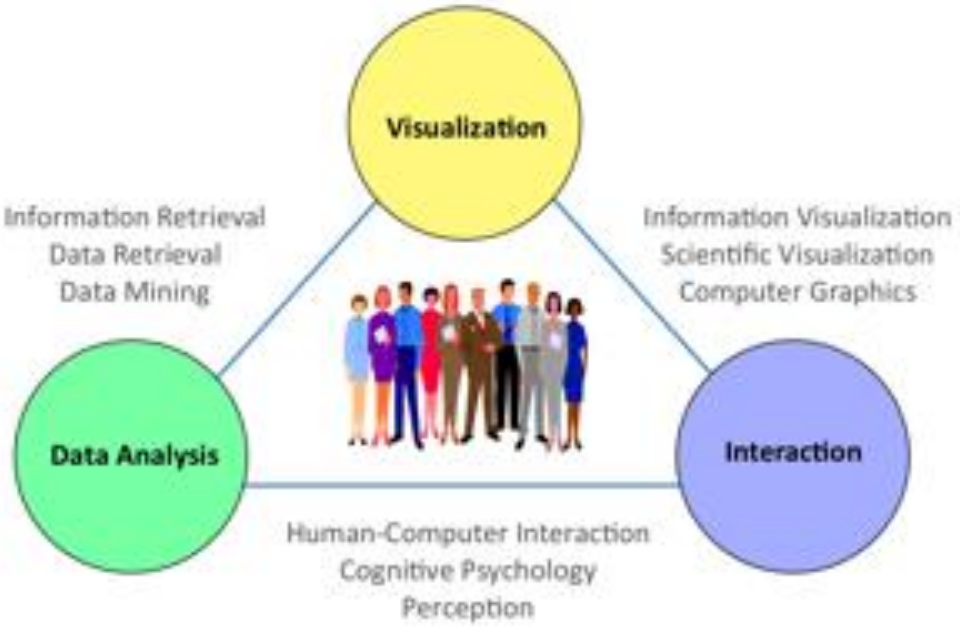
- Since the **user is directly involved**, shifting and adjusting the exploration goals is automatically done if necessary

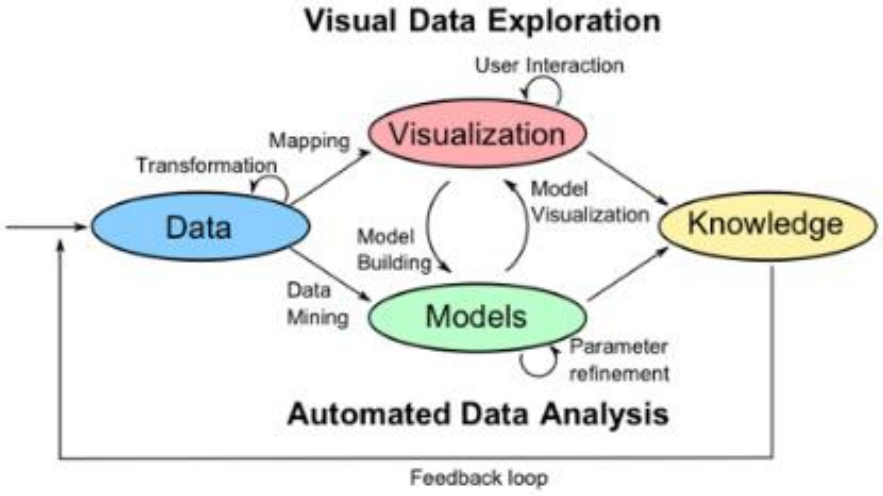
- **Main advantages** of visual over automatic data mining techniques (statistics or machine learning):
  - can easily deal with highly inhomogeneous and noisy data
  - is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters.
- Visual data exploration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

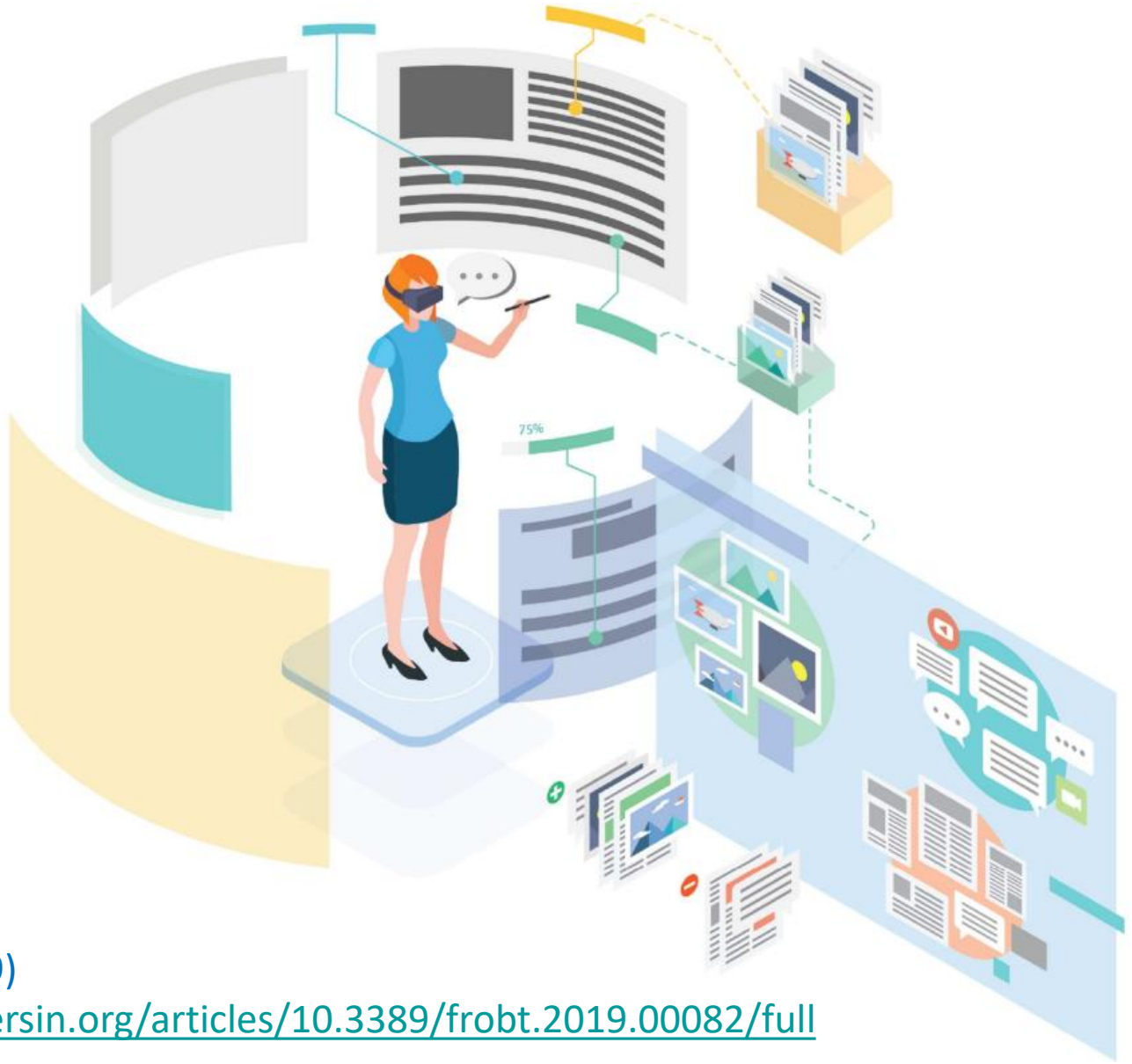




“The Visual Analytics Process **combines automatic and visual analysis methods** with a tight coupling through **human interaction** in order to gain knowledge from data.”

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

Immersive Visual Analytics



(Skarbez et al., 2019)
https://www.frontiersin.org/articles/10.3389/frobt.2019.00082/full

Information 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

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 …

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)

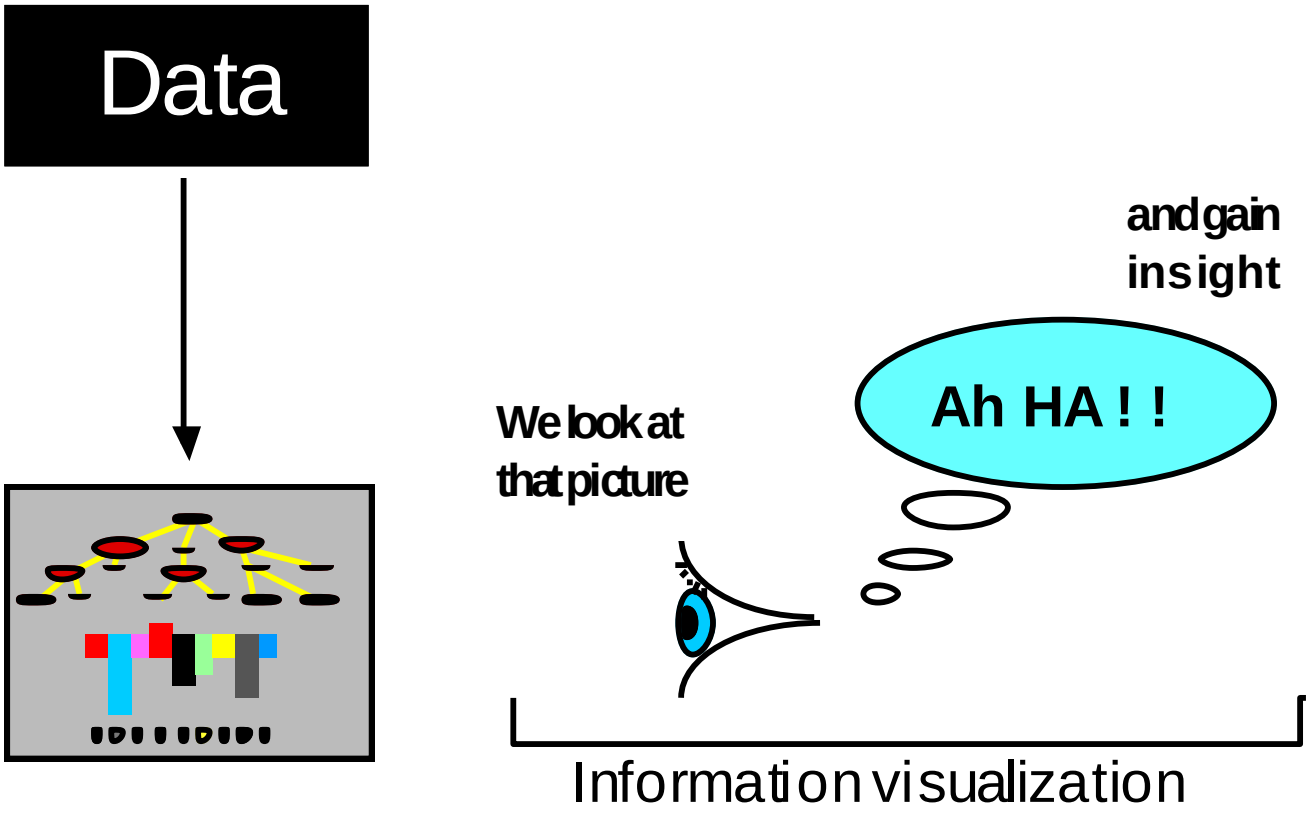
- for transitional use where the goal is to “**work itself out of a job**", by helping the designers of future purely computational solutions, etc.

Visualization to communicate to the general public

Hans Rosling's 200 Countries, 200 Years, 4 Minutes: 120 000 values

Income (x), Age expectancy (y) , Time (t), Continent (colour), Population (size)

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



The process of information visualization: graphically encoded data is viewed in order to form a mental model of that data (Spence, 2014)

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**

- Start by **finding what are the questions** users need to answer!

Example: how to select simple charts?

Max and Min temperatures along the month of February in a specific location (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 maximum MaxT value attained?

Q4- How were the daily temperature ranges?

Q5 – What was the maximum temperature range?

Q6 - Were all the maximum temperatures higher then highest minimum temperature?

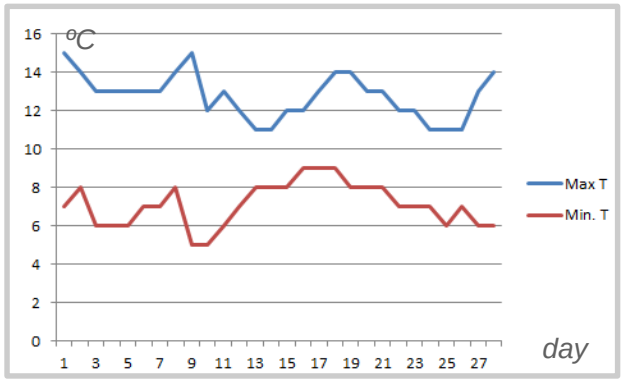
- What type of chart would you use to answer Q1?

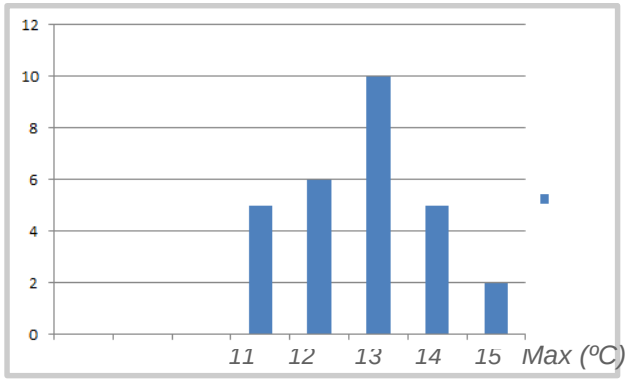
- And the other questions?

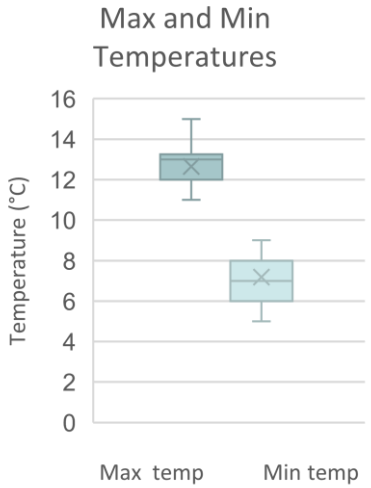
Example: how to select simple charts?

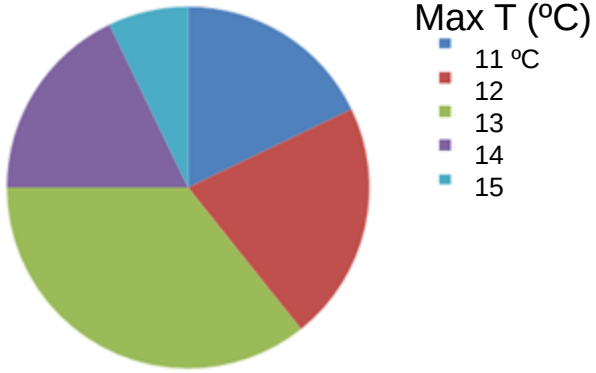
Temperatures along the month of February (in ºC): a few possible charts

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





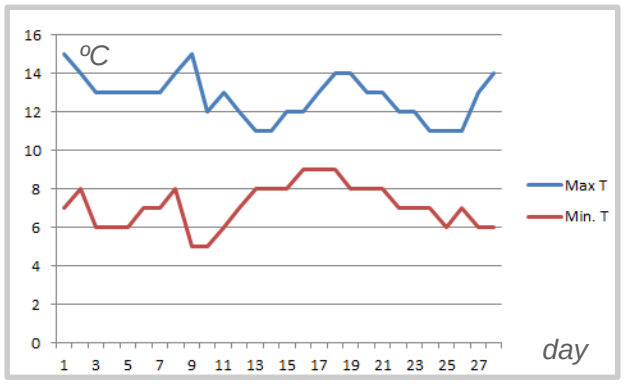




To what questions are they adequate?

Simple example

Temperatures along the month of February (in ºC):

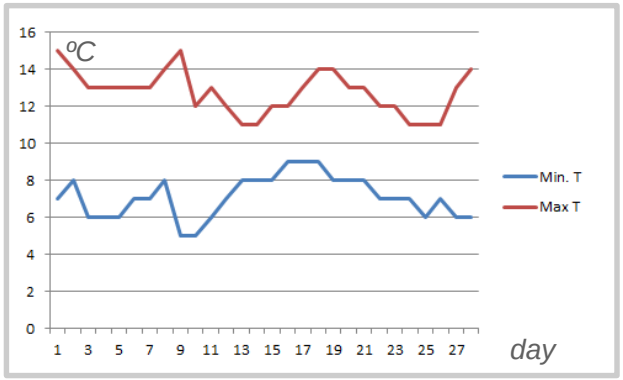


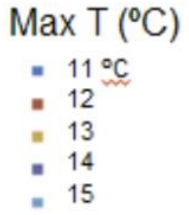
Anything “odd” about this chart?

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!

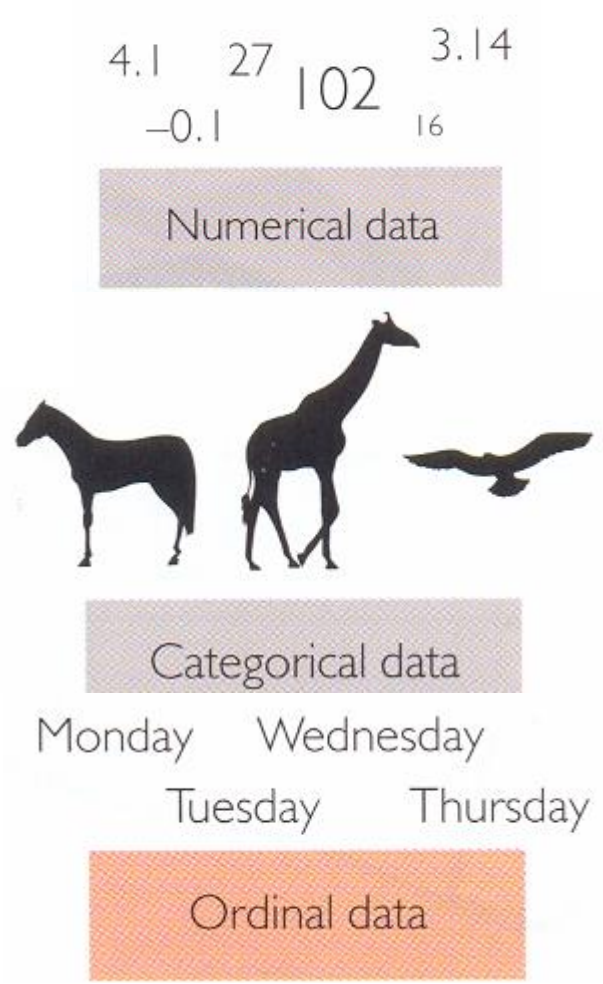
Data Characteristics

(wordcloud of chap.1, Mazza, 2009)

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

- Data nature:
  - Continuous
  - Discrete

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



(Spence, 2007)

- Examples of measuring scales and types of data:
  - **Nominal/categorical -->** 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)

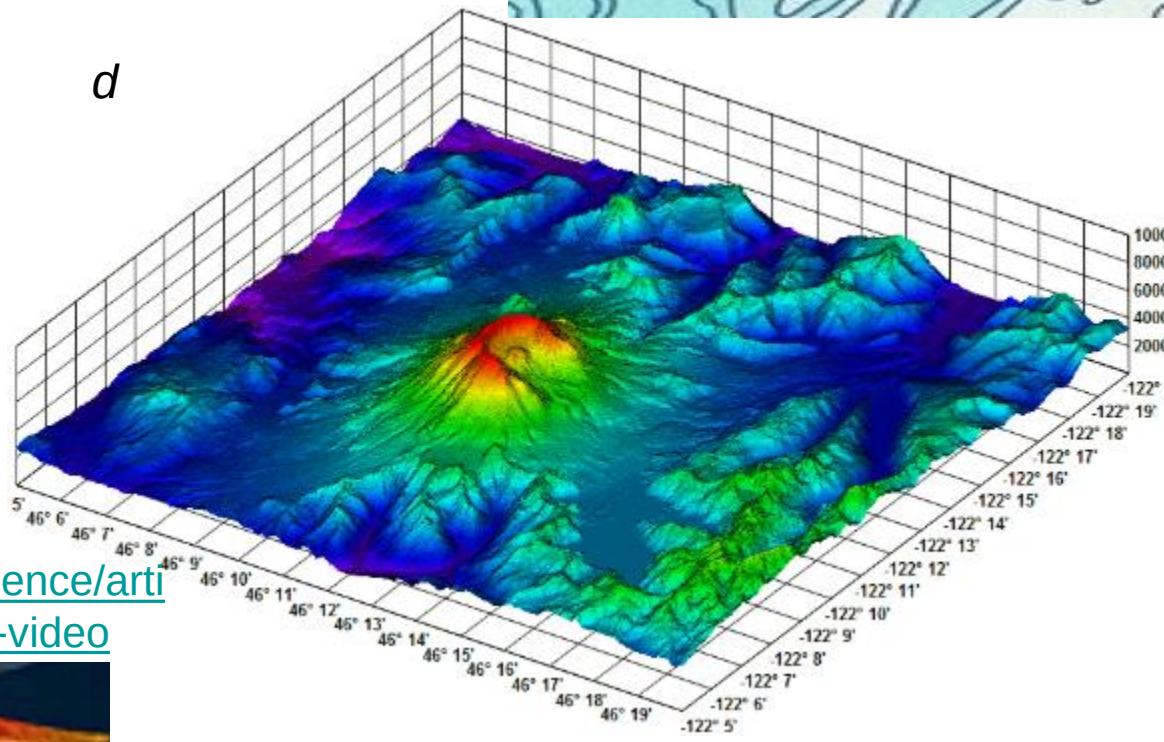
- This is a general classification and may be used to select the statistical methods to use with the data

http://lsc.cornell.edu/wp-content/uploads/2016/01/Intro-to-measurement-and-statistics.pdf
https://www.youtube.com/watch?v=KIBZUk39ncI

Example: beyond the structure of the data to Visualize

- Consider a data set with three columns:

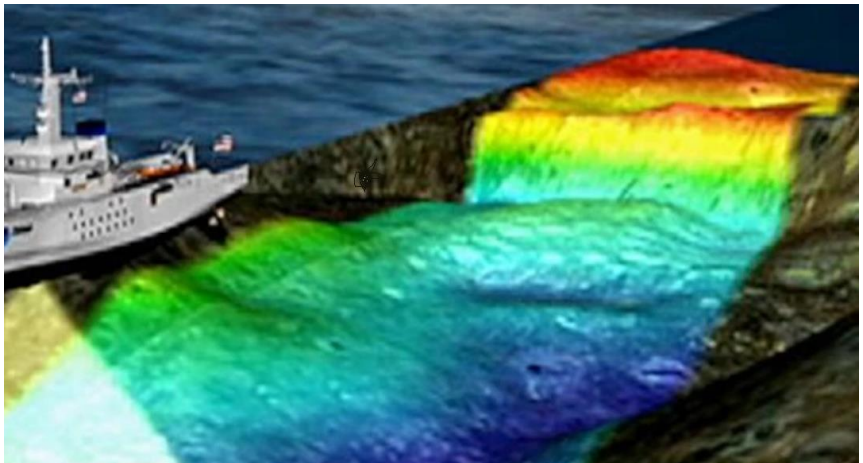
  latitude *longitude* *d*



- How would you visually represent these data?

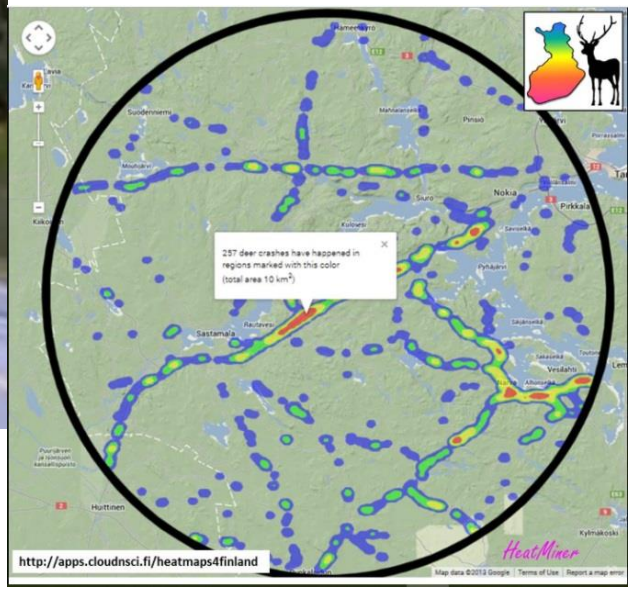
- If *d* is sea depth?

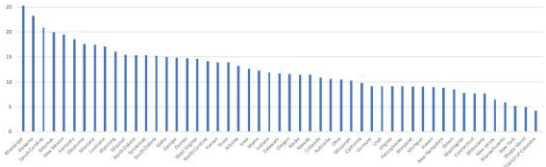
https://www.nationalgeographic.com/science/article/110421-us-ocean-floor-mapping-vin-video



Note:
You may use a visualization technique involving interpolation because the phenomenon is continuous
(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

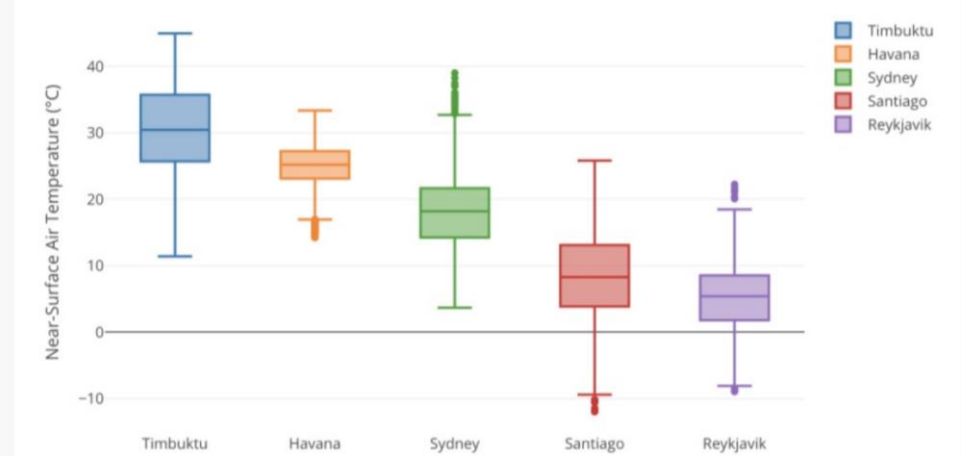
- Interpolation and contours don’t make sense!

Know the data structure is not enough

It is necessary to **know the phenomenon behind the data**

Visualization Literacy Quiz

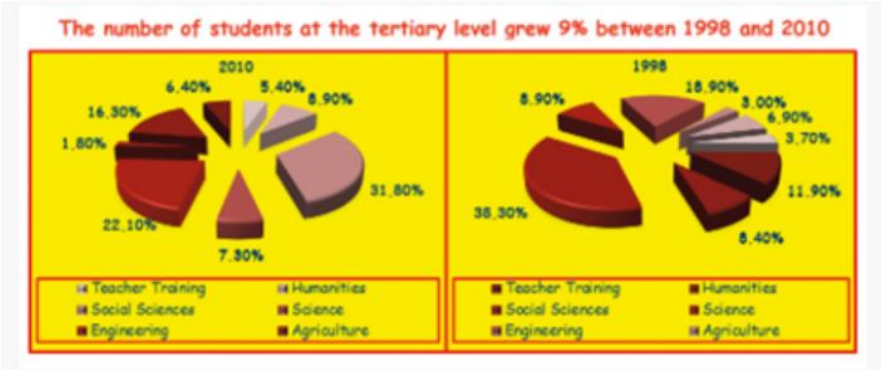
Does this type of data representation look at all familiar?



Choose one of the following answers

- not at all familiar
- not familiar
- reasonably familiar
- very familiar

Analyze these charts and comment on the adequacy of its elements …



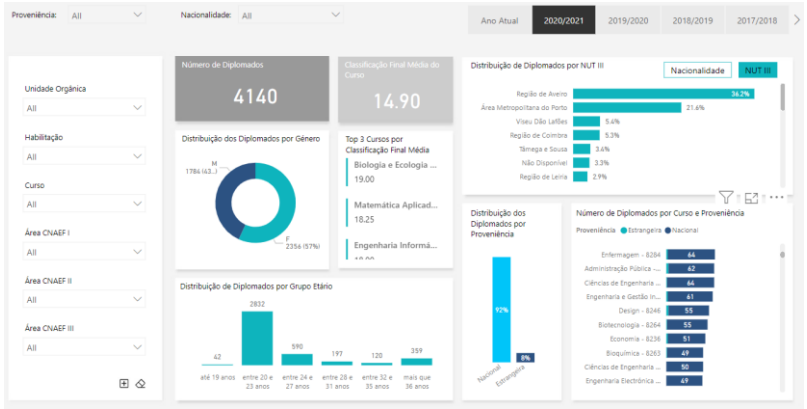
(Camões, 2016)

https://forms.ua.pt/index.php?r=survey/index&sid=127657&lang=en

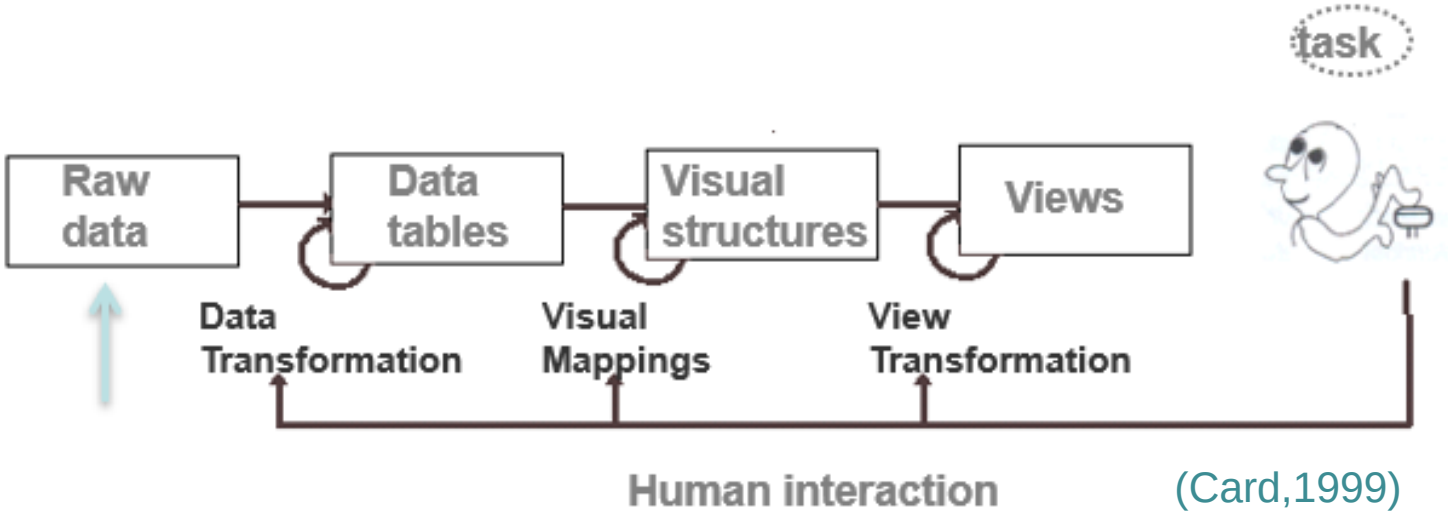
In a nut shell:
Do you have a lot of data?

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

Creating a Visualization



Visualization reference model



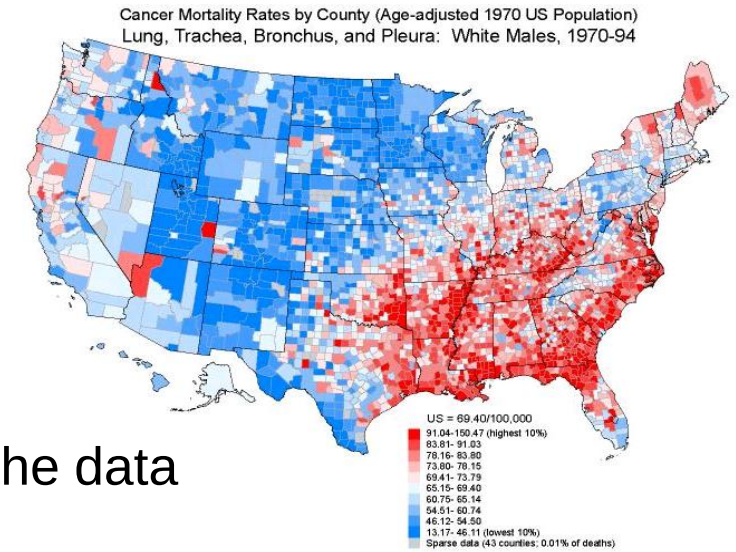
(Card,1999)

Visualization is a **Human-in-the loop process**

Visualizations should help the user **perform the task/ answer the question**

The user should get **insights** from the visualization and in visual data exploration scenarios may be involved in all phases

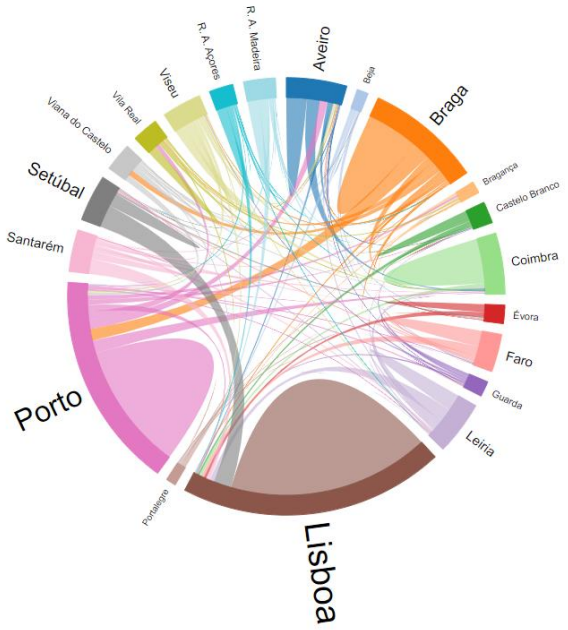
Visual mapping



- It is necessary to decide:
  - which visual structures use to **represent** the data

- Some types of data can be easily mapped to a spatial location
  e.g. data with a physical or geographical structure

- Abstract data don't have an easy correspondence with the dimensions of the physical space around us



Three **structures** must be defined in the **visual mapping/encoding**:

- spatial substrate

- graphical elements

- graphical properties

- **Spatial substrate** - dimensions in physical space where the visual representation is created (can be defined in terms of axes and type of data)

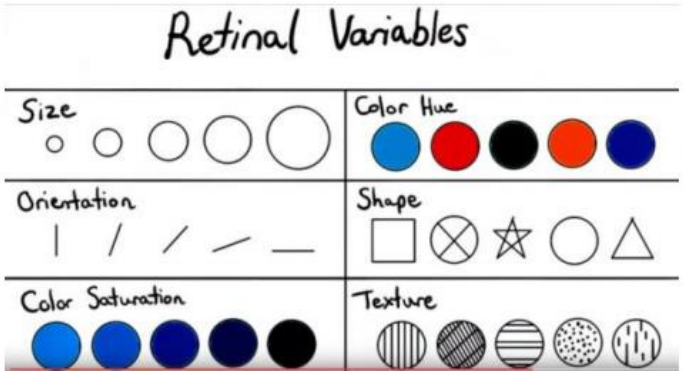
- **Graphical elements** - anything visible appearing in the space
  points, lines, surfaces, volumes

- **Graphical properties** – properties of the graphical elements to which the human retina is very sensitive - **retinal variables:**
  size, orientation, color, texture, and shape

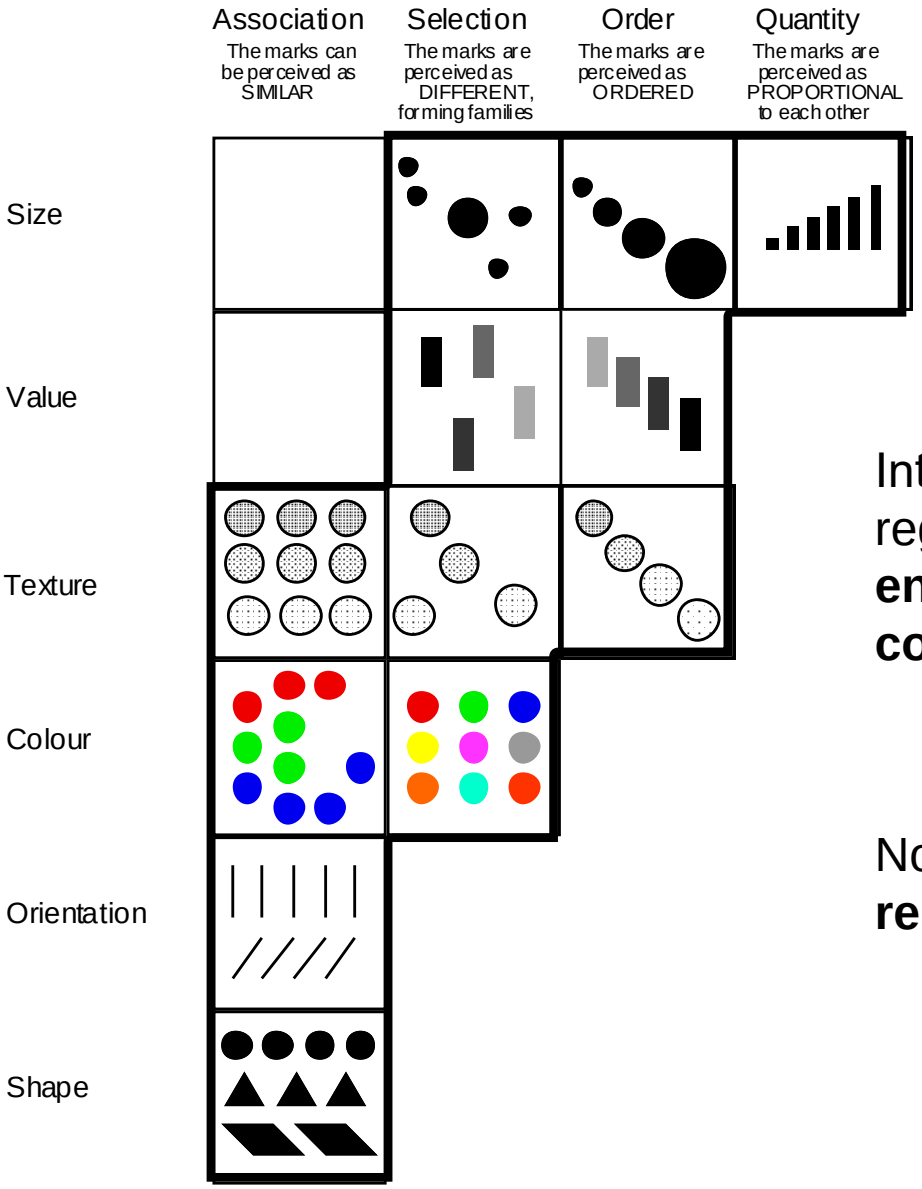
- **Spatial substrate** axes (x, y, …)
  type of data (quantitative, ordinal, categorical)

- **Graphical elements** points
  lines
  surfaces
  volumes

- **Graphical properties** retinal variables:
  size,
  orientation
  color (depends on physiology and culture)
  texture
  shape



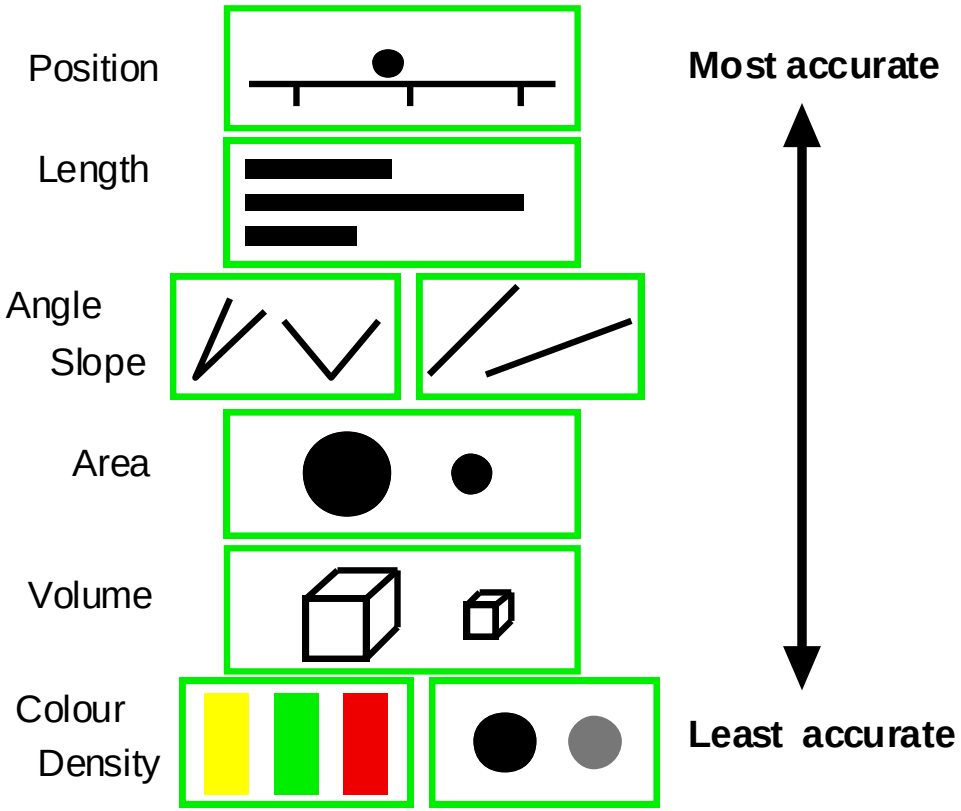
How to select visual encodings?



Interpretation of Bertin's guidance regarding the suitability of various **encoding methods to support common tasks** (Spence, 2014)

Note that **only size is adequate to represent quantity accurately**

How to select visual encodings to accurately represent quantity?



The relative difficulty of **assessing quantitative value** as a function of encoding mechanism, as established by Cleveland and McGill (Spence, 2014)

Important aspects to consider to select a Visualization technique:

nature of the problem
- communicate
- explore
- confirm

nature of the data to represent
- quantitative
- ordinal
- categorical

number of attributes
- univariate
- bivariate
- trivariate
- multivariate

Next: visualization techniques organized according the n. of attributes

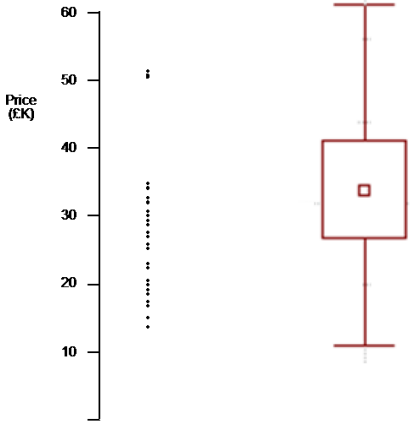
dataset types
- tables
- networks
- spatial or geographical
- fields
- geometry

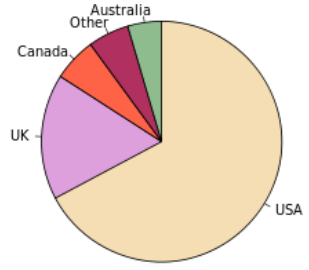
of tabular data

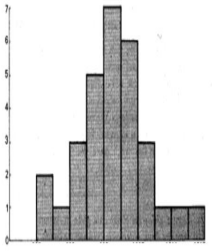
...

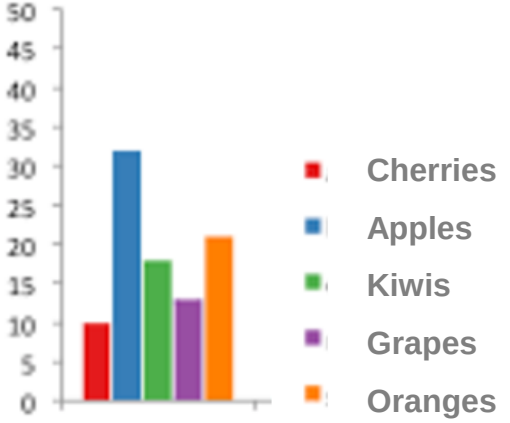
Common Visualization Techniques to visually represent univariate, bivariate data

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



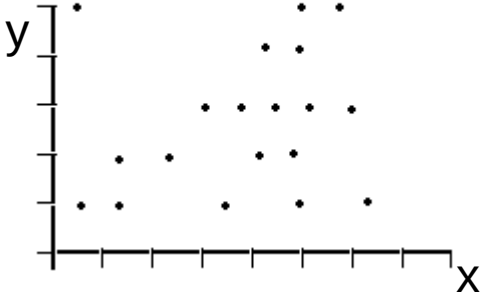


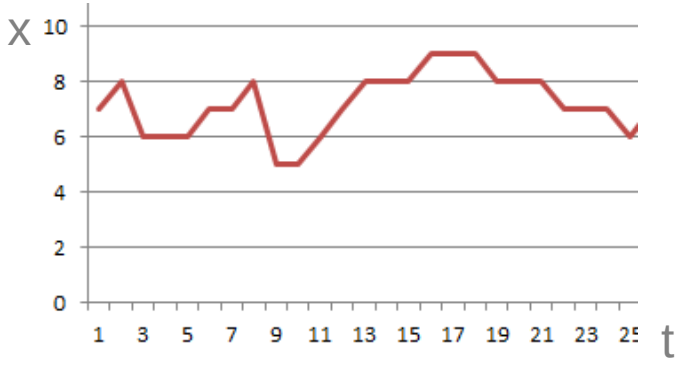




Bivariate data
- scatter plot
- line plot
- time series
- …

…





Representing univariate data

- A more common situation consists in representing a set of values
- Well established techniques exist
- But new ones can be invented!

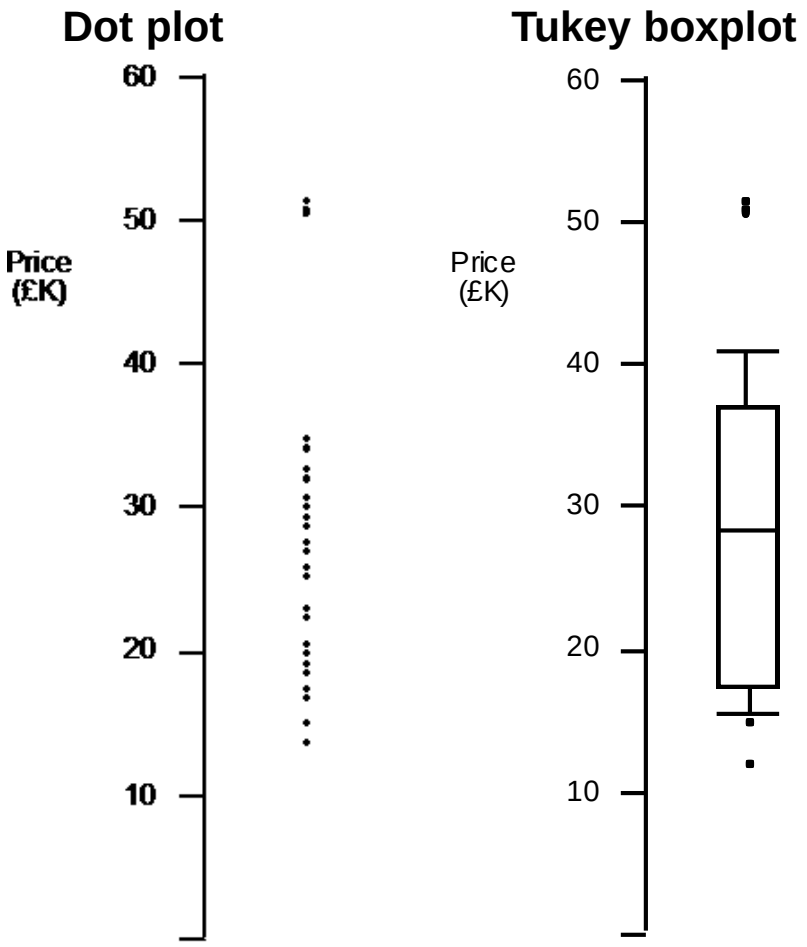
Example:

Price for a number of cars:

- dots on a linear scale
- box plot

(that will answer many questions: median value, outliers,...)

(Spence, 2007)



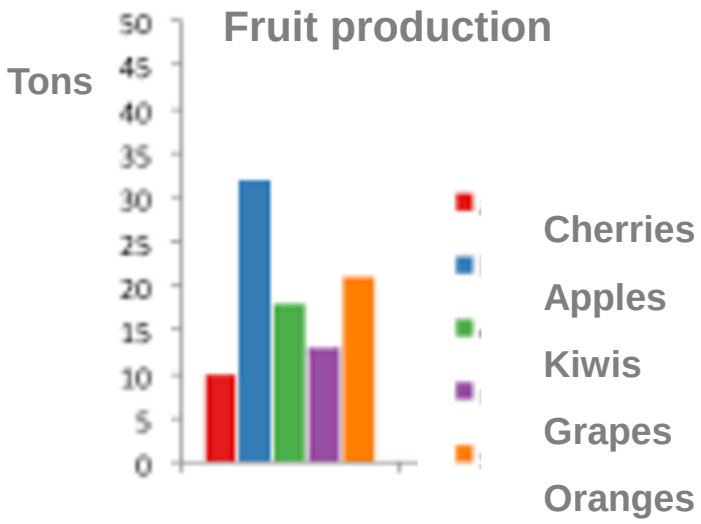
https://www.data-to-viz.com/caveat/boxplot.html

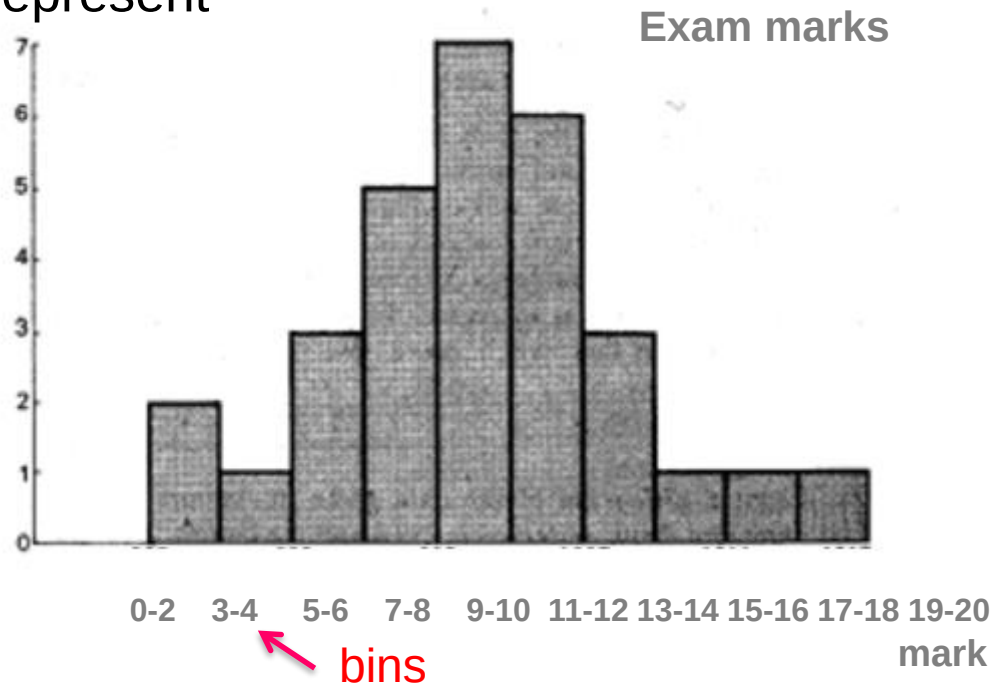
Two common techniques not to be confounded !

Histogram ⇨ represents a distribution of numerical data

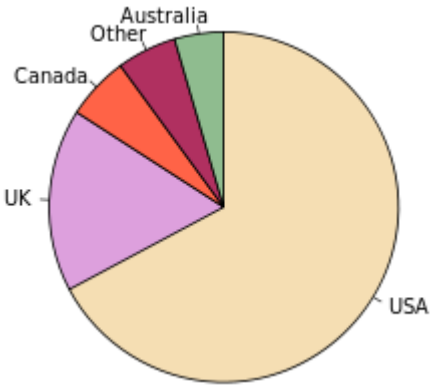
Bar chart ⇨ represents the number of occurrences of a categorical/ ordinal data

Both represent data by rectangular bars (vertical or horizontal) with length proportional to the values they represent





Another simple (and too common) technique



Native English speaking population

- Pie Chart

Represents numerical proportion, **parts of an whole**

The arc length of each slice (its central angle and area), is proportional to the quantity it represents

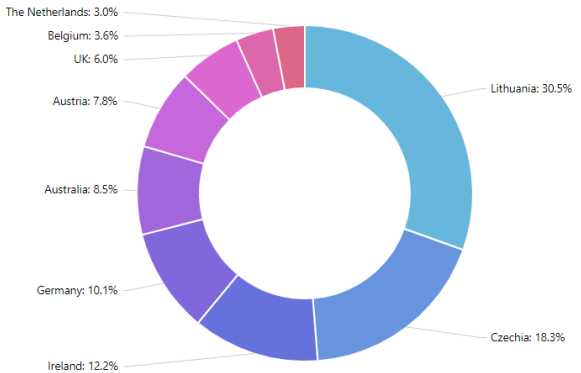
Are much controversial:

many experts recommend avoiding them

http://www.perceptualedge.com/articles/08-21-07.pdf

It is difficult to compare different sections of a pie chart, or to compare data across different pie charts

Variations of pie charts:



- Simple criteria to determine whether a pie chart is acceptable
- Consider it **only if**:
- **The parts make up a meaningful whole**
- **The parts are mutually exclusive**
- **There are <6 parts and slices have not very different sizes**

If the main purpose is to compare between the parts,
use a different chart!

https://eagereyes.org/techniques/pie-charts

Q: can we use a pie chart to represent the students of different nationalities enrolled at our university?

Representing bivariate data

- The **scatterplot** is the conventional representation

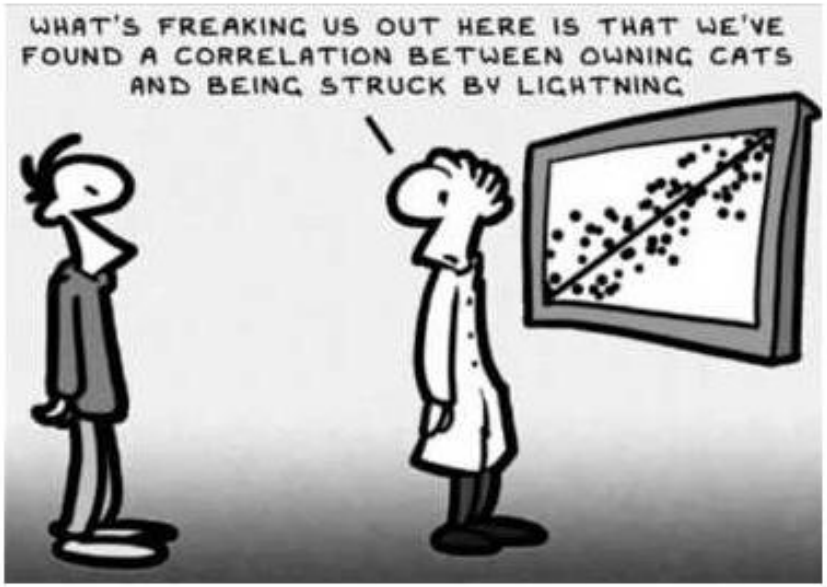
Each observation is represented by a point on a two dimensional space

The axes are associated with these two attributes

This representation affords awareness of:

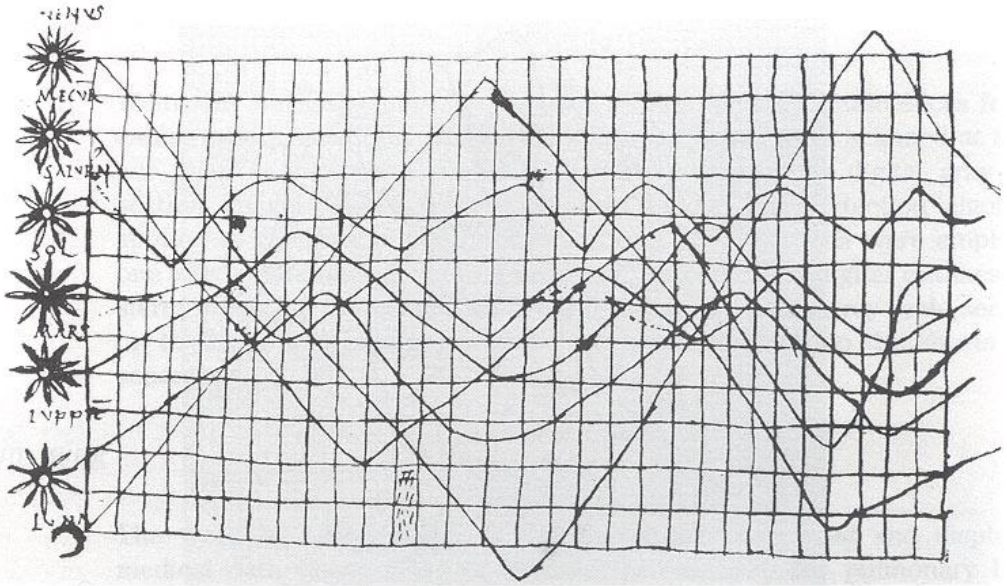
- general trends
- local trade-offs
- outliers

Correlation is not causation!



Representing bivariate data

The **line chart**
One of the oldest known and ubiquitous Visualizations

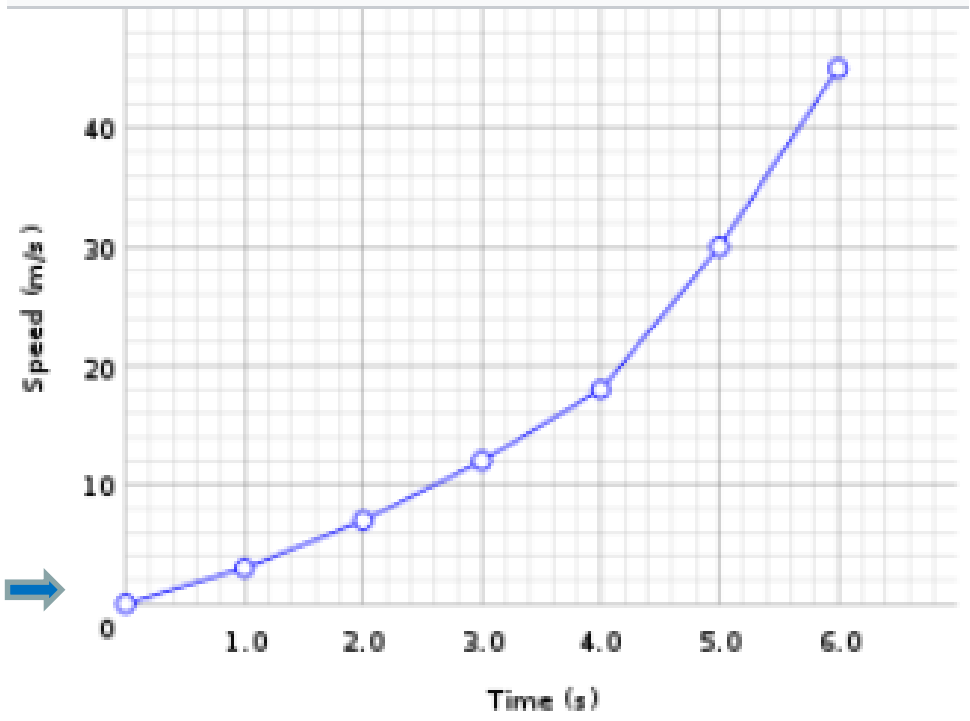


Inclination of orbits along the time - Xth century (Tufte,1983)

- A **line chart** or **line plot** or **line graph** or **curve chart** displays information as a series of data points called 'markers' connected by straight line segments

- Basic type of chart common in many fields

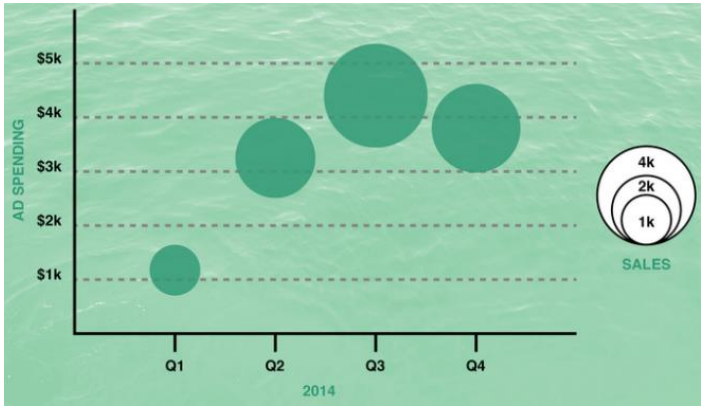
- Often used to visualize a trend in data over intervals of time



Q: in this case shall we use a solid line connecting the points?

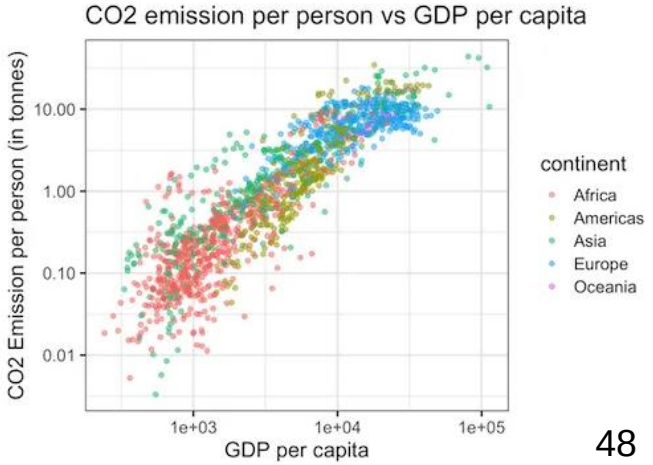
Simple (and common) representations of trivariate data

- In a **bubble chart** data are represented as a disk that expresses two of the values through the disk's *xy* location and the third through its size (radius or area?)



- Mapping the variable to size must be done carefully. The interpretation of size may be ambiguous

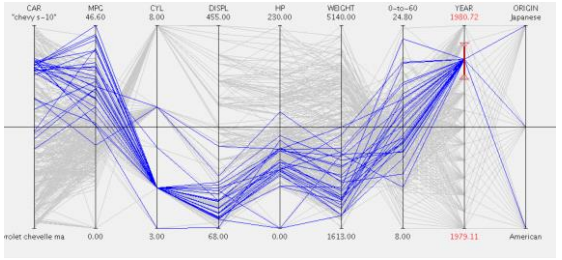
- Representing one more dimension through color



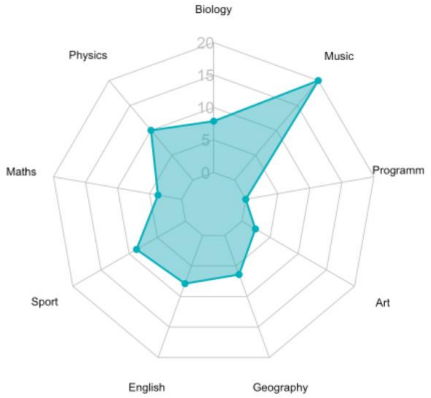
https://visage.co/data-visualization-101-bubble-charts/

Techniques for Multivariate (or Hypervariate) data

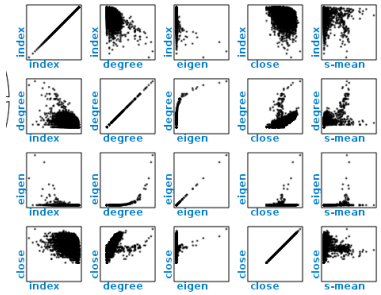
- Coordinate plots
  - parallel coordinate plots



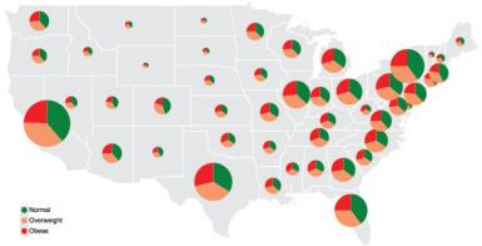
  - star (radar/spider) plots



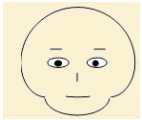
- Scatterplot Matrix



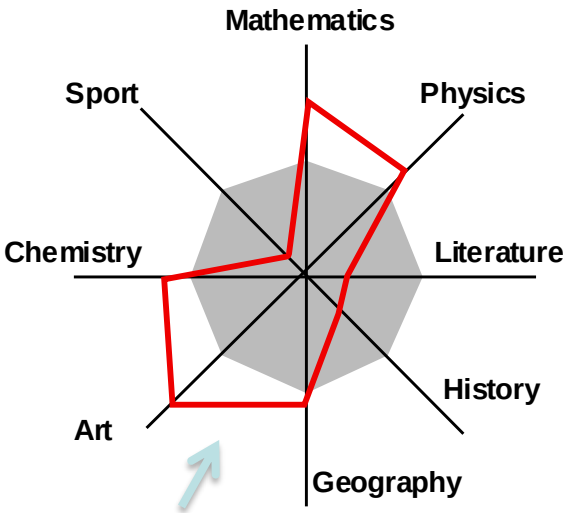
- Maps



- Icons/glyphs



- In **Star plots** attributes are represented by a point on a coordinate semi-axis

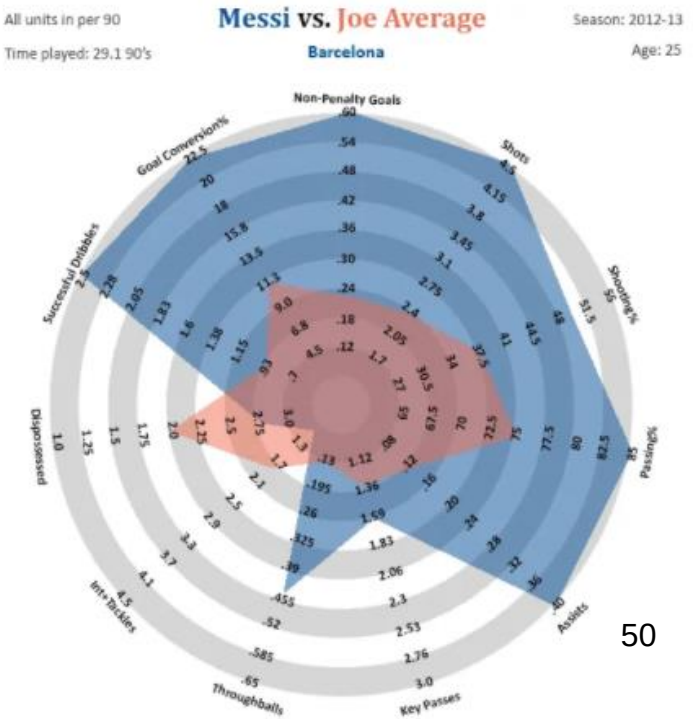


(Spence, 2007)

- Attribute axes radiate from a common origin

- For a given object, points are joined by straight lines

- Other useful information such as average values or thresholds can be encoded



- The **scatterplot matrix** (SPLOM) is applicable to higher n. of variables

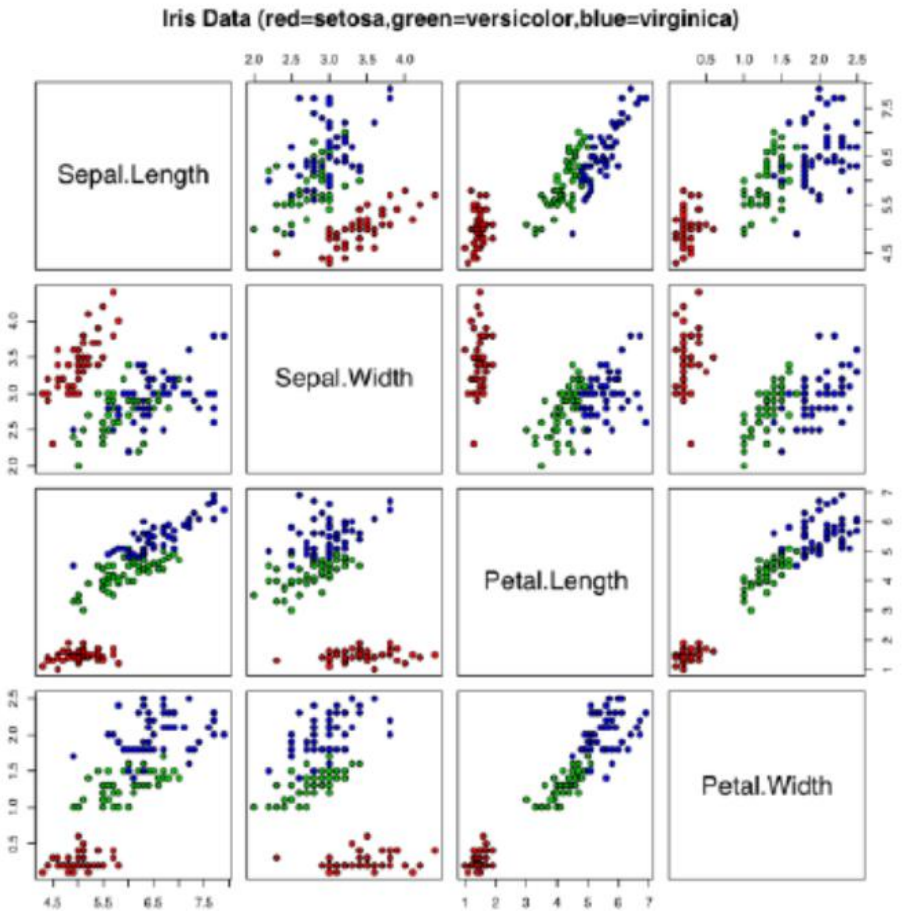
- However, as the number of attributes increase, the number of different pairs of attributes increases rapidly:

  – 2 variables-> 1 scatterplot

  – 3 variables -> 3 scatterplots

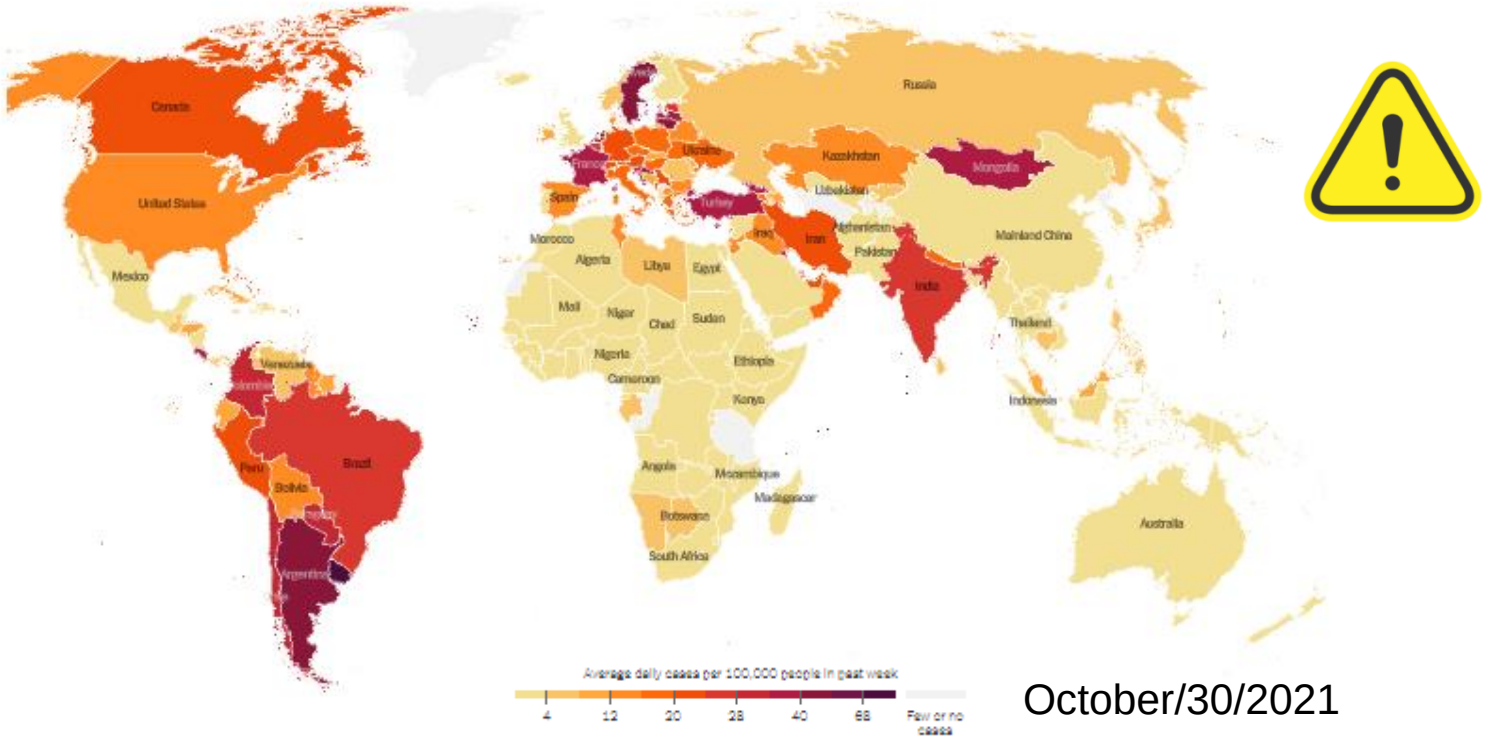
  – 4 variables -> 6 scatterplots

We may try to reduce the number of dimensions keeping the more relevant: **Dimensionality reduction!**



Choropleth maps - A standard approach to communicating aggregated data by geographical areas using color encoding of the geographic area

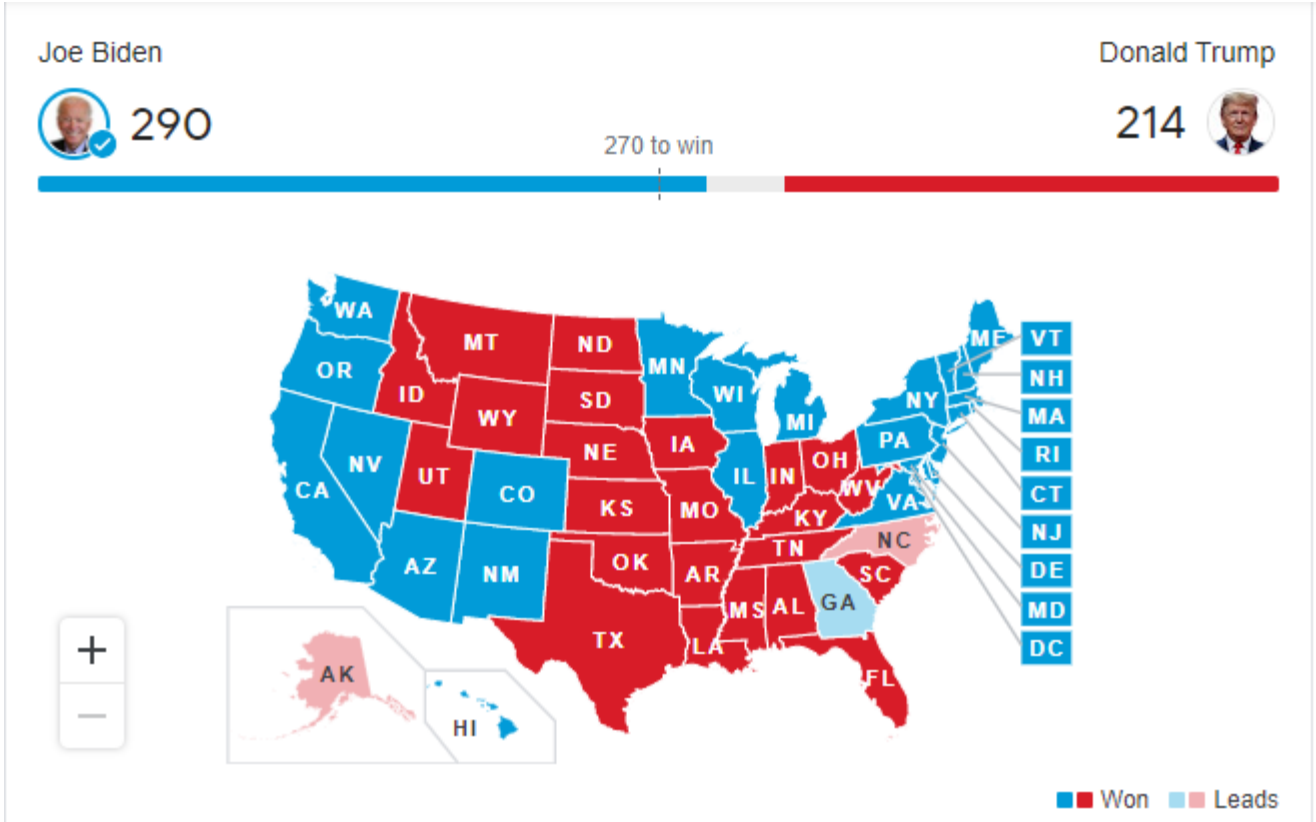
They require some care **Q: what are the possible issues?**



October/30/2021

https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html

Visualizations of the US 2020 Election
(choropleth + bar)
the bar helps better understand the ratio of votes



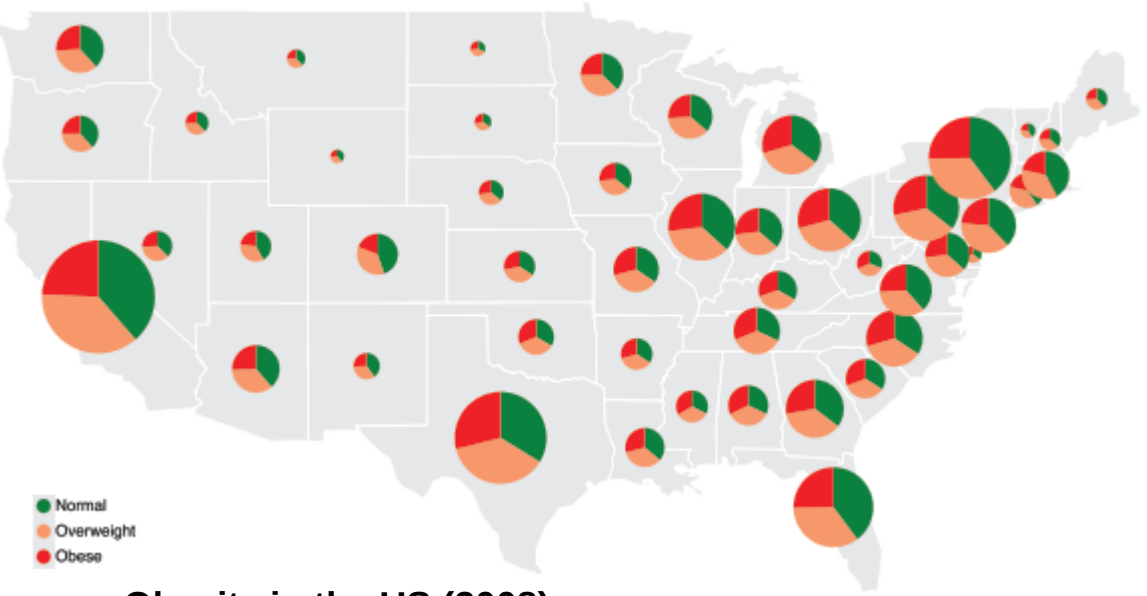
Candidates	Electoral votes	Vote %	Vote count
Joe Biden Democratic Party	290	50.6%	75,198,127
Donald Trump Republican Party	214	47.7%	70,804,457

Simple representations of attributes on a map

- **Graduated Symbol Maps** are an alternative to the choropleth map;

- Symbols are placed over an underlying map; may show more dimensions

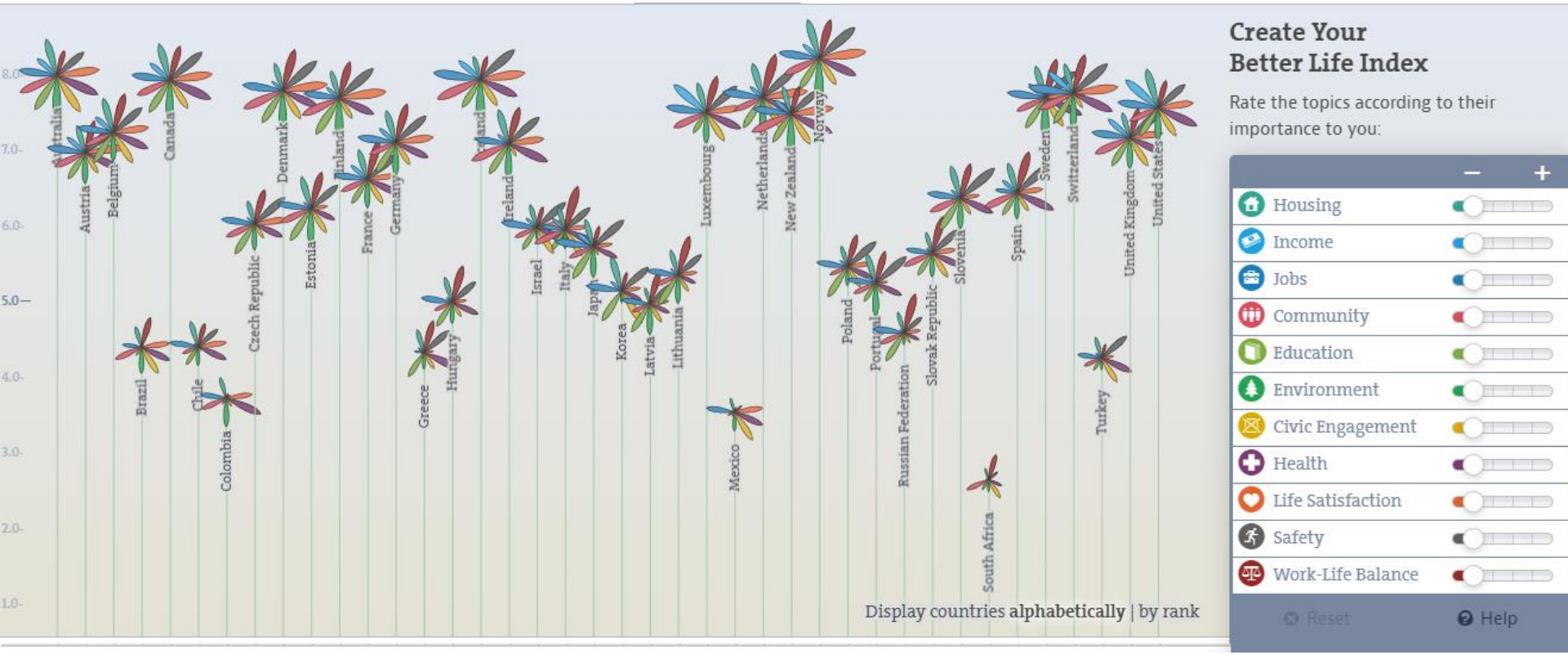
- Avoid confounding geographic area with data values



Obesity in the US (2008)
(Heer et al., 2010)

Glyph chart example

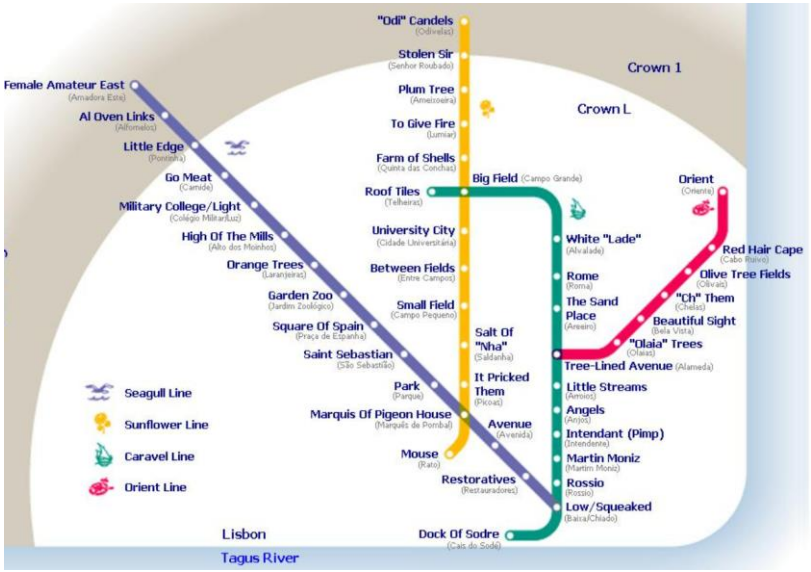
The physical properties of the shape represent different categorical variables sized according to the associated quantitative value and distinguished through color

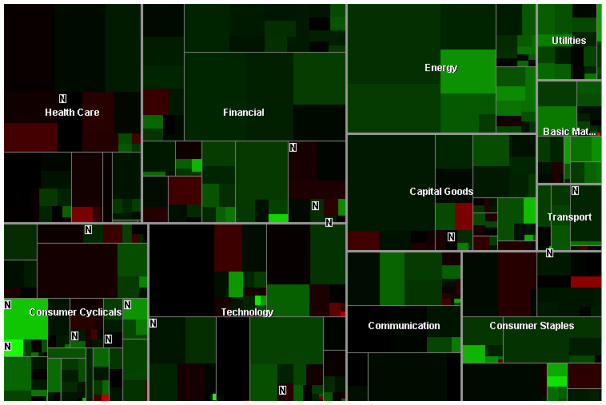


(Kirk, 2012)

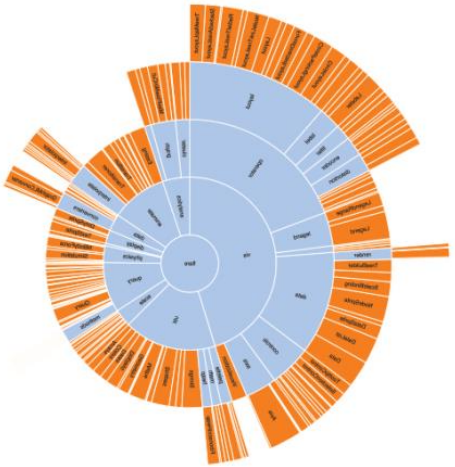
http://oecdbetterlifeindex.org

- These are only some of the visualization techniques to represent a value
- There are a lot more …
- And we may want to visually represent beyond value: **relation**
  - Networks
  - Hierarchical data





https://www.nngroup.com/articles/treemap/

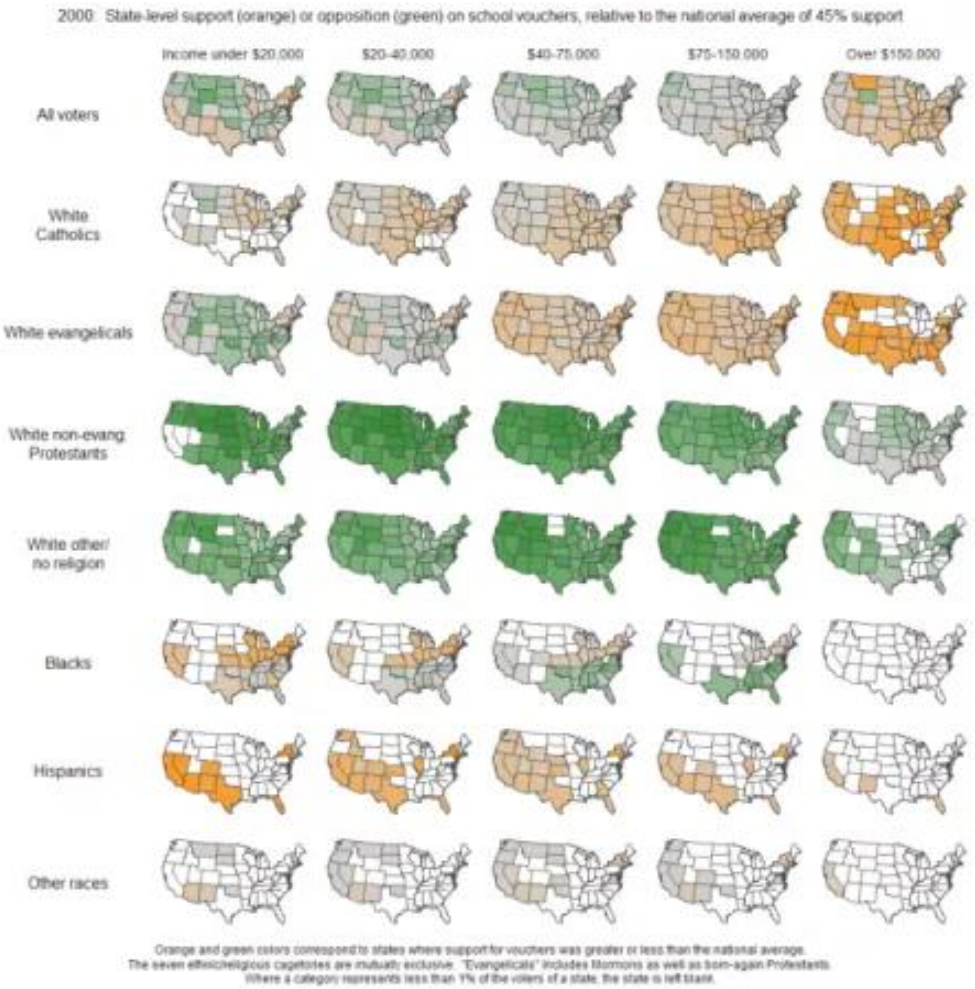


https://plotly.com/python/sunburst-charts/

Small multiples:

arrangement approach that facilitates efficient and effective comparisons

(Kirk, 2012)

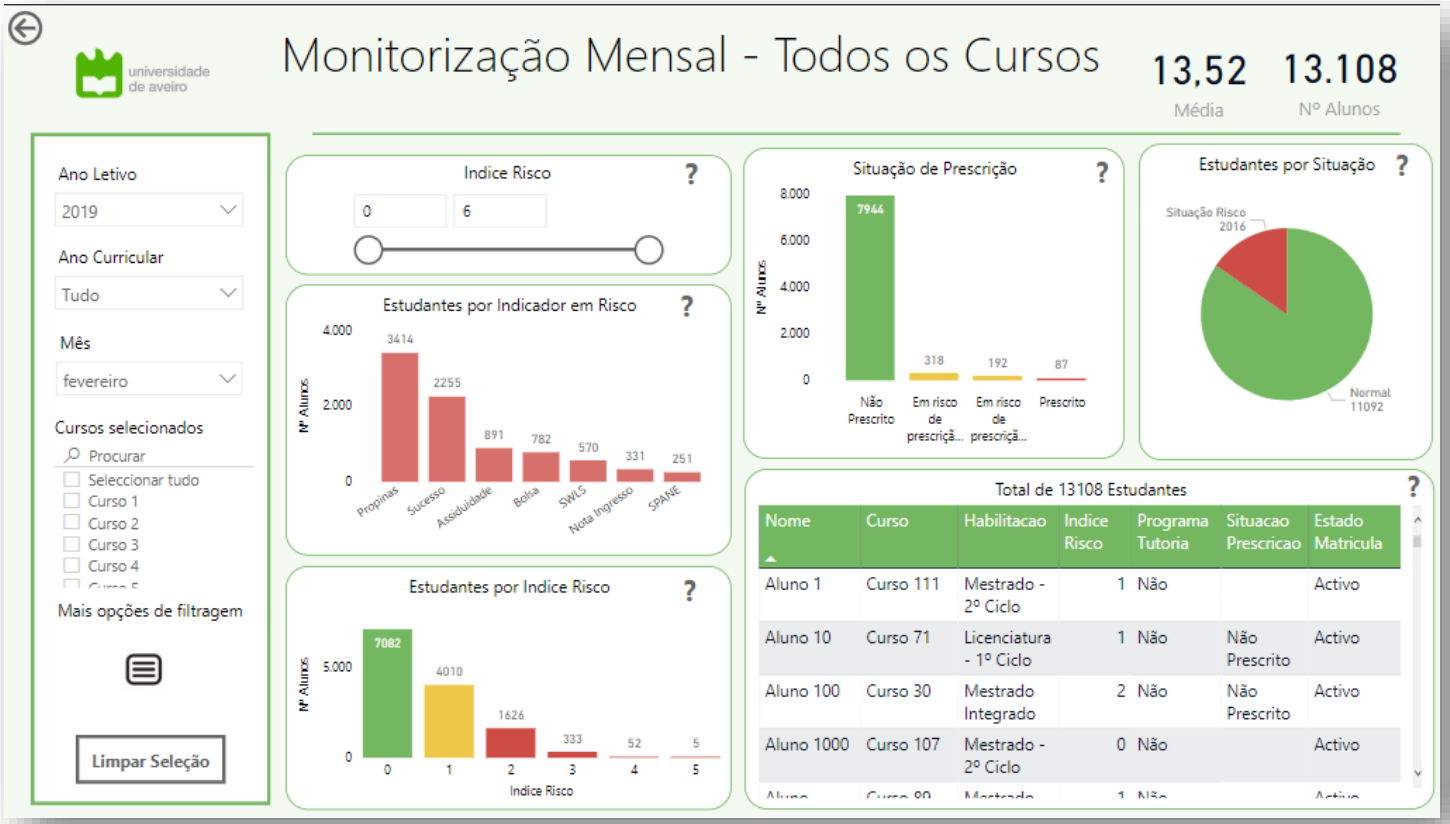


Dashboards

Visual display summarizing a dataset providing information at-a-glance (e.g. KPIs)

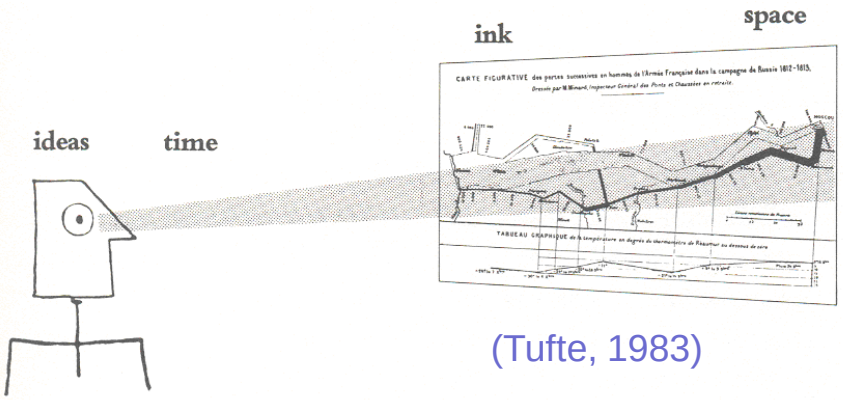
https://www.academia.edu/1380138/Information_dashboard_design_The_effective_visual_communication_of_data

Prototype: "*Portal dos indicadores, UA*"



Universidade de Aveiro
Departamento de Electrónica,
Telecomunicações e Informática

Effective Visualization and evaluation

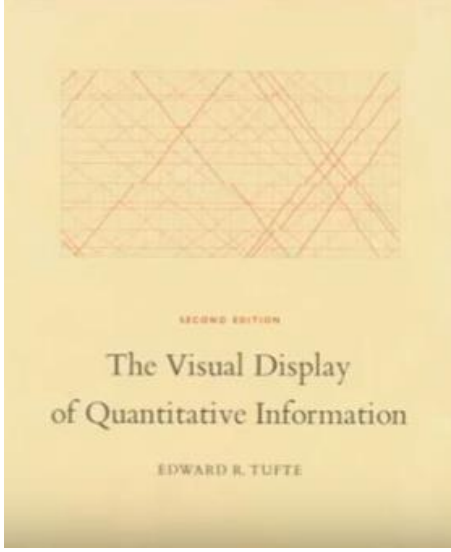


(Tufte, 1983)

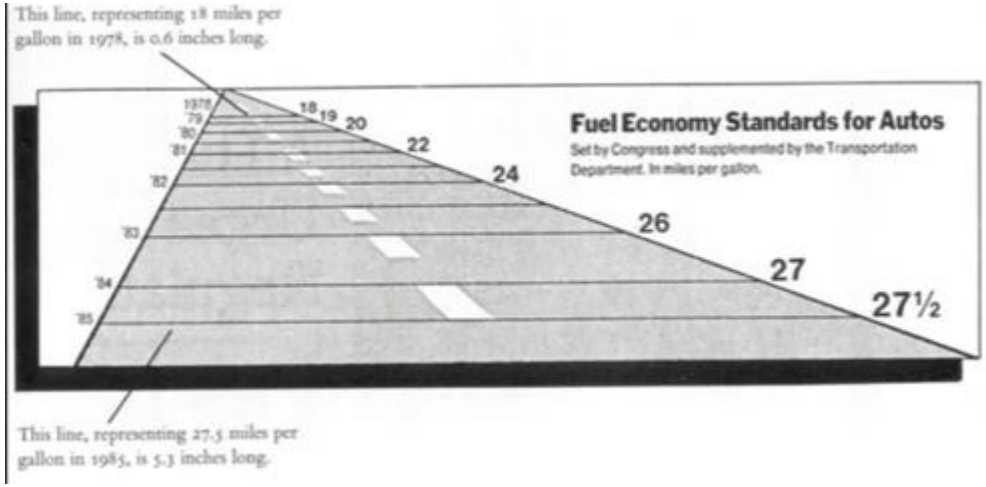
Effective visualization

Implies saying the **truth** about the data

Tufte presents a lot of commented examples in his book:



Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983

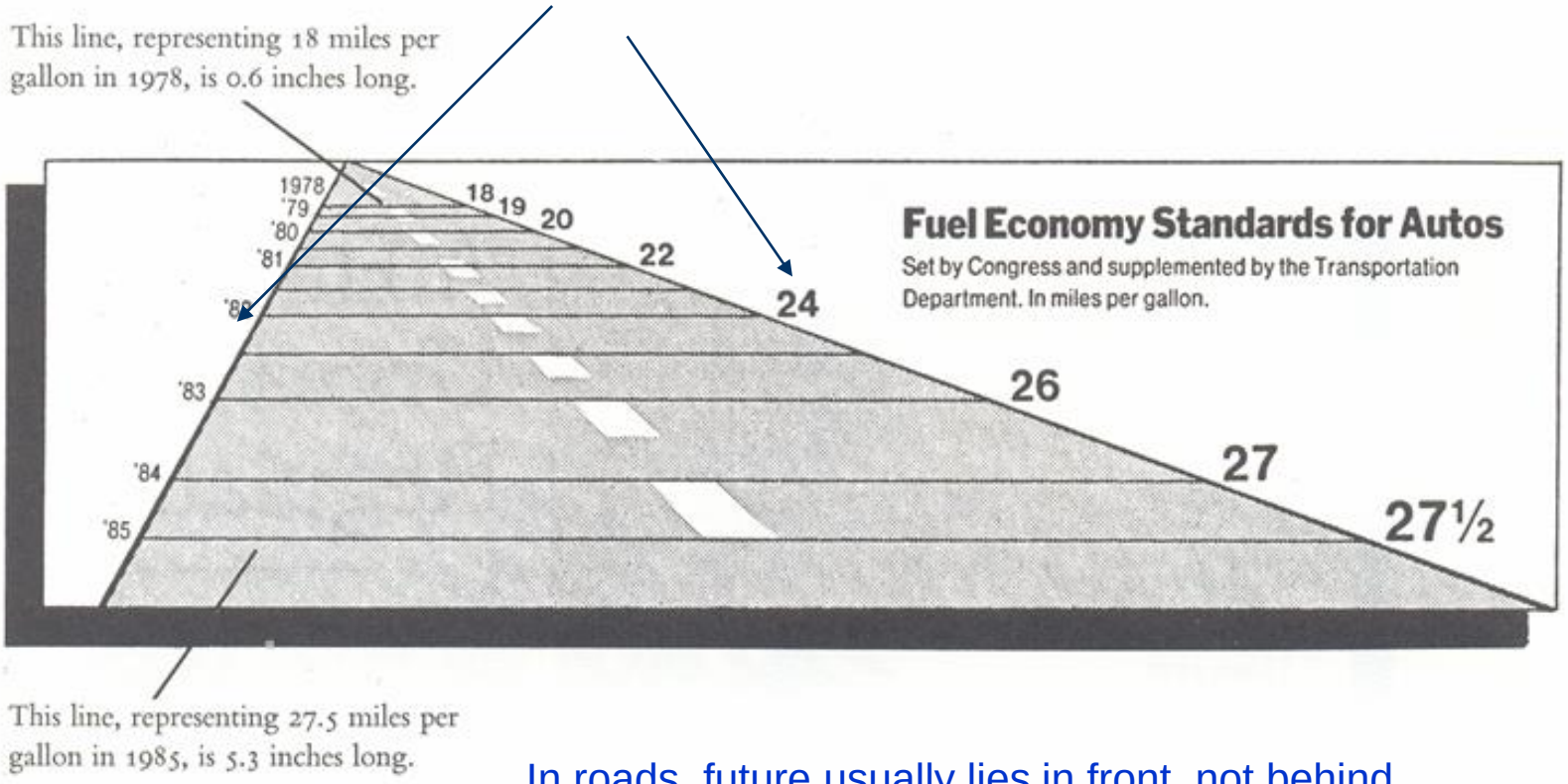


There are methods to evaluate visualizations that should be used along the process of creating a visualization

https://infovis-wiki.net/wiki/Lie_Factor

this example has several problems:

Legends have a constant size in one side and variable in the other



In roads, future usually lies in front, not behind

- Perception varies among people and with

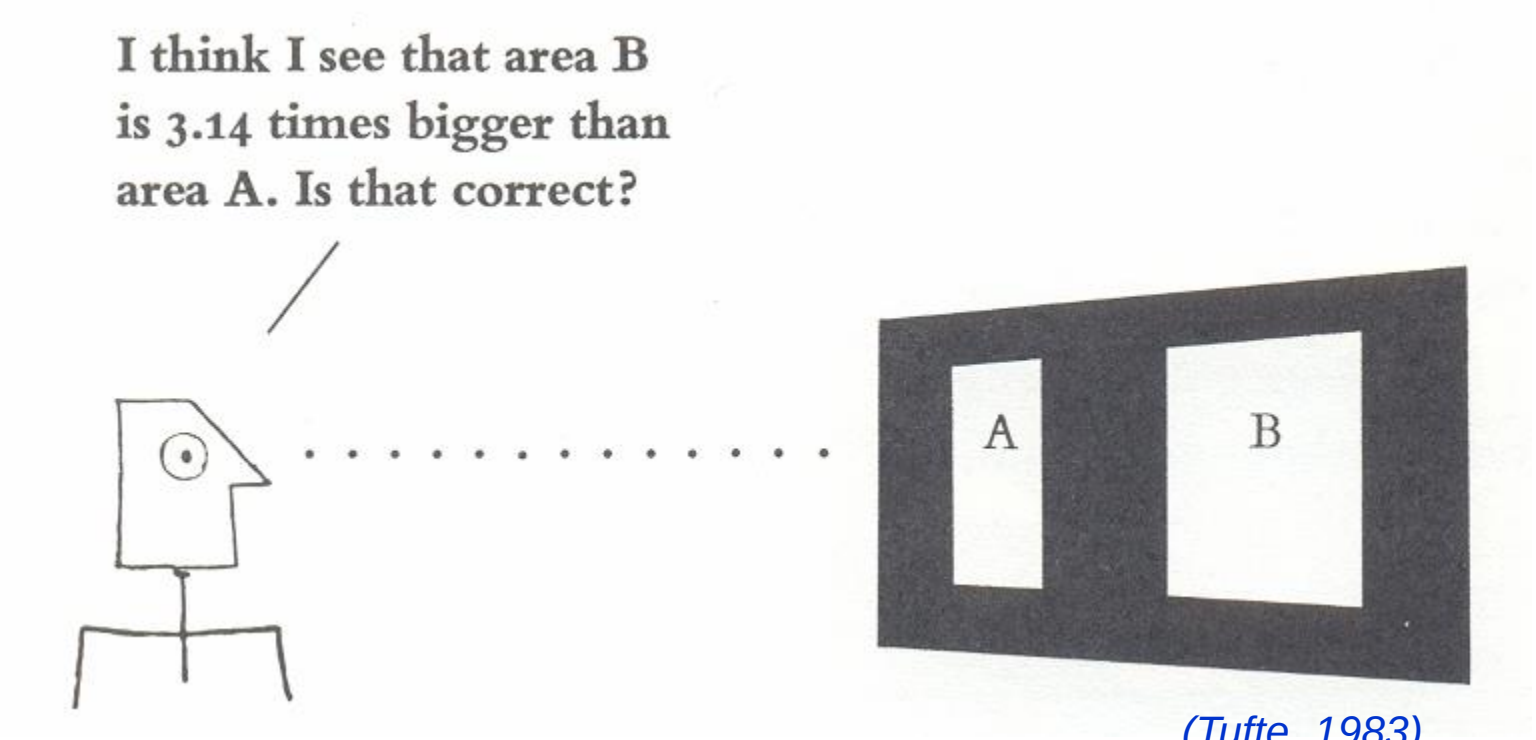
  - context

  - experience

(Tufte, 1983)

• How do we know that the visual image represents the underlying numbers?

• One way to try to answer these questions is to conduct experiments on the visual perception of graphics



(Tufte, 1983)

- What to do when we want to represent data in a graphic ?

- According to Tufte there are two fundamental principles to get graphical integrity:

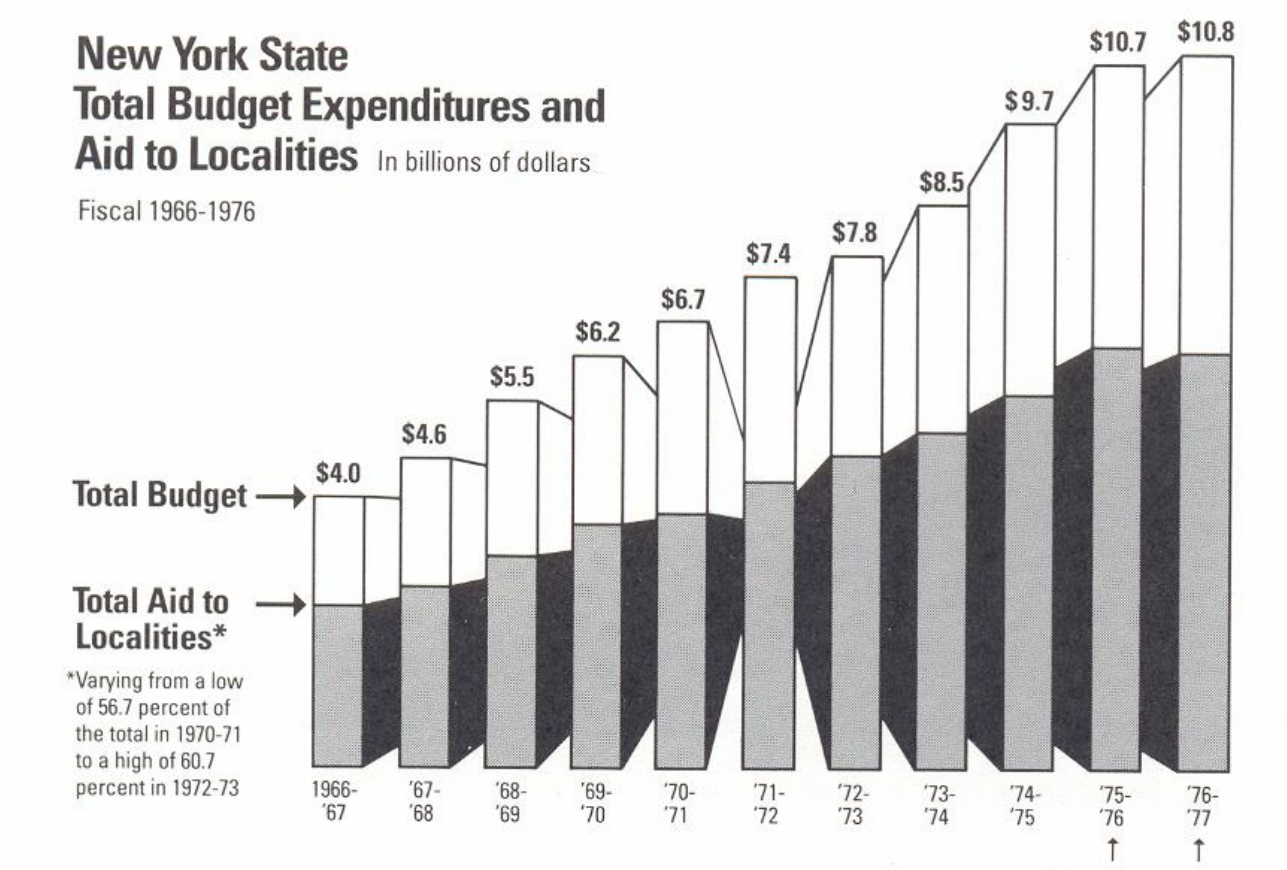
  - represent numbers, as physically measured on the surface of the graphic itself, directly proportional to the numerical quantities represented

  - Clear and thoroughly label to defeat graphical distortion and ambiguity

  Note:
  Visual representations must be **tested** as to their efficiency and efficacy for the target users to perform their tasks

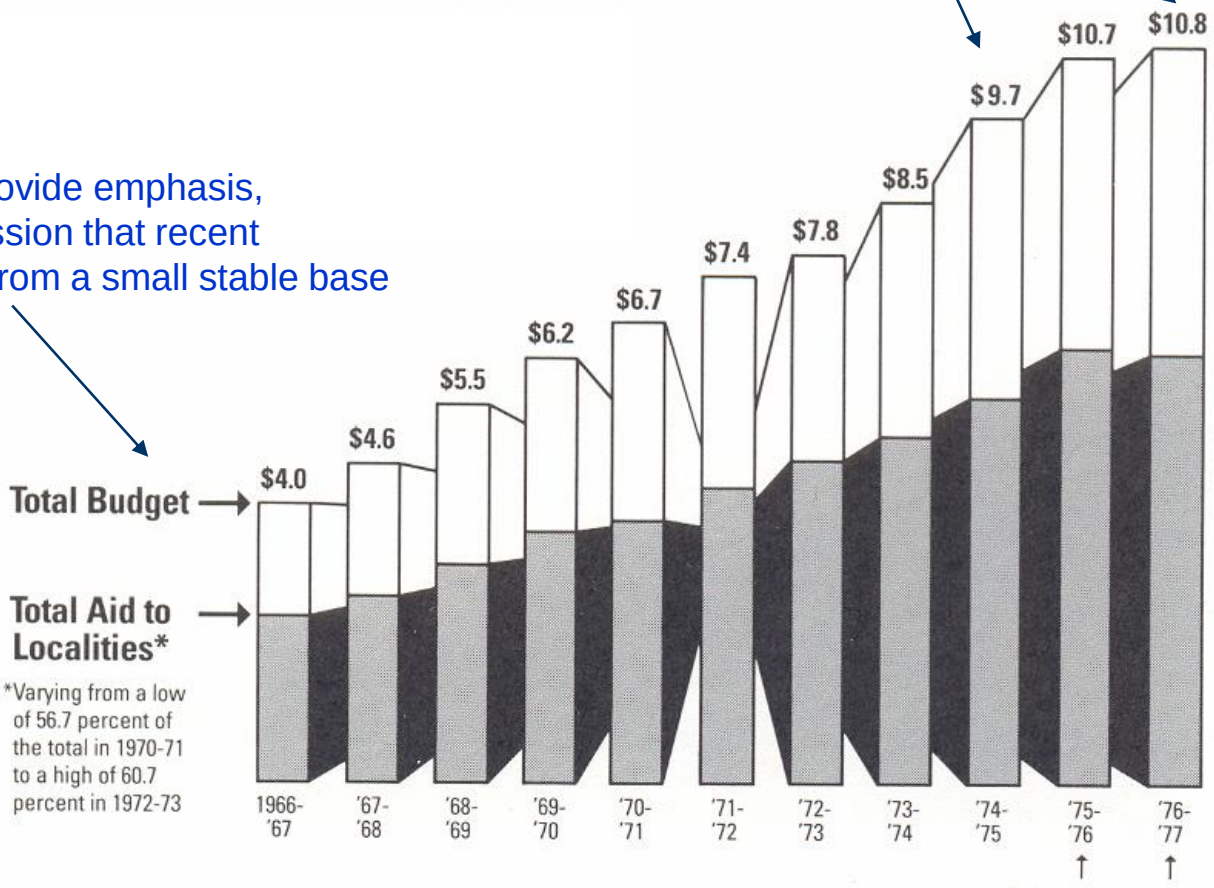
Effective visualization: anatomy of poor examples

Another example having several issues:



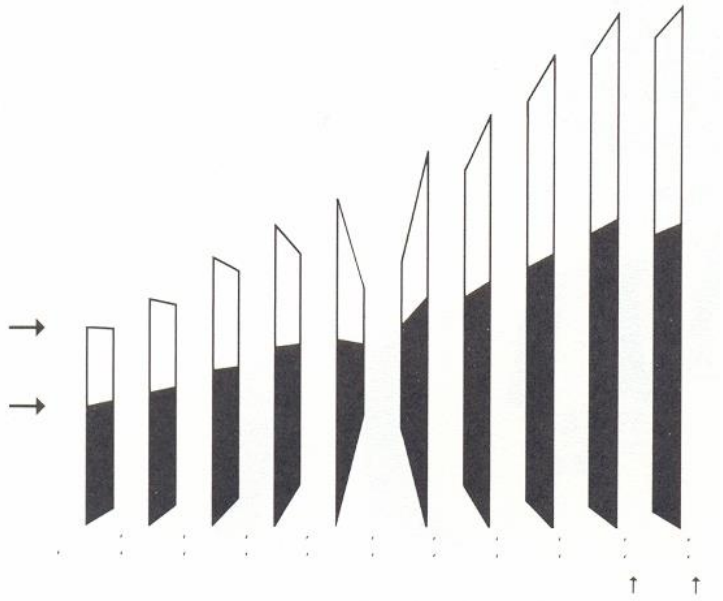
These three parallellipeds have been placed in an optical plane in front of the other eight, creating the image that the newer budgets tower over the older ones

Horizontal arrows provide emphasis, Encorage the impression that recent years have shot up from a small stable base

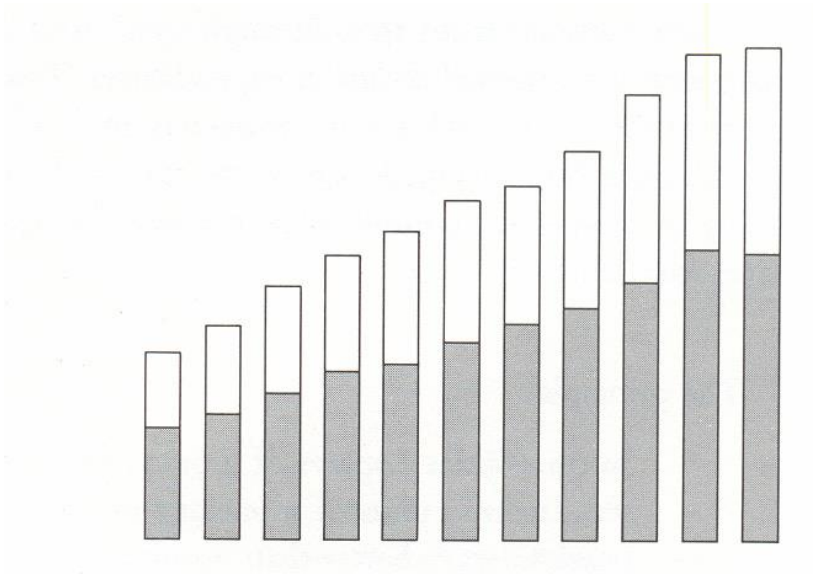


Arrows pointing straight up emphasize recent growth

Leaving behind the distortion



we have a calmer view:

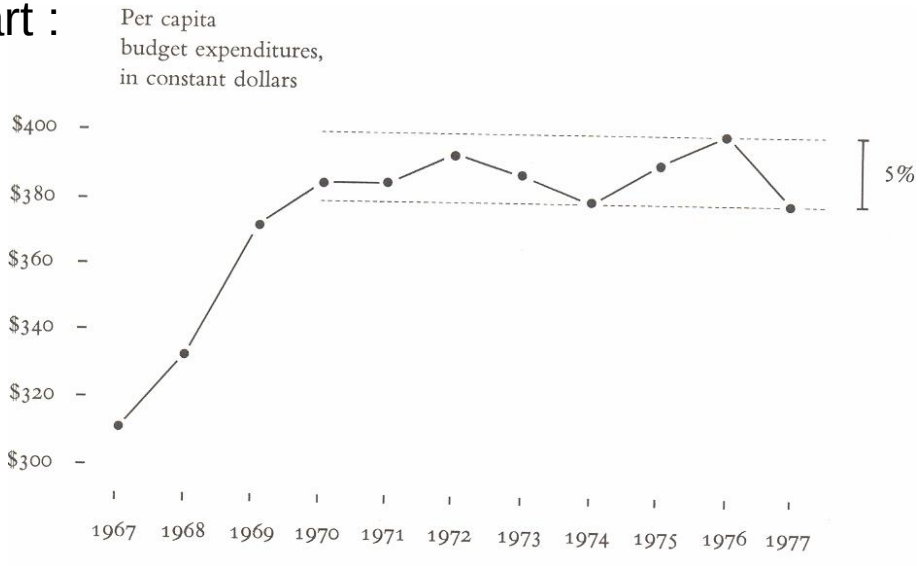


Two statistical lapses also bias the chart :

- Population increased10%

- there was substantial inflation

Final result →



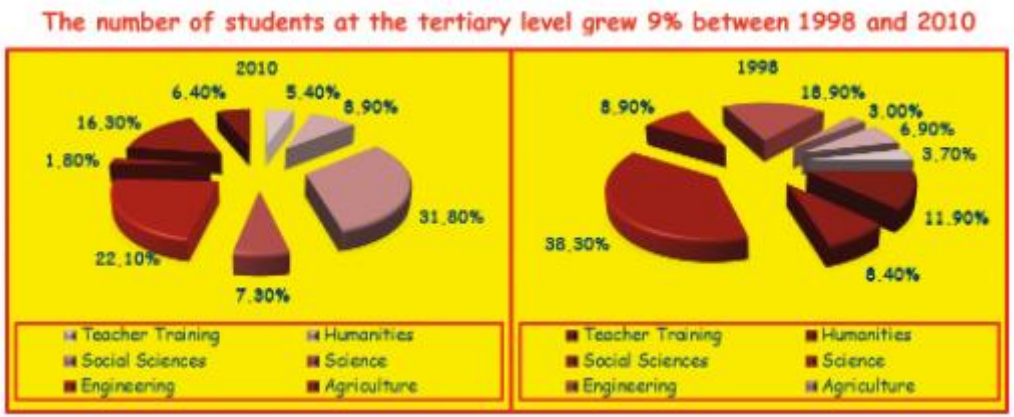
Effective Visualization: another poor example

Consider the following questions to be answered based on a data set:

- Which area of study grew the most?
- Which area of study decreased the most?
- How did Humanities behave?
- How many areas of study are increasing and how many are decreasing?

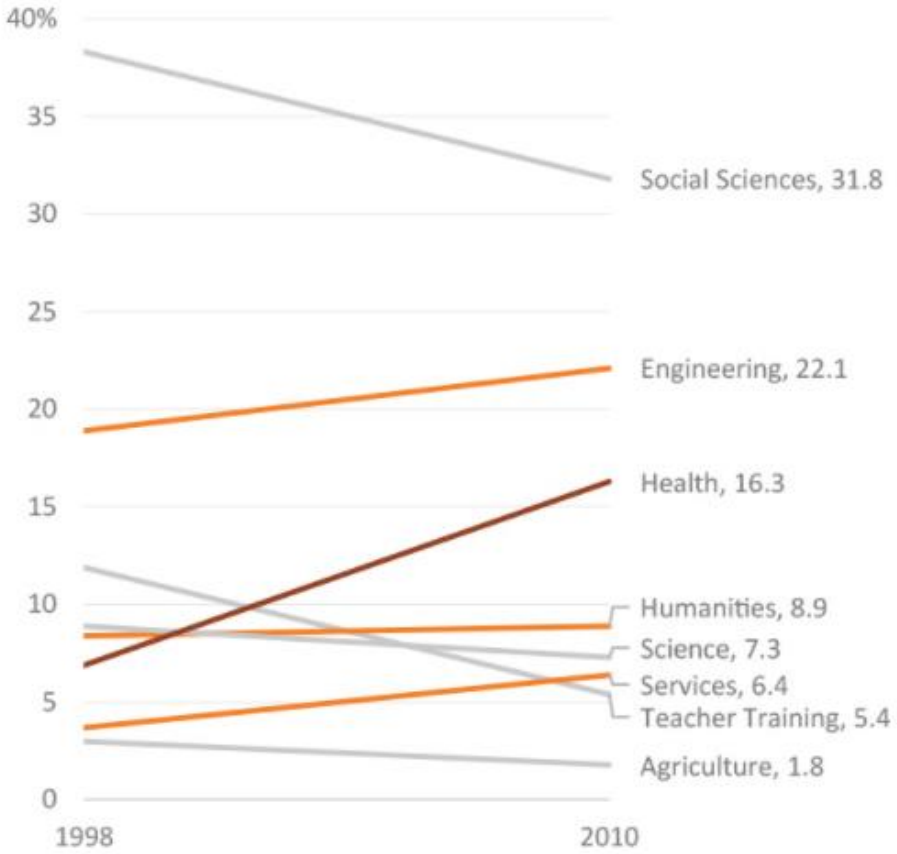
These pie charts have several issues:

- Chart type
- Time direction
- 3D effect
- Exploded slices and n. of slices
- Color usage …



The “graphenstein” (Camões, 2016)

A better way to answer the questions and provide insights from the same data:

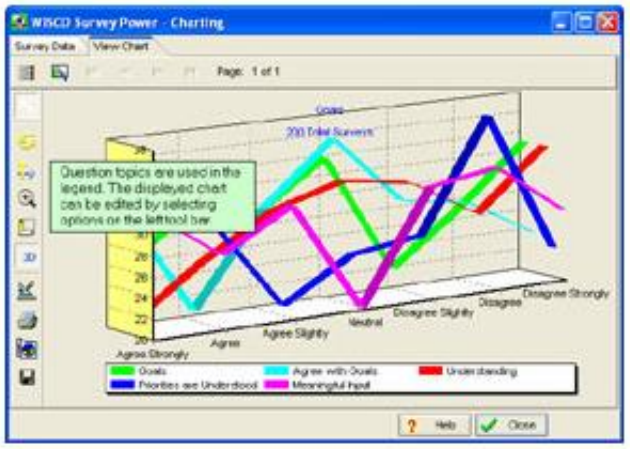


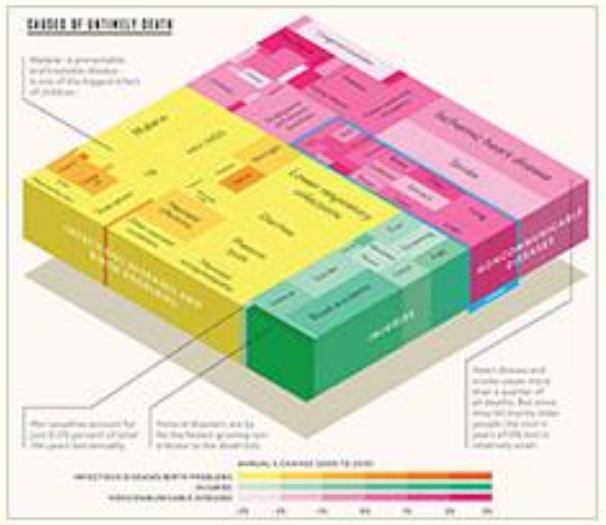
(Camões, 2016)

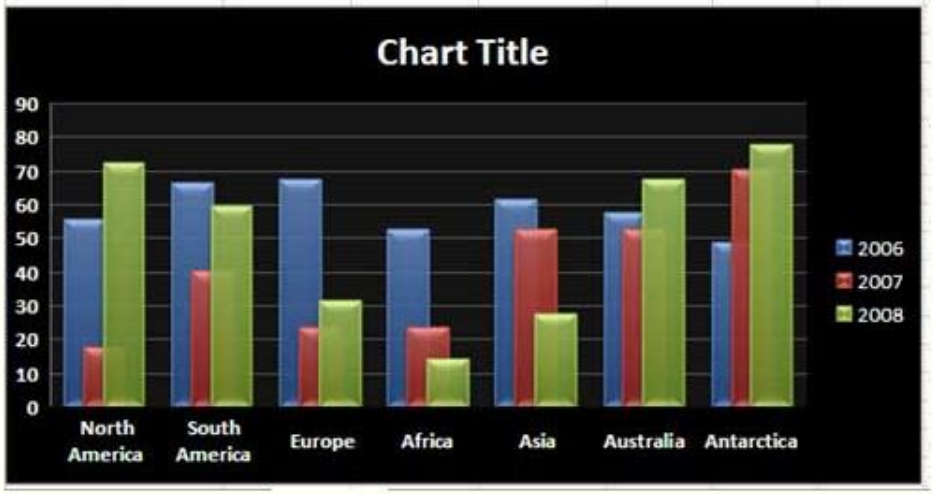
Hard-to-spot variations in the pie charts are now obvious:

- Line slopes display changes clearly
- color is used to make the chart easier to read.

Effective visualization: more poor examples analyzed







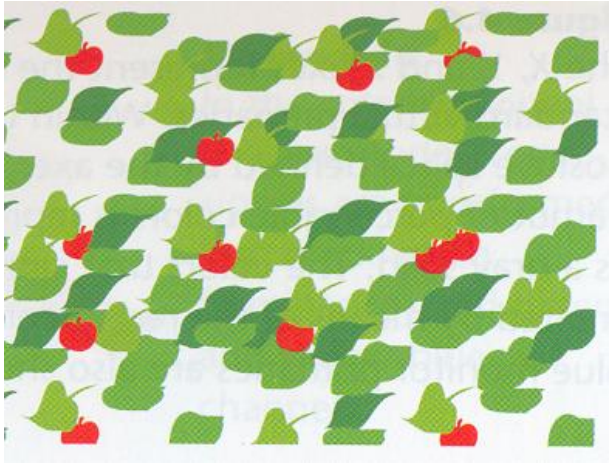
Remember: There are methods to evaluate visualizations that should be used along the process of creating a visualization

http://www.perceptualedge.com/examples.php

Color may help a lot in some tasks

…

How many cherries?

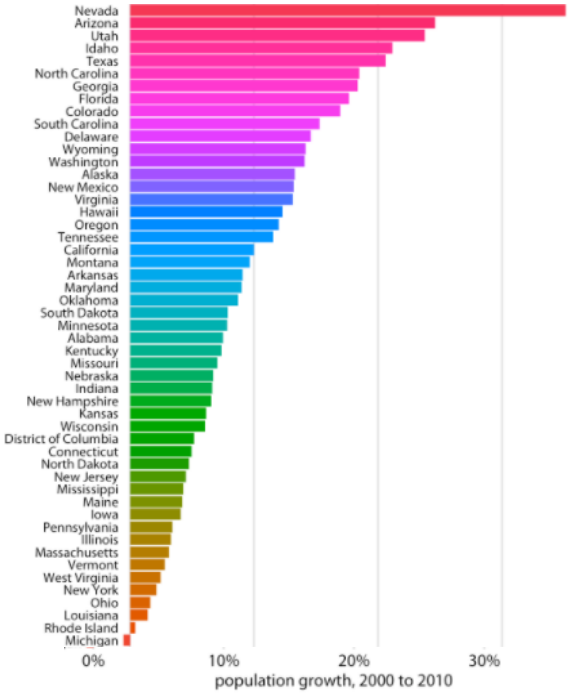


It may hinder if not properly applied …

It should be applied sparingly and carefully

should serve a purpose, be clear, and not distract

Common pitfalls of color use

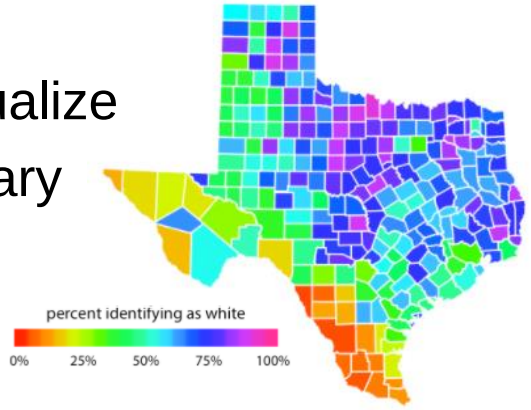


- **Encoding too much or irrelevant information**

  (the coloring of states serves no purpose, colors are over saturated; it is distracting)

- **Using non-monotonic color scales to encode data values**

  (The rainbow color scale is not an appropriate to visualize continuous data values; it tends to emphasize arbitrary features of the data)



- **Not designing for color-vision deficiency**

https://clauswilke.com/dataviz/color-pitfalls.html

Color Blindness

- The most common type is deuteranopia (“daltonism”)
  (affects ~10% of men; is residual in women)

- There are color blindness simulators and colorblind-friendly color scales

http://www.color-blindness.com/coblis-color-blindness-simulator

Normal vision

Deuteranopia

Tritanopia

http://www.colourblindawareness.org/

Guidelines to use color

- Design first in black and white; then color apply sparingly and carefully
- Use direct labeling instead of colors when you need to distinguish between more than about eight categorical items
- Avoid large filled areas of overly saturated colors. They make it difficult for the reader to carefully inspect the figure
- To make sure figures work for people with cvd, don't just rely on specific color scales. Instead, test figures in a cvd simulator
- Do not use blue to color small objects (will be difficult to see)

https://clauswilke.com/dataviz/color-pitfalls.html

- Visualizations produced to be used by many users **should be evaluated** to improve their effectiveness

- There are several methods that can be used to evaluate:

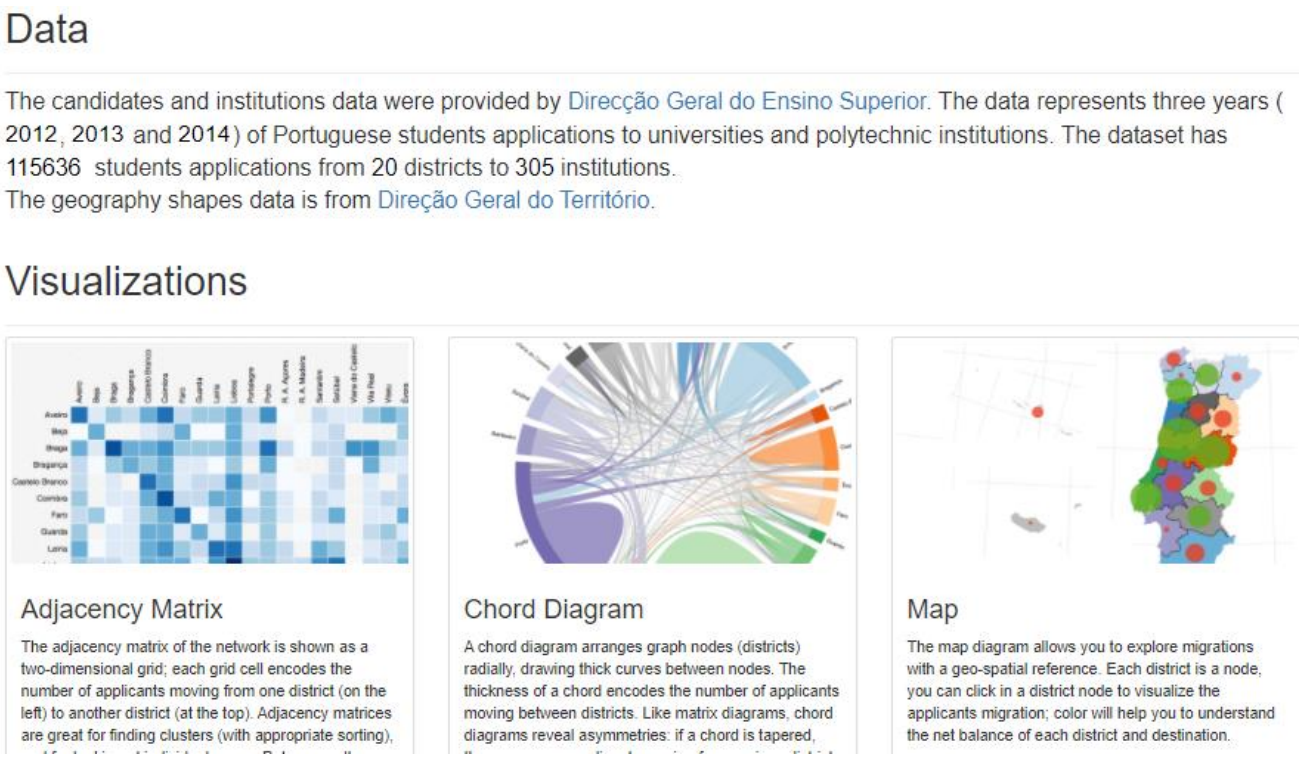
  - **Analytical** (involving only analysts, not users)
    e.g. heuristic evaluation

  - **Empirical** (involving users)
    user tests

At least review your visualizations considering the questions they should help users to answer and ask some target users to use them and provide feedback… Then, improve your visualizations!

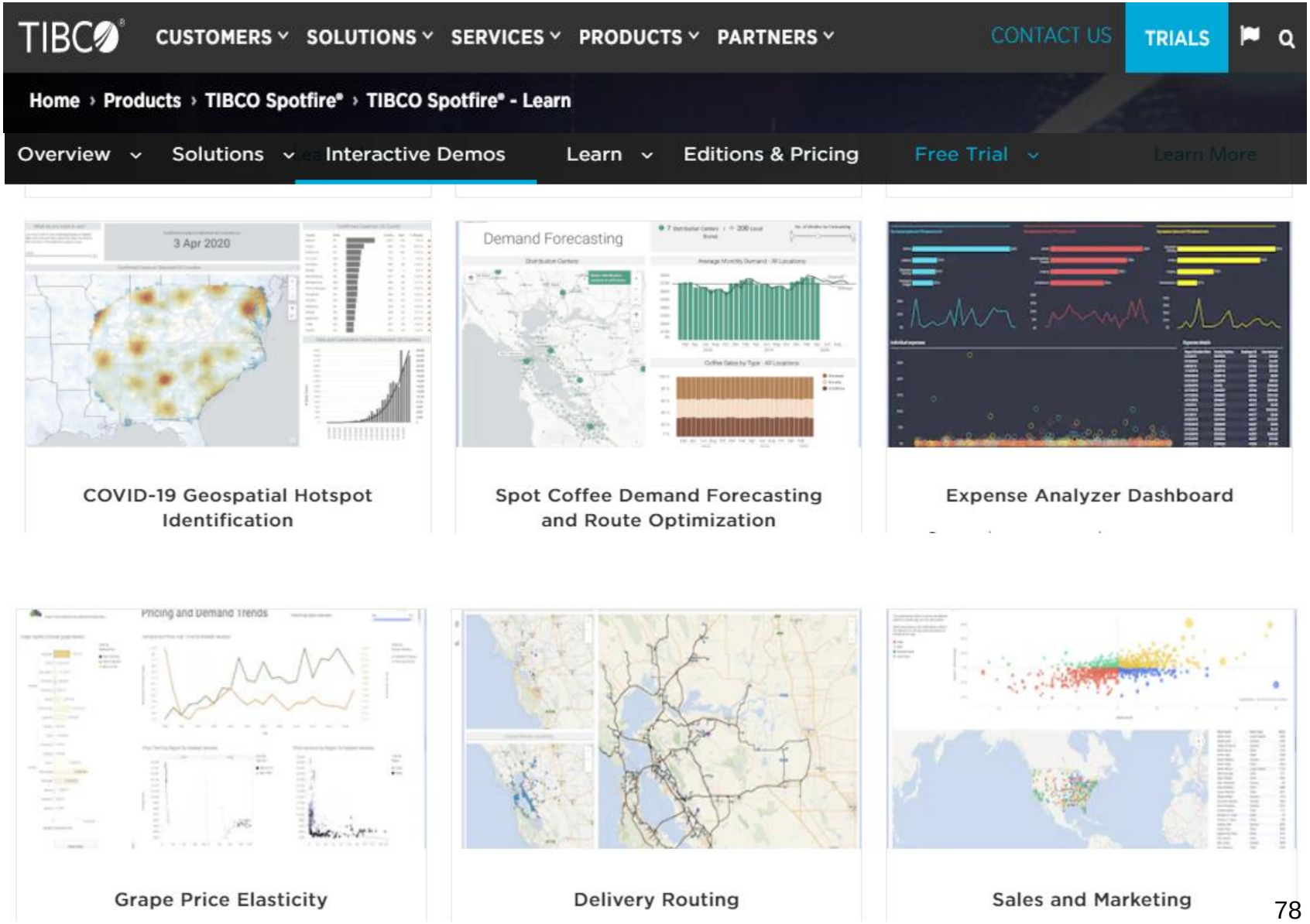
Example: usability test of a visual exploration app

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



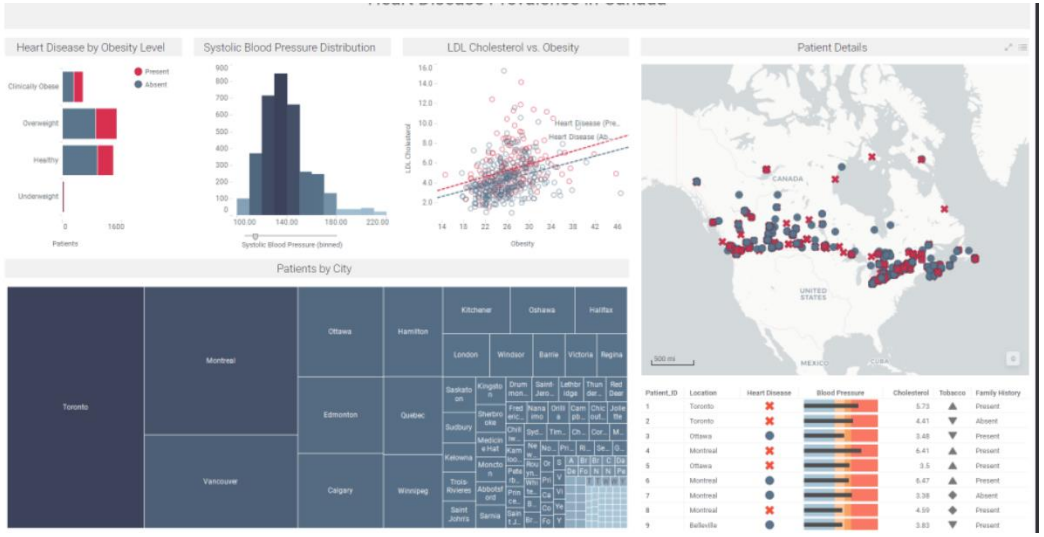
https://forms.ua.pt/index.php?r=survey/index&sid=489227

Examples of visual data mining simple applications



https://www.tibco.com/products/tibco-spotfire/learn/demos

Visualization Tools

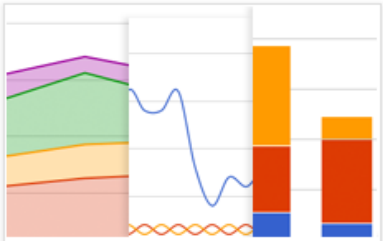


https://www.tibco.com/products/tibco-spotfire/learn/demos

Visualization S/W

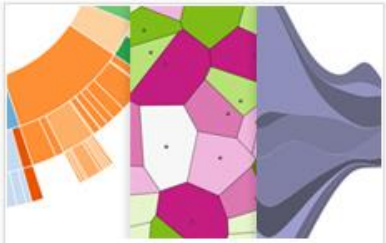
- There are a lot of tools, of different types and with different purposes from very simple to very complex …

- 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.

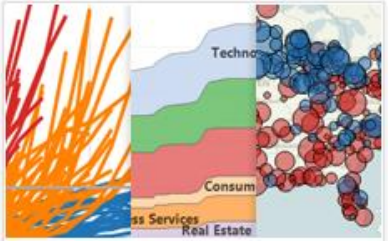
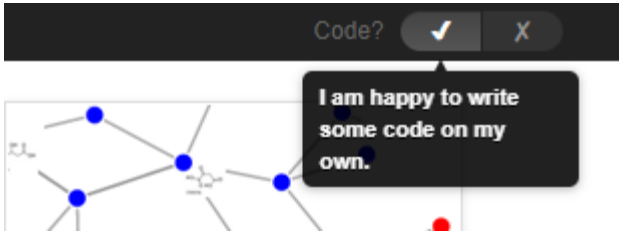


Tableau Public

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/



Which Visualization Tool?
It depends …

- Exploratory (discover patterns, 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?
- Public data or private 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

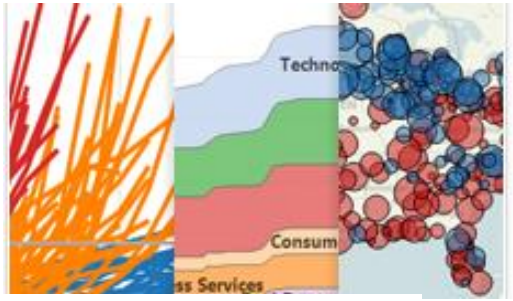
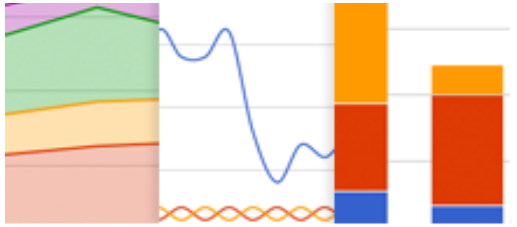


Tableau Public

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.

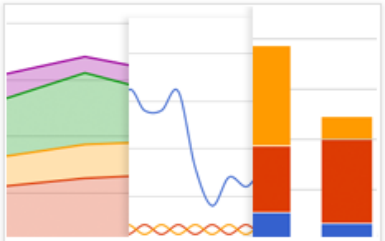
D3.js

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

Visualization Tools

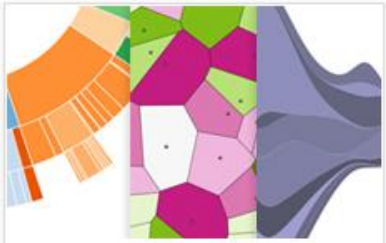
- There are a lot, of different types and with different purposes
  from very simple to very complex …

- Some interesting 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.

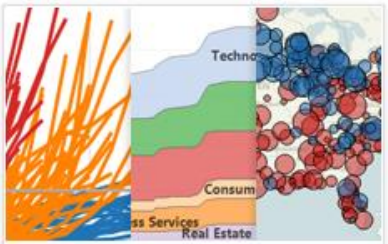
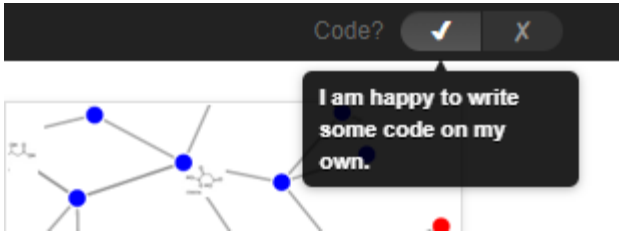


Tableau Public

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

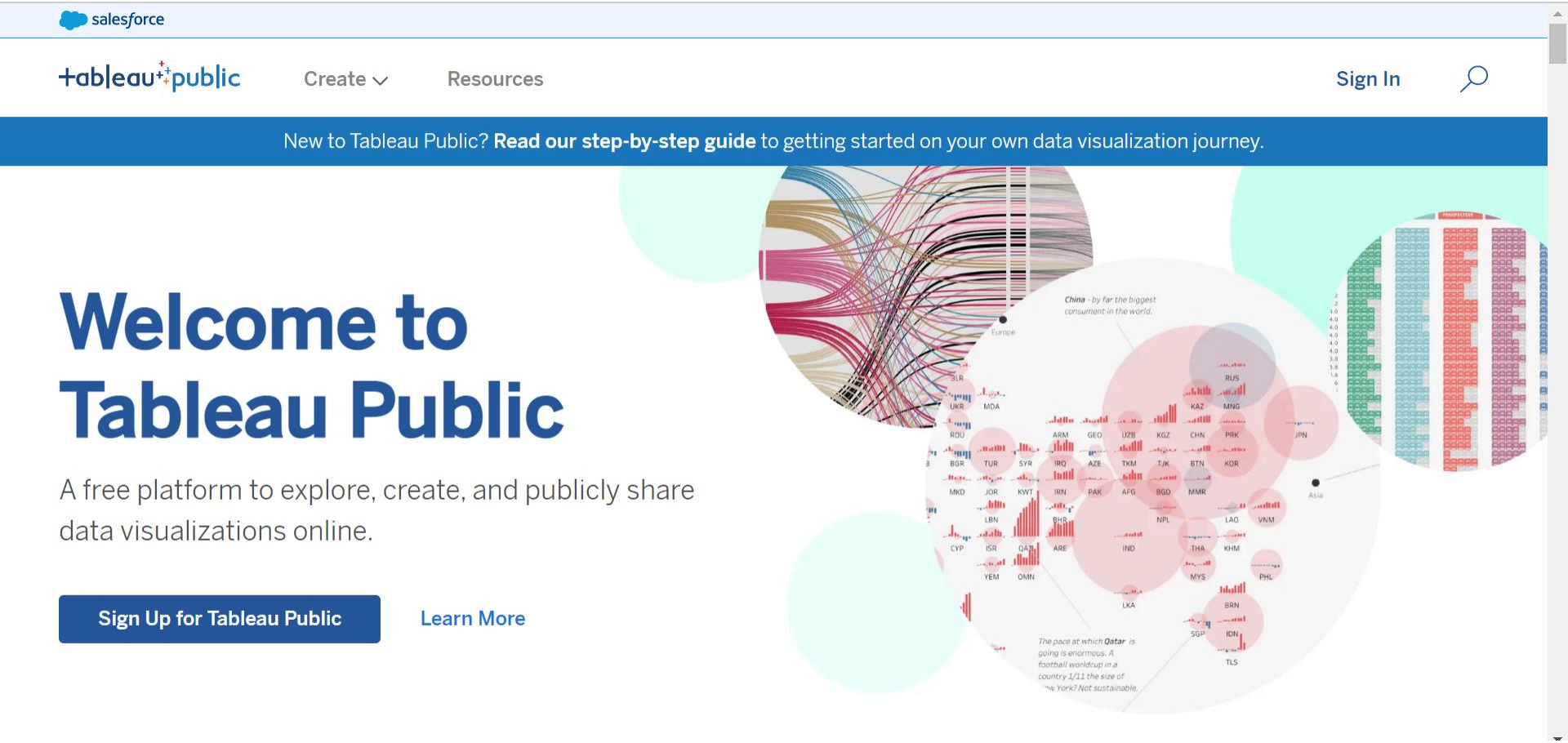
see e.g.
https://www.kdnuggets.com/2020/05/top-10-data-visualization-tools-every-data-scientist.html
http://selection.datavisualization.ch/



2022 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms

Note: the report is available upon request

An application you may use for your Visualizations (they will be public…)



https://public.tableau.com/app/discover

Example of a Tableau Public application

LITERACY

By Jenny Nguyen

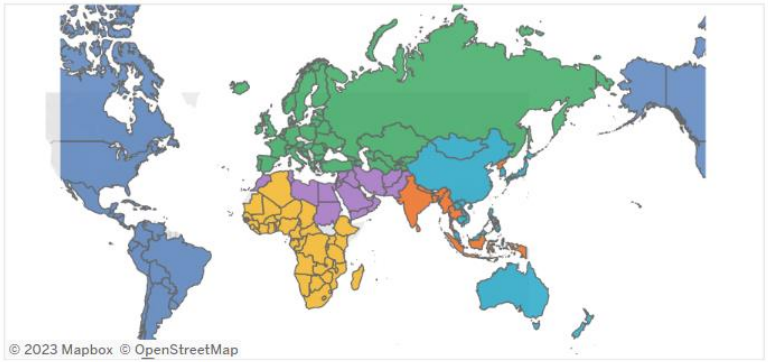
INTRODUCTION

Literacy is an education measurement of a demographic. In this dashboard, we aim to explore how literacy rates correlate with these measurements of a demographic: life expectancy, fertility rate, GDP per capita and GDP. We will also look at youth literacy and literacy by sex.

The timeline for this visualisation is from the year 2000 and onwards.

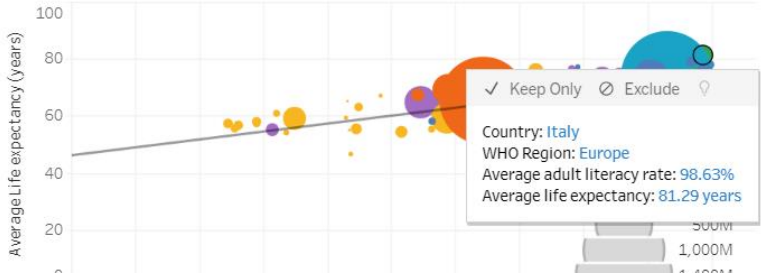
Below are the regions that will be used throughout the dashboard, based on WHO classification.

Africa | Americas | Eastern Mediterranean | Europe | South-East Asia | Western Pacific



GLOBAL LITERACY RATE

This map shows the literacy rate of each country. Many countries have a literacy rate of 90% and above, which means that more than 90% of the population can read and write. Unfortunately, the literacy rate of some countries are still below 50%, particularly in Africa.



ADULT LITERACY RATE vs. LIFE EXPECTANCY

This plot shows the correlation between adult literacy rate and life expectancy. As we can see, there is a positive correlation between the two. The higher the adult literacy rate, the higher the life expectancy.

Many countries in Europe have a high adult literacy rate and life expectancy. This is largely due to good socioeconomic factors in the region, such as having good healthcare systems.

However, there are countries in Africa and South-East Asia having high adult literacy but low life expectancy. This is due to poorer

[Literacy - #IronVizStudentEdition 2022 | Tableau Public]

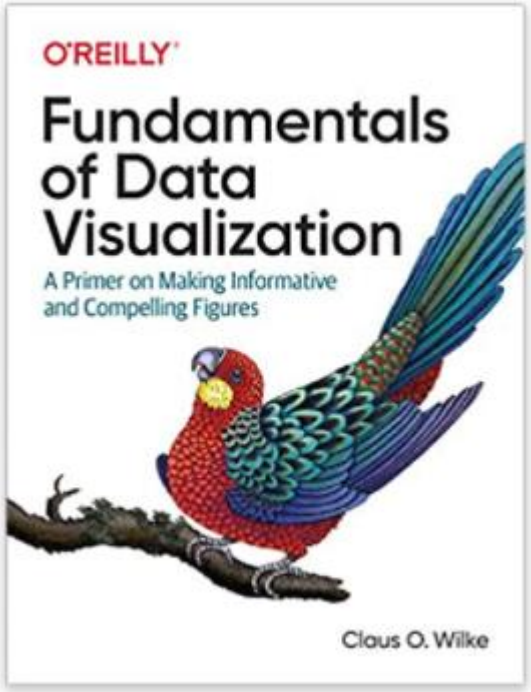
Learning resources

Introductory 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://clauswilke.com/dataviz/ https://learning.oreilly.com/library/view/fundamentals-of-data/9781492031079/

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

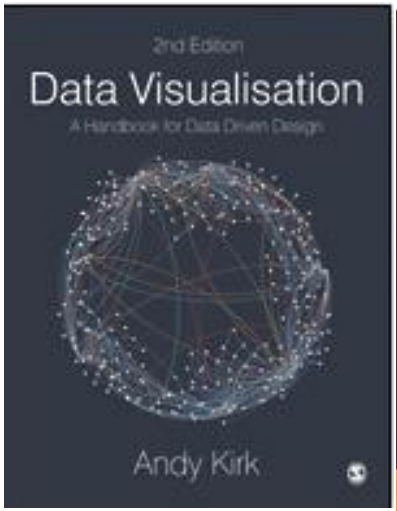


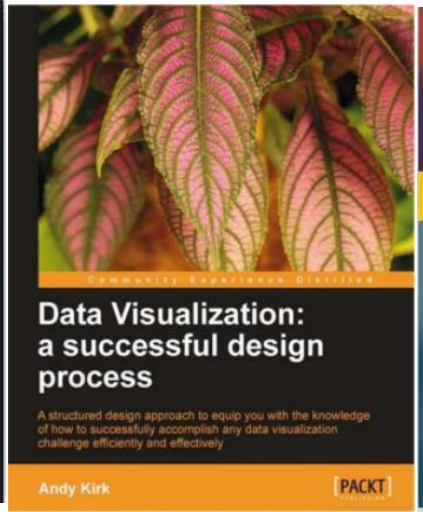
More advanced bibliography

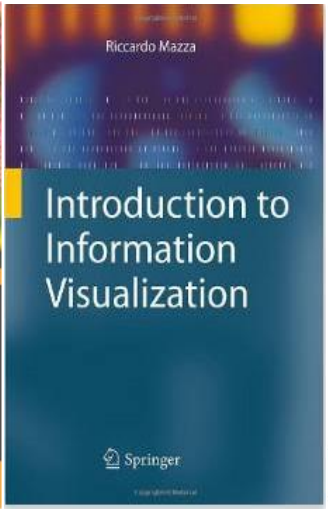
- Kirk, A., *Data Visualisation A Handbook for Data Driven Design*, 2nd. Ed., Sage, 2019
- Spence, R., *Information Visualization, An Introduction*, Springer, 2014
- Ware, C., Information Visualization, Perception for Design, 3rd ed., Morgan Kaufman, 2012
- Mazza, R., *Introduction to Information Visualization*, Springer, 2009
- Kirk, A., *Data Visualization: A successful design process*, Pack Publishing, 2012

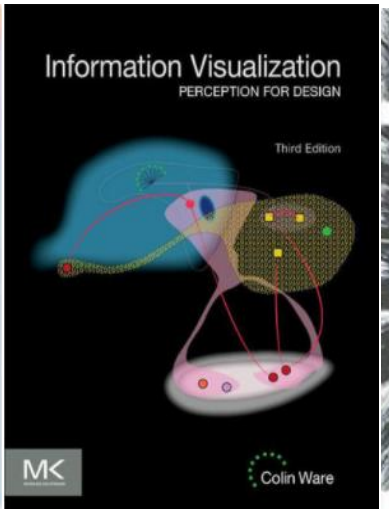
Explore other books available at the playlist:

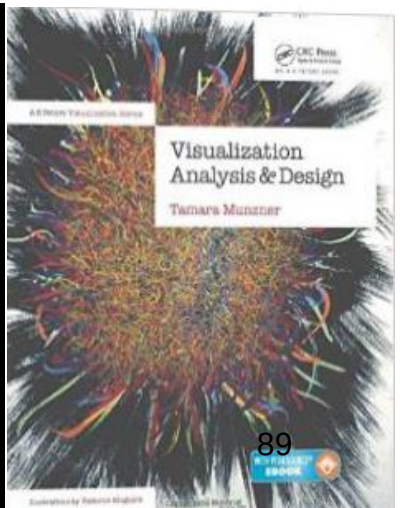
- https://learning.oreilly.com/playlists/dfba5bea-f37f-49cd-a964-f46626b24299/











InfoVis Seminal Bibliography

- 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.
- Bederson, B. , B. Shneiderman, *The Craft of Information Visualization: Readings and Reflections*, Morgan Kaufmann, 2003
- Keim, D., “Information visualization and visual data mining,” *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 1–8, 2002
- Card, S., J. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think,* Morgan Kaufmann, 1999

Information Design Bibliography

- Few, S., *Now you see it: An Introduction to Visual Data Sensemaking*, Analytics Press; 2nd ed., 2021
- Cairo, A., *The Functional Art*, New Riders, 2012
- Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983
- Tufte, E., *Envisioning Information,* Graphics Press, 1990
- 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

Bibliography: some InfoVis interesting papers

- 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.

- Heer, J.,, Michael Bostock, M., Ogievetsky, V., “A Tour through the Visualization Zoo A survey of powerful visualization techniques, from the obvious to the obscure”, ACM Queue, vol. 8, no.5, 2010.

- Endert, A., Ribarsky, W., Turkay, C., Wong, B., Nabney, I., Blanco, I., 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.

- Y. Lu, Y., Garcia, R., Hansen, Gleicher, B., Maciejewski, R., “The State-of-the-Art in Predictive Visual Analytics,” *Comput. Graph. Forum*, vol. 36, no. 3, pp. 539–562, 2017.
- Skarbez R., Polys N., F., Ogle J.T., North, C., Bowman, D. A., “Immersive Analytics: Theory and Research Agenda”, *Frontiers in Robotics and AI*, vol 6, 2019.

- A. Kirk, “A Recipe of Capabilities for Pursuing Expertise in Data Visualization: A Practitioner's Perspective”, *IEEE Comput. Graph. Appl.*, vol. 41, no. 1, pp. 58 - 62, 2021.

To probe further
Scientific Journals/Conferences

IEEE Transactions on Visualization and Computer Graphics

IEEE Computer Graphics and Applications

Computer Graphics Forum

Computers & Graphics

Information Visualization

IEEE Computer Graphics and Applications – Visualization Viewpoints

IEEE Vis (http://ieeevis.org/)

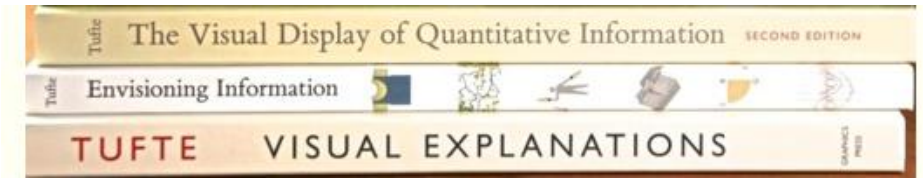
Eurovis (https://www.eurovis.org/)

Information Visualization (http://www.graphicslink.co.uk/IV2023/)

Interesting links

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

Visual Business Intelligence
for enlightening analysis and communication

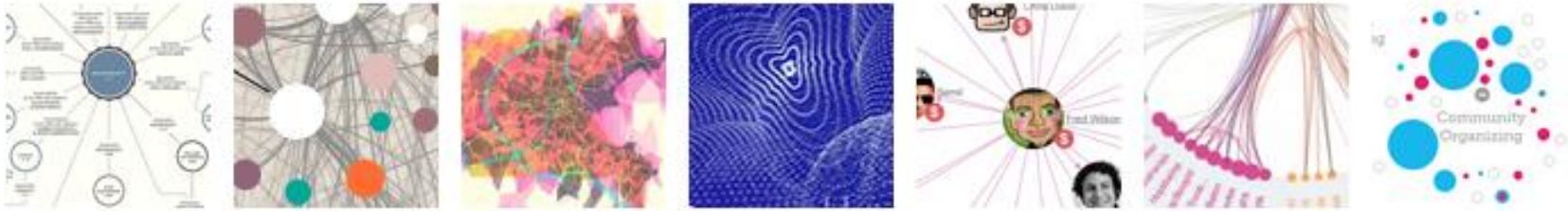


Interesting links

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

- http://seeingdata.org/

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



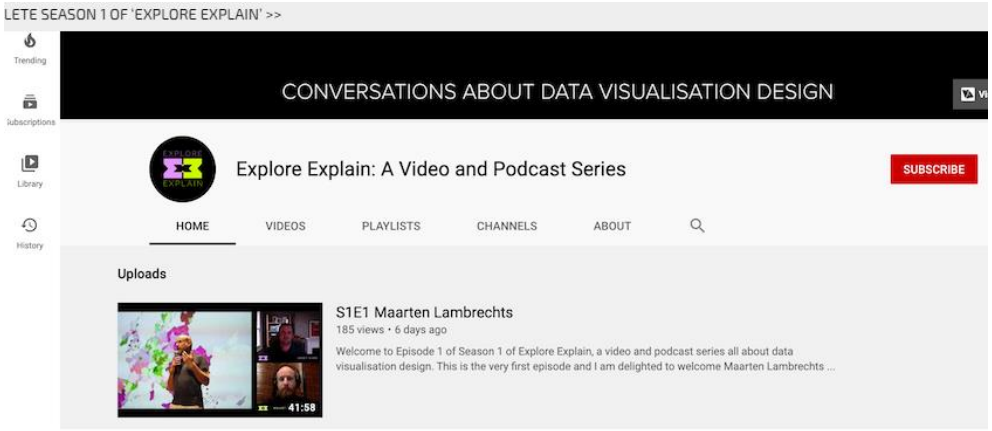
Podcasts and Online courses

https://datastori.es/

DATA STORIES

A podcast on data visualization with Enrico Bertini and Moritz Stefaner

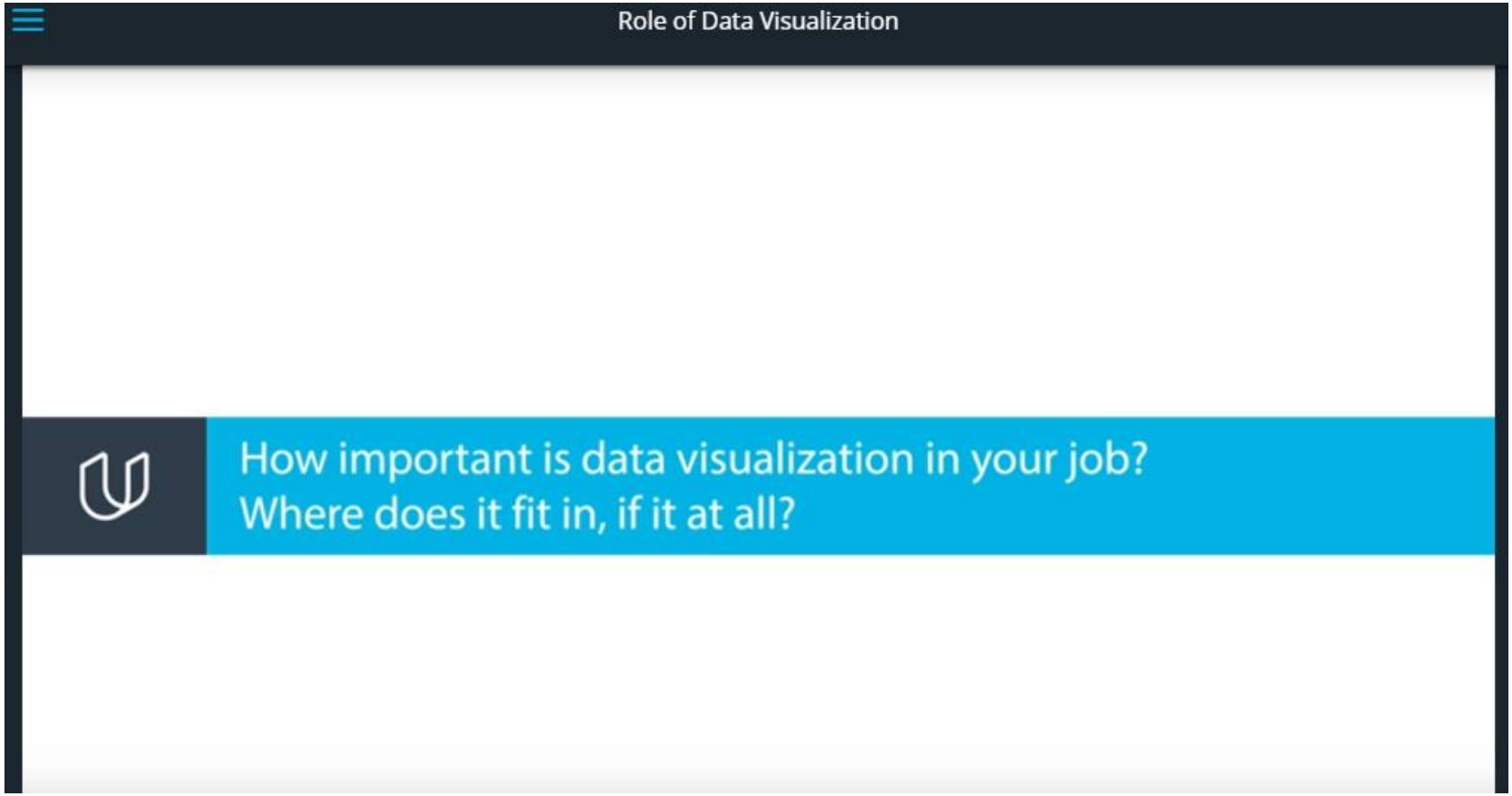
https://www.visualisingdata.com/podcast/



https://www.coursera.org/specializations/information-visualization

NYU

Interviews with Netflix Data Scientists



https://classroom.udacity.com/courses/ud404/lessons/9239573934/concepts/91687840320923

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

...

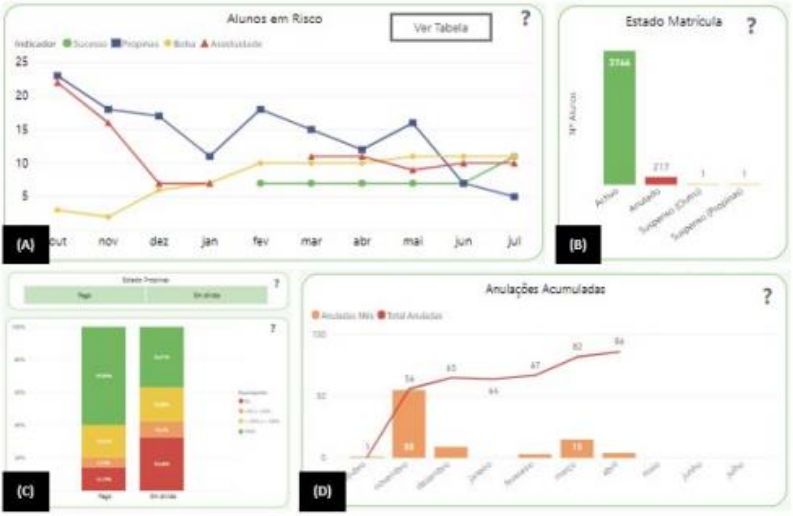
- it is impossible to build a model unless I understand what the data means

- You may do some boxplots, scatterplots, trend analysis …

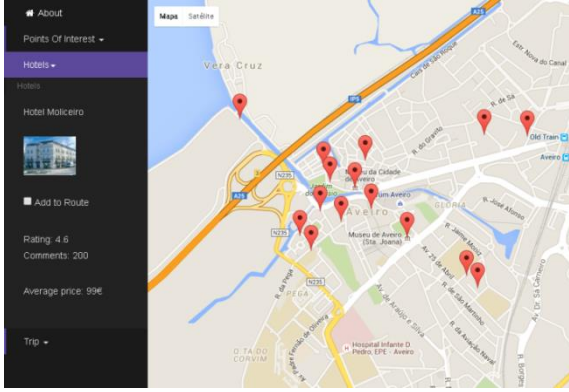
- Domain scientists play a very important role

Information Visualization **(examples “made in UA”)**

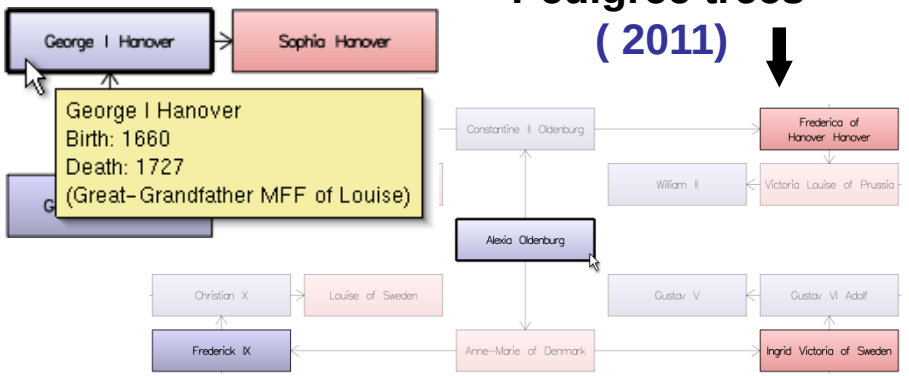
Academic data dashboard
(2020)



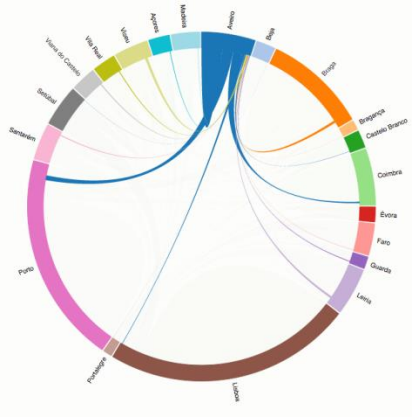
Location Routing in trip planning (2016)

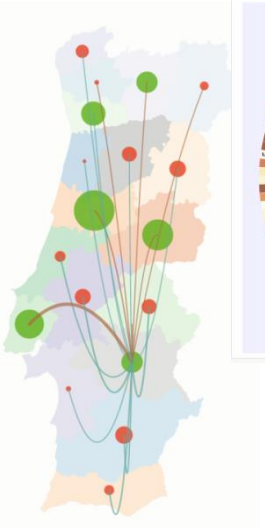


Pedigree trees
(2011)



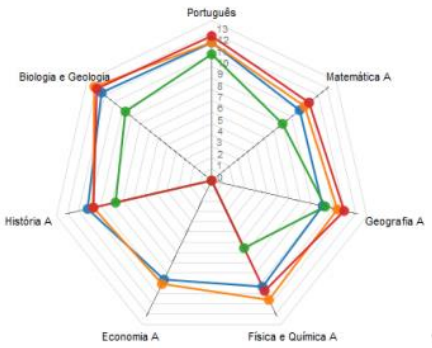
Human Migrations
(2015)

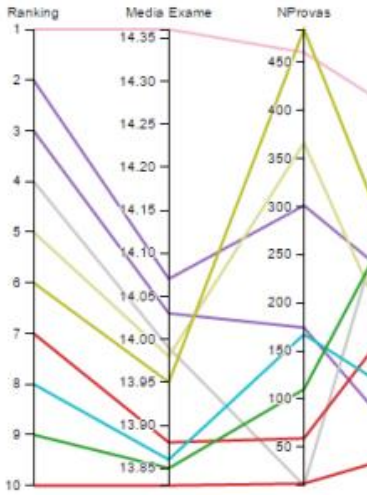




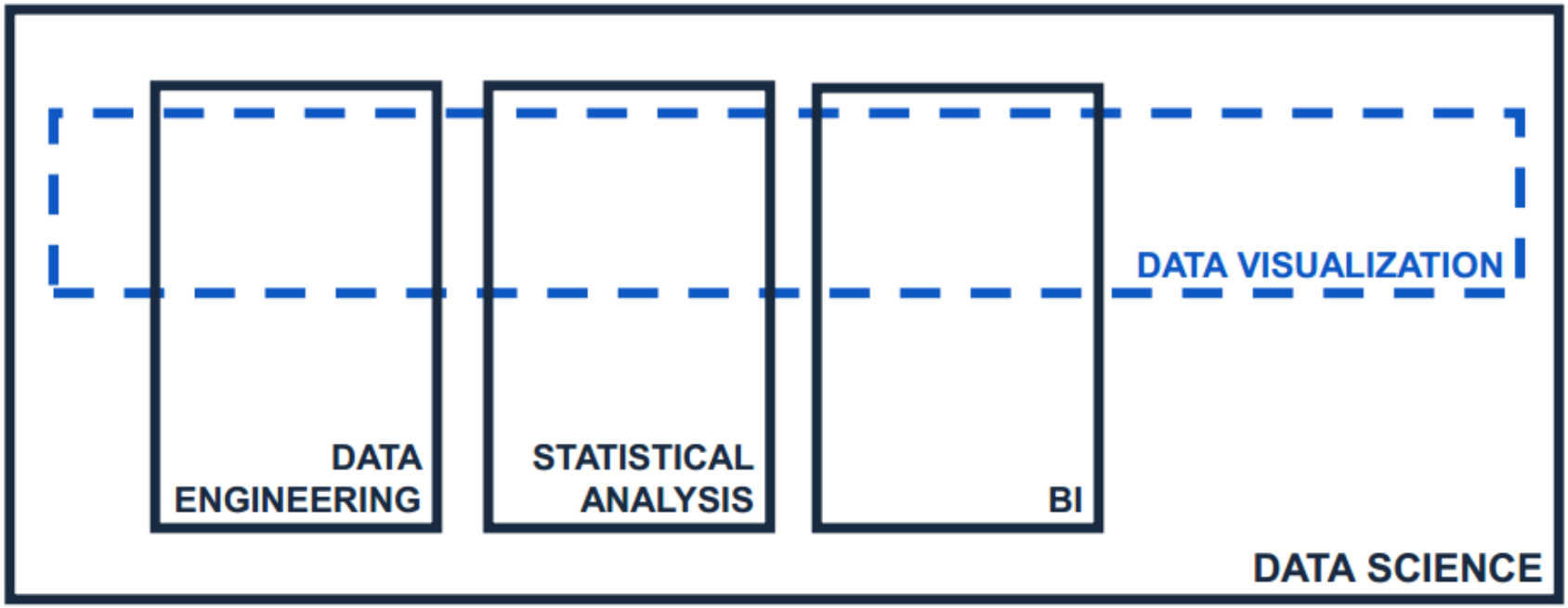
Papers Taxonomy
(2021)

School Ranking
(2016)

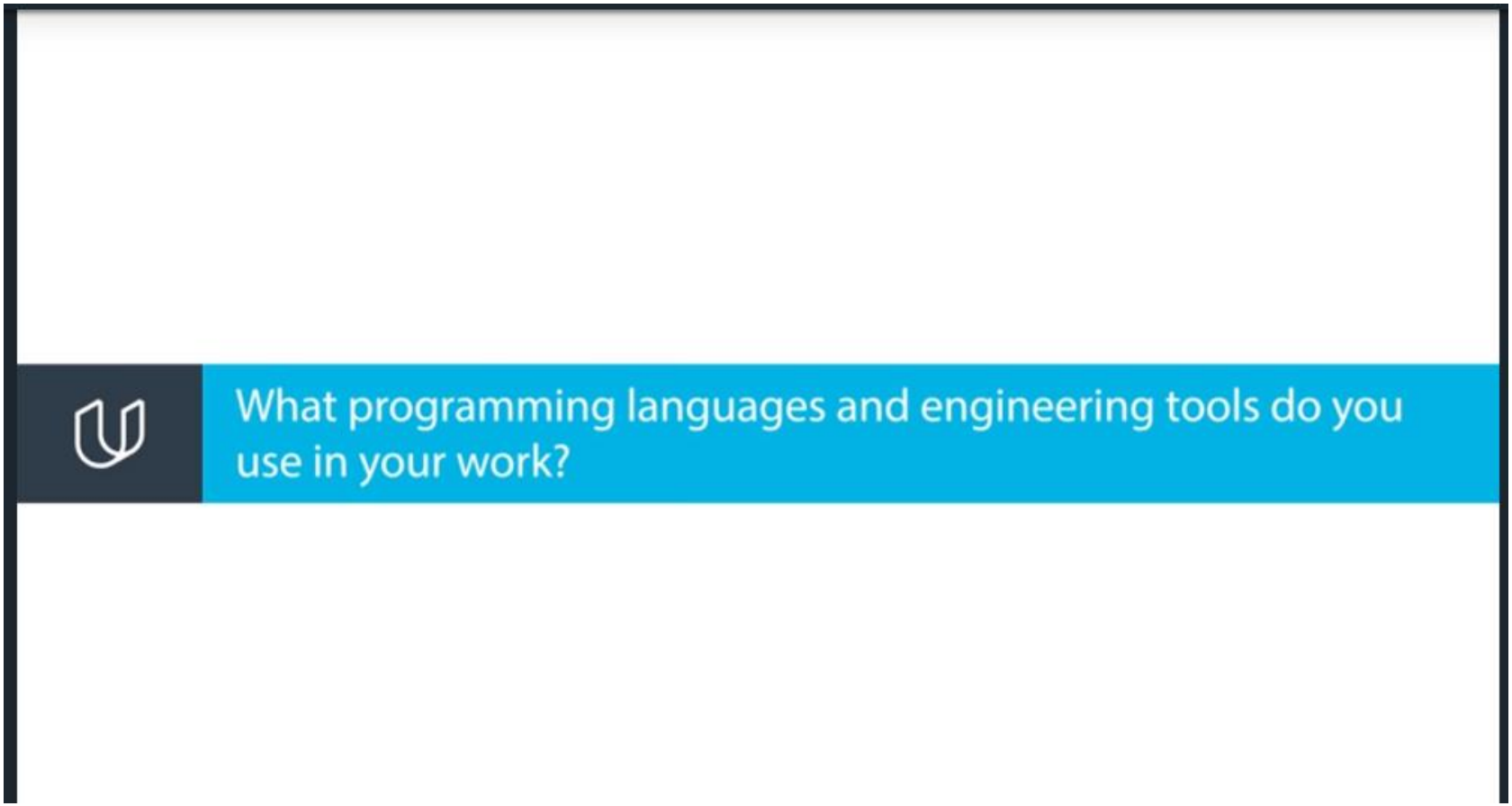




Visualization and Data Science



Interviews with Netflix Data Scientists



https://classroom.udacity.com/courses/ud404/lessons/9259930027/concepts/91687840330923