Introduction to data visualization

DATA VISUALIZATION IN TABLEAU

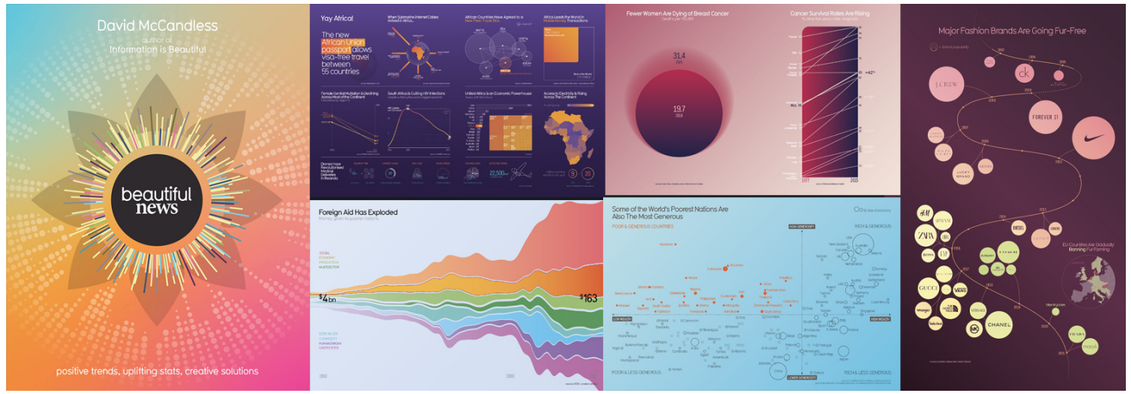


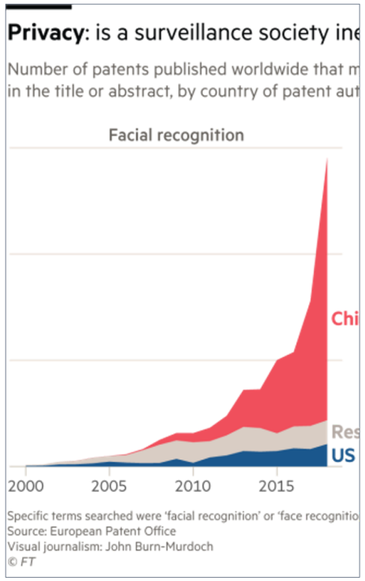
Agata Bak-Geerinck
Senior Data Analyst, Telenet

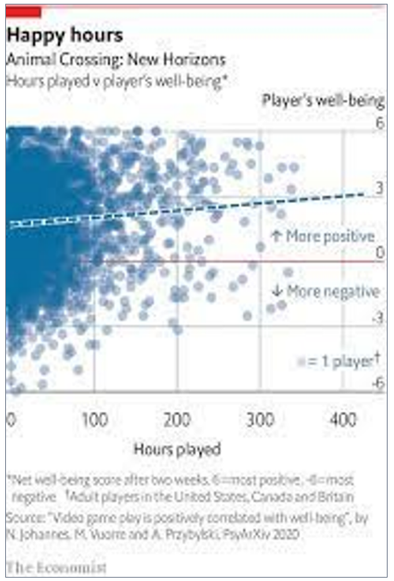
datacamp

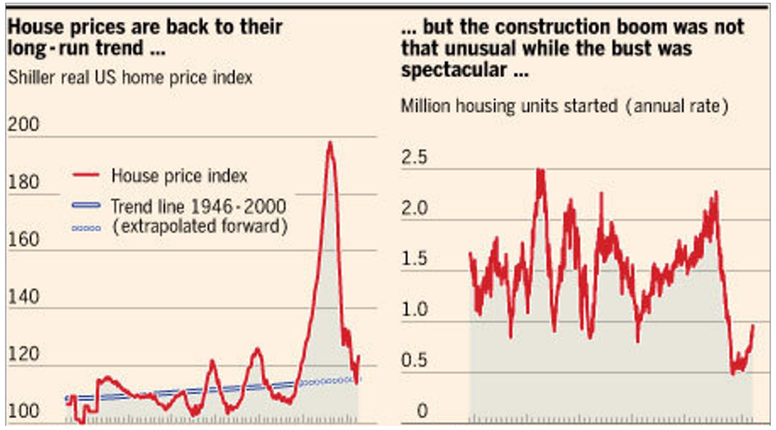
What is data visualization?

- A graphical representation of information
- Presentation of (complex) data to wide audiences
- Selection of most effective charts and elements of design
- May be considered a form of art









[1] Pictures: The Economist, The Financial Times, David McCandless books.

Why we need data visualization?

Visual information is here to stay:

- Humans - visual creatures, made for pattern & trend spotting
- Enhances understanding and retention of information
- Essential in exploratory data analysis
- Critical part of the decision-making process of data-driven organizations

90 % of the information absorbed by the brain is **VISUAL.**



Pitfalls of data visualization

What to look out for?

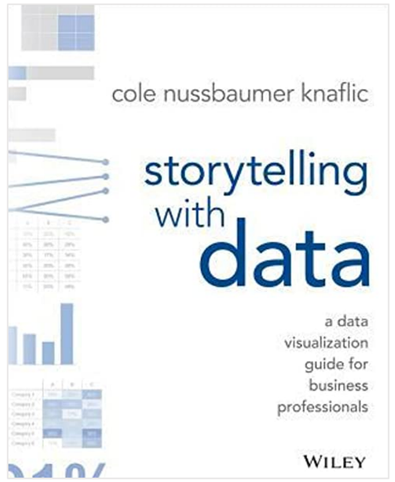
- Data is for everyone... proceed with caution
- Poor numerical literacy among non-data professionals
- Detachment between data and the context
- Lack of experience of data analysts
- Misleading visualizations

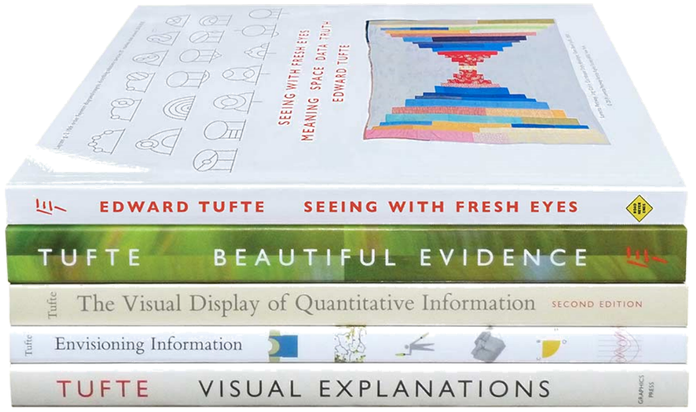


How to succeed in data visualization?

Elements of success:

- Experimenting with instant results
- Prioritizing useful over beautiful
- Always keeping the audience in mind
- Challenging oneself and investing in self-study



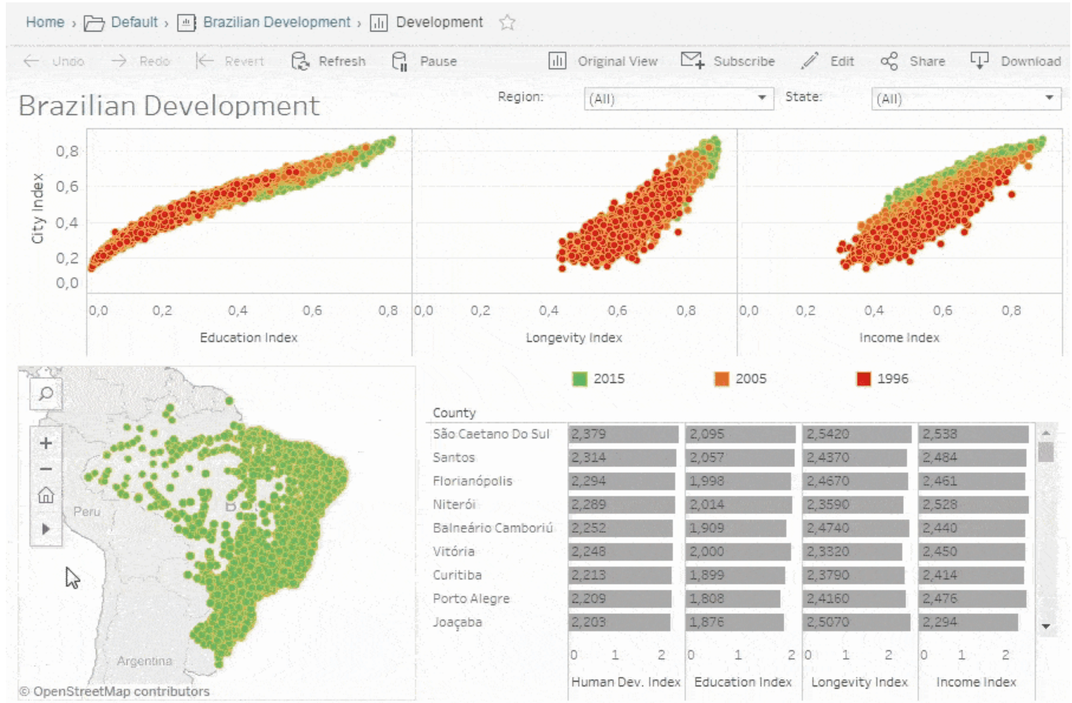


[1] Check out https://www.storytellingwithdata.com/ and accompanying book by Cole Nussbaumer Knaflic

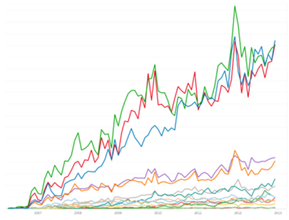
Tableau as one of leading data visualization tools

Advantages of Tableau for data visualization:

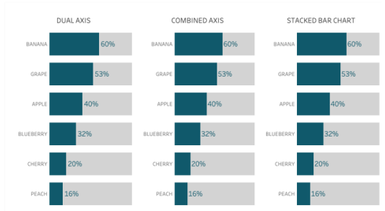
- Quick development of charts and dashboards
- Focus on simple, intuitive, and highly interactive outputs
- Plenty of formatting features and beautiful visualizations
- Responsive design for various screens and devices



A (wide) variety of charts



Line chart



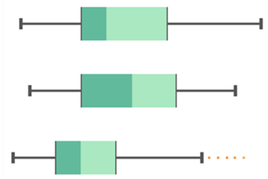
Bar chart



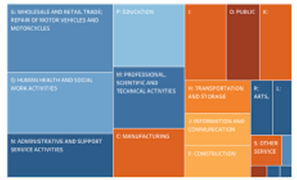
Word cloud



Heatmap



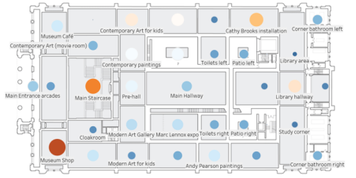
Box plot



Tree map



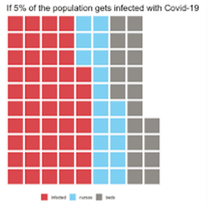
Geomap



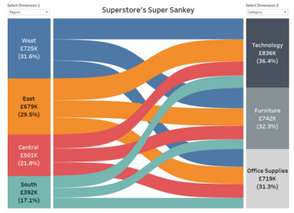
Non-geo maps



DNA chart



Waffle chart



Sankey diagram



Waterfall chart

From basic to advanced data visualization

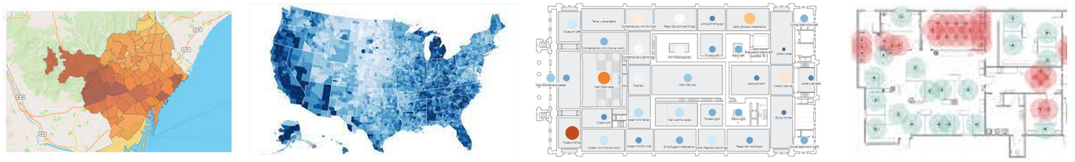
Chapter 1: Introduction to Data Visualization



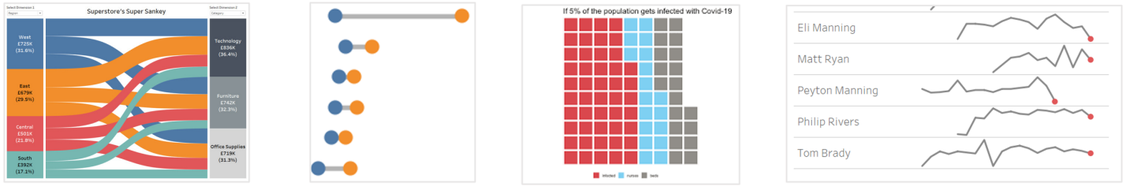
Chapter 2: Best Practices in Visualization



Chapter 3: Visualization of Spatial Data



Chapter 4: Advanced Charts and Storytelling



Our dataset

Internet Movie Database (IMDb)

- Top films and series across the years
- Actors and film people
- Genres, media types
- Ratings, votes, scores, and gross earnings



Let's practice!

DATA VISUALIZATION IN TABLEAU

Building first visualizations in Tableau

DATA VISUALIZATION IN TABLEAU



Agata Bak-Geerinck
Senior Data Analyst, Telenet

datacamp

Let's practice!

DATA VISUALIZATION IN TABLEAU

More advanced visualizations

DATA VISUALIZATION IN TABLEAU

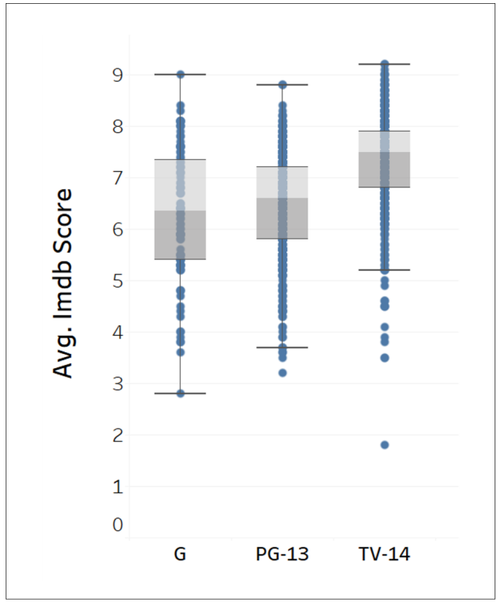


Agata Bak-Geerinck
Senior Data Analyst, Telenet

datacamp

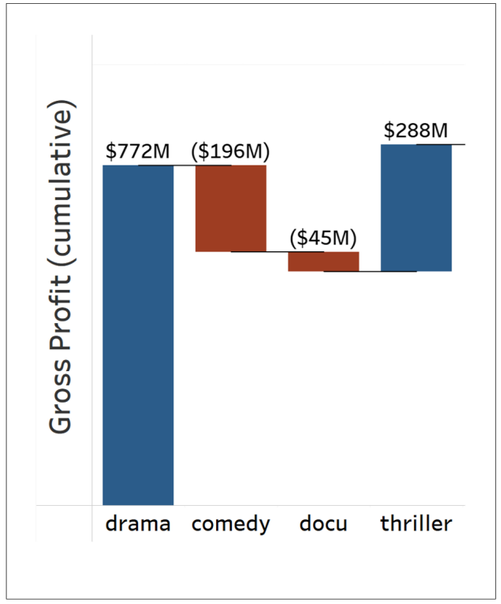
Visualizing for various use cases

Distribution



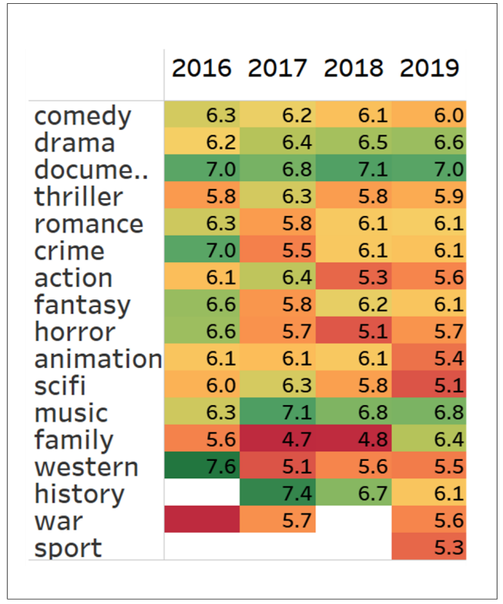
Box Plot

Contribution, change



Waterfall / Bridge

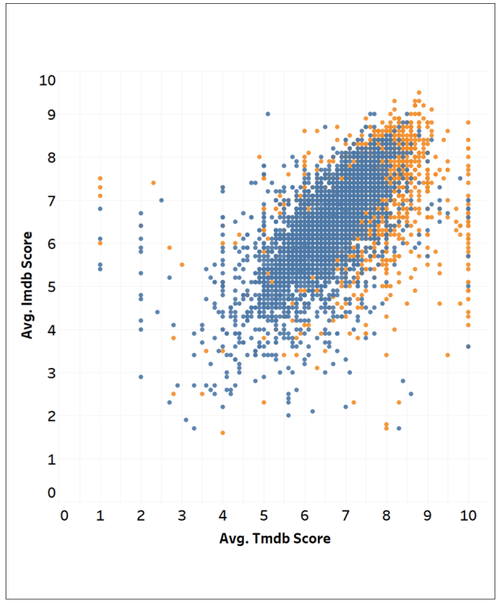
Density, matrix



	2016	2017	2018	2019
comedy	6.3	6.2	6.1	6.0
drama	6.2	6.4	6.5	6.6
docume..	7.0	6.8	7.1	7.0
thriller	5.8	6.3	5.8	5.9
romance	6.3	5.8	6.1	6.1
crime	7.0	5.5	6.1	6.1
action	6.1	6.4	5.3	5.6
fantasy	6.6	5.8	6.2	6.1
horror	6.6	5.7	5.1	5.7
animation	6.1	6.1	6.1	5.4
scifi	6.0	6.3	5.8	5.1
music	6.3	7.1	6.8	6.8
family	5.6	4.7	4.8	6.4
western	7.6	5.1	5.6	5.5
history		7.4	6.7	6.1
war		5.7		5.6
sport				5.3

Heat map

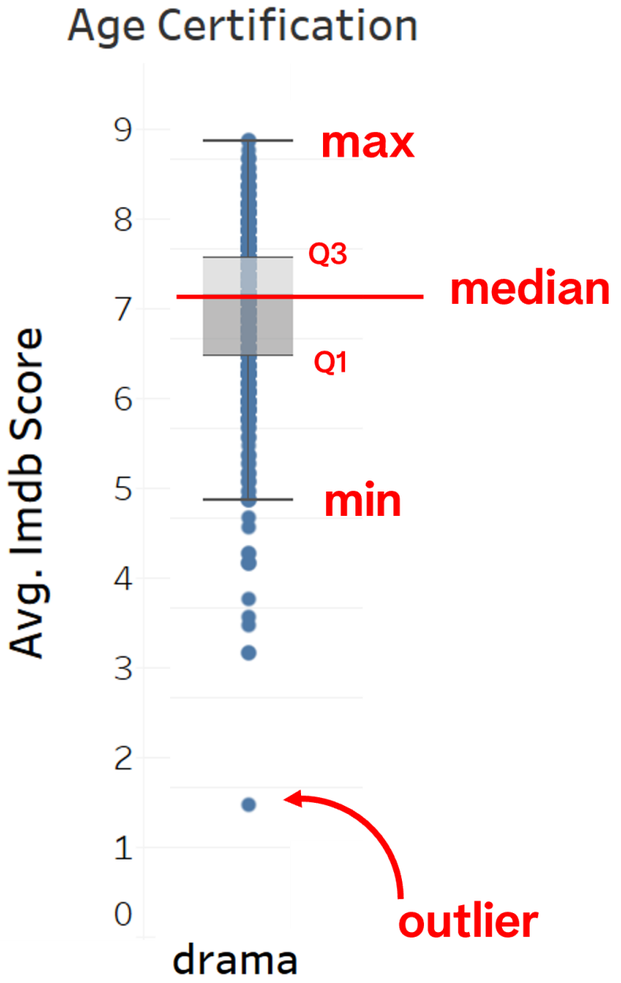
(Cor)relation



Scatter plot

Box plot - standardized chart for distribution of data

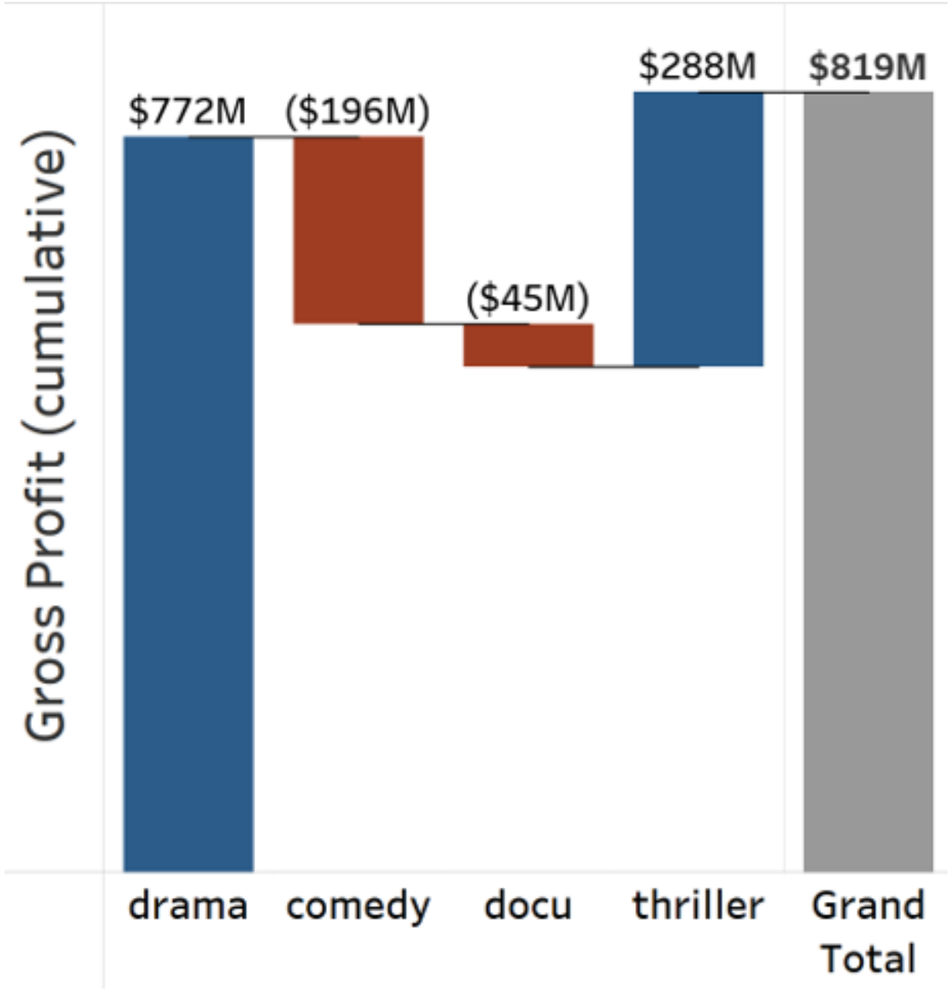
- A five-number summary:
  - Minimum
  - First quartile (Q1)
  - Median
  - Third quartile (Q3)
  - Maximum
- How data is distributed, grouped?
- Is it symmetrical/skewed?
- Are there any outliers?



Waterfall (bridge) chart - contribution and change

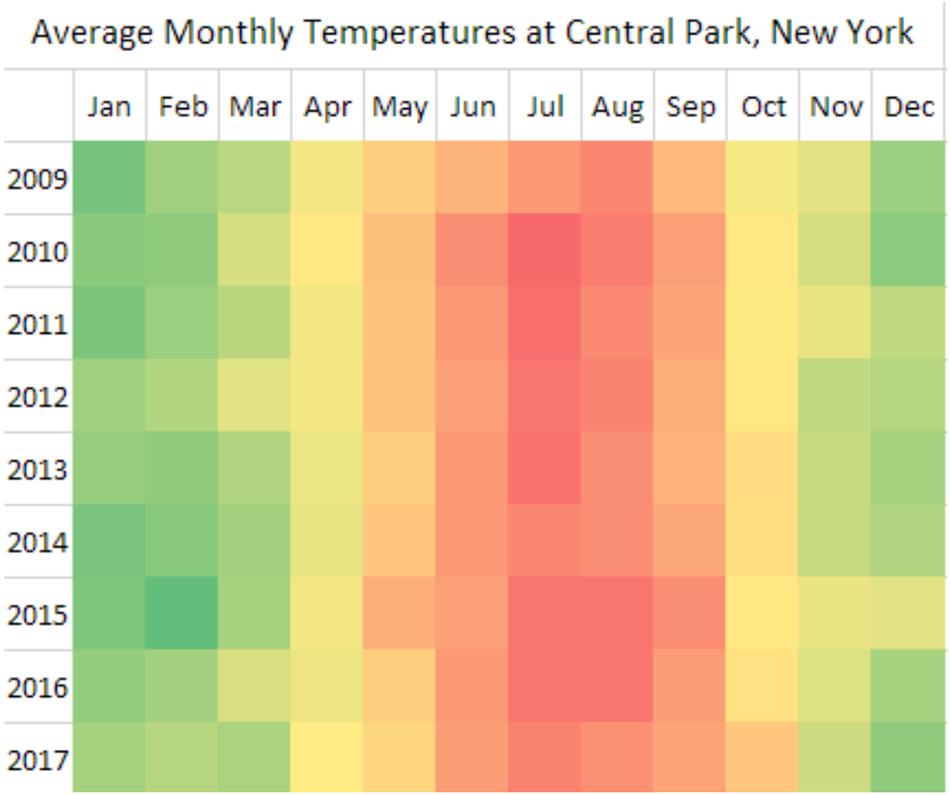
> A chart explaining the net change in value between two points, split over categories.

- Presents category contribution to the total
- Uses color to separate positive versus negative values
- Exposes the complexity hidden behind an aggregated number
- Often used with financial values
- Works well with a limited amount of categories



Heat map - density and matrix comparisons

> A colored chart for presenting density and comparisons, often in a matrix form.



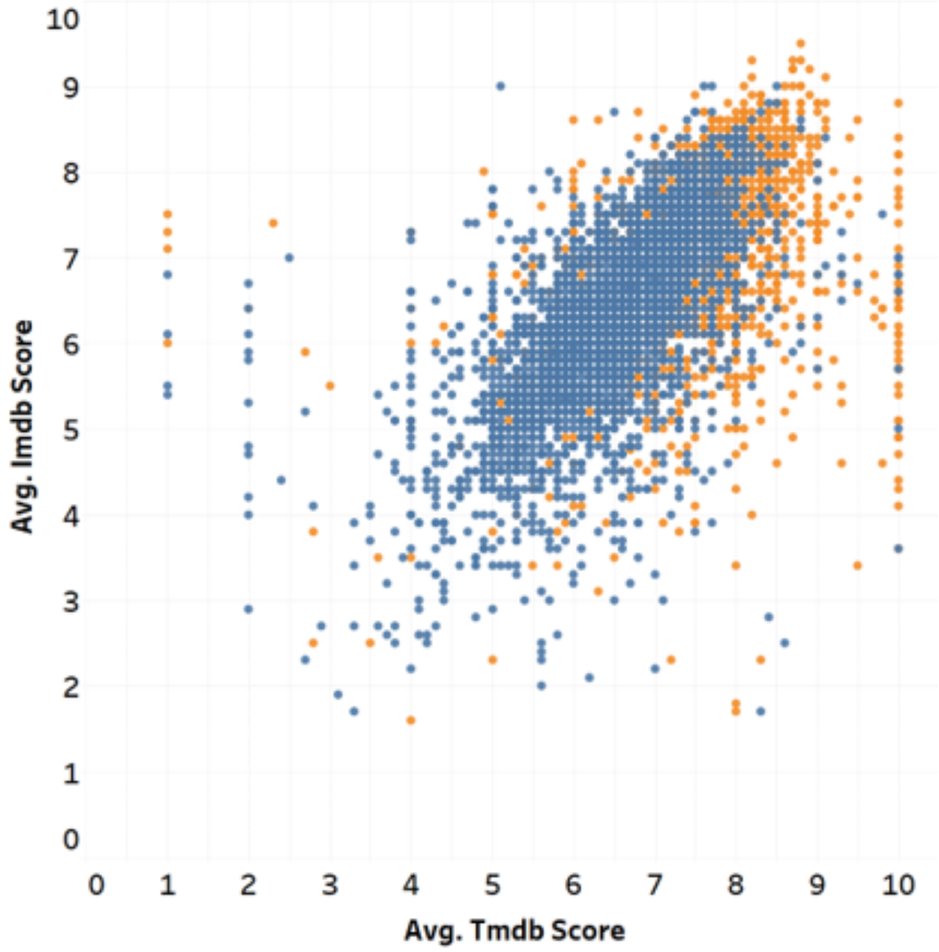
Main forms and types of heat maps:

- Colored (geo-)map
  - E.g., density of the population
- Matrix comparison (as on the picture)
  - E.g., comparing temperatures per day and month
- Uses easy-to-understand color gradations
- Hard to interpret without detailed labels

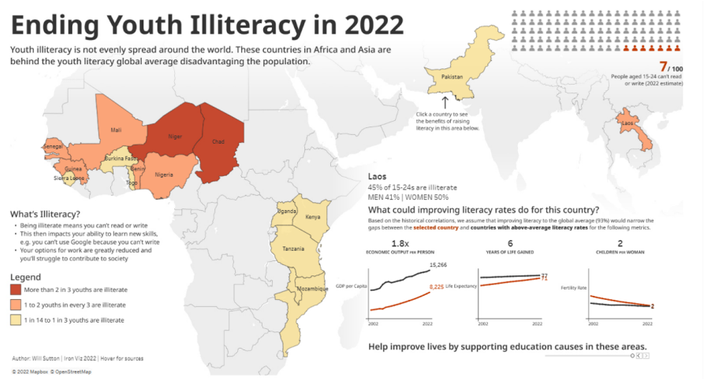
Scatter plot - relationship and correlation

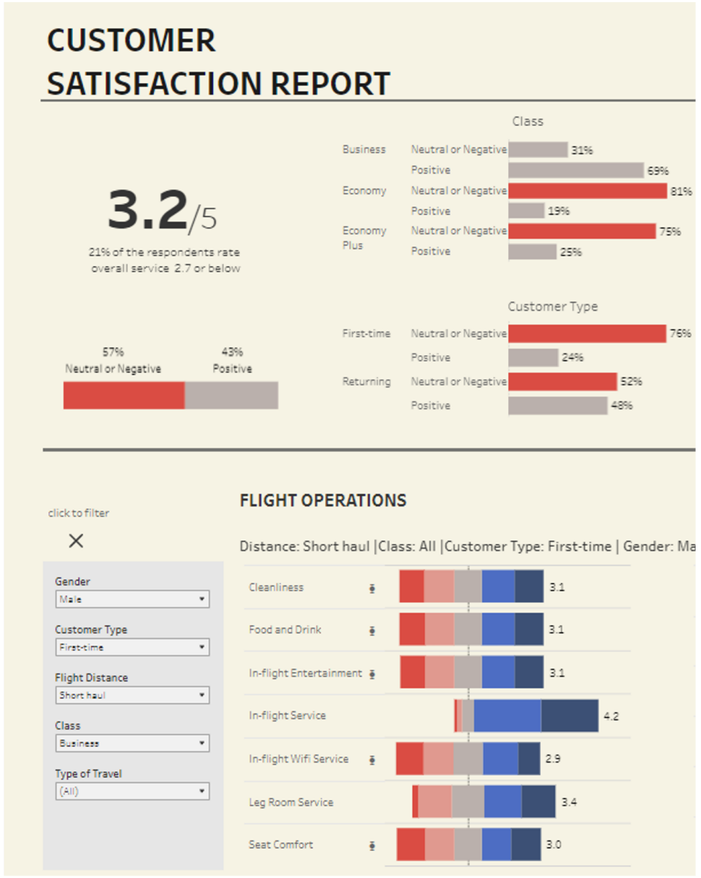
> A chart showing the relationship between two numerical variables plotted along both the X and Y axis.

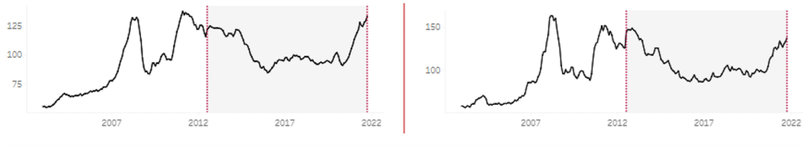
- Often used in exploratory data analysis
- Used to plot data on a quadrant
- Can be multidimensional, with the use of color, size, and shape
- Can store a relatively high number of data marks
- Difficult to communicate the data if they contain many data marks

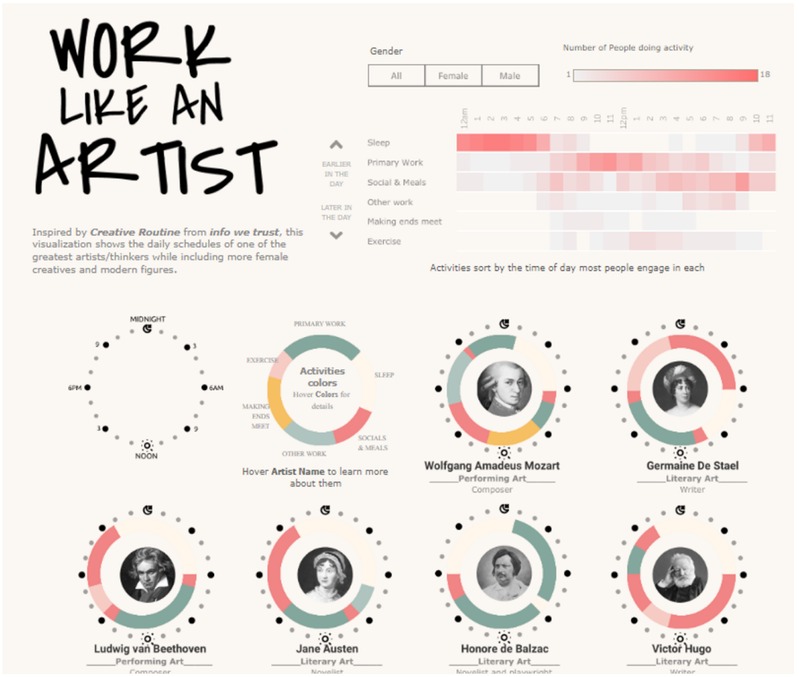


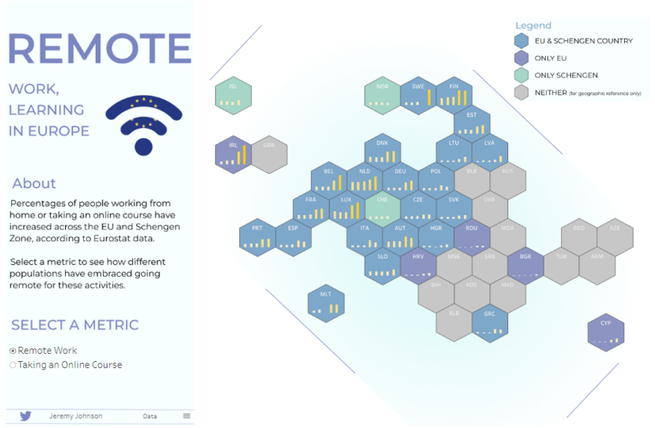
Get inspired - Tableau's Viz of The Day











[1] https://public.tableau.com/app/discover/viz-of-the-day

Let's practice!

DATA VISUALIZATION IN TABLEAU

Building more visualizations in Tableau

DATA VISUALIZATION IN TABLEAU



Agata Bak-Geerinck
Senior Data Analyst, Telenet

Let's practice!

DATA VISUALIZATION IN TABLEAU