Lecture 11
Infographics, Other Systems

Data Visualization · 1-DAV-105

Lecture by Broňa Brejová

Infographics

Some examples of infographics

Several examples that are close to data visualization:

- Income by religious group in US (image, website)
- Deadliest pandemics (image, website)
- War casualties (website)
- Game of Thrones relationships (image, website)

Some explain other types of information:

- Famous people with learning disabilities (website)
- Preference polling in Slovakia (website)

Data visualization (DV) vs infographics (IG)

- **Target audience:** IG general public, DV often experts
- **Storytelling:** often in IG, can be created from multiple DV
- **Design and aesthetics:** more elaborate in IG, includes graphics elements and clipart (considered chart junk in DV)
- **Process of creation:** many simple tools for DV, IG time consuming, often created by collaboration of data analysis, domain experts and graphic designers

https://www.statsilk.com/blog/real-difference-between-infographics-and-data-visualizations

Other visualization tools

Other visualization tools

Spreadsheets

- Excel
- Google sheets

Well-know commercial tools

- Tableau: Advanced visualization tools
- Flourish
- Microsoft Power BI: Interactive data visualization with a focus on business intelligence (our university); see elective course Data Analysis in Power BI

Other visualization tools

System R

- Programming language for statistical computing
- Together with Python, very popular in data science, built-in plots
- Colab has R runtype
- Library ggplot2 based on Grammar of Graphics (cheatsheet)

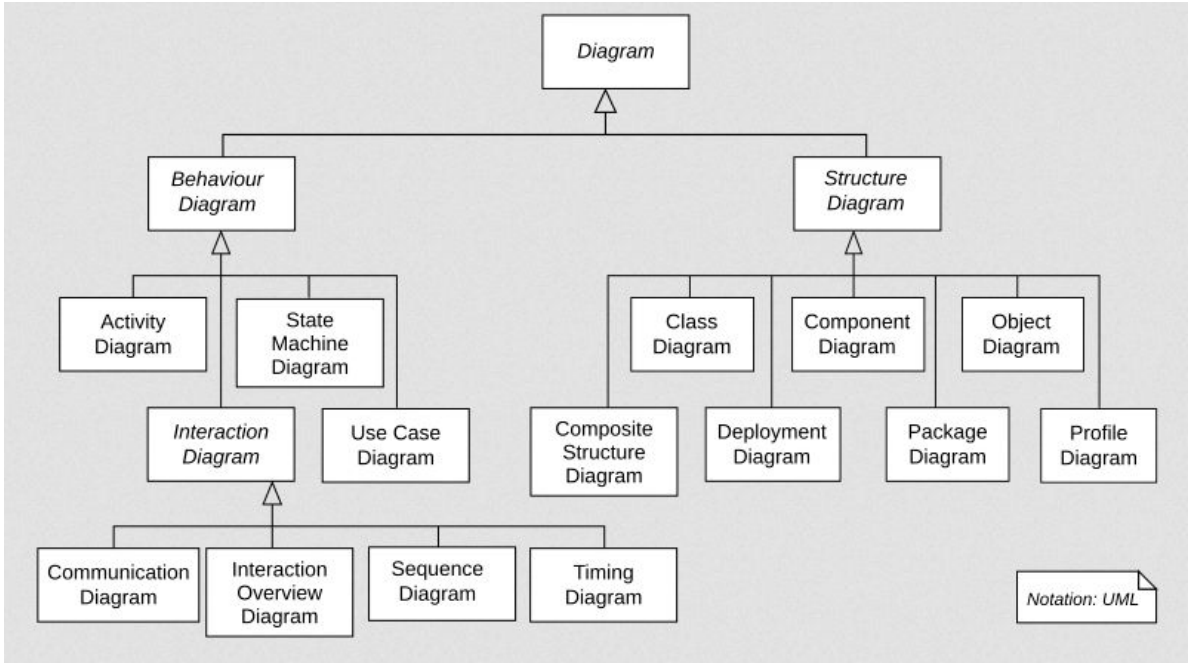
Javascript

- Programming language popular in web programming
- Google charts for Javascript (see Data management course)
- D3.js library (Data-Driven Documents)
- Vega-Lite uses javascript to embed plots specified as json

Specialized visualizations

UML diagrams in computer science

Display relationships between components and aspects of software



https://commons.wikimedia.org/wiki/File:UML_diagrams_overview.svg

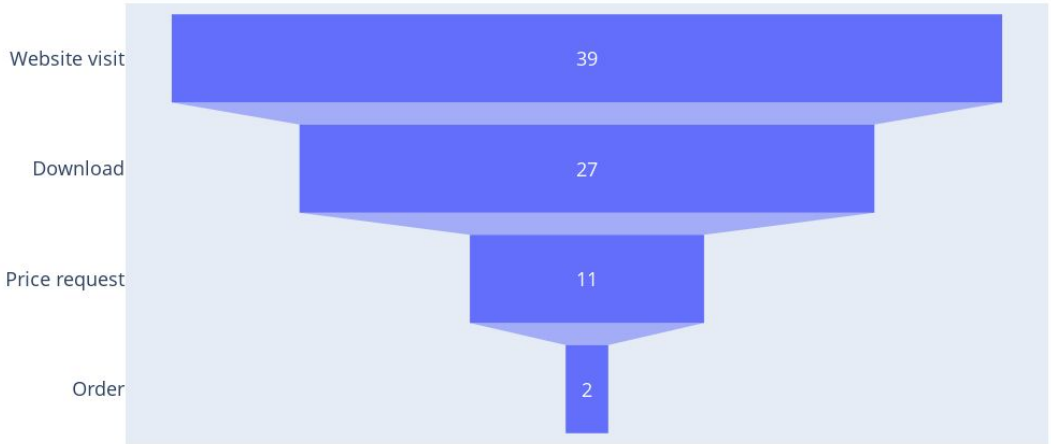
Waterfall chart

- Used in business analysis: financial, inventory, human resources etc.
- Displays effects decreasing or increasing a given value
- The first and last bars display starting and final value
- Intermediate columns float, displaying changes from previous total (more info)

https://commons.wikimedia.org/wiki/File:Waterfallchart_ex2.jpg

Funnel charts

- Display losses within a business process, e.g. from website visit to actual purchase
- Horizontal bar chart with centered bars
- Different funnel plot in medical meta-analyses of multiple publications



Made in plotly

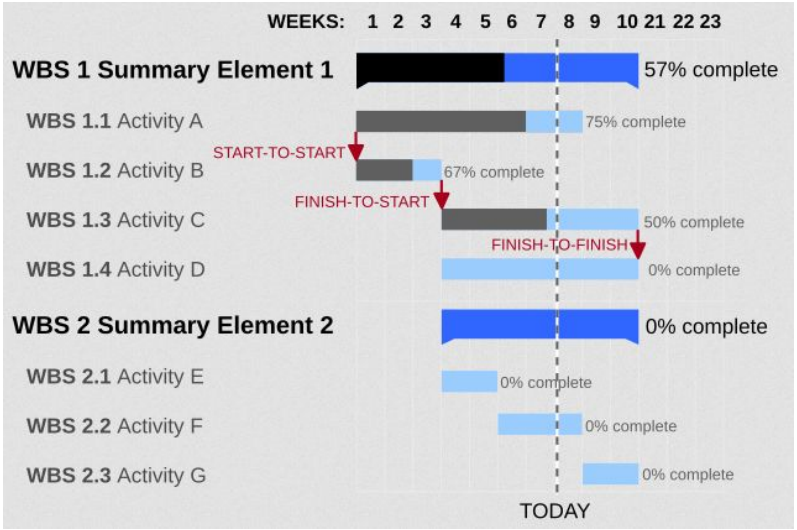
Candlestick chart

- Similar to boxplot, used in financial data, e.g. stocks, currency exchange rates
- Line: minimum and maximum, box: opening and close, color: increase or decrease

https://commons.wikimedia.org/wiki/File:Candlestick_Chart_in_MetaTrader_5.png

Gantt chart

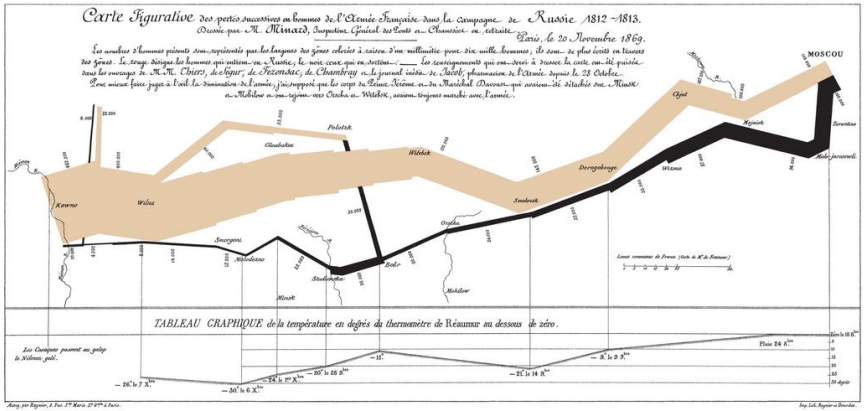
- Used in management to display project schedule with different tasks and their planned duration
- Can also display current status of tasks and their dependencies



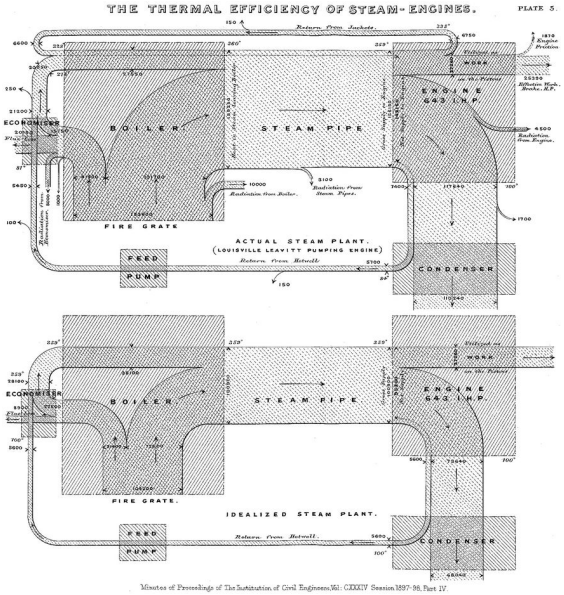
https://commons.wikimedia.org/wiki/File:GanttChartAnatomy.svg

Sankey diagram

- flows in a system, line width proportional to amount
- Minard's invasion of Russia 1869, Sankey's energy in steam engine 1898
- [Energy in EU]



https://commons.wikimedia.org/wiki/File:Minard.png



https://commons.wikimedia.org/wiki/File:JIE_Sankey_V5_Fig1.png

Another interesting book

Claus O. Wilke 2019 Fundamentals of Data Visualization

The author is a scientist (actually working bioinformatics)
The book is written in R, with source code available

We will look at several examples from this book

Forensic glass composition

- Dataset
- 214 glass samples
- Content of 7 elements (Na, Mg, Al, Si, K, Ca, Ba, Fe)
- Classification into 4 classes: window, headlamp, tableware, container
- Any ideas for visualization?
- What if we are interested in associations among variables?

Forensic glass composition

- Dataset
- 214 glass samples
- Content of 7 elements (% of mass)
- Classification into 4 classes: window, headlamp, tableware, container
- From the book:
  - correlogram 1 (image)
  - correlogram 2 (image)
  - PCA (image); meaning of axes (image)
- How can we use PCA to learn something about a new sample?
- Notice that PCA plot uses both shape and color - what is the advantage?

CO_2 emissions per person

- 166 countries
- Years 1970 and 2010 for each country (called paired data)
- How would you visualize?
- What questions can you ask?
- What if we are interested in only a small subset of countries?

CO2 emissions per person

- 166 countries
- Years 1970 and 2010 for each country (called paired data)
- From the book:
  - scatterplot (image)
  - slopegraph 1 (image)
  - slopegraph 2 (image) for years 2000, 2005, 2010
- Scatterplot of paired data has x-y diagonal
- Hard to judge distance from diagonal
- Mean-difference or sum-difference plot transforms coordinates (example)

Comparing two time series

- Average temperature in different days of a year in two cities
- How would you plot this?

Comparing two time series

- Plot each against time in one linegraph
- Plot each against time in two graphs next to each other
- Scatterplot of one temperature against other
  - loses time information
  - can be added back as color
- Book uses connected scatterplot (image)
  - another example (image); compared to two line graphs (image)

Follow-up courses

- Data Management various techniques for processing data BIN, DAV 2S
- Database Systems BIN, DAV 3W

- Fundamentals of Probability and Statistics BIN 3W, DAV 2W
- Statistical Methods DAV 3W

- Principles of Data Science DAV 3W
- Network Science DAV 3S