Lecture 11
Interactivity, Other Systems

Data Visualization · 1-DAV-105

Lecture by Broňa Brejová

Interactivity

Interactivity

Interactive visualization engages audience, allows them to explore data in depth and according to their interest.

Examples

- [US cities with the same name](#) (also the animated explainer)
- [PhD gender gap](#)
- [Making it big](#) (more animated than interactive)

Techniques in interactivity visualization

Similar to decisions made in designing a static plot (Stephen Few)

- Selecting variables (x, y, color, …)
- Filtering data (selecting table rows)
- Aggregating (display countries or region summaries)
- Zooming / panning
- Rescaling (log-scale) / reexpressing (e.g. % instead of counts)
- Sorting (e.g. bars in bargraphs)
- Highlighting points or groups
- Displaying details (tooltips)
- Annotating
- Bookmarking

Dashboard

A display consisting of multiple plots, summarizing the current state of important indicators (e.g. of a business, pandemics, …)

- Inspired by dashboards in cars and planes
- Often interactive, but main features in default view

Two SARS-CoV-2 examples:

- WHO
- Nextstrain (selecting color, filtering, highlighting, aggregating, zooming and panning maps and tree, rescaling time vs divergence, tooltips, bookmarking)

Interactivity in Plotly Express

All Plotly plots by default have some interactivity:

- Filtering groups
- Zooming / panning
- Details (tooltips / hover label)
- Spike lines

See two examples in the notebook

We have also seen how to build more advanced interactive plots with Plotly Dash (adding selectors, sliders, buttons, …)

Other visualization tools

Other visualization tools

Spreadsheets
- Excel
- Google sheets

Well-know commercial tools
- Tableau: Advanced visualization tools
- Microsoft Power BI: Interactive data visualization with a focus on business intelligence (our university)

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)

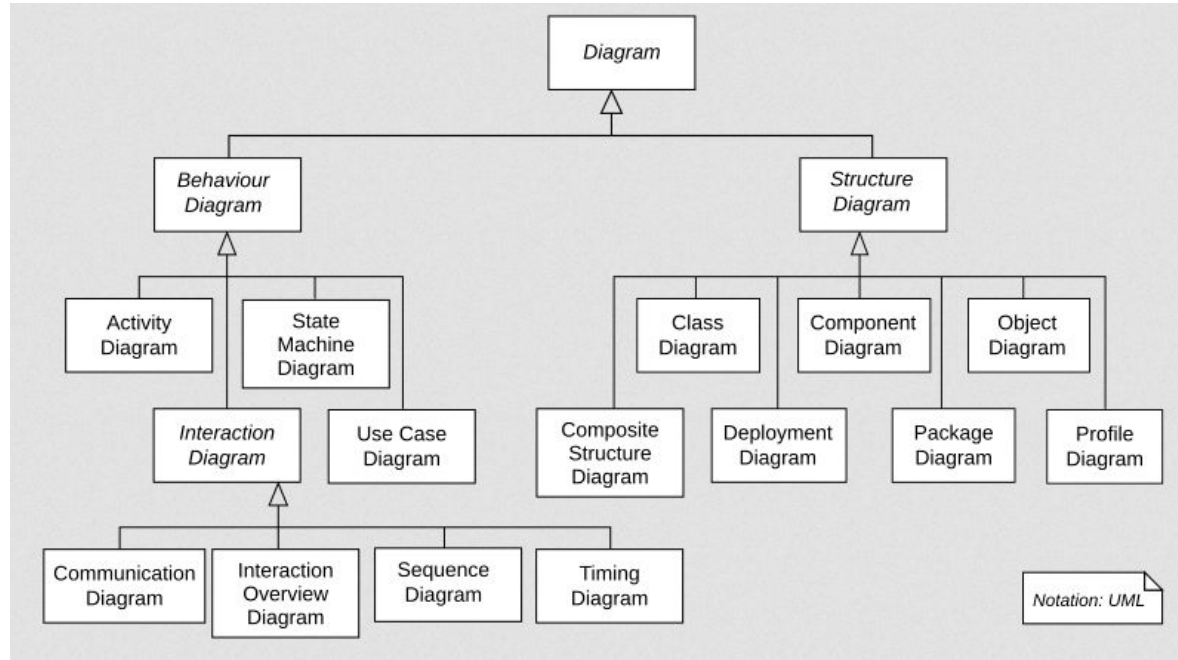**Javascript**

• Programming language popular in web programming

• Google charts for Javascript

• 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

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
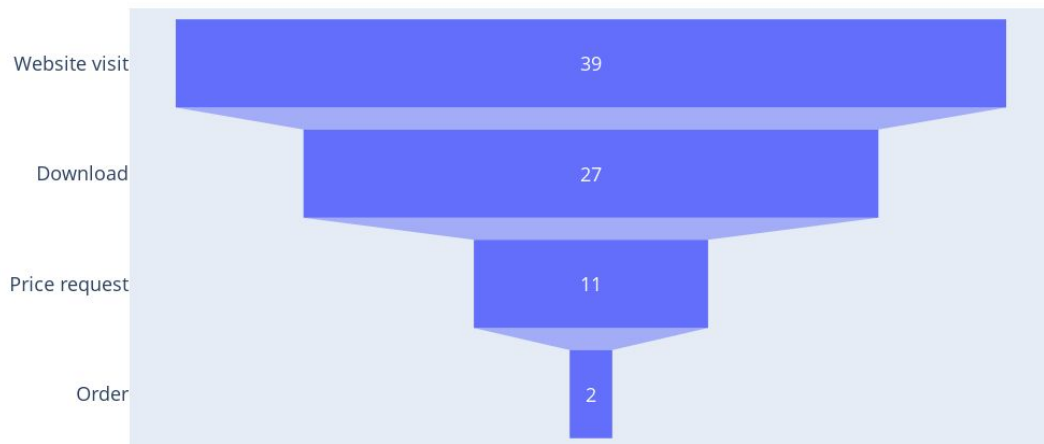
Made in plotly,
see the notebook

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
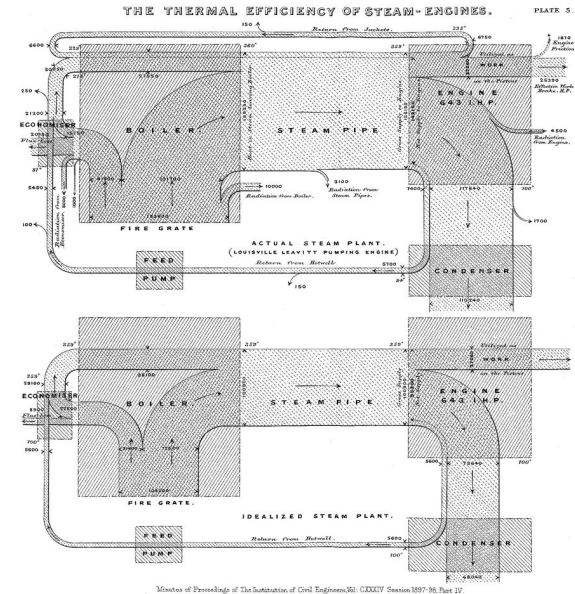

https://commons.wikimedia.org/wiki/File:GanttChartAnatomy.svg
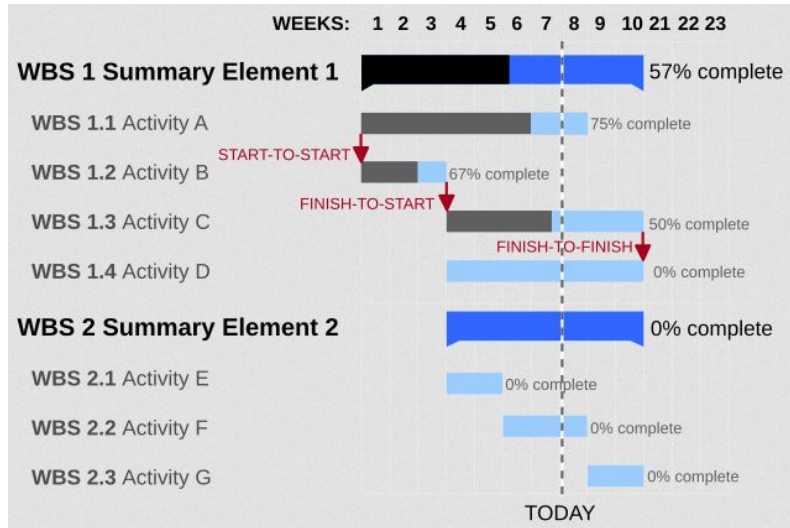
Sankey diagram

- 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](#)
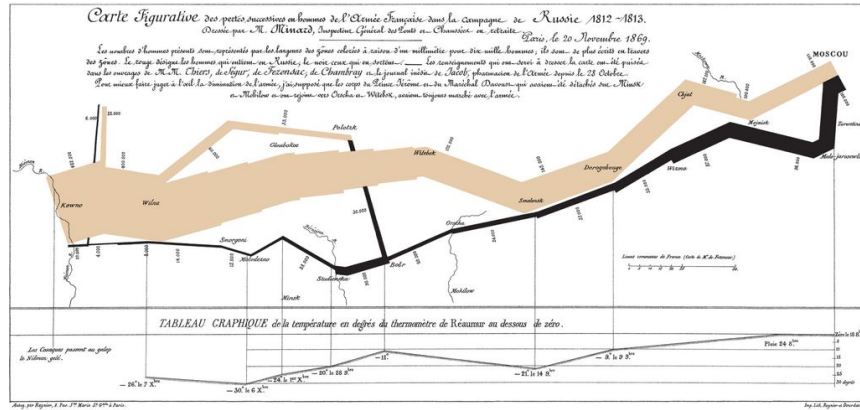
https://commons.wikimedia.org/wiki/File:Minard.png

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 his book

Forensic glass composition

- [Dataset](#)
- 214 glass samples
- Content of 7 elements
- 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](#)
  - [correlogram 2](#)
  - [PCA](#) ([meaning of axes](#))
- 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?

CO2 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
  - slopegraph 1
  - slopegraph 2 (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
- [Connected scatterplot](#) from the book (and [another](#))