

# 1 Lecture 11: Infographics, interactivity, other tools, specialized plots

Data Visualization · 1-DAV-105

Lecture by Broňa Brejová

Acknowledgement: some materials inspired by lectures from Martina Bátorová in 2021

[ ]:

## 1.0.1 Several examples of infographics

Several examples that are close to data visualization:

- Income by religious group in US ([image](#), [website](#))
- Deadliest pandemics ([website](#))
- War casualties ([website](#))
- Game of Thrones relationships ([website](#))
- Emergency medical services in Slovakia 2019 ([website](#))

Some explain other types of information:

- Sitting and standing is bad ([website](#))

## 1.1 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

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

## 1.2 Interactivity

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

### 1.2.1 Examples

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

### 1.2.2 Techniques in interactivity visualization

Similar to decisions made in designing a static plot:

- Selecting variables (x, y, color, ...)

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

(Stephen Few)

### 1.2.3 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](#)
  - many options: selecting color, filtering, highlighting, aggregating, zooming and panning (maps and tree), rescaling (time vs divergence), tooltips, bookmarking

### 1.2.4 Interactivity in Plotly Express

All Plotly plots by default have some interactivity:

- Filtering groups
- Zooming / panning
- [Details](#)
- [Spike lines](#)

Example 1: Country indicators from World Bank, <https://databank.worldbank.org/home> under CC BY 4.0 license.

Regions can be switched on and off.

```
[1]: import plotly.express as px
import pandas as pd
url = 'https://bbrejova.github.io/viz/data/World_bank.csv'
countries = pd.read_csv(url)

px.scatter(
    countries, x="GDP2020", y="Expectancy2020", color="Region",
    hover_data=['Country'],
    title="Country indicators 2020", log_x=True,
    width=800, height=500
)
```

Example 2: Life expectancy data, based on free data from World bank via [gapminder.org](https://gapminder.org), CC-BY license (years 1900-2017) and [World bank](https://data.worldbank.org) directly (years 2018-2021).

Compare data along the x coordinate.

```
[2]: url = "https://bbrejova.github.io/viz/data/life_expectancy_years.csv"
orig_expectancy = pd.read_csv(url, index_col=0)
orig_expectancy = orig_expectancy.iloc[:, 1:].reset_index()
expectancy = pd.melt(orig_expectancy, id_vars=["Country"], var_name="Year",
    value_name="Expectancy")
expectancy['Year'] = expectancy['Year'].astype(int)
display(expectancy.head())
```

	Country	Year	Expectancy
0	Afghanistan	1900	29.4
1	Albania	1900	35.4
2	Algeria	1900	30.2
3	Angola	1900	29.0
4	Antigua and Barbuda	1900	33.8

```
[5]: selected = expectancy.query("Country=='Slovak Republic' or Country=='Portugal'")
fig=px.line(
    selected, x="Year", y="Expectancy", color="Country",
    width=800, height=500
)
fig.update_layout(hovermode="x unified")
```

### 1.2.5 More interaction with Dash by Plotly

- Dash library by Plotly allows adding control elements (selectors, sliders, buttons, ...)
- We have seen an example in L01

## 1.3 Other visualization tools

Non-programmers typically create plots in spreadsheets:

- Excel ([examples](#))
- Google sheets ([examples](#))

Tableau

- Advanced visualization tools, commercial
- [Gallery](#)

Microsoft Power BI

- Interactive data visualization software with a focus on business intelligence
- An [example](#)

System R: programming language for statistical computing

- Together with Python, very popular in data science

- Built-in [plots](#)
- Colab has R runtime
- Also other libraries, notably [ggplot2](#) based on system called Grammar of Graphics ([cheat-sheet](#))

Javascript

- Programming language popular in web programming
- Google charts for Javascript ([examples](#))
- [D3.js](#) library (Data-Driven Documents)
- Vega-Lite uses javascript to embed plots specified as json ([examples](#), [embedding in html](#))

## 1.4 Several specialized visualization types

### 1.4.1 UML diagrams in computer science

- Display relationships between different classes or other components and aspects of software

[https://commons.wikimedia.org/wiki/File:UML\\_diagrams\\_overview.svg](https://commons.wikimedia.org/wiki/File:UML_diagrams_overview.svg) Derfel73; Pmerson

### 1.4.2 Waterfall chart

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

[https://commons.wikimedia.org/wiki/File:Waterfallchart\\_ex2.jpg](https://commons.wikimedia.org/wiki/File:Waterfallchart_ex2.jpg) FusionCharts Blog, CC BY-SA 4.0

### 1.4.3 Funnel charts

- Display losses within a business process, e.g. from website visit to actual purchase
- Horizontal bar chart with centered bars
- Beware: different from [funnel plot](#) in medical meta-analyses of multiple publications

```
[22]: # inspired by https://plotly.com/python/funnel-charts/
data = dict(
    number=[39, 27, 11, 2],
    stage=["Website visit", "Download", "Price request", "Order"])
fig = px.funnel(data, x='number', y='stage')
fig.update_layout(font=dict(size=18), yaxis_title="", margin=dict(l=10, r=10, u
    ↪t=30, b=10))
fig.show()
```

### 1.4.4 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](https://commons.wikimedia.org/wiki/File:Candlestick_Chart_in_MetaTrader_5.png)

#### 1.4.5 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>