PLIN064 Úvod do *digital humanities*

Zuzana Nevěřilová
xpopelk@fi.muni.cz

Centrum zpracování přirozeného jazyka, B203
Fakulta informatiky, Masarykova univerzita

6. listopadu 2019

Why visualization

Data visualization using graphs, charts, diagrams etc. is the expression of quantitative information in graphic form.

Good visualization = persuasive visual arguments

Mapping visualization (inspired by carthography) is a symbolic representation of the reality (Digital Cultural Mapping, Spatial Humanities)

- reveal patterns, anomalies, concurrences
- illustrate findings or distillate argument

Visualization Tools for Python

- `matplotlib`
- `pandas`
- `seaborn`
- `yellowbrick`
- `bokeh`

Visualization Methods have to Fit the Data

(almost) continuous data: temperature, age, time

- line charts

discrete data: categories, places

- bar charts

continuous data made discrete: weekdays, age groups

- bar charts – histograms (probability distribution for given intervals)
- box plots – five number summaries (0, 25, 50, 75, 100)

Co-occurrence visualization

- continuous 2D: correlation
- continuous 3D: depth, density plots
- discrete 2D: scatter plot
- discrete multi-D: heat map, confusion matrix

to visualize multi-D, we can

- be more creative: up to 6D, we can use colors, sizes, shapes in a 3D visualization space
- draw several charts for one dataset: faceting
- reduce the dimensionality
  e.g. using t-SNE (t-Distributed Stochastic Neighbor Embedding = probabilistic technique)
  or PCA (Principal Component Analysis = matrix-based technique)

More Information & Try it Yourself

<https://towardsdatascience.com/introduction-to-data-visualization-in-python-89a54c97fbed>
<https://towardsdatascience.com/the-art-of-effective-visualization-of-multi-dimensional-da>

Free datasets, e.g. Iris + any python module to try it yourself.