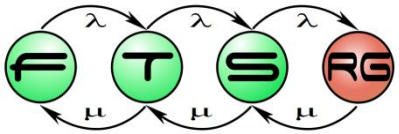
Visual Analysis Basics

Visual Analysis of Measurement Data

2019.09.26.

Budapest University of Technology and Economics
Fault Tolerant Systems Research Group

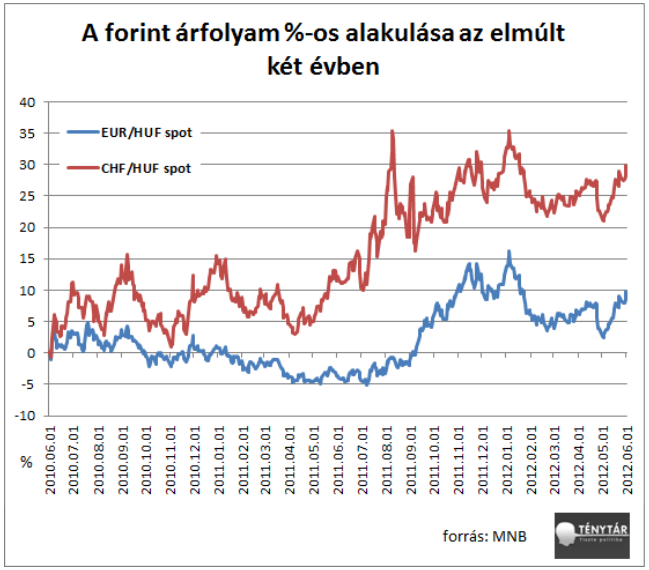


Aims of visualization

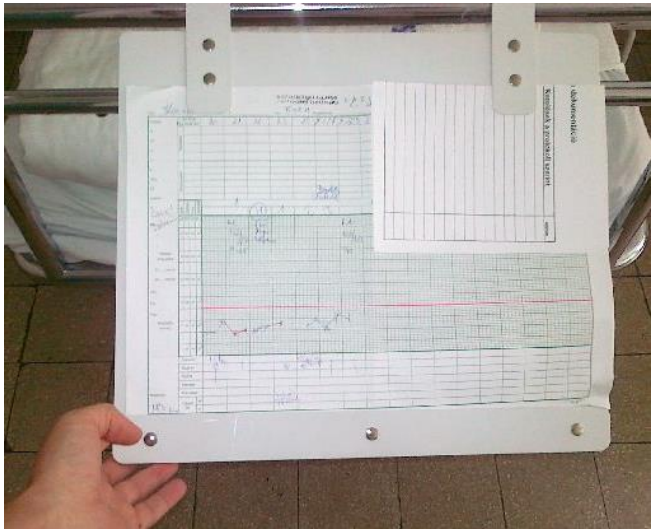
- Assumption formulation
- Communication of results
- Share research in a reproducible manner
- Give a tool to browse the data
- Create a high level report
- Evaluate a system/solution
- Tell a story
- …

Applications

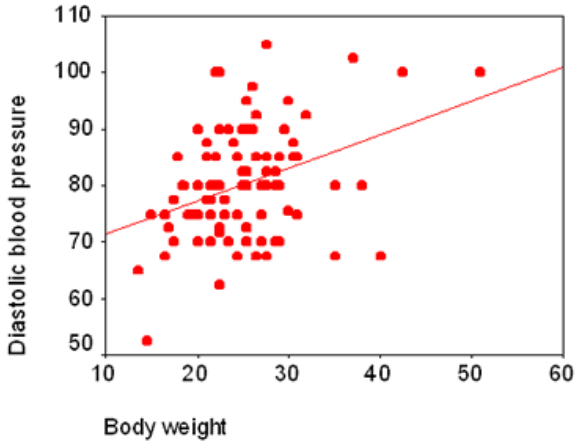
Trend analysis



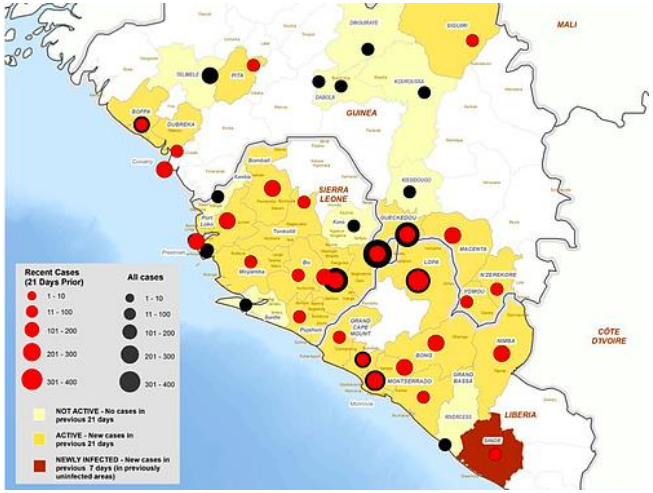
Time series analysis



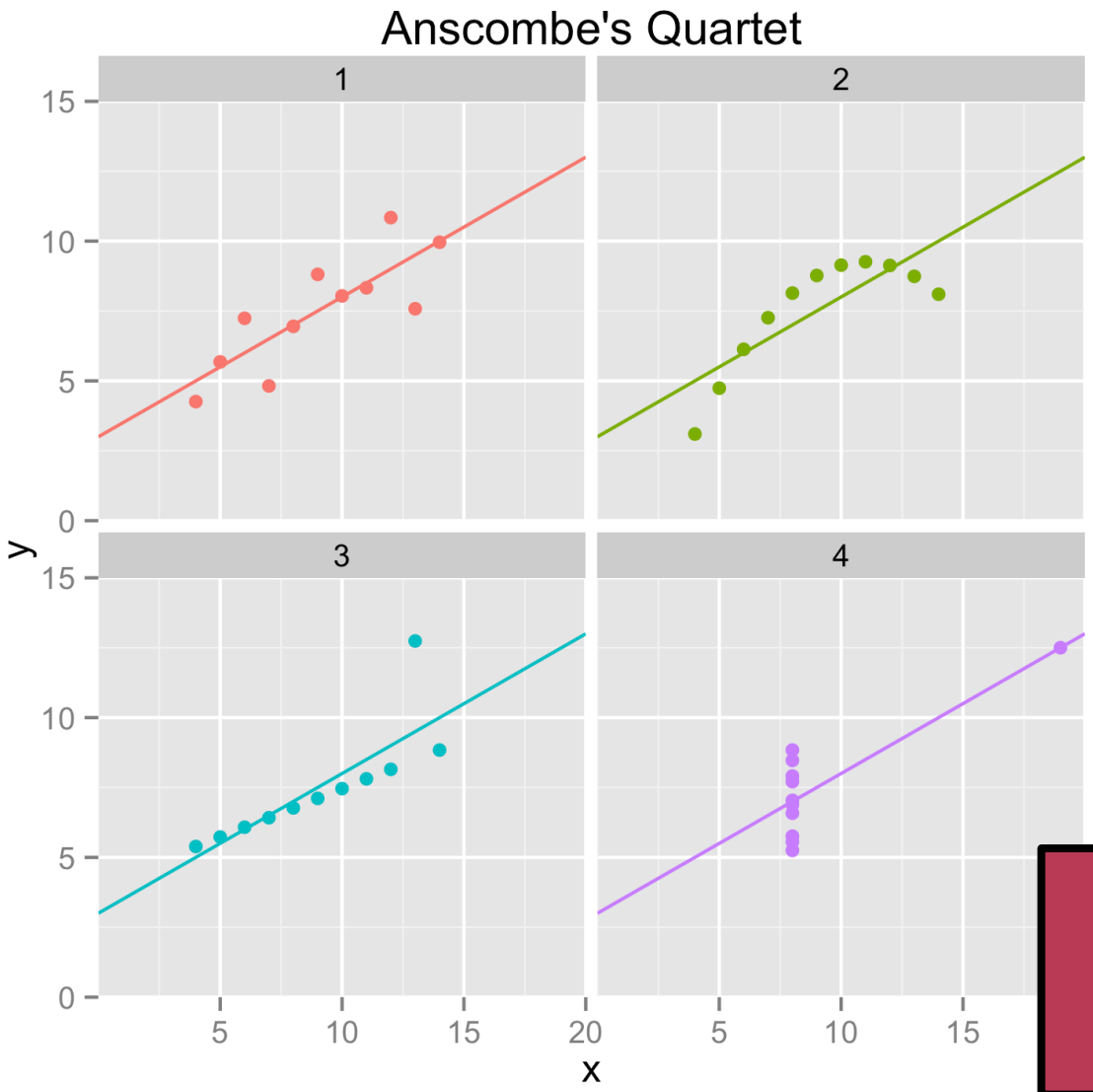
Correlation analysis



Spatial analysis

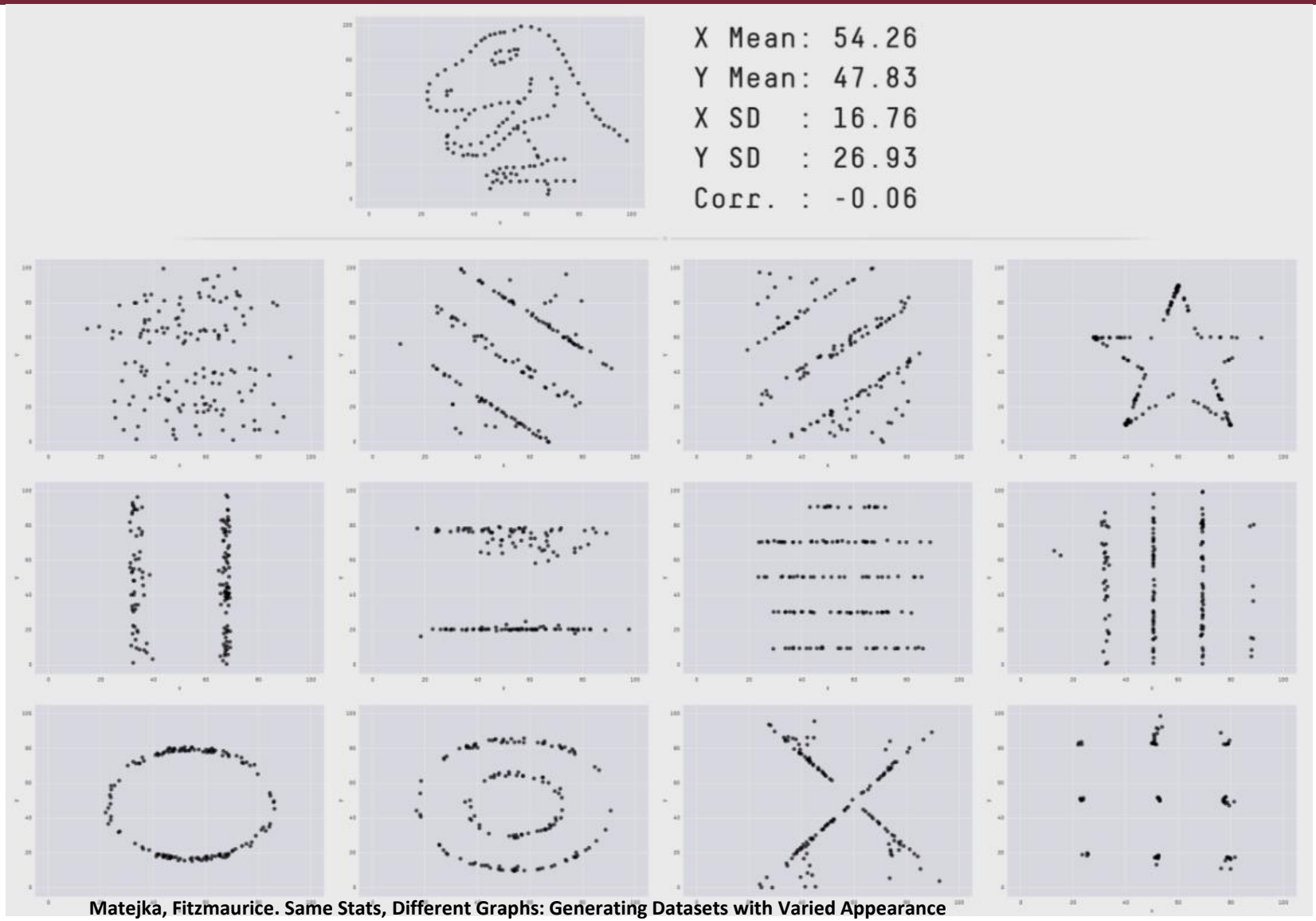


Validation for automated methods



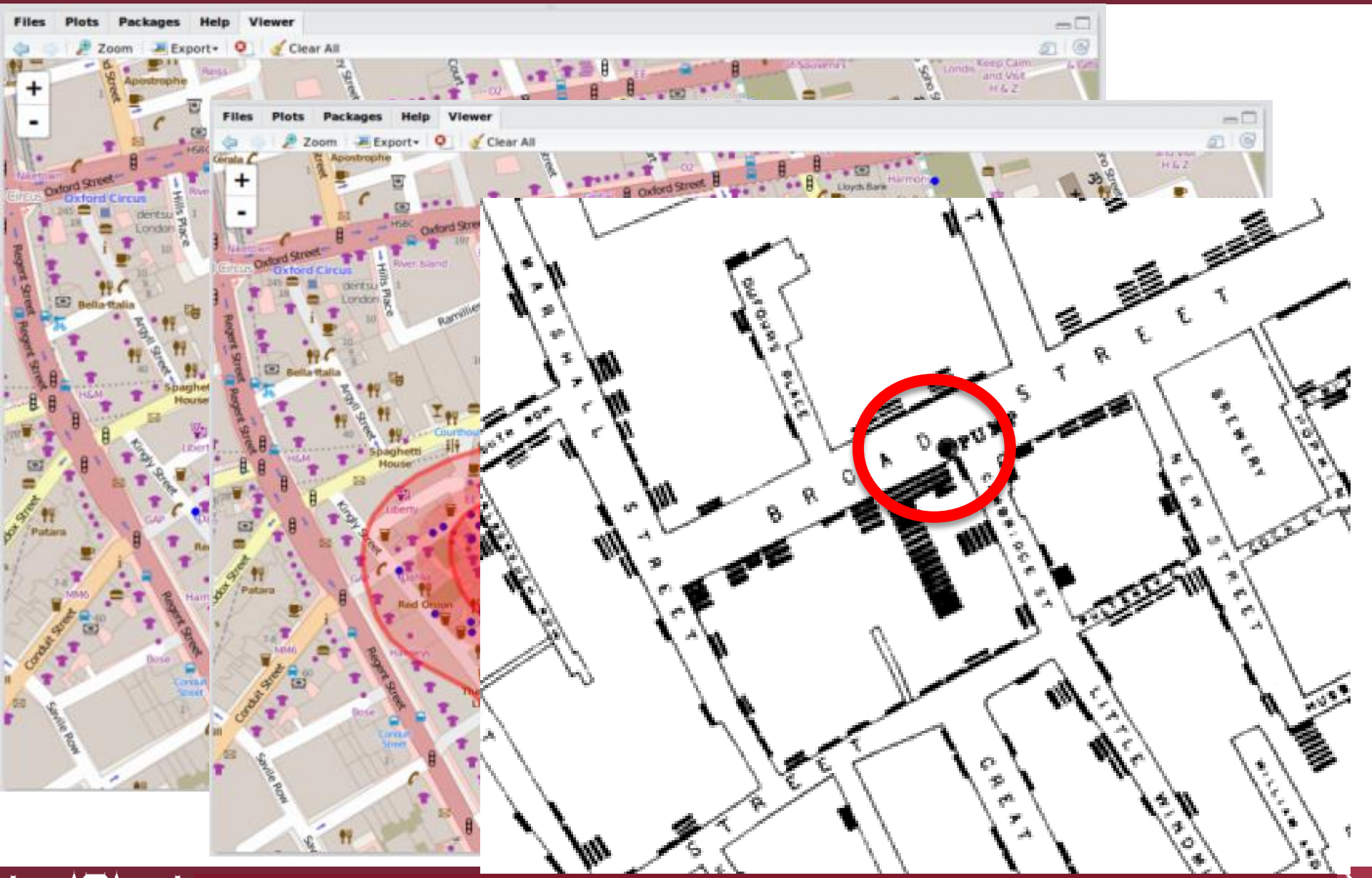
Avoid false assumptions
Use human intuition
(1973)

… and some more



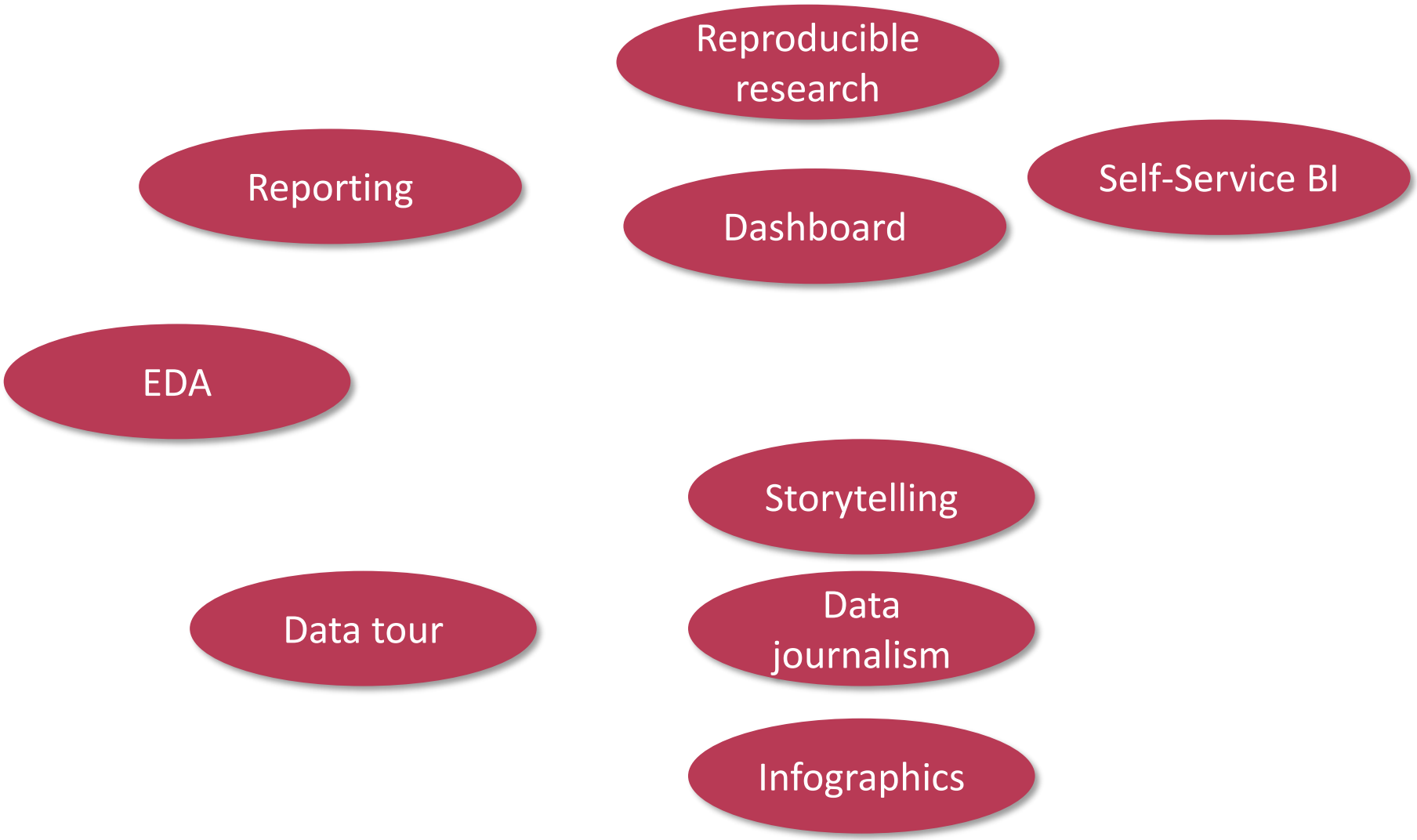
Matejka, Fitzmaurice. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. ACM SIGCHI Conference on Human Factors in Computing Systemshttps://www.autodeskresearch.com/publications/samestats

Find patterns/correspondence

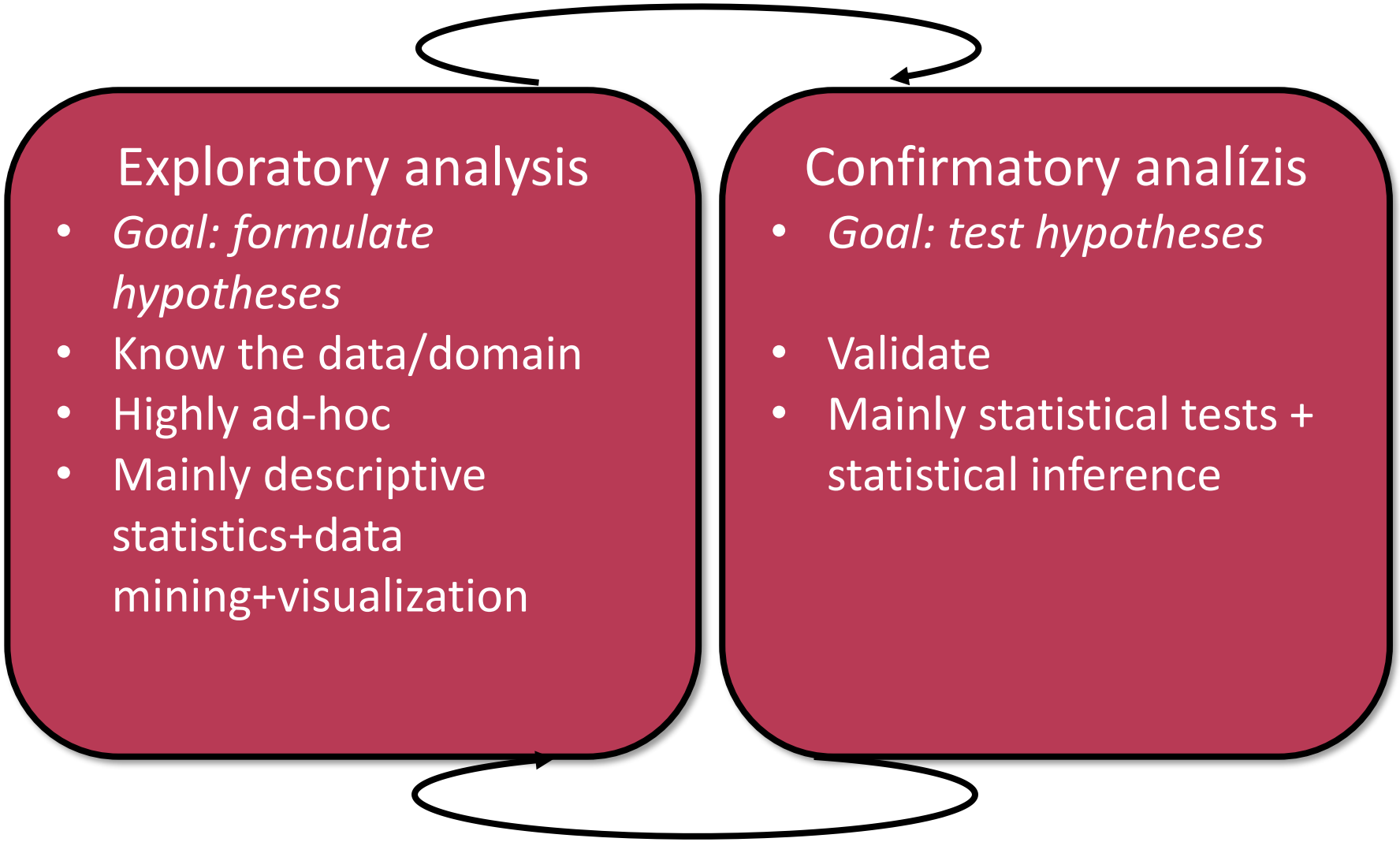


https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

Visualization types



Data analysis

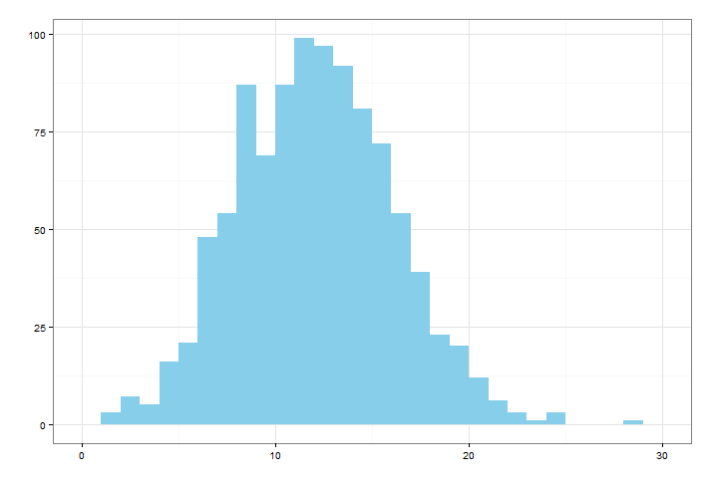


Data analysis

- E.g. distribution analysis

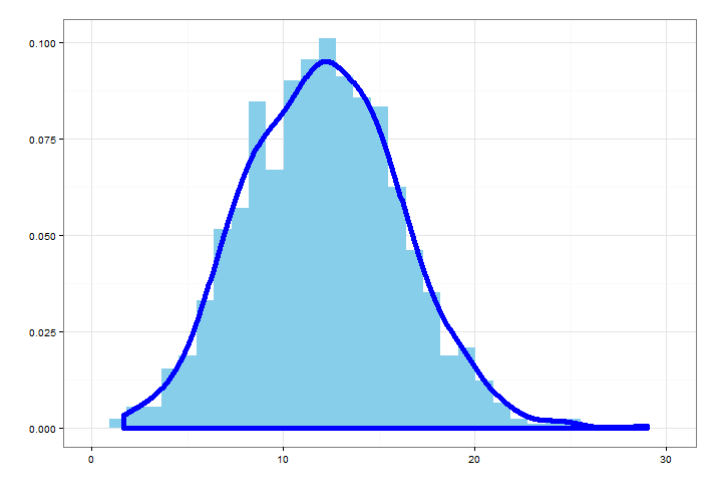
Exploratory

Hypothesis: variable x follows normal distribution



Confirmatory

Variable x follows $N(12, 4)$ distribution

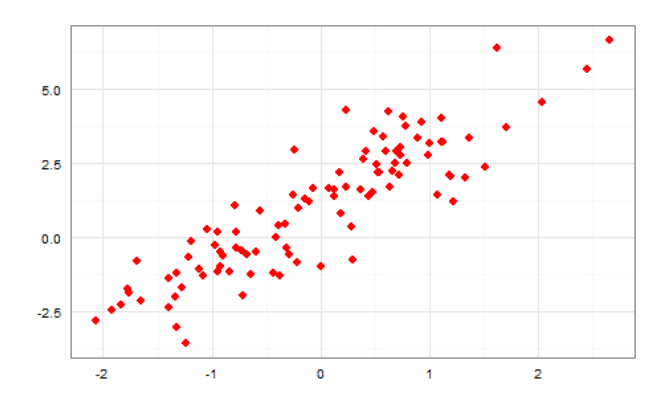


Data analysis

- E.g., linear regression

Exploratory

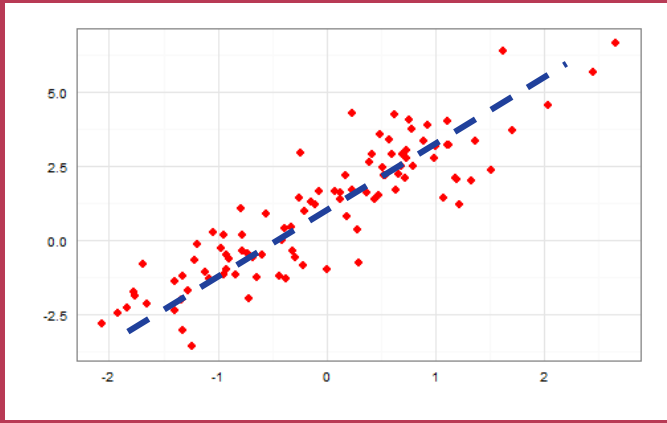
Hypothesis: variables x and y are linearly correlated



Confirmatory

$$Y = 3.2 \times x - 1.2$$

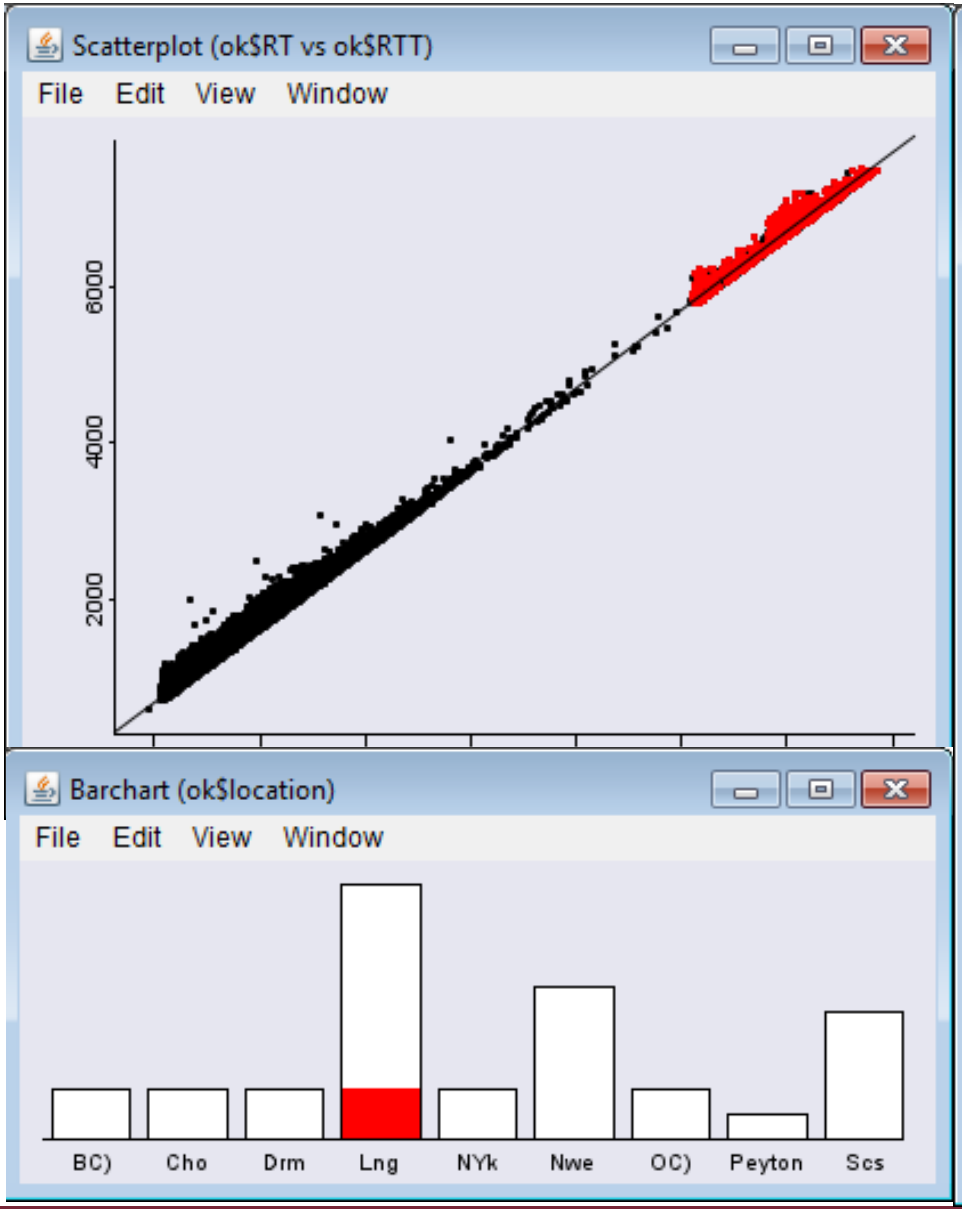
$$R^2 = 0.1234$$

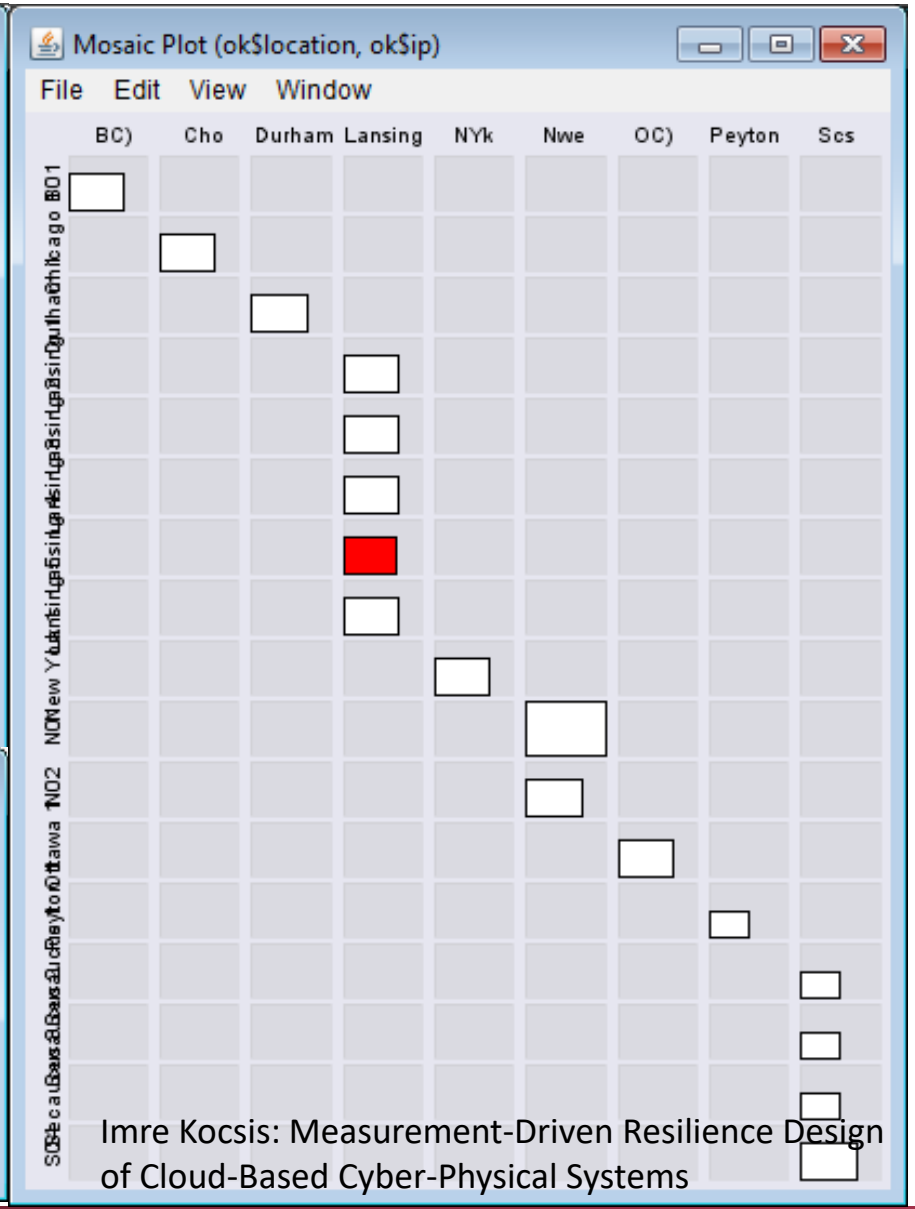


Exploratory Data Analysis

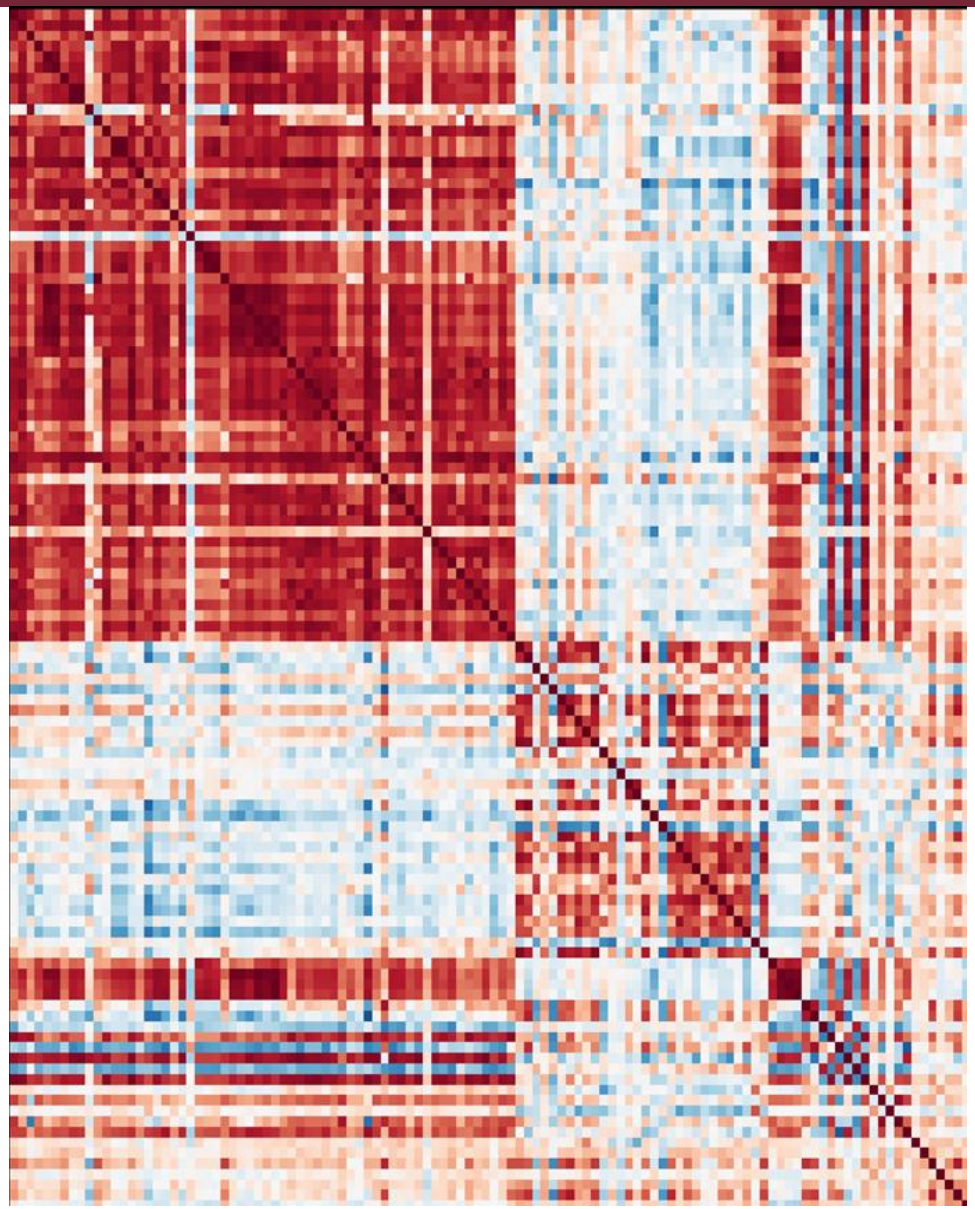
- Goal: hypothesis formulation
- Pattern recognition
- Early validation
- „Sensors of type X are sensitive to high temperature”
- „Application of Type Y is sensitive to CPU load”
- Interactive, human expert needed
- Later: automated support(IBM Watson Analytics)

EDA example

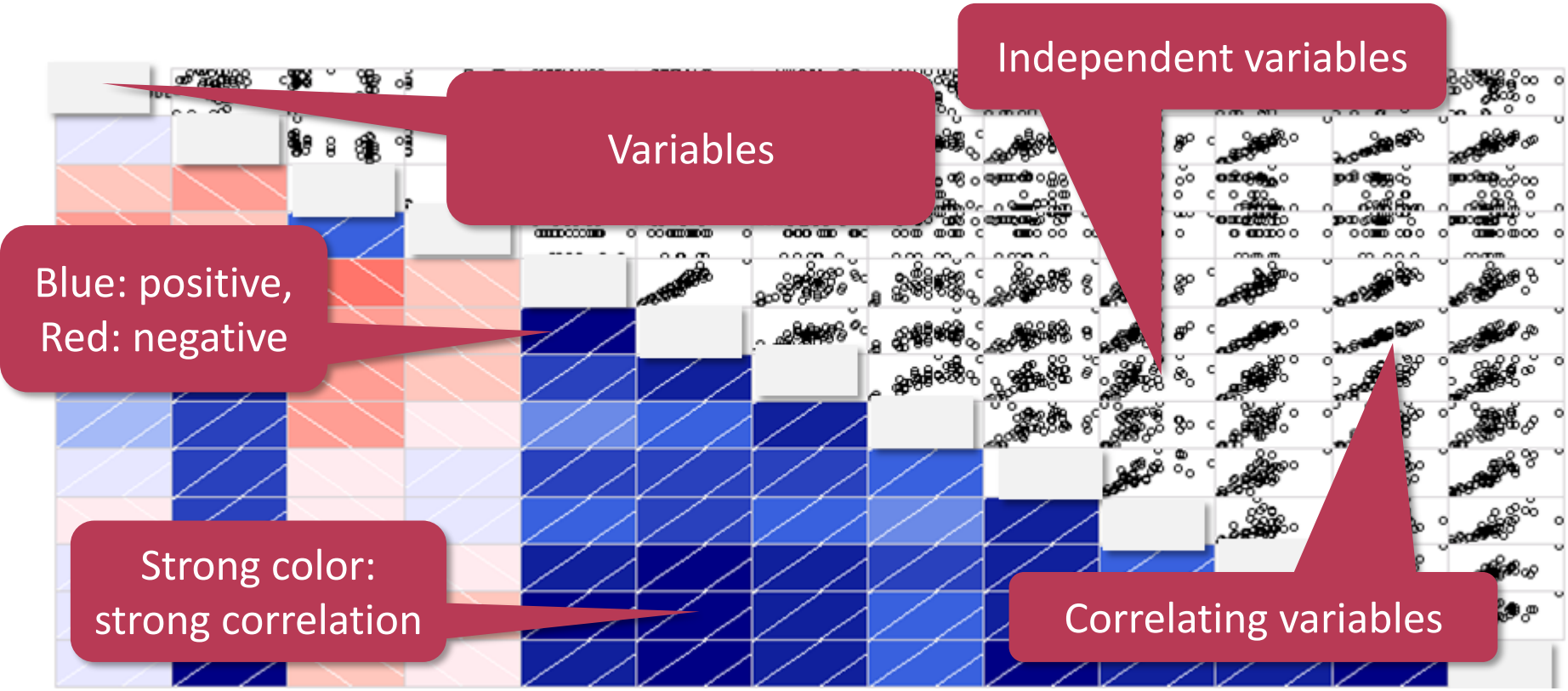

Scatterplot (ok$RT vs ok$RTT)
File Edit View Window
6000
4000
2000
Barchart (ok$location)
File Edit View Window
BC)
Cho
Drm
Lng
NYk
Nwe
OC)
Peyton
Scs


Mosaic Plot (ok$location, ok$ip)
File Edit View Window
BC)
Cho
Durham
Lansing
NYk
Nwe
OC)
Peyton
Scs

EDA example2

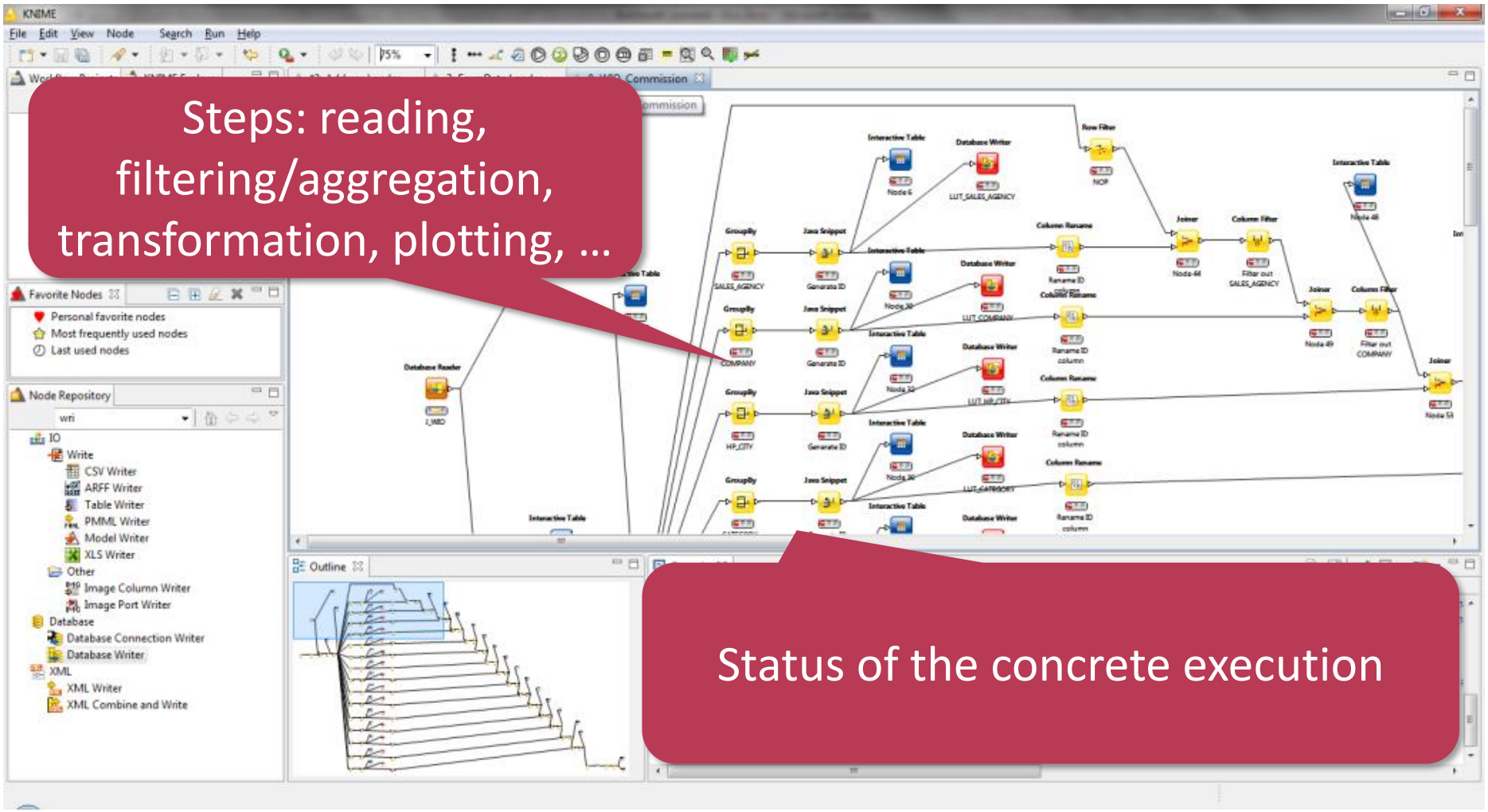


EDA example3: pairwise correlation



R „corrgram package”
Pearson linear correlation
Scatterplot matrix
Goal: identify correlating variables, outliers
→ Dimension reduction
→ Feature selection

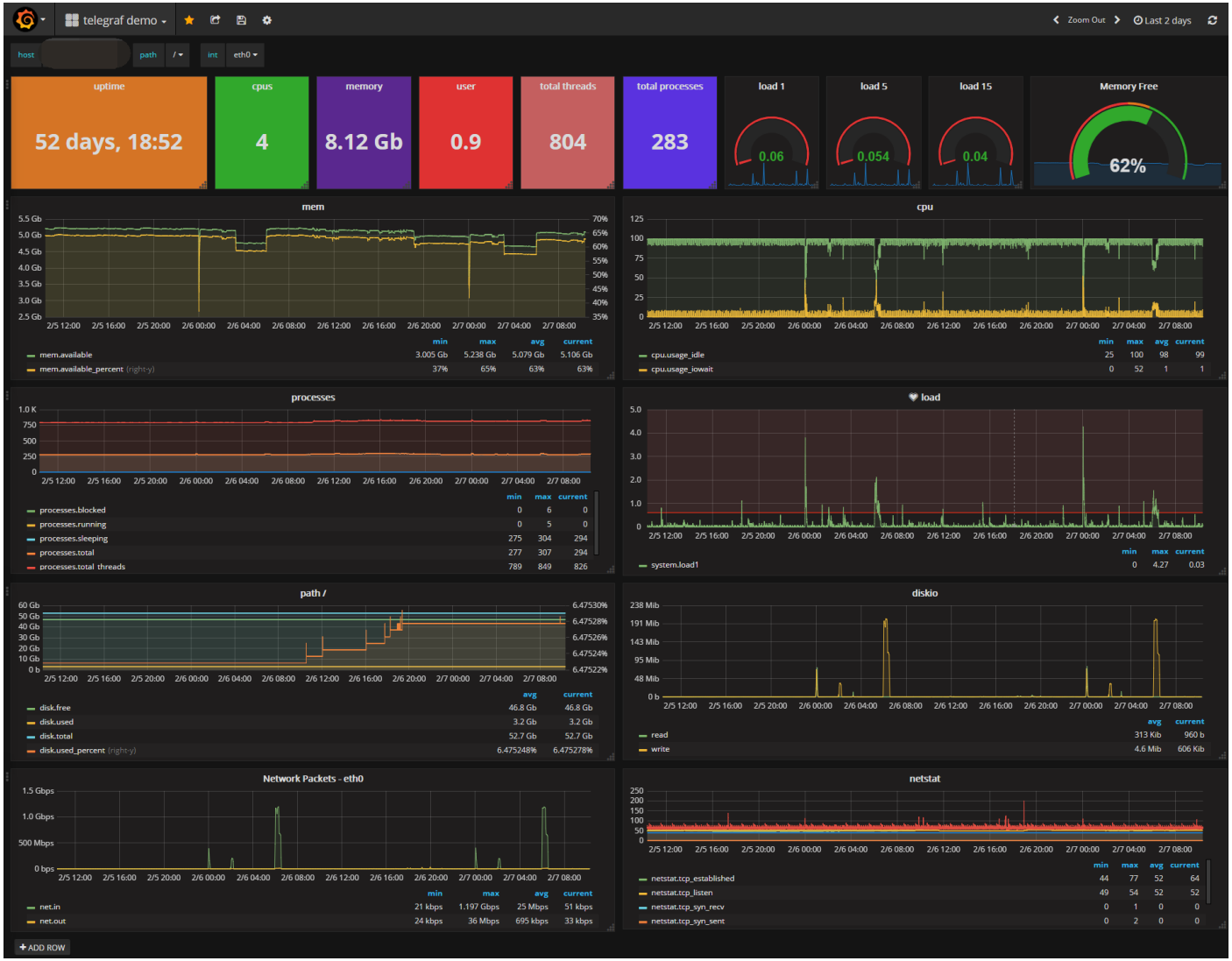
Example data processing workflow (KNIME)



Reporting/dashboard

- Goal: communication
- Report: snapshot
- Dashboard
  - Interactive: filtering, selection
  - „Drill-down” like in OLAP

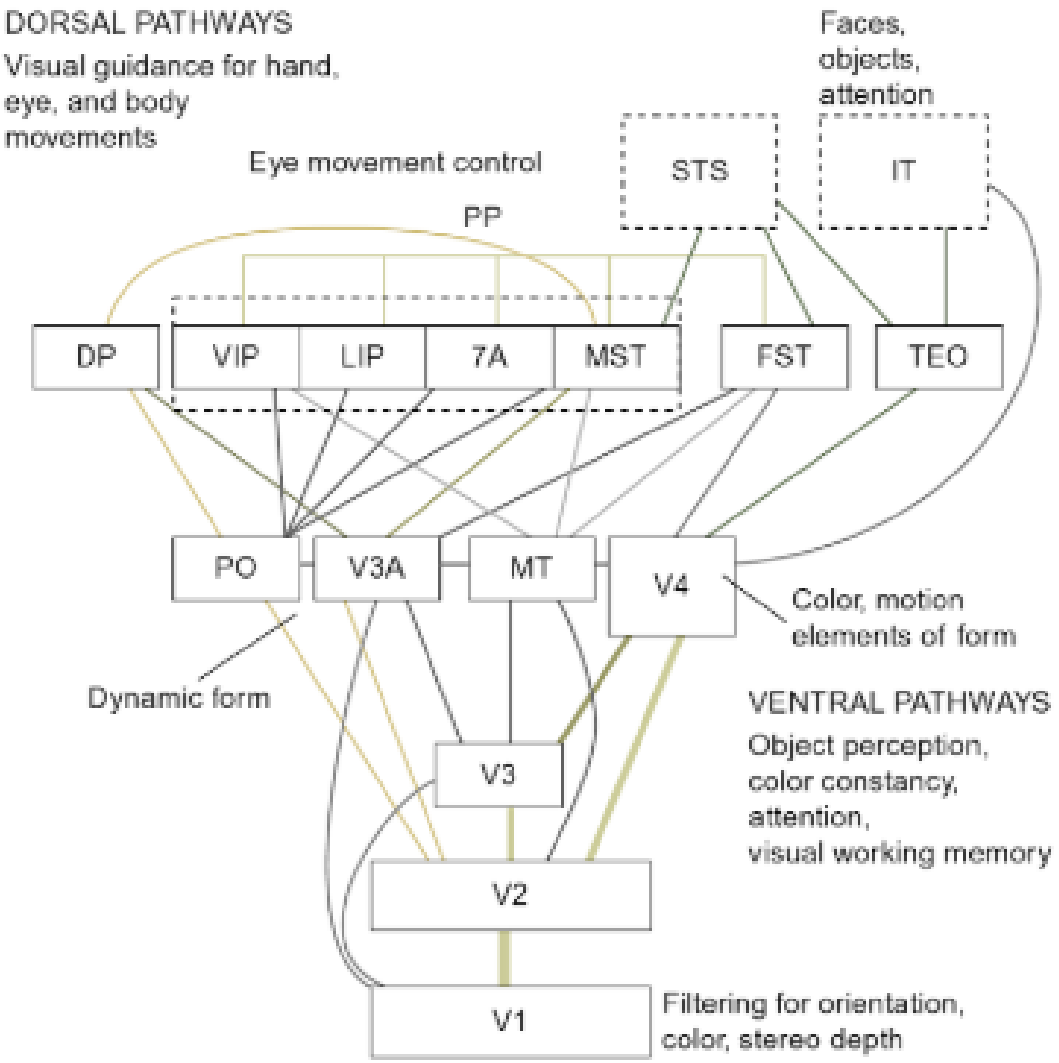
Dashboards… (e..g Grafana)



https://grafana.com/dashboards/1443

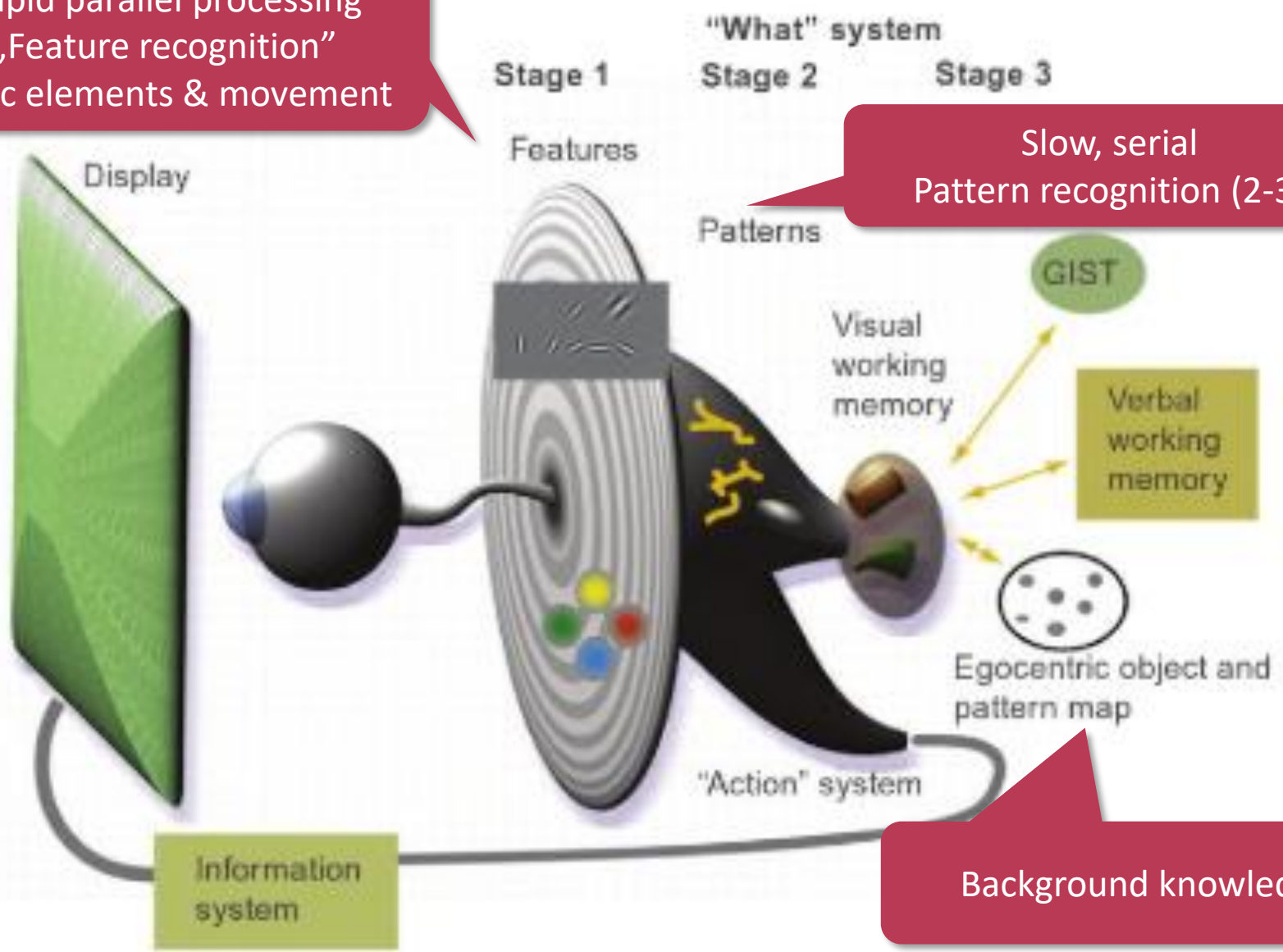
VISUALITAZION/PERCEPTION

Visual pathways



3-stage model

Rapid parallel processing
„Feature recognition”
Basic elements & movement



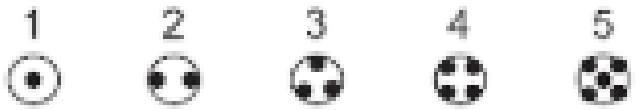
Slow, serial
Pattern recognition (2-3)

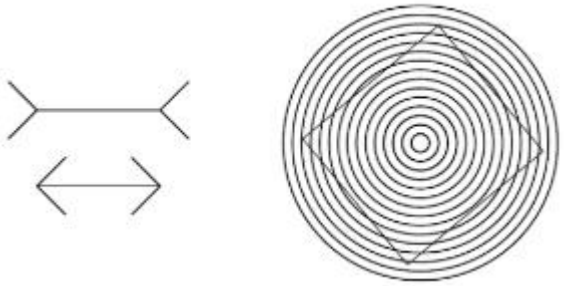
Background knowledge

Sensory vs symbols

- Sensory
  - Without training
  - Hard to give alternative meaning
  - Fast to understand
  - No cultural dependence
- Symbols
  - Harder to learn
  - Easy to forget
  - Cultural dependence BUT
  - Formally powerful

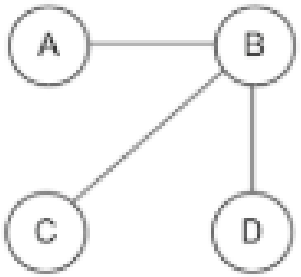
Examples

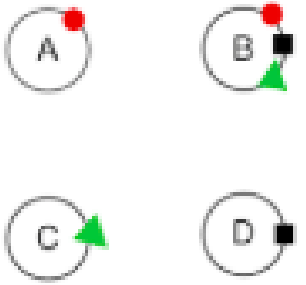




Muller-Lyer

Escher, Penrose triangles: Waterfall 1961

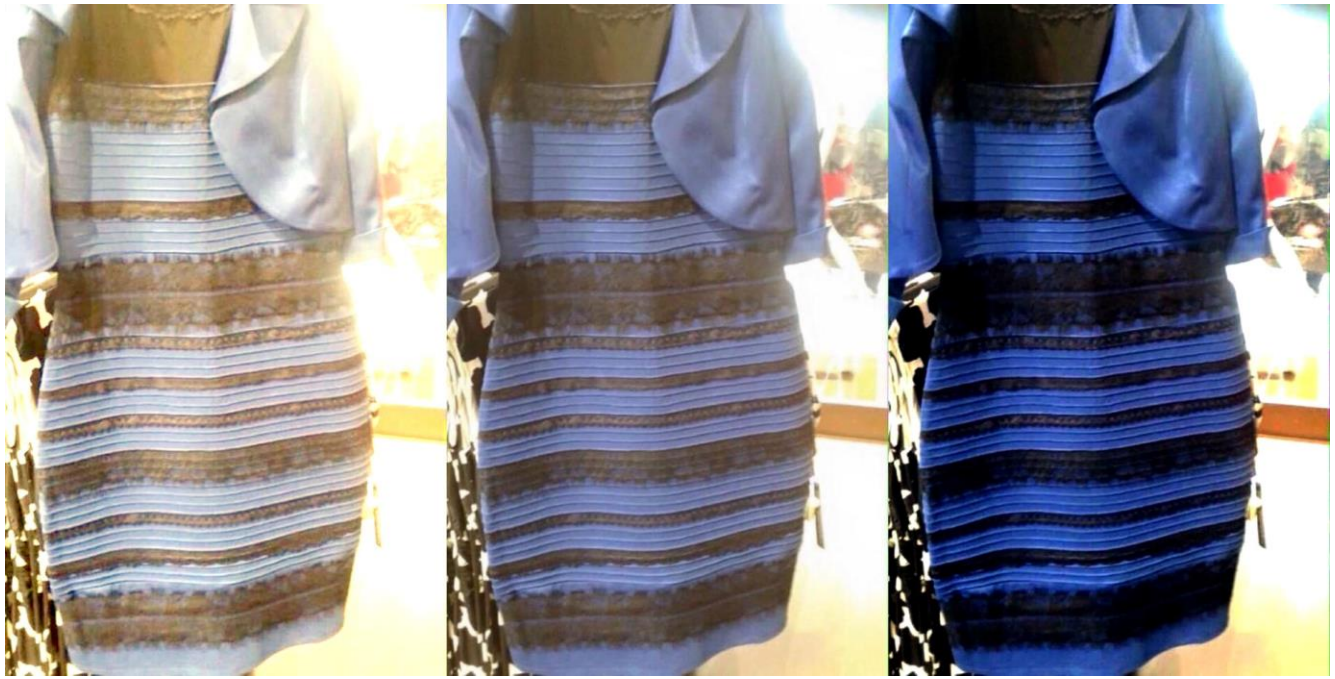




Ware: Information Visualization: Perception for Design

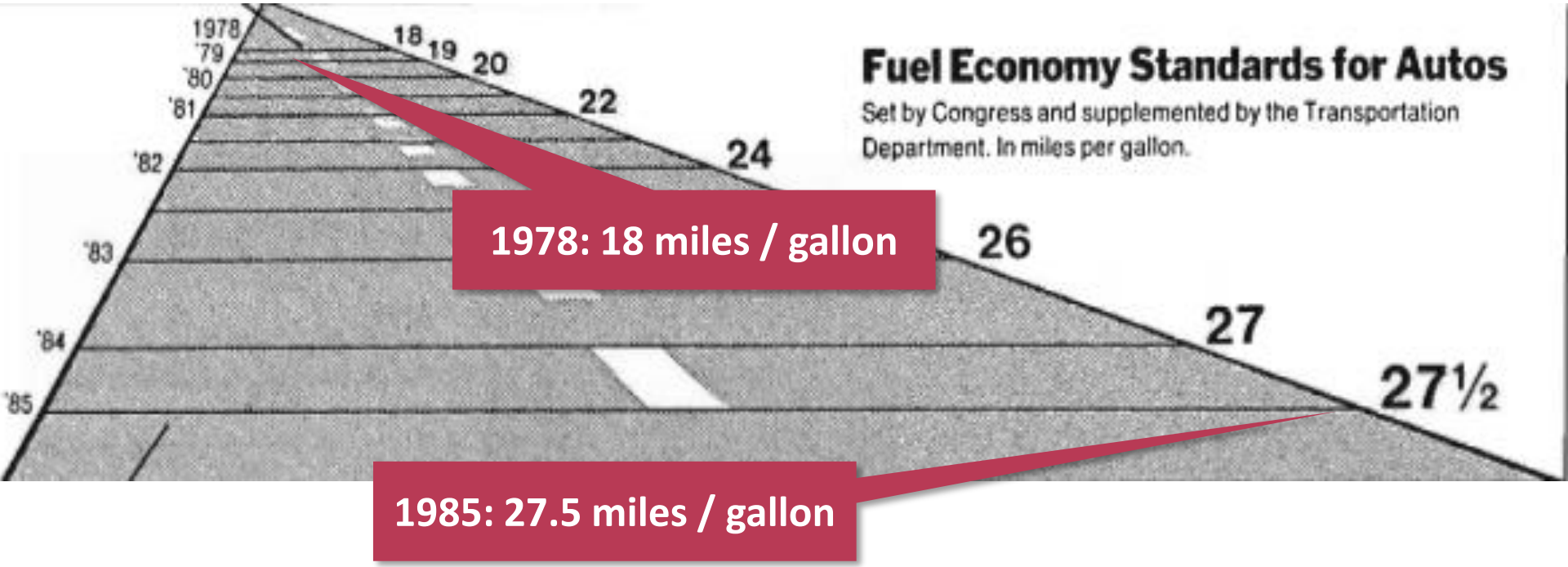
Colors?

Most people see the red closer than the blue, but some see the opposite effect



https://www.wired.com/2015/02/science-one-agrees-color-dress/

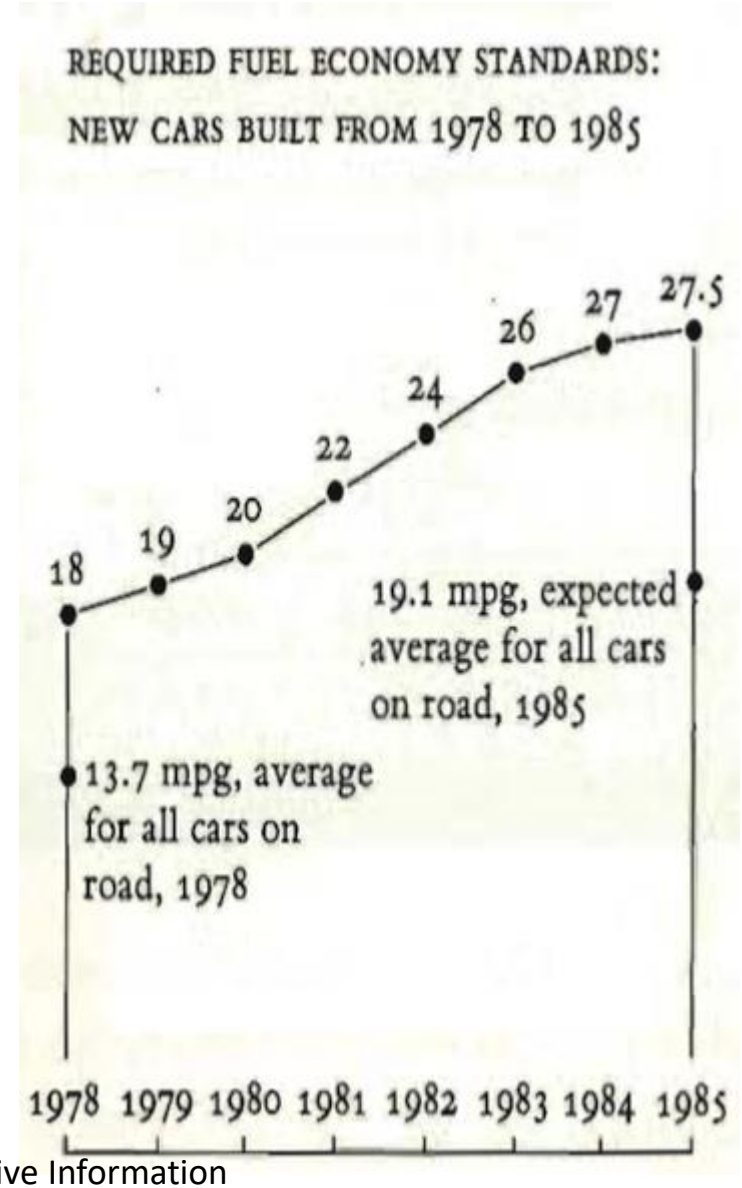
Power of distortion



Lie Factor = Size of Effect in Graphic / Size of Effect in Data = 7.85 / 0.53 = 14 !!!

New Yort Times, August 9, 1978
Tufte: The Visual Display of Quantitative Information

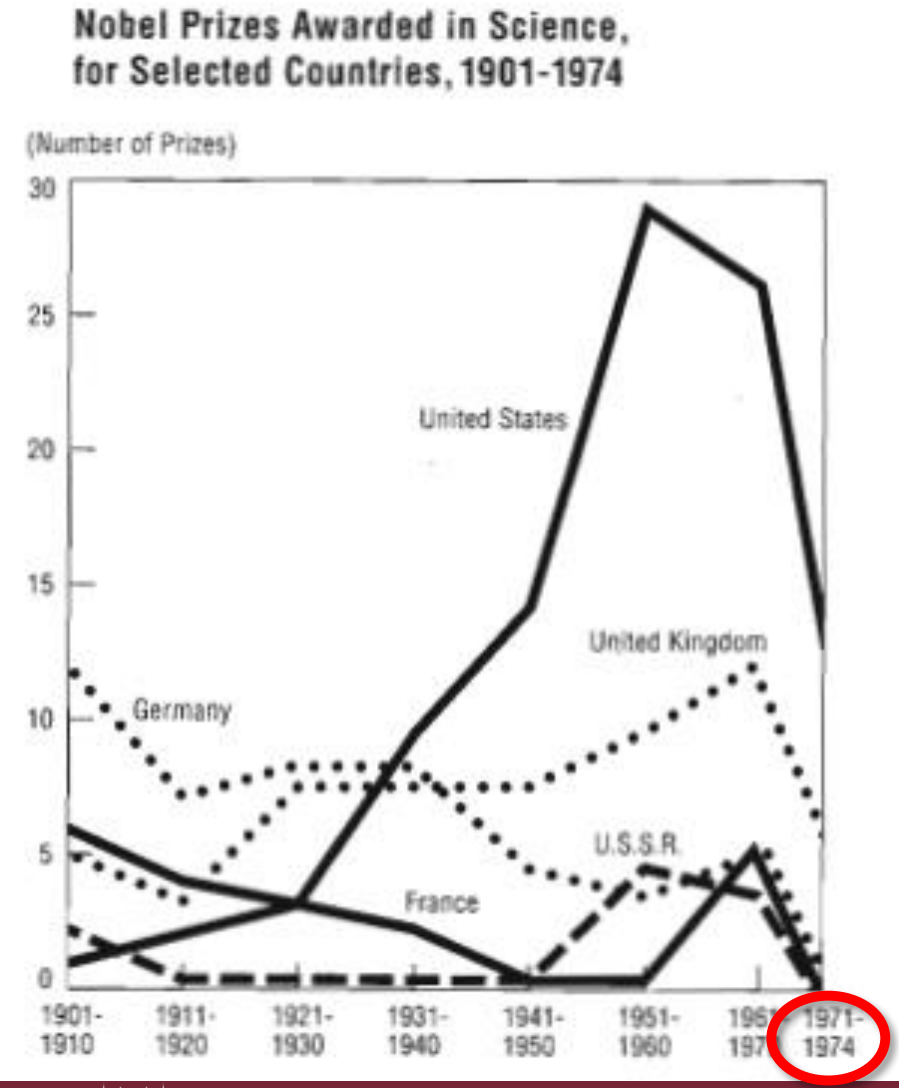
A better visualization

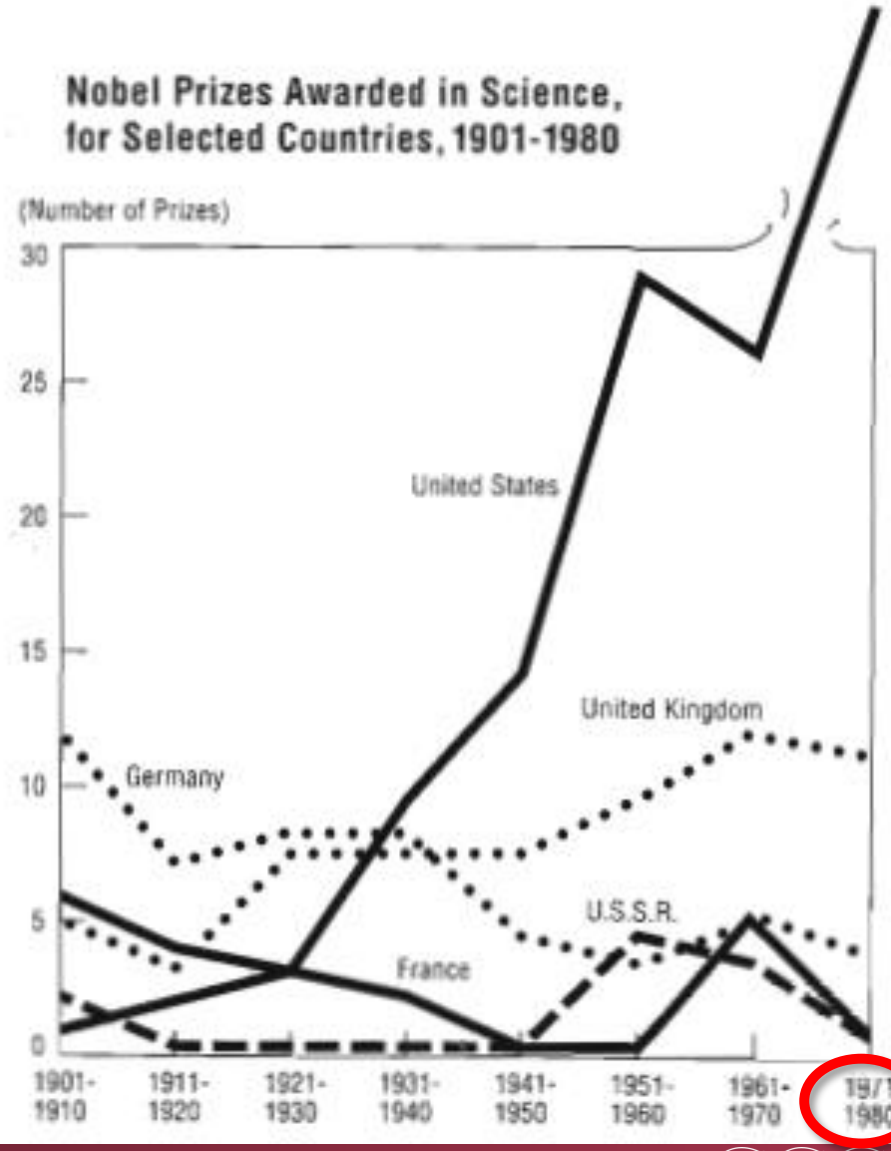


Tufte: The Visual Display of Quantitative Information

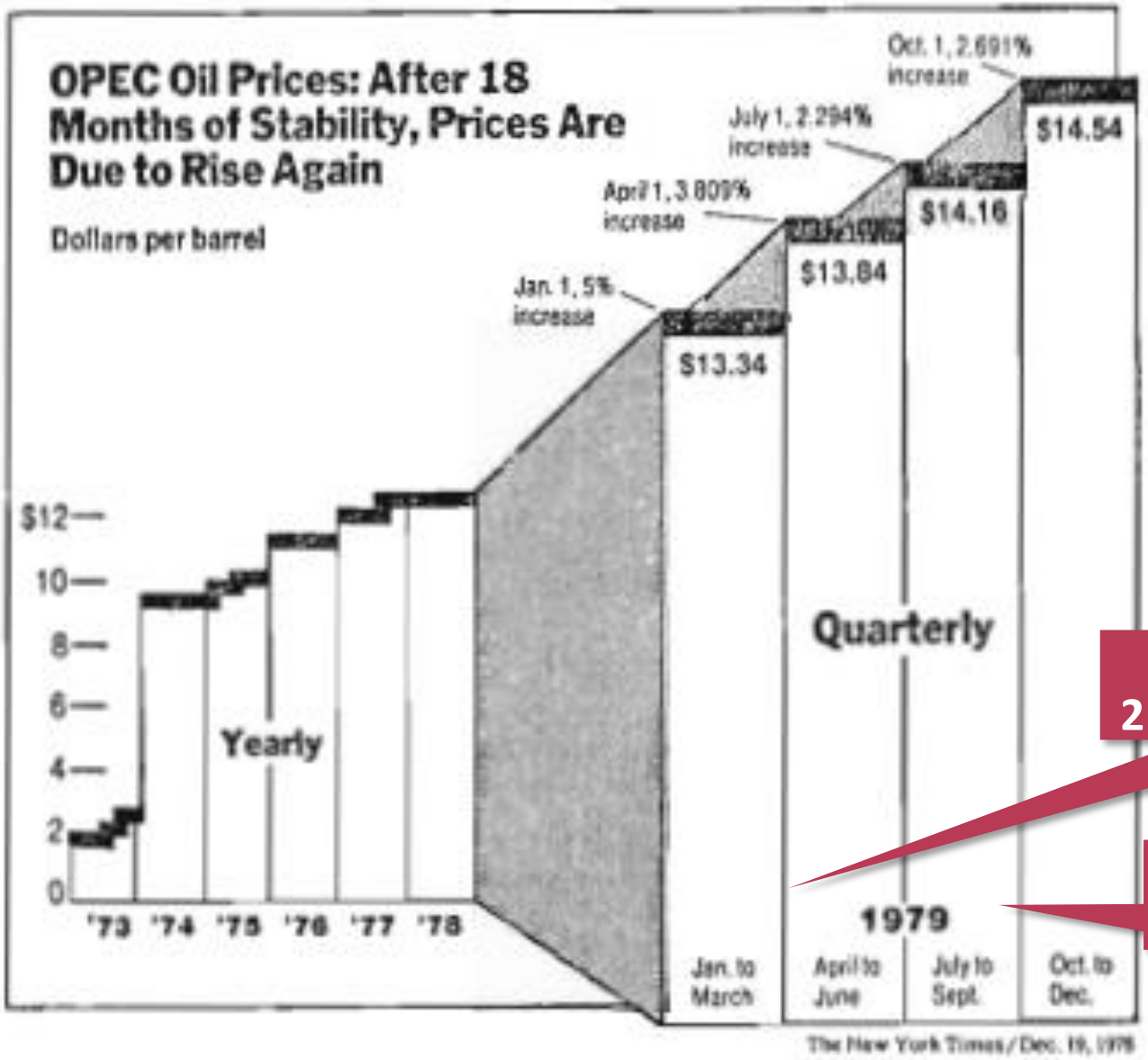
Uniform scale + extrapolation

National Science Foundation, *Science Indicators*, (Washington DC, 1976), p15.
Tufte: The Visual Display of Quantitative Information





Design variation vs data variation



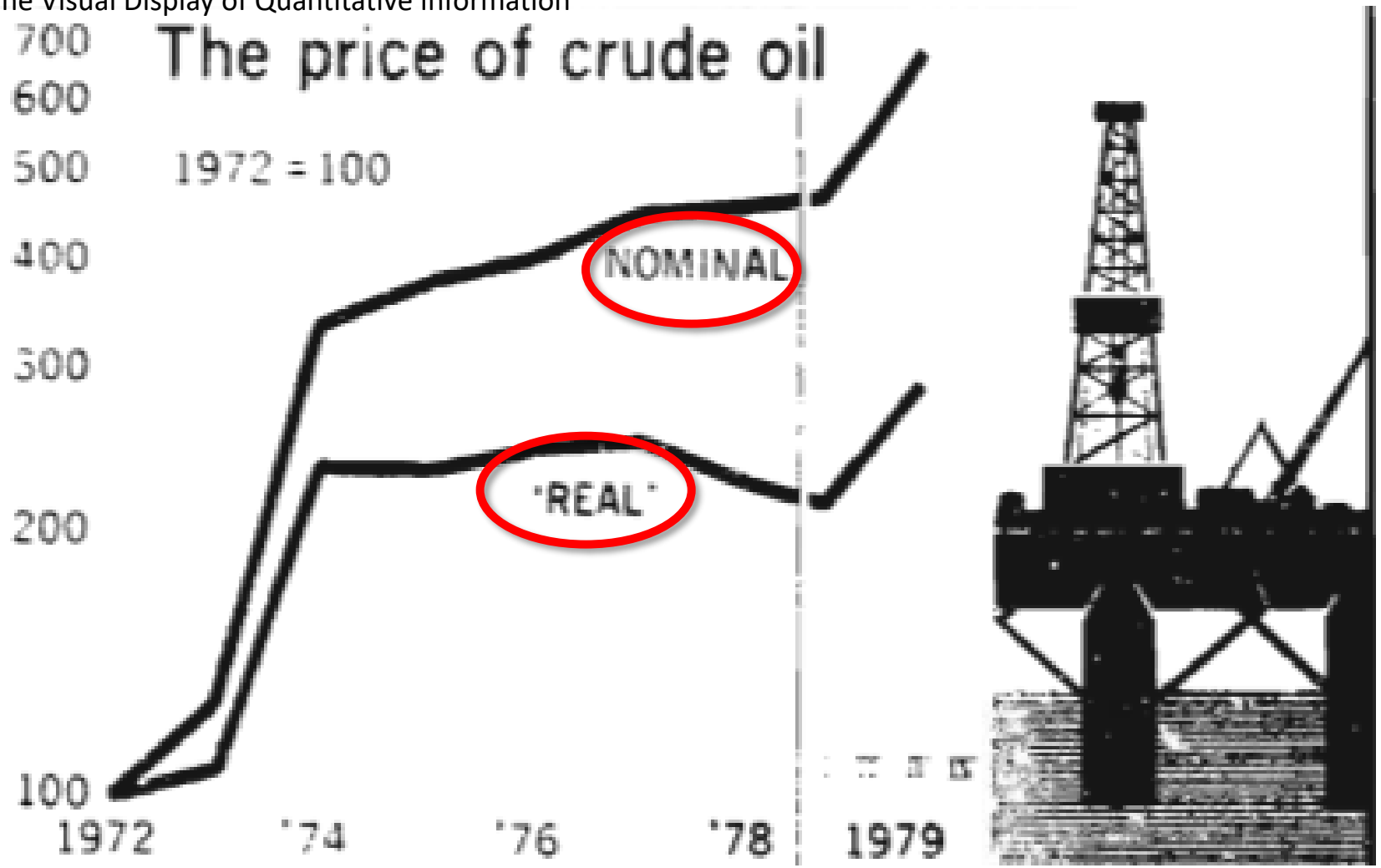
New Yort Times, Dec 19, 1978, p D-7
Tufte: The Visual Display of Quantitative Information

5 vertical scales (>1:2)
2 horizontal scales (>1:6)

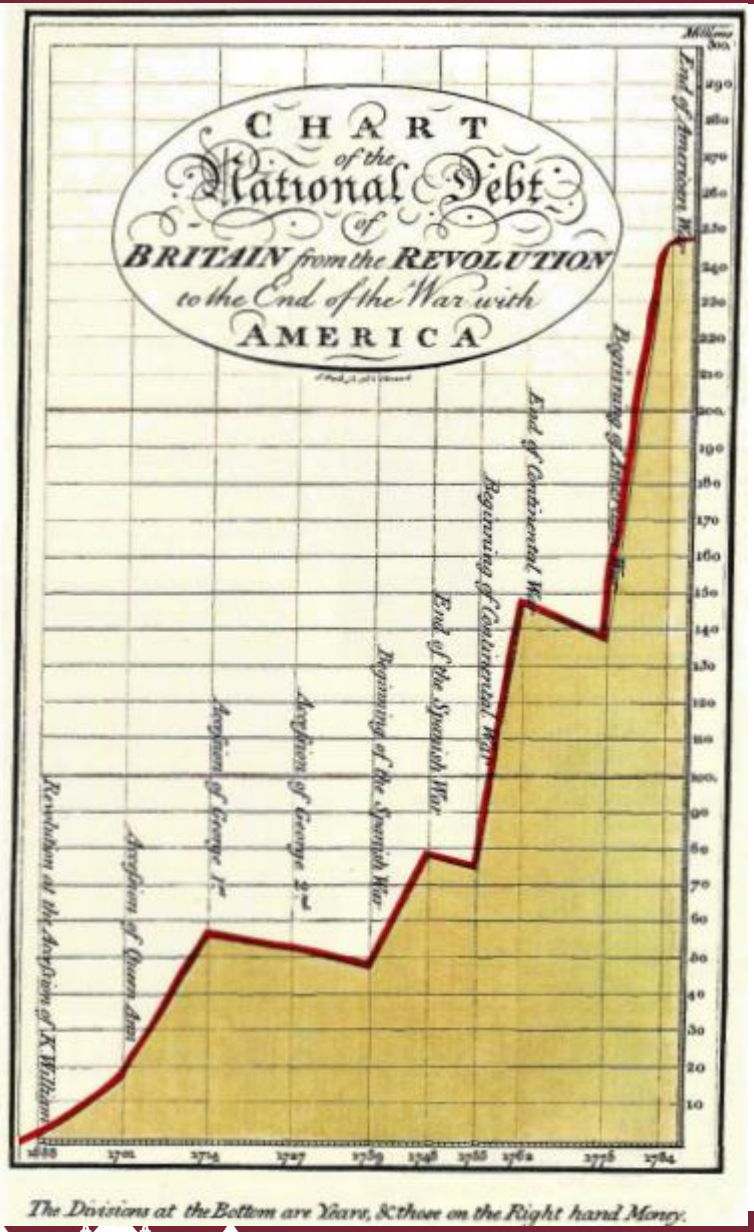
Lie Factor = 15.1 !

Basis of comparison? Standardization?

Sunday Times, (London), Dec 16 1979, p 54
Tufte: The Visual Display of Quantitative Information

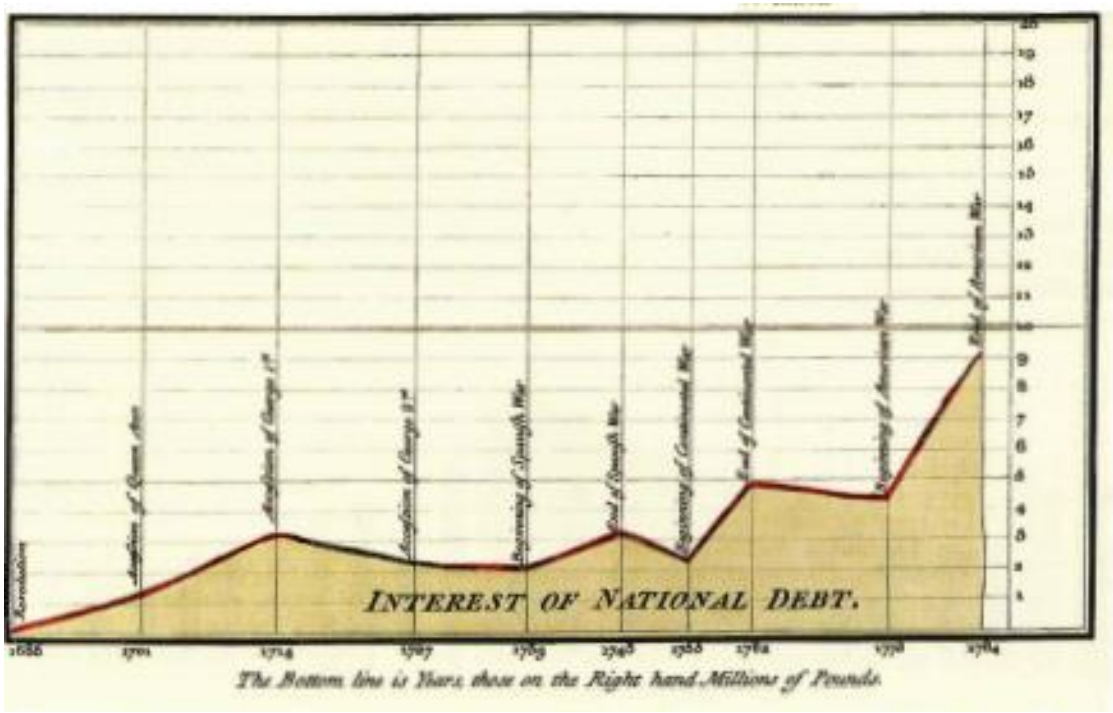


Plot ratio



Nominal vs Real money
Effect of distortion

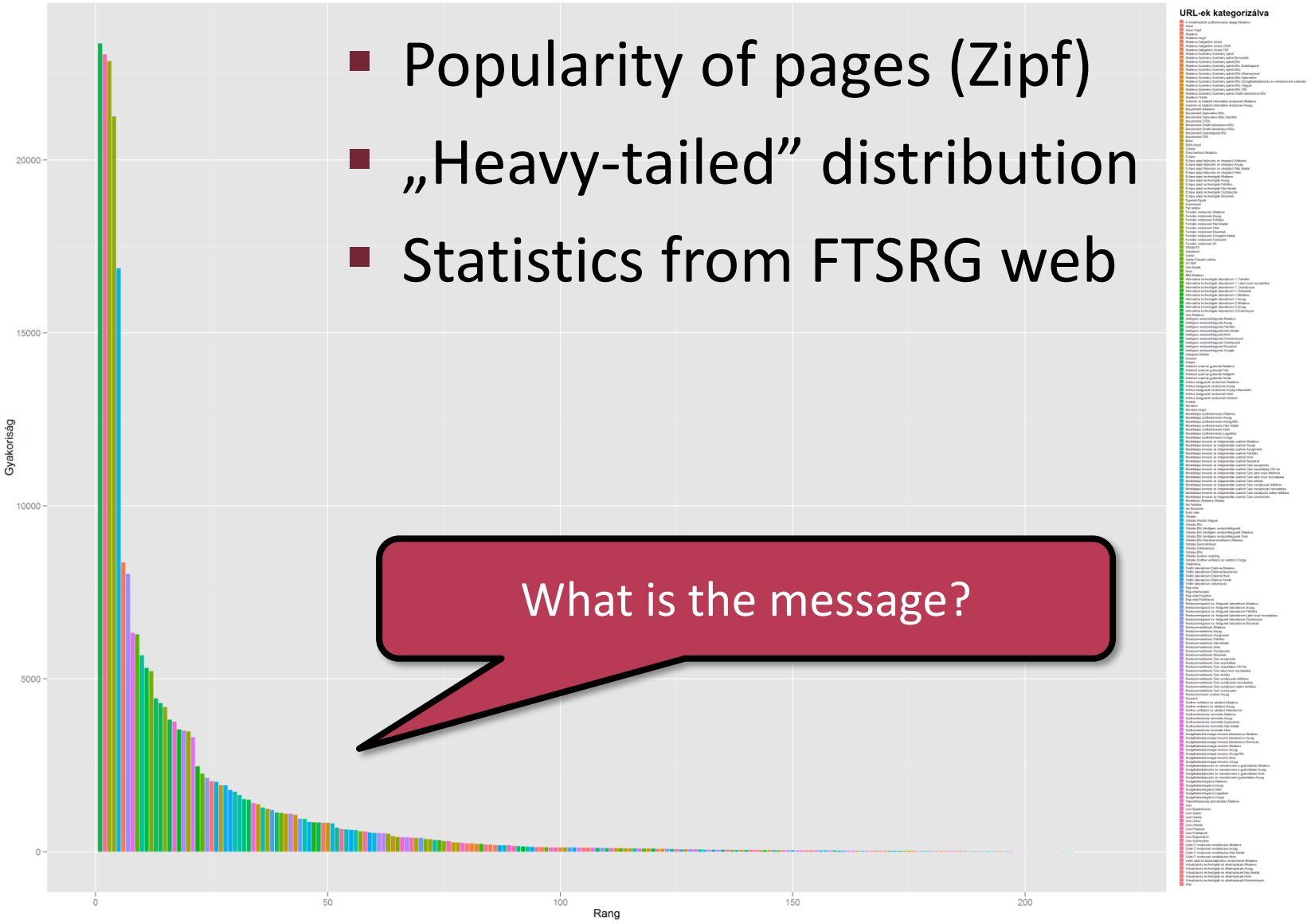
William Playfair, 1786 (!)



EXAMPLES

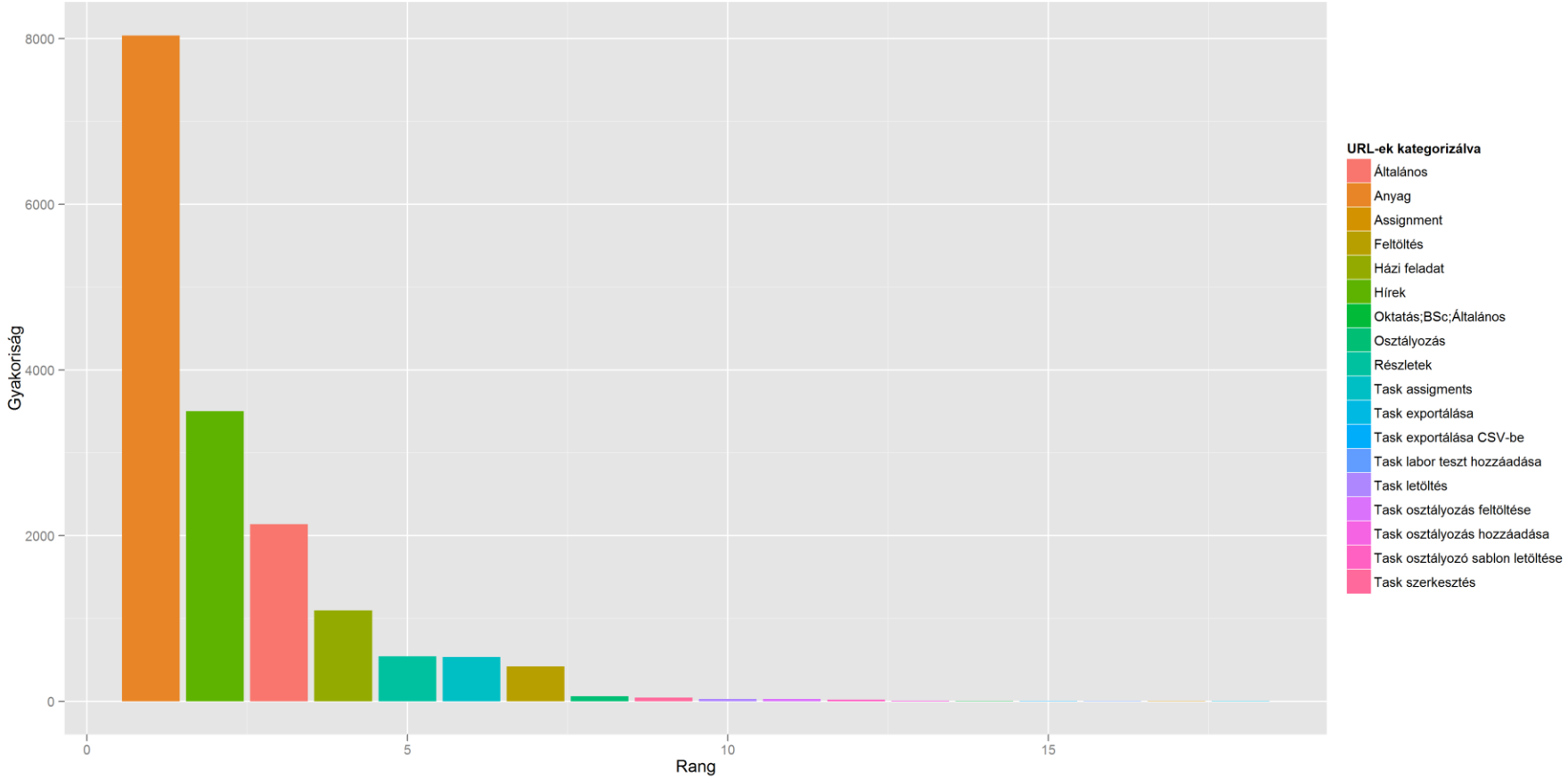
Data granularity & focus

- Popularity of pages (Zipf)
- „Heavy-tailed” distribution
- Statistics from FTSRG web



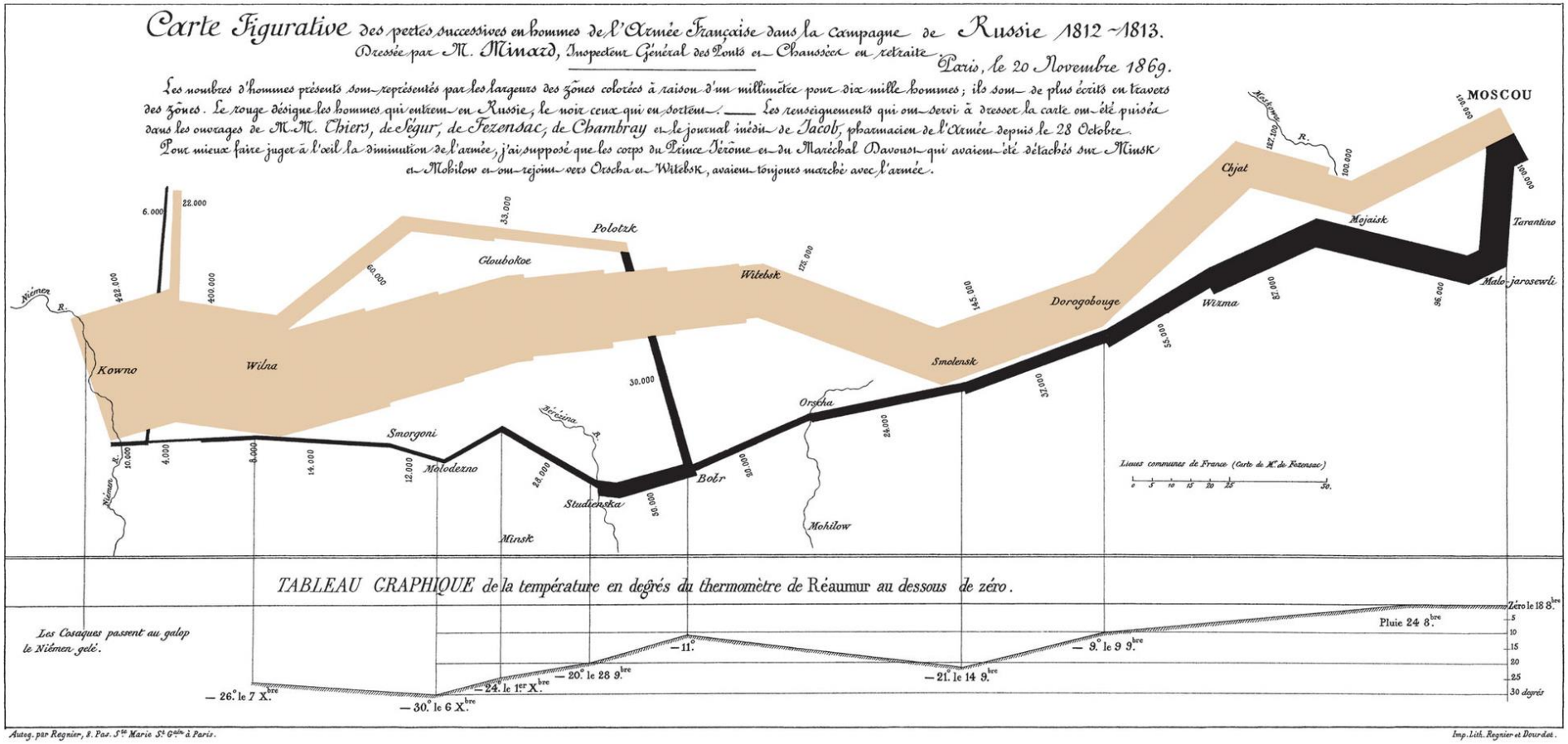
Better focus: statistics of a course

- Filtering of data
- Categorization



INFOGRAPHICS

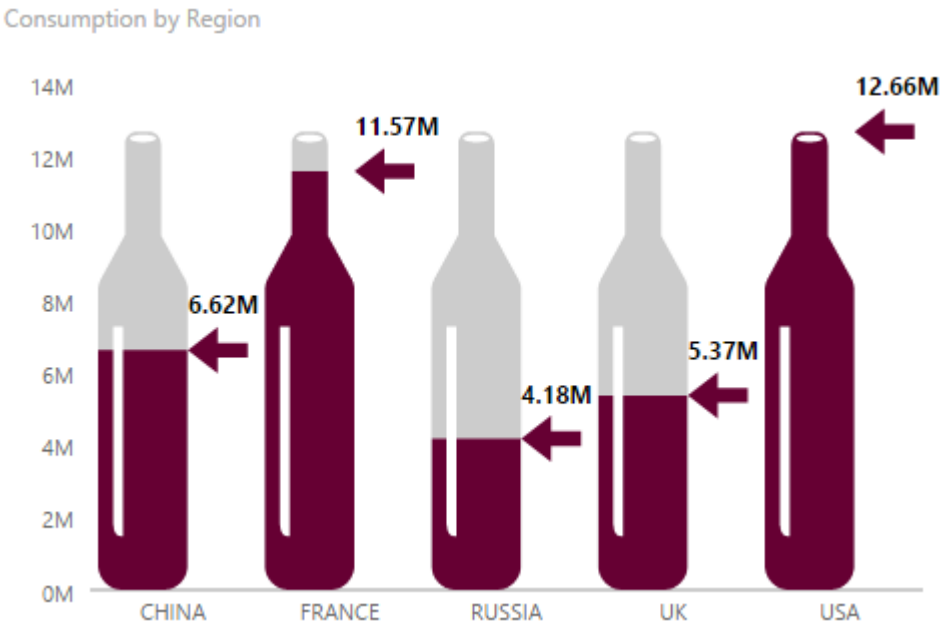
First infographic :2 dimensions, 6 types

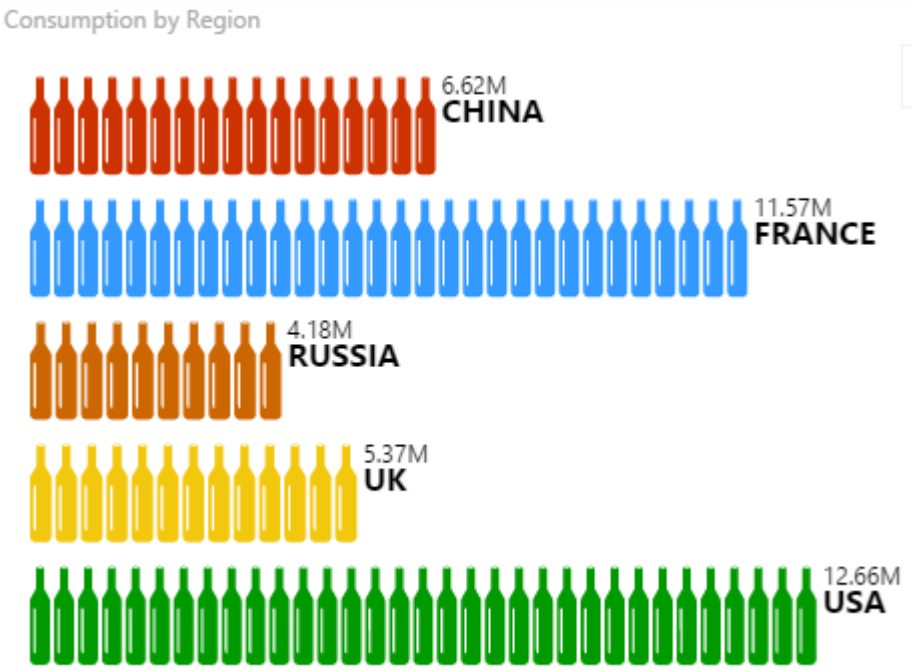


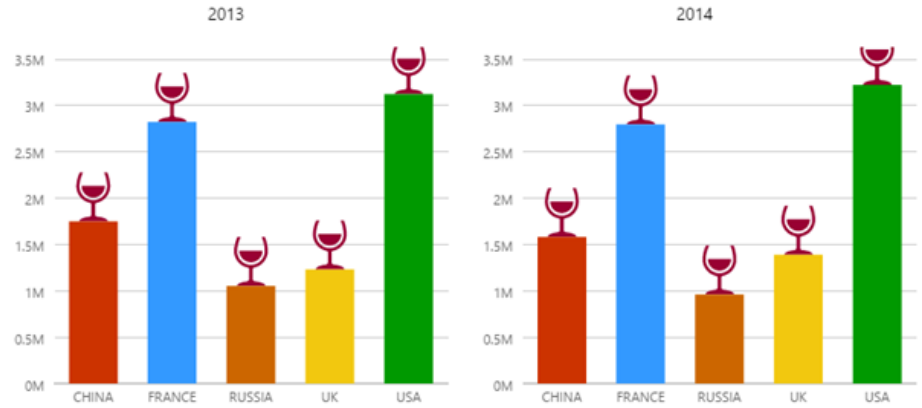
Charles Minard, 1869, https://en.wikipedia.org/wiki/Charles_Joseph_Minard

https://robots.thoughtbot.com/analyzing-minards-visualization-of-napoleons-1812-march

Infographics combined with plots

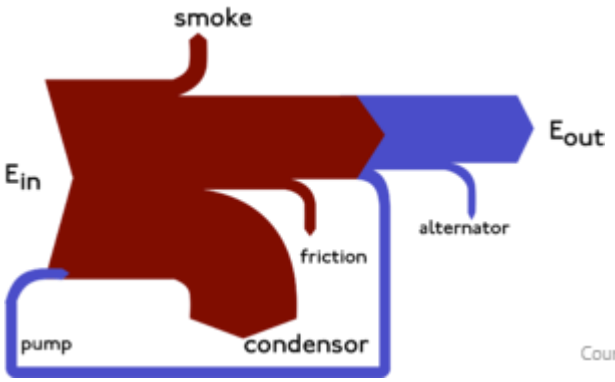


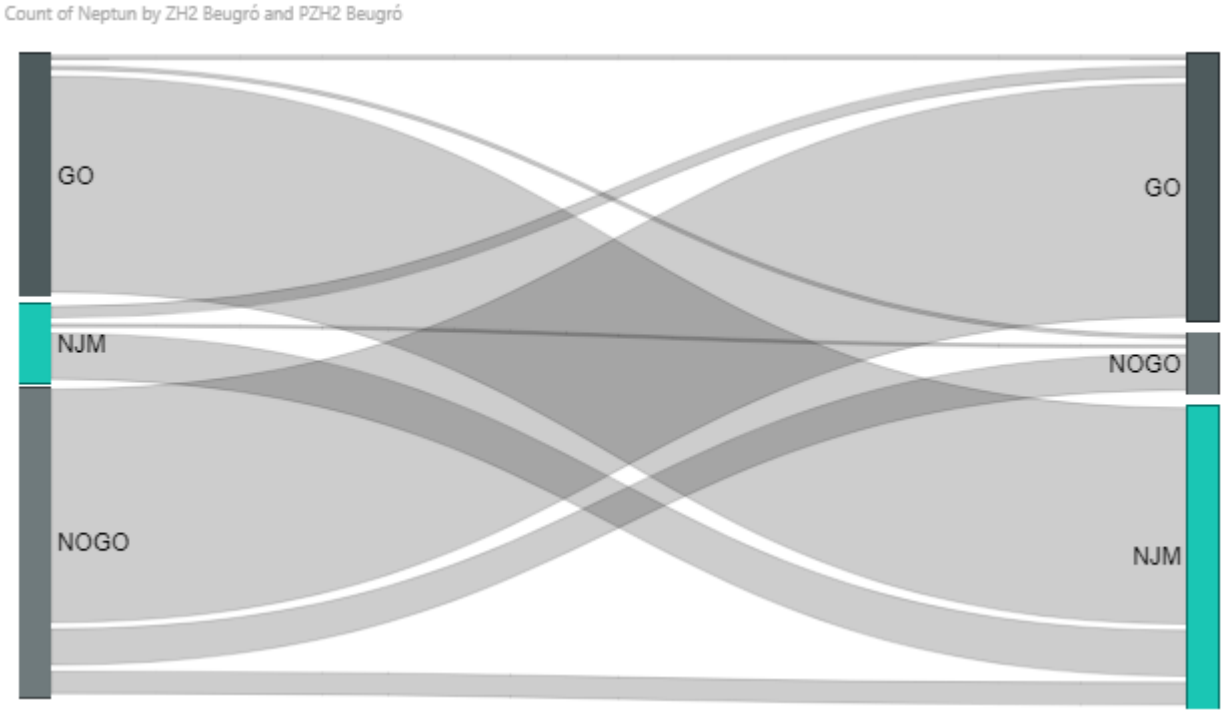




https://powerbi.microsoft.com/en-us/blog/quickly-create-infographics-with-the-infographic-designer-custom-visual-for-power-bi/

Later: Sankey





References

- Colin Ware: Information Visualization: Perception for Design, Morgan Kaufmann, 2012.
- https://guides.library.duke.edu/datavis/vis_types
- https://datavizcatalogue.com/
- Tukey, John W. *Exploratory data analysis*. Vol. 2. 1977.
- Tufte, Edward, and P. Graves-Morris. "The visual display of quantitative information.; 1983."
- Inselberg, Alfred, and Bernard Dimsdale. "Parallel coordinates for visualizing multi-dimensional geometry." *Computer Graphics 1987*. Springer, Tokyo, 1987. 25-44.
- Kocsis Imre. Vizuális analízis. In: Antal Péter, Antos András, Horváth Gábor, Hullám Gábor, Kocsis Imre, Marx Péter, Millinghoffer András, Pataricza András, Salánki Ágnes. Intelligens adatelemzés. 141 p. Budapest: Typotex Kiadó, 2014. pp. 58-69. (ISBN:978-963-2791-71-5)

What next?

- Plot types
- Variables
- Data structure
- Data quality
- Data cleaning
- Some stats