Visual Analytics

Lecture BigData Analytics

Julian M. Kunkel

julian.kunkel@googlemail.com

University of Hamburg / German Climate Computing Center (DKRZ)

2017-10-20

Disclaimer: Big Data software is constantly updated, code samples may be outdated.

Outline

Statistical Graphics [44]

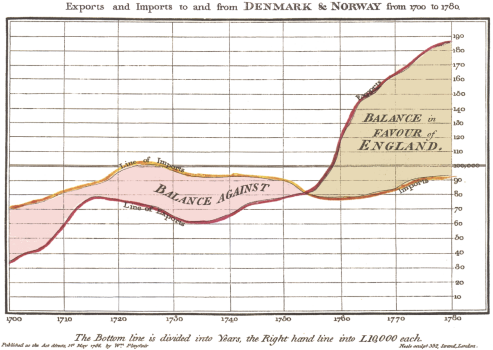
Definition: Graphics in the field of statistics used to visualize quantitative data

Objectives

- The exploration of the content of a data set
- The use to find structure in data
- Checking assumptions in statistical models
- Communicate the results of an analysis

Plots (Excerpt)

- Scatter, box, histograms
- Statistical maps
- Probability plots
- Spaghetti plots
- Residual plots



Source: William Playfair's Time Series of Exports and Imports of Denmark and Norway [44]

Visual Analytics [32]

Definition [33]

*The science of **analytical reasoning** facilitated by **interactive visual interfaces**.*

Objective

- Solve complex questions/time critical problems **applying the scientific method**
- Present gained insight / communicate it visually

Analytical tasks

- Understanding past situations; trends and events that caused current conditions
- Monitoring events for indicators for an emergency
- Identifying possible alternative future scenarios and their warning signs
- Determining indicators of the intent of an action or an individual
- Supporting decision makers in times of crisis

Visual Analytics Workflow

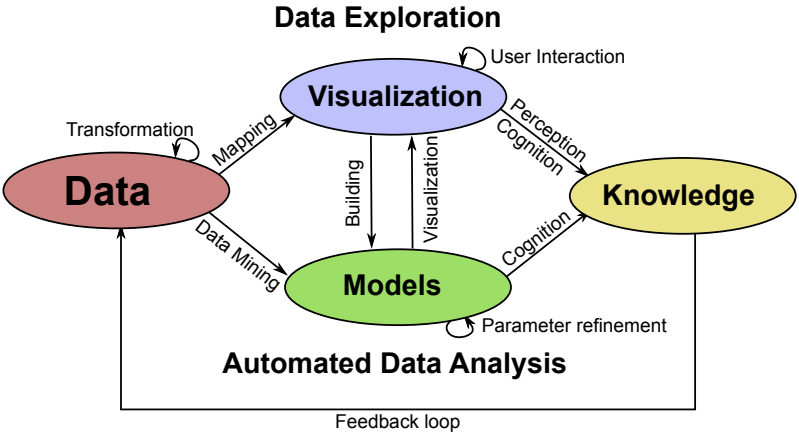
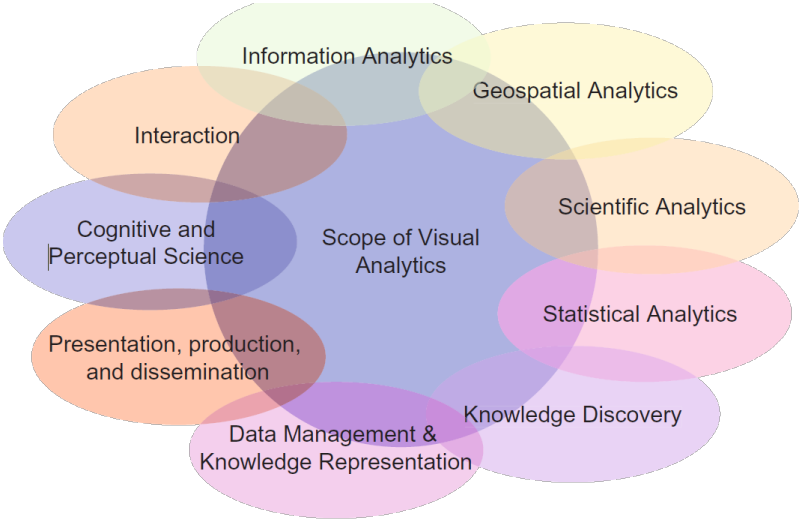


Figure based on [48]

Motto: Analyse First – Show the Important; Zoom, Filter and Analyse Further – Details on

Fields of Visual Analytics



Source: Visual analytics: Scope and challenges [48]

Human-Computer Interaction

Why do we team humans and computers using a visual interface?

Comparing capabilities of humans and computers

- Human brain processing power is enormous
  - 100 billion neurons, linked together by many synapses
  - Synapses fire with $4.3 \cdot 10^{15}$ spikes/s; data rate of $1.1 \cdot 10^{16}$ bits/s = 125 TiB/s; 20 Watt [6]
  - Fastest supercomputer in the world [7]: Sunway TaihuLight: 125 TFlop/s, 15 MW
  - Estimation: Simulating one second of human brain activity requires 83k processors
- Strength of humans and computers:

Human	Computer
Pattern recognition	Execution of algorithms
Creative thinking	Accuracy
Processing new infos	

- Visual perception and analysis capabilities exceed computers, e.g., computer vision
  - Vision uses 30-50% of the brain's capabilities

$\Rightarrow$ Visual representation and analytics is key for efficiency

Example Analysis Session: Demo

Based on a real case [35]

- 1854, Broad Street, London
- Within a few days people died mysteriously
- Dr. John Snow investigated the cause to stop “disease”
  - He analyzed data visually with the scientific method
- We will follow his analysis steps
  - Using modern data analytics tools

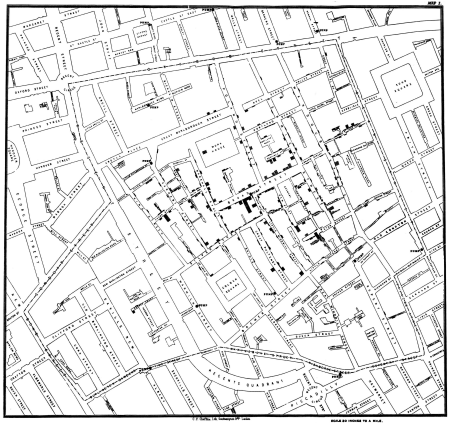
Interactive lab notebook

- Record notes/hypothesis, type code, store it together with results
- The notebook is prepared using Jupyter with Python

Analysis Results

- John found the source of the Cholera: The pump
  - He claimed the disease is spread by the water
  - John is one of the founders of our Germ theory
- They unmounted the pump handle
  - But could not proof theory
- Board of health did not believe his analysis
  - They believed “Miasma” is the cause

  $\Rightarrow$ Convincing documentation is important!

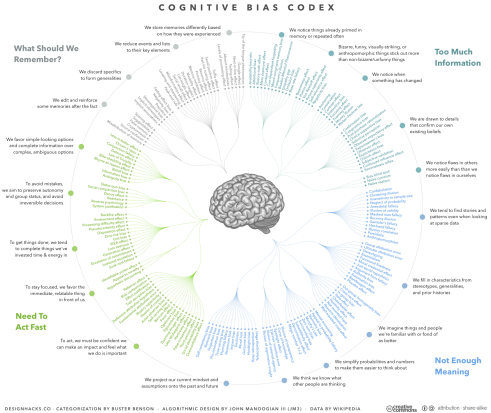


Original map made by John Snow in 1854.
Cholera cases are highlighted in black. [35]

Cognition

Definition: The mental action or process of **acquiring knowledge** and **understanding** through thought, experience, and the senses [46]

- **Communicated** information and **interpretation** is biased by humans due to:
  - Perception
  - Information processing
  - Subjective knowledge
- Psychology knows many **cognitive biases** [40]
- Categories of cognitive biases:
  - Limits of memory
  - Too much information
  - Not enough meaning
  - Need to act fast
- Categories serve as guidelines for visual analytics
- We will focus on visual perception

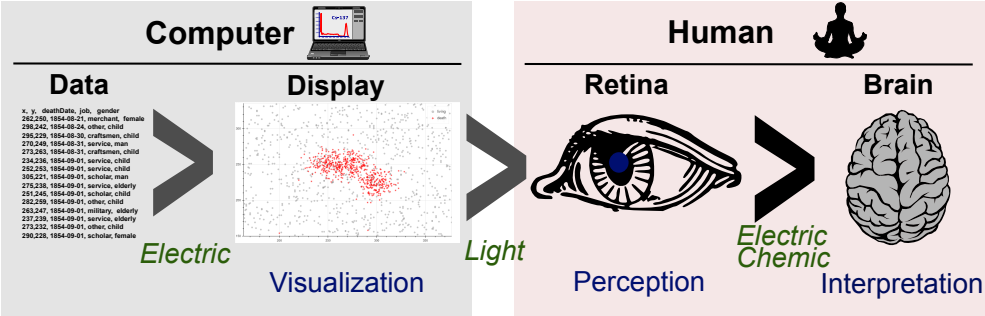


Source: Wikipedia's complete (as of 2016) list of cognitive biases, beautifully arranged and designed by John Manoogian III (jm3). Categories and descriptions originally by Buster Benson. [40]

Visual Perception: Information Pipeline

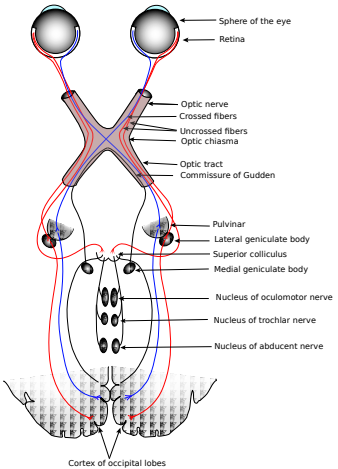
Information Communication

- Information is transformed several times from digital data to human
- The retina and brain interprets visual information
- Efficient communication requires to understand **human perception**



Optical Illusions [38]

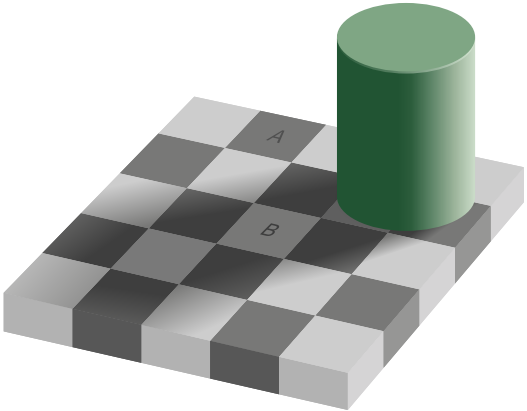
- Definition: visually **perceived images** that differ from **objective reality**
  - They are caused by the **visual system**
- They are many different types of illusions
  - Perceived colors and contrasts
  - Size and shapes of objects
  - Interpretation of objects
  - Depth perception
  - Moving of objects
  - Afterimages
  - ...



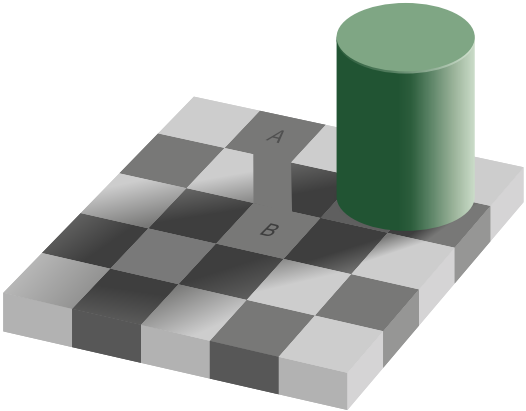
Source: Gray's Anatomy depiction of the optic nerves & nuclei... KDS444 [39]

Color Illusion

Field A and B have the same gray tone



Source: The checker shadow illusion. Edward H. Adelson [38]



Proof: Breaking the illusion.
Source: Edward Adelson [38]

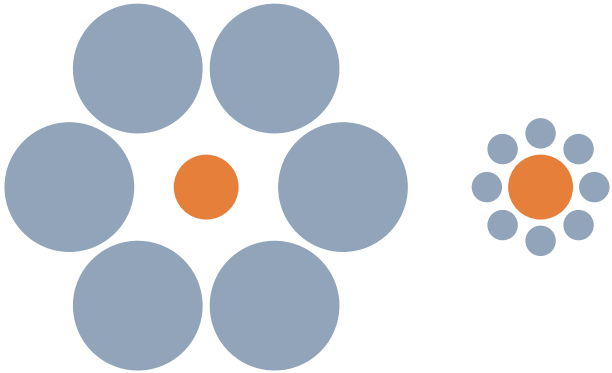
Color Illusion (2)

Form that seems to be filled in yellow instead of white

Source: Blue-bordered cookie that misleadingly seems to be filled with light yellow water-color. Jochen Burghardt. [38]

Shapes of Objects

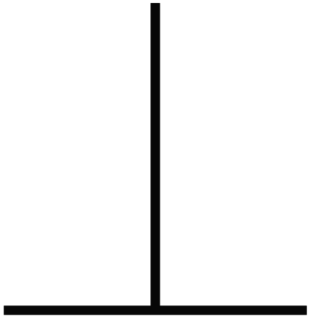
Both orange circles are the same size

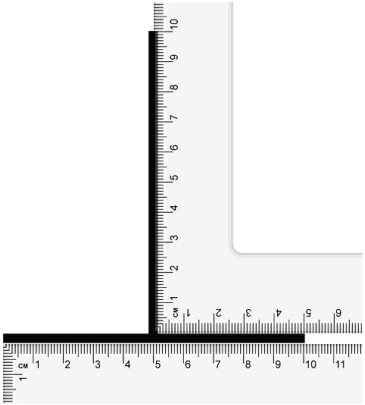


Source: Optical illusion: The two orange circles are the same size. [38]

Shapes of Objects (2)

Vertical and horizontal lines have the same length

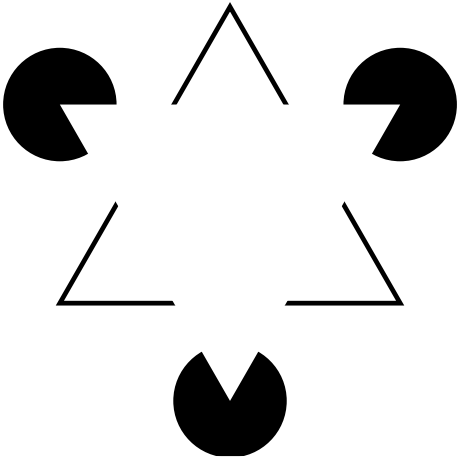




Source: Vertical–horizontal illusion, S-kay [38]

Shapes of Objects (3)

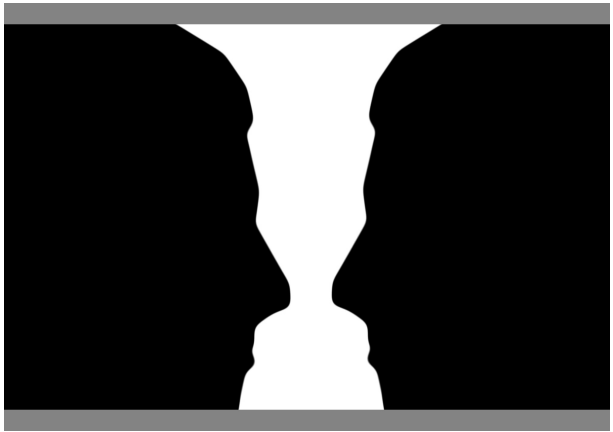
Imaging a white triangle in the center



Source: Kanizsa triangle. Fibonacci [38]

Interpretation of Images

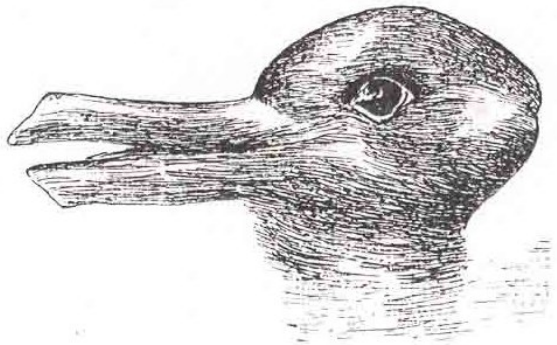
Vase or two faces



Source: Two silhouette profiles or a white vase?, Brocken Inaglory [38]

Interpretation of Images (2)

Duck or rabbit

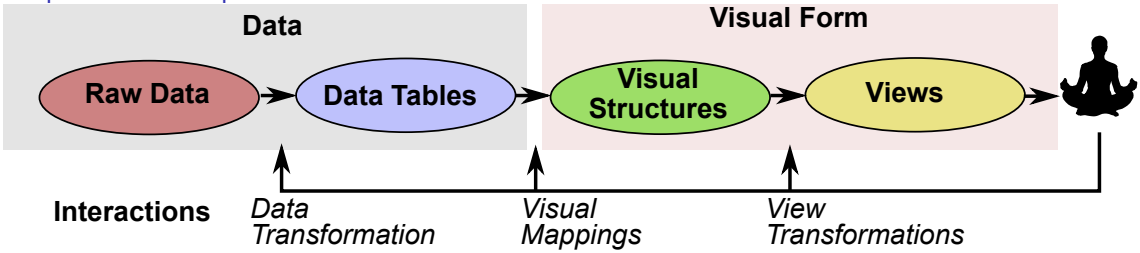


Source: Jastrow, J. (1899). The mind's eye. Popular Science Monthly, 54

Design of (Interactive) Graphics

- Designing a good visualization is non-trivial
- There exist many guidelines and languages to “program” graphics
- Considerations: limitations of the visual system and cognitive biases
  - Limits of memory
  - Too much information
  - Not enough meaning
  - Need to act fast

Graphics creation process



Components of Visual Mappings / Encodings [43]

- Spatial substrate: mapping variables to space (and axes)
  - Depends on the type of data: structured, unstructured
  - Values: nominal, ordinal, quantitative
- Marks: visible elements: points (0D), lines, areas, volumes (3D)
- Connection: uses points and lines to show relationships
- Enclosure: boxes around elements; useful to encode relationships
- Retinal properties:
  - Spatial: Size, orientation
  - Object: Gray scale, color, texture, shape
- Temporal encoding: Animations

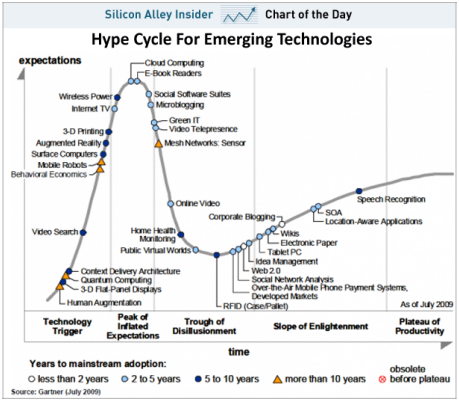
Guidelines

Goals of **graphical displays** according to [42]

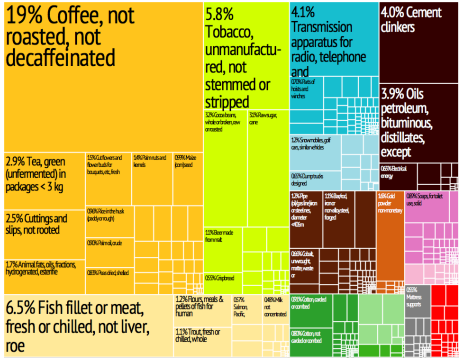
- show the data
- induce the viewer to **think about the substance** rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set

Information Graphics (Infographics) [41]

Definition: Graphic visual representations of information, data or knowledge intended to present information **quickly and clearly**



Source: Gartner Hype Cycle for Emerging Technologies. Jeff McNeil [41]



Source: Uganda Export Treemap from MIT Harvard Economic Complexity Observatory. R. Haussmann, Cesar Hidalgo, et.al. [41]

Guidelines

Simple rules

- Use the right visualization for the for data types
- Use building blocks for graphics (known plot styles)
- Reduce information to the essential part to be communicated
- Consistent use of building blocks and themes (retinal properties)

Promising concepts in expressing graphics

- ggplot2 (for R)
  - Follows the “Grammar of graphics”
  - Aesthetics define data used for the plot
  - Geometry are visual elements organizing the data
  - Faceting generates multiple subplots based on properties
- Vega https://vega.github.io/vega/
  - Declarative language for interactive graphics
  - Specified in JSON format; suitable for browser visualization

Interactive Data Visualization

Typical interactions with a view [50]

- **Brushing**: selecting elements individually/with a lasso
- **Painting**: create a group from selected elements
  - Allows to perform subsequent operations with the group
- **Identification**: cursor/mouse provides details about marked element(s)/groups
- **Scaling**: navigate plots, rescale, zoom, drill-up/down aggregated data
- **Linking**: interactions are performed on all connected plots
  - An element/group marked in one plot is highlighted on other plots
  - Scaling operations affect connected plots

Summary

- Visual perception is efficient for communication of information
- Understanding limitations of cognition (the visual system) is important
- Visual analytics follows the scientific method
  - **Interactive** data exploration, modeling & **experimentation**
  - Extends **exploratory data analytics**
- Graphics design follows principles

Bibliography

31 https://en.wikipedia.org/wiki/Scientific_method

32 https://en.wikipedia.org/wiki/Visual_Analytics

33 James Thomas, Kristin Cook. 2005. Illuminating the Path: The R&D Agenda for Visual Analytics National Visualization and Analytics Center

34 Keim D. A, Mansmann F, Schneidewind J, Thomas J, Ziegler H. 2008. Visual analytics: Scope and challenges. Visual Data Mining

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

36 Martins N., Erlhagen W., Freitas R. 2011. Non-destructive Whole-brain Monitoring using Nanorobots

37 http://www.top500.org (Nov. 2016)

38 https://en.wikipedia.org/wiki/Optical_illusion

39 https://en.wikipedia.org/wiki/Visual_system

40 https://en.wikipedia.org/wiki/List_of_cognitive_biases

41 [https://en.wikipedia.org/wiki/Infographic]

42 Edward Tufte. 1983. The Visual Display of Quantitative Information.

43 Scott Card. 2009. Information visualization. In A. Sears & J. A. Jacko (Eds.), Human-Computer Interaction: Design Issues, Solutions, and Applications

44 https://en.wikipedia.org/wiki/Statistical_graphics

45 https://en.wikipedia.org/wiki/Exploratory_data_analysis

46 https://en.oxforddictionaries.com/definition/cognition

47 https://de.wikipedia.org/wiki/Visual_Analytics

48 D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, H. Ziegler. 2008. Visual analytics: Scope and challenges. Visual Data Mining

49 Comparison of Open Source Visual Analytics Toolkits, http://www.sandia.gov/~pjcross/papers/Part1.pdf

50 https://en.wikipedia.org/wiki/Interactive_data_visualization