Interactive Data Visualization

Basic Introduction to Visualization and Tools

Benjamin WEYERS
University of Trier

Forthnightly Video Conference on Visualization

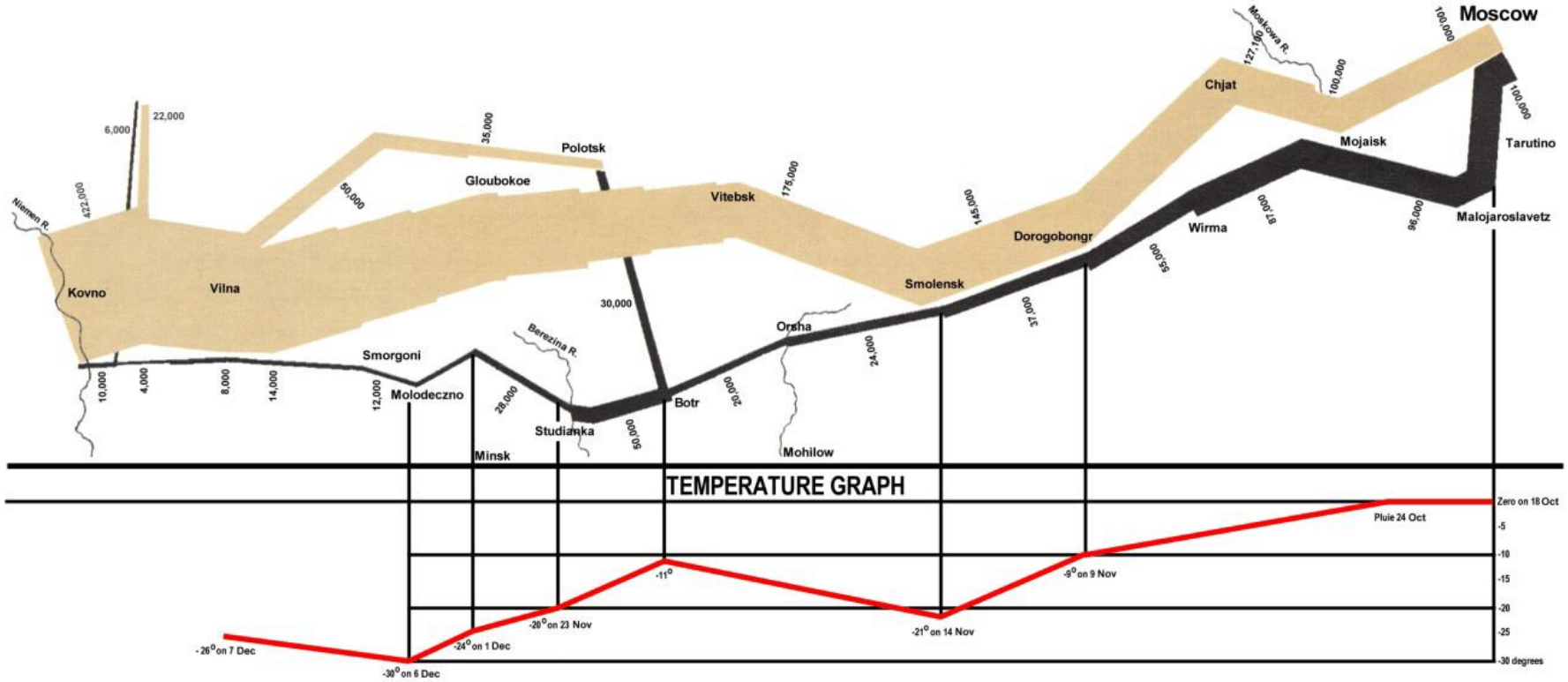
- Send your questions, topics, or requests to weyers@uni-trier.de
- They will move into the agenda and discussed or answered directly
- Search the Collaboratory for the „Visualization Video Conference“ collab
- Takes place each 1st and 3rd Thursday
- 9-10am
- Is announced over Twitter!

Literature and References

- Robert Spence, (2014), *Information Visualization: An Introduction*, Springer.
- Edward R. Tufte, (1991), *Envisioning Information,* Graphics Press.
- Edward R. Tufte, (2001), *The Visual Display of Quantitative Information,* Graphics Press.
- Scott Murray, (2013), *Interactive Data Visualization for the Web*, O'Reilly

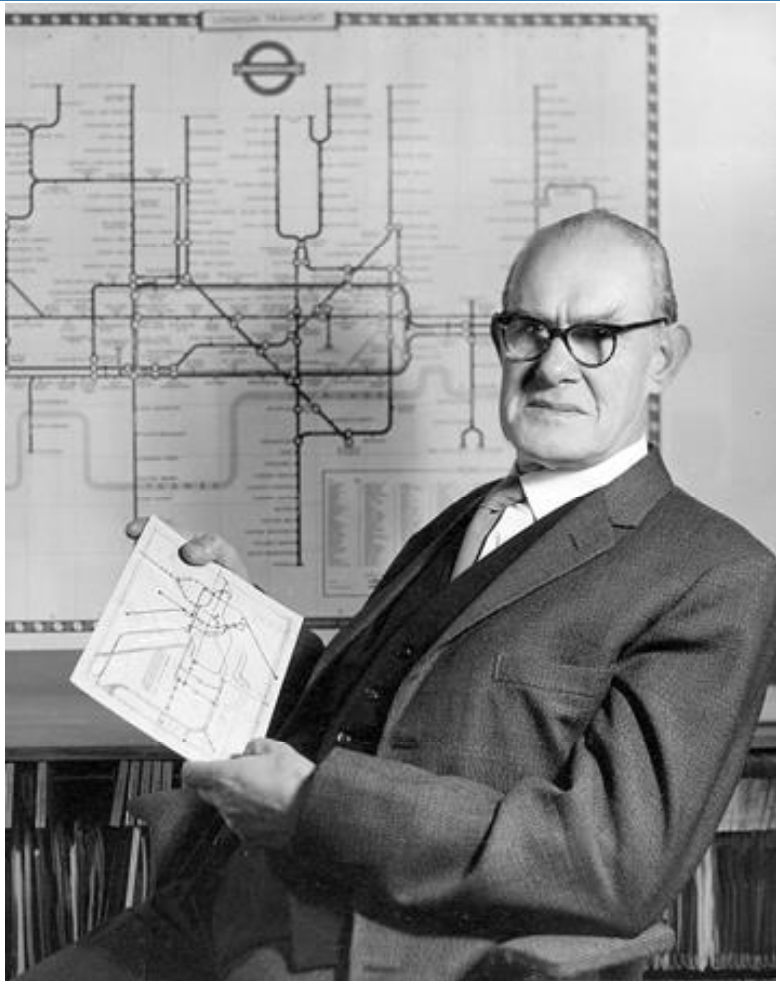
All material not equipped with additional references (URL) on the slides is taken from the above books.

Introduction

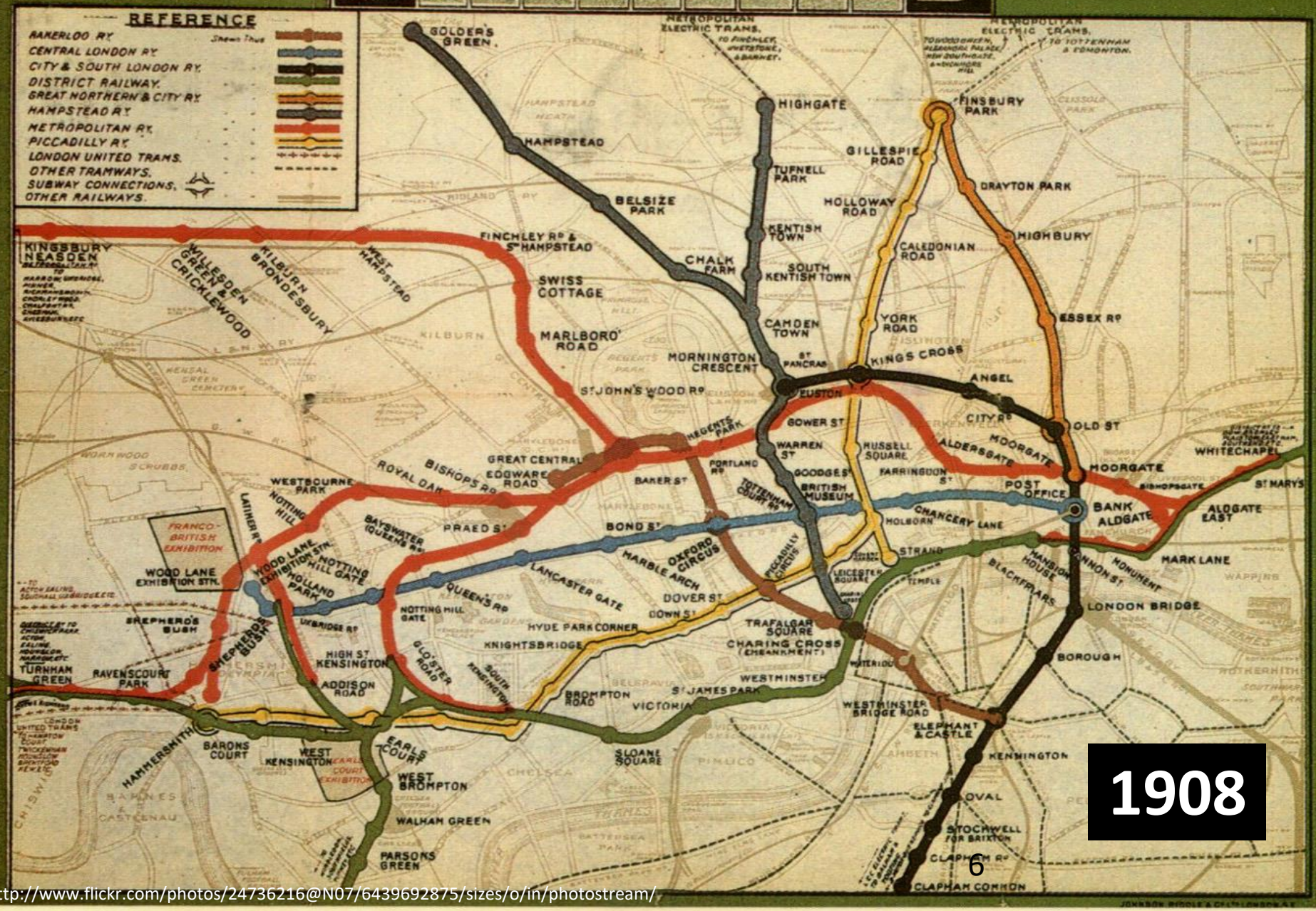


Graph drawing by Charles Joseph Minard showing the decreasing size of the Grande Armée as it marches to Moscow and back with the size of the army equal to the width of the line.

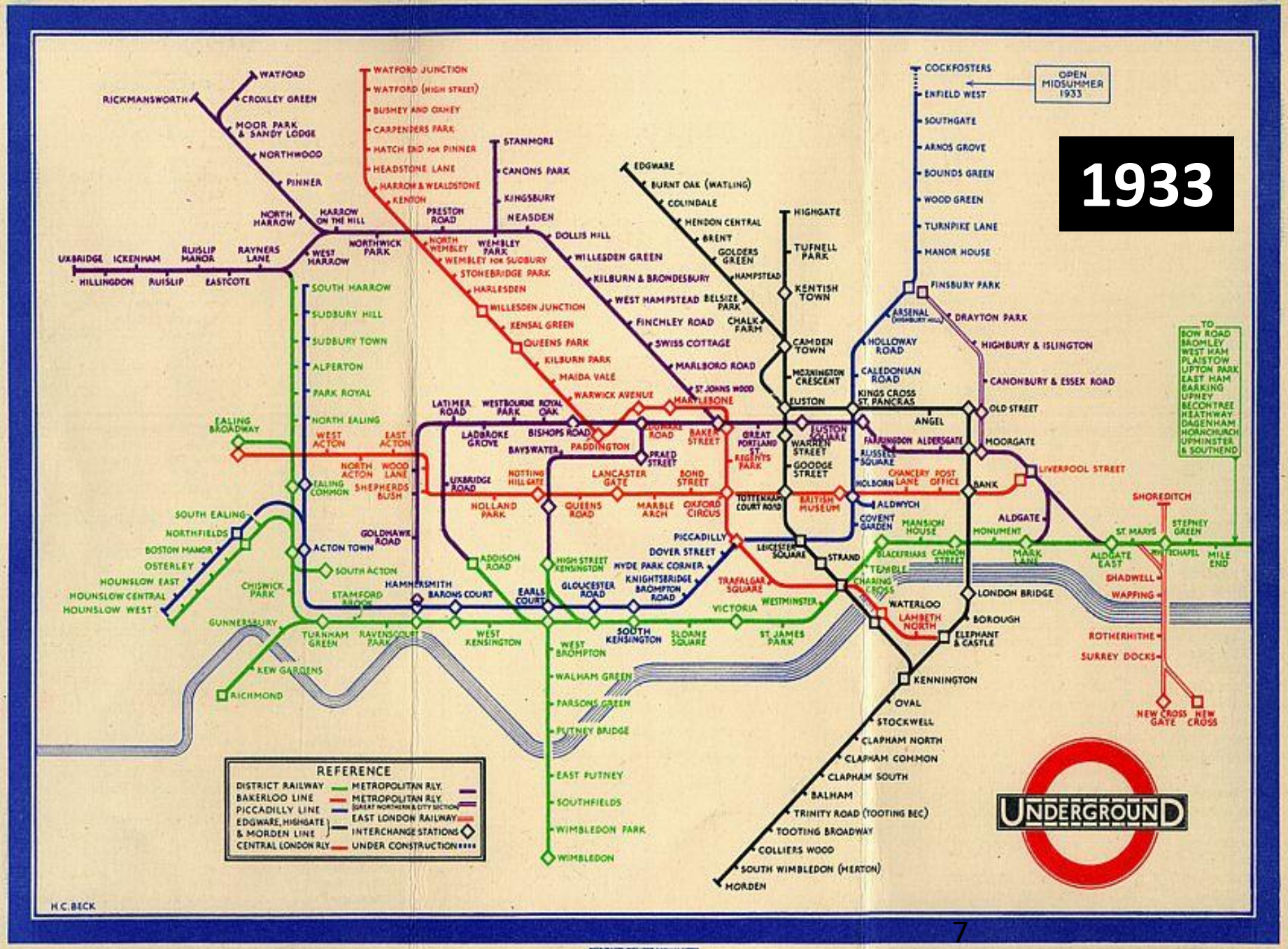
Introduction



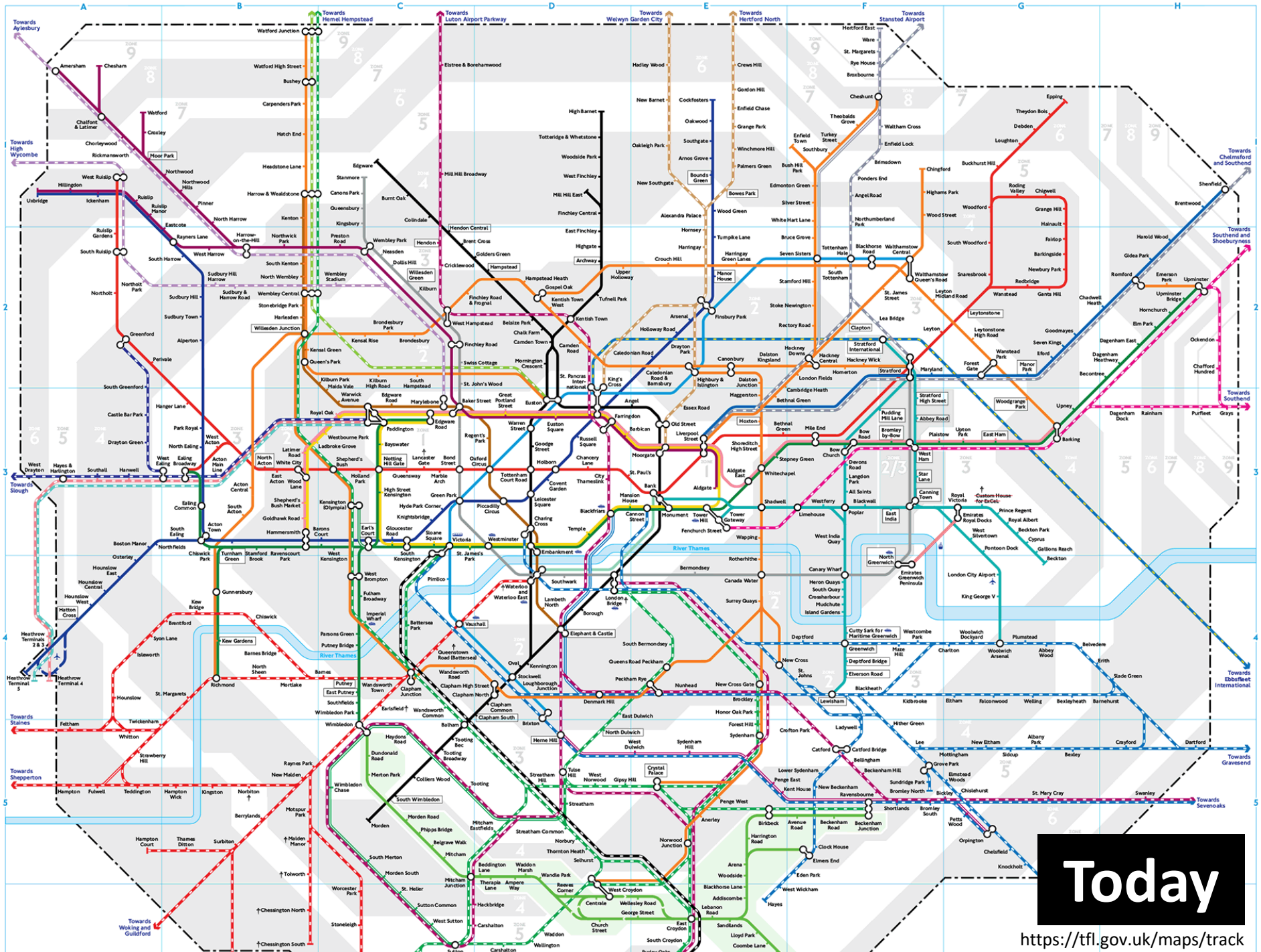
Harry Beck: „When you are underground it does not matter where your are“



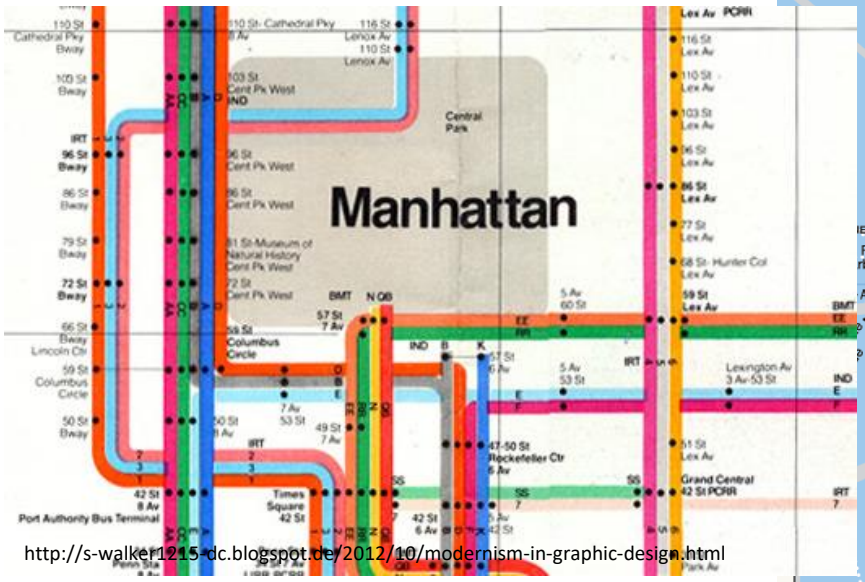
http://www.flickr.com/photos/24736216@N07/6439692875/sizes/o/in/photostream/



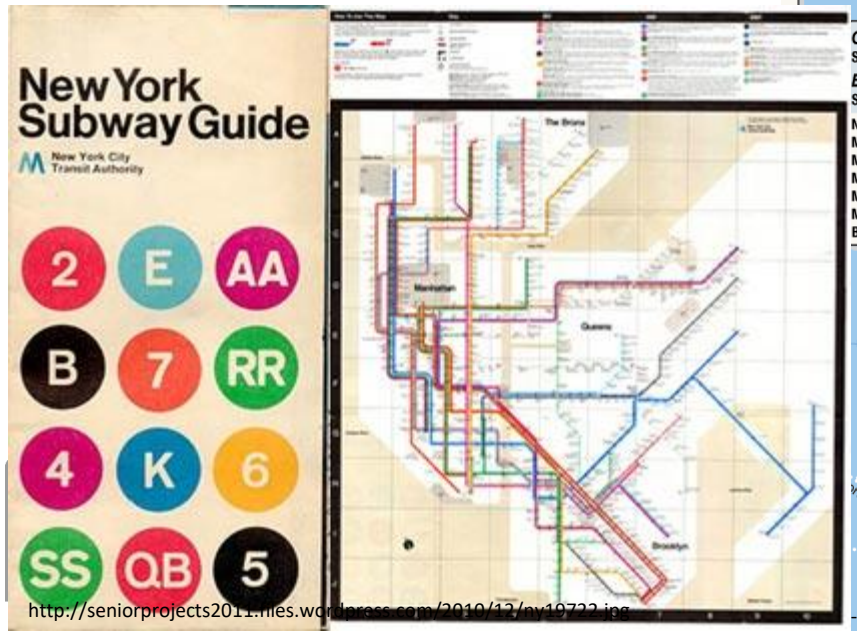
http://sovibrantopinion8.blogspot.de/2011/04/design-classic-no145-london-underground.html


Towards Aylesbury
Towards Hemel Hempstead
Towards Luton Airport Parkway
Towards Welwyn Garden City
Towards Hertford North
Towards Stansted Airport
Towards High Wycombe
Towards Chelmsford and Southend
Towards Southend and Shoeburyness
Towards Southend
Towards Slough
Towards Ebbsfleet International
Towards Staines
Towards Shepperton
Towards Gravesend
Towards Sevenoaks
Towards Woking and Guildford
River Thames
Today
https://tfl.gov.uk/maps/track

Museum of Modern Art, New York, USA – Massimo Vignelli



http://s-walker1215-dc.blogspot.de/2012/10/modernism-in-graphic-design.html



http://seniorprojects2011.miles.wordpress.com/2010/12/ny19722.jpg

-York-City-Subway-Map-2.gif

Information Visualization

- **“to visualize”** [Oxford, 2010]
  - form a mental image of; imagine
  - make (something) visible to the eye
- not only produce pretty images, but **aid the understanding of data**
- visualization is **interdisciplinary by definition**
- the **difference** between **InfoVis and SciVis** is the type of data being visualized: Abstract Data vs. Measured Spatial Data
  - Nevertheless, the differentiation is not 100% clear…
  - Very different definitions out there
- **in general:** Information visualization‘s objective (as well as scientific visualization) is to represent data in a way that the human is able to gain insight and get enabled to understand its structure *easily*.

Information Visualization Guidelines

- A good visualization considers

  1. The user (cognitive skills, pre-knowledge – expert vs. novice, disabilities, …)
  2. The task (goals, operations, information, context, …)
  3. The data

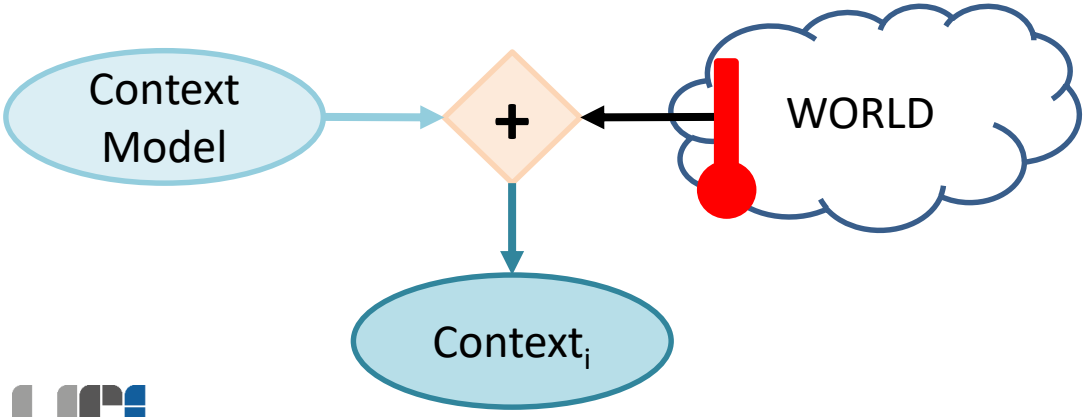
- The creation process of visualization applications and data tools need to involve an **incremental development a**
  - identify the **user group** and investigate the/their **pr**
  - identify the task and to investigate the **task's s** d derive the **relevant analysis steps** from it
  - create a **prototype** and to **evaluate** it in a
  - repeat

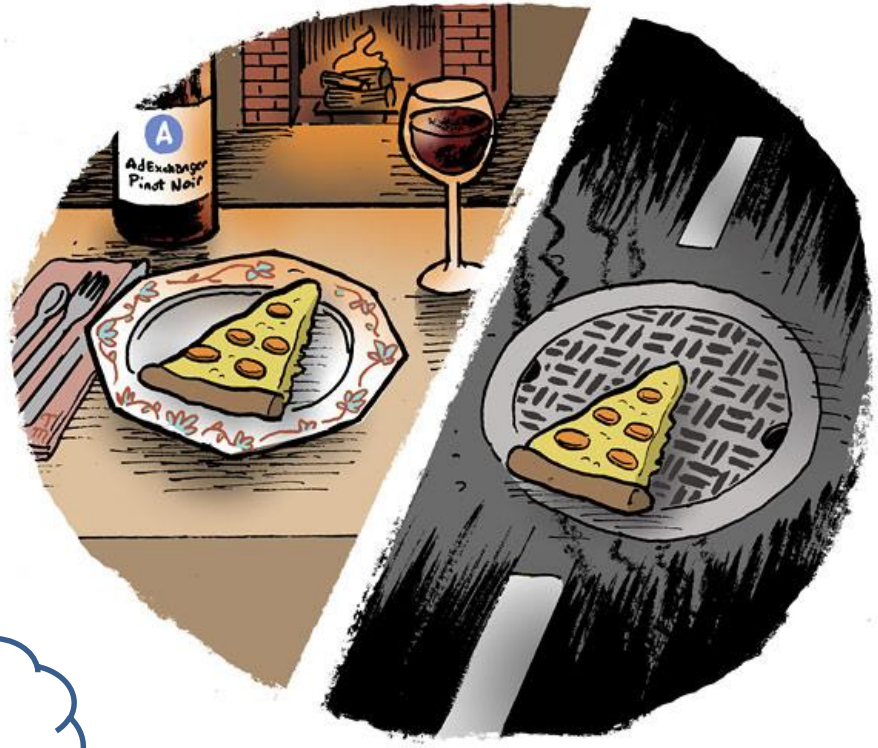
User Centered Design

The Context

- What you need:
  - Understanding **what context is**
  - A description of context that is **understandable by the computer**
  - A system that brings the **description and data together, which characterises the context**

→ Makes context understandable by the computer-based system





Context Matters

http://adexchanger.com/comic-strip/adexchanger-context-matters/

The Data

- Before starting to discuss the visual representation of data and values, the different types of data should be classified first

- The goal of this classification is to describe concepts in InfoVis not like

„Color encoding is well suited for the representation of the development in the stock market“

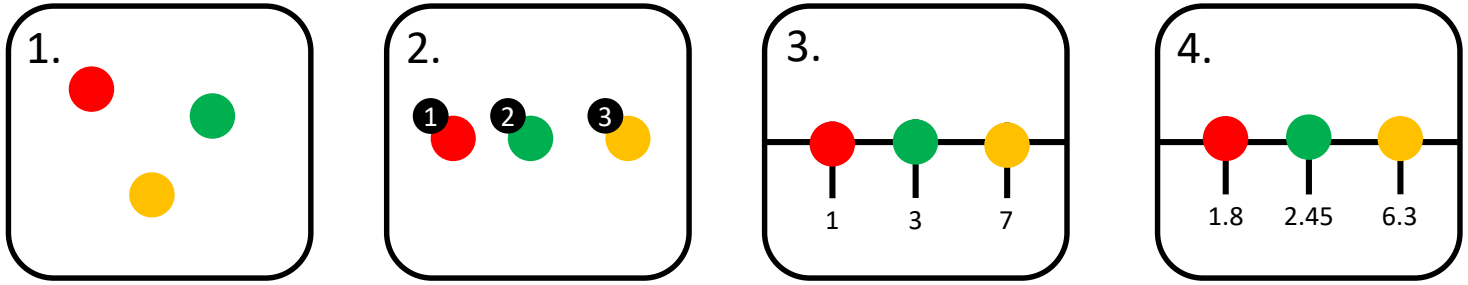
but

„Color encoding is well suited for the representation of categories“

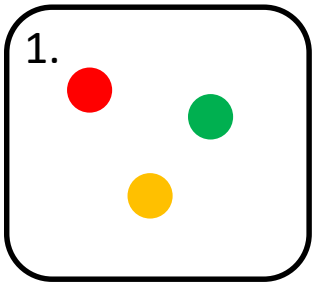
Visual Representation of Data and Values – Data Classification

- Stevens (1946) proposed a qualitative classification of data attributes along numeric scales:

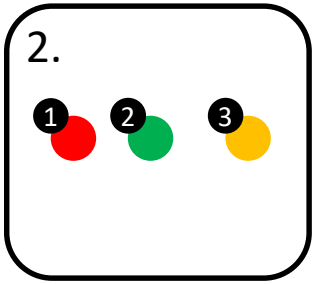
  1. **Nominal:** Elements that could not be logically ordered, e.g., apple, orange, grape, banana
  2. **Ordinal:** Elements that could be logically ordered but where the ordering has no distance metric, e.g., weather situations along a scale of favor
  3. **Interval:** Ordinal attributes that can be equipped with a (discrete) metric, e.g., arrival time of planes
  4. **Ratio:** Extended Interval scale to the whole range of real numbers, e.g., the mass of solid physical objects



Visual Representation of Data and Values – Data Classification

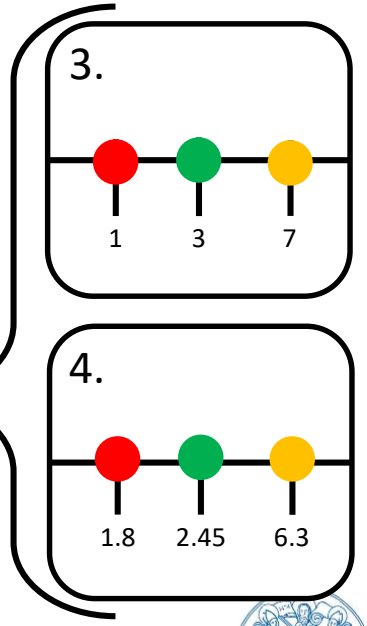


Nominal ⇔ Category Data ⇔ **`char`**



Ordinal ⇔ Discrete Data ⇔ **`enum`**

Ratio + Interval ⇔ Continous Data ⇔ **`int, float`**



Visual Representation of Data and Values – Data Classification

- The attribution of entities can be of higher dimension
- In general, an entity can be represented as field of attributes of different dimensions -> **Data Objects**

- **Operations** applied to data objects or more general than entities, attributes, or relations, which can not be defined as operations
  - Examples of operations are:
    - Mathematical operations
    - Merge of two lists
    - Invert values
    - Instantiation of entities or relations
    - …

Visual Representation of Data and Values

- Central **requirements of the visualization** of data and values are:
  - Present **more than one value** at a time
  - Relevant **dependencies and correlation** should be visible at one glance
  - Should be **intuitive and simple** to understand
  - Should **match the basic visual and perceptual characteristics** of the human visual cognition

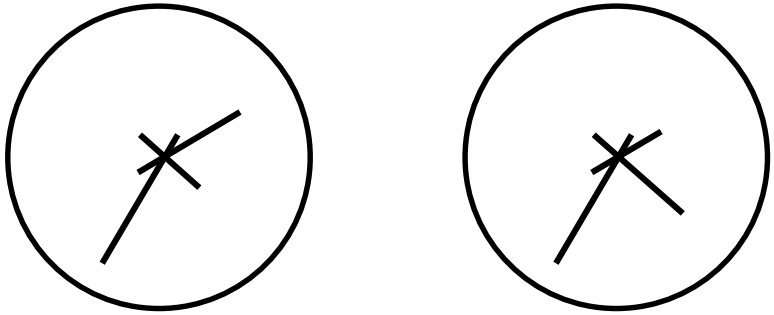
- In the following, the visual representation of **single values** will be followed by the representation of **multiple values** up to **multi-dimensional data representations**

Single Value Representation (0-dimensional)

Single Value Representation

- A simple example for the representation of a single value is the altimeter used in airplanes.

- This instrument is responsible for various accidents
- By changes in attention and focusing to other contexts, changes in the altimeter can overseen very easily



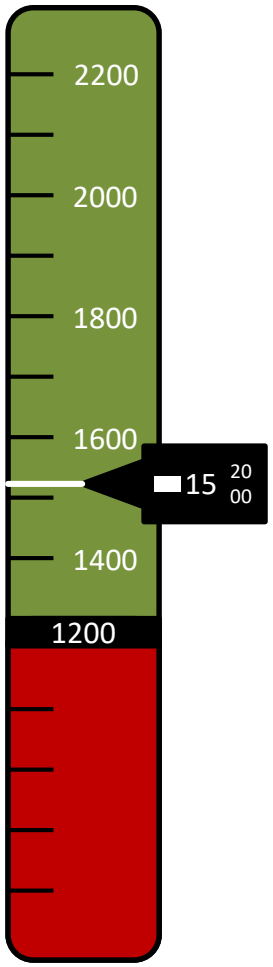
http://www.m0a.com/altimeter/

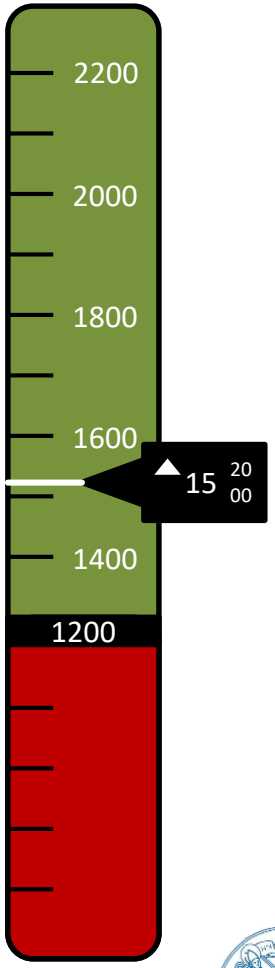
Video

Change Blindness

- This effect is known as **change blindness**. Human perception is unaware of small changes in complex environments or in more or less complex representations.

Single Value Representation





1-Dimensional Value Representation

Representation of Value Sets – One Dimension

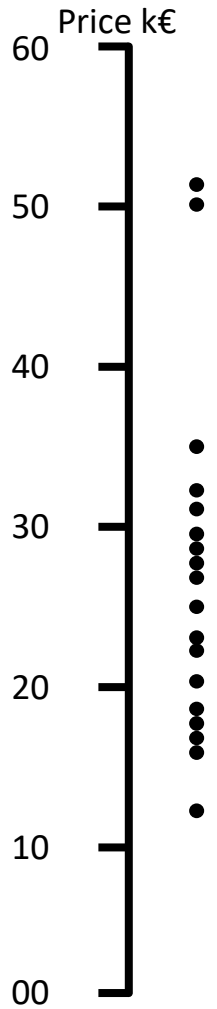
- **Example:** Visualize the prize of various cars, which are related to a linear scale

- **Question:** What is most effective? What should be identified in the data?

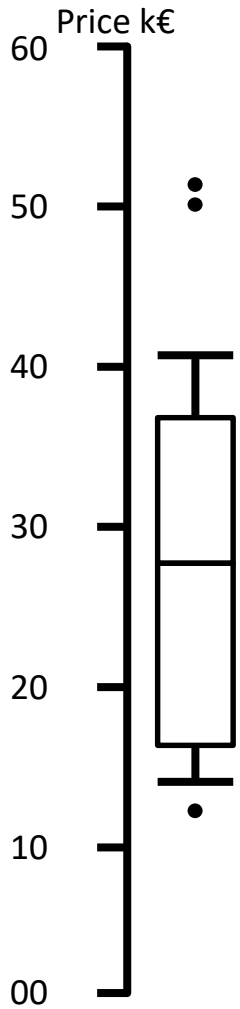
- **Focusing on**
  - Mean values
  - Distribution
  - Min and Max (price/s)
  - …

- Possible Visualization Methods: **Dot Plot, Box Plot, Histogram**

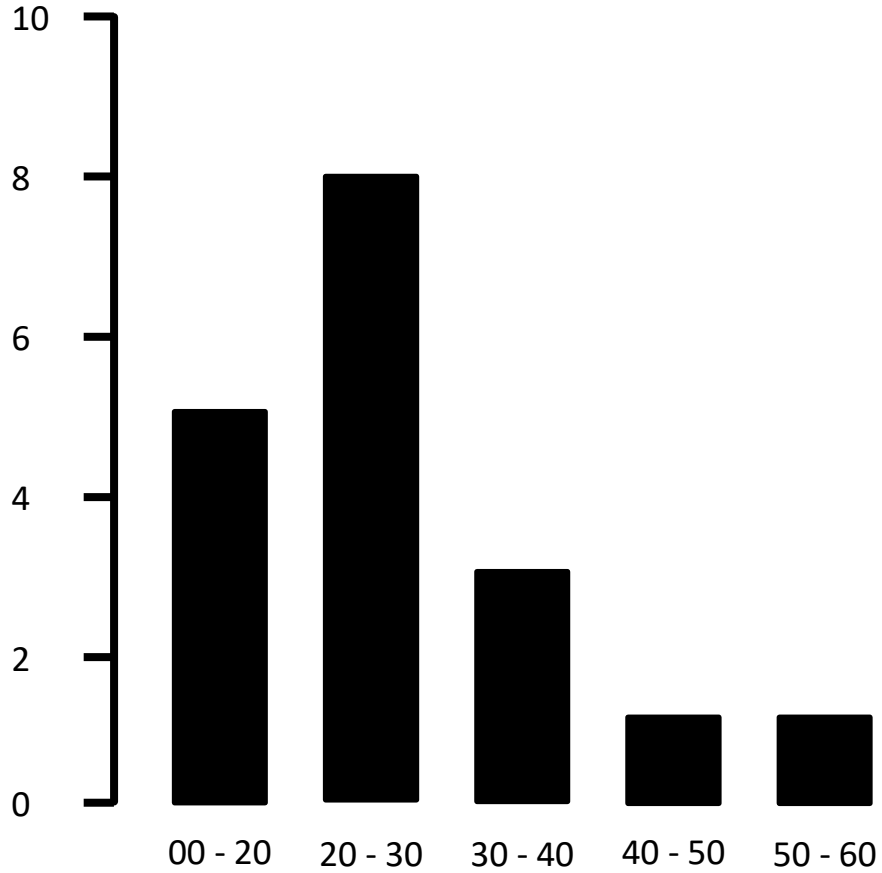
Representation of Value Sets – One Dimension



Dot Plot

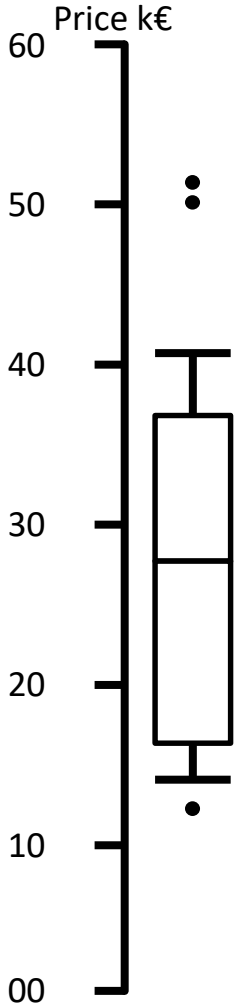


Box Plot



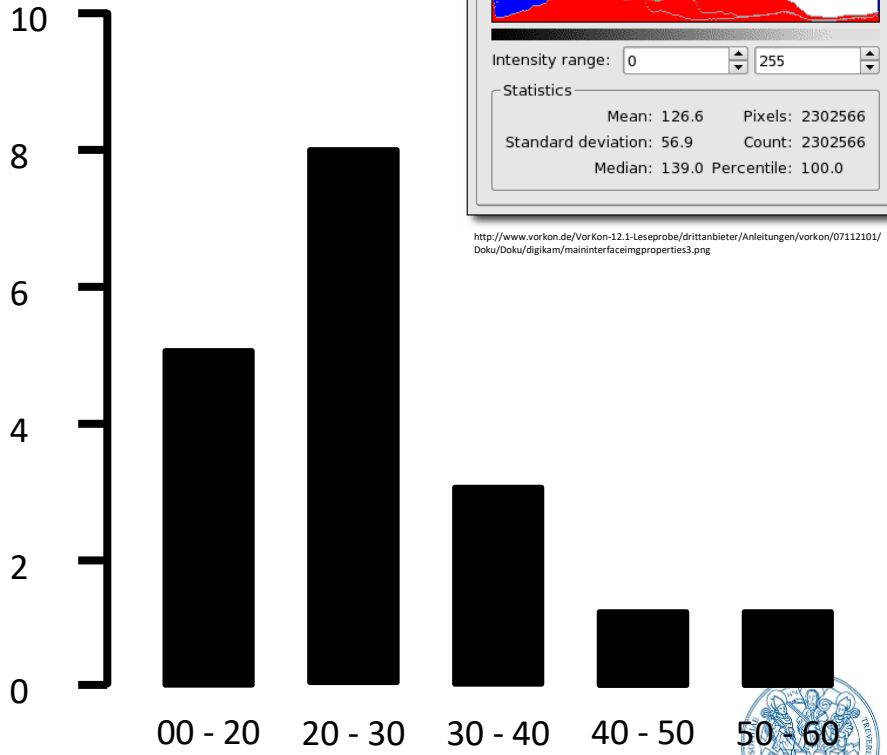
Histogram

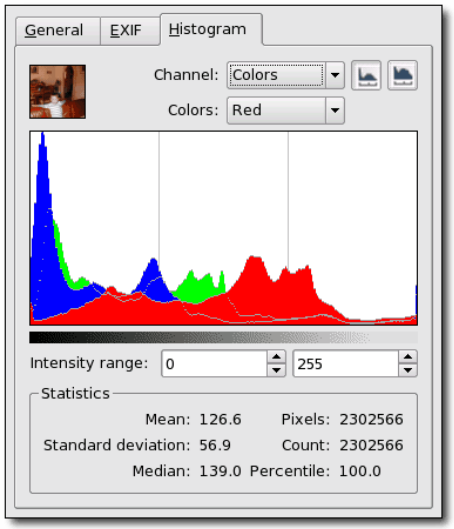
Box Plot and Histogram



- Box Plot:
  - Central Line specifies the Median
  - End of Boxes specify 25 and 75 percentile
  - End of Lines specify 5 and 95 percentile
  - Dots show outliers

- Histogram:
  - Represent frequency/occurrence of values in a value set or of specific characteristics





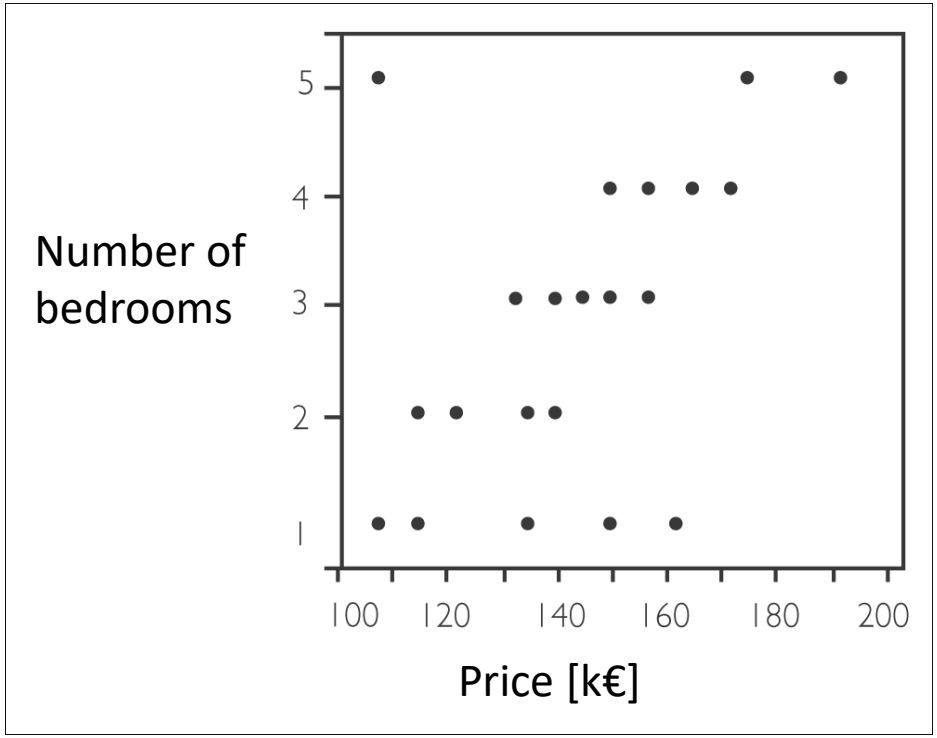
http://www.vorkon.de/VorKon-12.1-Leseprobe/drittanbieter/Anleitungen/vorkon/07112101/Doku/Doku/digikam/maininterfaceimgproperties3.png

2-Dimensional Value Representation

Presentation of Values Sets – Two Dimension

- Example: 2 dimensions (numberOfBedrooms [], price [k€])

  (1, 108), (1, 115), (1, 135), (1,150), …, (4, 150), …, (5, 195)

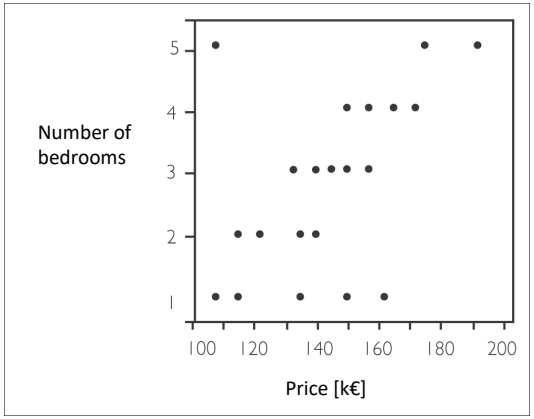


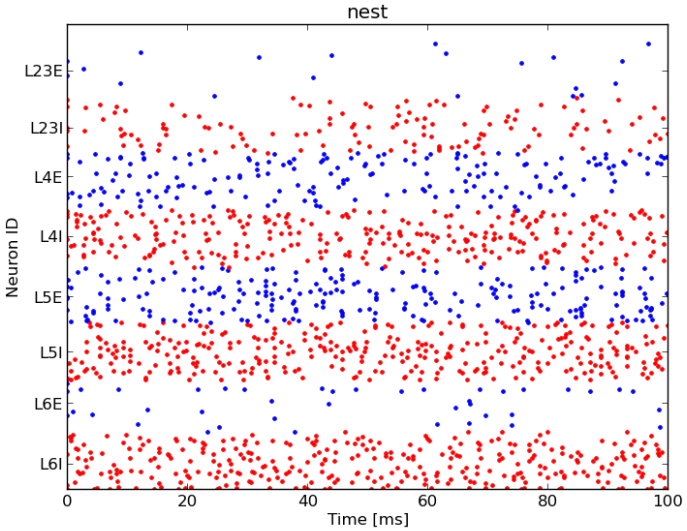
Scatter Plot

- Basic **Scatter Plot** can visualize two-dimensional data

- Possible interpretations are
  - Identification of trends
  - Local trade-offs
  - Outliers

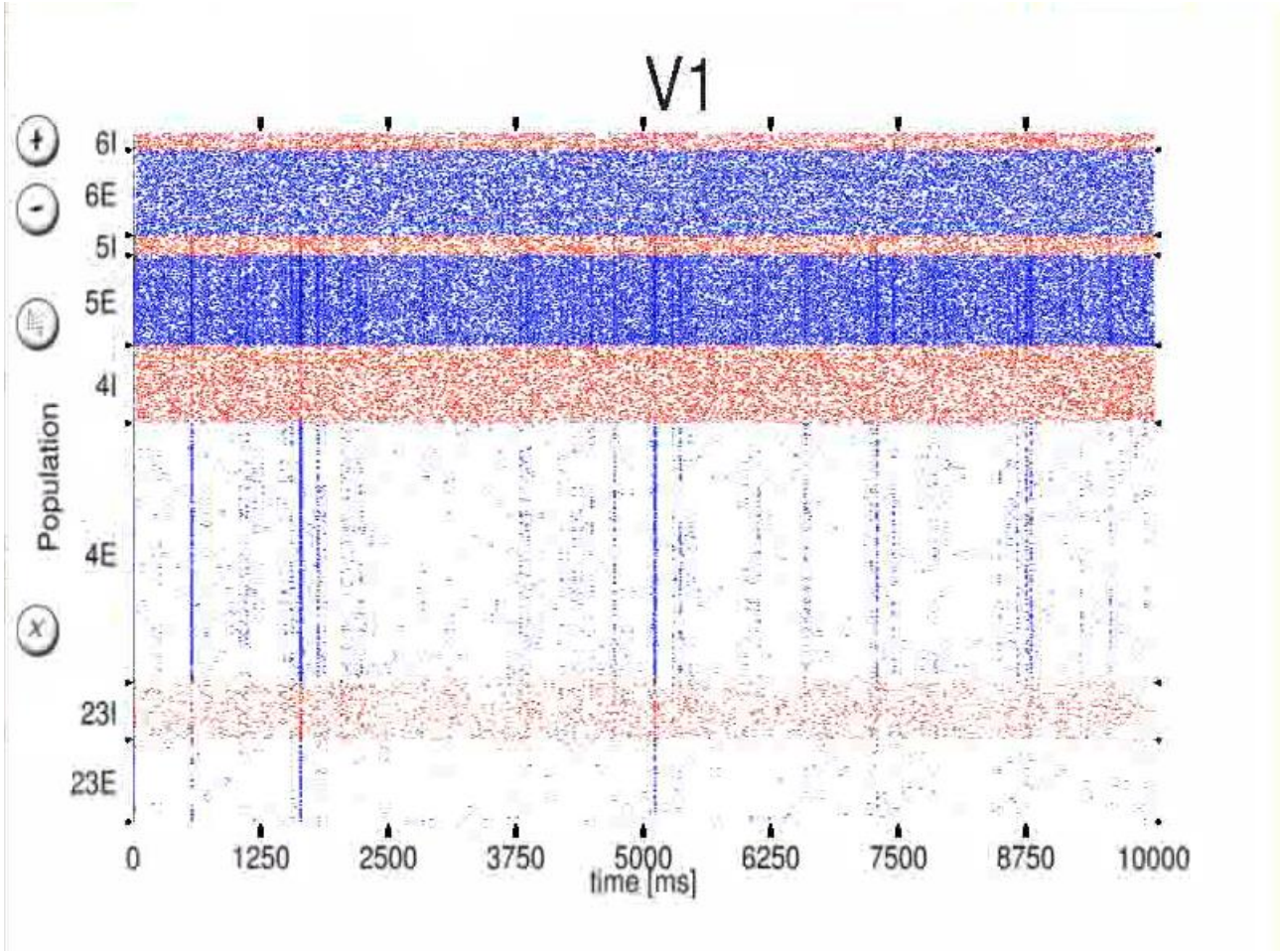
- A specific area for scatter plots is in time-dependent data, such as **Spiking Plots** in Neuroscience





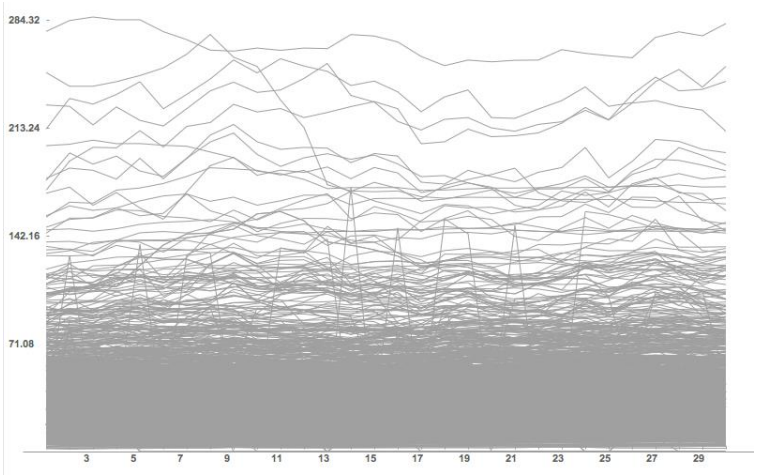
https://capocaccia.ethz.ch/capo/wiki/2013/spinnaker13

Spiking Plots



Presentation of Value Sets – Two Dimension

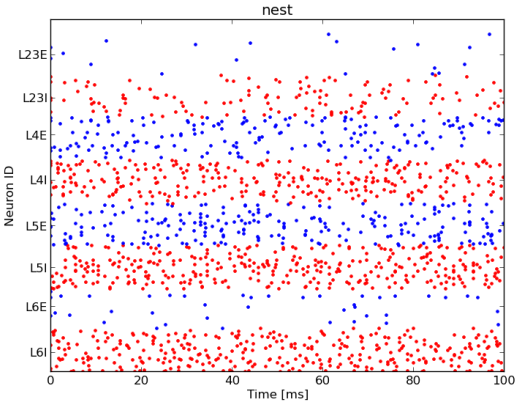
Function Plot



Bar Chart

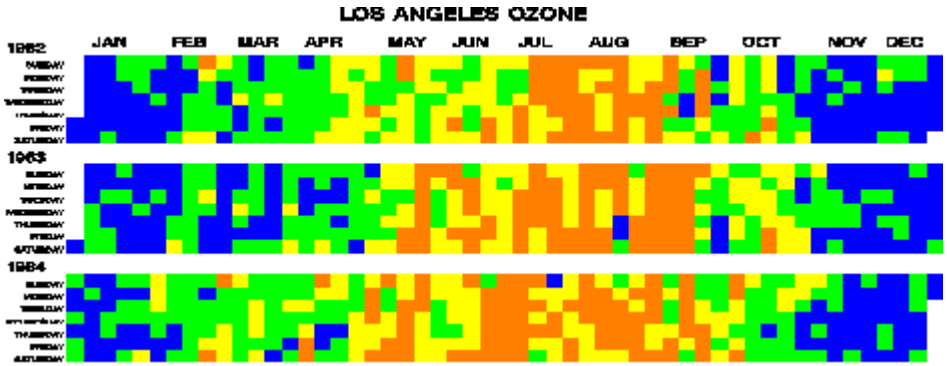
jcharts.sourcefourge.net

Scatter Plot



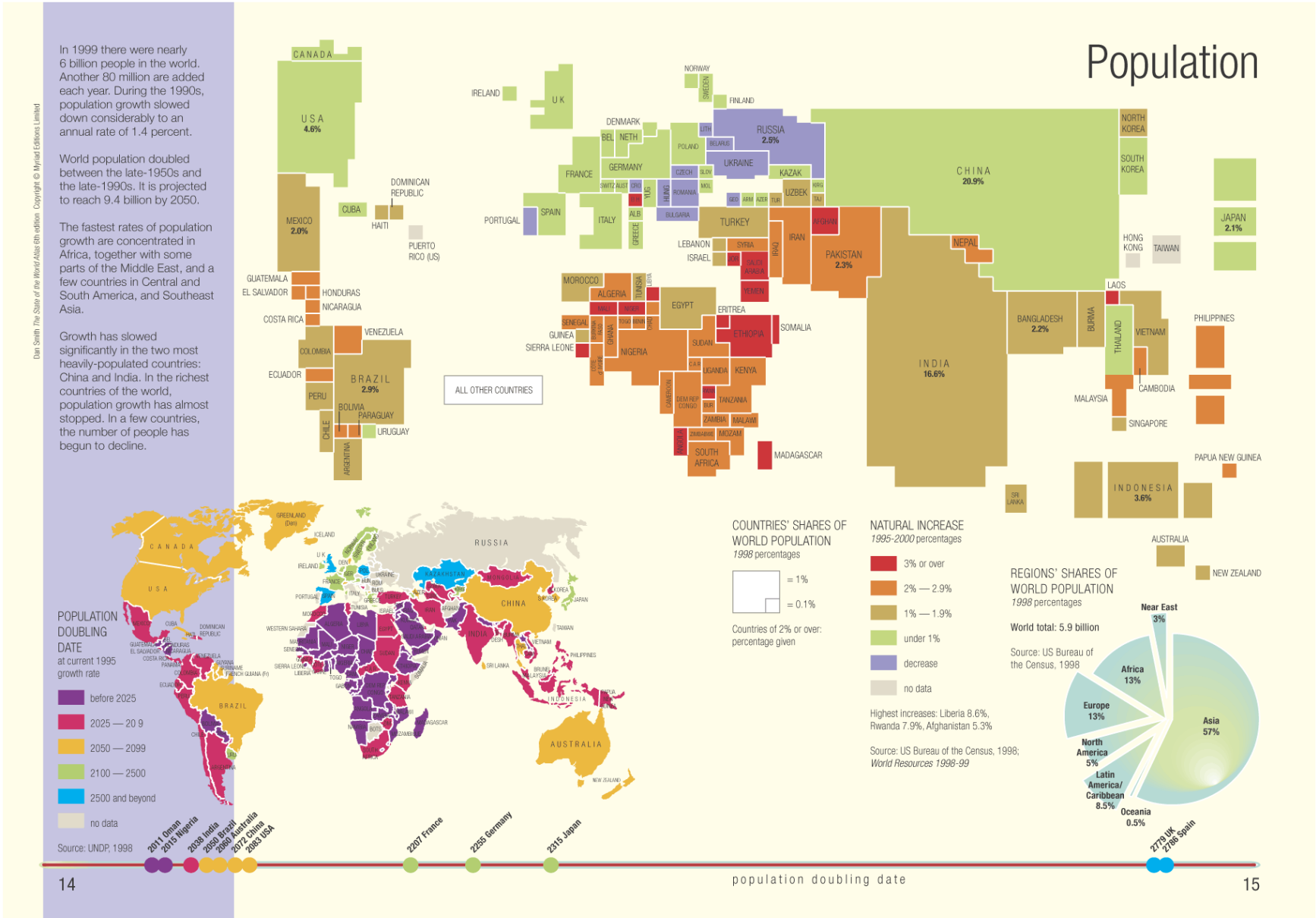
https://capocaccia.ethz.ch/capo/wiki/2013/spinnaker13

Heat Maps



http://www.infovis.info/

Presentation of Values Sets – Two Dimension

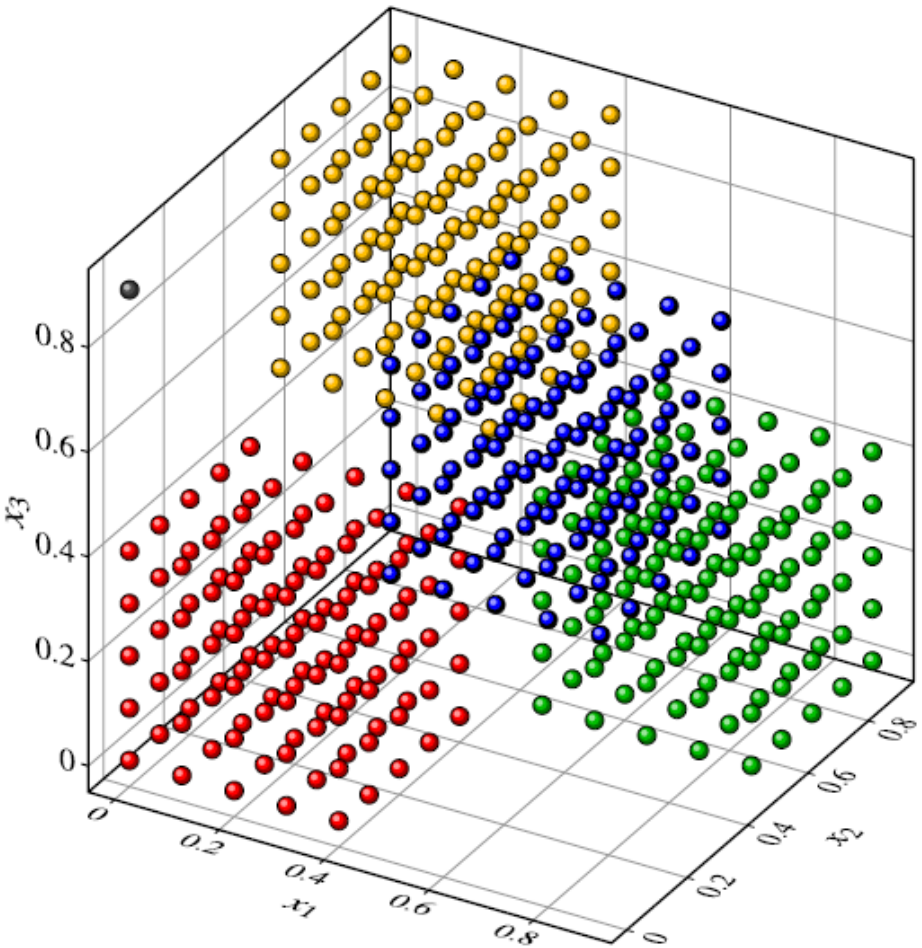


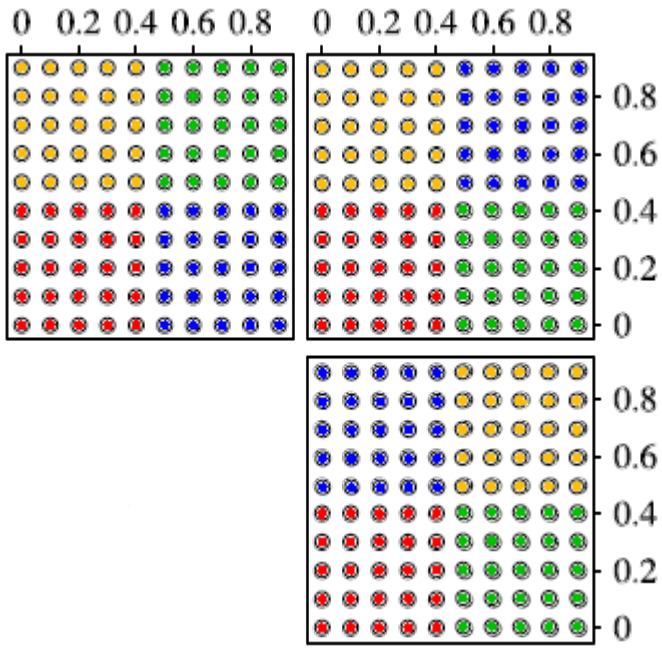
Presentation of Values Sets – Two Dimension

Presentation of Values Sets – Two Dimension

3-Dimensional Value Representation

3D Data





http://stats.stackexchange.com/questions/70569/interpreting-3d-scatter-plot

N-Dimensional Value Representation & Interaction Concepts for Visualization

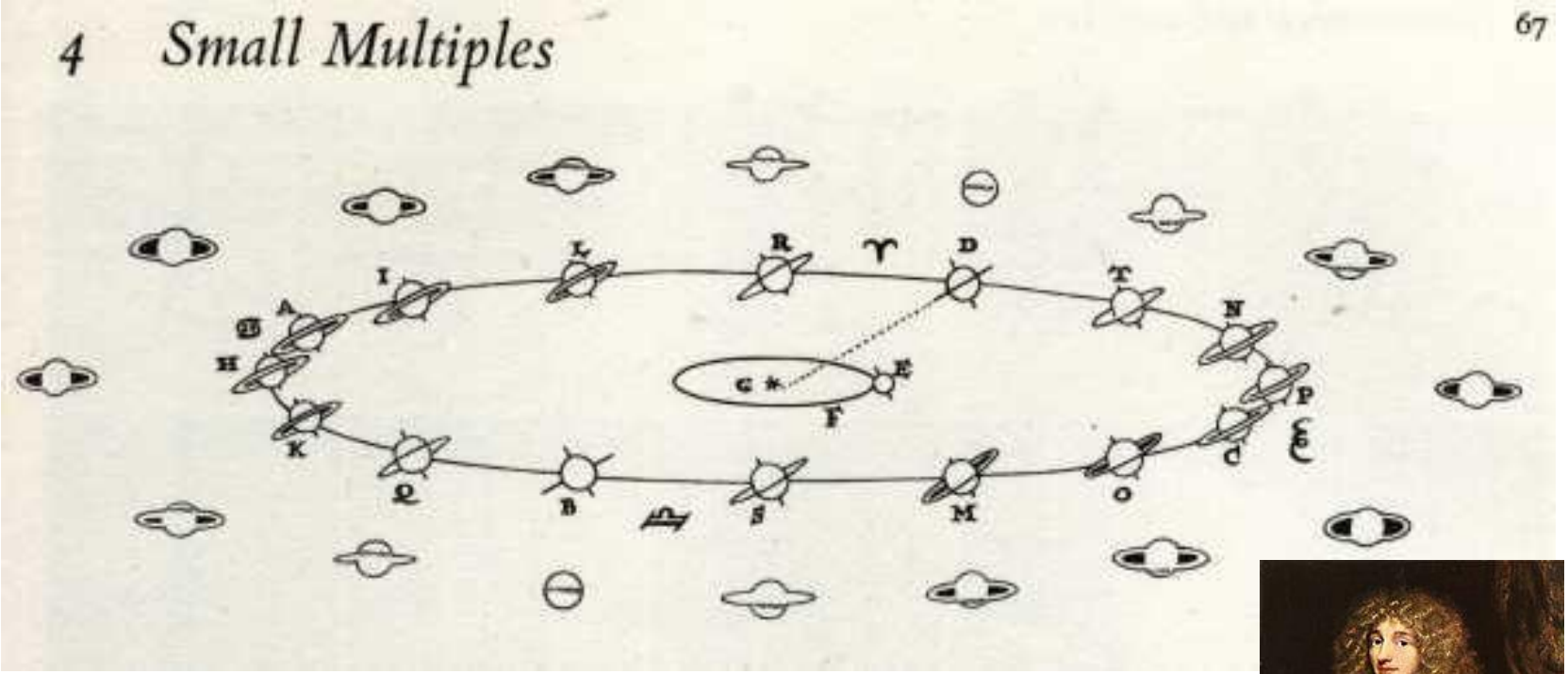
Presentation of Value Sets – N Dimensions

- A well known and nice concept is the so called „Small Multiples“ described by Tufte, 2013

- In general: A data set is presented in many small drawings, where one dimension is altered between all drawings, such as time, position, etc.

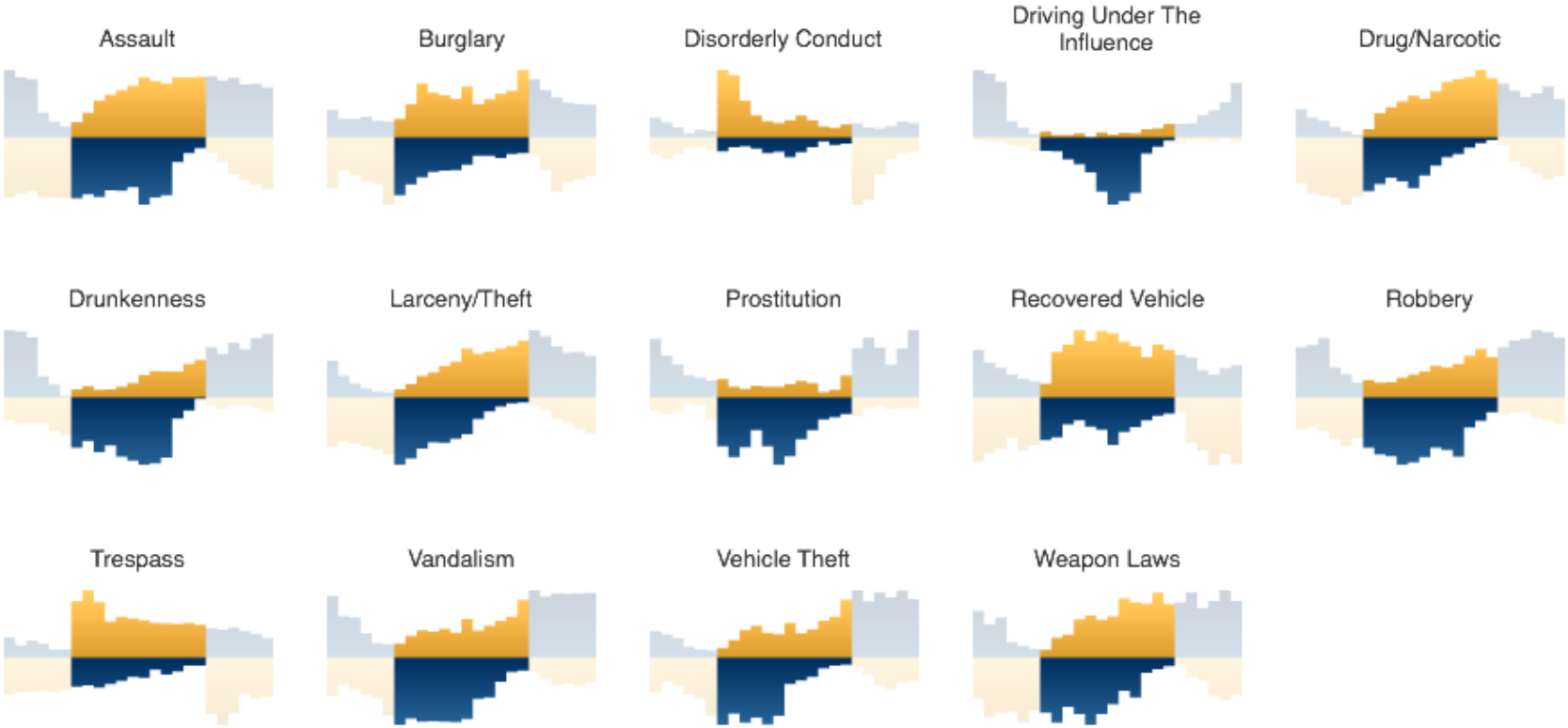
- Small Multiples are very good in making **changes visible** along the altered dimension
  - It enables the user to compare different views to the data with each other in a very simple and convenient way.

Small Multiples – Historic Example



- Christiaan Huygens, Systema Saturnium (The Hague, 1659), p. 55. Cited by Edward R. Tufte, (1991) p.67. *Envisioning Information* Graphics Press. Cheshire, Connecticut

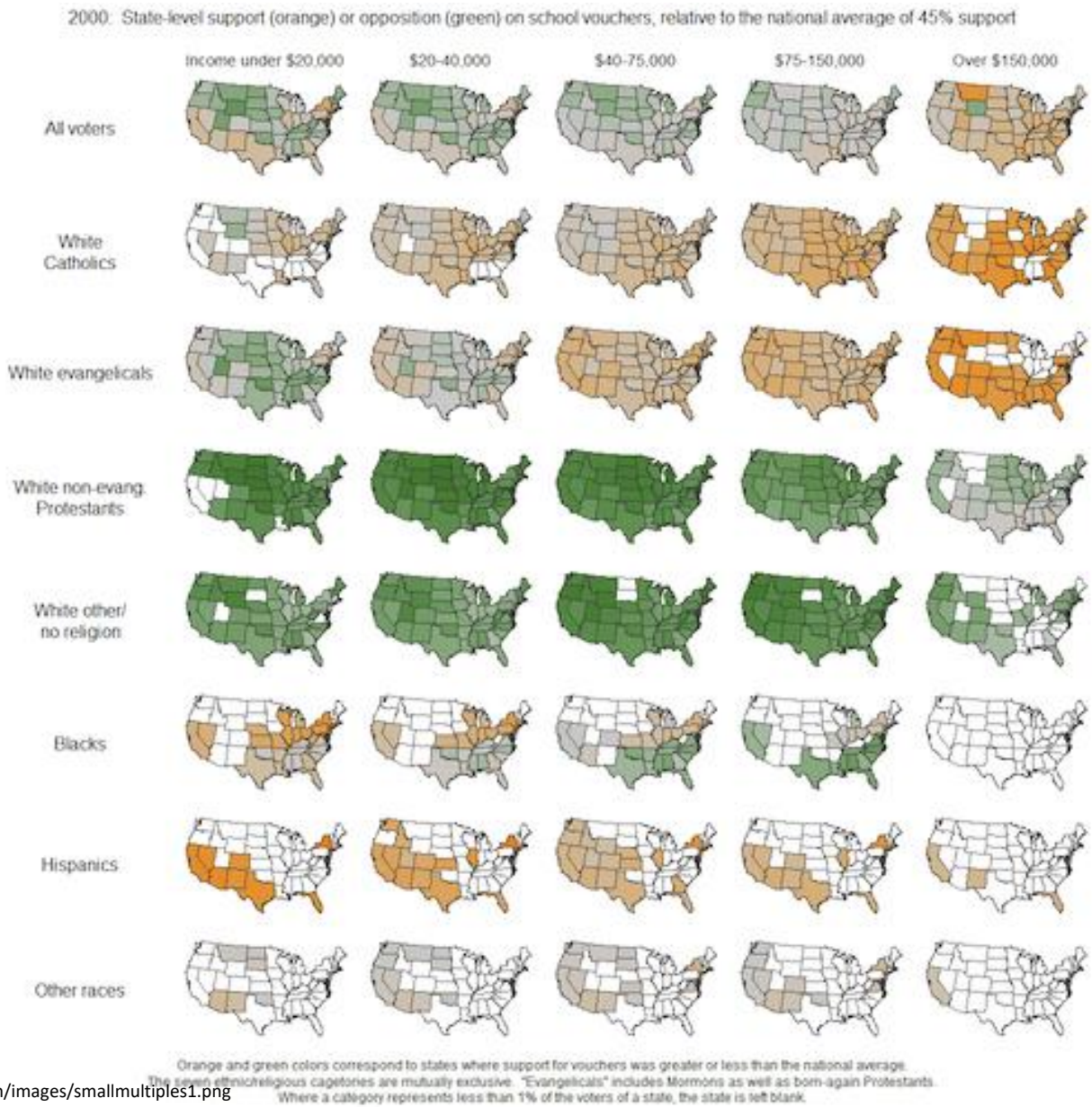
Small Multiples – Examples



http://dougmccune.com/blog/wp-content/uploads/2011/04/small_multiples_small.png

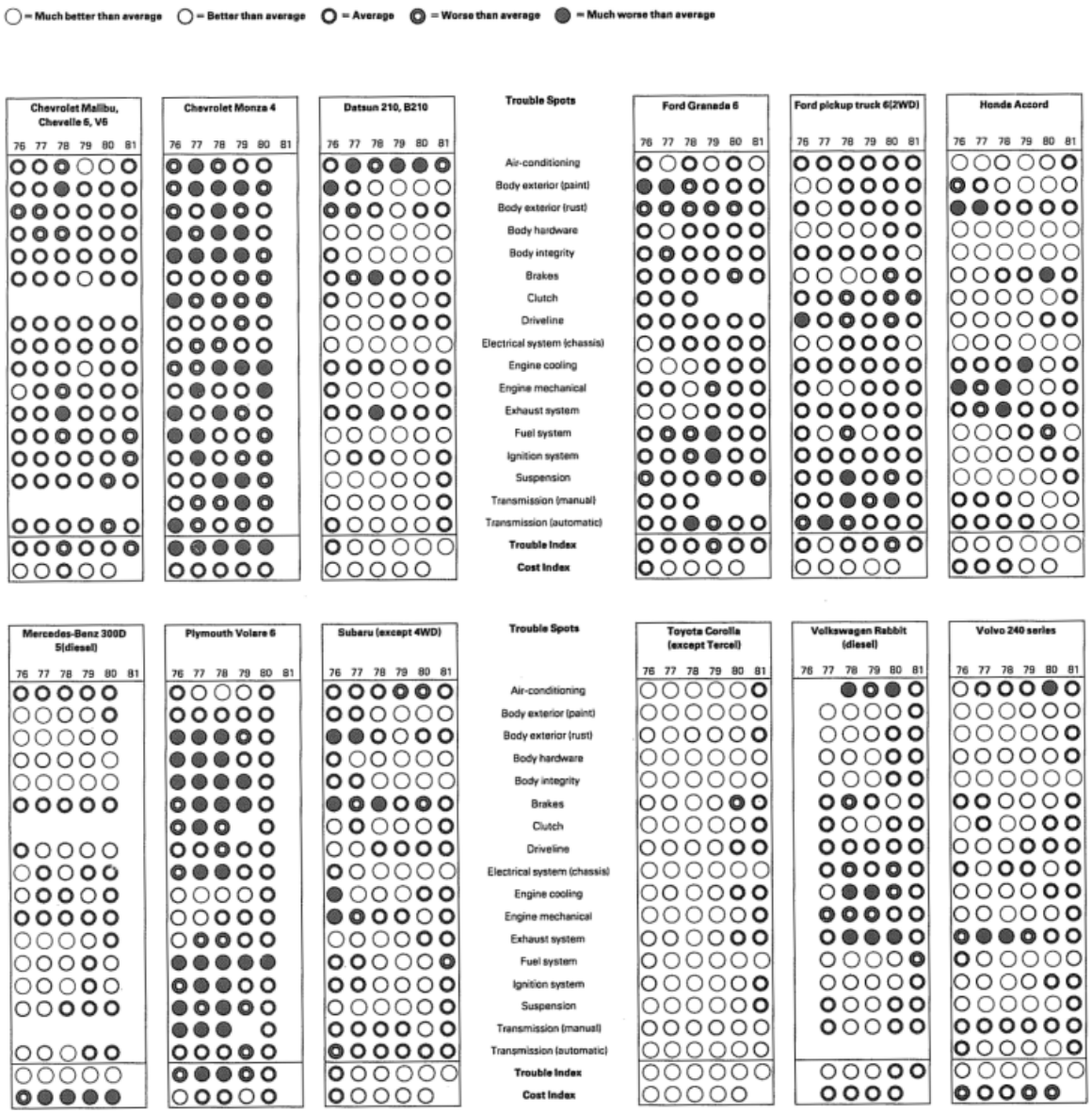
- Crime data from San Francisco http://apps.sfgov.org/datafiles/index.php?dir=Police&by=name&order=asc
- Double Bar Charts

Small Multiples – Examples



http://media.juiceanalytics.com/images/smallmultiples1.png

Small Multiples – Examples



http://www.infovis-wiki.net/index.php?title=Teaching:TUW_-_UE_InfoVis_WS_2008/09_-_Gruppe_06_-_Aufgabe_1_-_Small_Multiples

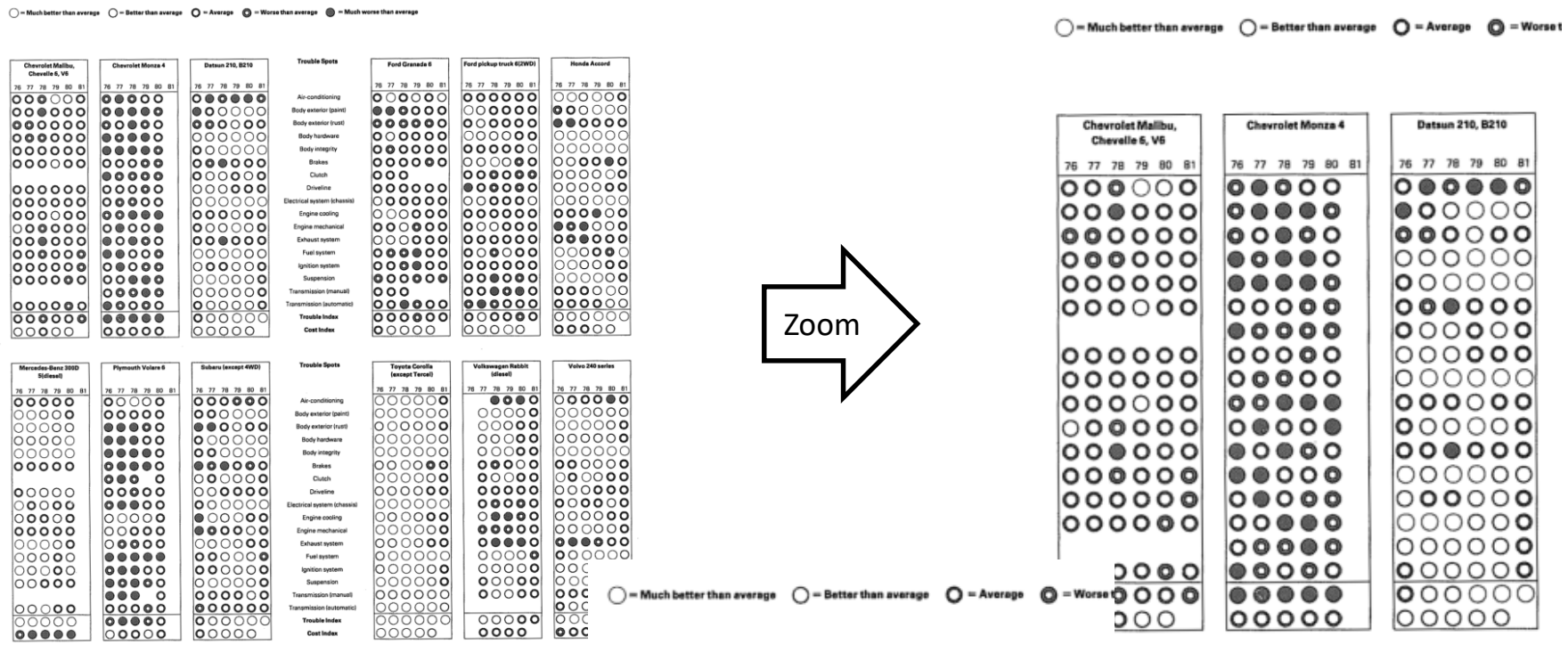
Visual Information-Seeking Mantra (Shneiderman Mantra)

Overview first, zoom and filter, then details-on-demand

- **Overview**: Gain an overview of the entire collection.
- **Zoom**: Zoom in on items of interest
- **Filter**: Filter out uninteresting items.
- **Details-on-demand**: Select an item or group and get details when needed
- **Relate**: View relationships among items
- **History**: Keep a history of actions to support undo, replay, and progressive refinement
- **Extract**: Allow extraction of sub-collections and of the details when needed.

Ben Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336-343, Washington. IEEE Computer Society Press, 1996.

Visual Information-Seeking Mantra (Shneiderman Mantra)



Details

Filter

Interactive Data Analysis – Selection

- Adding **interaction capabilities** to information visualizations extends the possible **interpretation and working scope** of these kinds of visualization massively

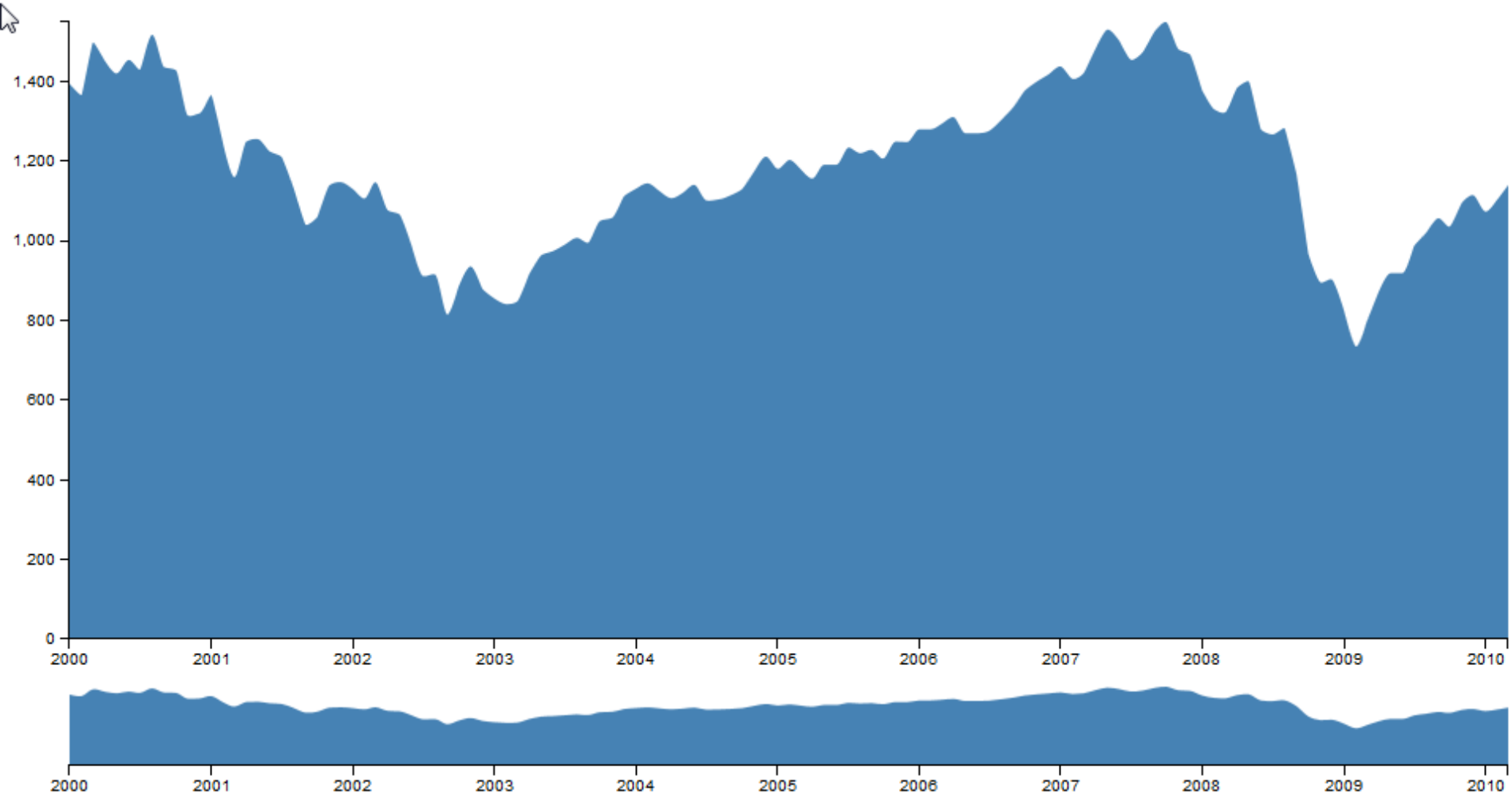
- One well known interaction paradigm in information visualization is the **multi-view paradigm** or **multiple coordinated views**

- The multi-view paradigm is (mainly comprised) of two interaction metaphors:

  1. **Linking** – Selection in one view gets propagated to other connected representations
  2. **Brushing**

http://bl.ocks.org/mbostock/raw/1667367/

Multiple View – Brushing

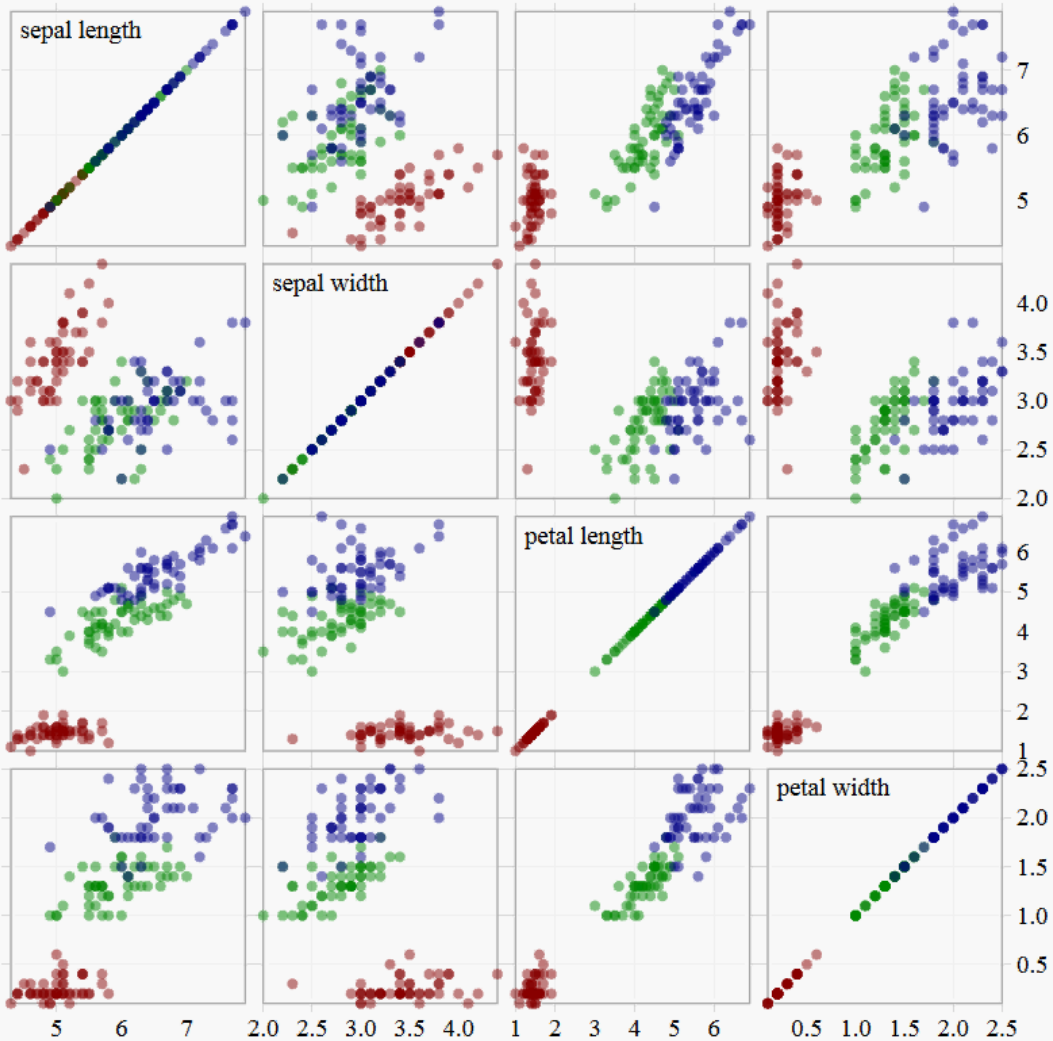


http://mbostock.github.io/d3/talk/20111116/iris-splom.html

Multiple View – Brushing

Iris Setosa Iris Virginica Iris Versicolor

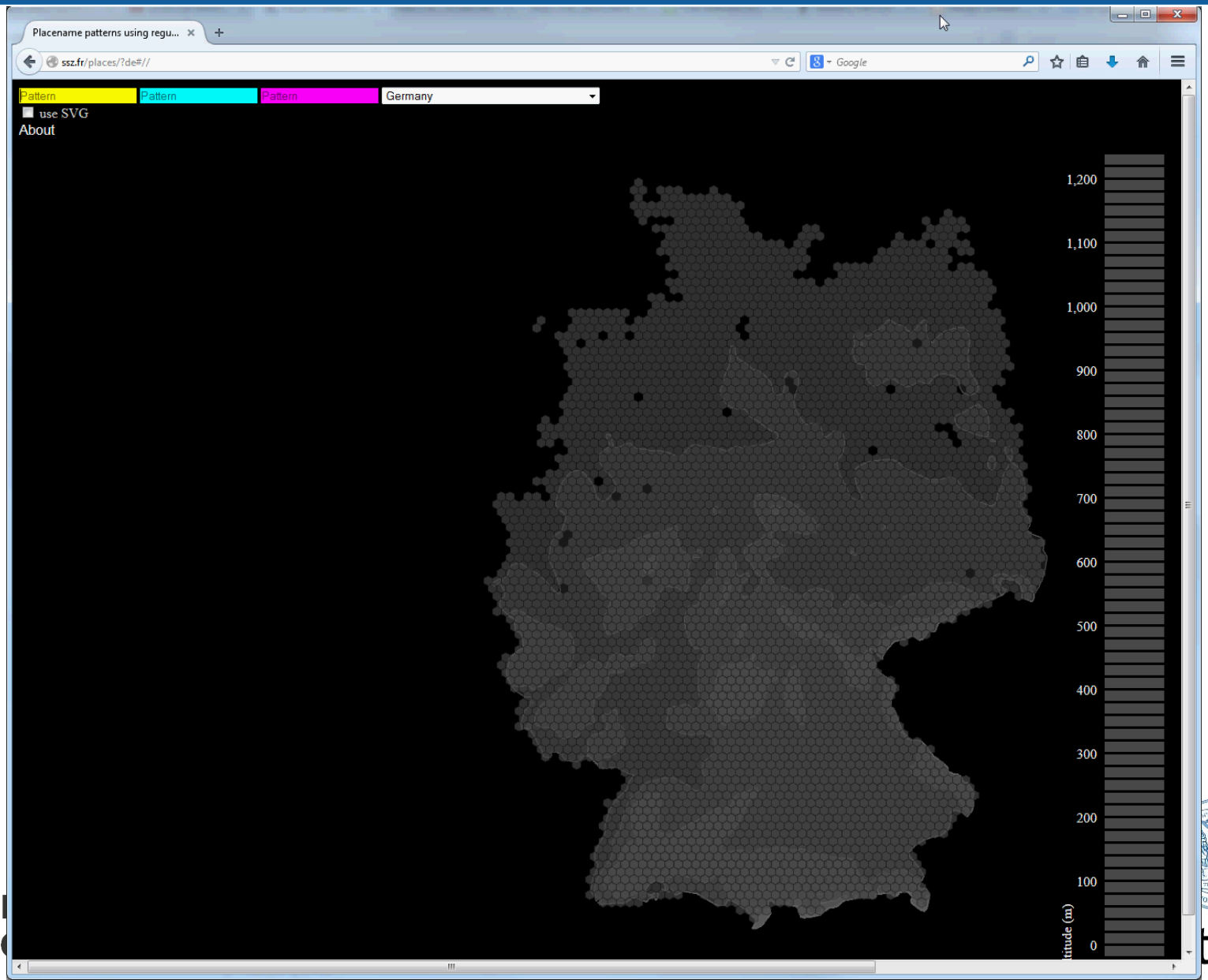
https://c.mql5.com/2/3/Figure10_Iris_flower.png



- *Iris setosa*
- *Iris versicolor*
- *Iris virginica*

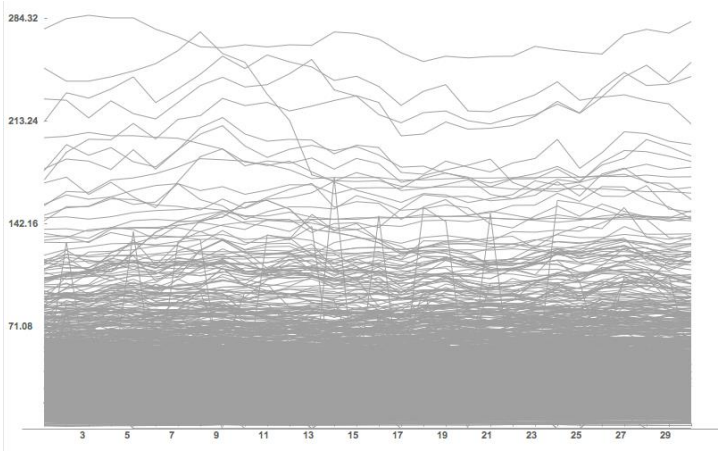
Edgar Anderson's *Iris* data set scatterplot matrix

Interactive Data Analysis – Filtering

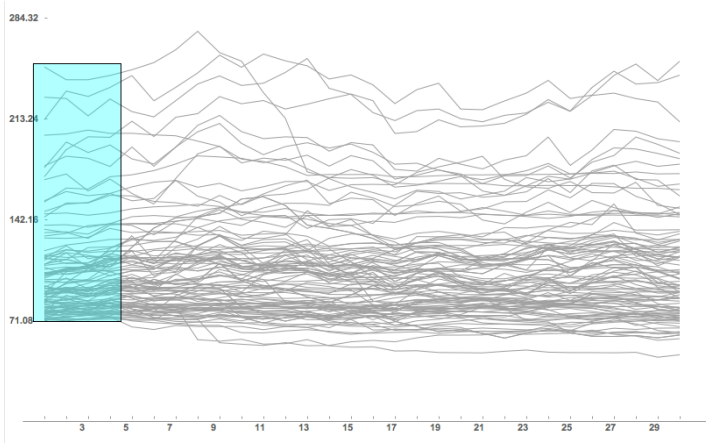


Interactive Data Analysis – Filtering

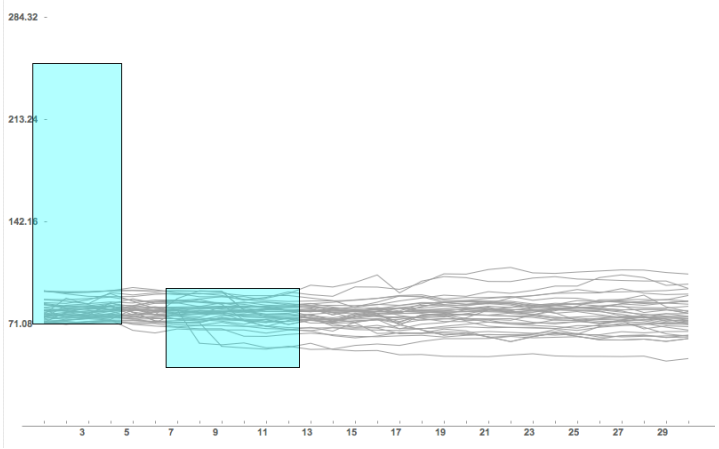
(a)



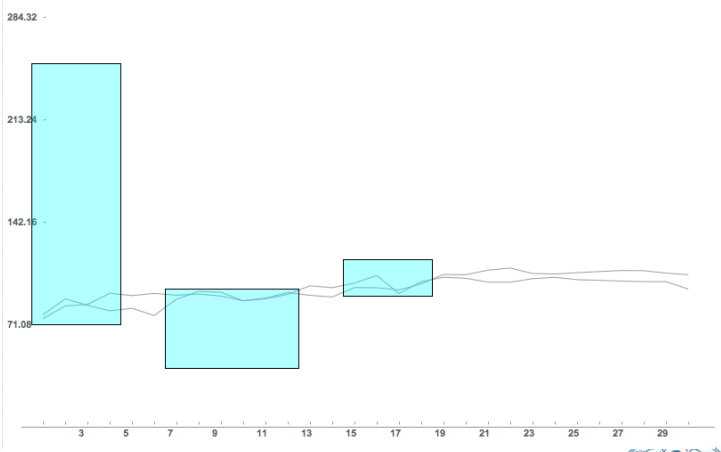
(b)



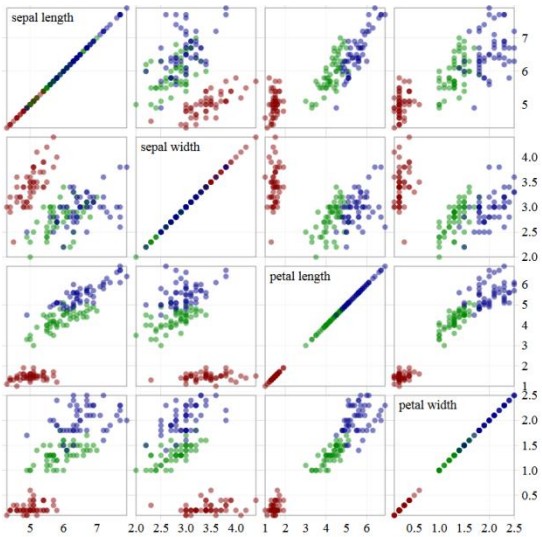
(c)



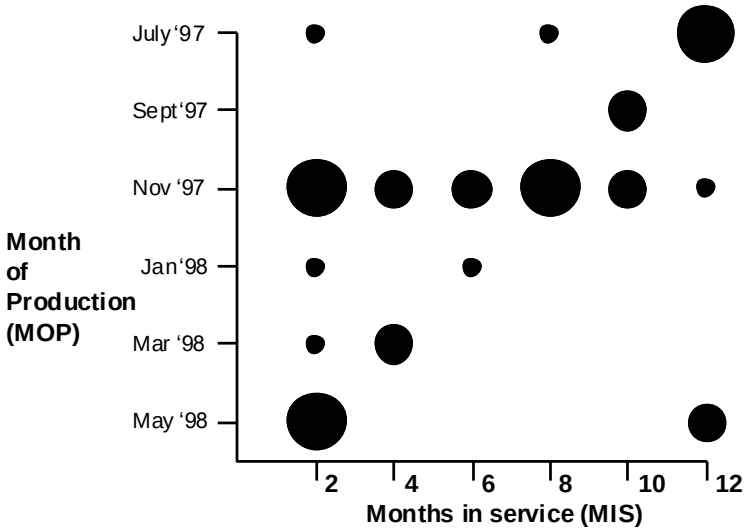
(d)



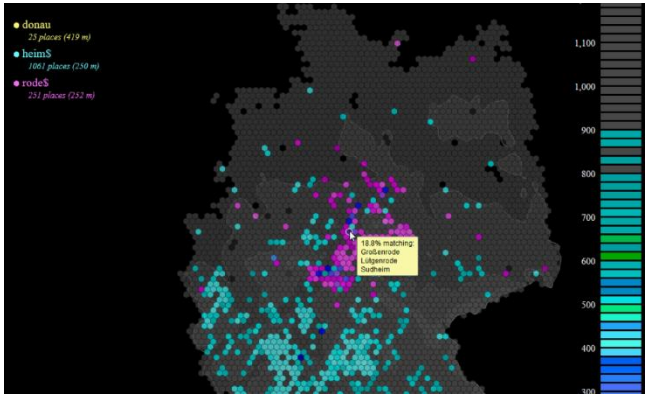
Presentation of Values Sets – N Dimension



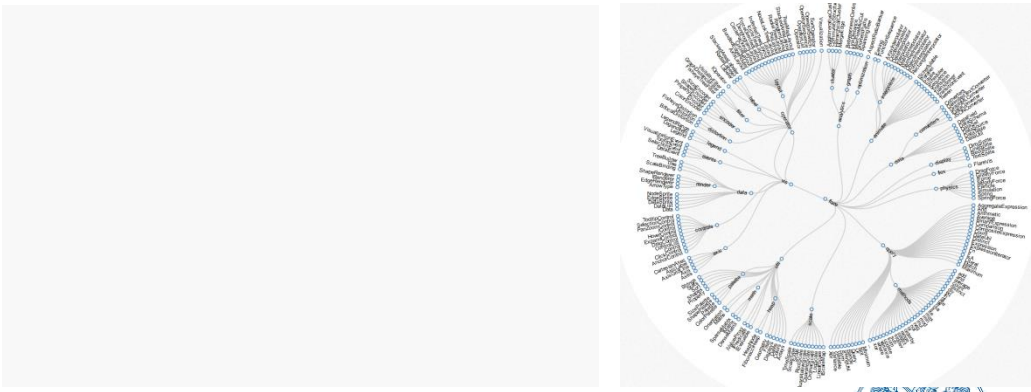
Scatter Plot Matrix



Extended Scatter Plots



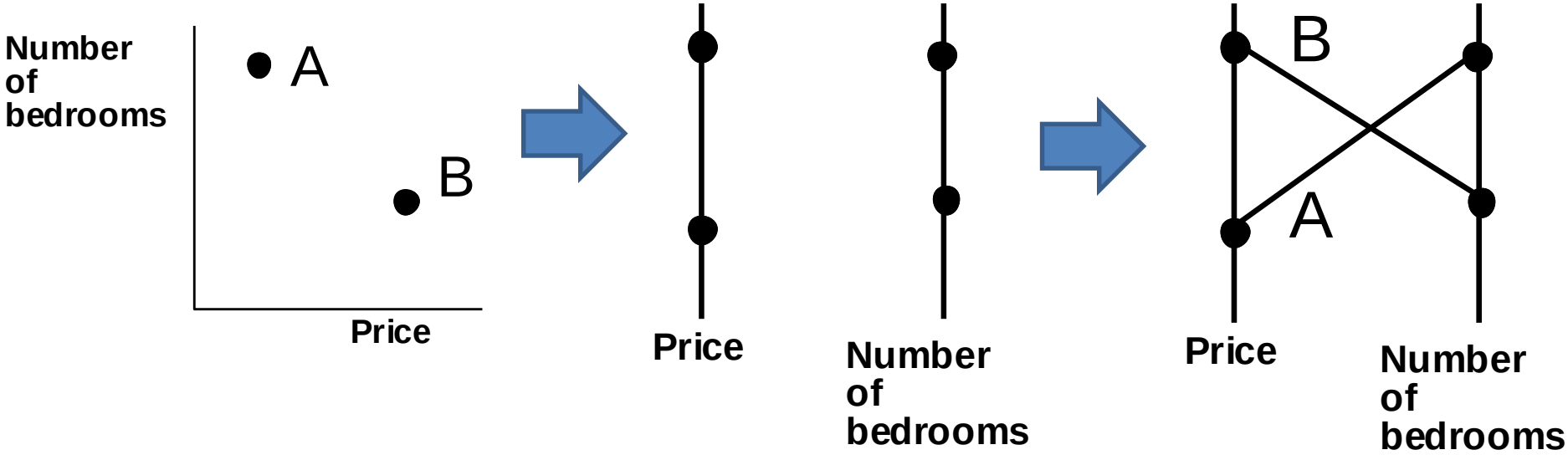
Geographical Hexbins



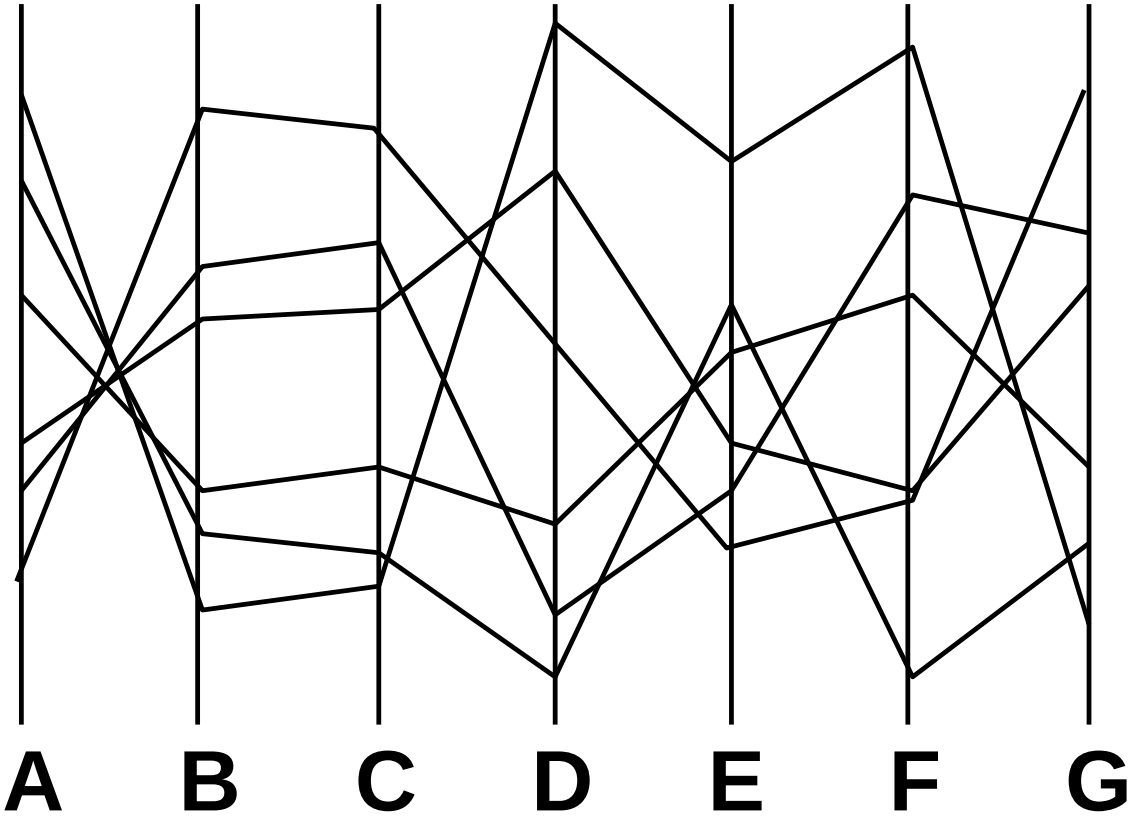
Trees & Graphs

Parallel Coordinates

- Is a simple solution to present multi-dimensional data
- There are some problems with this type of representation but is still a way to present multi dimensions in one chart

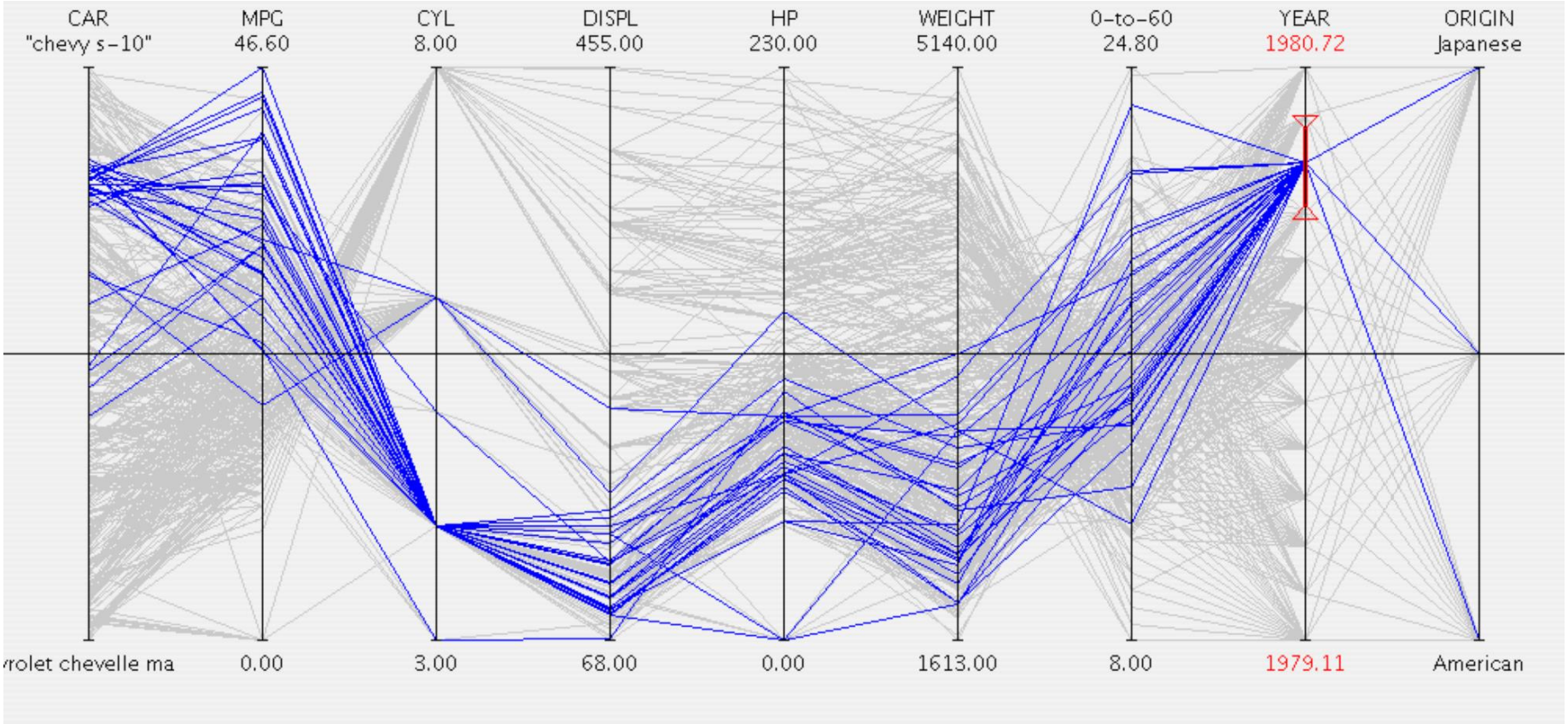


Coordinate Plots



- It does not become clear that there is a trade-off between B and E, where C and G show a strong correlation
- Interaction can be solution to this draw back

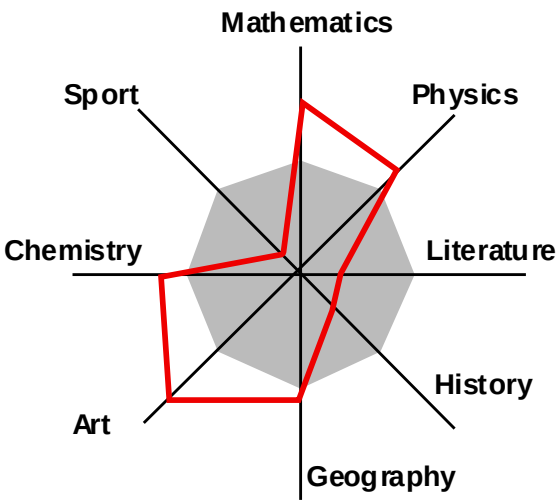
Coordinate Plots – Parallel Coordinates

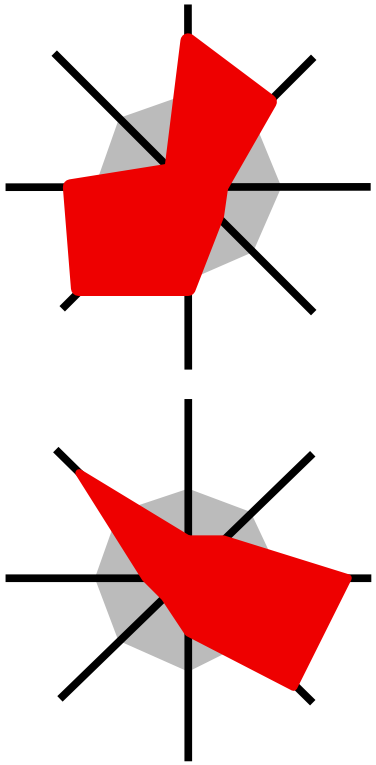


- Brushing, Filtering, Sorting…

Star Plots

- Close related to parallel coordinates
- In case of star plots, parameter axises are arranged in radial fashion





- They can be used for comparing different data sets, e.g., the performance of students
- They support the visibility of object and structures in the data

Dust Magnet – Interactive Data Analysis Approach

Dust & Magnet:
Interactive Visualization for Everyday Data

Ji Soo Yi[1], Rachel Melton Ponder[2], John Stasko[2], & Julie A. Jacko[1]

[1]Industrial and Systems Engineering, Georgia Tech
[2]GVU Center & College of Computing, Georgia Tech

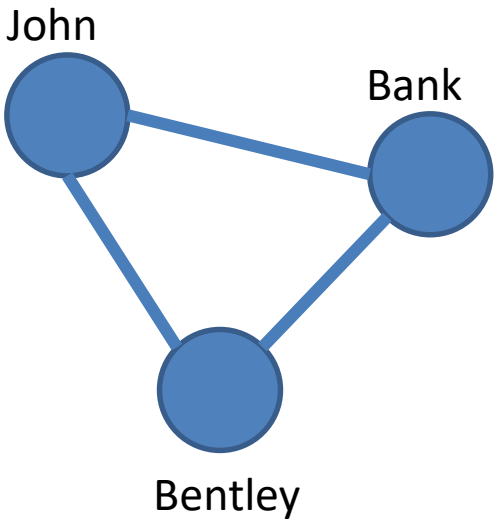
Relational Data

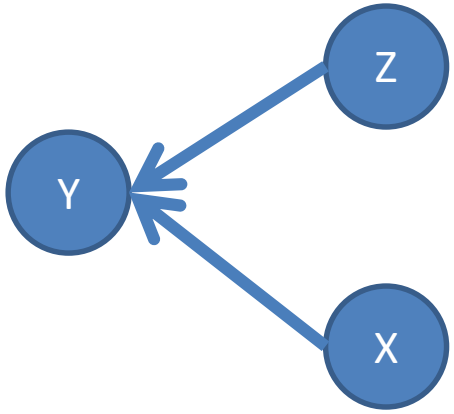
Literature

- Guiseppe Di Battista, Peter Eades, Roberto Tamassia, Ioannis G. Tollis, (1998), *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice Hall.
- Robert Spence, (2014), *Information Visualization: An Introduction*, Springer.
- Matthew O. Ward, Georges Grinstein, Daniel Keim, (2015), Interactive Data Visualization – Foundations, Techniques, and Applications, CRC Press.

Representation of Relations

- A relation can be defined as logical or natural association between two or more things; as relevance of one to something else; as connection

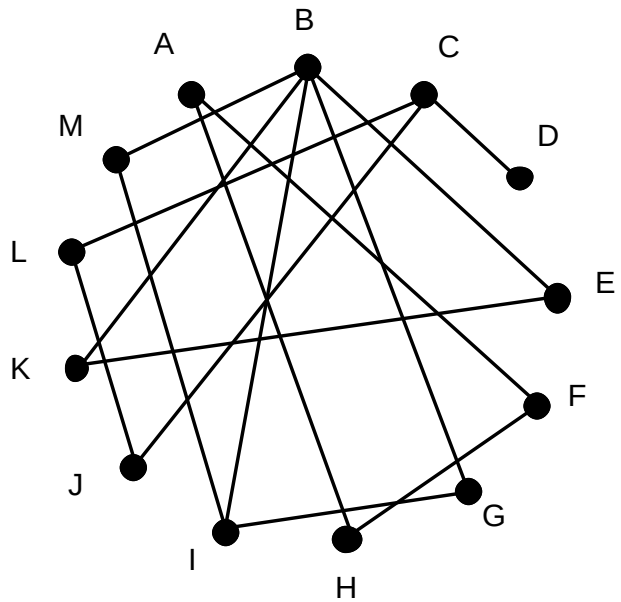




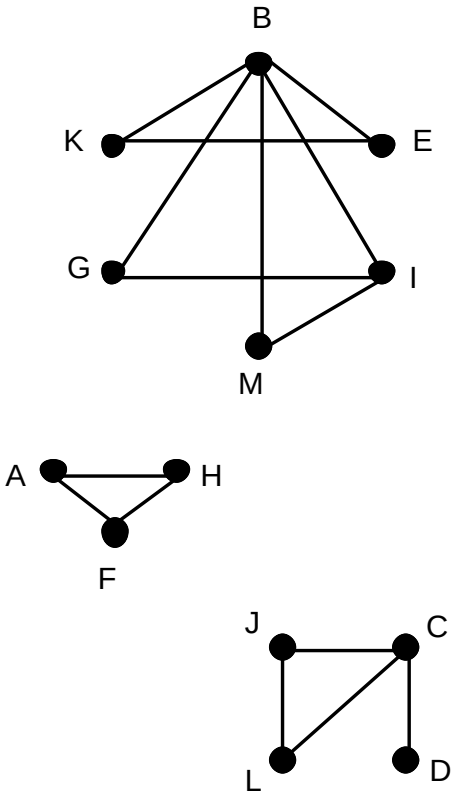
Lines

Originator	Receiver
A	H
C	L
I	M
B	E
F	H
G	I
I	B
B	M
K	B
G	B
K	E
C	J
D	C

(a)



(b)

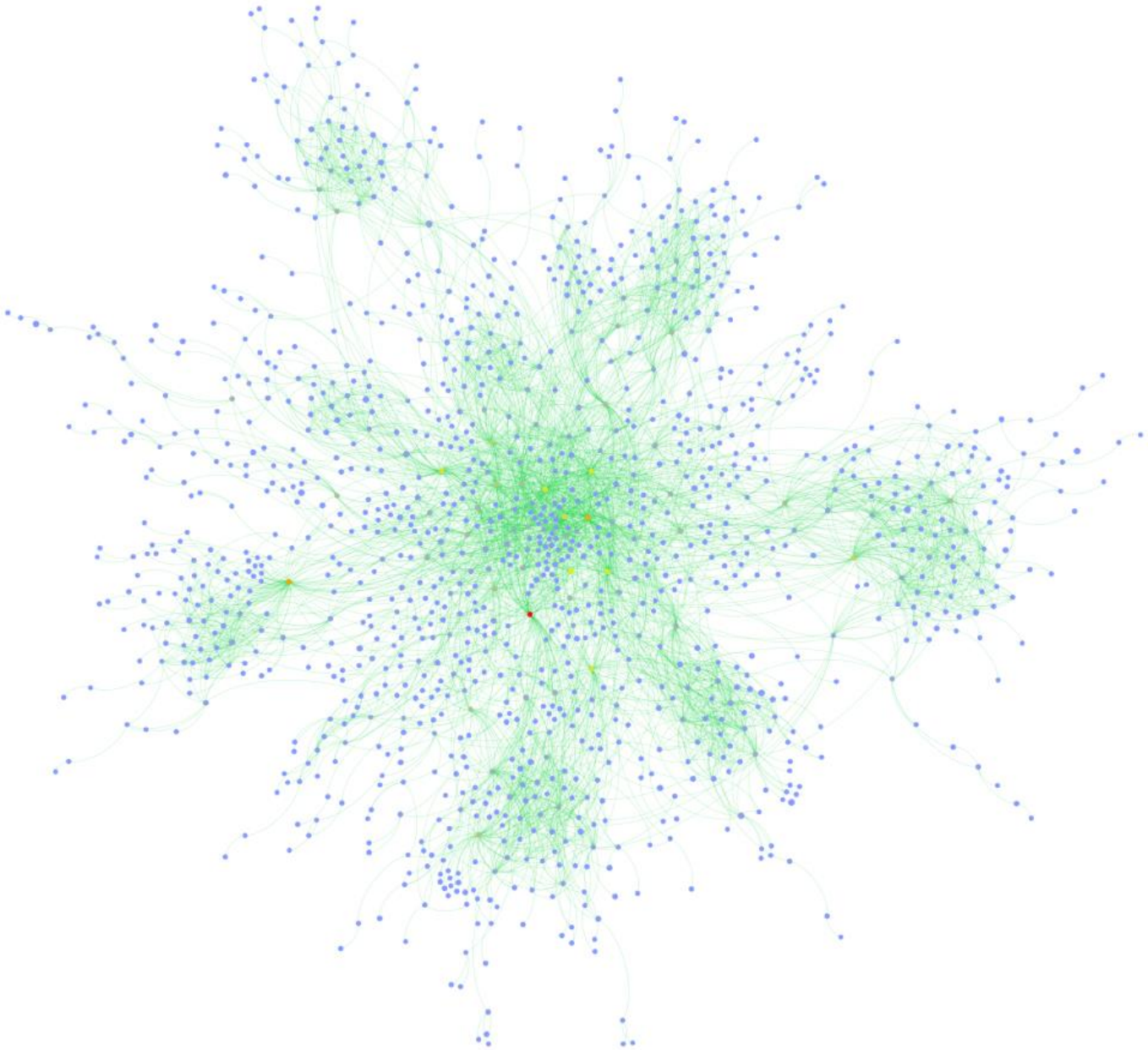


(c)

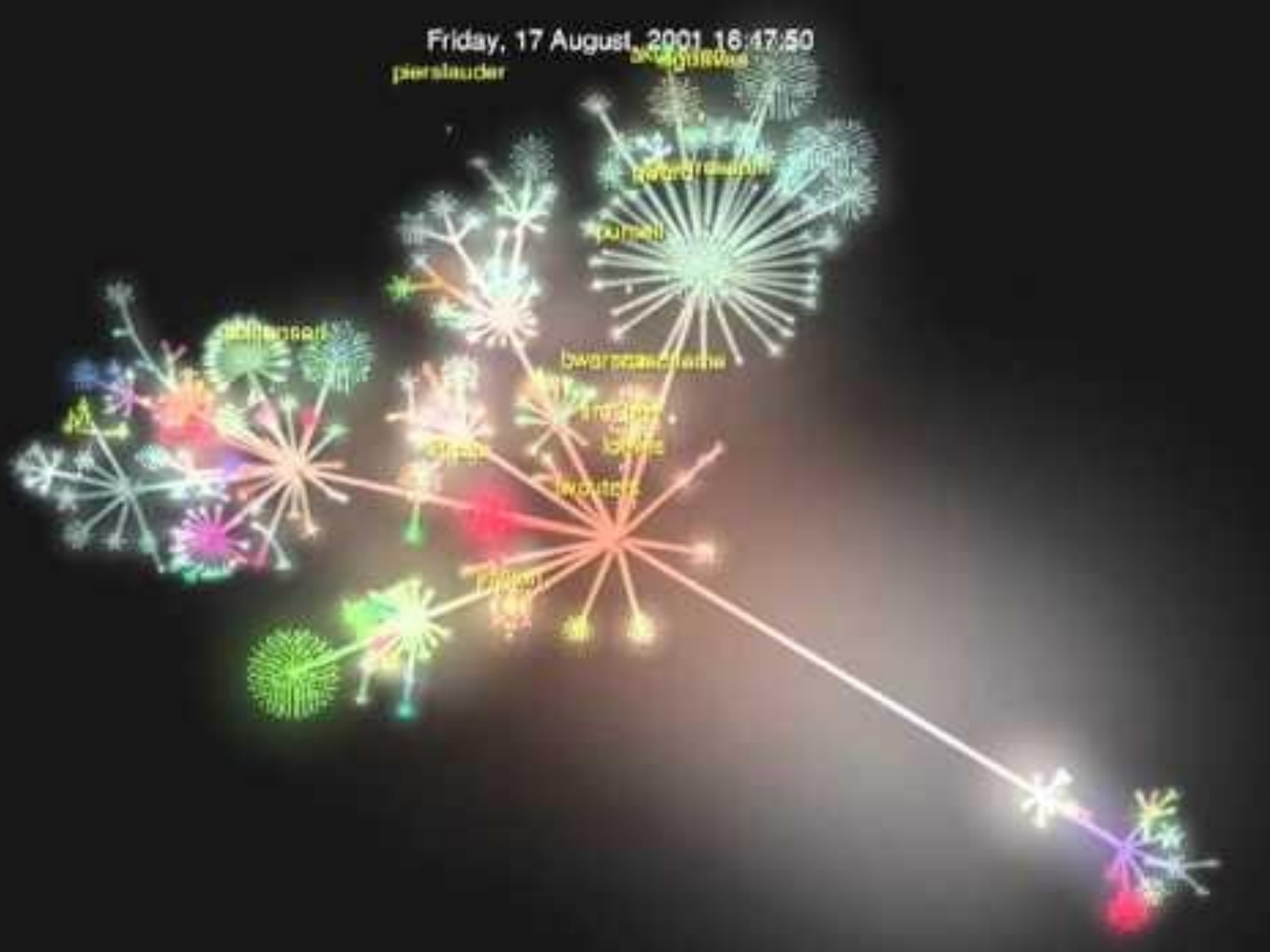
Lines - Example

- The resulting representation is less meaningful

Examples

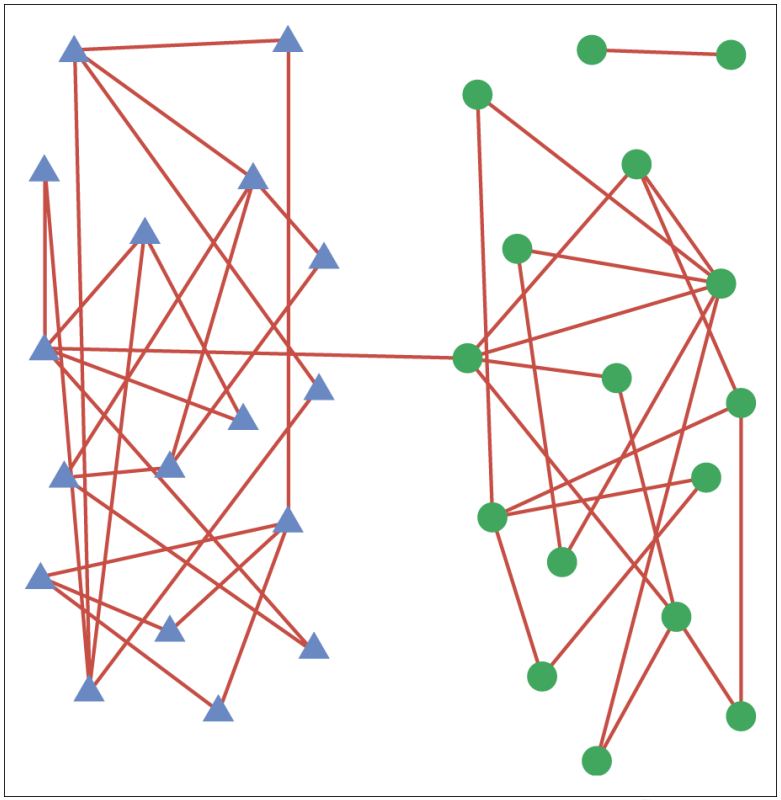


http://thepoliticsofsystems.net/wp-content/uploads/2014/02/netvizz_europeangreens_last50posts_users_betw-1024x1024.png


Friday, 17 August, 2001 16:47:50
pierslauder
purnell

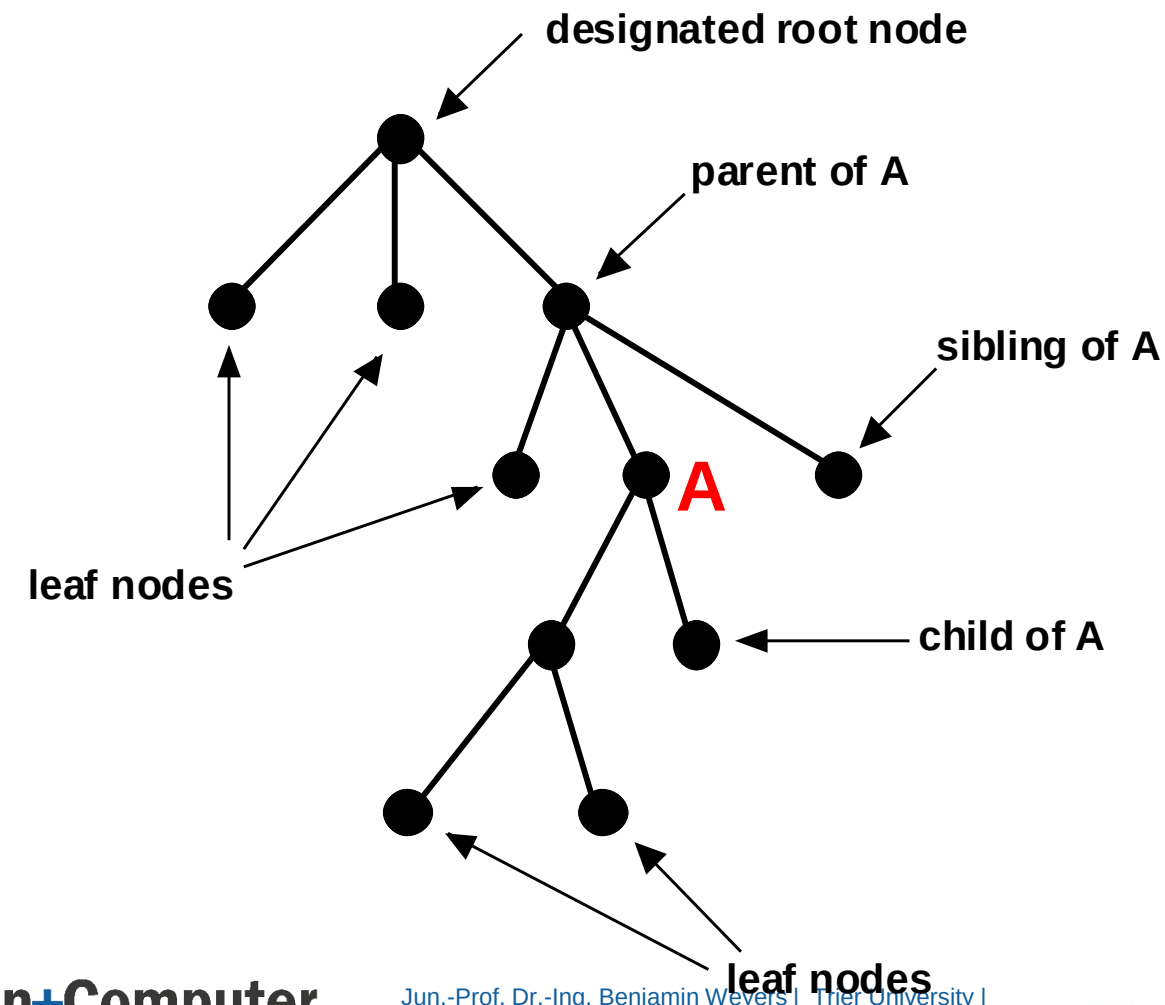
Node-Link Diagrams

- **Node-Link Diagrams** are visual presentations of relational data structures
- Relational data structures refer to graphs
- Special types of graphs are **trees**



Trees

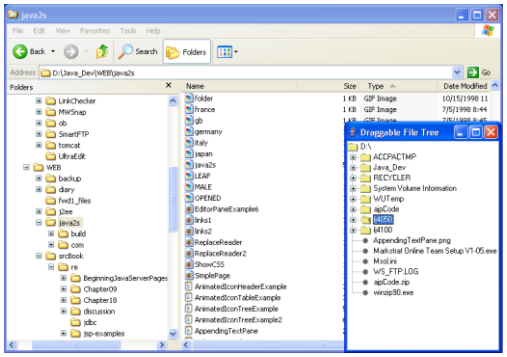
- Trees are a sub-group of graphs with a specific structure



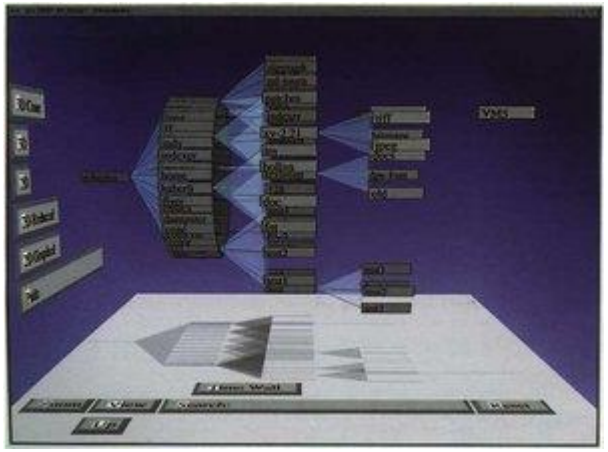
3D Cone Trees

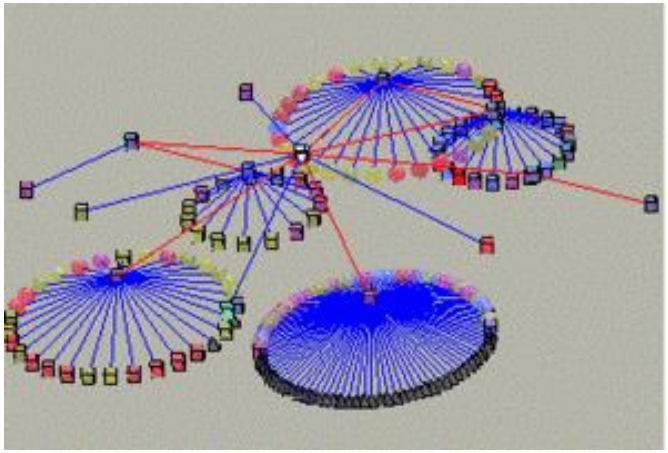
3D Cone Trees

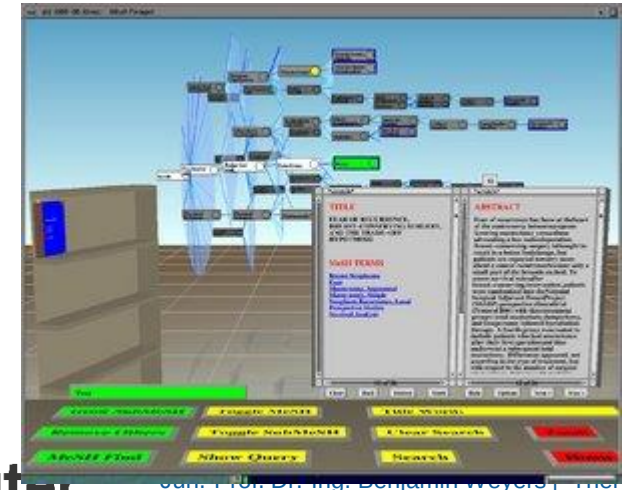
- Cockburn et al. 2000 evaluated that Cone Trees are really attractive to work with, but are less efficient than standard 2D tree (file browsers)

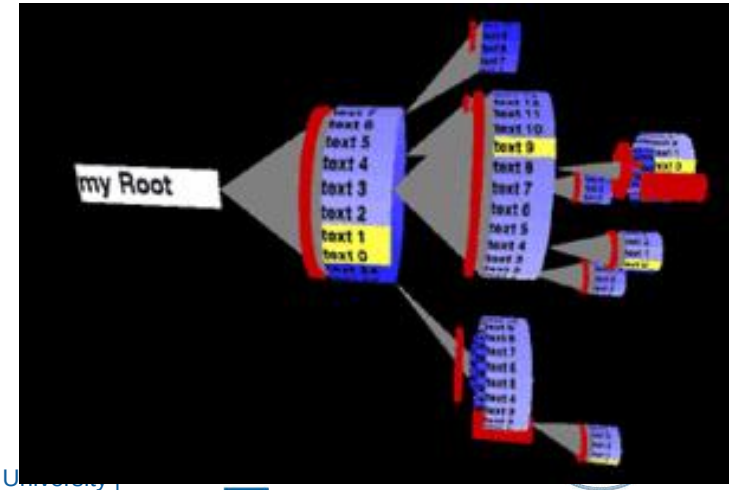


http://www.java2s.com/Code/JavaImages/FileTreeDragSource.PNG





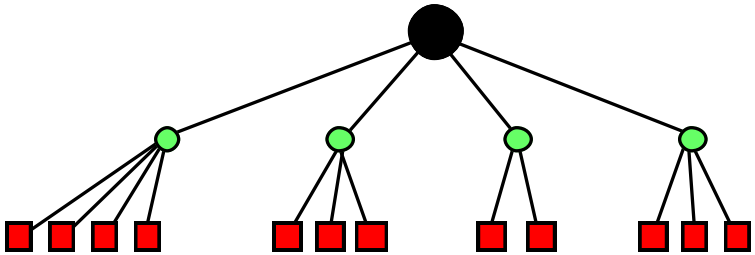




http://www.infovis-wiki.net/index.php?title=Cone_Trees

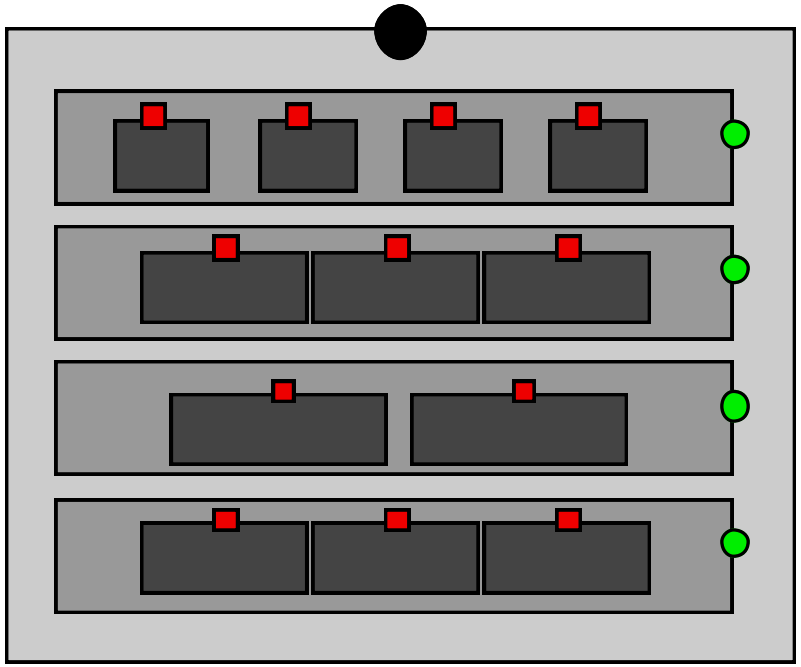
Tree Maps

- Depth of encapsulation directly relates to depth in the tree

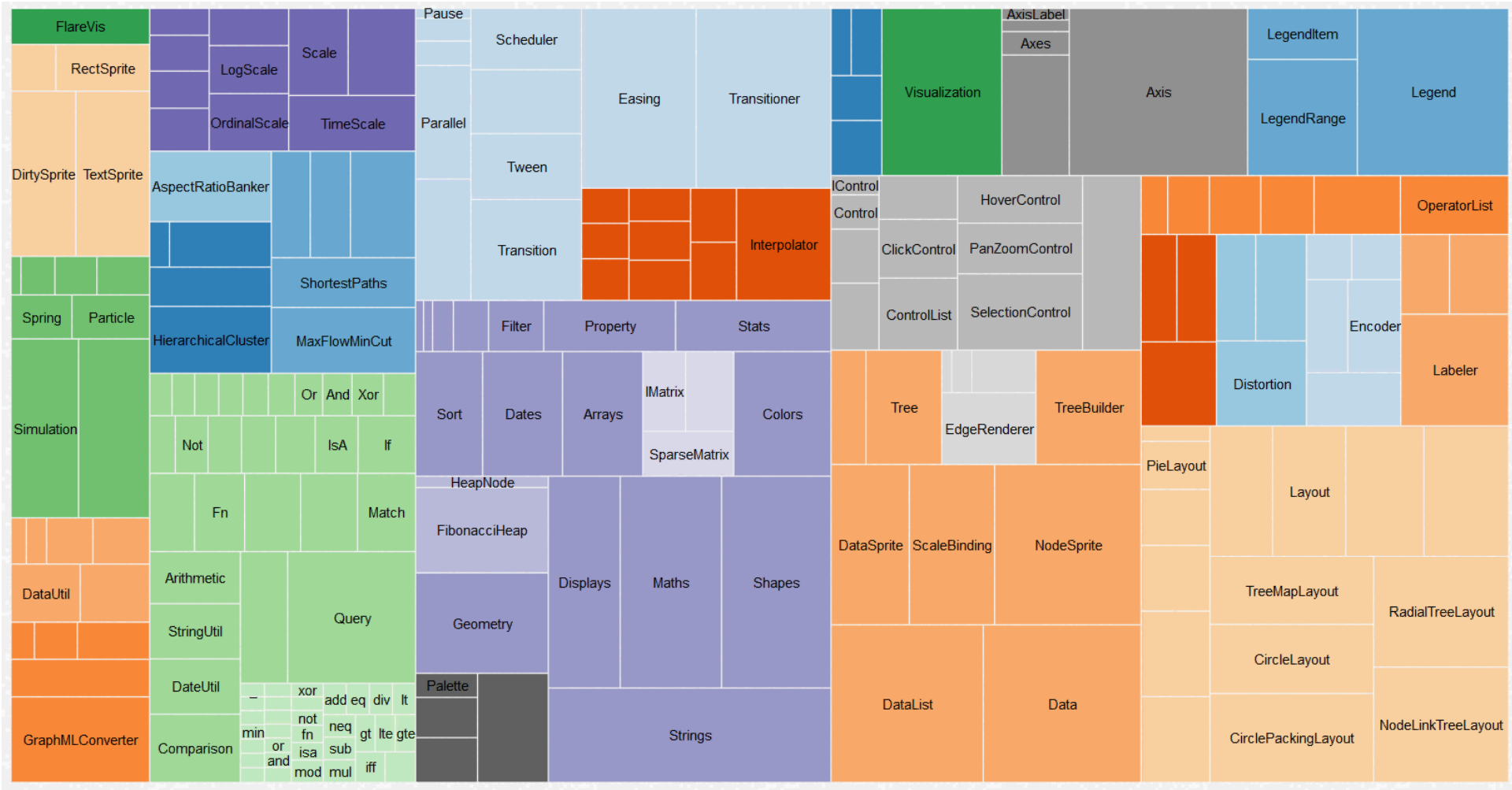


Tree

Tree Map



Tree Maps

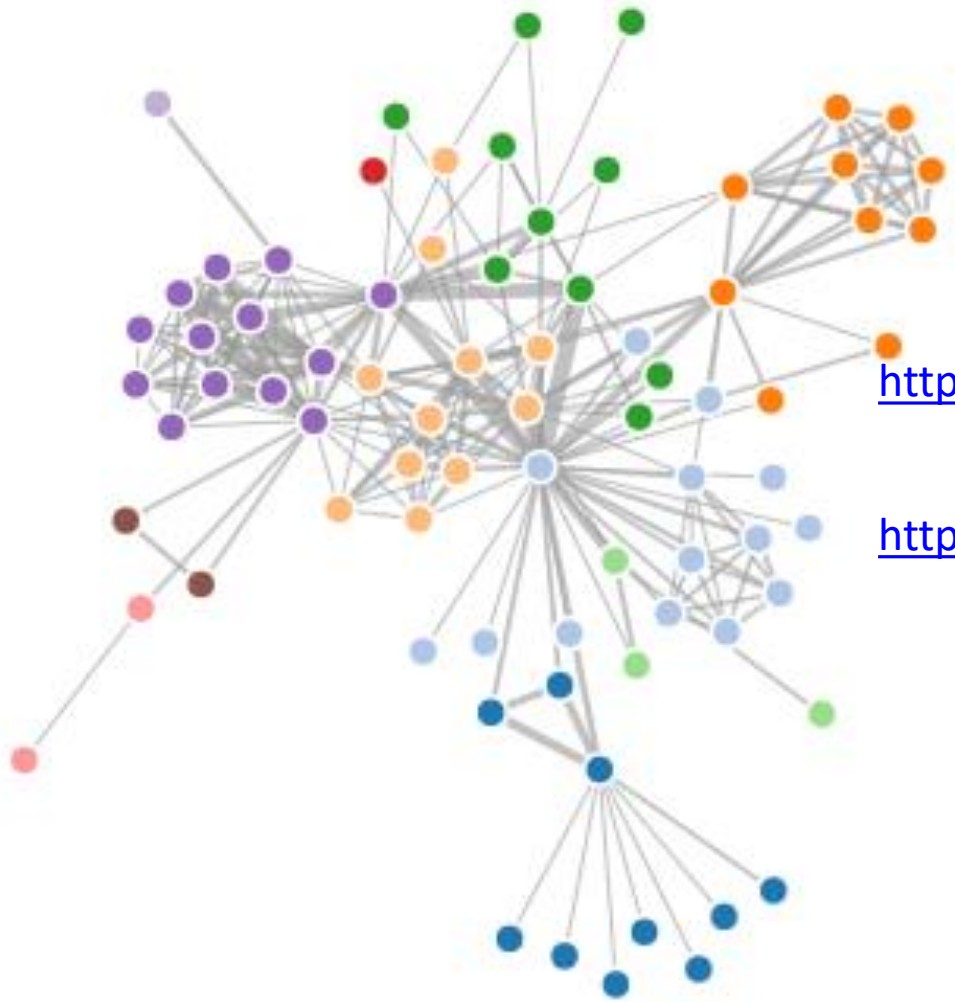


d3.layout.treemap

click or option-click to descend or ascend

Size

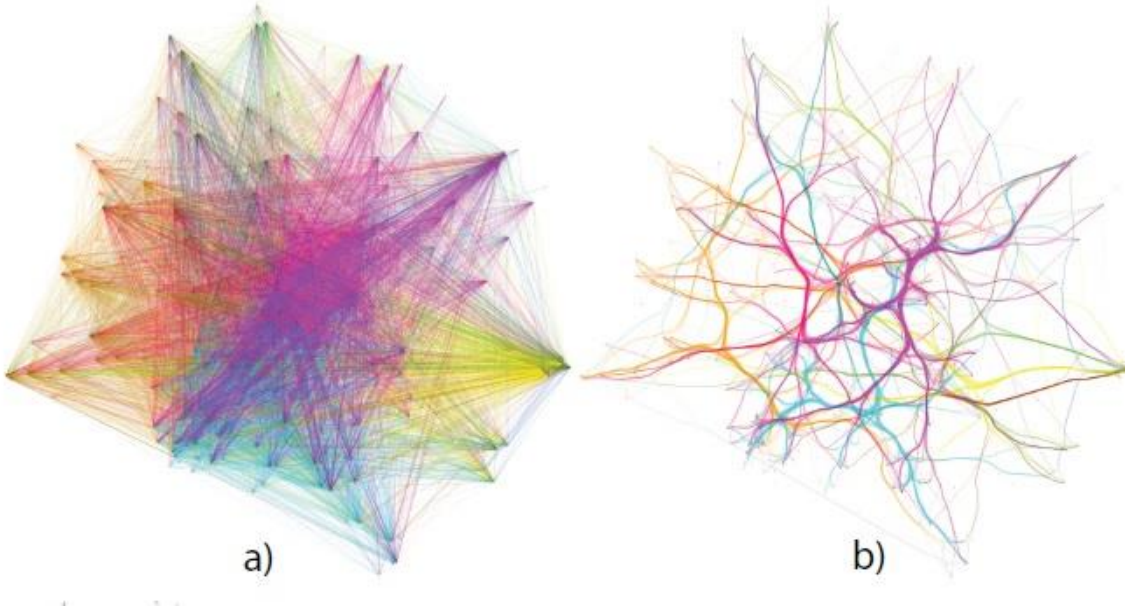
Force-Directed Approach



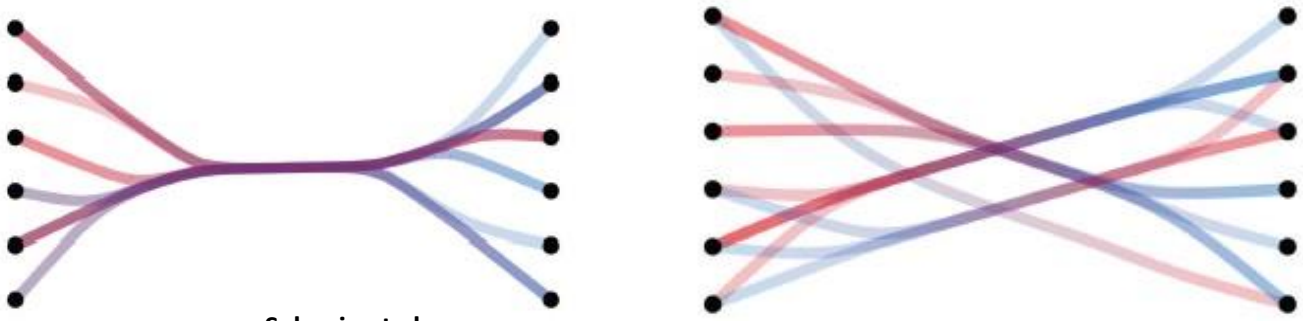
http://bl.ocks.org/couchand/raw/6420534/

https://vida.io/documents/N4jSip7n68yQ48DXp

Edge Bundling

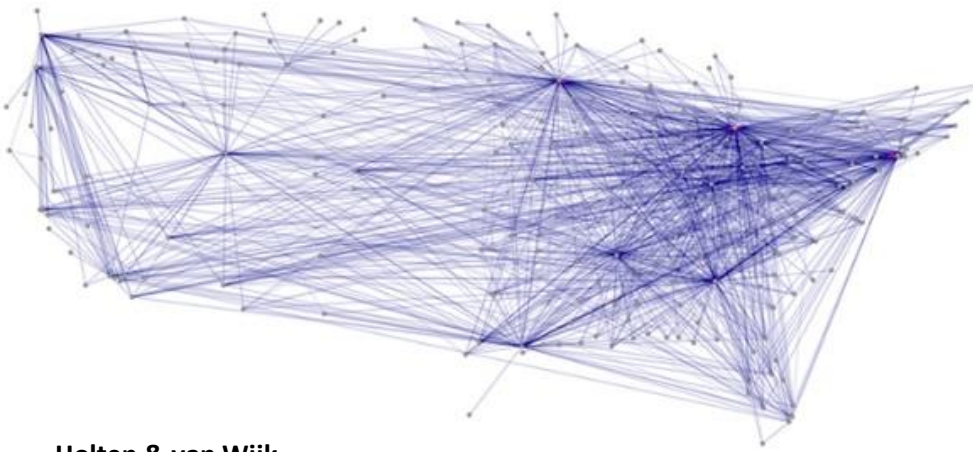


Ersoy et al.
Skeleton-based edge bundling for graph visualization



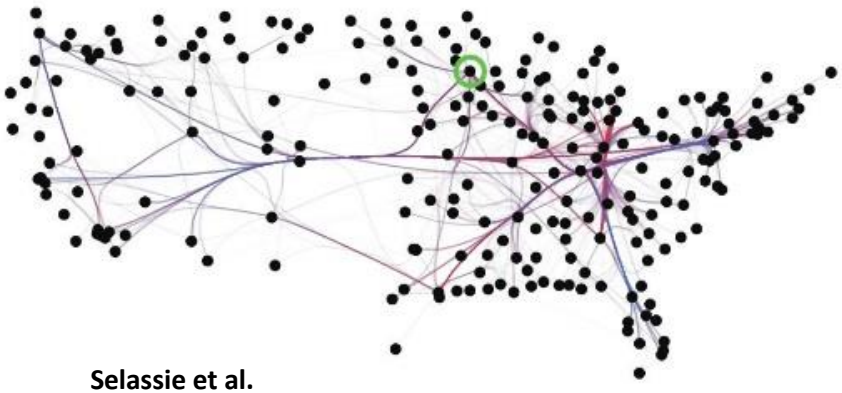
Selassie et al.
Divided Edge Bundling for Directional Network Data

Edge Bundling



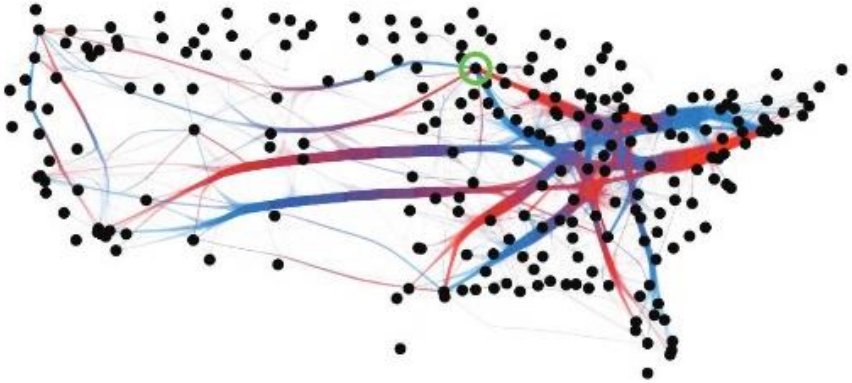
Holten & van Wijk
Force-Directed Edge Bundling for Graph Visualization

Force Directed Edge Bundling

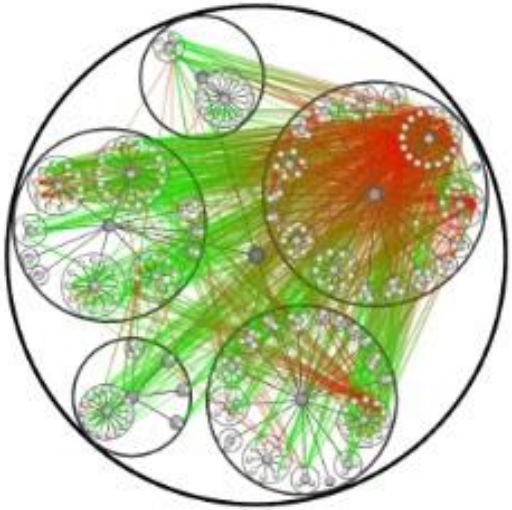


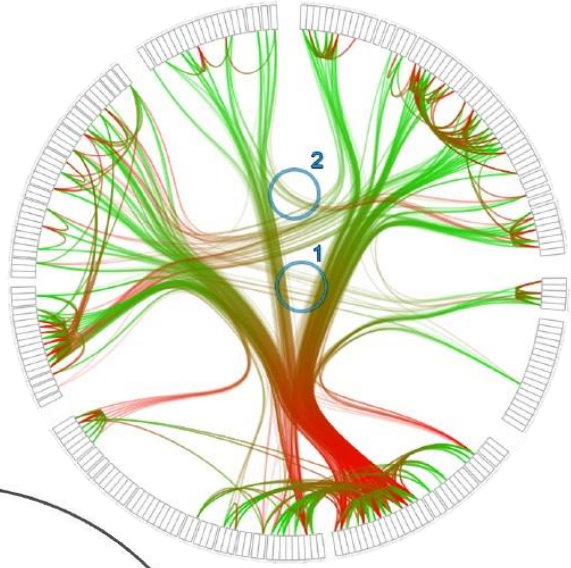
Selassie et al.
Divided Edge Bundling for Directional Network Data

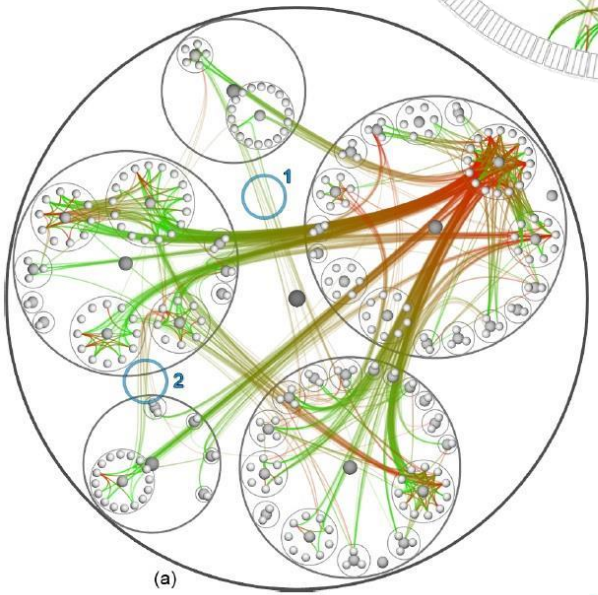
Divided Edge Bundling



Edge Bundling

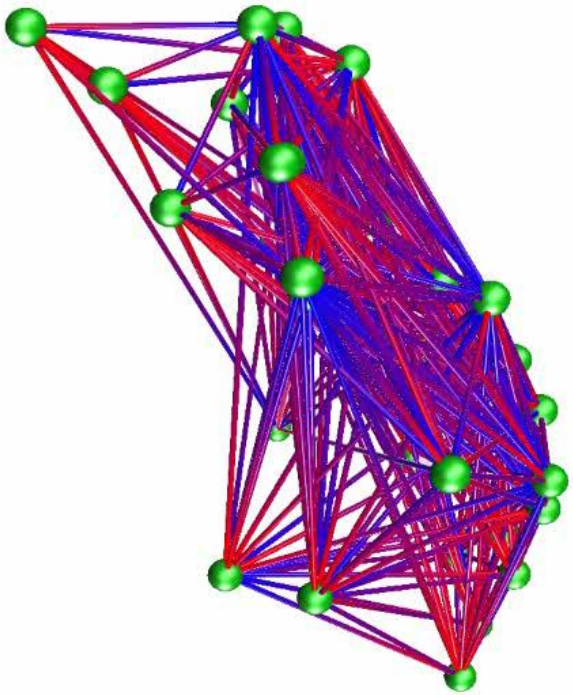






Holten
Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data

Edge Bundling – short view to 3D graph drawing

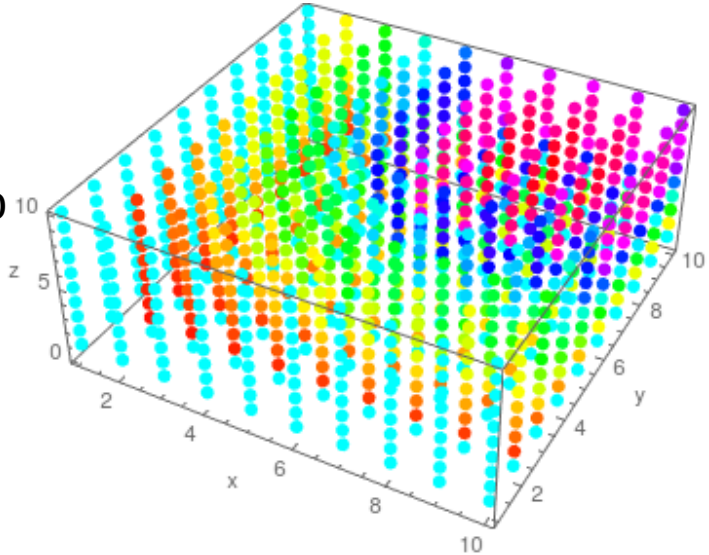


Spatial Data

Scientific Visualization

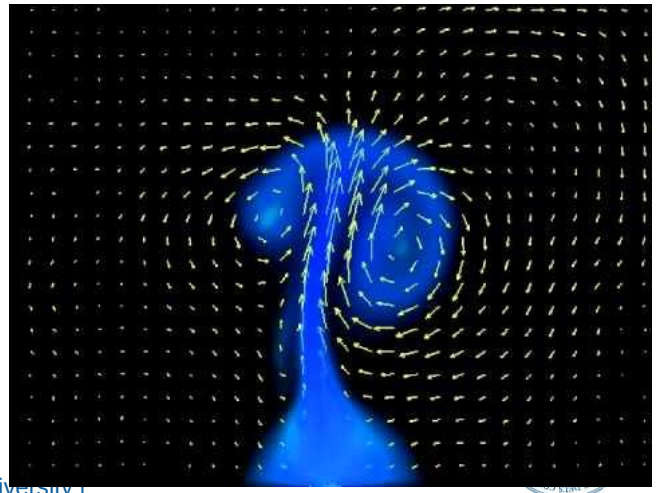
Data Types

- Scalar Fields (structured/unstructured grid)
  - Explicit Volume Rendering → Ray Casting
  - Implicit Volume Rendering → Extraction of Iso Surface/Marging Cubes



https://i.stack.imgur.com/187c9.png

- Vector Fields (structured/unstructured grid)
  - Glyphs
  - Streamlines
  - Patricle Tracing
  - Line Integral Convolution
  - Topology Methods

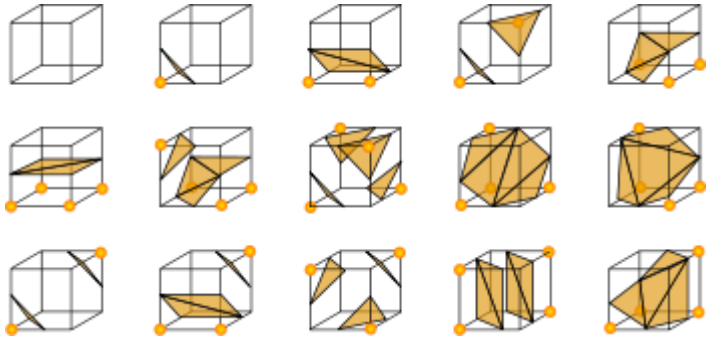


https://i.ytimg.com/vi/S7c7QYufPo8/hqdefault.jpg

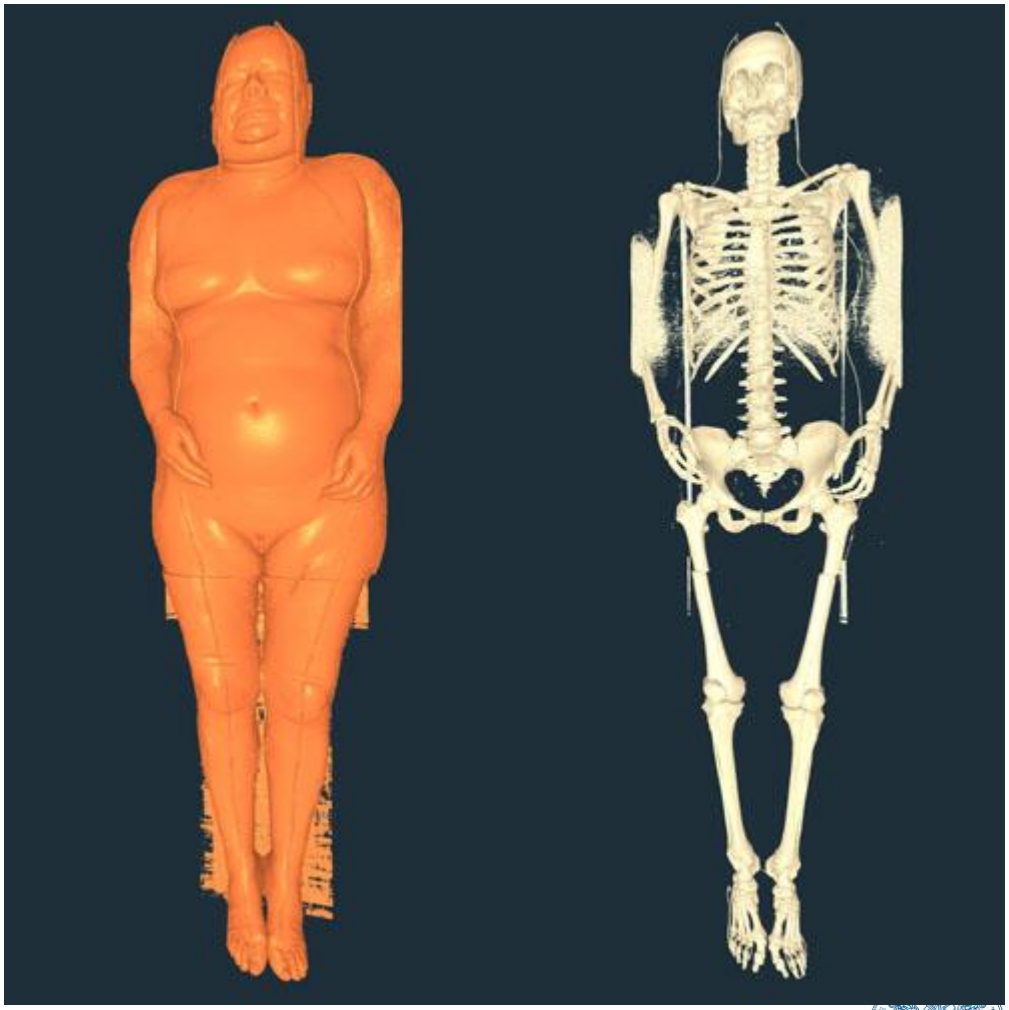
- Tensor Fields

Iso Surface – Implicit Volume Rendering

- Use of ray tracing or geometry rendering
- Extraction of Iso Surfaces: Marching Cubes

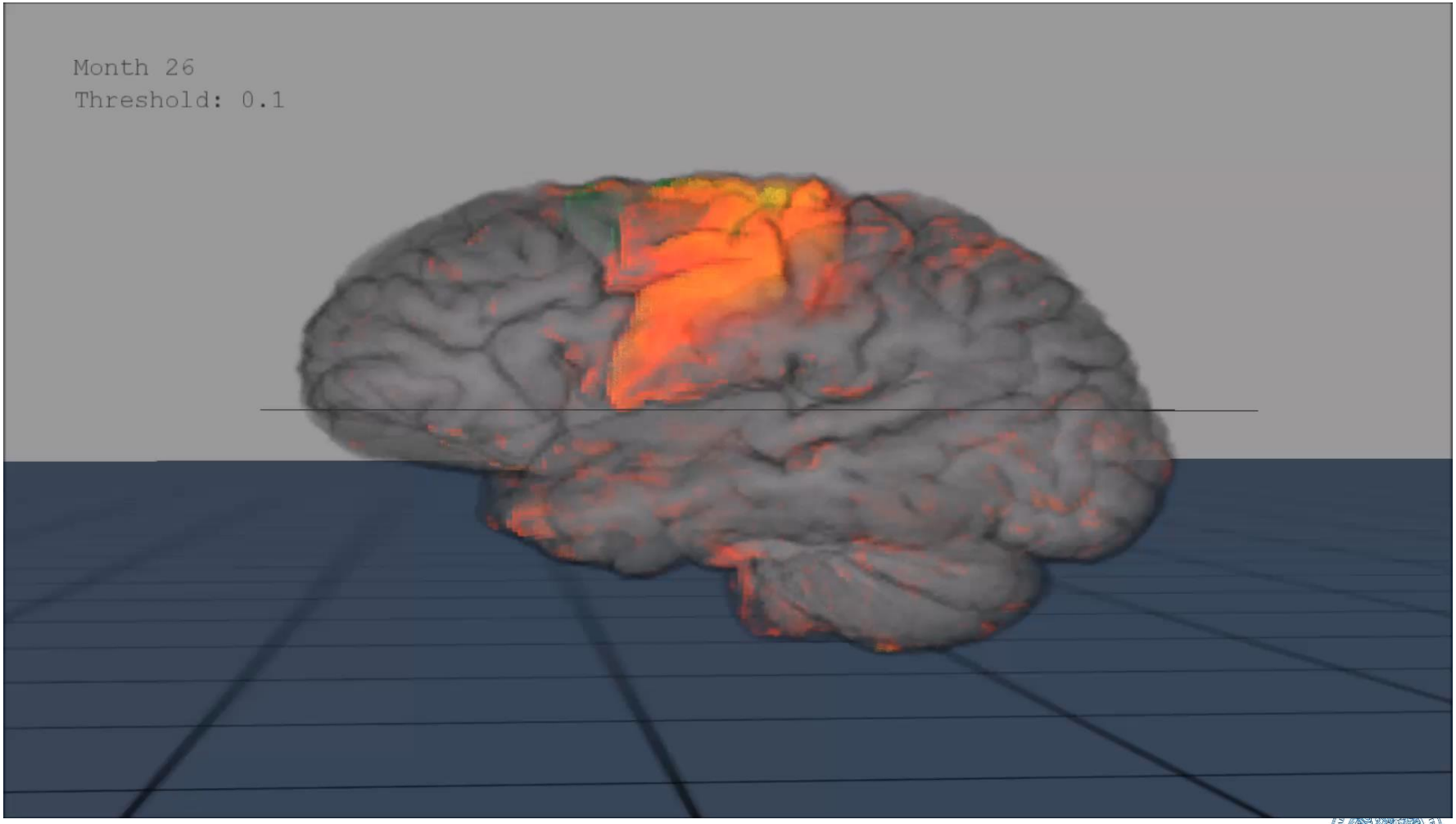


https://upload.wikimedia.org/wikipedia/commons/thumb/a/a7/MarchingCubes.svg/350px-MarchingCubes.svg.png

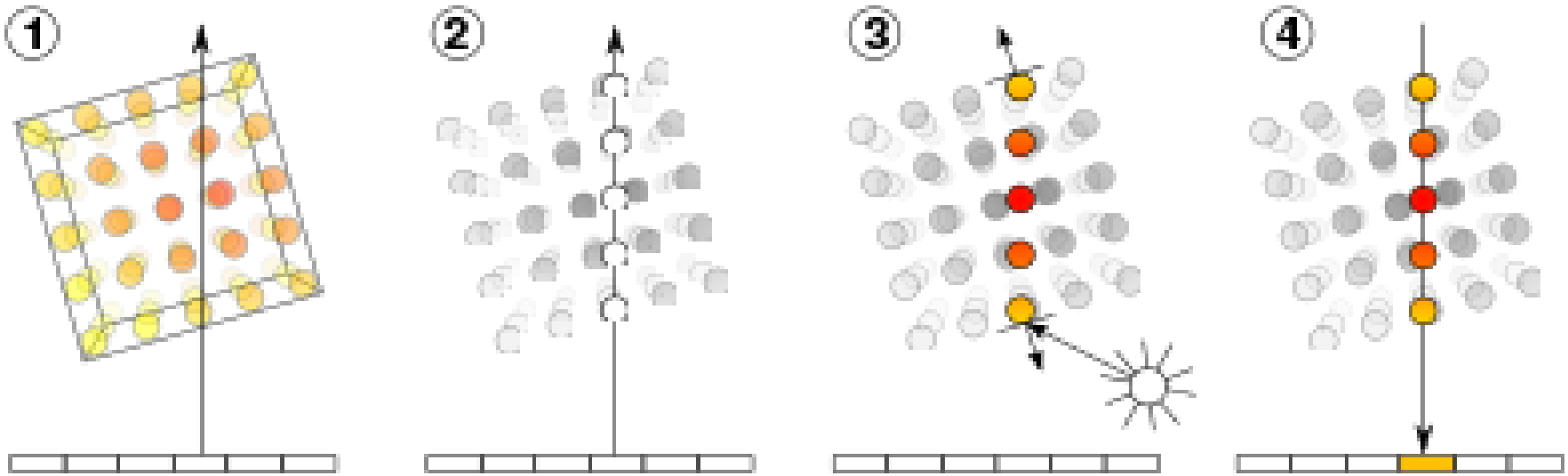


http://www.cs.utah.edu/~shirley/papers/iso/both.jpg

Volume Rendering – Ray Casting



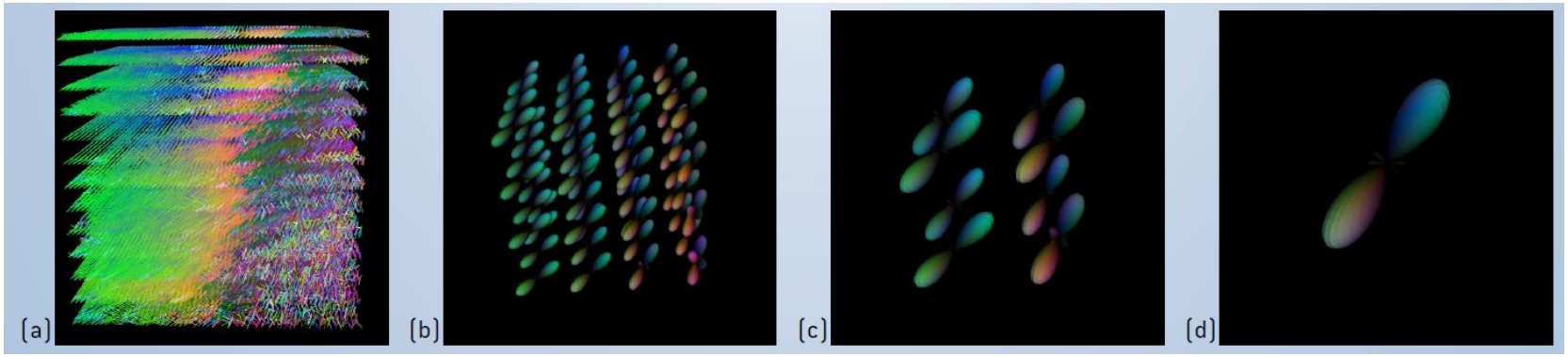
Ray Casting



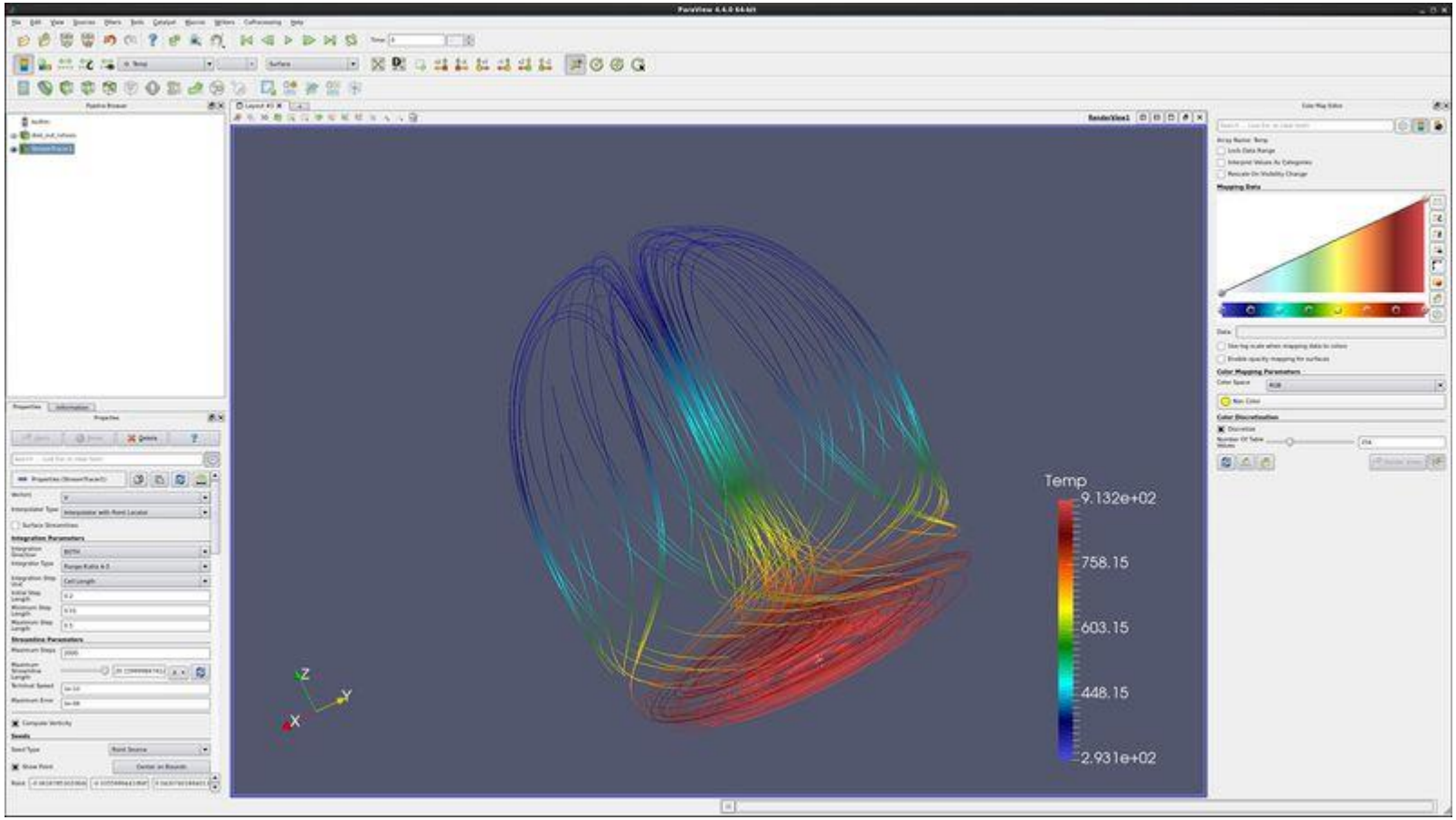
https://upload.wikimedia.org/wikipedia/commons/thumb/7/76/Volume_ray_casting.svg/400px-Volume_ray_casting.svg.png

1. Ray Casting
2. Sampling
3. Shading
4. Compositing

Vector Field - Glyphs



Vector Field – Stream Lines



https://www.paraview.org/Wiki/images/thumb/8/89/Beginning_filters_7.png/800px-Beginning_filters_7.png.jpeg

Examples from Neuroscience

VisNEST

PLIViewer

NeuroScheme

NeuroScheme

(this demo uses cortex data)

http://gmrv.es

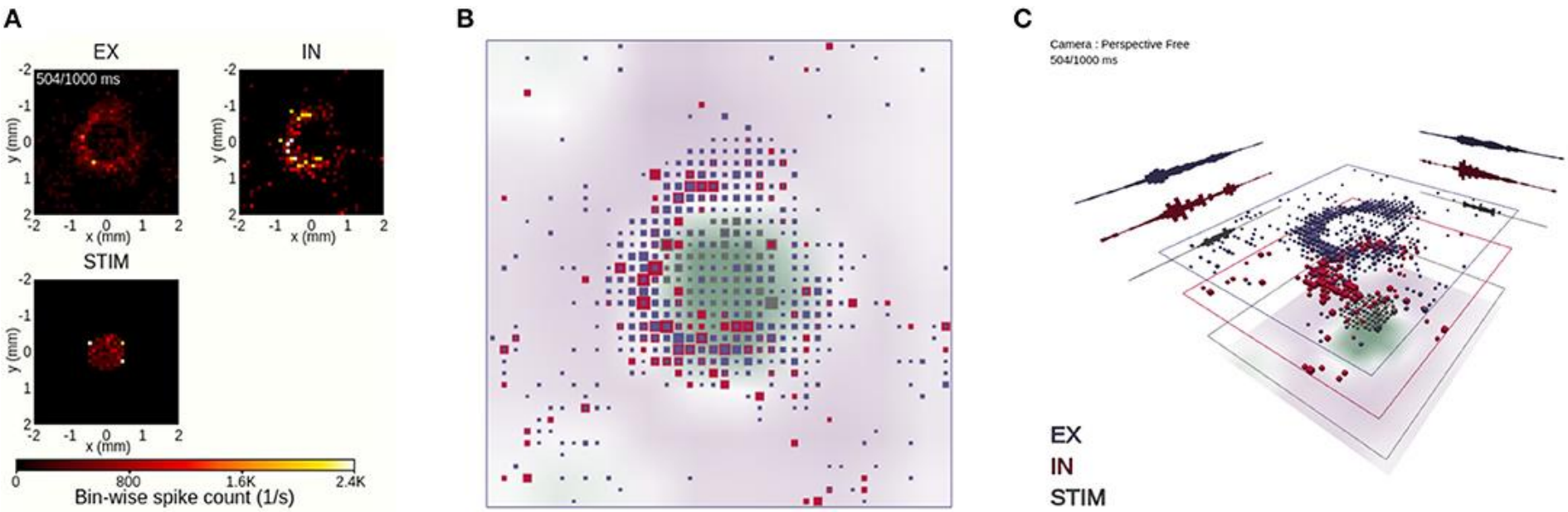
Universidad Rey Juan Carlos

POLITÉCNICA

VIOLA – Life Demo

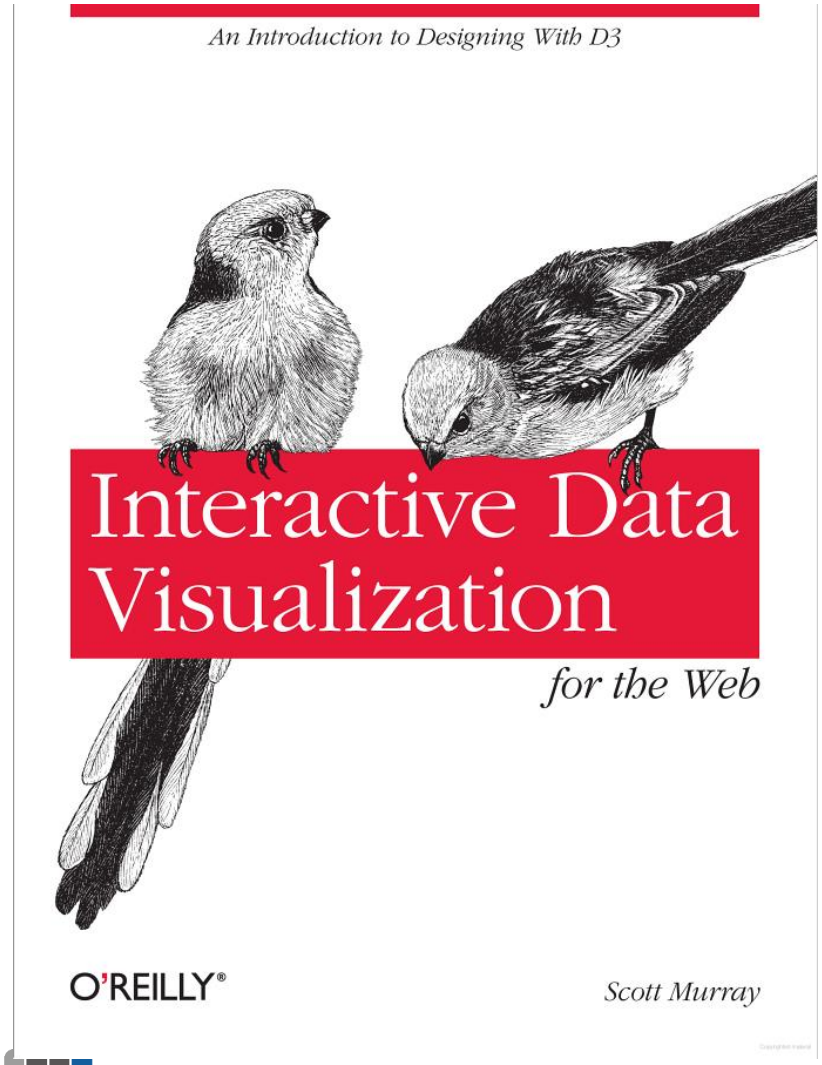
https://github.com/HBPVIS/VIOLA

http://hbpvis.github.io/VIOLA/



Tools and Libraries

Data Driven Documents – D3.js



- The following content is based on the Book „Interactive Data Visualization for the Web“ written by Scott Murray

and

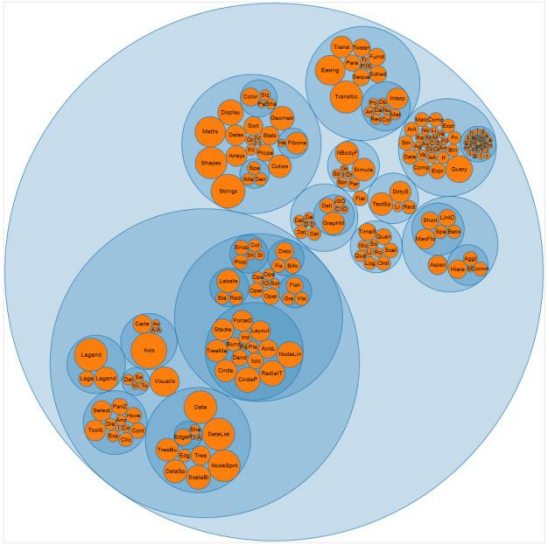
- his tutorial to find on *http://alignedleft.com/tutorials/d3*

- The home of D3 is http://d3js.org/

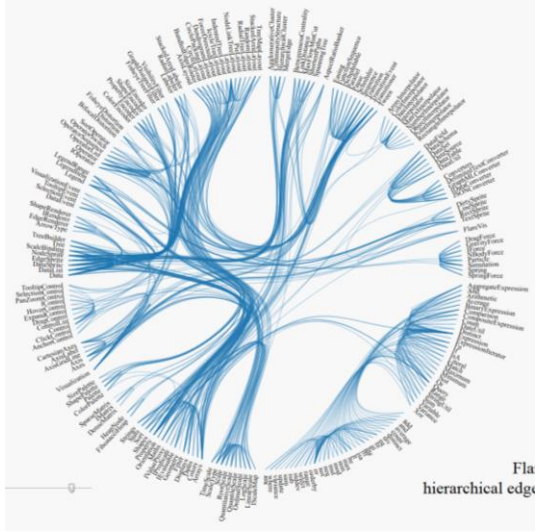
and

https://github.com/mbostock/d3

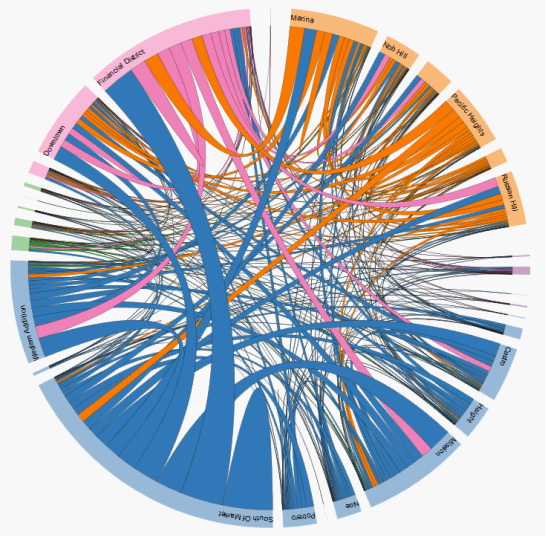
Add-On Libs



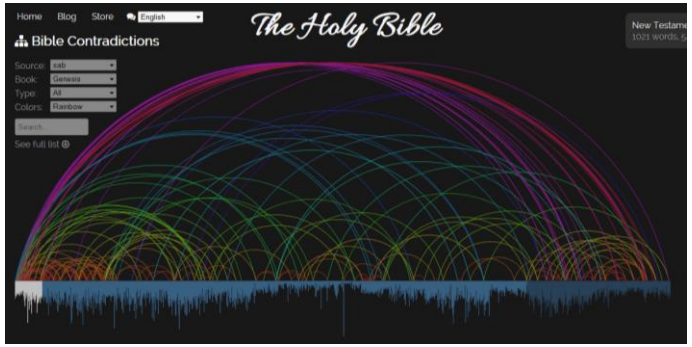
Circle Packing

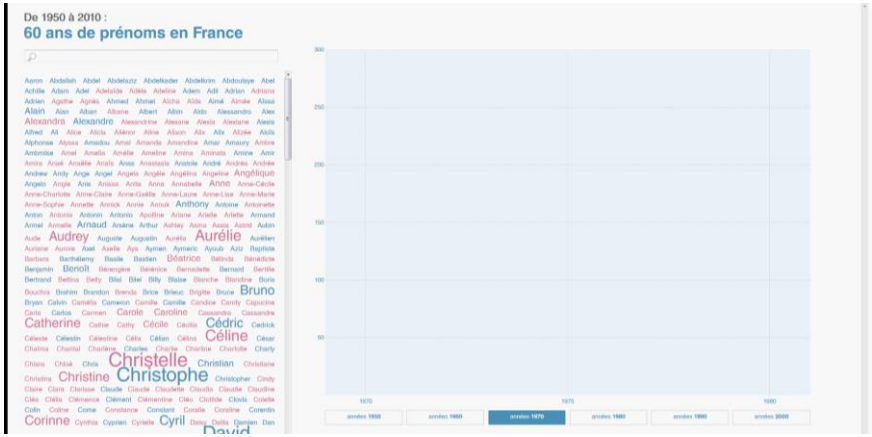


Hierarchical Edge Bundling



Chord Diagram

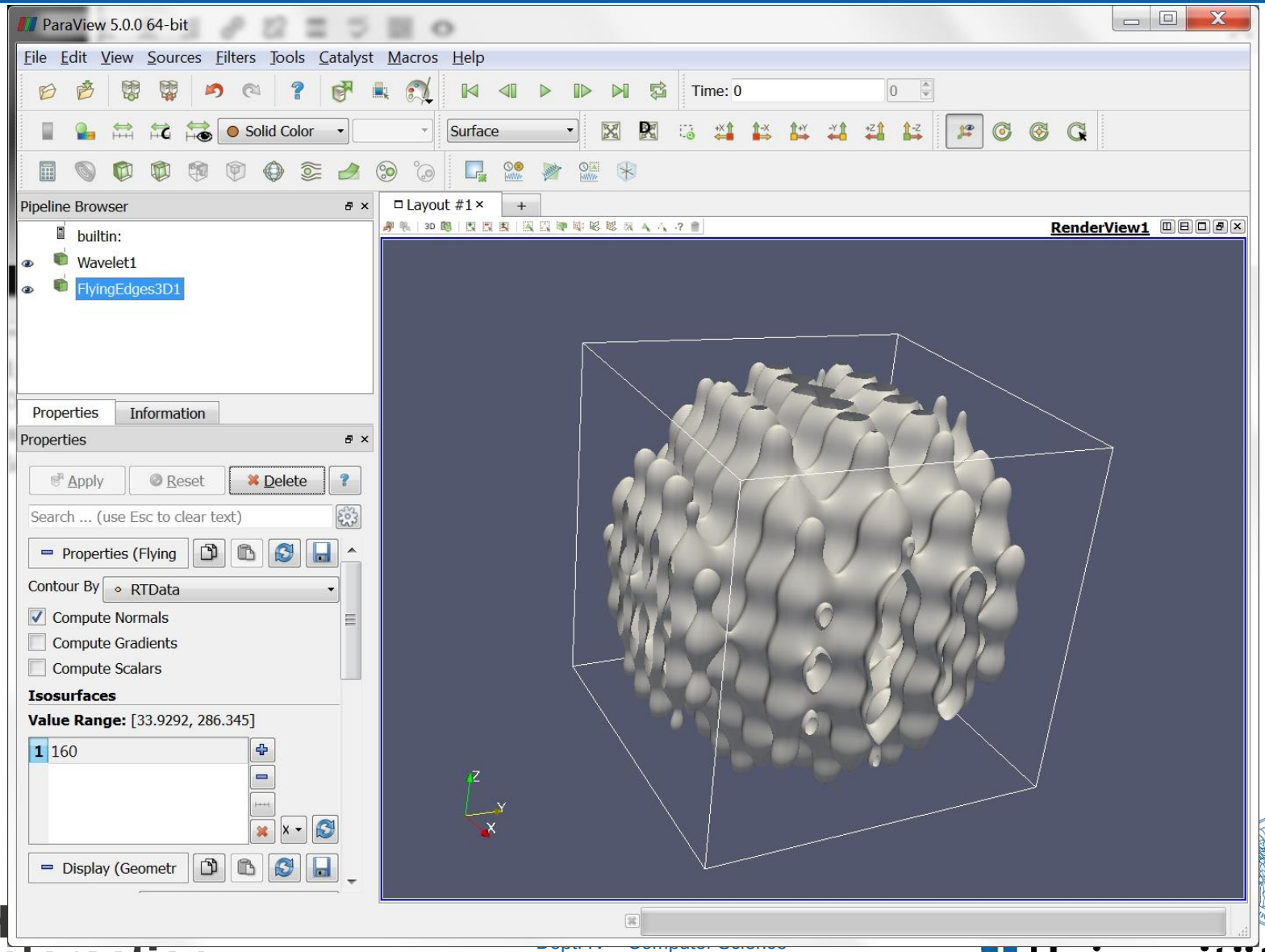




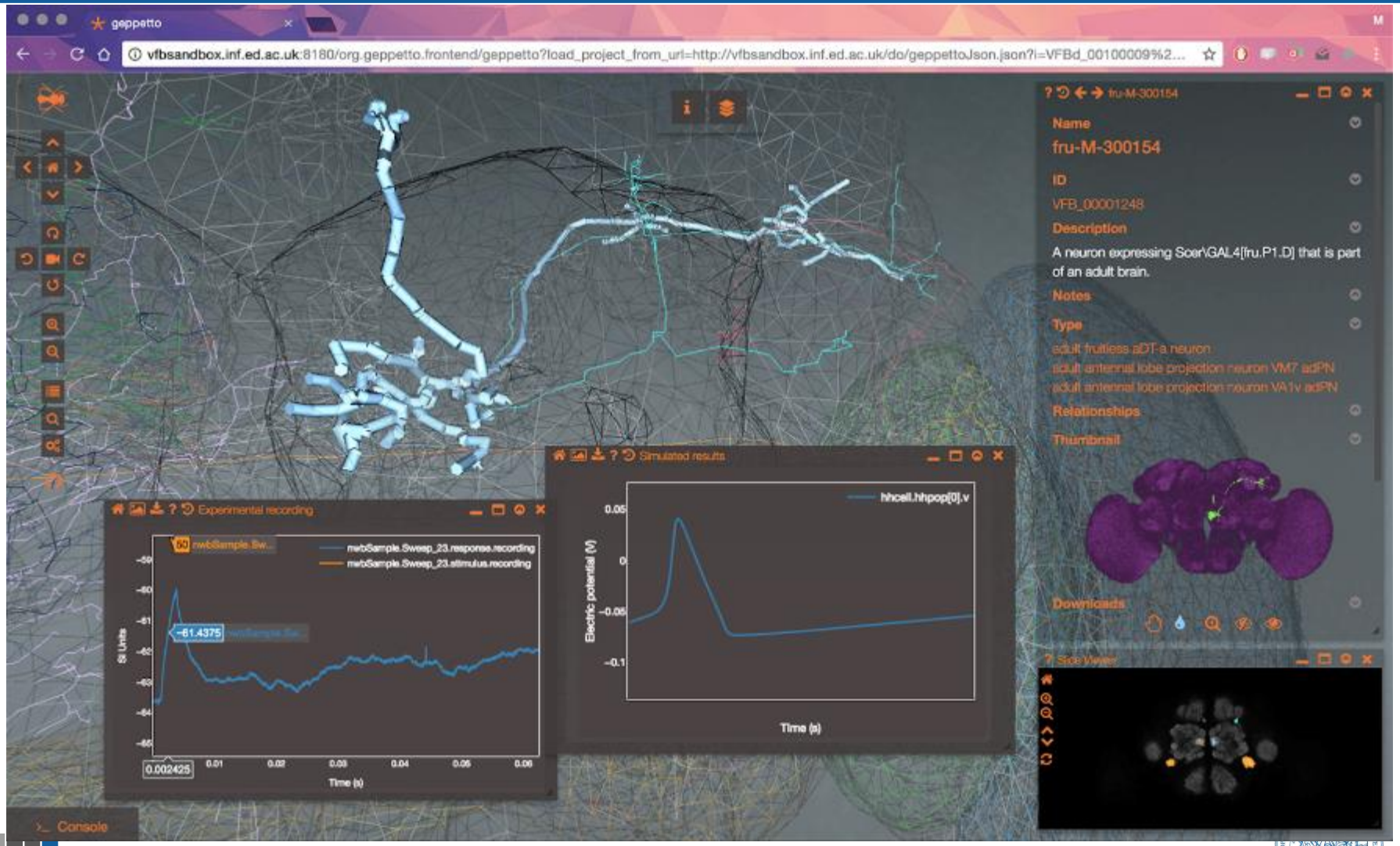
Interactive Bible Contradictions & Bar Charts

Interactive Function Plots

ParaView


ParaView 5.0.0 64-bit
File Edit View Sources Filters Tools Catalyst Macros Help
Time: 0
Solid Color
Surface
Pipeline Browser
builtin:
Wavelet1
FlyingEdges3D1
Layout #1
RenderView1
Properties
Information
Apply
Reset
Delete
Search ... (use Esc to clear text)
Properties (Flying
Contour By
RTData
Compute Normals
Compute Gradients
Compute Scalars
Isosurfaces
Value Range: [33.9292, 286.345]
1 160
Display (Geometr

Geppetto


geppetto
vfbsandbox.inf.ed.ac.uk:8180/org.geppetto.frontend/geppetto?load_project_from_url=http://vfbsandbox.inf.ed.ac.uk/do/geppettoJson.json?i=VFBd_00100009%2...
fru-M-300154
Name
fru-M-300154
ID
VFB_00001248
Description
A neuron expressing Soer\GAL4[fru.P1.D] that is part of an adult brain.
Notes
Type
adult fruitless aDT-a neuron
adult antennal lobe projection neuron VM7 adPN
adult antennal lobe projection neuron VA1v adPN
Relationships
Thumbnail
Downloads
Experimental recording
mwbSample.Sweep_23.response.recording
mwbSample.Sweep_23.stimulus.recording
-61.4375
0.002425
Time (s)
Simulated results
hhcell.hhpop[0].v
Electric potential (V)
Time (s)
Side Viewer
Console

Visual Integrity

Graphical Integrity

- Graphical integrity specifies the aspect in how far graphical representations lie about the underlying data

- Tufte specifies a „lie factor“ as given below

$$\textbf{lie factor} = \frac{\textbf{size of effect shown in grahpic}}{\textbf{size of effect in data}}$$

- If the Lie Factor is 1, the graphic is valid

Lie Factor – Example

- The graphic was intended to present a regulation by the US Dept. of Transportation defining an increase of miles per gallon a car has to meet until 1985 from 18.0 miles per gallon in 1978 to 27.5 in 1985

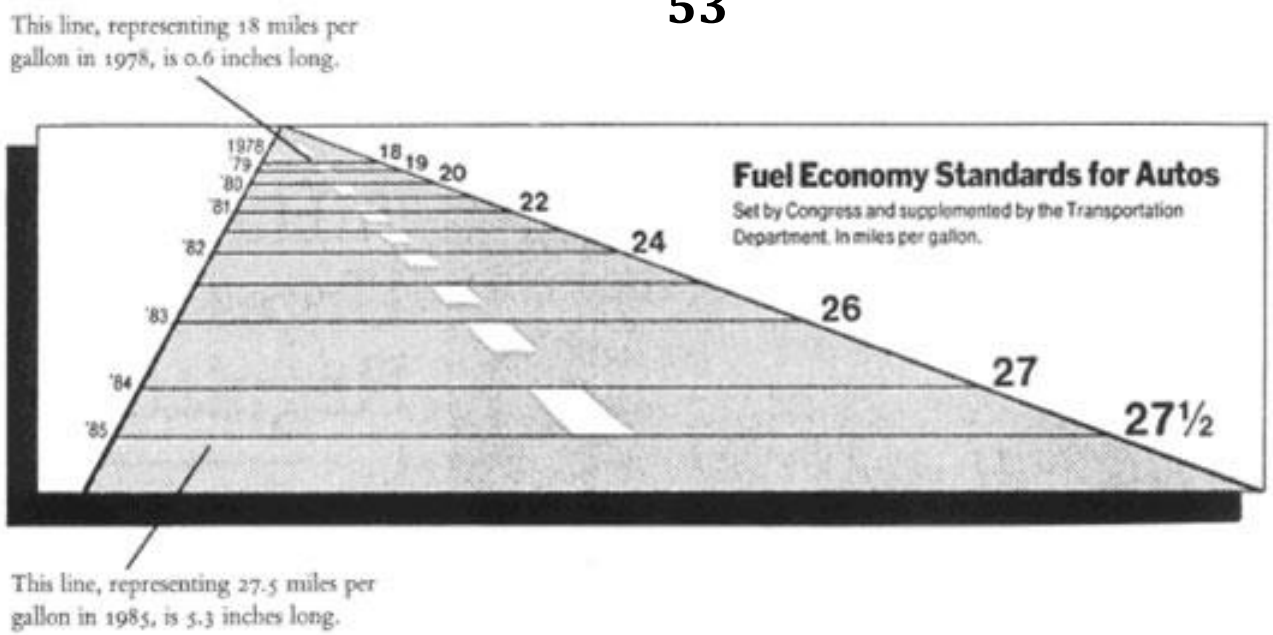
- Real increase is:

$$\frac{27.5 - 18.0}{18.0} \times 100 = \mathbf{53}\%$$

Shown is (relative lengths of lines):

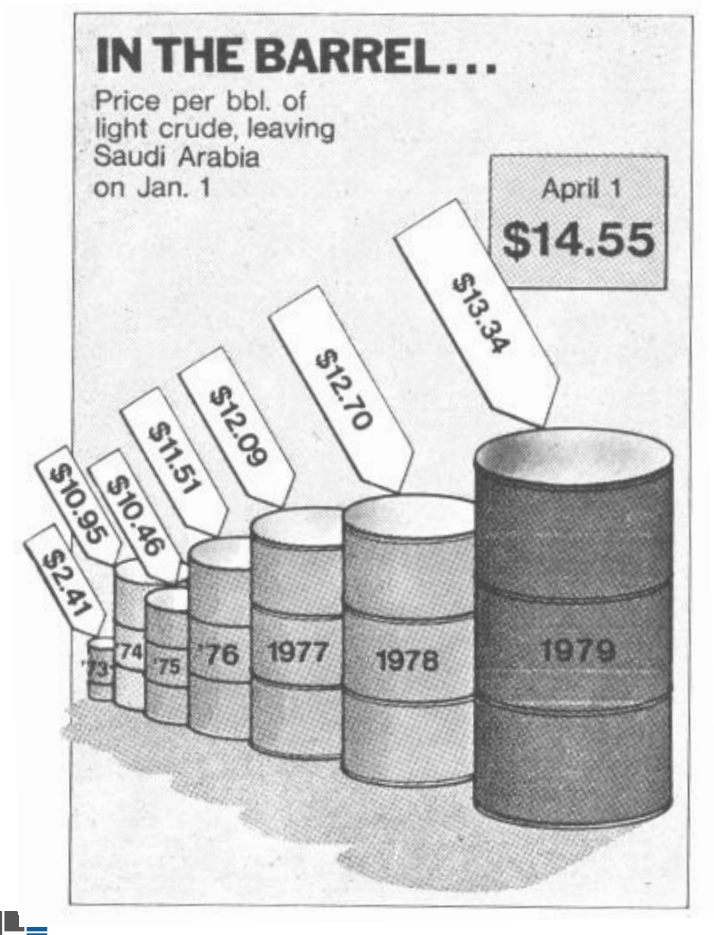
$$\frac{5.3 - 0.6}{0.6} \times 100 = \mathbf{783}\%$$

$$\boldsymbol{Lie\ Factor} = \frac{\mathbf{783}}{\mathbf{53}} = \mathbf{14.8}$$

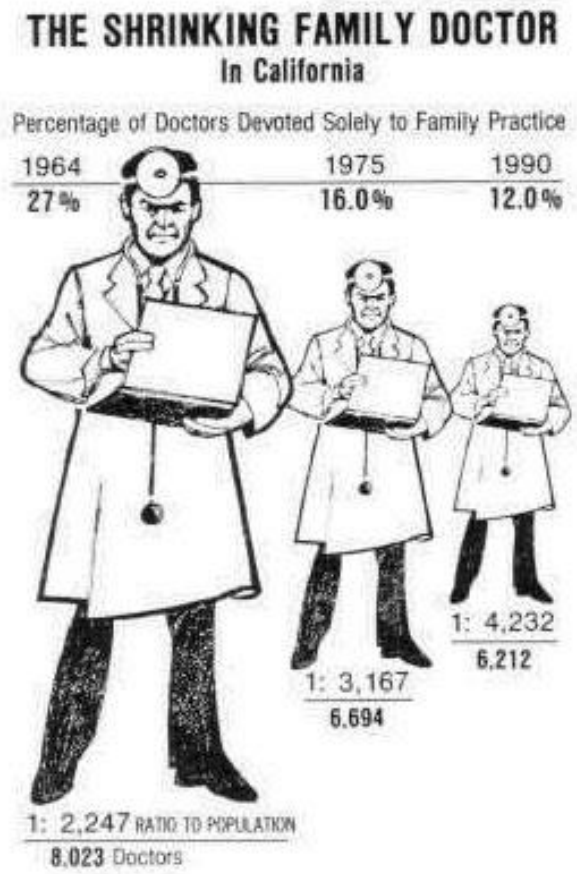


Lie Factor – Examples

- Lie Factor of 9.4:



- Lie Factor of 2.8:



Forthnightly Video Conference on Visualization

- Send your questions, topics, or requests to weyers@uni-trier.de
- They will move into the agenda and discussed or answered directly
- Search the Collaboratory for the „Visualization Video Conference“ collab
- Takes place each 1st and 3rd Thursday
- 9-10am
- Is announced over Twitter!


1 minute paper

Contactinformation

The video conference will be held online. Connection details can be found below.
Telephone Number: 004910097956411 Conference Number: 97956411
In order to enter the conference use one of the following numbers/addresses:
1) H.323 dial-in using a local gatekeeper (Global Dialling Scheme, GDS) - dial 004910097956411
2) H.323 dial-in using DFN gatekeeper - configure your email address as H.323 alias and your phone number as E.164 number - configure gk.vc.dfn.de as the gatekeeper - dial 004910097956411
3) H.323 dial-in without any gatekeeper (URI dialling) - for Tandberg, Mirial dial 97956411@vc.dfn.de - for Polycom, Lifesize dial vc.dfn.de##97956411 Additional information about URI dialling and other end systems can be found at: http://vcc.zih.tu-dresden.de/index.php?linkid=11100&lang=en
4) SIP dial-in (For clients such as Cisco SX10 or Polycom Real Presence Mobile Client at Android or iOS) - dial 97956411@vc.dfn.de Registration to a SIP proxy is not required. If your endpoint supports BFCP protocol, you can use the second video stream for presentations.
5) Telephone/Mobile (POTS/ISDN) dial-in: - dial +49-30-20097956411 You will be directed to the conference.
You can test all of this anytime as the only participant to the conference to see if your VC system is working well with the MCUs. If you need a peer for testing two-way video and audio with more than your own signals give us a shout and we can schedule a common test slot some days before the meeting for anybody who wants to test that thoroughly in advance. For troubleshooting please contact DFNVC Hotline at +49-711-63314-214.