Information Visualization - Introduction

Eduard Gröller, Manuela Waldner

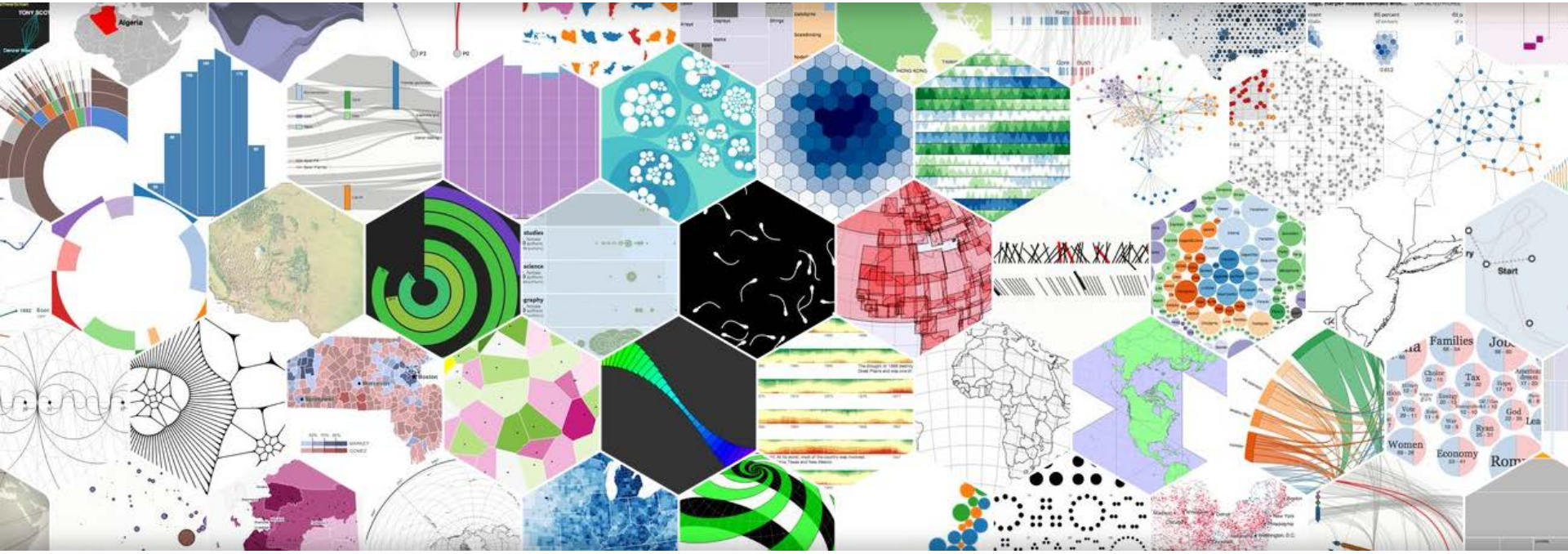
Institute of Computer Graphics and Algorithms

Vienna University of Technology

Information Visualization

"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition"

[Card et al., Readings in Information Visualization: Using Vision to Think, 1999]



[http://d3js.org/]

Outline

- Introduction
- Knowledge crystallization
- InfoVis reference model
  - Visual mappings, visual structures
  - View transformations
  - Interaction

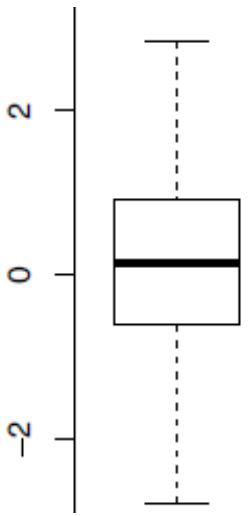
Why visualize?

[Munzner, 2014]

Exploration **Confirmation** **Presentation**

Exploratory Data Analysis

- Starting point:
  - No hypothesis about the data
- Process:
  - Searching and analyzing data to find potentially useful information
- Result:
  - Hypotheses extracted from data
- Introduced by statistician **John Tukey** (1915-2000)
  - Invented box plots



[Munzner, 2014]
[Wickham and Stryjewski, 40 years of boxplots, 2012]

Confirmatory Analysis

- Starting point:
  - one or more hypotheses about the data
- Process:
  - goal-oriented examination of these hypotheses
- Result:
  - confirmation or rejection of hypotheses

Presentation

- Starting point:
  - Facts to be presented (fixed a priori)
- Goal: efficiently and effectively communicate results of analysis

„In the last 30 years, about 80 percent of four-year forecasts have been too optimistic.“

The New York Times
www.nytimes.com

Published: February 2, 2010

Budget Forecasts, Compared With Reality

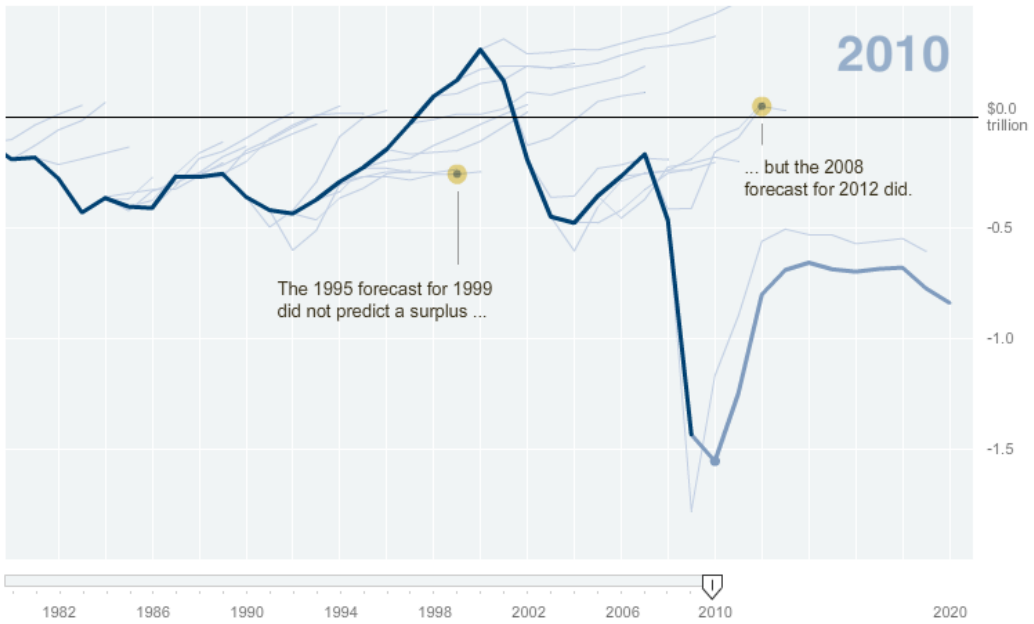
Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

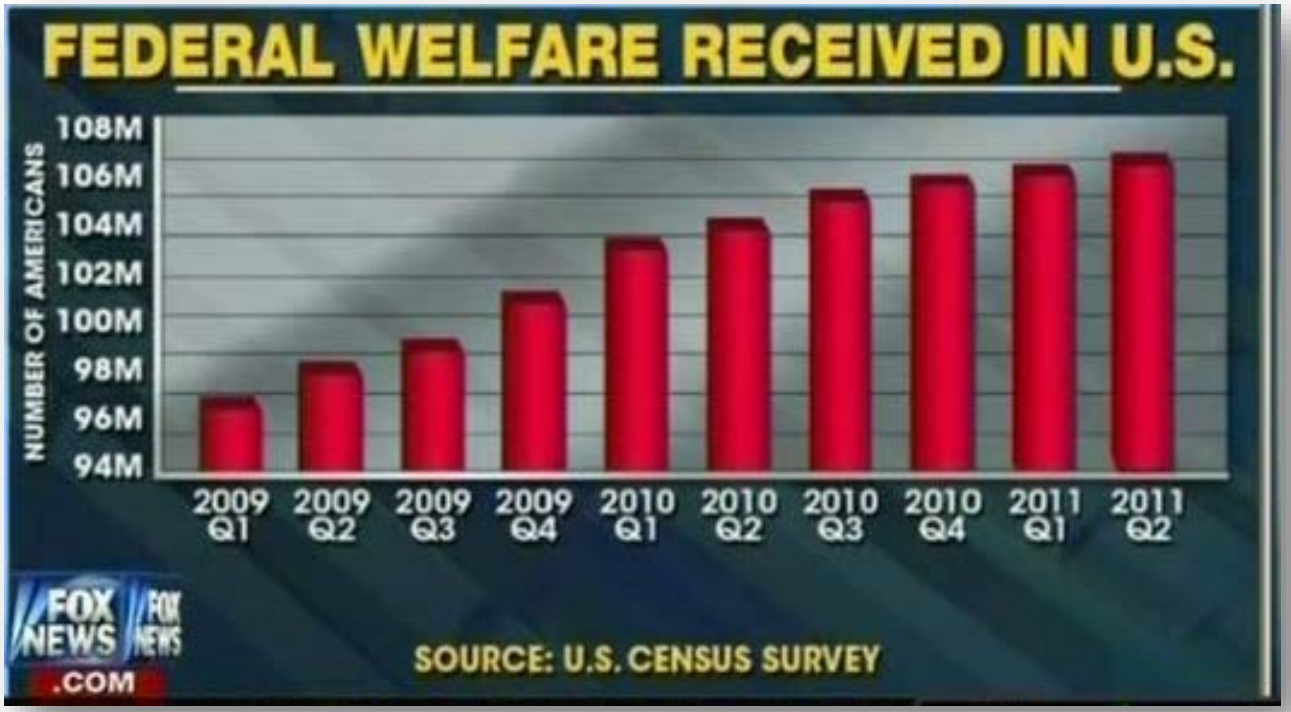
1 2 3 4 5 6 NEXT

Past forecasts

Even that may be an understatement. In the last 30 years, about 80 percent of four-year deficit forecasts have been too optimistic.

The early Clinton budgets — which failed to predict the surpluses that were generated, in part, by a stock market bubble — are the only major exception.





"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented."

[Edward Tufte, The Visual Display of Quantitative Information, Second Edition, Graphics Press, USA, 1991]

Information Visualization (InfoVis)

External Cognition

use external world to accomplish cognition

Information Design

design external representations to amplify cognition

Visualization

computer-based, interactive

Scientific Visualization

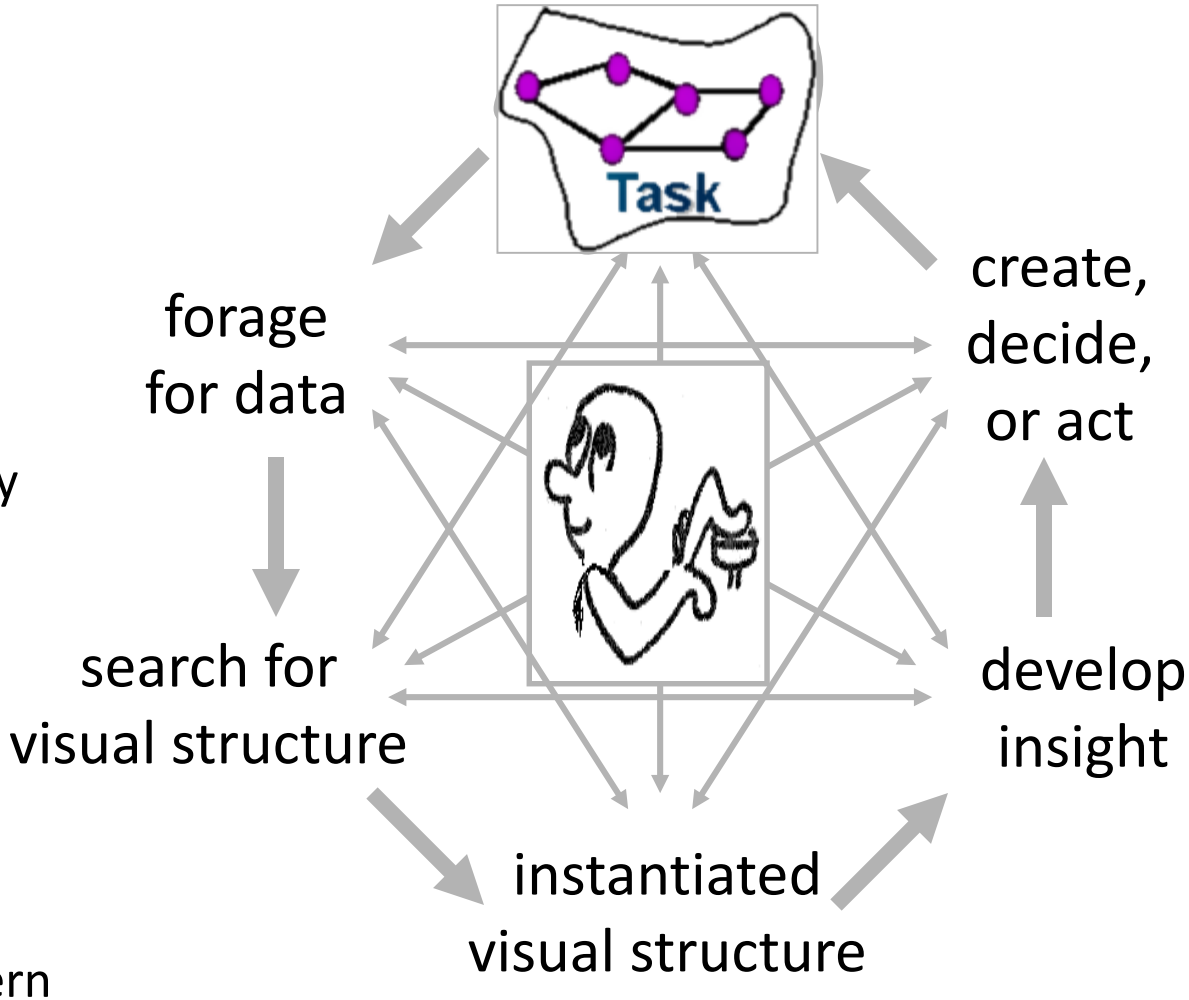
typically physical data

Information Visualization

abstract, nonphysical data

Courtesy of Jock Mackinlay

Knowledge Crystallization



Overview
Zoom
Filter
Details
Browse
Search query

Reorder
Cluster
Class
Average
Promote
Detect pattern
Abstract

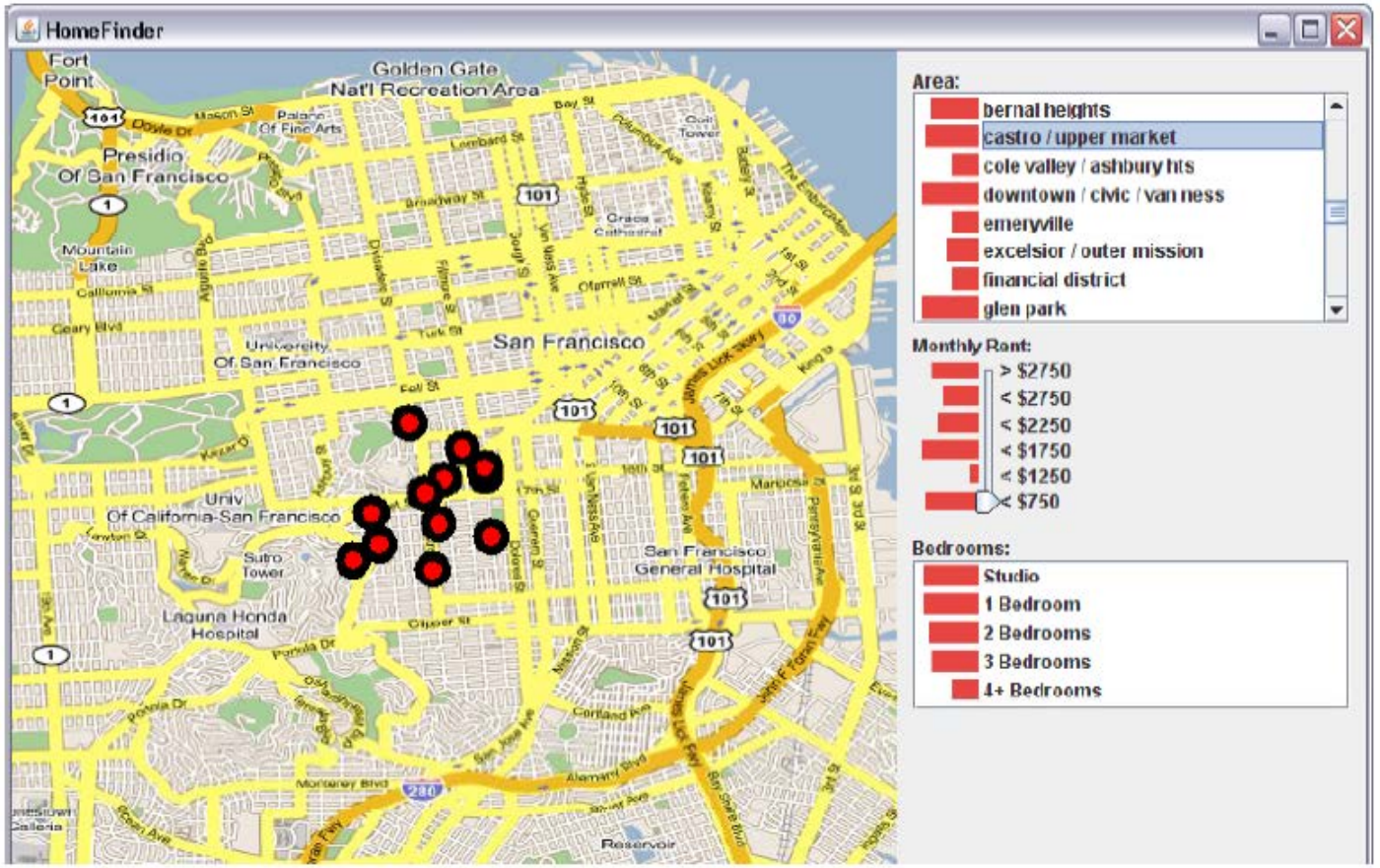
Extract
Compose
Present

Create
Delete
Manipulate
Read fact
Read pattern
Read compare

Courtesy of Jock Mackinlay

HomeFinder

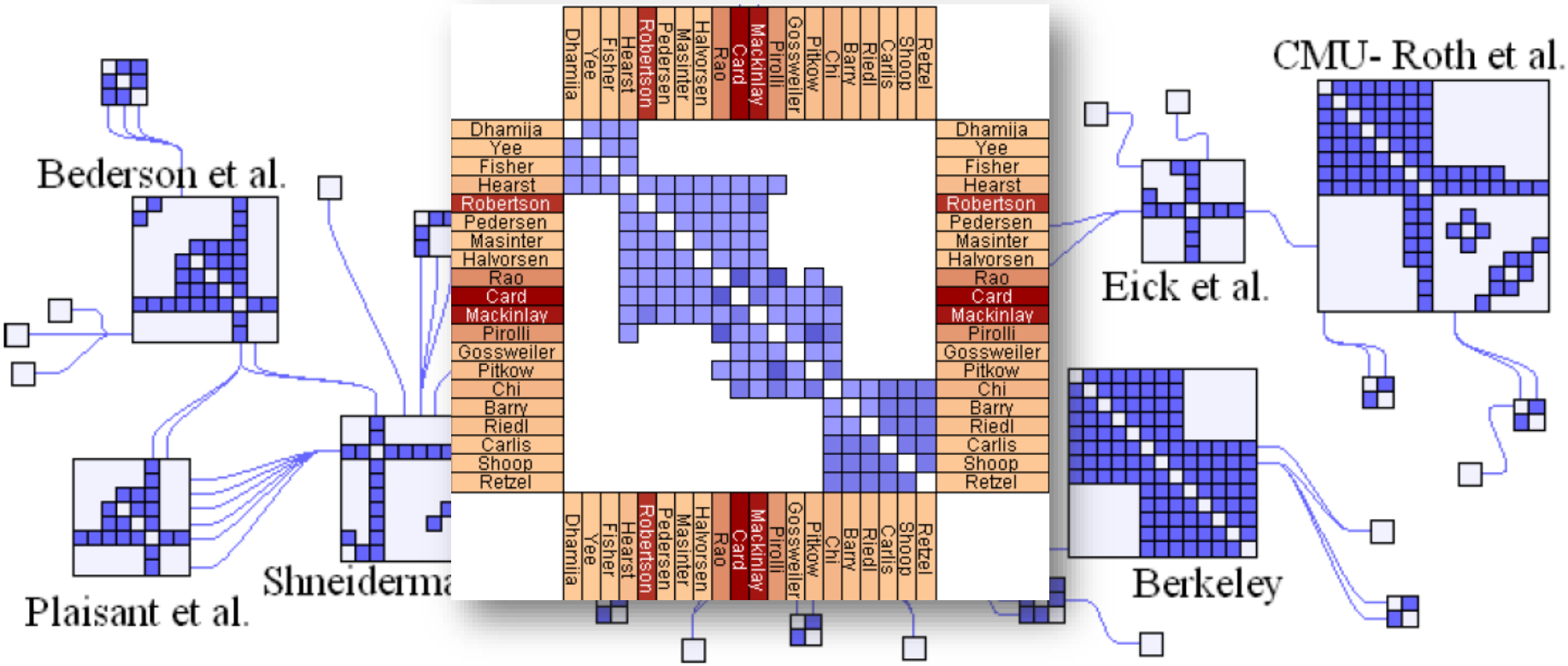
- Browsing housing market
- Data, schema (structure), task



[Willett et al., Scented Widgets, TVCG 2007]

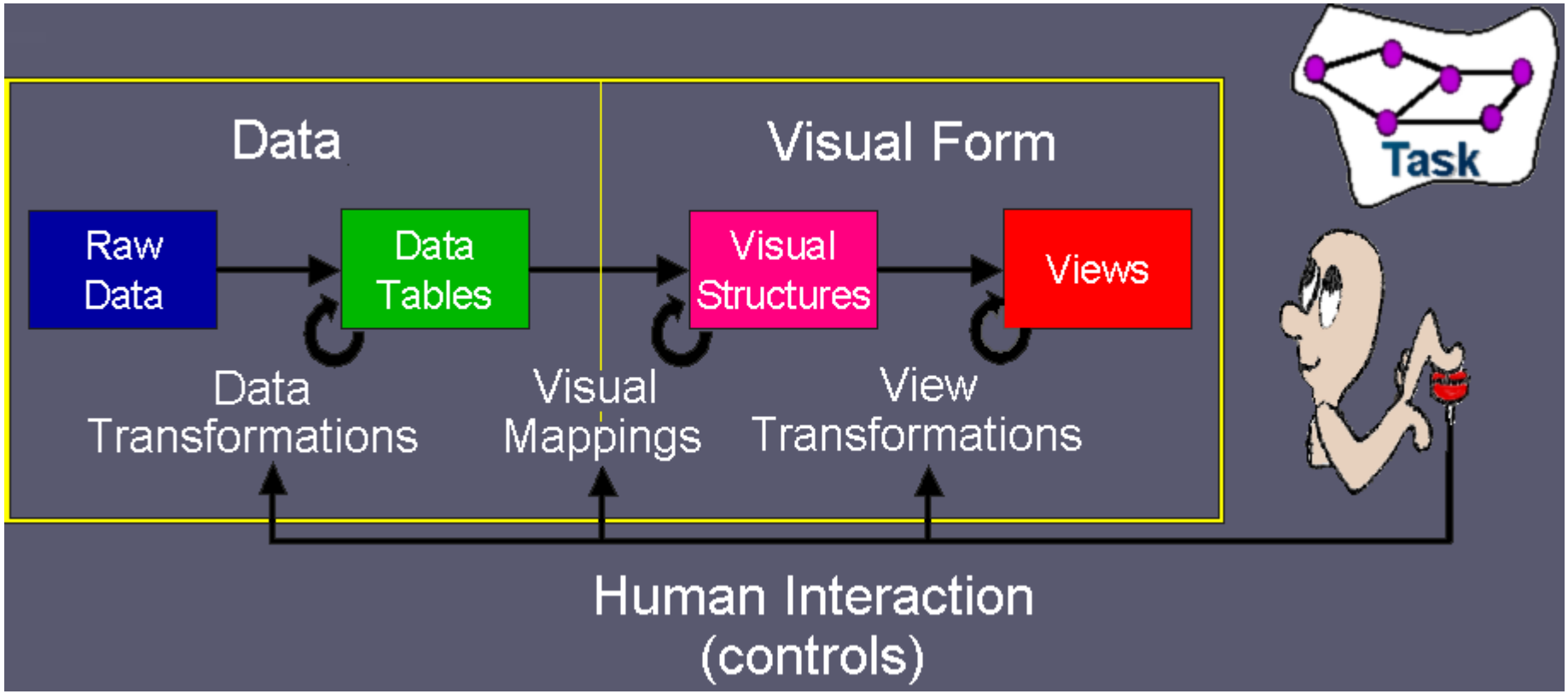
NodeTrix

- Large social network visualization
- Aggregate / split, order, merge matrices



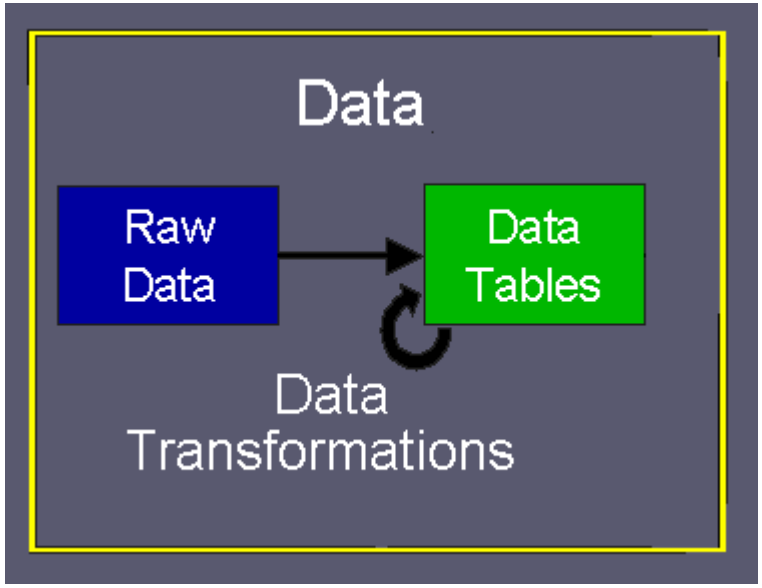
[Henry et al., NodeTrix, TVCG 2007]

InfoVis Reference Model

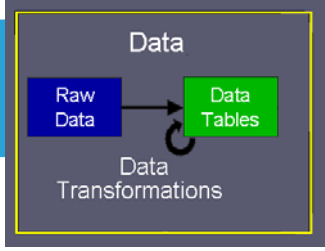


- Raw Data: idiosyncratic formats
- Data Tables: relations(cases by variables)+metadata
- Visual Structures: spatial substrates + marks + graphical properties
- Views: graphical parameters (position, scaling, clipping, zooming,...)

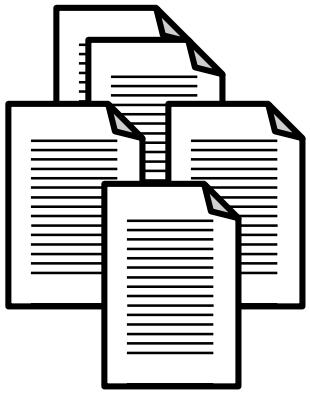
Data



Raw Data



Documents



Words

aardvark
billion area book
anode
answer bay
arrow
about
absent
are anonymous
Aarhus

Word Vectors

Document	D1	D2	D3	…
aardvark	1	0	0	…
Aarhus	0	1	0	…
about	1	0	1	…
…	…	…	…	…

Other units

- Sentence
- Paragraph
- Section
- Chapter
- Characters
- Pictures

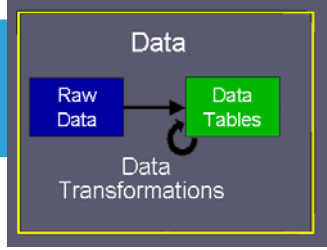
Meta-data

Document	D1	D2	D3	…
Length	4	3	6	…
Author	John	Sally	Lars	…
Date	16/8	11/4	24/7	…
…	…	…	…	…

Meaning

Jock Mackinlay's Slide

Raw Data Issues



- Errors
- Variable formats
- Missing data
- Variable types
- Table Structure

Document	D1	A	D3	…
Length	4	3.5	6	…
Author	John		Lars	…
Date	16/8	Fall	24/7	…
…	…	…	…	…

Document	D1	D2	D3	…
TUWIEN	1	0	0	…
UNIWIEN	0	1	0	…
about	1	0	1	…
…	…	…	…	…

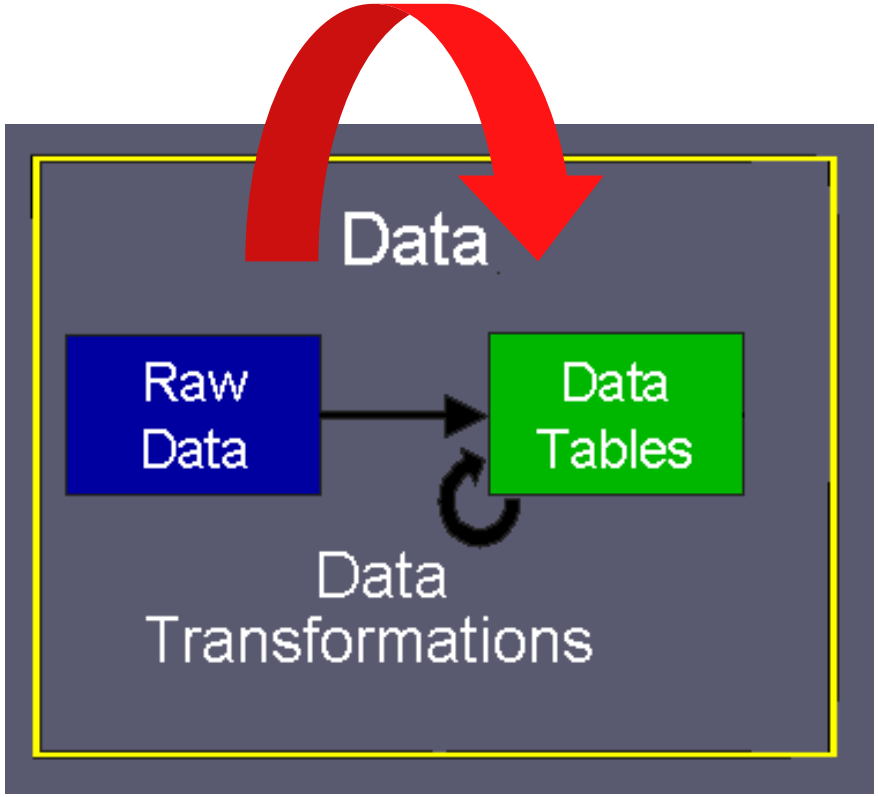
VS

TUWIEN	D1,...
UNIWIEN	D2,…
about	D1, D3, …
…	…

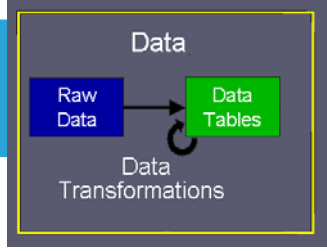
Courtesy of Jock Mackinlay

Data Transformations

- Process of converting Raw Data into Data Tables.
- Used to build and improve Data Tables



Data Tables



- Data Tables:
  - Cases / Items
  - Variables / Attributes
    - Nominal
    - Quantitative
    - Ordinal
  - Values
  - Metadata

Hans
46
ID-22222

Anna
17
ID-11111

Peter
15
ID-33333

attribute

Name	N	Anna	Hans	Peter
Age	Q	17	46	15
ID	O	11111	22222	33333

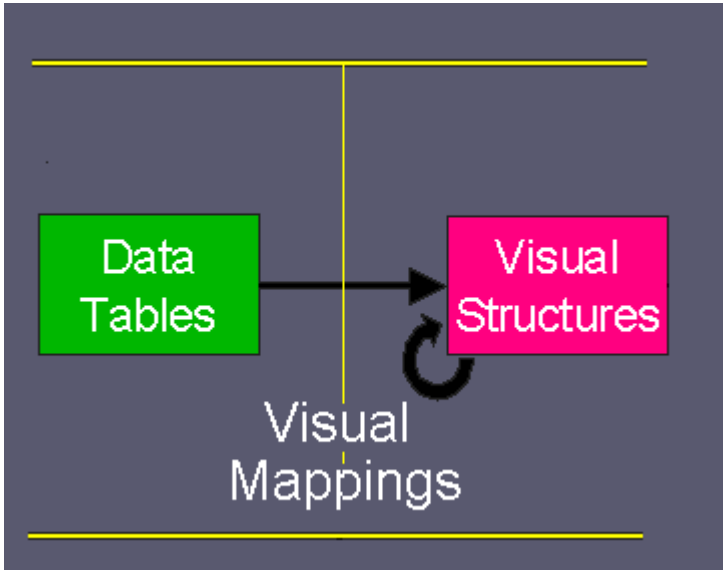
item

Data Transformations

- Values → Derived Values
- Structure → Derived Structure
- Values → Derived Structure
- Structure → Derived Values

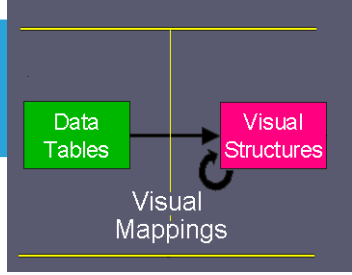
	Derived value	Derived structure
Value	Mean	Sort Class Promote
Structure	Demote	X,Y,Z→P xzy

Visual Mappings

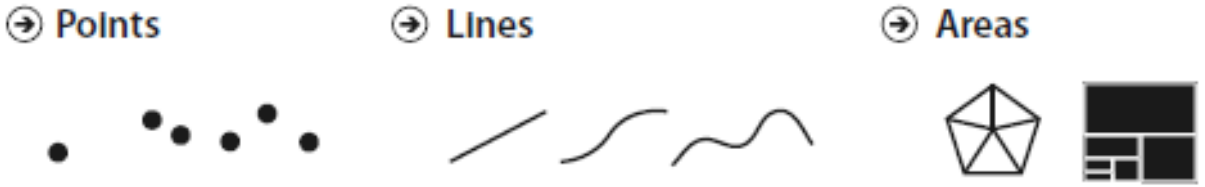


- Expressiveness
- Effectiveness

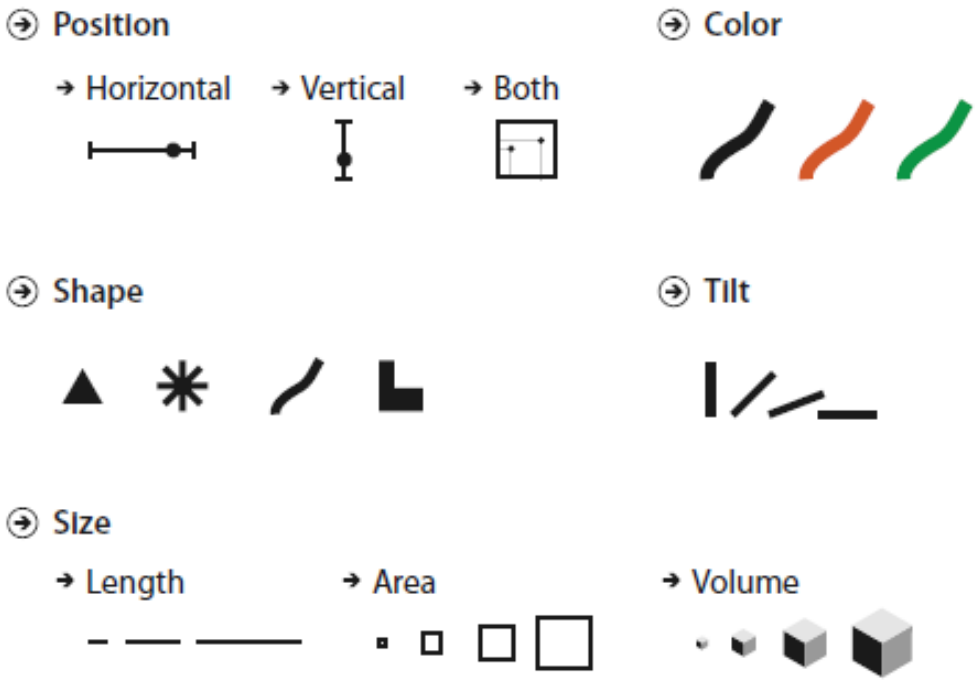
Marks and Channels



- Building blocks for visual encodings
- Marks:

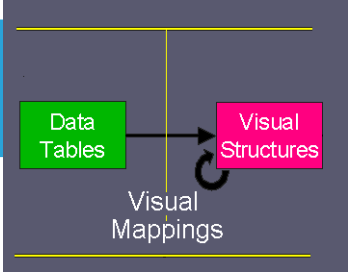


- Visual channels control appearance of marks



[Munzner, 2014]

Channel Effectiveness



Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Most ▲ Effectiveness ▼ Least

Same (Color luminance, Color saturation)

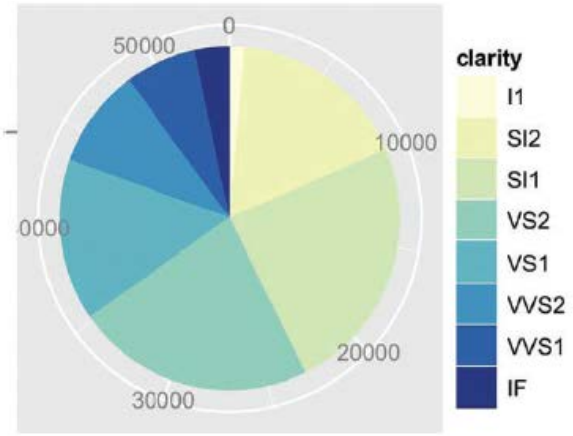
Same (Curvature, Volume (3D size))

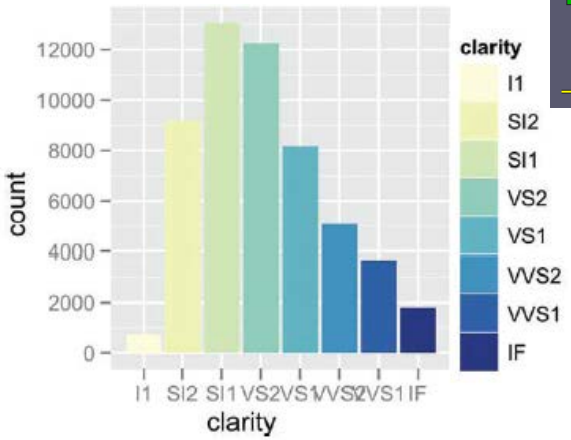
Identity Channels: Categorical Attributes

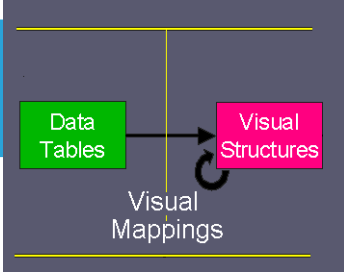
- Spatial region
- Color hue
- Motion
- Shape

[Munzner, 2014]

Channel Effectiveness: Example





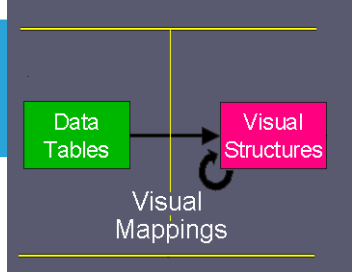


[Munzner, 2014]
[Wickham, A Layered Grammar of Graphics, 2010

- Pie Chart
  - Visual mark: area
  - Attribute 1: color (categorical)
  - Attribute 2: angle (quantitative)

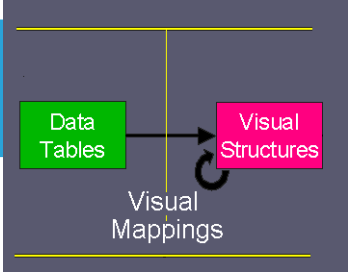
- Bar Chart
  - Visual mark: line
  - Attribute 1: horizontal position + color (categorical)
  - Attribute 2: vertical position / length (quantitative)

Visual Encoding

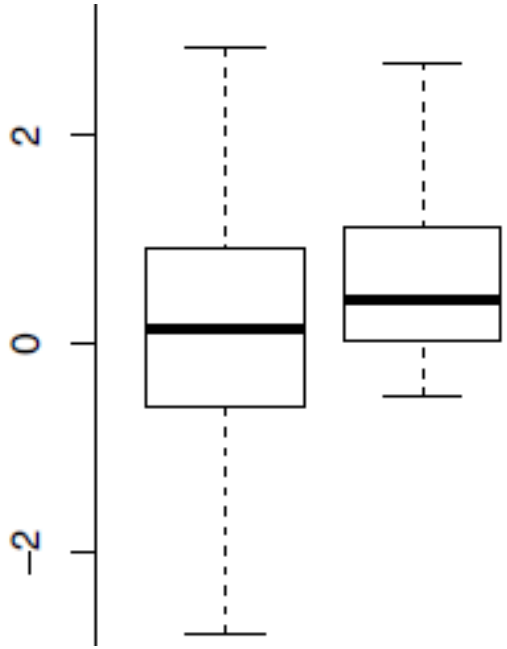


- Classification by data
  - Number of dependent variables / values:
    - Univariate / bivariate / multivariate data
  - Number of independent variables / keys:
    - One-dimensional... multidimensional data
  - Sets
  - Networks
  - Trees
  - Text:
    - documents / corpus / streams

Univariate Data



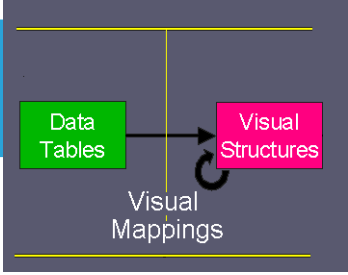
- 1 dependent variable



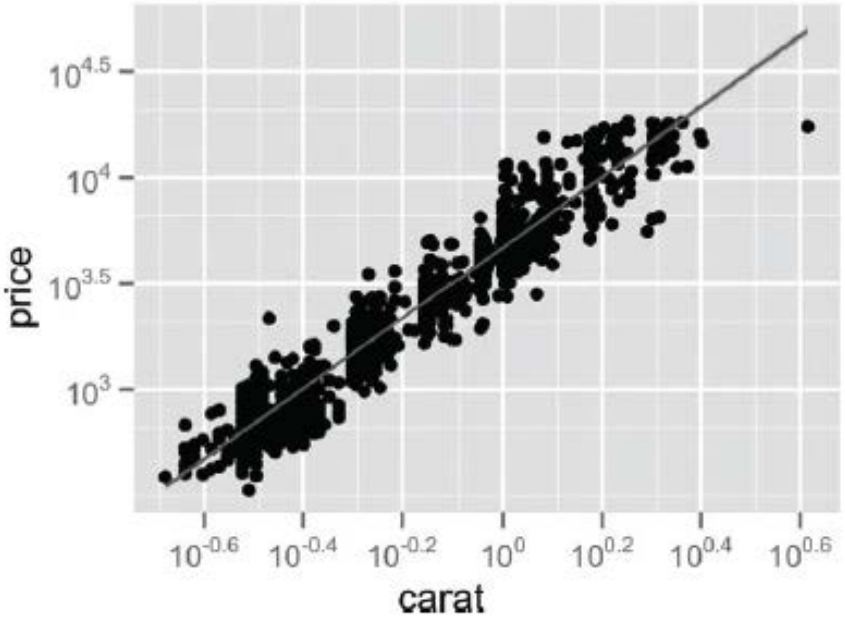
[Munzner, 2014]
[Wickham and Stryjewski, 40 years of boxplots, 2012]

- **Box plot**
  - Quantitative value attributes
  - Median, lower and upper quartiles, fences

Bivariate Data



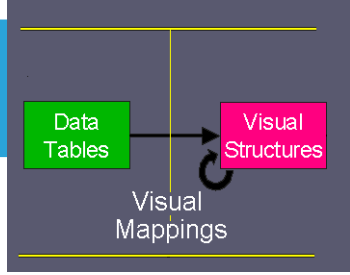
- 2 dependent variables



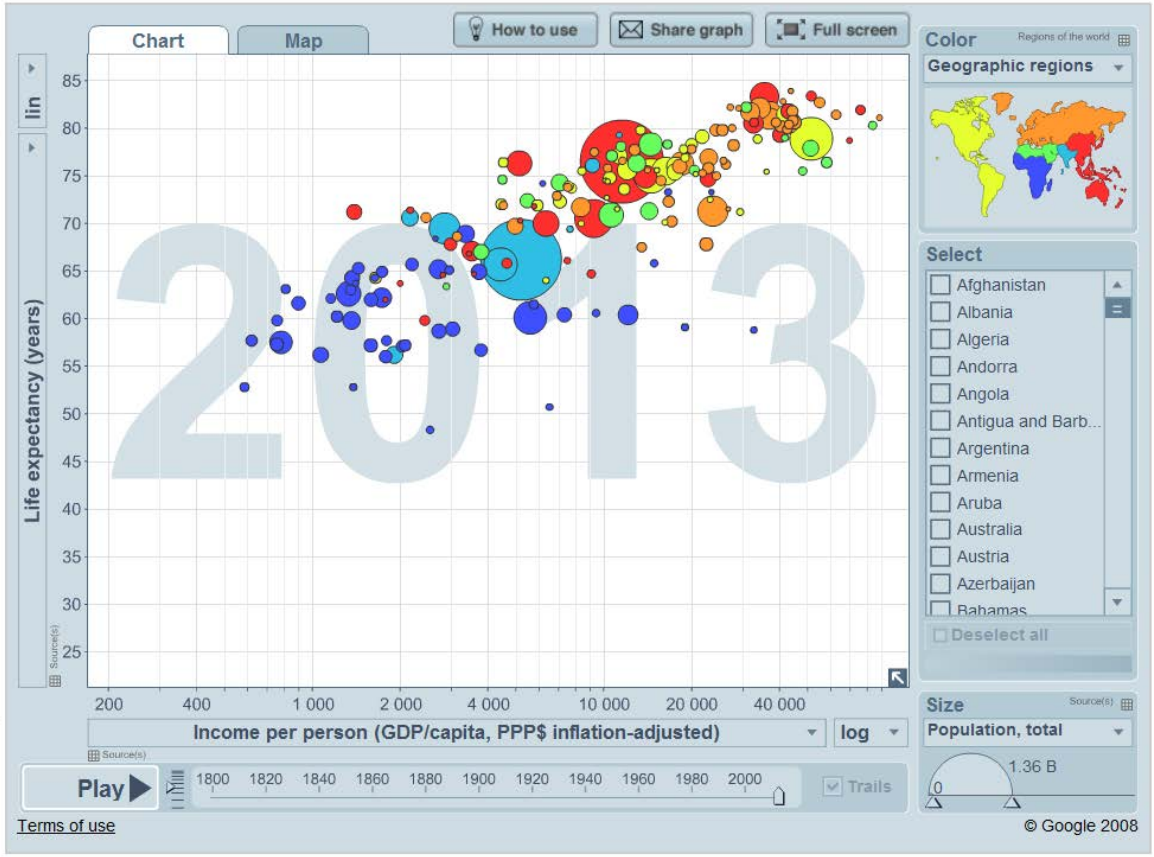
[Munzner, 2014]
[Wickham, A Layered Grammar of Graphics, 2010

- **Scatterplot**
  - 2 value attributes (quantitative)
  - characterizing distributions, finding outliers, correlations, extremes, clusters

Multivariate Data

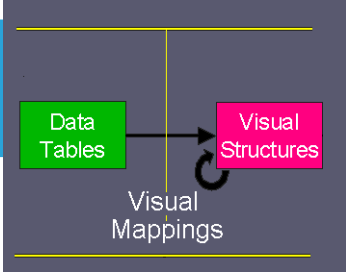


- N dependent variables
- **Scatterplot / „bubble chart“**
  - Additional quantitative attribute: size
  - Additional categorical attribute: color

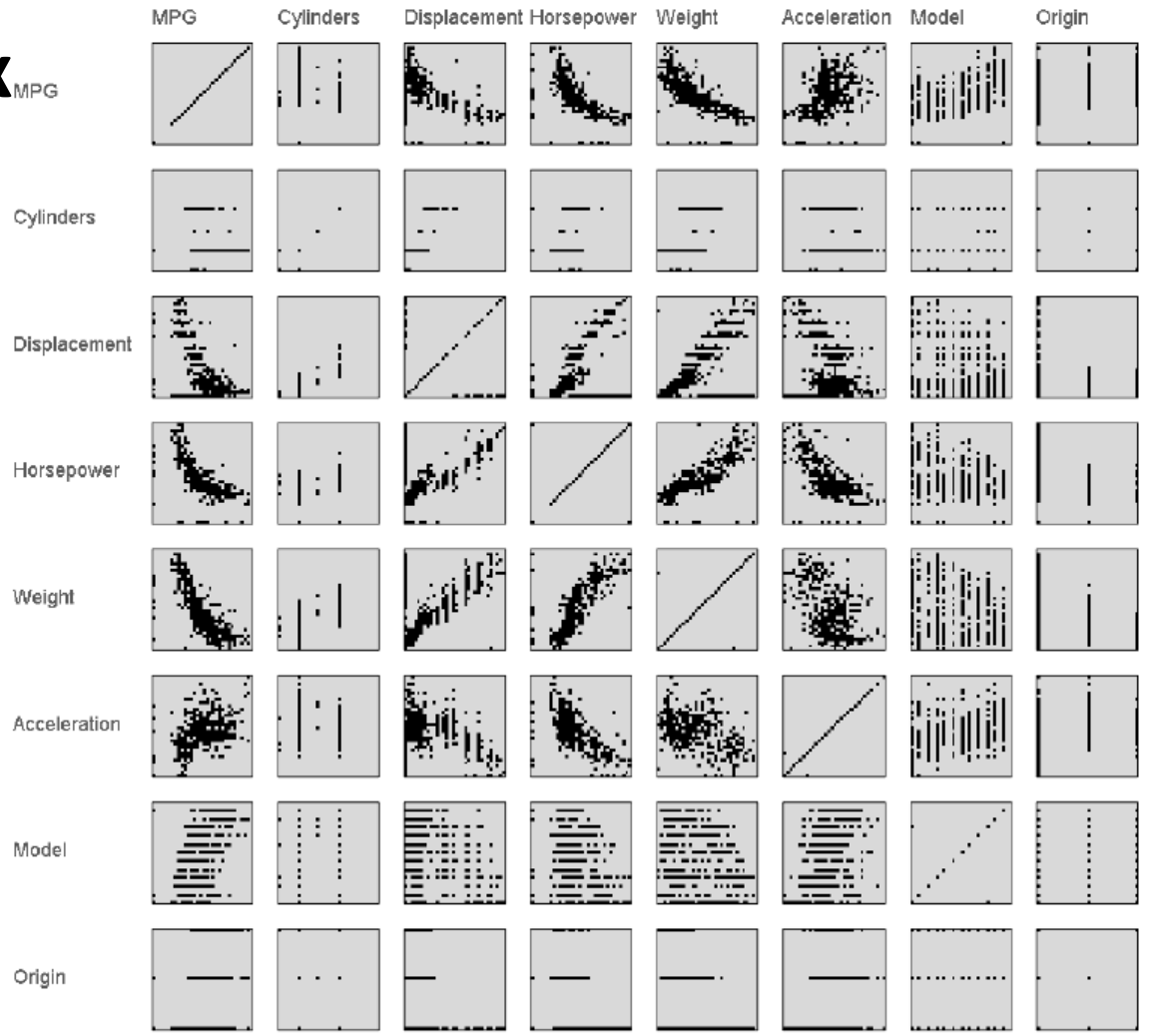


http://www.gapminder.org/

Multivariate Data

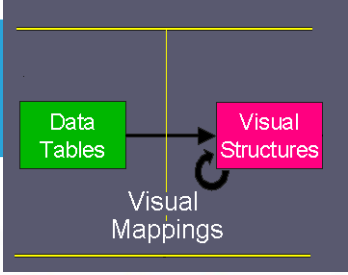


- N dependent variables
- **Scatterplot matrix (SPLOM)**
  - Rows and columns are all attributes
  - Each matrix cell contains a scatterplot

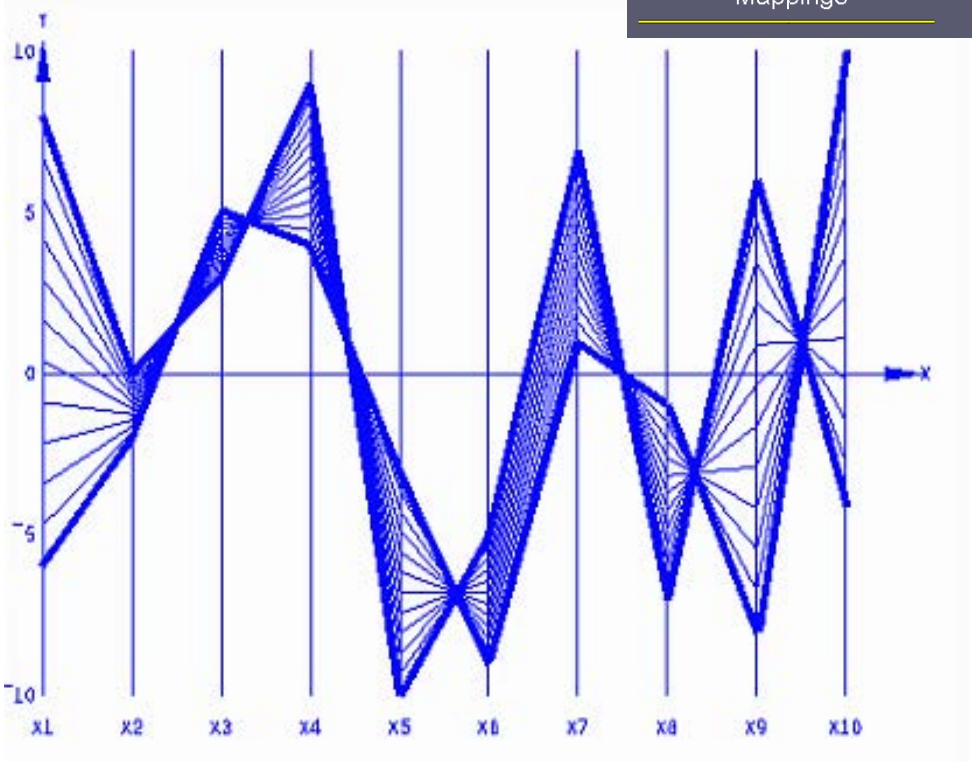


[Elmqvist et al., TCVG 2008]

Multivariate Data

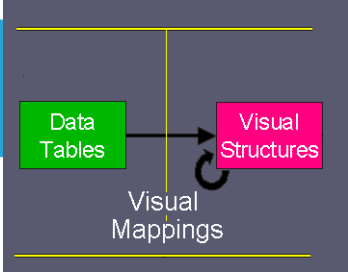


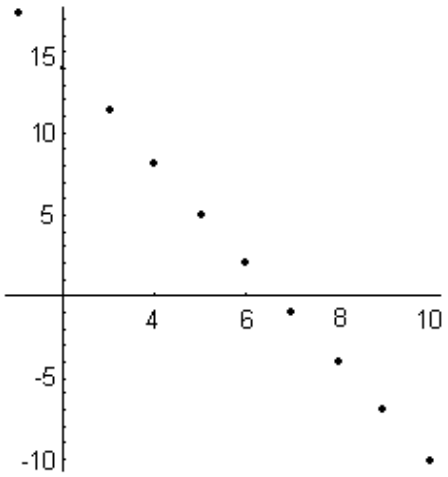
- N dependent variables



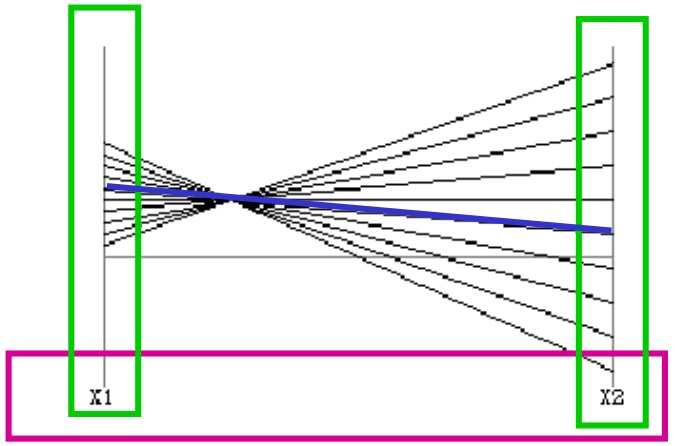
- **Parallel coordinates**
  - Parallel 2D axes
  - Add/Remove data
    - Establish patterns
    - Examine interactions
  - Useful for recognizing patterns between the axes
  - Skilled user

Parallel Coordinates





Dataset in a Cartesian graph



Same dataset in parallel coordinates

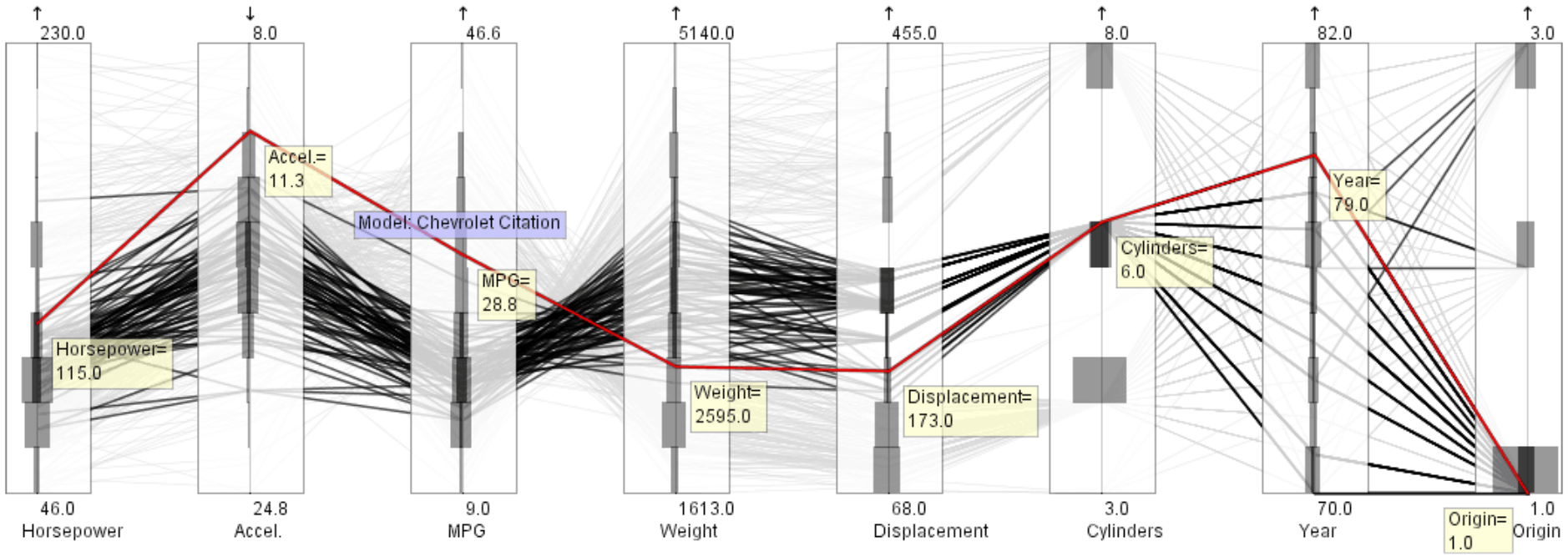
[Inselberg]

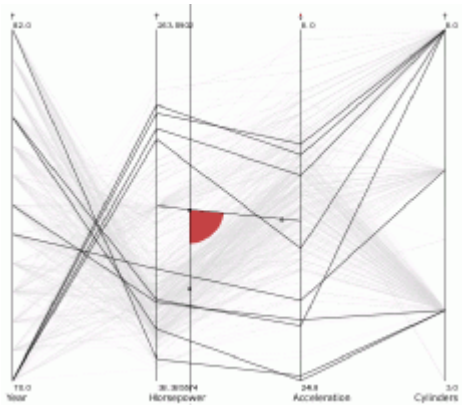
Encode variables along a horizontal row

Vertical line specifies single variable

Blue line specifies a case

Parallel Coordinates



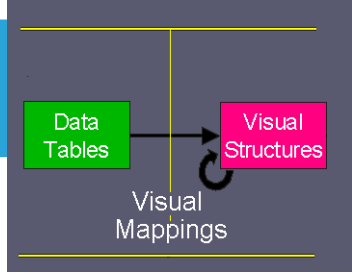


- Greyscale, color
- Histogram information on axes
- Smooth brushing
- Angular brushing

vr vis

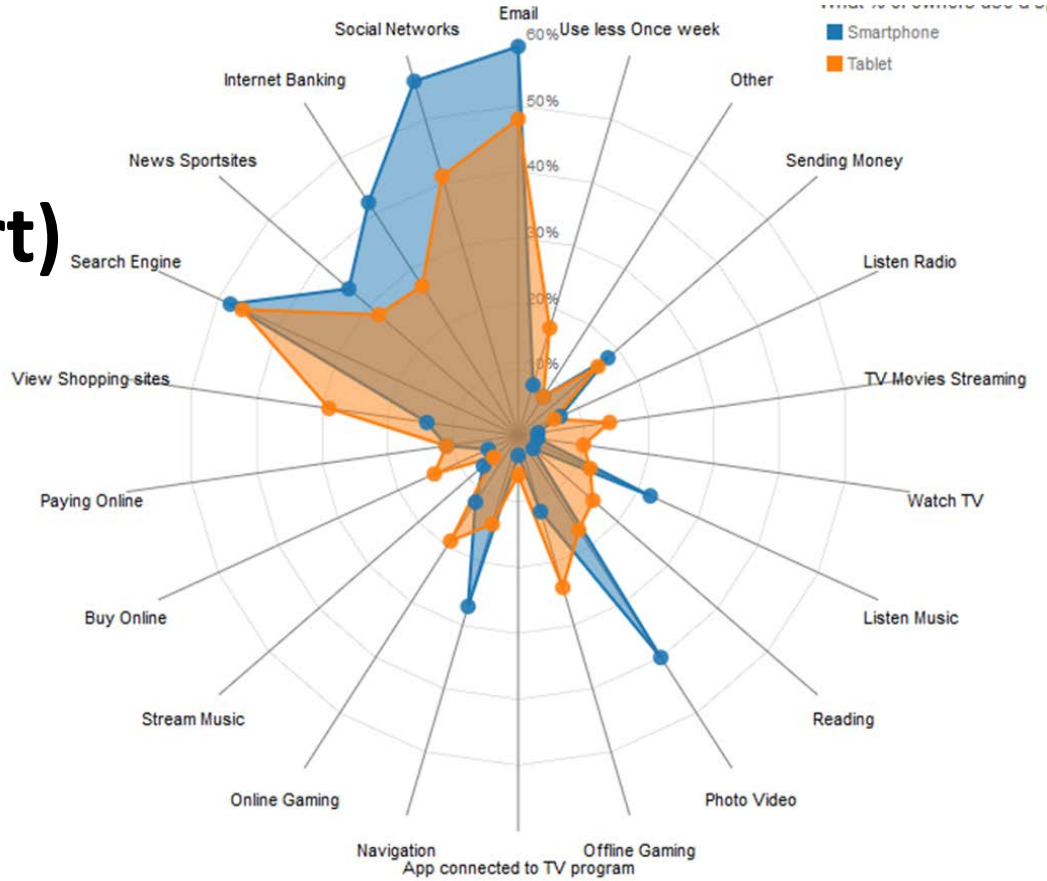
[Hauser et al.]

Multivariate Data



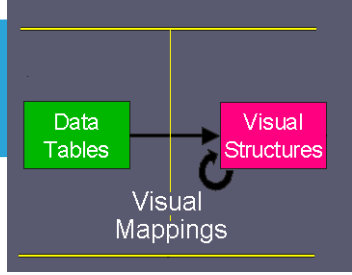
- N dependent variables

- **Radar chart (star plot, spider chart)**
  - Radial axes arrangement
  - Items are polylines



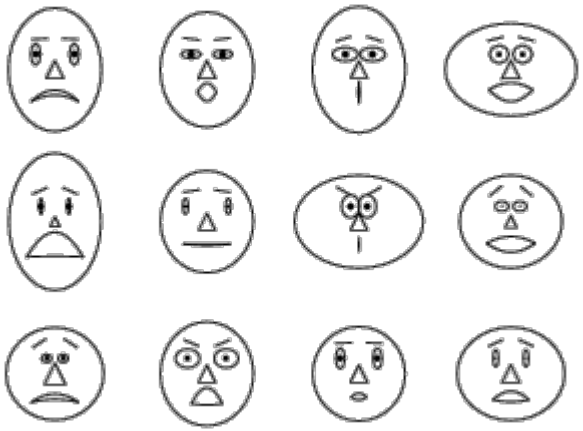
http://bl.ocks.org/nbremer/6506614

Multivariate Data



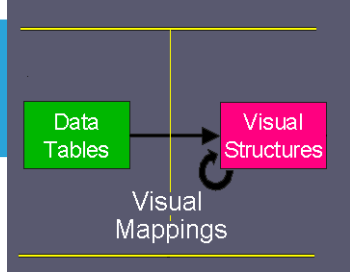
- N dependent variables

- **Chernoff Faces**
  - Icon-based display technique / glyphs
  - Each item is associated with one face
  - Quantitative value attributes control face characteristics



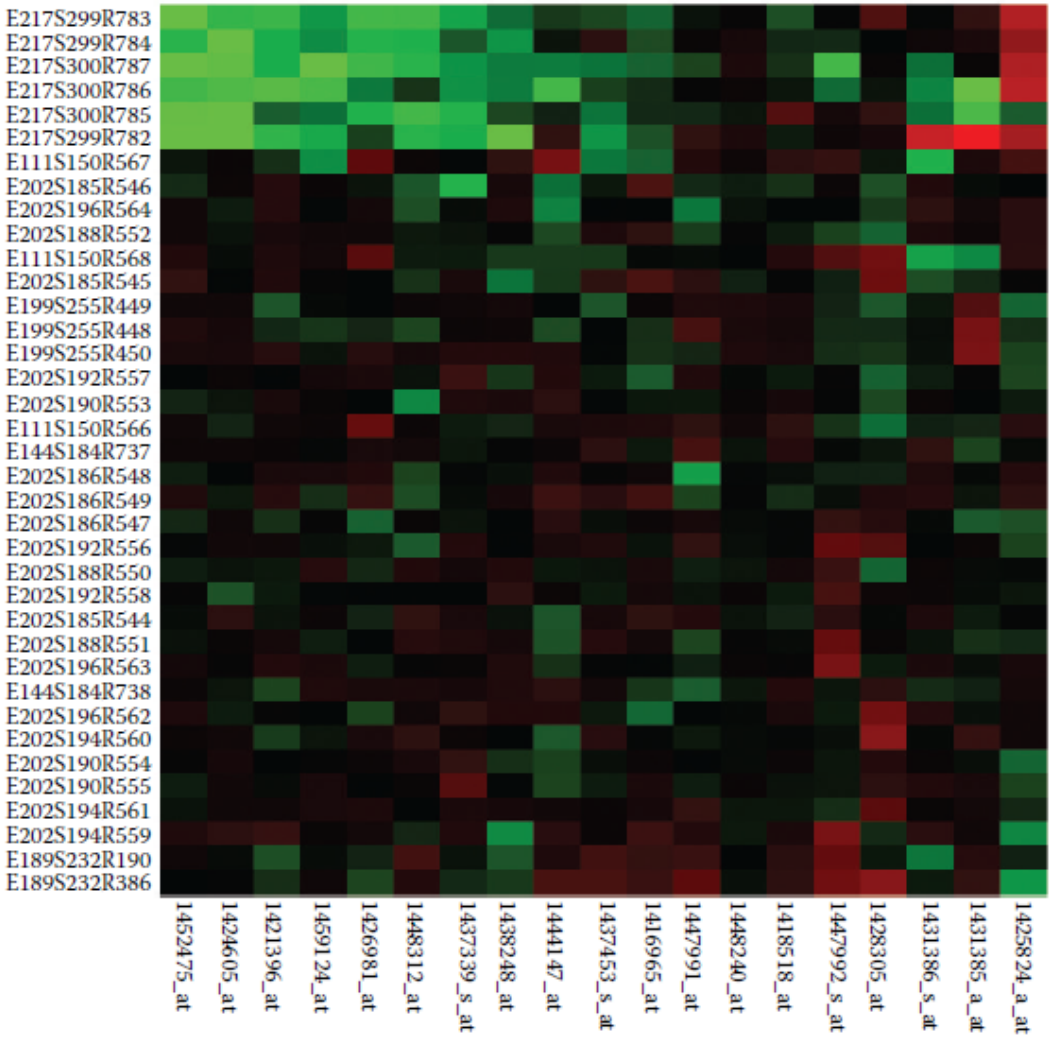
http://mathworld.wolfram.com/ChernoffFace.html

Multidimensional Data



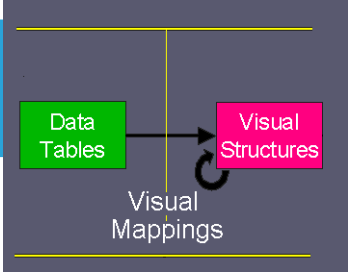
- 2 independent variables

- **Heatmap**
  - Quantitative value attribute (diverging color)
  - bioinformatics

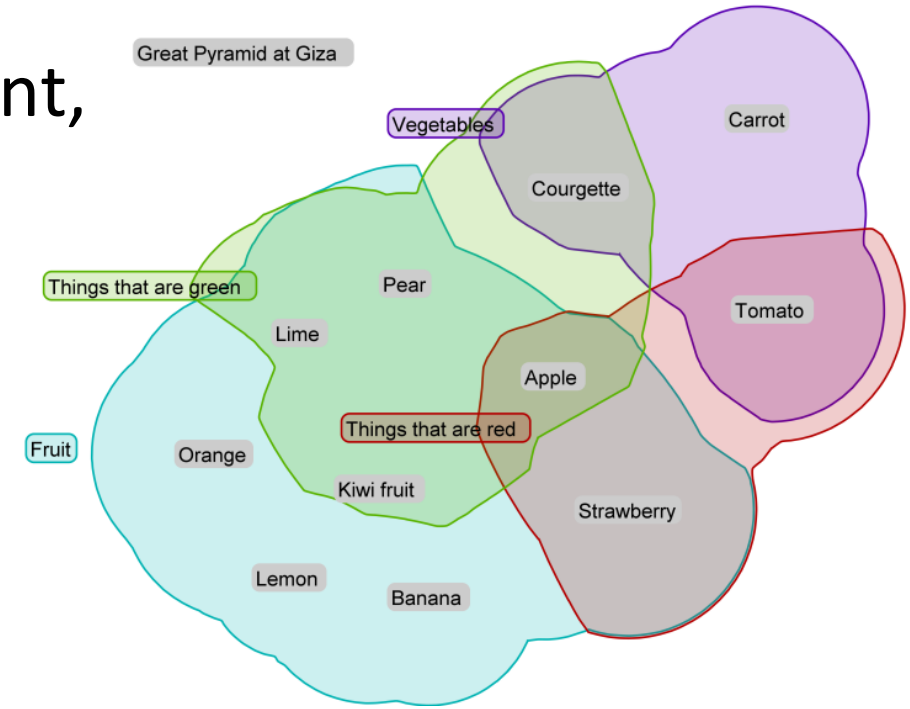


[Munzner, 2014]

Set-typed Data

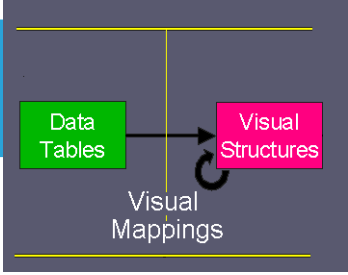


- Sets
  - Items classified into one or more categories
- **Euler Diagram**
  - Represents containment, intersection, exclusion
  - Uses closed curves
  - Only small number of sets possible

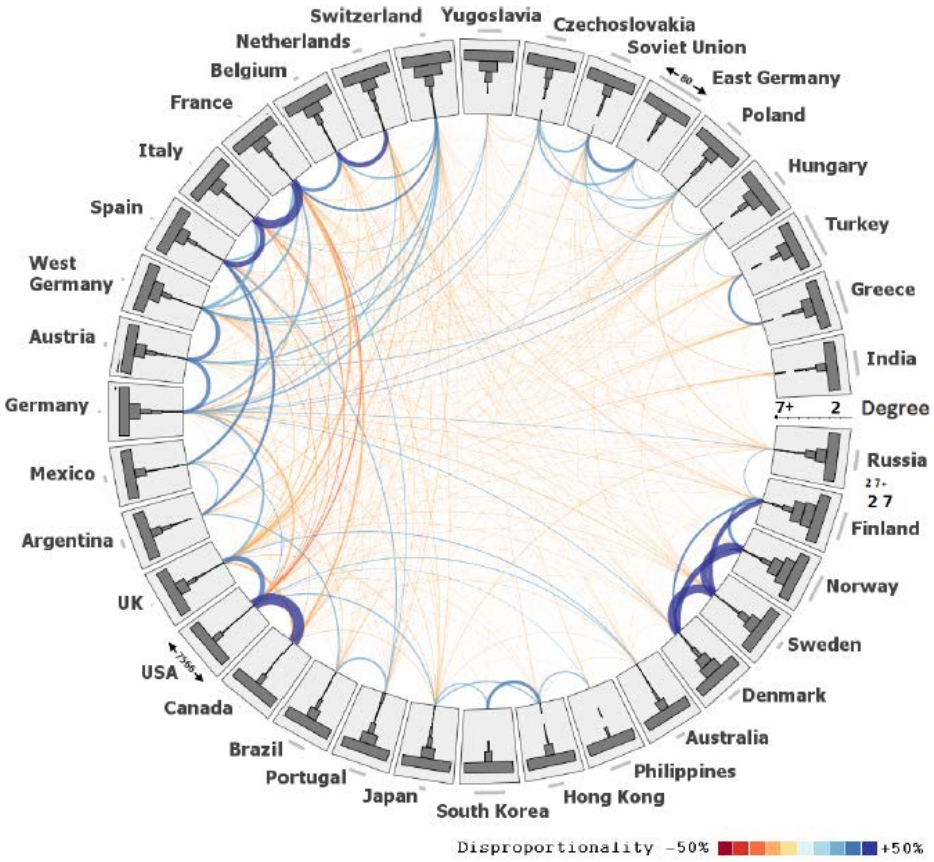


http://www-edc.eng.cam.ac.uk/tools/set_visualiser/

Set-typed Data

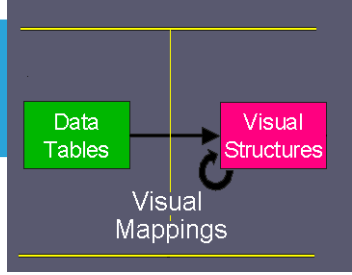


- Sets
  - Items classified into one or more categories
- **Radial Sets**
  - For large number of items
  - Sets: Radially arranged regions
  - Overlaps: links between regions

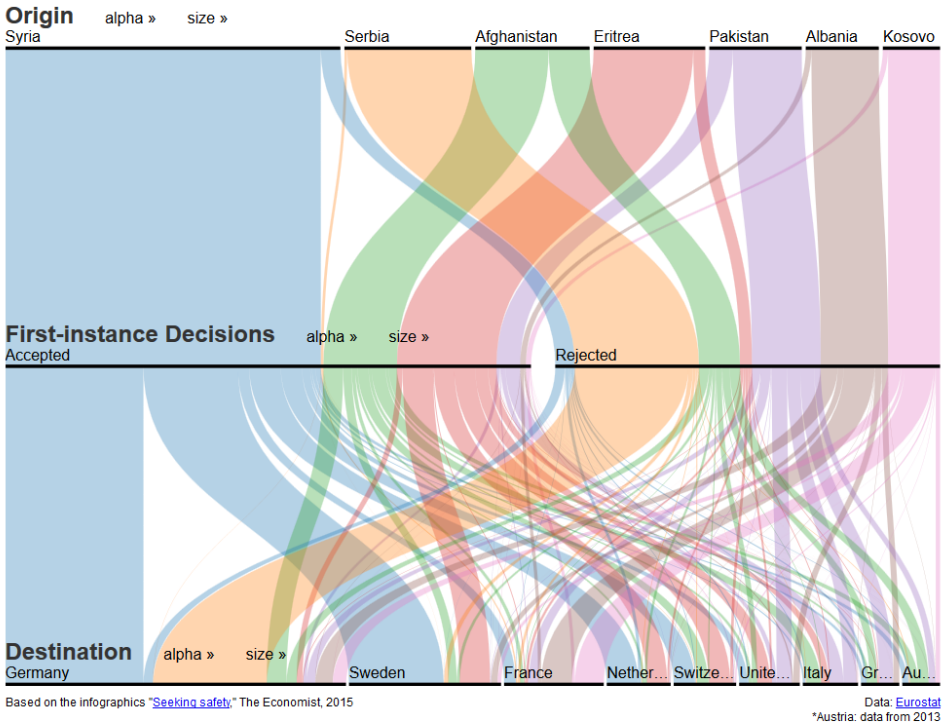


[Alsallakh et al., TVCG 2013]

Set-typed Data



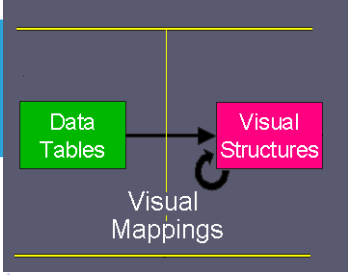
- Sets
  - Items classified into one or more categories
- **Parallel Sets**
  - Axis layout of parallel coordinates
  - Boxes: categories
  - „Parallelograms“ / „ribbons“ between axes: relations between categories



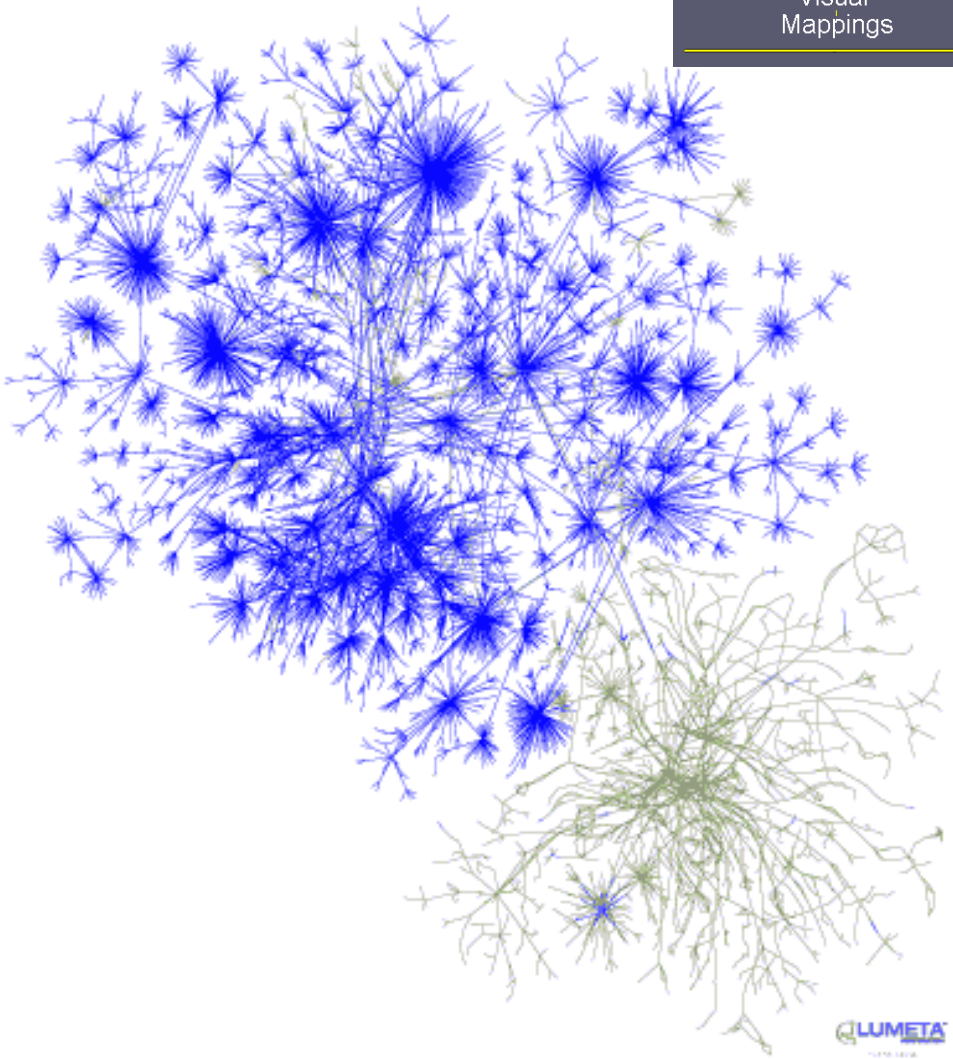
http://multivis.net/lecture/parallel-sets.htm

Johannes Kehrer

Networks

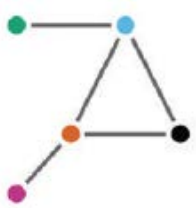


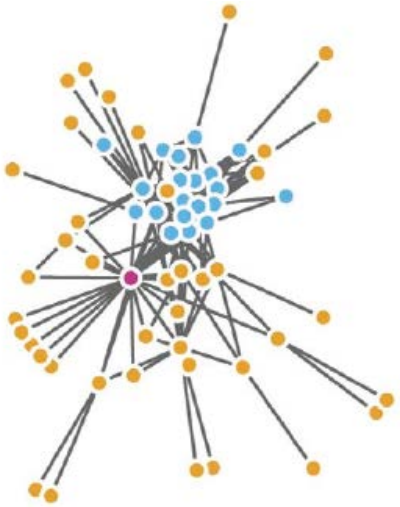
- Used to describe Communication Networks, Telephone Systems, Internet
- Nodes
  - Unstructured
  - Nominal
  - Ordinal
  - Quantity
- Links
  - Directed
  - Undirected

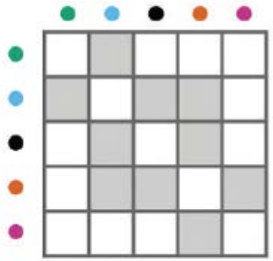


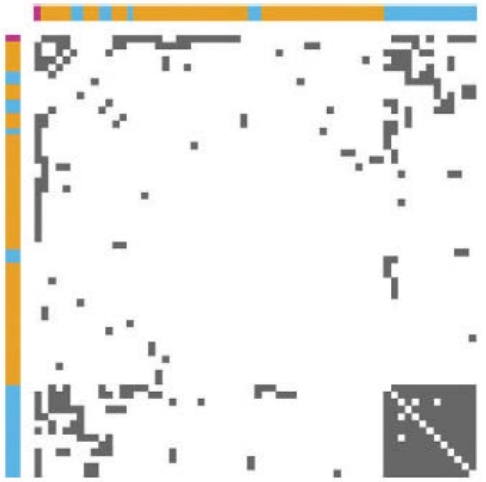
[Branigan et al, 2001]

Networks





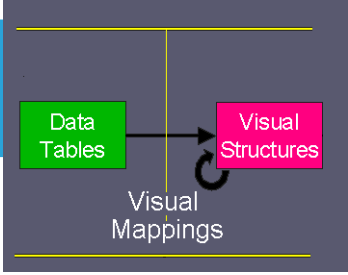




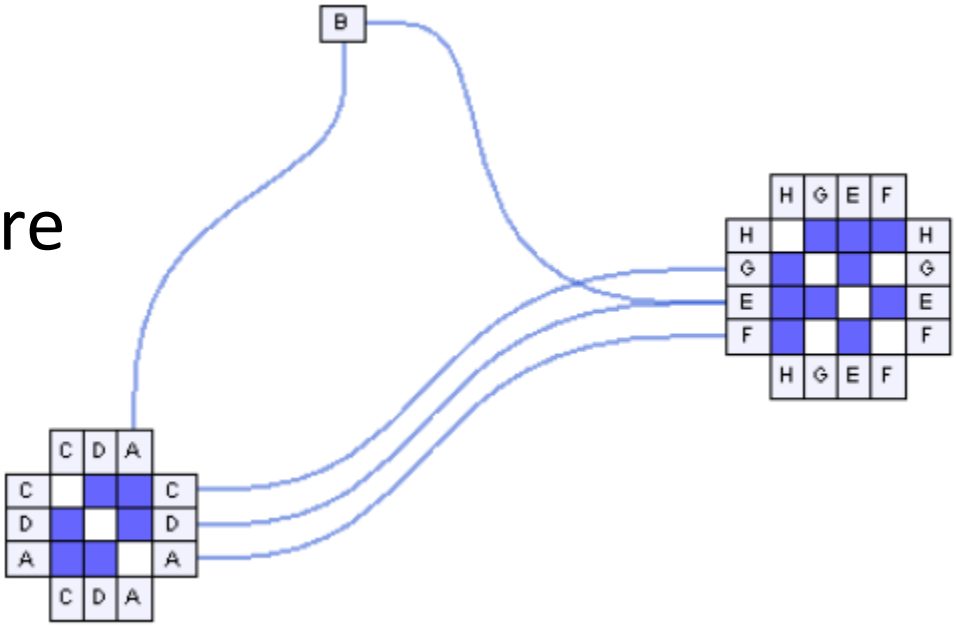
- Node-link diagram
  - Nodes: point marks
  - Links: line marks
  - Force-directed layout
- Adjacency matrix
  - Nodes: table keys
  - Links: cell entries
  - Symmetric for undirected networks

[Munzner, 2014]
[Gehlenborg and Wong, Points of View: Networks, 2012]

Large Networks



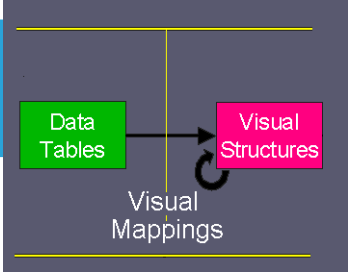
- Hybrid node-link and matrix representation
- **NodeTrix**
  - Node-link diagram: overall graph structure
  - Adjacency matrices: communities



[Von Landesberger et al., Viusal Analysis of Large Graphs, CGF 2011]

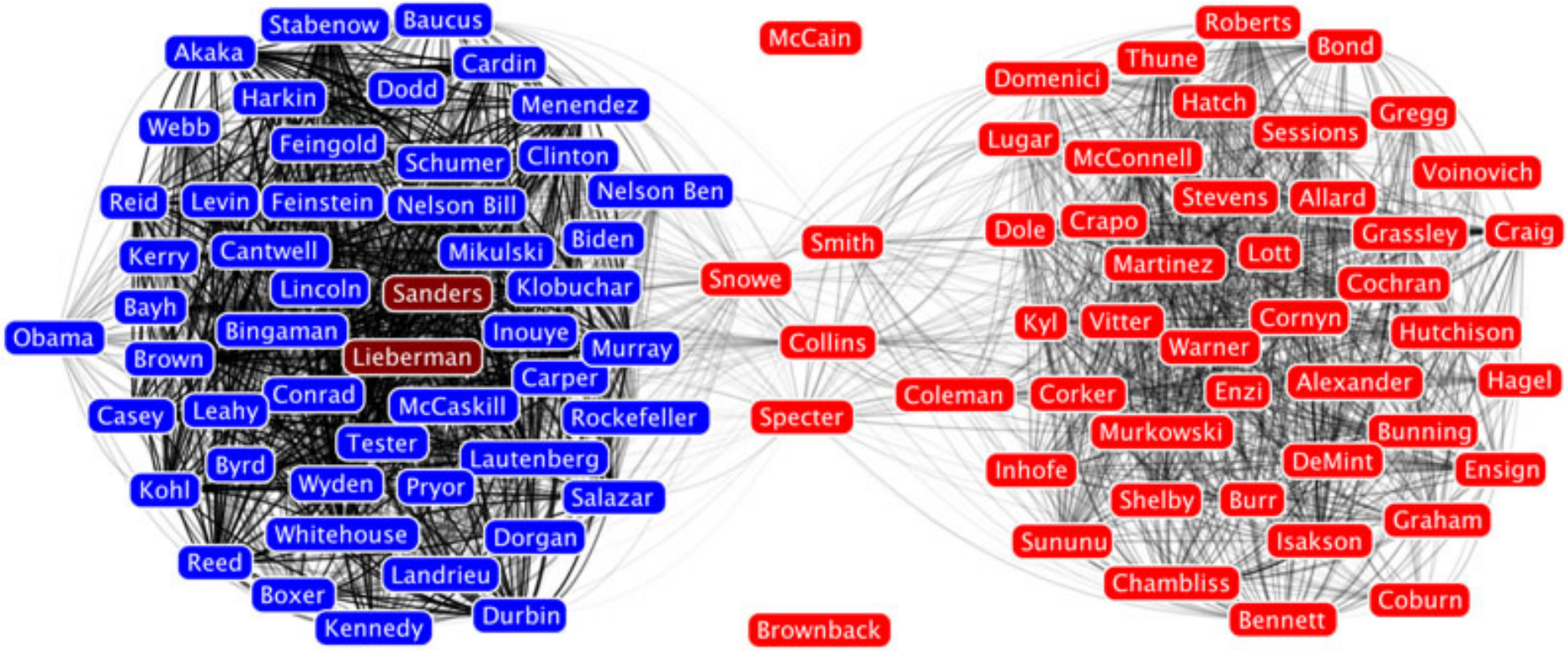
[Henry et al., NodeTrix, TVCG 2007]

Large Networks



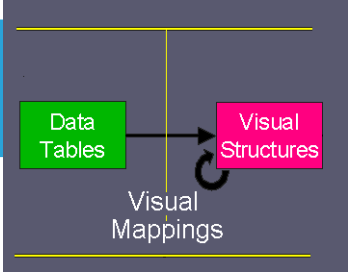
Example:

U.S. Senate 2007 co-voting network

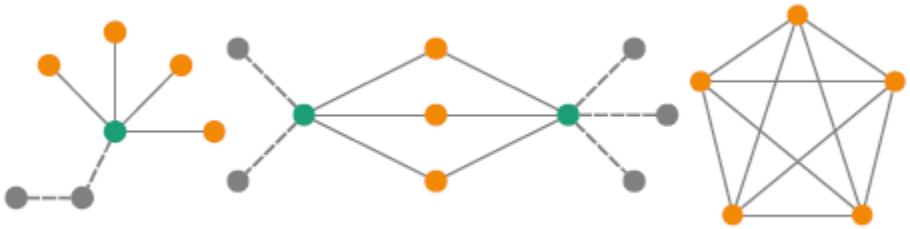


http://www.cs.umd.edu/hcil/science20/

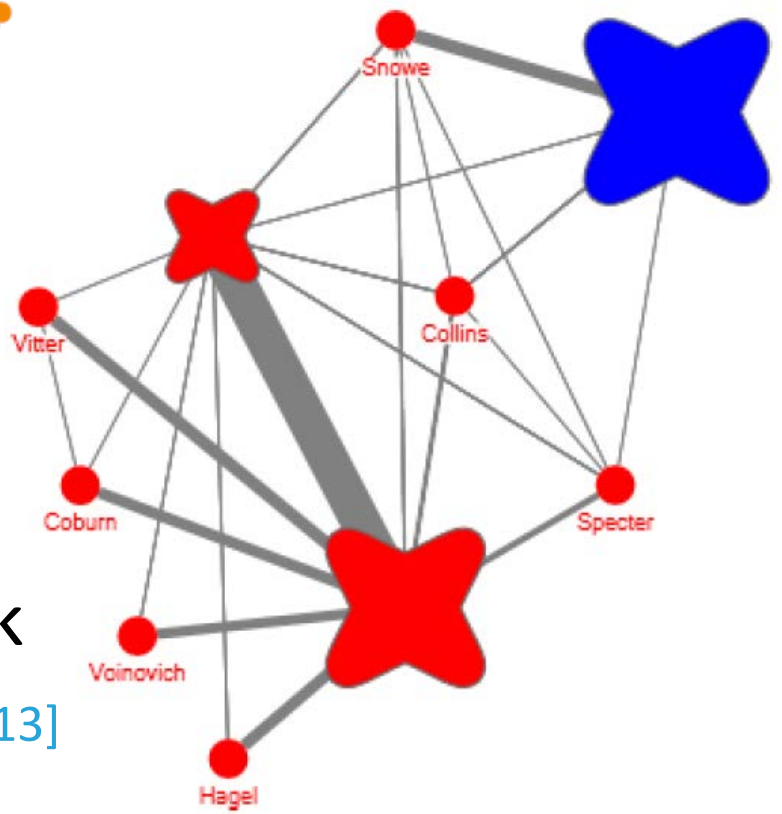
Large Networks



- **Motif simplification**
  - Motifs: subnetworks with common patterns of nodes and links

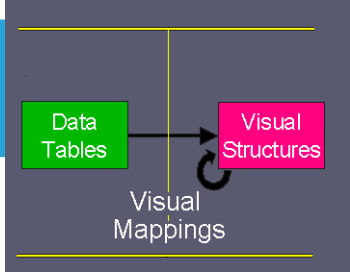


  - Common motifs replaced by glyphs
  - Example: clique motif glyphs in co-voting network



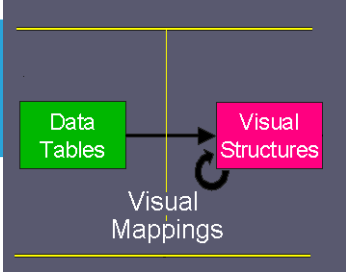
[Dunne et al., Motif simplification, CHI 2013]

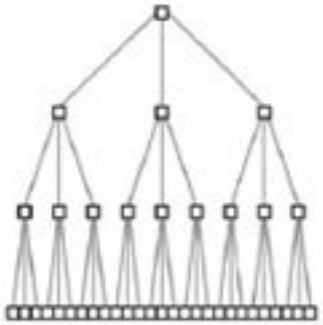
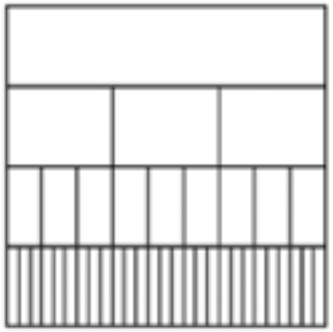
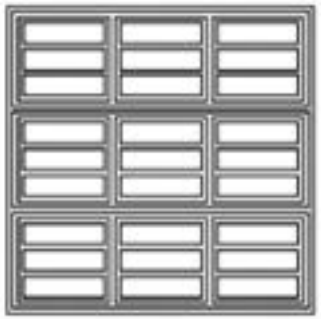
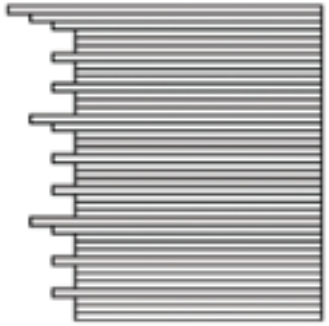
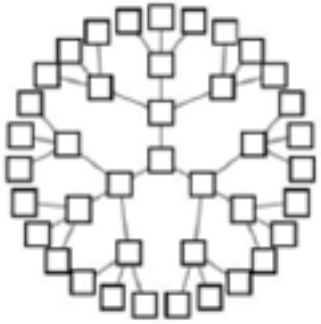
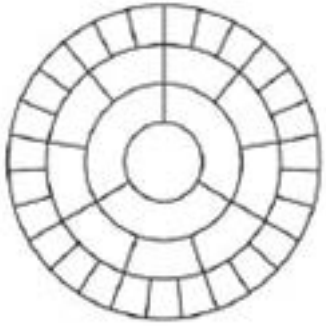
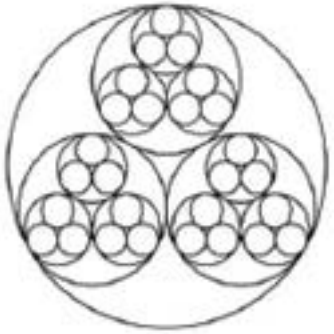
Trees



- Visual Structures that refer to use of connection and enclosure to encode relationships among cases
- Desirable Features
  - Planarity (no crossing edges)
  - Clarity in reflecting the relationships among the nodes
  - Clean, non-convoluted design
  - Hierarchical relationships should be drawn directional

Trees

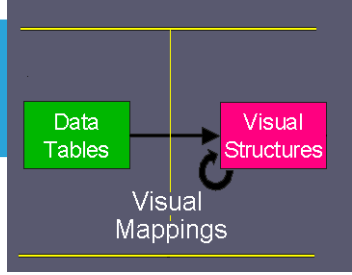


	node-link	icicle	treemap	Indented outline
rectilinear				
radial				
	node-link	concentric circles	nested circles	

[Munzner, 2014]

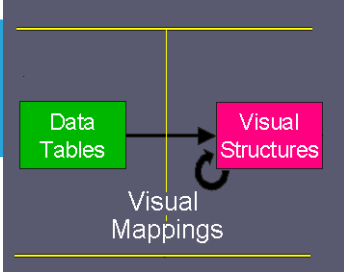
[McGuffin and Robert, Quantifying the Space-Efficiency of 2D Graphical Representations of Trees, 2010]

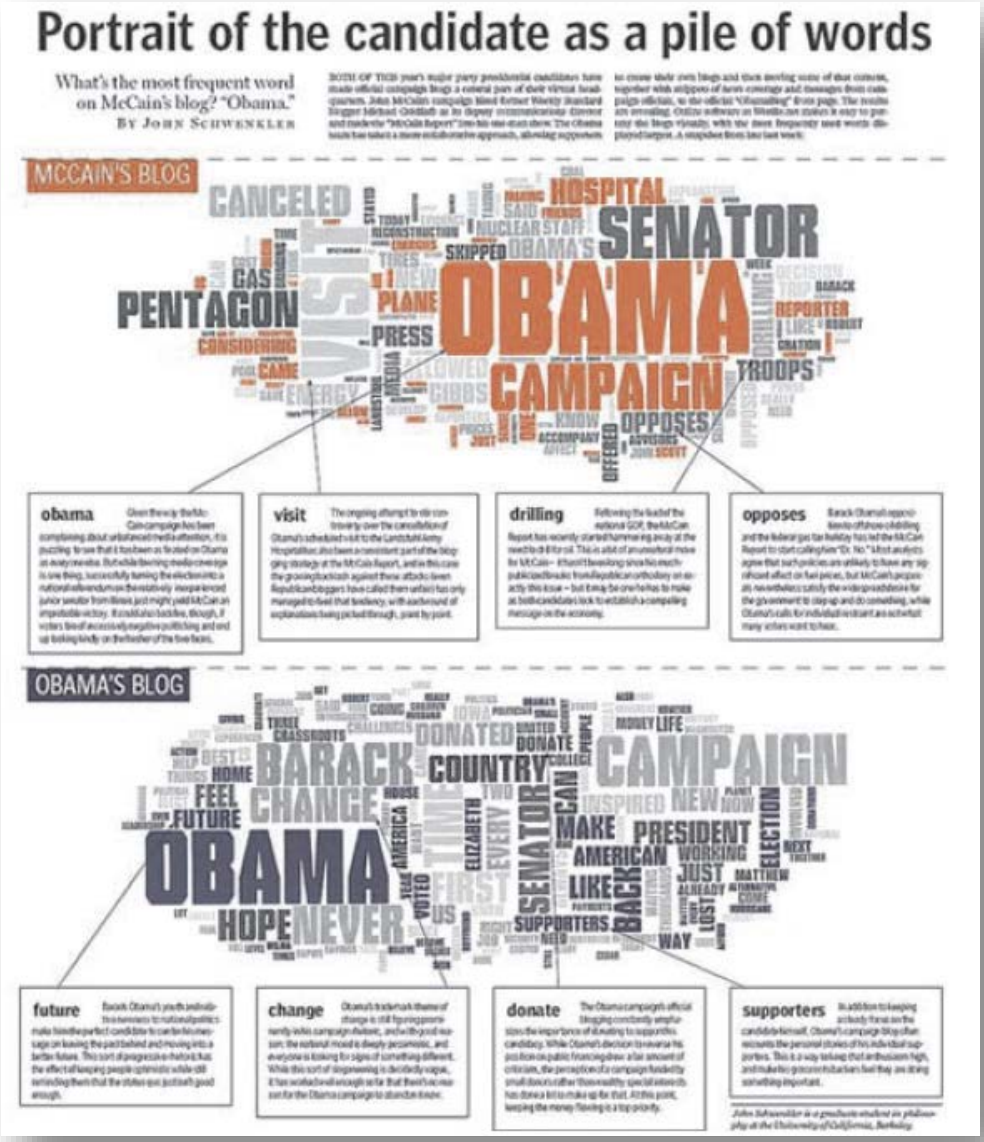
Visualization of Text Documents



- Transforming text information into spatial representation to reveal:
  - Thematic patterns
  - Relationships between documents

Text



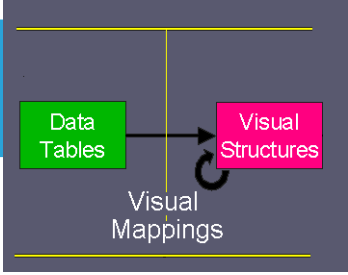


- **Wordle**
  - Font size: word frequency in documents
  - Removal of frequent „stopwords“ (*the, of, in*...)
  - Layout: tight packing of words
  - Similar: tag clouds

http://www.wordle.net/

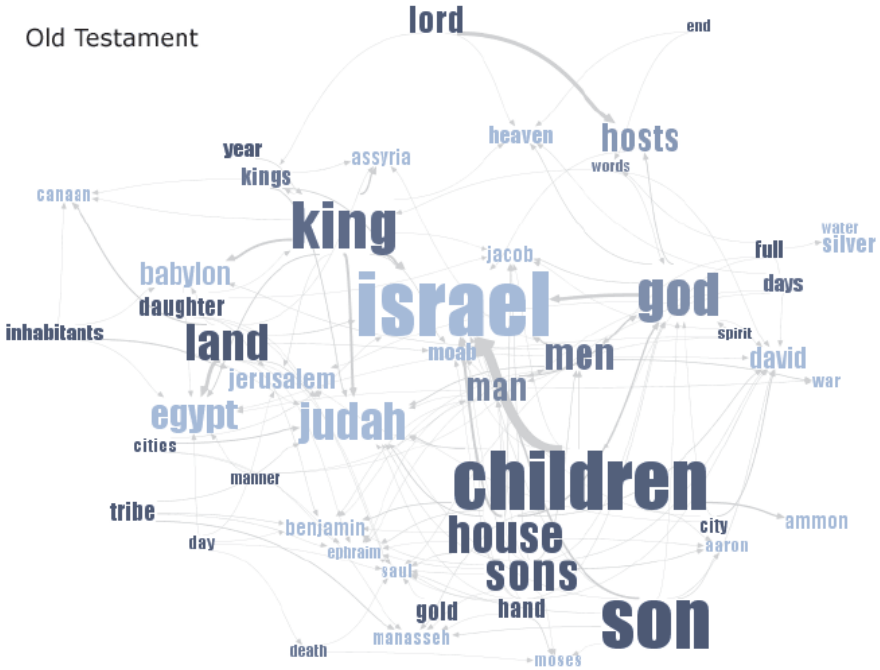
[Viegas et al., TVCG 2009]

Text



- **Phrase Nets**
  - Visualizes text patterns in documents
  - Nodes: words (node size: frequency)
  - Edges: user-specified relation

[van Ham et al., TVCG 2009]



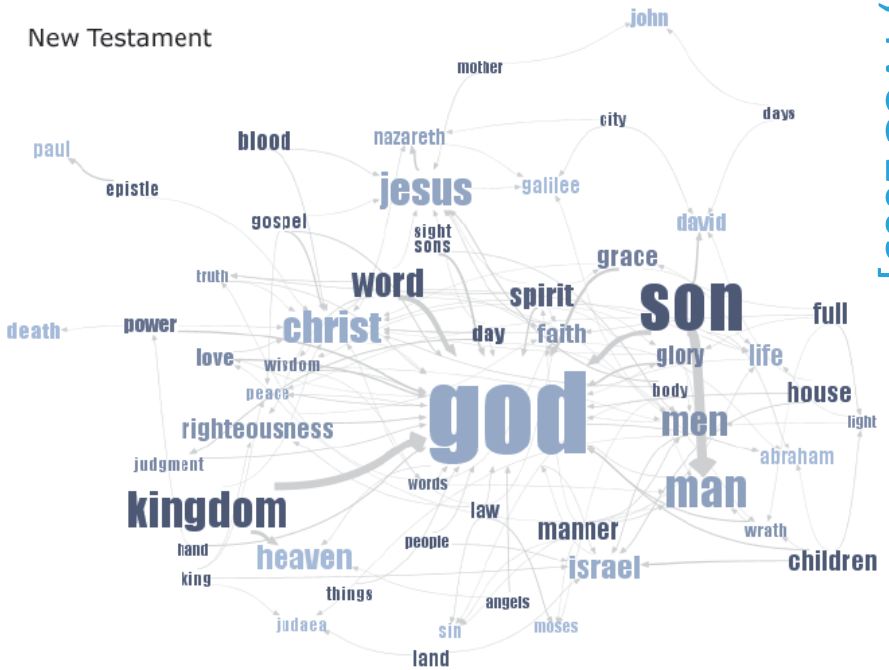
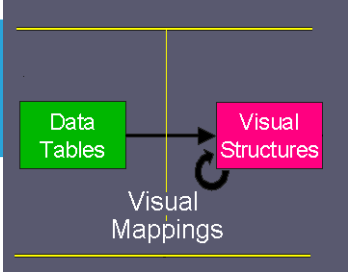
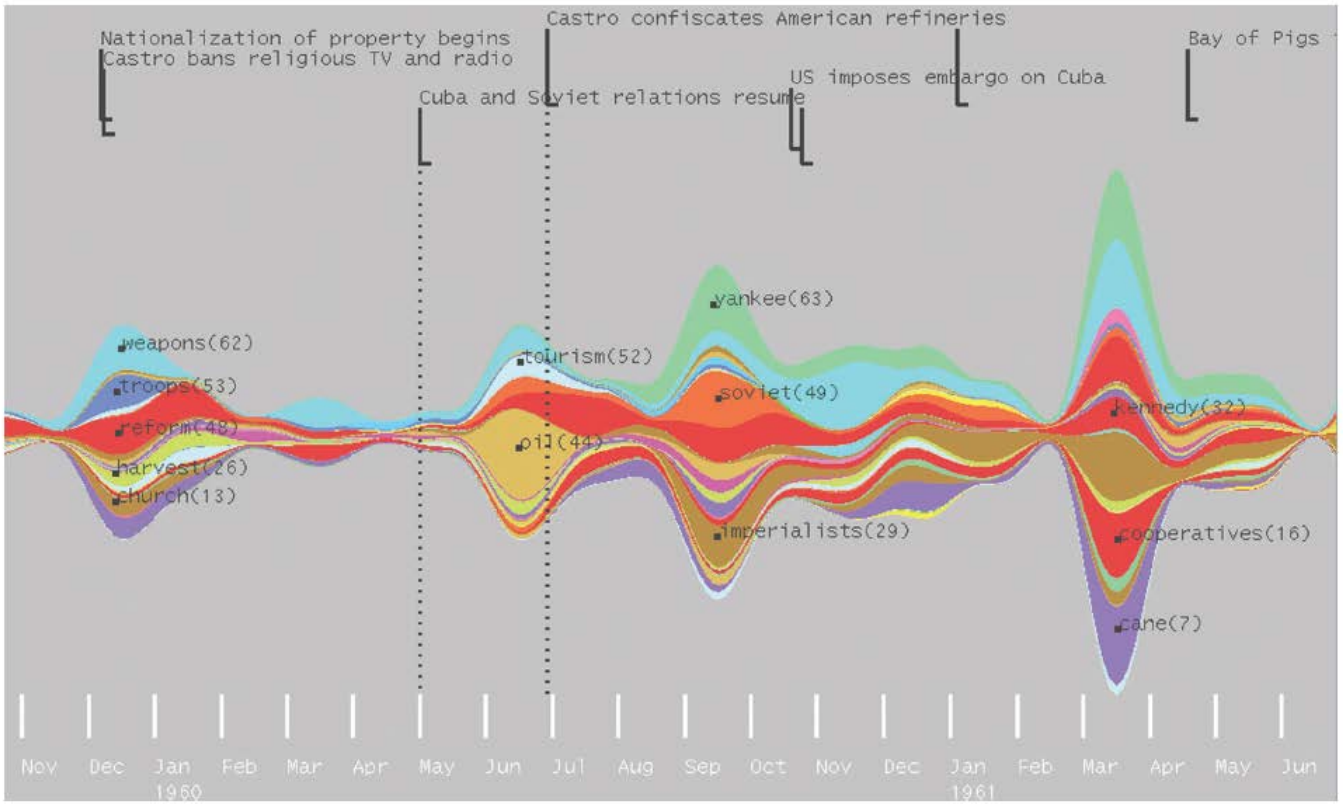


Fig 4. Matching the same pattern on different texts. Here we used the pattern "X of Y" to compare the old and new testaments. Israel takes a central place in the Old Testament, while God acts as the main pattern receiver in the New Testament.

Text



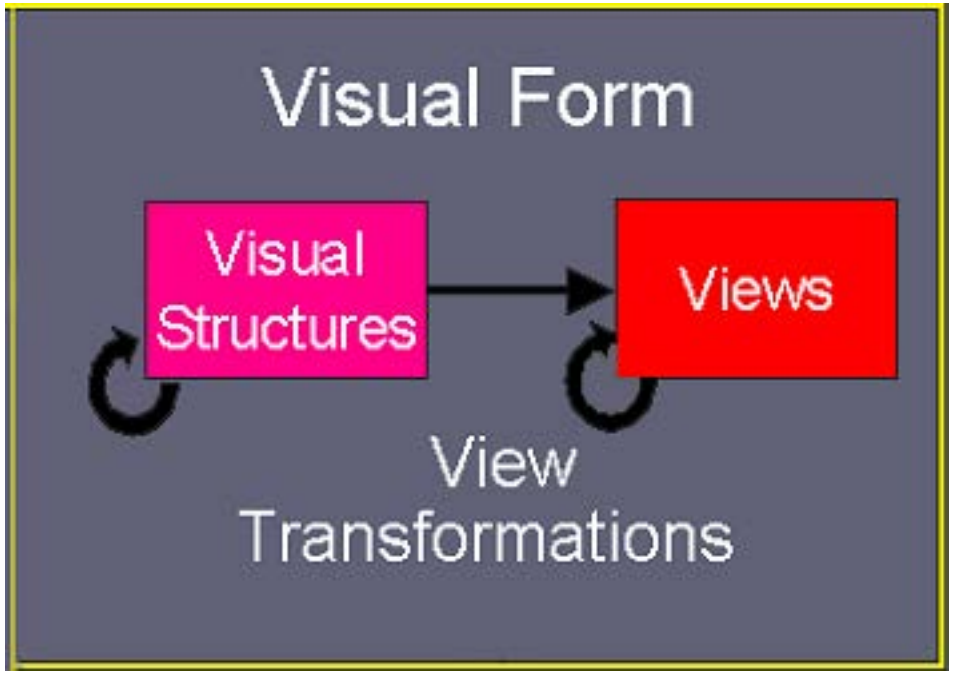
- **ThemeRiver**
  - Thematic changes in document collections over time



[Havre et al., TVCG 2002]

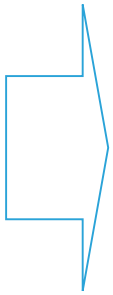
Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.

View Transformations



View Transformations

- Problems:
  - Scale
  - Region of Interest
  - How to specify focus?
    - Find new focus
    - Stay oriented



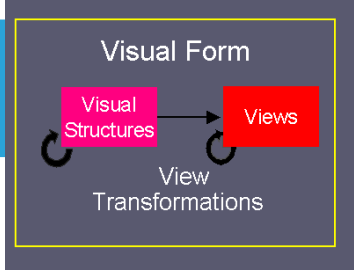
Overview + Detail

Zooming

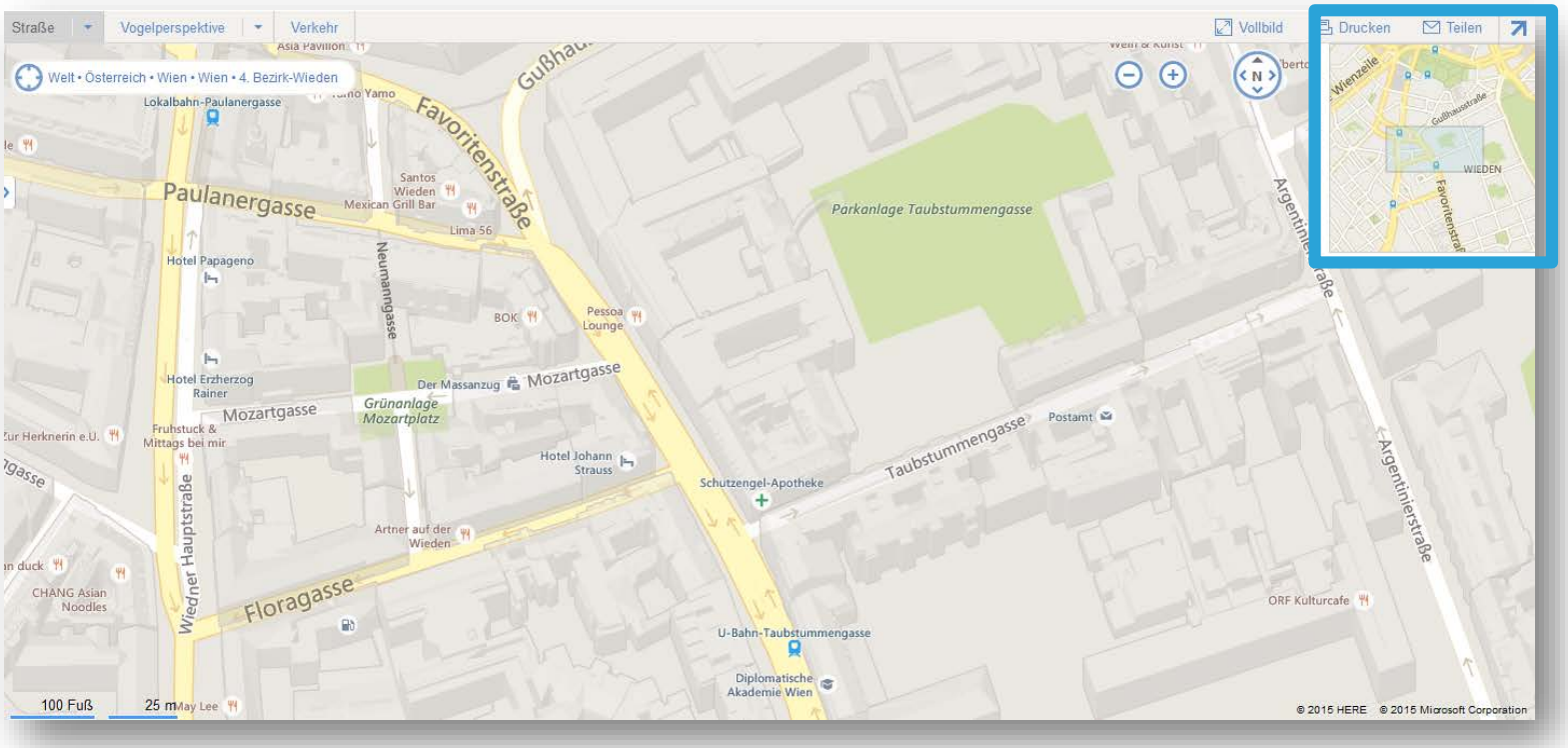
Focus + Context

- Ability to interactively modify and augment visual structures, turning static presentations into visualizations

Overview+Detail – Examples



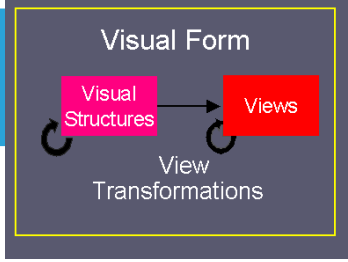
- Provide both overview and detail displays



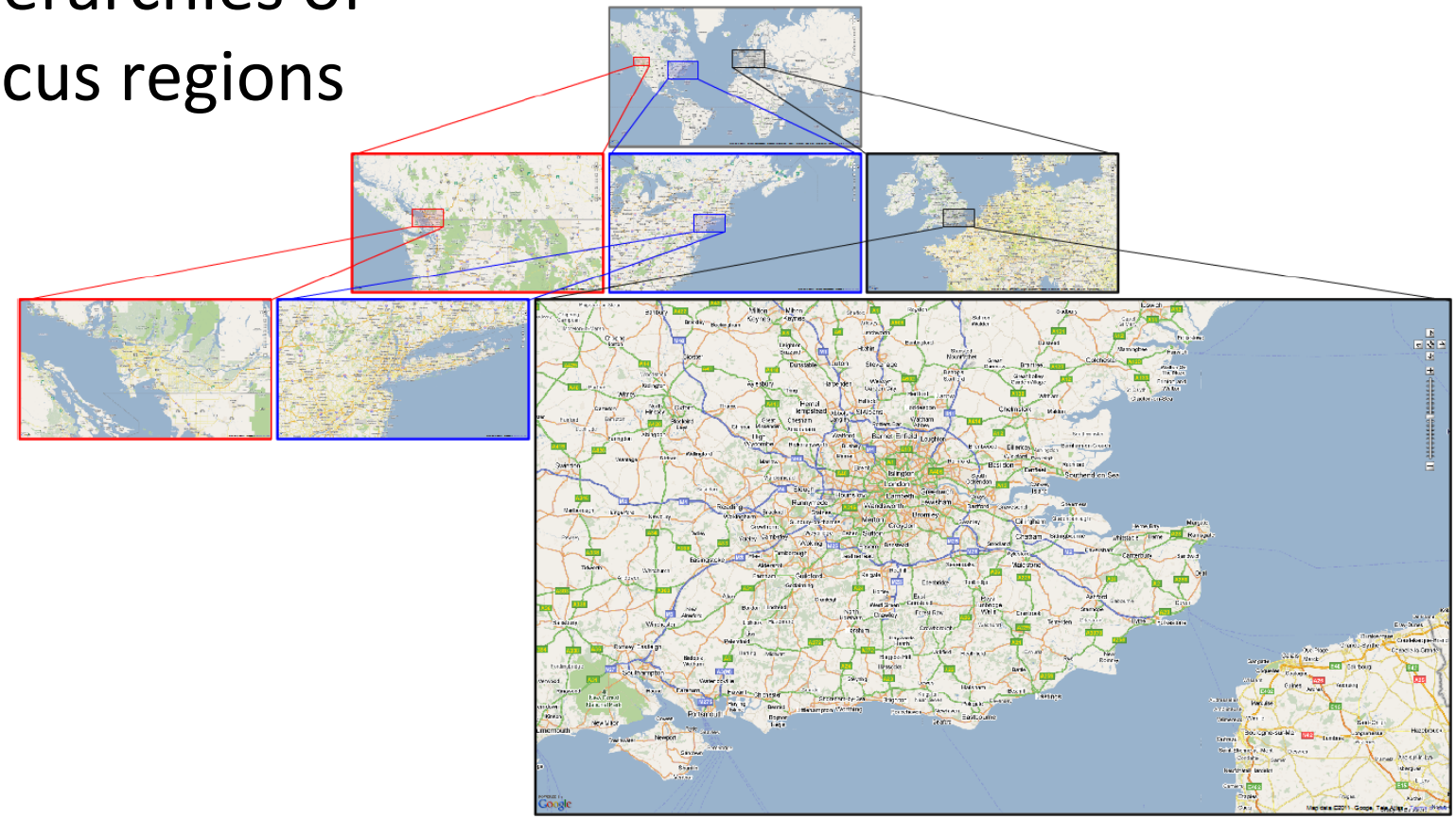
https://www.bing.com/maps/

- Semantic Zoom
  - Amounts of detail depending on zoom level

Overview+Detail – Examples

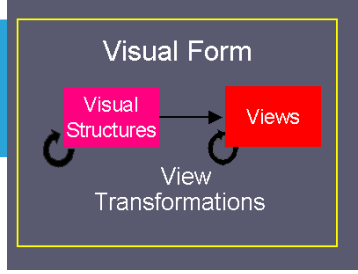


- Polyzoom
  - Hierarchies of focus regions

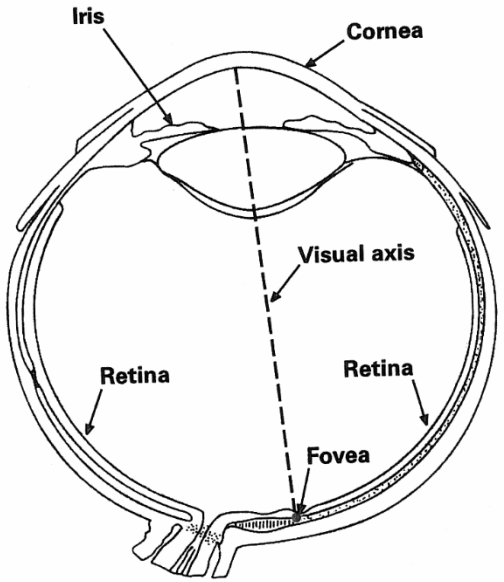


[Javed et al., CHI 2012]

Focus + Context



- Overview Content
- Detail Content
- Dynamical Integration

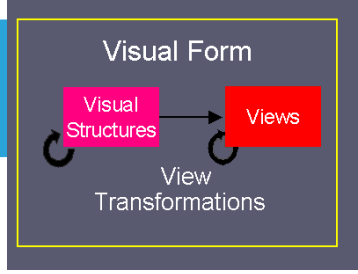


Rationale

- Zooming hides the context
- Two separate displays split attention
- Human vision has both fovea and retina

Courtesy of Jock Mackinlay

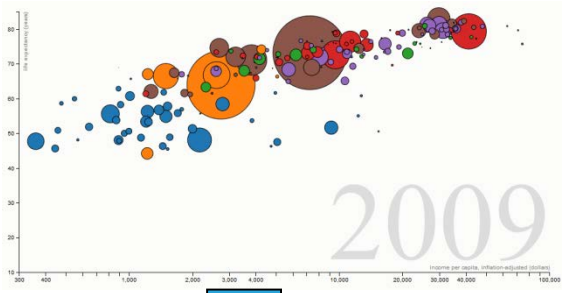
Focus + Context

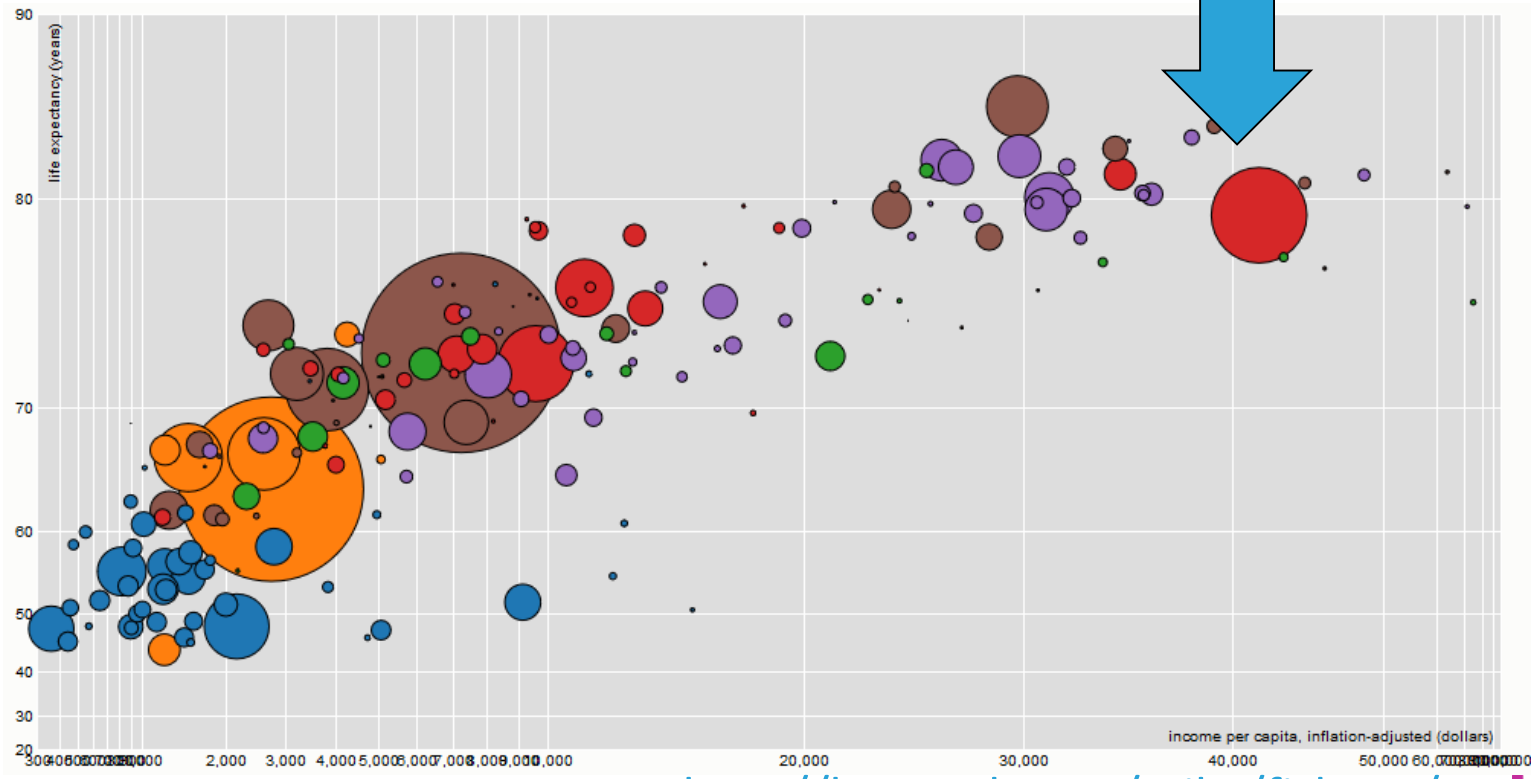


- Filtering
  - Selection of cases
  - Manually or dynamically
- Selective aggregation
  - New cases
- Distortion
  - Relative changes in the number of pixels devoted to objects in the space
  - Types of distortion:
    - Size of the objects representing cases
    - Size due to perspective
    - Size of the space itself

Focus + Context – Example

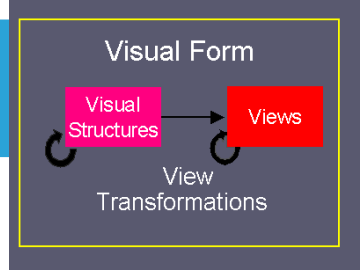
- Cartesian Fisheye Transformation
  - Local magnification around the mouse pointer

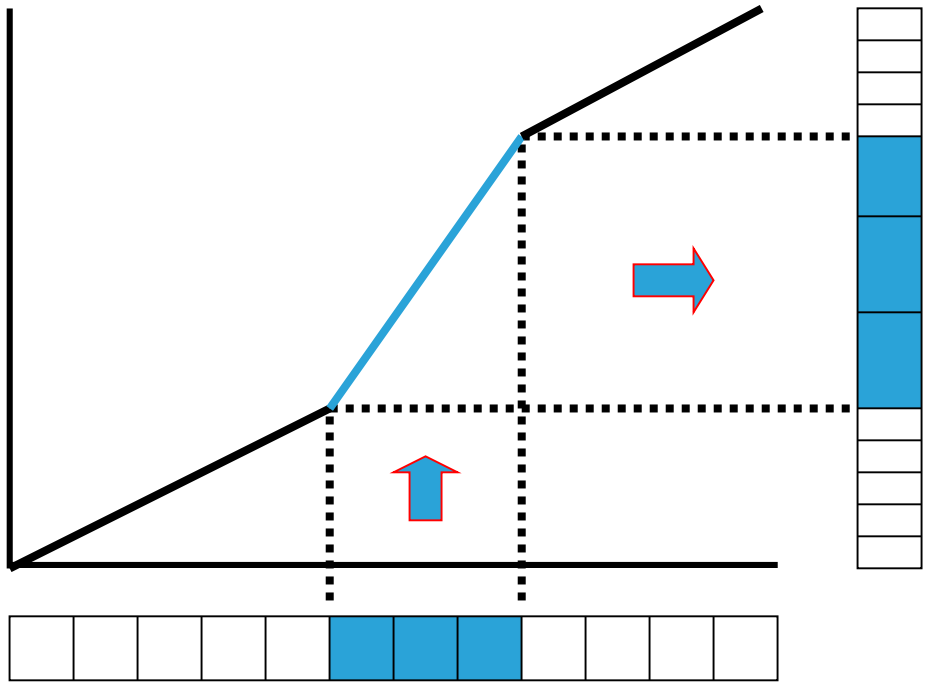




http://bost.ocks.org/mike/fisheye/

Visual Transfer Function

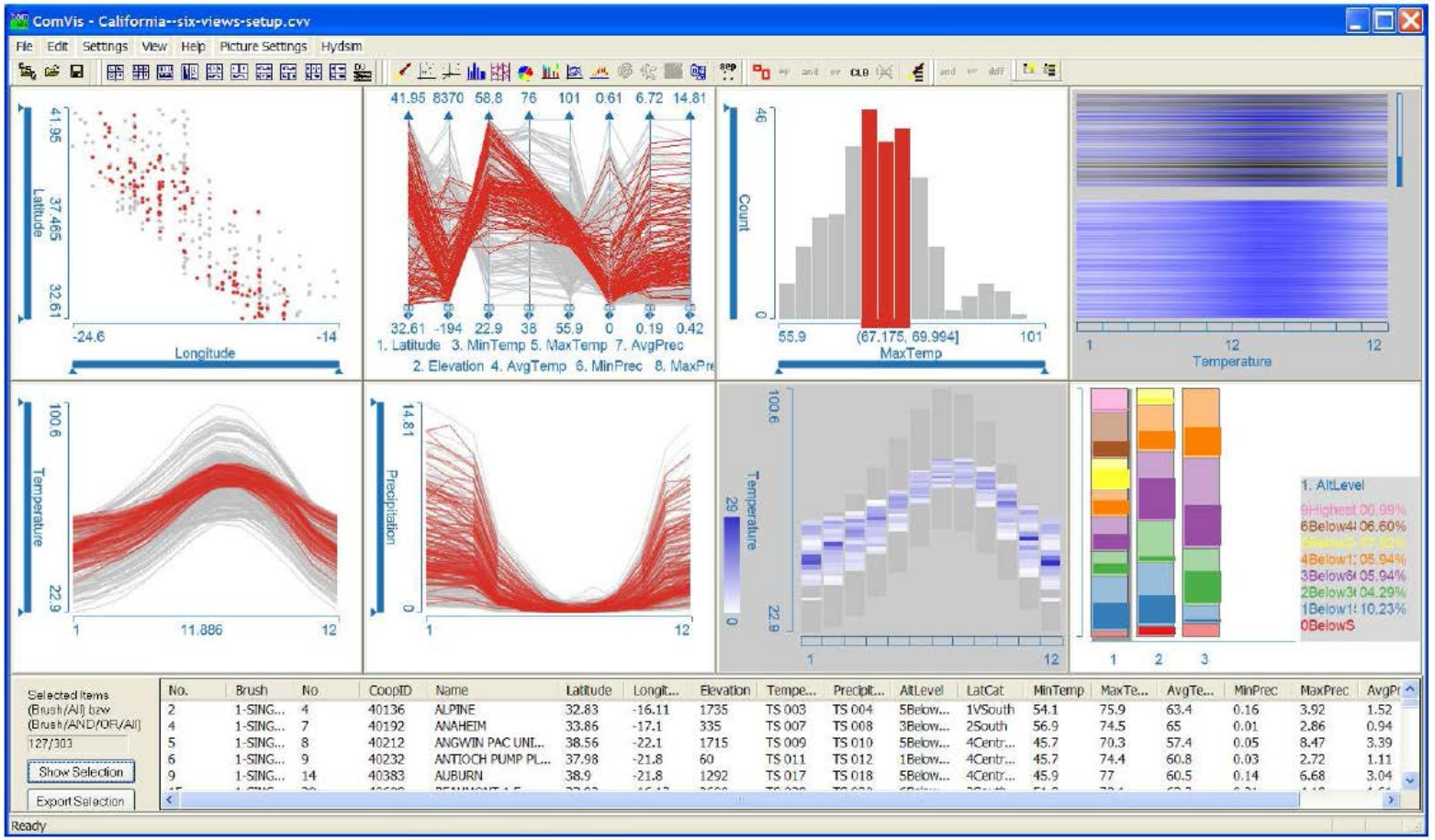




- Functions that distort visualizations by stretching or compressing them, giving the portion of visualization attended to more visual detail
- DOI - Degree Of Interest Function

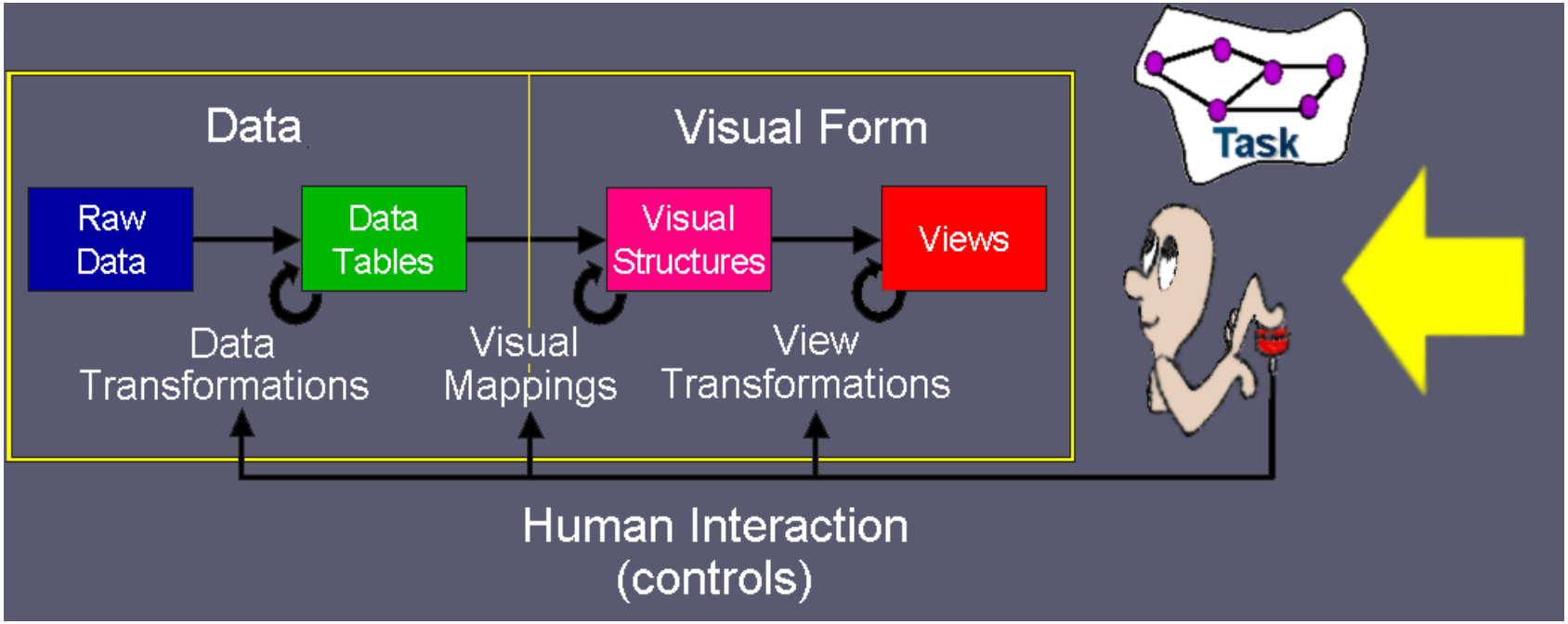
Multiple Coordinated Views

- Two or more (usually juxtaposed) views to support investigation of single conceptual entity



[Matkovic et al., IV 2008]

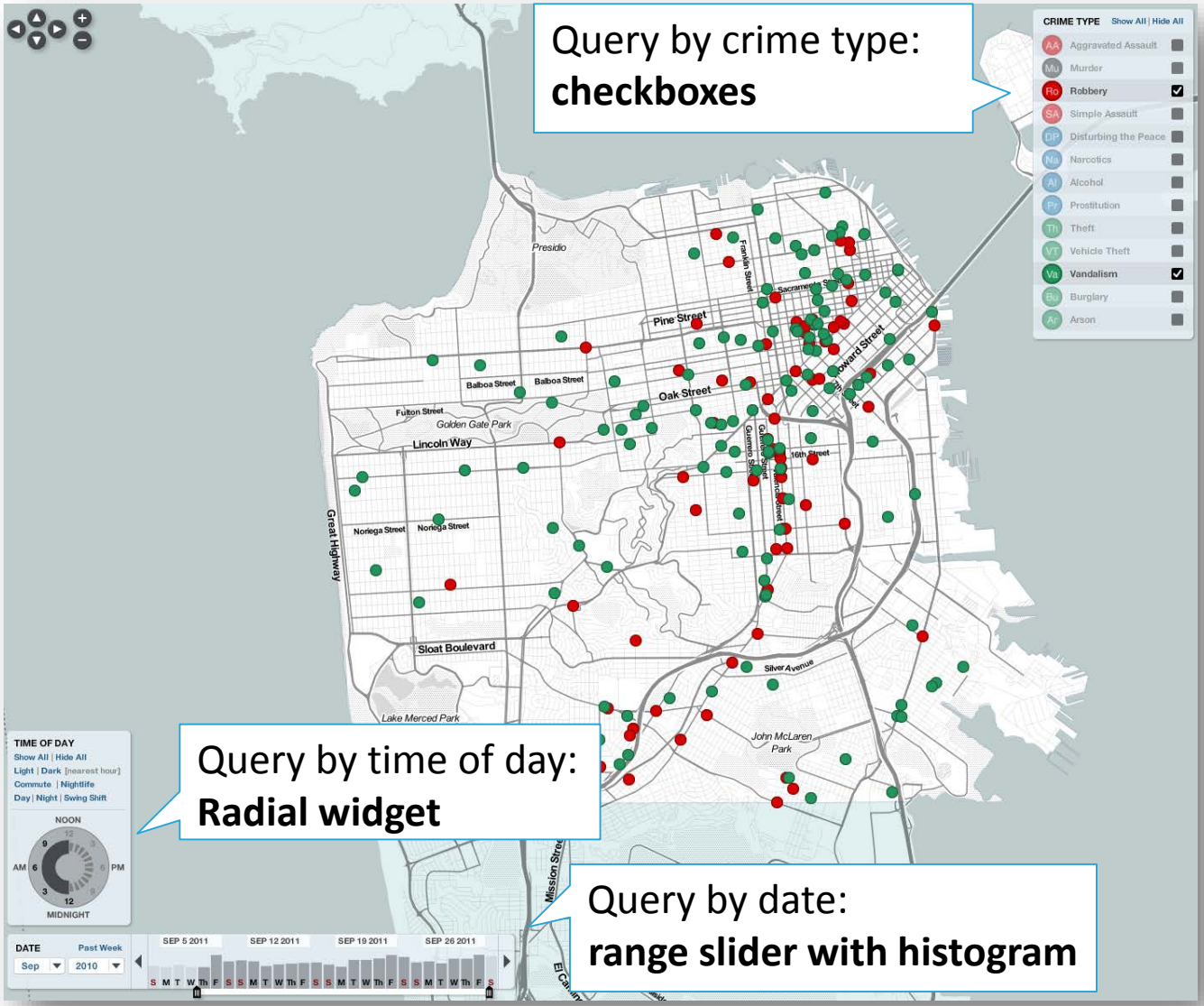
Interaction



- Dynamic Queries
- Details-on-Demand
- Brushing

Dynamic Queries

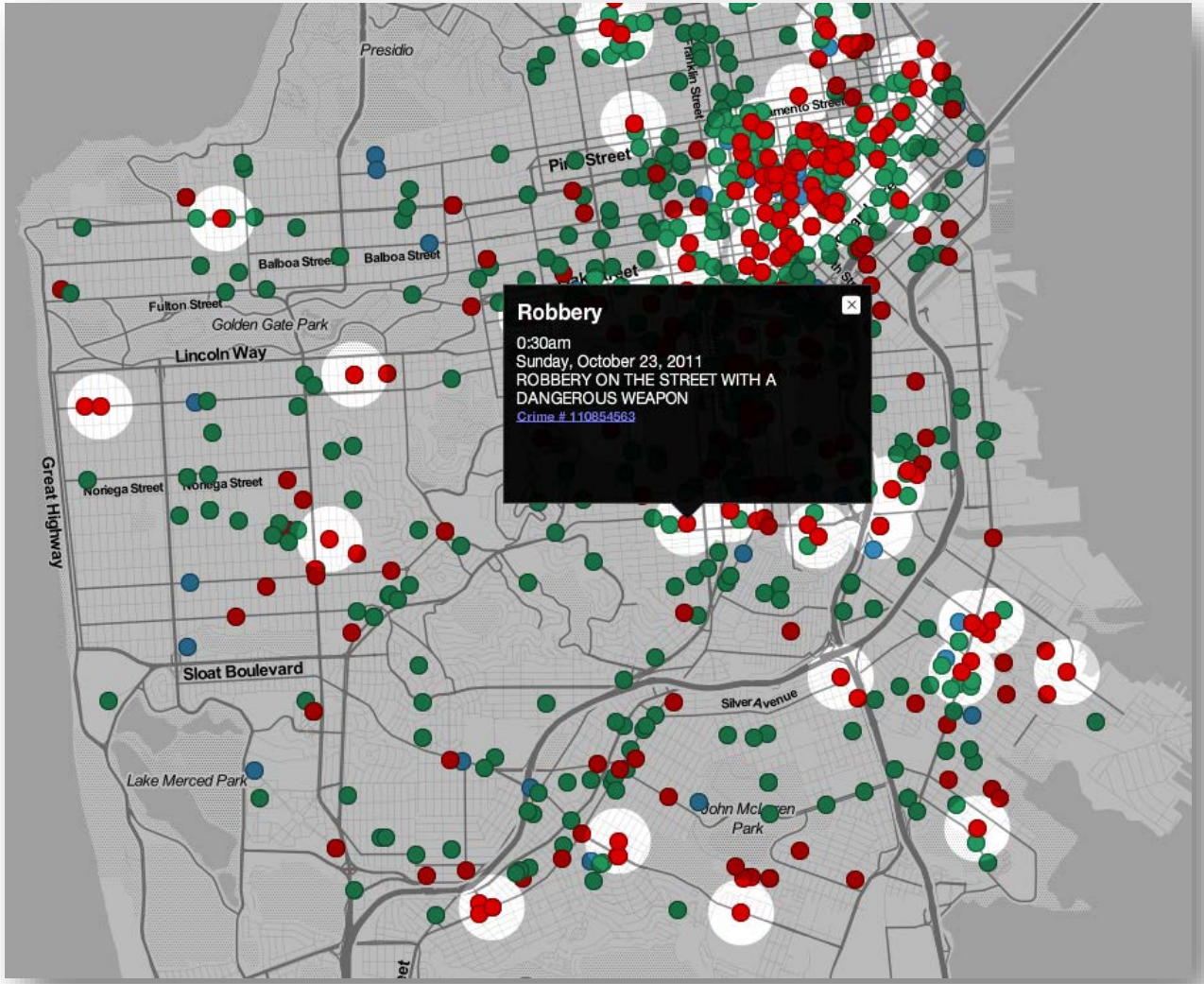
- Widgets to control item visibility



http://sanfrancisco.crimespotting.org/

Details-on-Demand

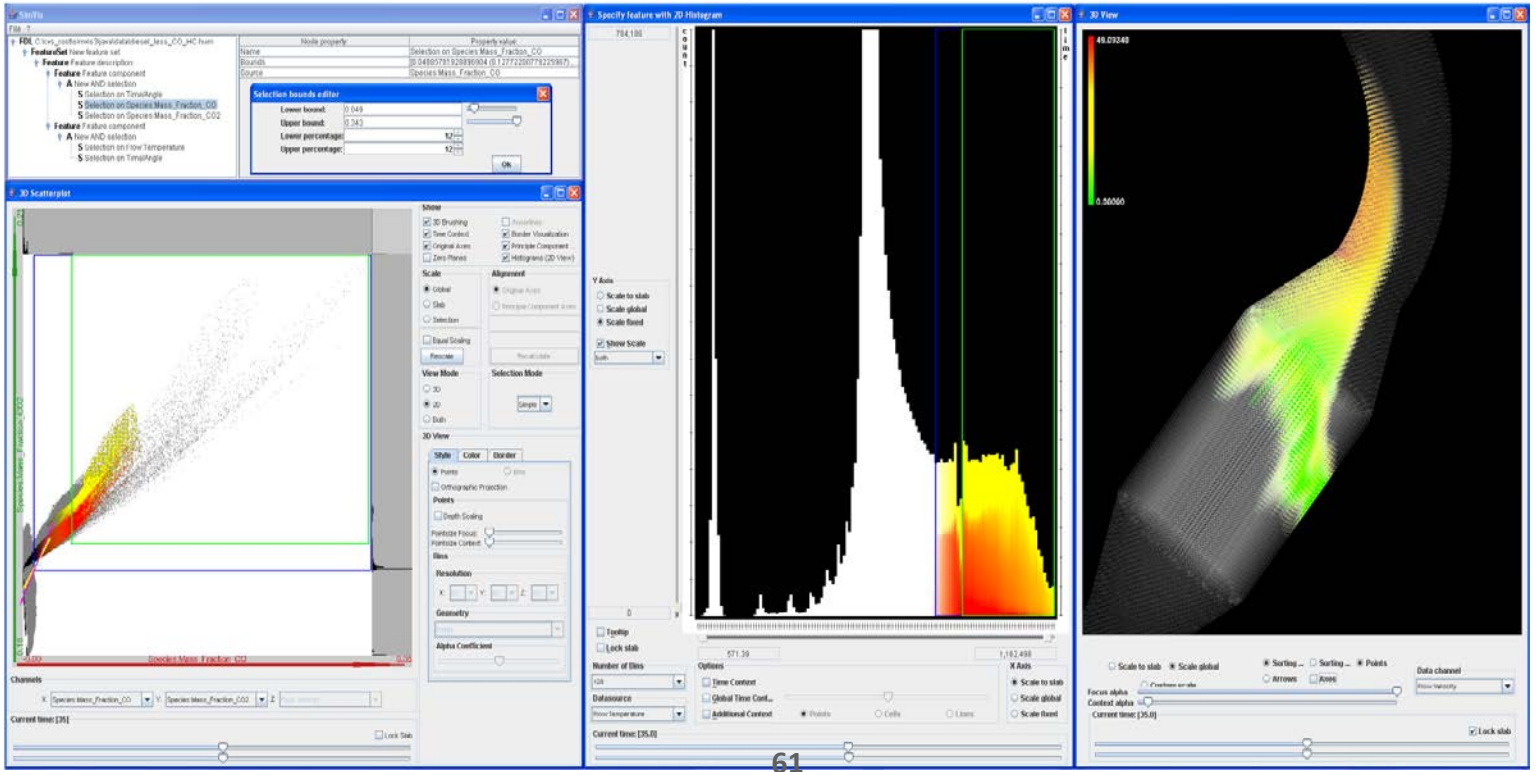
- More information about selected item



http://sanfrancisco.crimespotting.org/

Brushing

- Used with multiple coordinated views
- Highlighting one case from the Data Table selects the same case in other views
- Linking and Brushing



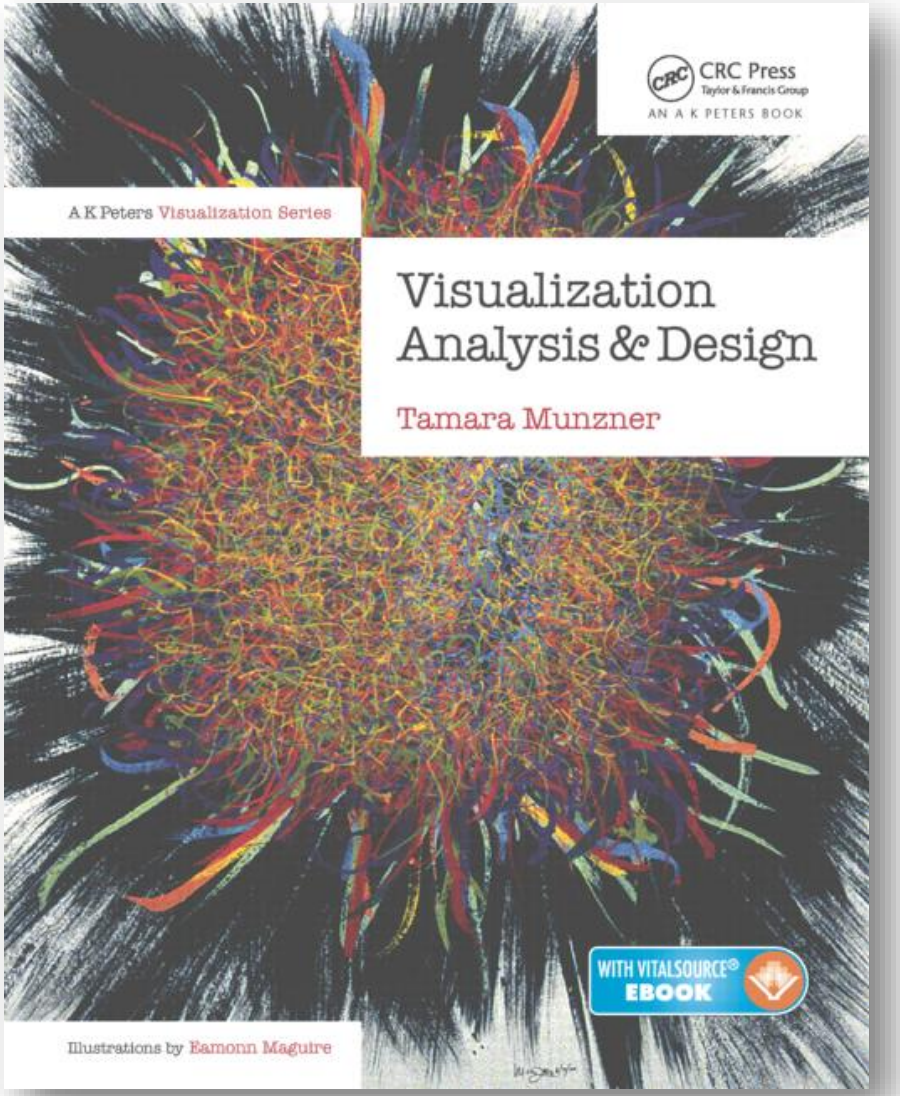
Books

[Munzner, 2014]

Tamara Munzner

Visual Analysis and Design

A K Peters / CRC Press 2014

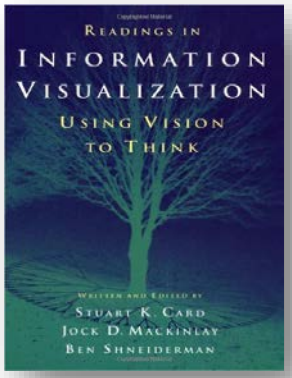


Books

Card, Mackinlay, Shneiderman

Readings in Information Visualization Using Vision to Think

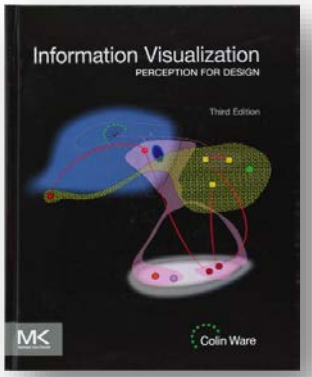
Morgan Kaufmann, 1999



Colin Ware

Information Visualization: Perception for Design

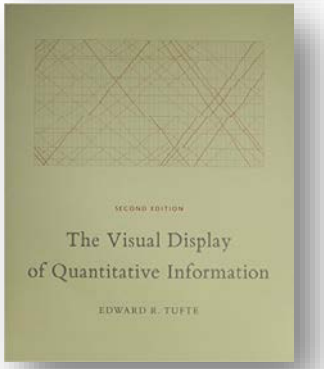
Morgan Kaufmann, 2012 (3rd edition)



Edward Tufte

The Visual Display of Quantitative Information

Graphics Pr., 2001 (2nd edition)



Interesting Links

- The **Information Visualization** community platform
  - http://www.infovis-wiki.net/index.php/Main_Page
- Google Public Data
  - http://www.google.com/publicdata/directory
- Hans Rosling – Gapminder
  - http://www.ted.com/speakers/hans_rosling
  - http://www.gapminder.org/
- D3 – data-driven documents
  - http://d3js.org/
- Visual complexity
  - http://www.visualcomplexity.com/vc/
- Further links
  - https://www.cg.tuwien.ac.at/courses/InfoVis/index.html

Surveys

- Visualization Techniques for Time-Oriented Data
  - http://survey.timeviz.net/
- Visual Bibliography of Tree Visualization 2.0
  - http://vcg.informatik.uni-rostock.de/~hs162/treeposter/poster.html
- Visualizing Dynamic Graphs
  - http://dynamicgraphs.fbeck.com/
- Visual Survey of Text Visualization Techniques
  - http://textvis.lnu.se/
- Visualizing Sets and Set-typed Data
  - http://www.cvast.tuwien.ac.at/~alsallakh/SetViz/literature/www/index.html
- Biological visualization tools
  - http://bivi.co/visualisations
- Visualizing High-Dimensional Data
  - http://www.sci.utah.edu/~shusenl/highDimSurvey/website/

Acknowledgments

With selected contributions by Marc Streit (JKU Linz)