



Information Visualisation

Computer Animation and Visualisation
Lecture 13

Taku Komura
tkomura@ed.ac.uk

Institute for Perception, Action & Behaviour
School of Informatics





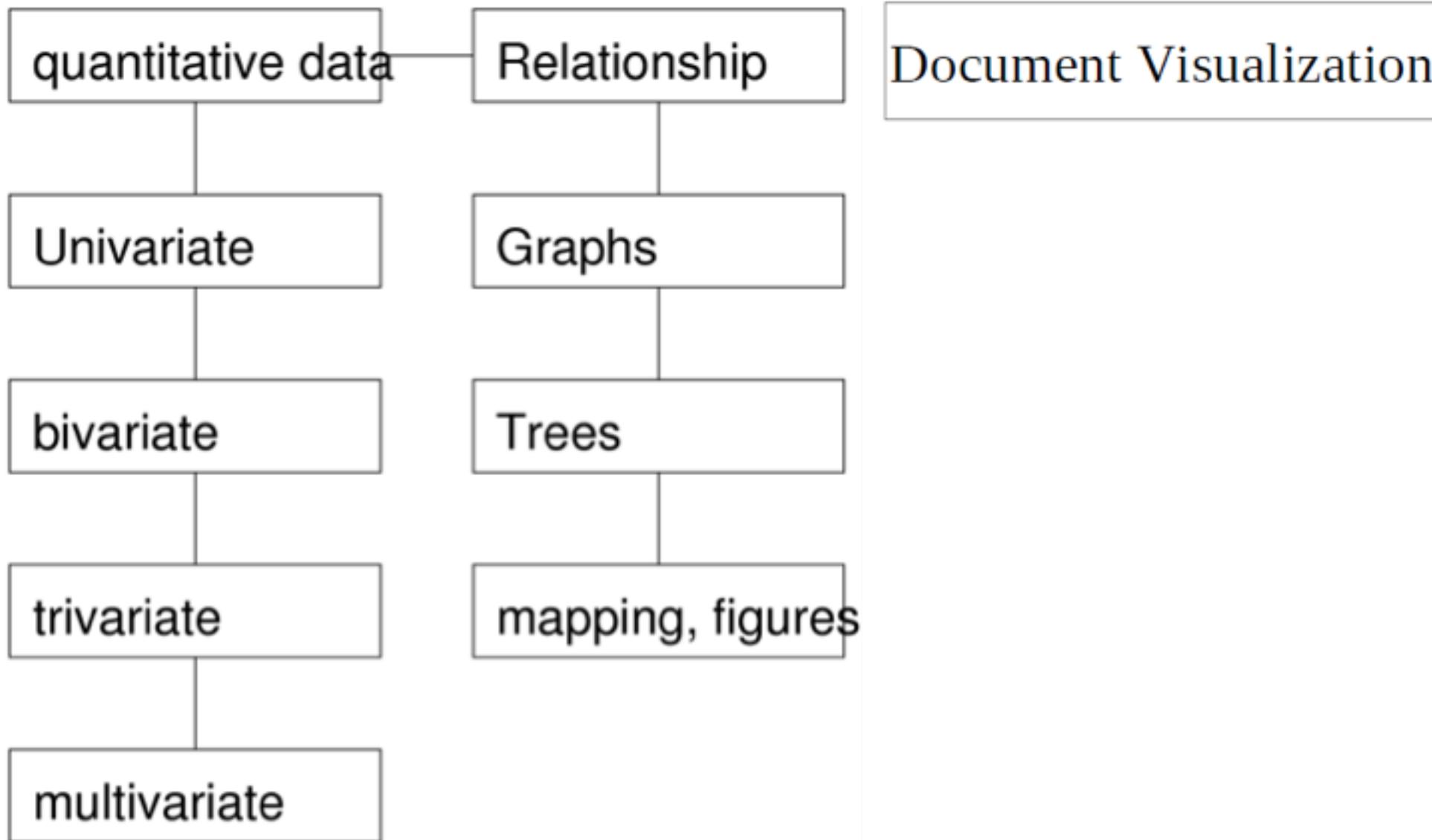
What is Information Visualisation ?

- Visualising ***discrete*** data with no or minor spatial information
- Visualisation of important information contained in abstract data types (trees, graphs)
 - Needs to be intuitive
 - Such that people can easily and quickly understand
- *Tools for*
 - *Extraction of information from the data*
 - *Discovery of new knowledge*





Data types

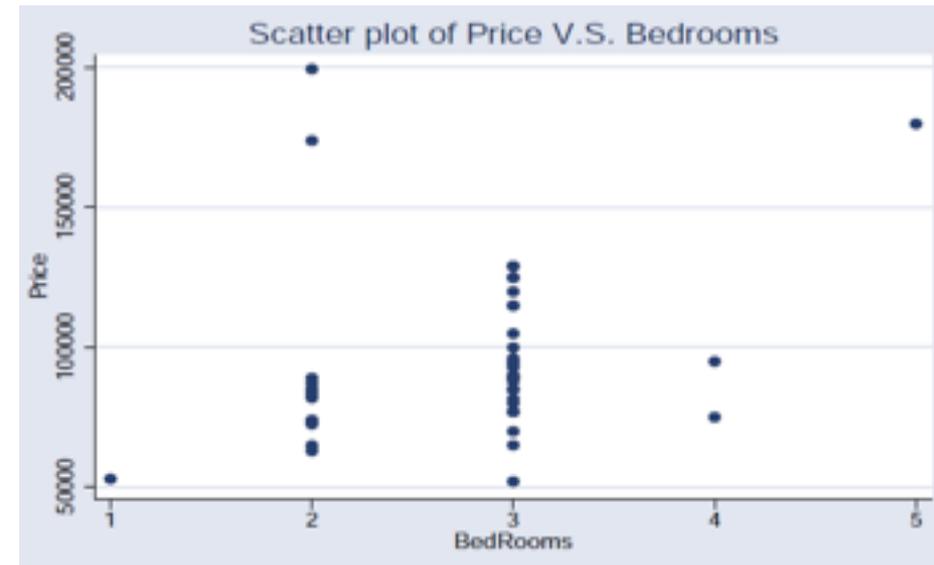
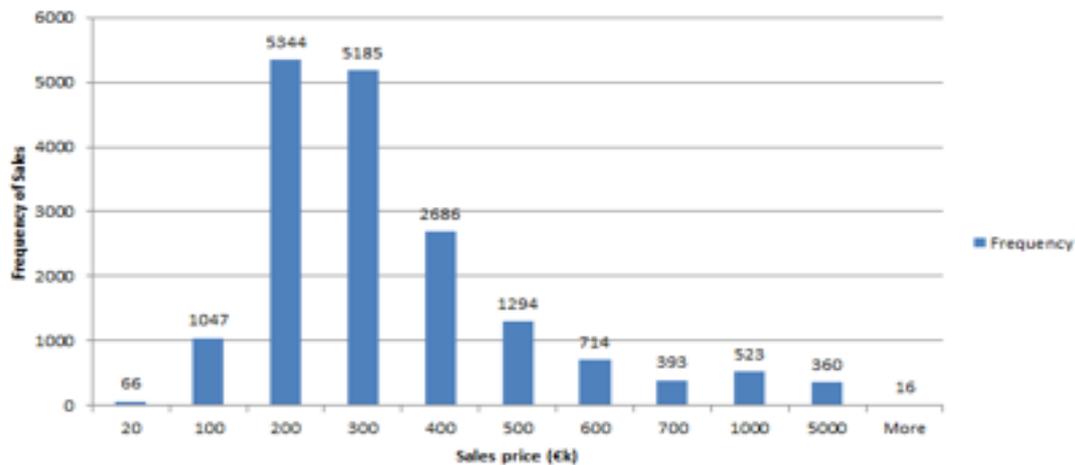




Univariate Data, Bivariate Data

- Data samples with one or two attributes
- Can use scatter plots, histograms

Dublin Property Sales 2010-2012 (17,628)

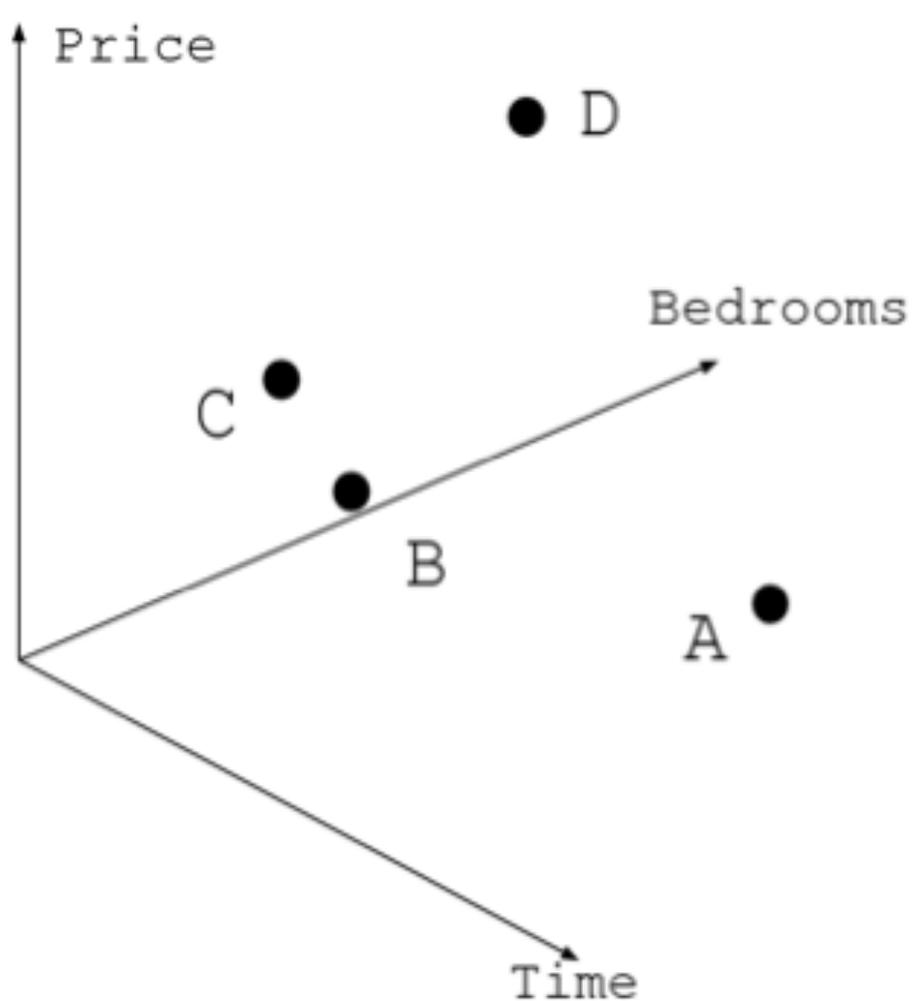




Trivariate Data

Scatterplots

Scatterplot matrix



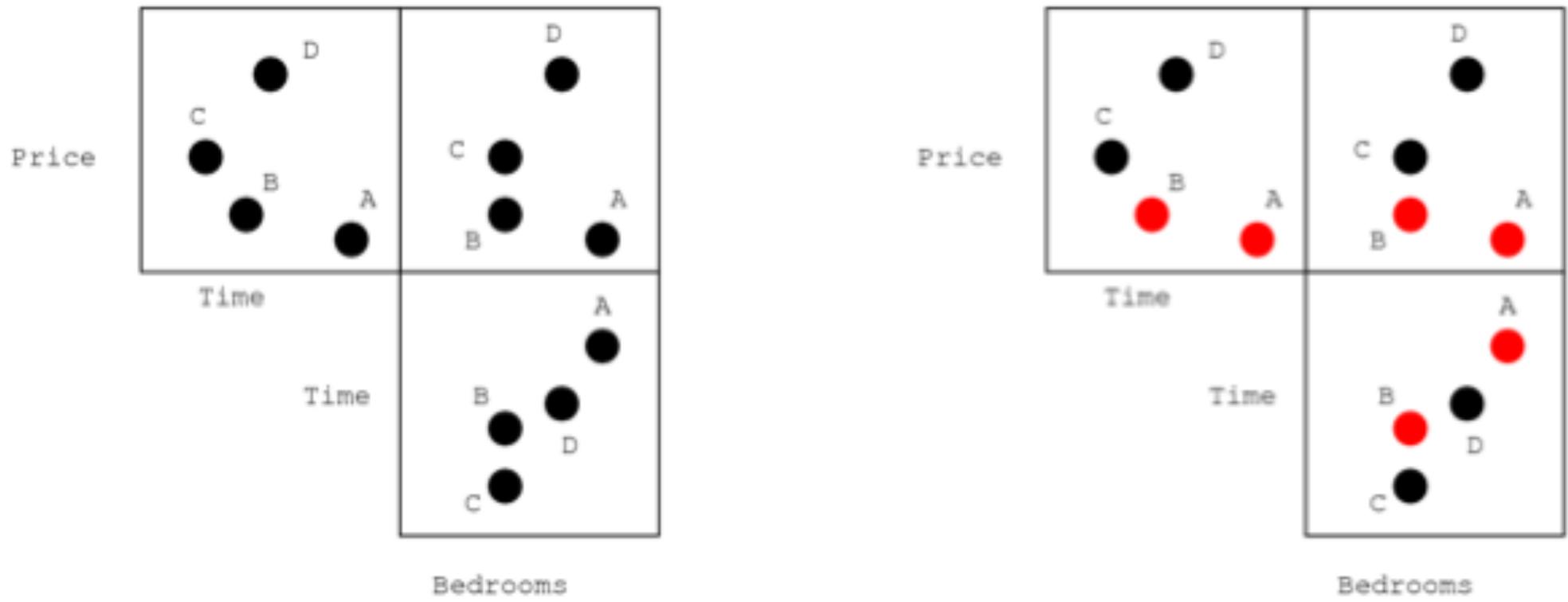
Not clear if D is more expensive
Than B and C





Trivariate Data

Scatterplot Matrix : Visualizing the relations of every two variables

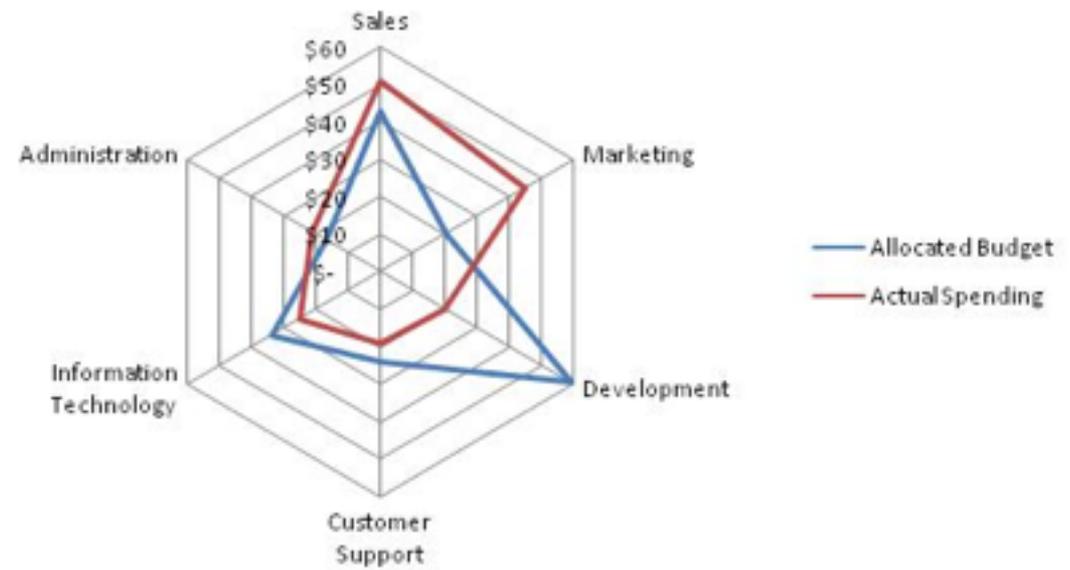
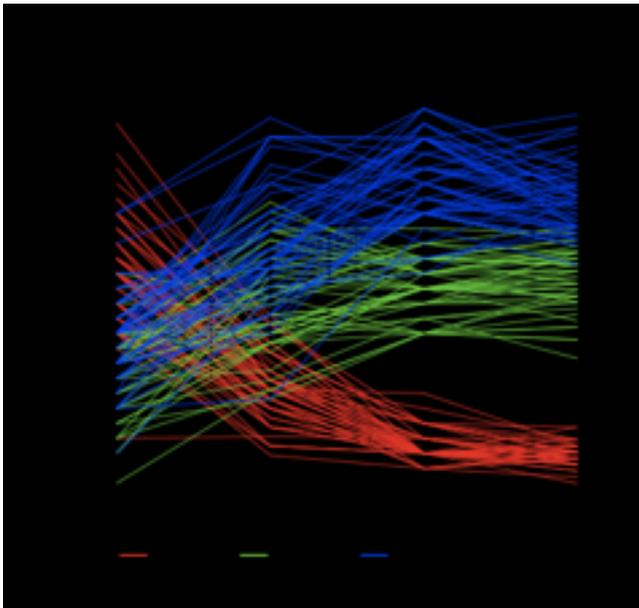




Multivariate Data

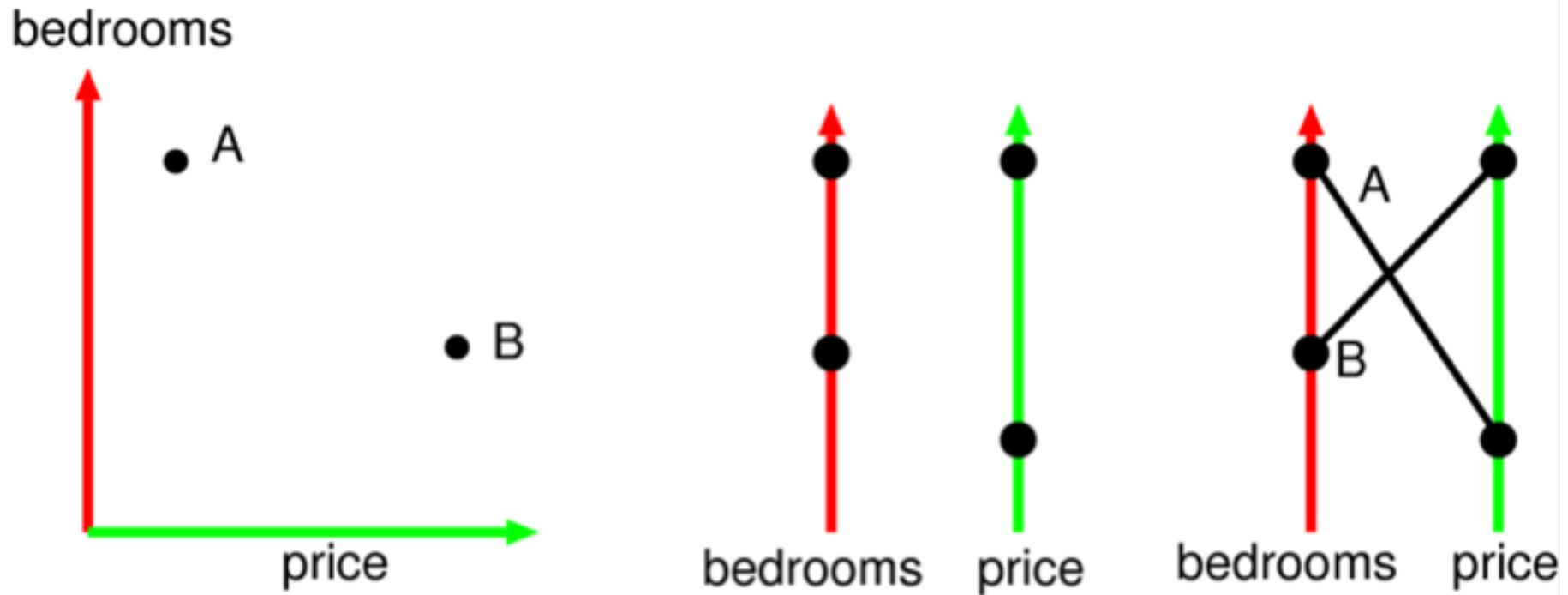
Parallel Coordinates

Star plots



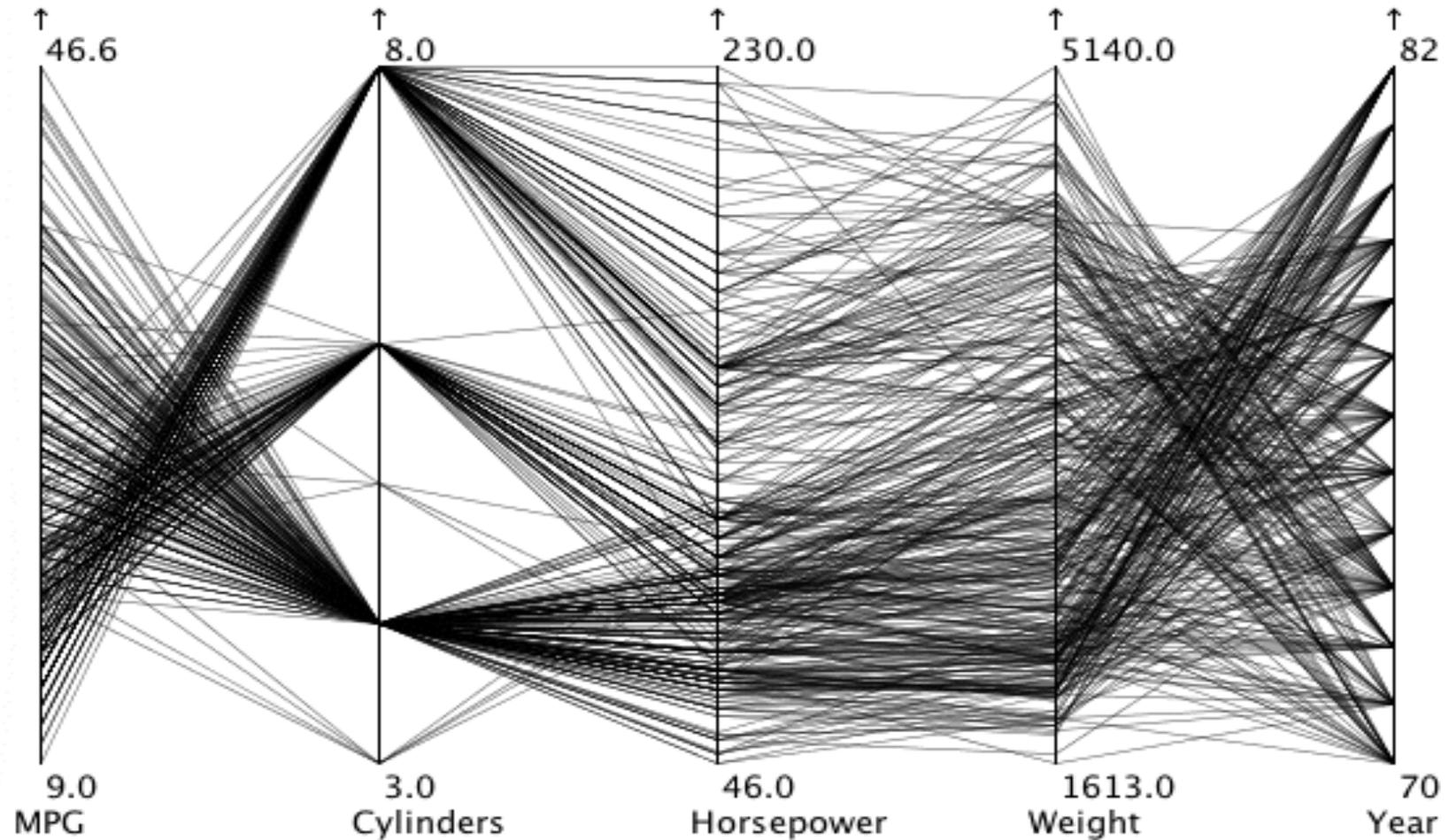


Parallel Coordinates





Parallel Coordinates



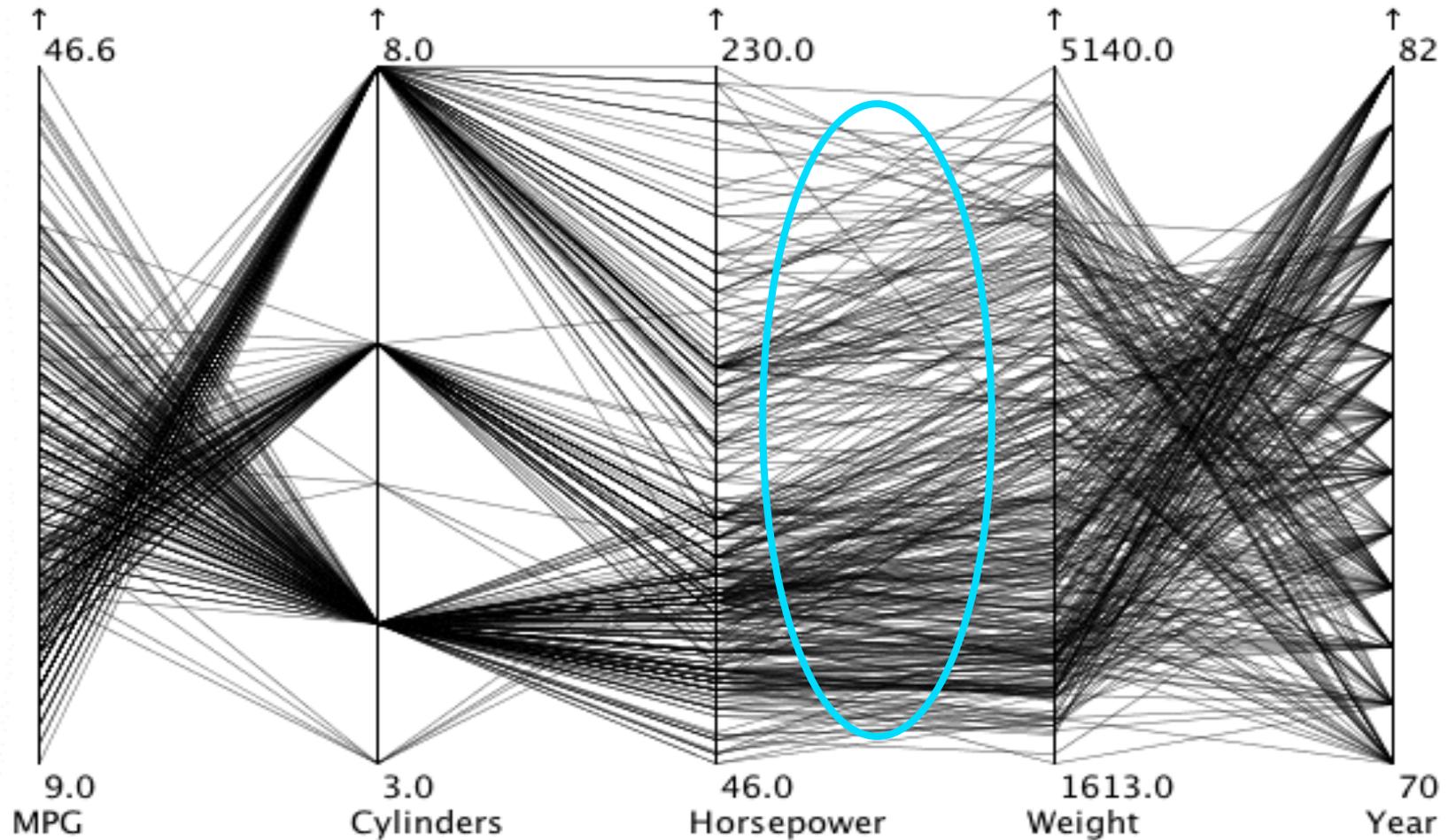
Car data :

<http://eagereyes.org/techniques/parallel-coordinates>





Parallel Coordinates

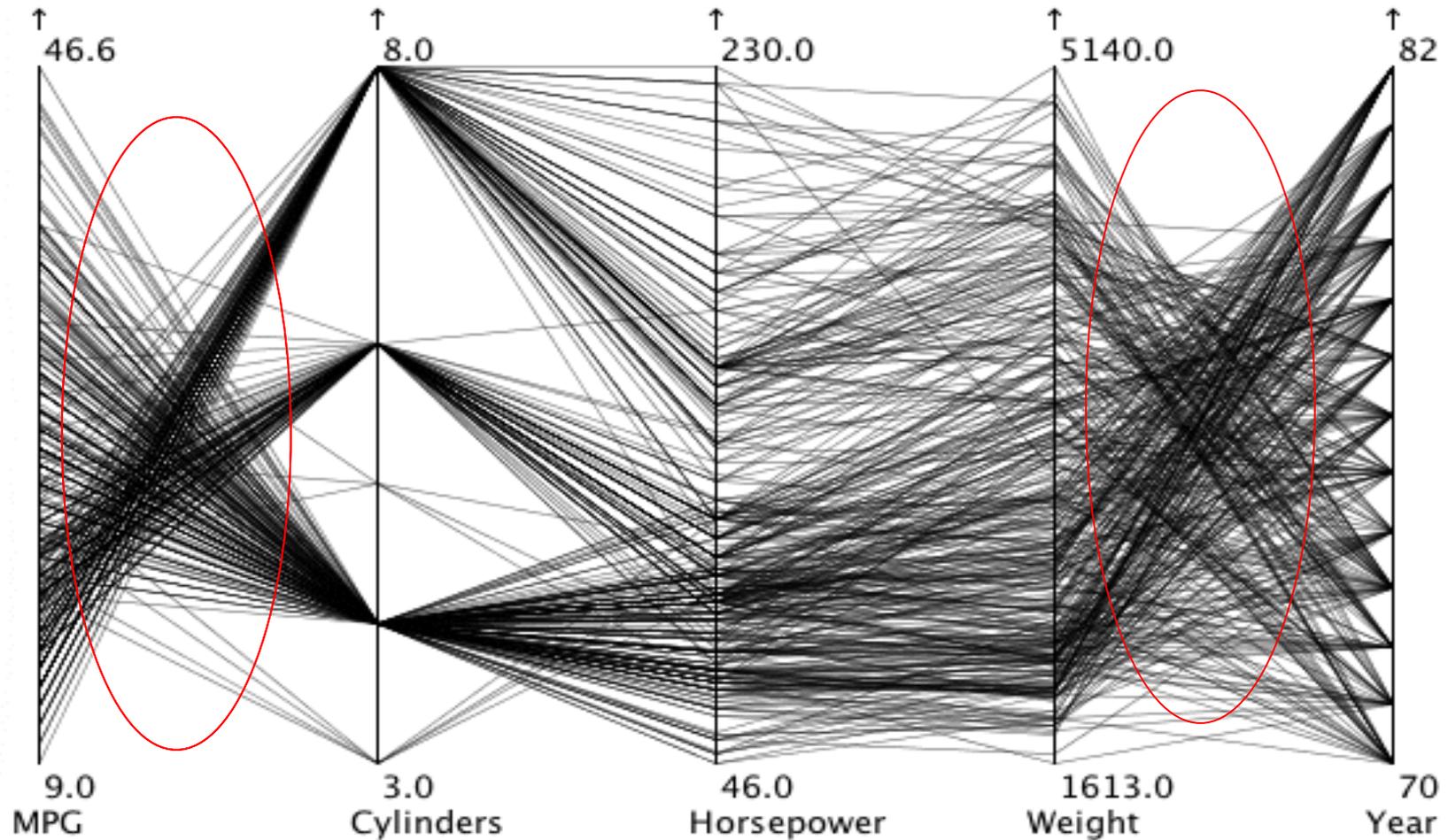


Direct correlation





Parallel Coordinates



Inverse Relations

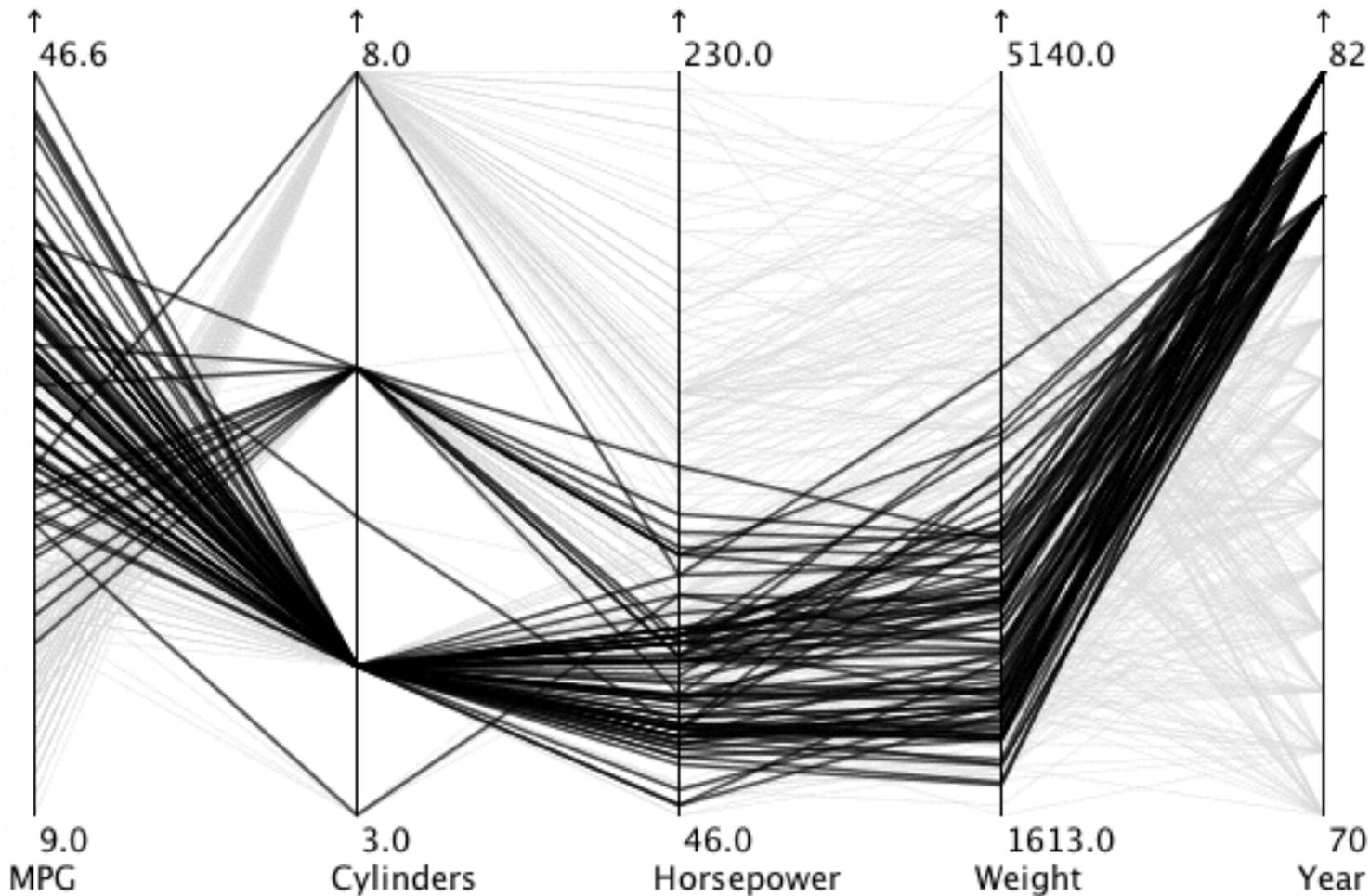




Brushing

Select some data using one of the coordinates

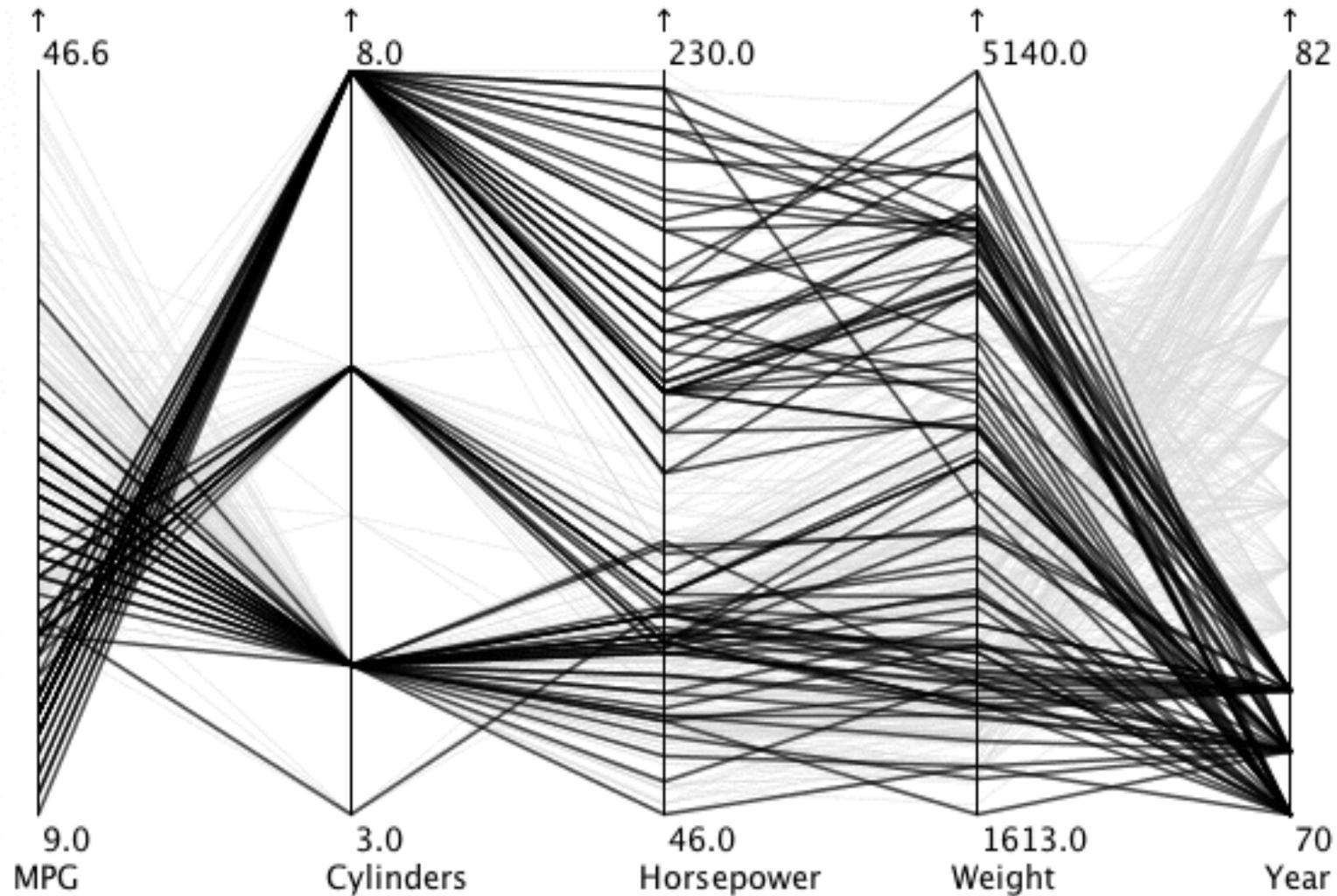
Brushing years 1980 to 1982





Brushing

Brushing the years 1970 to 1972

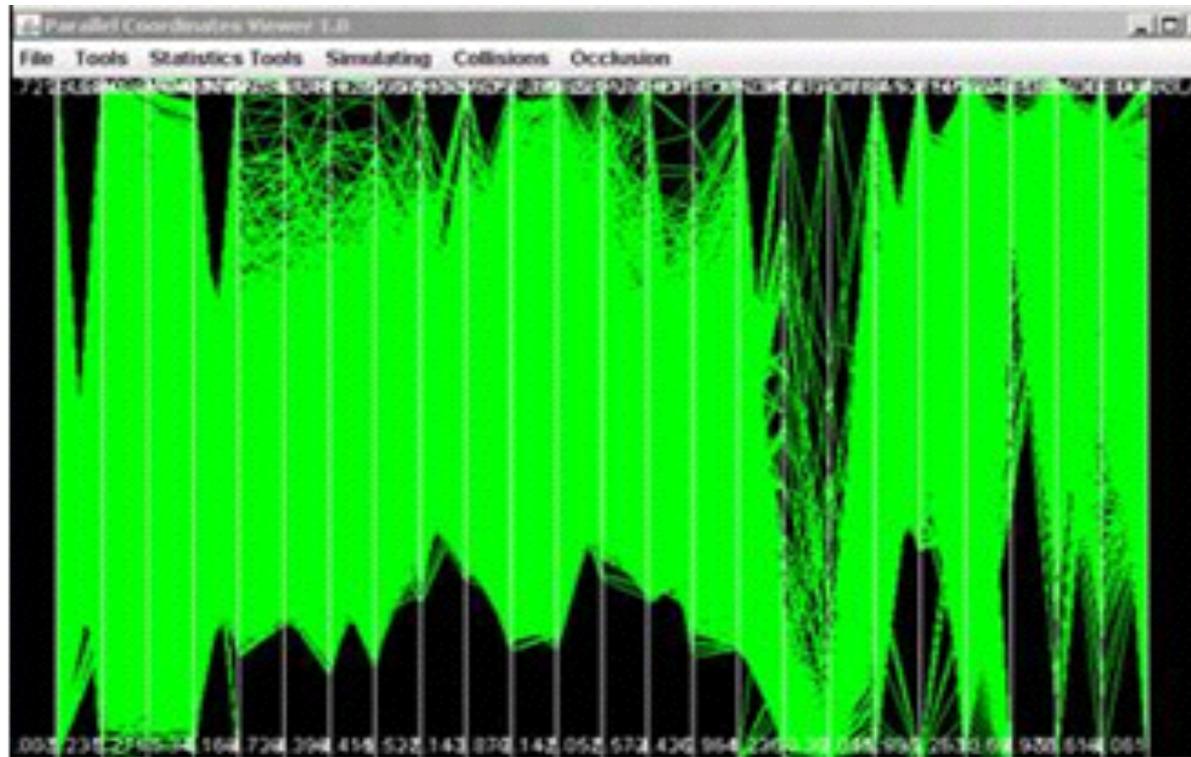




Limitations

Visual clutter

- Many lines cluttered together making it impossible to see anything
- Too many dimensions make things difficult to see



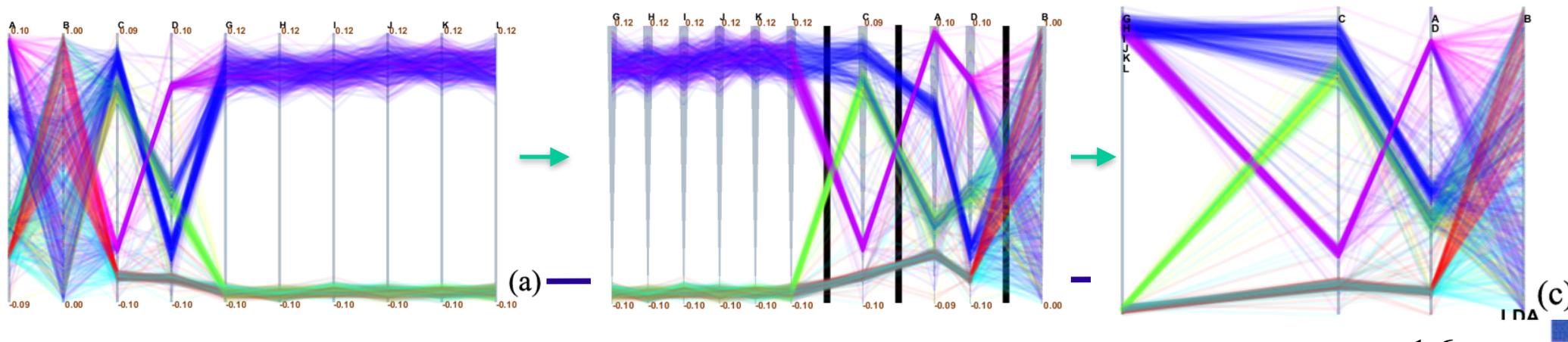


Solutions

- Re-ordering the axes to make the correlated coordinates adjacent to each other
- Negating the coordinates such that the inverse correlation becomes direct correlations
- Clustering the axes

We can use the biclustering algorithm that considers both the samples and the dimensions

Watanabe et al. 2015





K-Means Clustering

Computing clusters where the following criteria is minimized

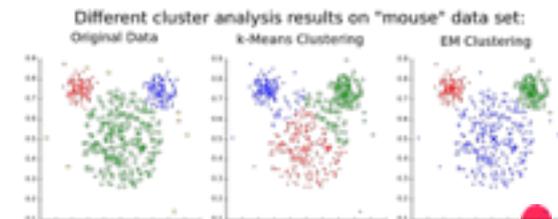
$$\sum_{i=1}^n \|r_i - \theta_{\kappa(i)}\|^2 = \sum_{i=1}^n \sum_{j=1}^d (x_{ij} - \theta_{\kappa(i)j})^2.$$

where

$\{r_1, \dots, r_n\}$, $r_i = (x_{i1}, \dots, x_{id}) \in \mathbf{R}^d$: samples

$\theta_k = (\theta_{k1}, \dots, \theta_{kd})$: cluster means

$\{\kappa(i) \in \{1, 2, \dots, K\}\}_{i=1}^n$: cluster labels





Biclustering

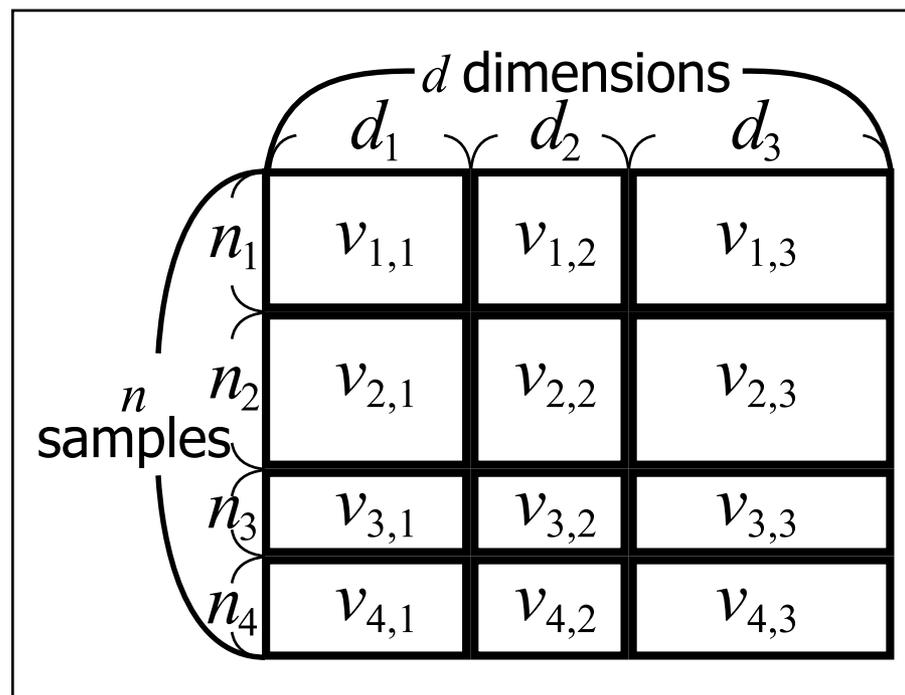
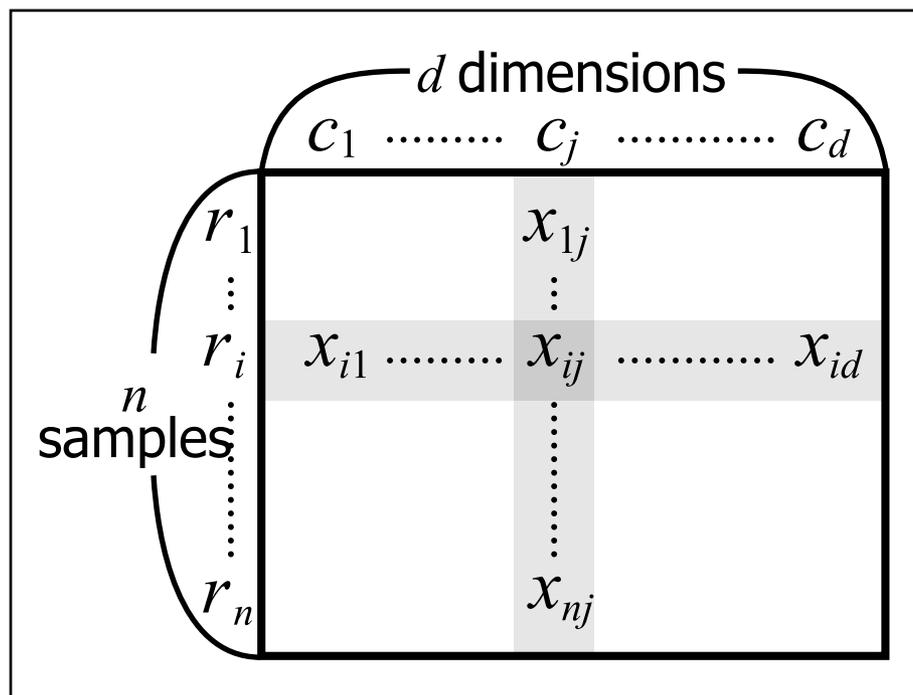
Clustering the samples and the dimensions so that the correlated dimensions are joined into the same dimension cluster

$$\sum_{i=1}^n \sum_{j=1}^d (x_{ij} - v_{\kappa(i), \lambda(j)})^2,$$

$\kappa(i) \in \{1, \dots, K\}$ ($i = 1, \dots, n$) : sample cluster

$\lambda(j) \in \{1, \dots, L\}$ ($j = 1, \dots, d$) : dimension cluster

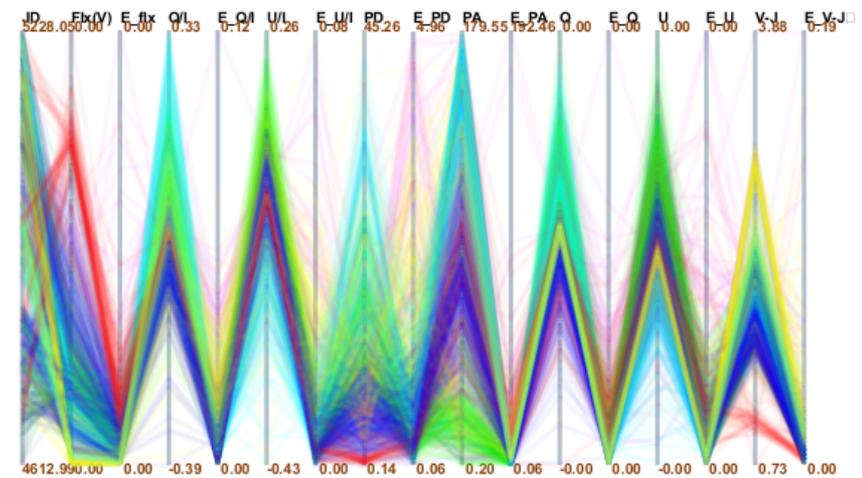
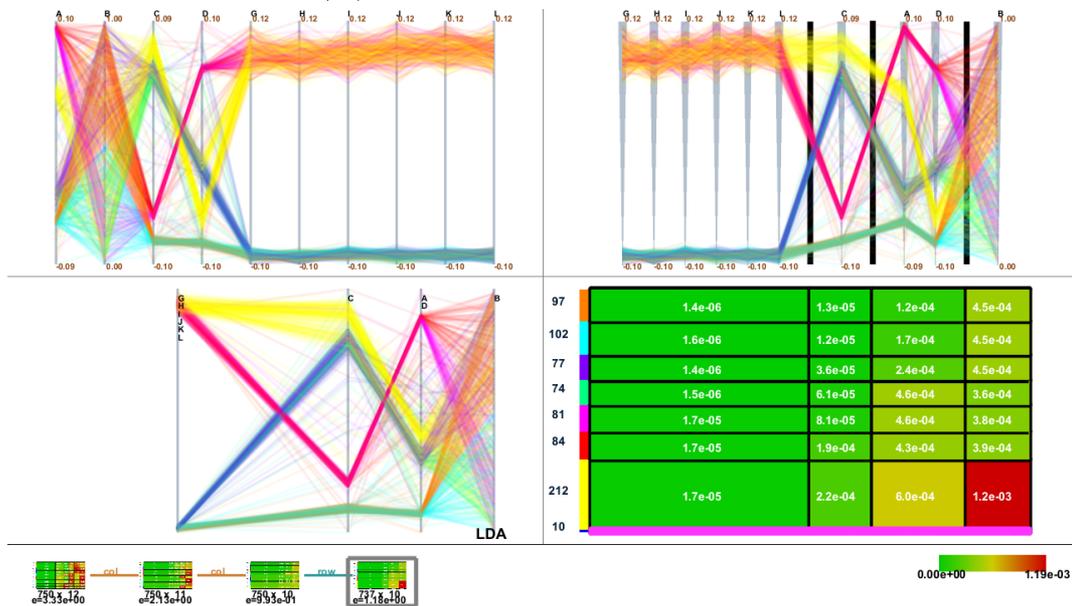




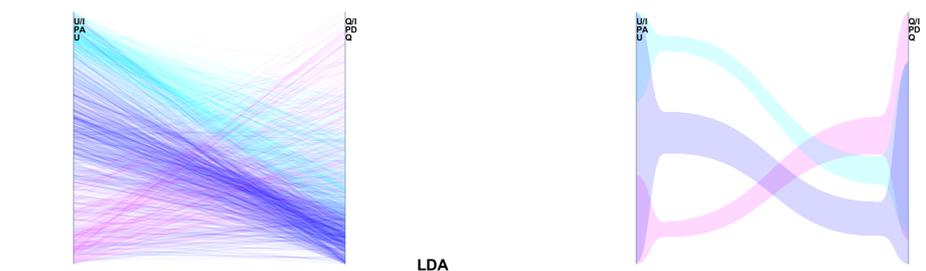


More results

(a) $K = 9$ and $L = 6$

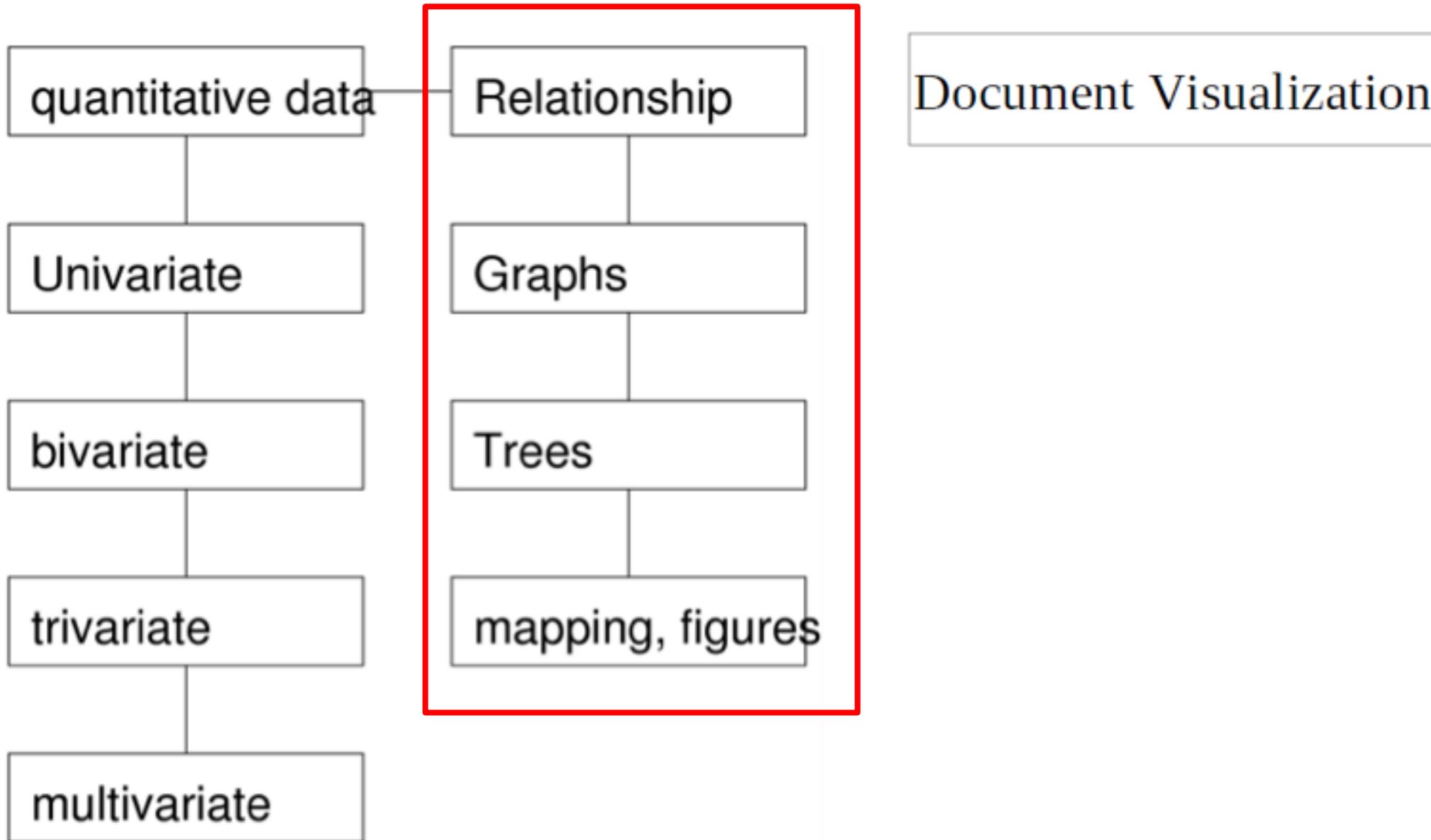


(a)





Data types

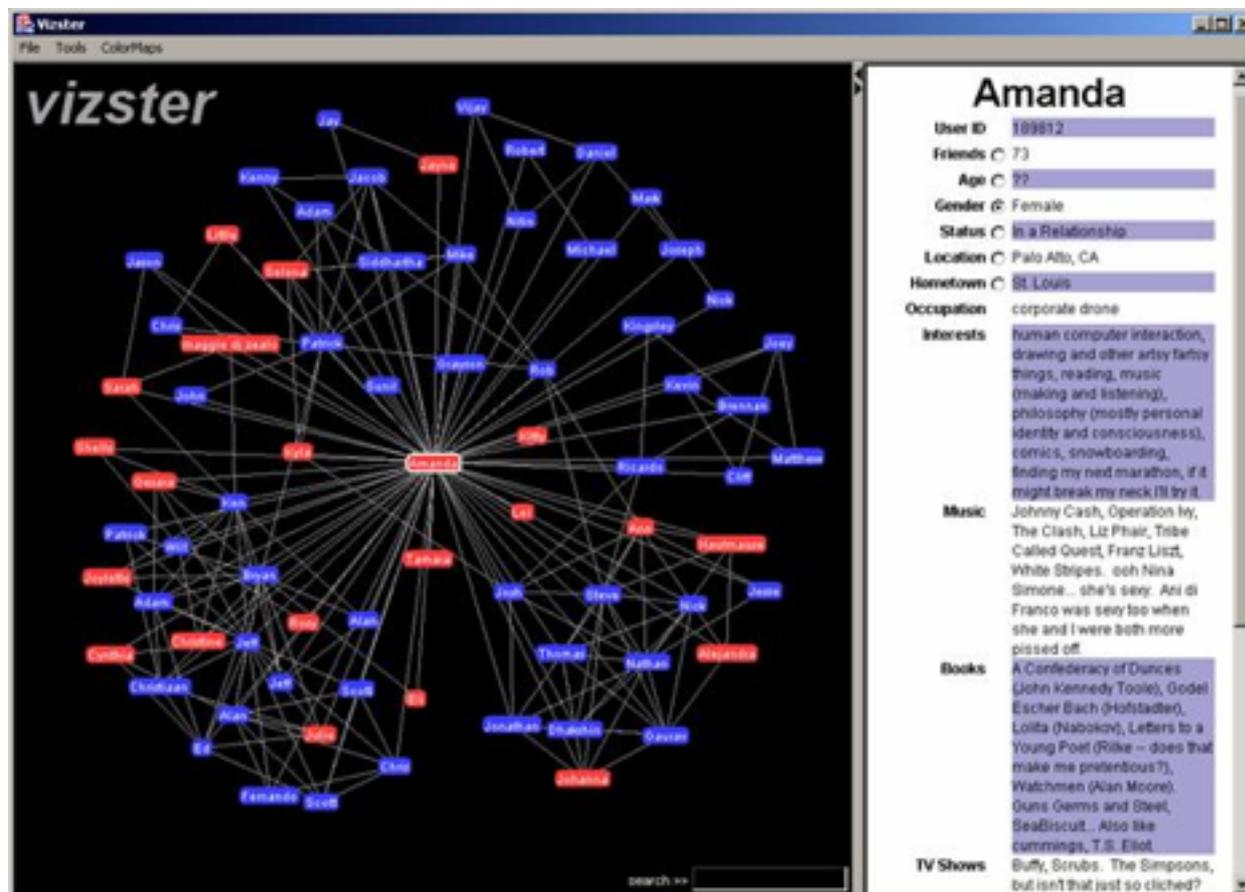




Visualizing Relations

Relation: A logical or natural association between two or more things; relevance of one to another; connection

Usually use lines to represent the relations



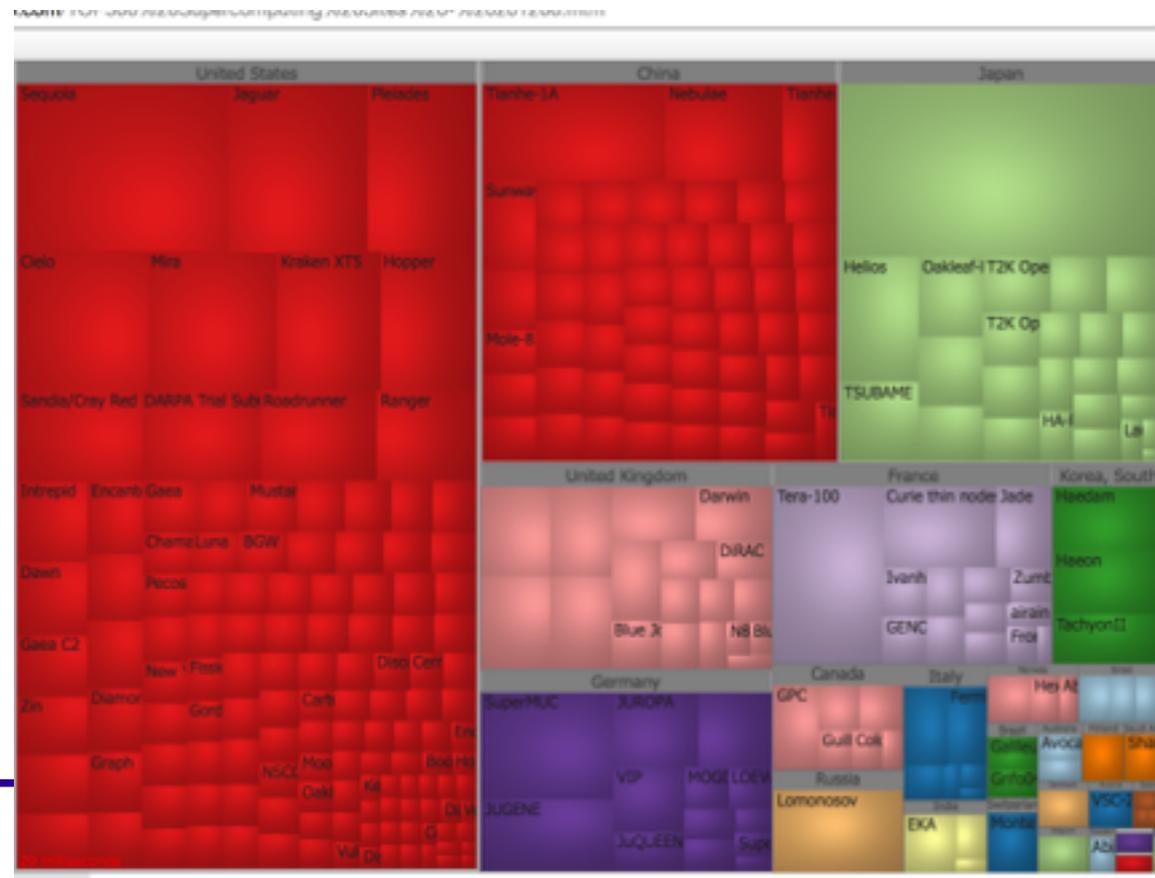


Treemaps

Display hierarchical (tree-structured) data as a set of nested rectangles

The area representing a scalar attribute

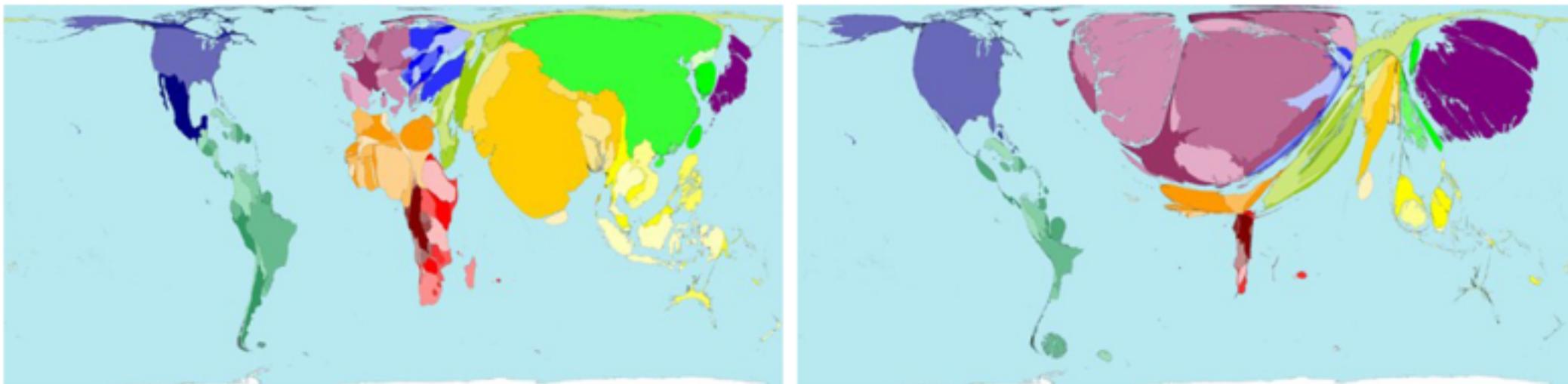
The leaf nodes are often colored to visualize another attribute data





Worldmapper

<http://sasi.group.shef.ac.uk/worldmapper/>



Distorted maps according to numbers: **Cartograms**



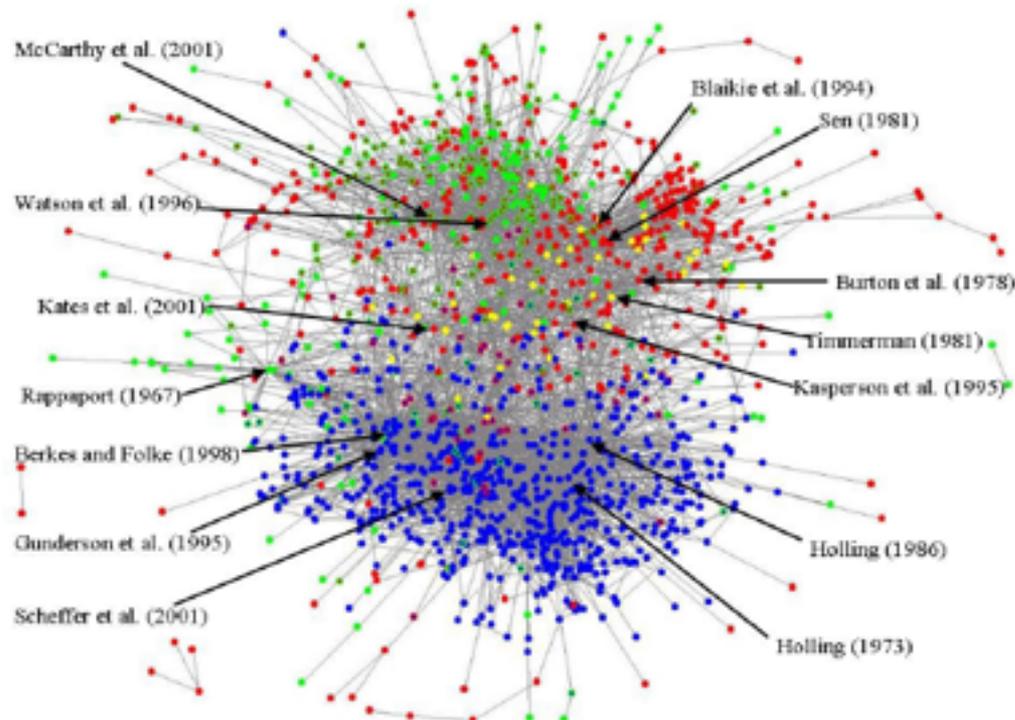


Graph Visualization

Visualizing correlation of different nodes

E.g.

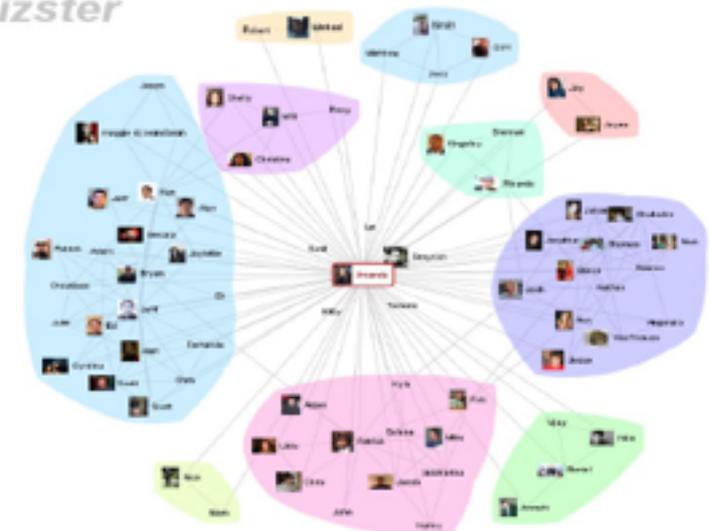
social networks
citation networks



vizster



vizster





Facebook relations





Facebook relations



“I defined weights for each pair of cities as a function of the Euclidean distance between them and the number of friends between them. Then I plotted lines between the pairs by weight, so that pairs of cities with the most friendships between them were drawn on top of the others. I used a color ramp from black to blue to white, with each line's color depending on its weight. I also transformed some of the lines to wrap around the image, rather than spanning more than halfway around the world.”





Formal Aesthetics

Minimize node-node / node-edge occlusions

Minimize edge crossings

Minimize edge bends

Maximize symmetry

Maximize the minimum angle between edges

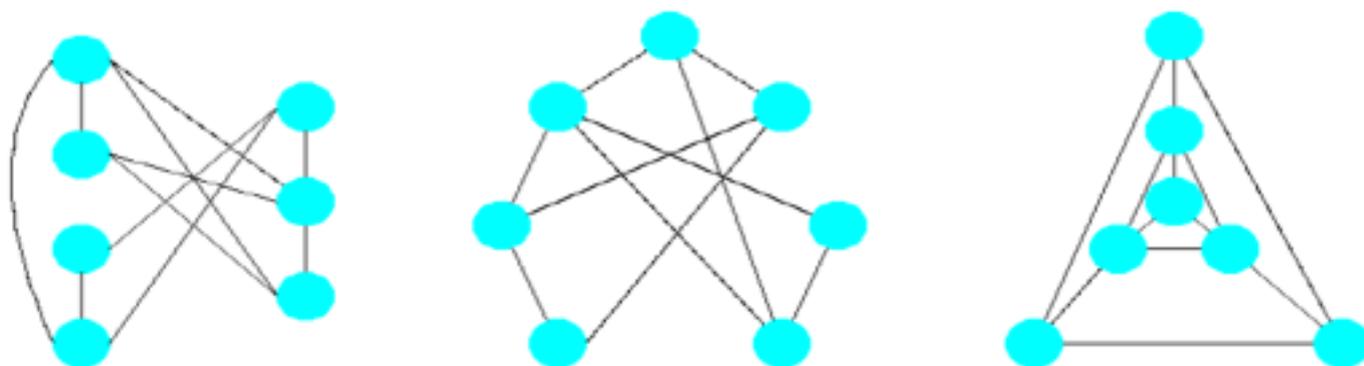
Maximize edge orthogonality



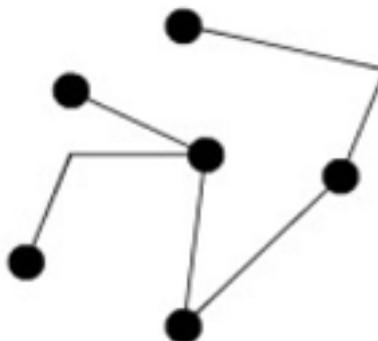


Formal Aesthetics Metrics

Minimize edge crossings



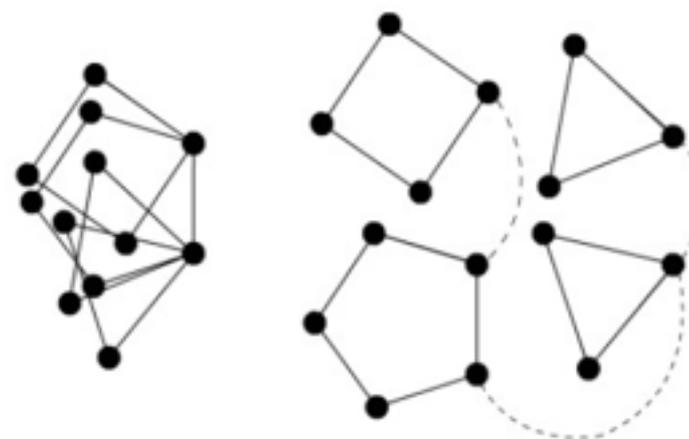
Minimize edge bends



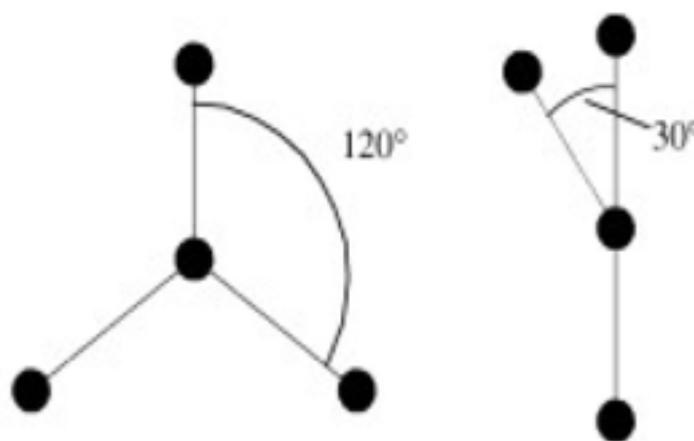


Formal Aesthetics Metrics

Maximizing symmetry



Maximizing the minimum angle between edges leaving a node





Another Application

metro maps : Wu et al. 2013

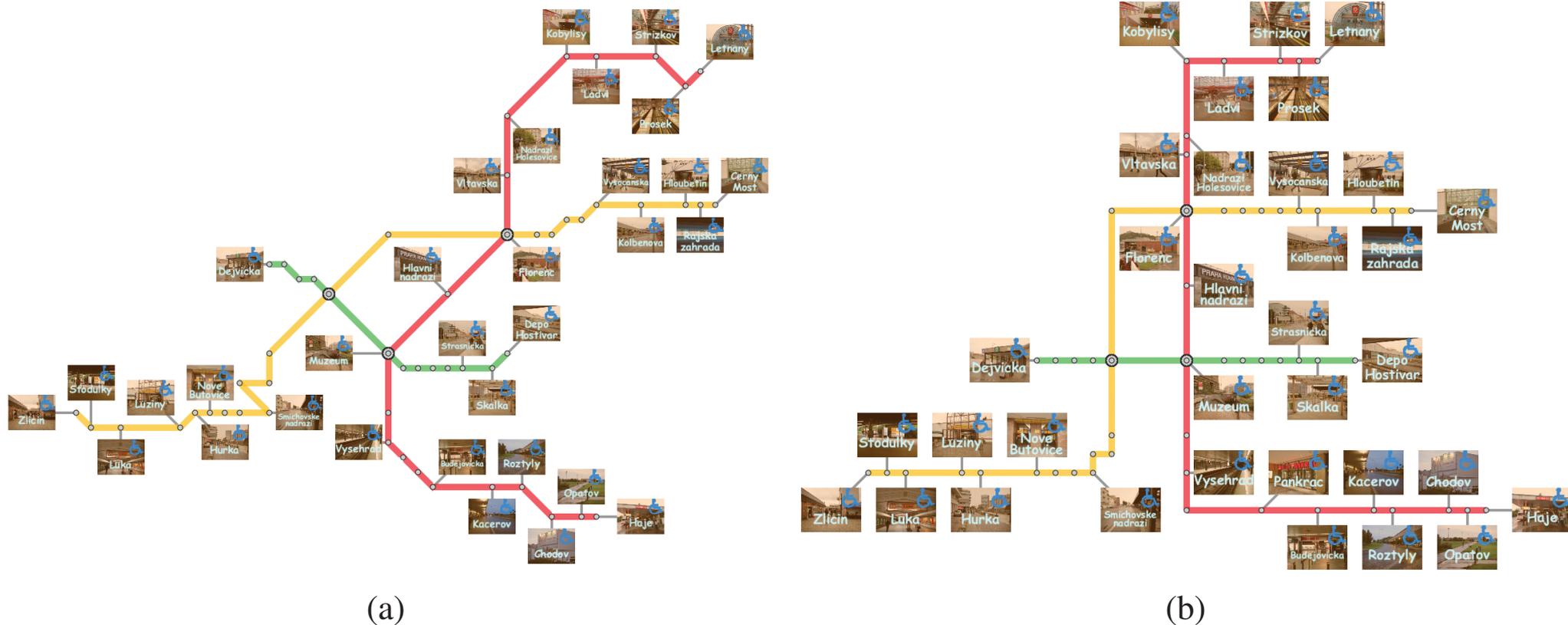
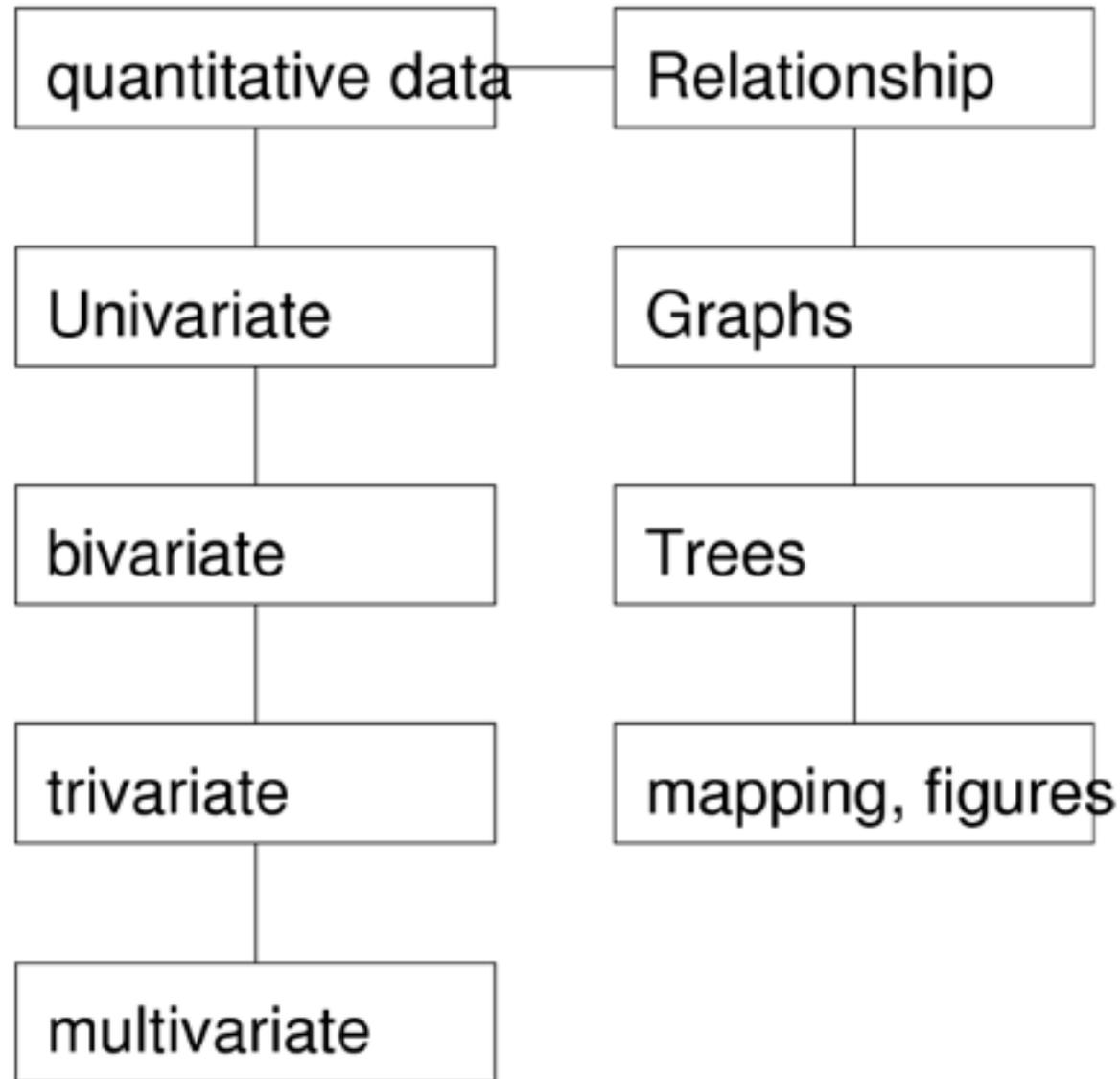


Figure 1: Wheelchair accessible stations in Prague metro. (a) Octilinear and (b) orthogonal annotated maps.



Data types



Document Visualization



Document Visualisation

- Motivation:

Action	Units of Information transfer
Typing at 10 bytes per second	1
Mouse Operations.	2
Reading	3-40
Hearing	60
Visualisation and Pattern Recognition	12,500

Source :
Silicon Graphics Inc.

- visualisation is considerably faster than hearing / reading!





Visualisation of Documents

- **Motivation : large bandwidth of human visual system**
 - 100s millions of documents available on-line
 - information only in textual form
- **‘Visualising the non-visual’**
 - searching for scientific papers
 - analysing witness statements
 - awareness of events in news bulletins





Document Visualisation - Stages

- **Representation of results**
 - form high-dimensional vector (one for each word, ~10000+)
 - cluster documents based on vector similarity (e.g. Nearest-Neighbour)
- **Visualisation of clustered results**
 - projection to lower dimensional space
 - 3D “galaxy” / 2D “theme-scape” / 1D “theme-river”

Query

“keywords” from user specification

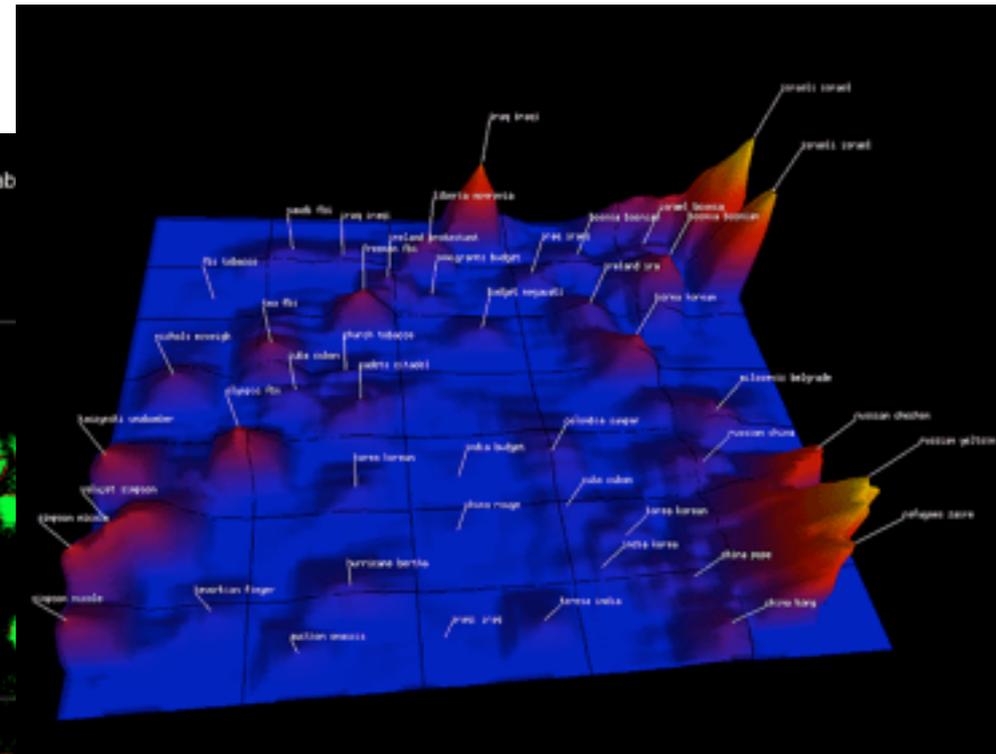
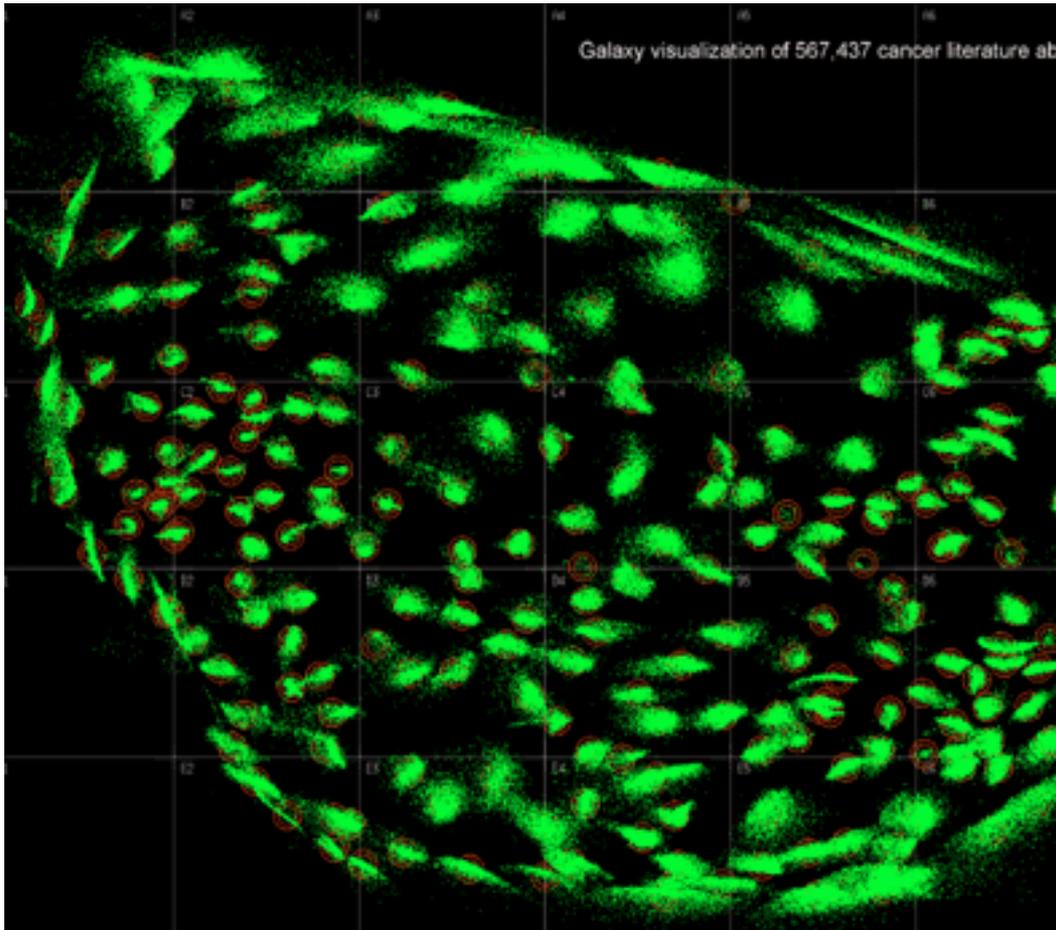
comparison to sample “reference” document





2D and 3D projections of documents

3D Visualisation of 567,000 cancer literature abstracts.



Pacific Northwest National Laboratory.





Multidimensional Scaling

A dimensionality reduction scheme useful for information visualization

Visualizing the information contained in a distance matrix

Computing the 2D coordinates of the projected samples such that the following stress is minimized

$$\text{Stress}_D(x_1, x_2, \dots, x_N) = \left(\sum_{i \neq j=1, \dots, N} (d_{ij} - \|x_i - x_j\|)^2 \right)^{1/2}$$

distance between sample i and j

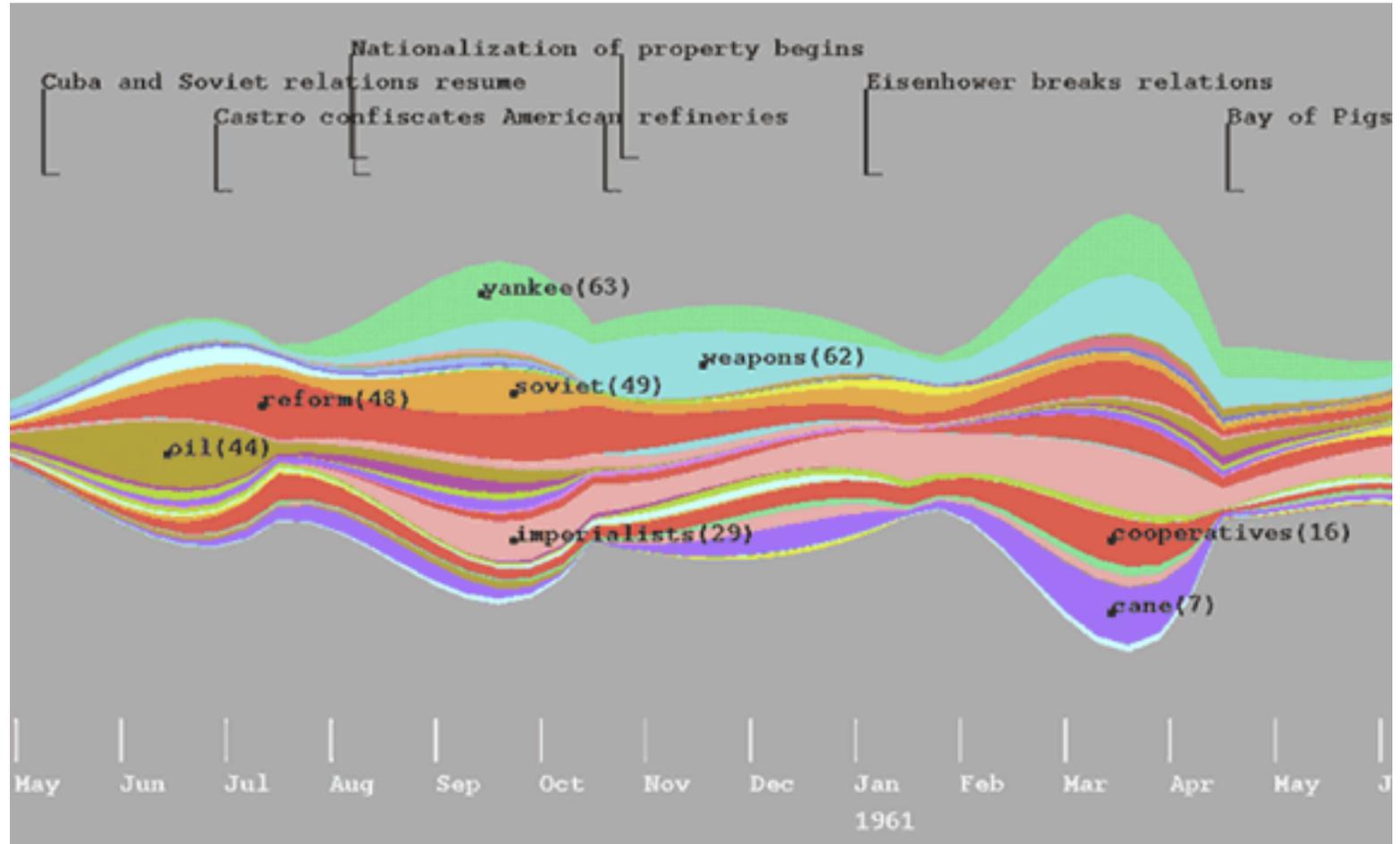
coordinates of i and j





1D visualisation of news articles

A 'Theme River' shows the relative importance of themes over the course of a year from press articles.



Pacific Northwest
National Laboratory.





Document Querying

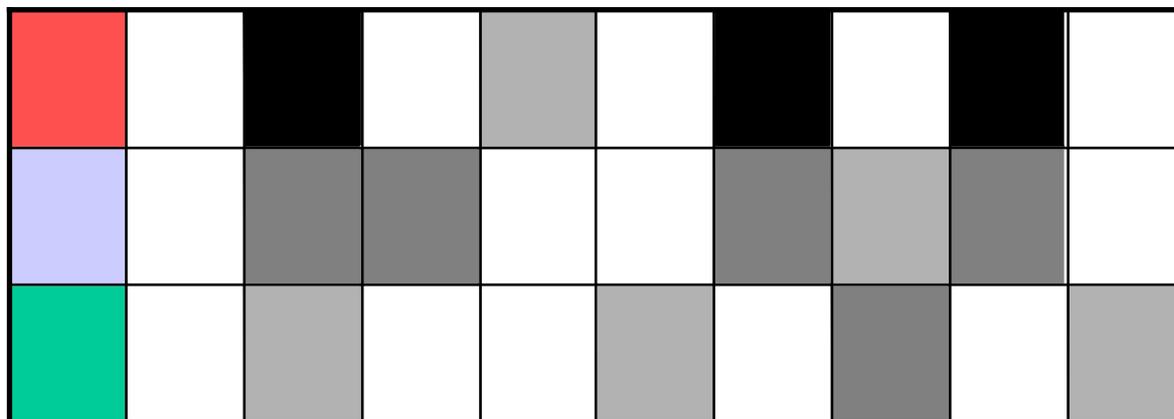
- **Keyword search is problematic**
 - **ambiguity**
 - ~7-18% of people describe same concept with same word (Barnard '91)
- Interested in
 - **distribution of keywords in the document**
 - **related articles** to the keyword entered
- **Tile bar scheme** (Hearst 1995)
 - display a list of documents with a **tile bar**
 - tile bar **shows the occurrence of keywords** in document





Title Bar Method

User query
Cancer
Prevention
Research



Columns **represent** paragraphs or pages in a document.
Shade indicates relevance shown by word occurrence.
 Shows length and likely relevance.
 System allows **interactivity** by clicking on box.

- Visualisation - Use of document topology / colour-mapping / interaction





Example : Title Bar Query / Result

Query terms: DBMS (Database Systems)

Reliability

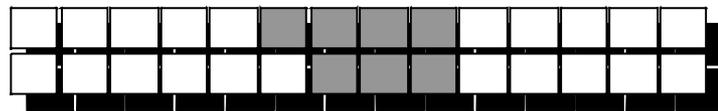
What roles do they play in retrieved documents?



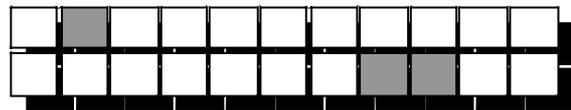
Mainly about both DBMS
& reliability



Mainly about DBMS, discusses
reliability



Mainly about, say, banking, with
a subtopic discussion on
DBMS/Reliability



Mainly about
something different





Summary

- Information Visualisation
 - Univariate, bivariate, trivariate, multi-variate data
 - Relations visualized by lines, tree visualization
 - Document visualization





Reading

- *Marti A. Hearst* **TileBars: Visualization of Term Distribution Information in Full Text Information Access**
- **Collins, Christopher**; Carpendale, Sheelagh; and Penn, Gerald. DocuBurst: Visualizing Document Content using Language Structure. Computer Graphics Forum (Proceedings of Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis '09)), 28(3): pp. 1039-1046, June, 2009
- Westin et al. '02, "Processing and visualization for diffusion tensor MRI"
- Watanabe et al., "Biclustering Multivariate Data for Correlated Subspace Mining", PacVis 2015
- Wu et al. Spatially Efficient Design of Annotated Metro Maps, EuroVis 2013
- <http://faculty.uoit.ca/collins/research/docuburst/index.html>
- http://searchuserinterfaces.com/book/sui_ch11_text_analysis_visualization.html

