



Andreas Holzinger
VO 709.049 Medical Informatics
11.01.2017 11:15-12:45



Lecture 09 Interactive Visualization and Visual Analytics

a.holzinger@tugraz.at

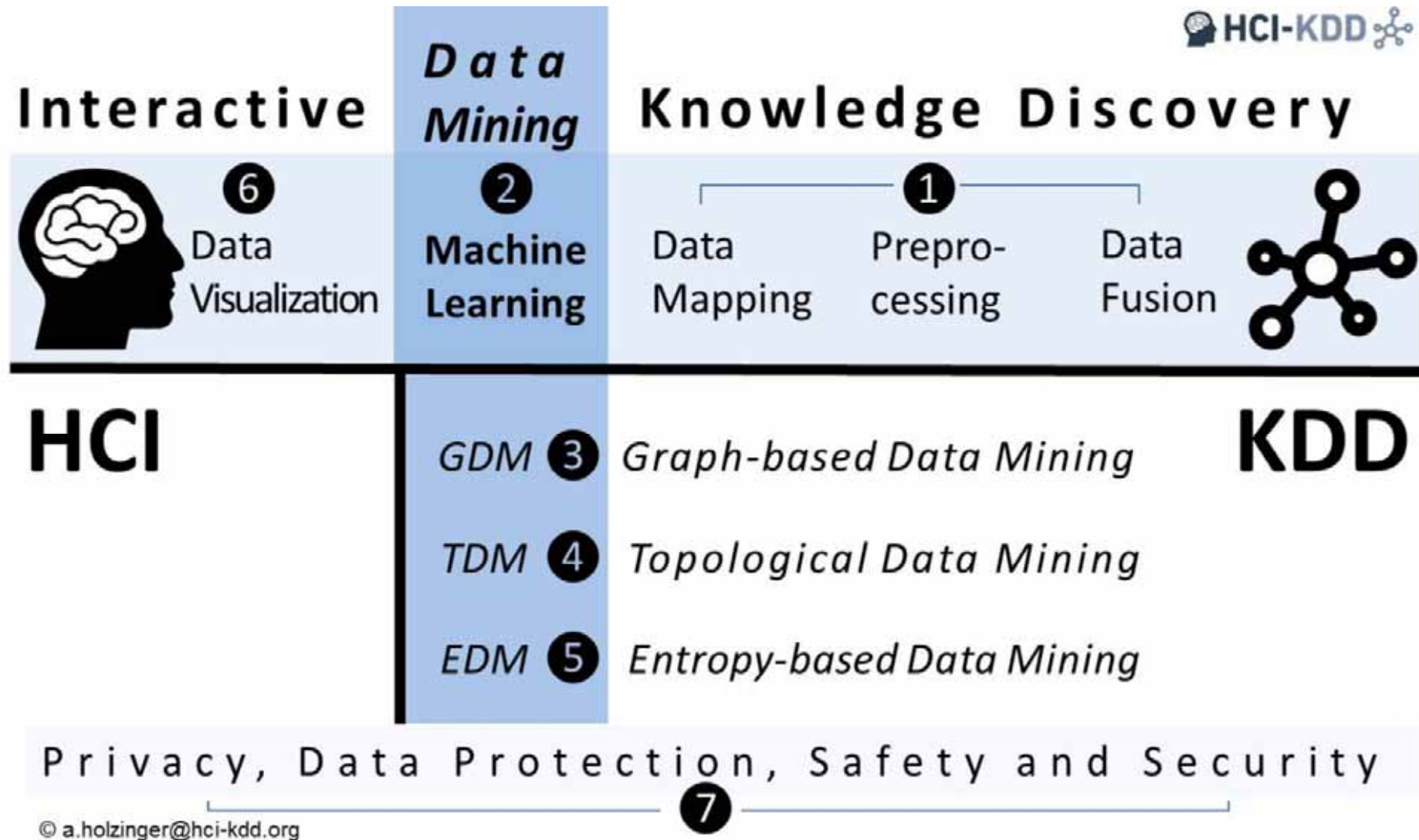
Tutor: markus.plass@student.tugraz.at

<http://hci-kdd.org/biomedical-informatics-big-data>

 R^2 

HCI-KDD

 R^n



- Data/Information/Knowledge visualization
- Flow cytometry
- Human-Computer Interaction (HCI)
- Information visualization
- Interactive information visualization
- k-Anonymization
- Longitudinal data
- Multivariate data
- Parallel coordinates
- RadViz
- Semiotics
- Star plots
- Temporal data analysis
- Visual analytics
- Visual information

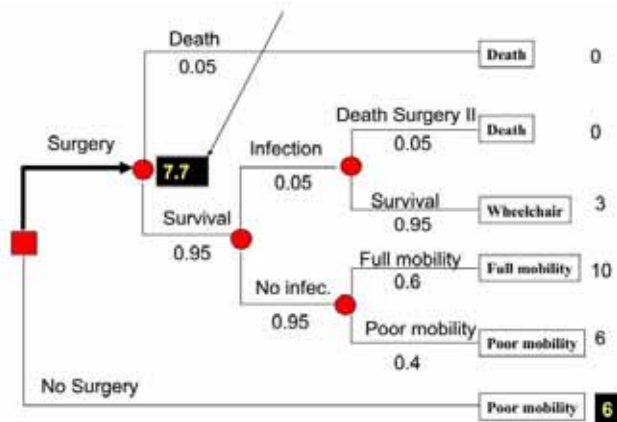
- **Biological data visualization** = as branch of bioinformatics concerned with visualization of sequences, genomes, alignments, phylogenies, macromolecular structures, systems biology, etc.
- **Clustering** = Mapping objects into disjoint subsets to let appear similar objects in the same subset;
- **Data visualization** = visual representation of complex data, to communicate information clearly and effectively, making data useful and usable;
- **Information visualization** = the interdisciplinary study of the visual representation of large-scale collections of non-numerical data, such as files and software, databases, networks etc., to allow users to see, explore, and understand information at once;
- **Multidimensional scaling** = Mapping objects into a low-dimensional space (plane, cube etc.) in order to let appear similar objects close to each other;
- **Multi-Dimensionality** = containing more than three dimensions and data are multivariate;
- **multivariate** = encompassing the simultaneous observation and analysis of more than one statistical variable; (Antonym: univariate = one-dimensional);

- **Parallel Coordinates** = for visualizing high-dimensional and multivariate data in the form of N parallel lines, where a data point in the n -dimensional space is transferred to a polyline with vertices on the parallel axes;
- **RadViz** = radial visualization method, which maps a set of m -dimensional points in the 2-D space, similar to Hooke's law in mechanics;
- **Semiotics** = deals with the relationship between symbology and language, pragmatics and linguistics. Information and Communication Technology deals not only in words and pictures but also in ideas and symbology;
- **Semiotic engineering** = a process of creating a semiotic system, i.e. a model of human intelligence and knowledge and the logic for communication and cognition;
- **Star Plot** = aka radar chart, spider web diagram, star chart, polygon plot, polar chart, or Kiviat diagram, for displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point;
- **Visual Analytics** = focuses on analytical reasoning of complex data facilitated by interactive visual interfaces;
- **Visualization** = a method of computer science to transform the symbolic into the geometric, to form a mental model and foster unexpected insights;

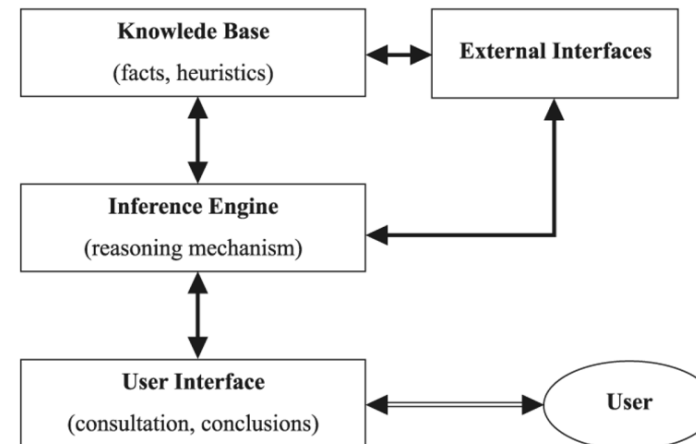
- ... have some background on visualization, visual analytics and content analytics;
- ... got an overview about various possible visualization methods for multivariate data;
- ... got an introduction into the work of and possibilities with parallel coordinates;
- ... have seen the principles of RadViz mappings and algorithms;
- ... are aware of the possibilities of Star Plots;
- ... have seen that visual analytics is intelligent Human-Computer Interaction at it finest;

- **00 Reflection – follow-up from last lecture**
- **01 Verbal vs. Visual Information**
- **02 Informatics as Semiotics Engineering**
- **03 Visualization Definitions**
- **04 Usefulness of Visualization**
- **05 Visualization Methods**
(long chapter but incomplete!)

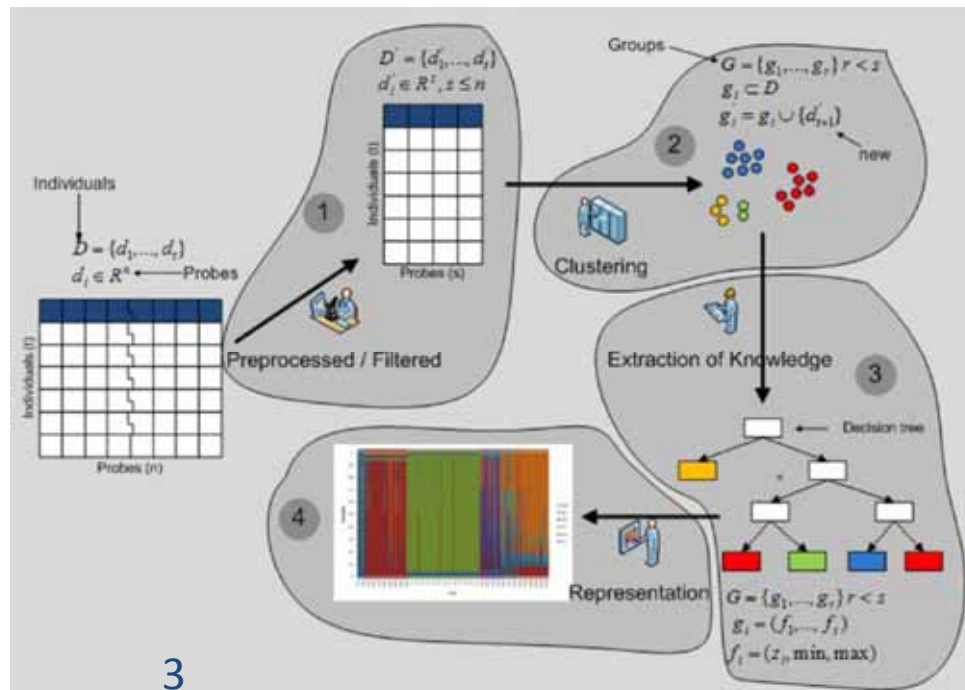




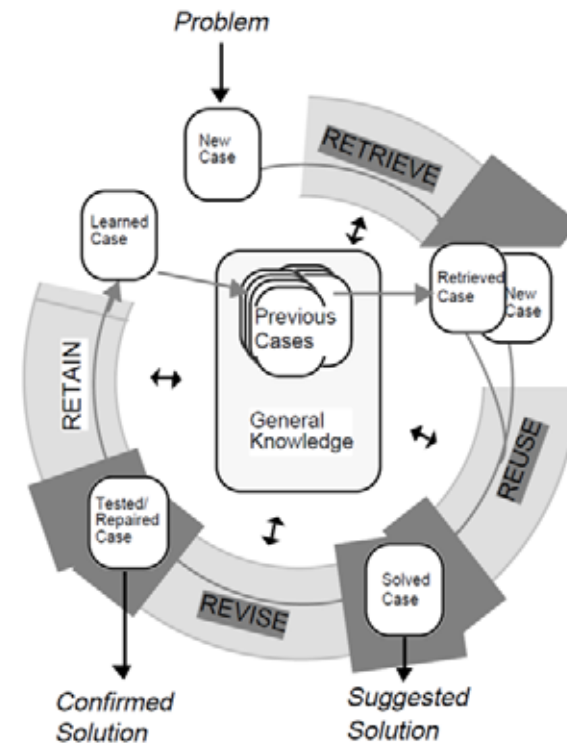
1



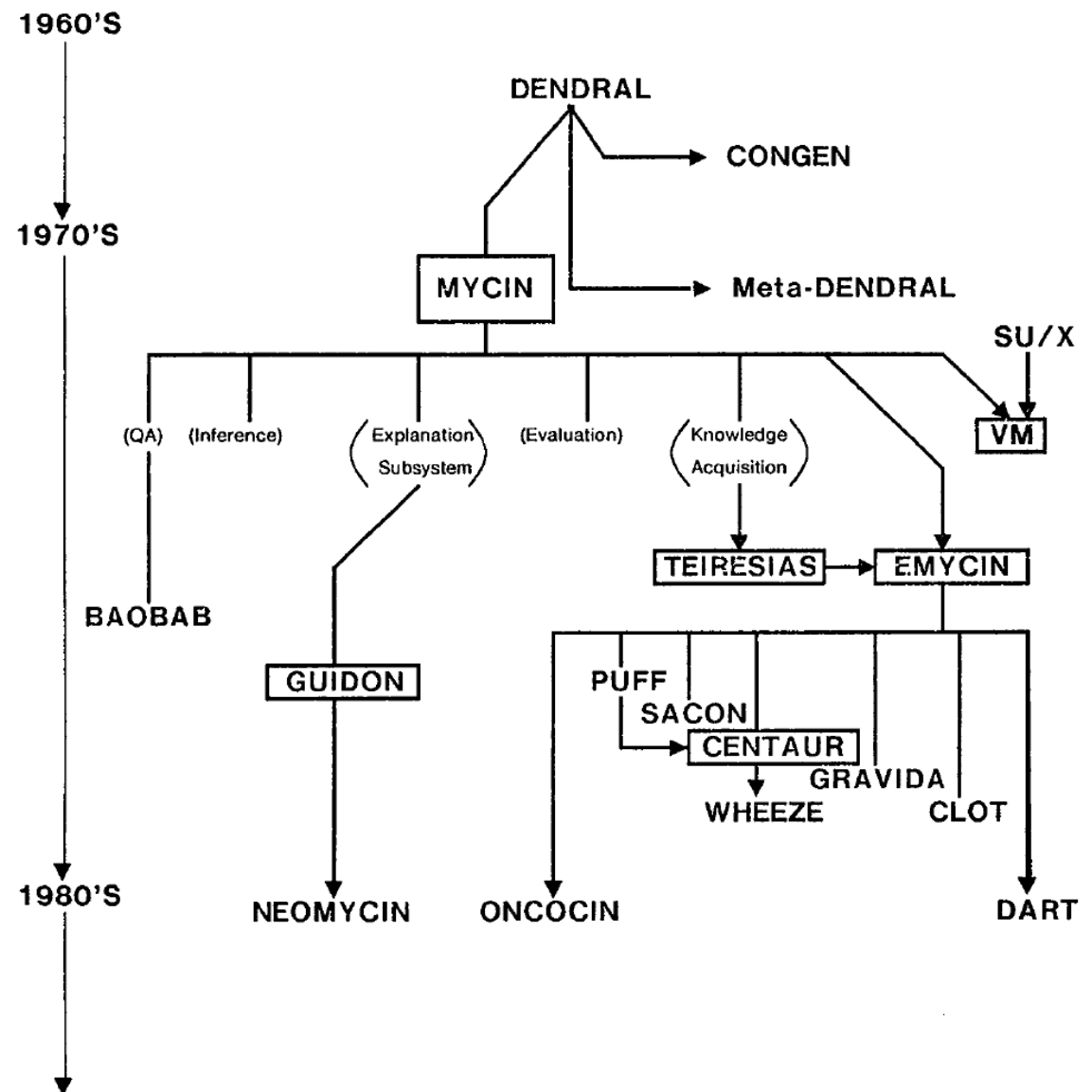
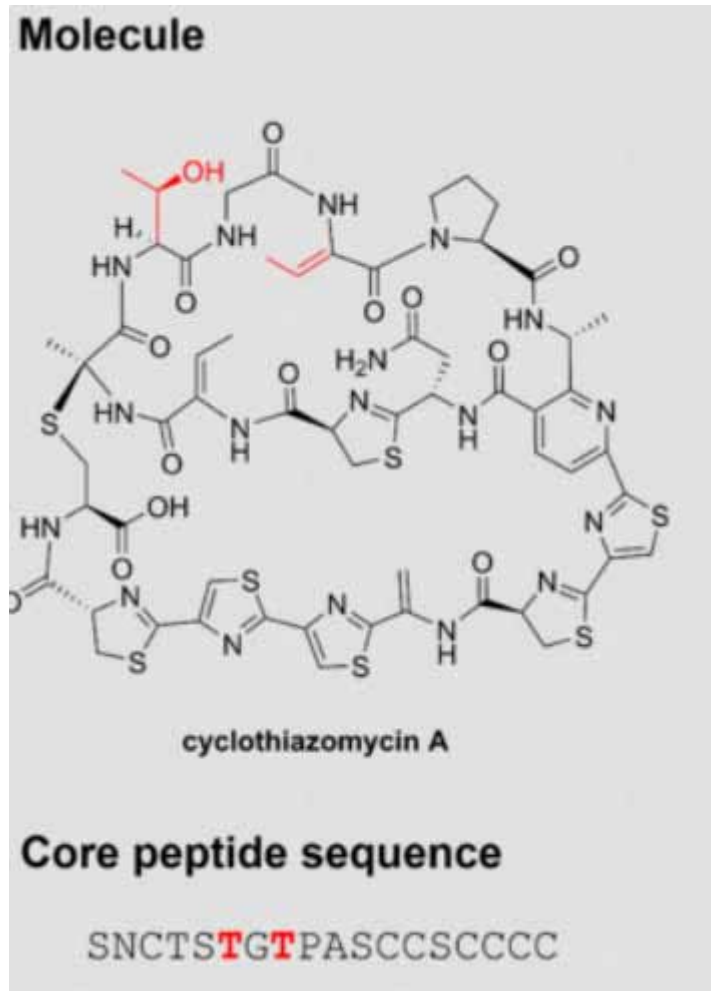
2

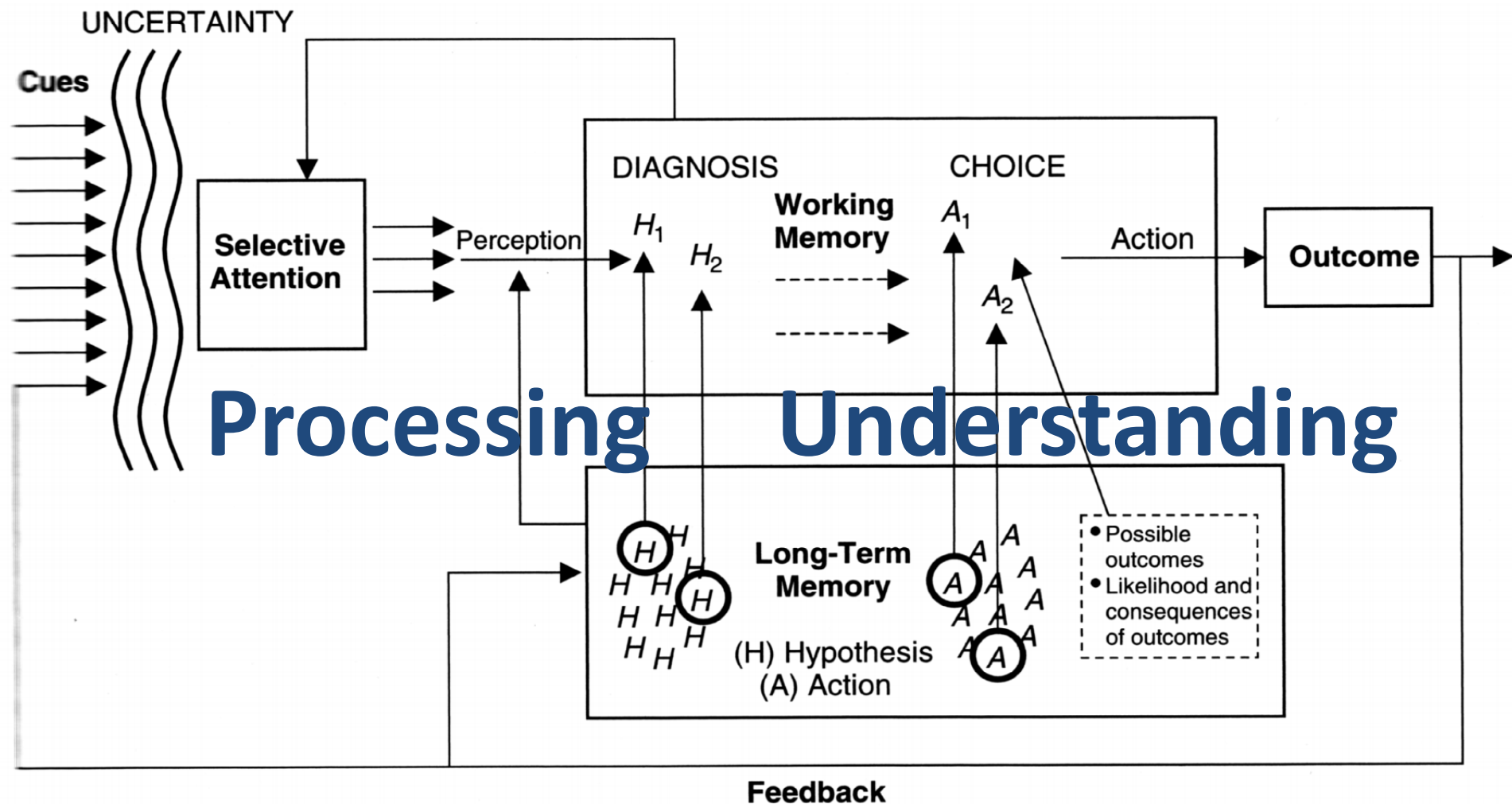


3



4





Wickens, C. D. (1984) *Engineering psychology and human performance*. Columbus (OH), Charles Merrill.

01

apple

Verbal Information

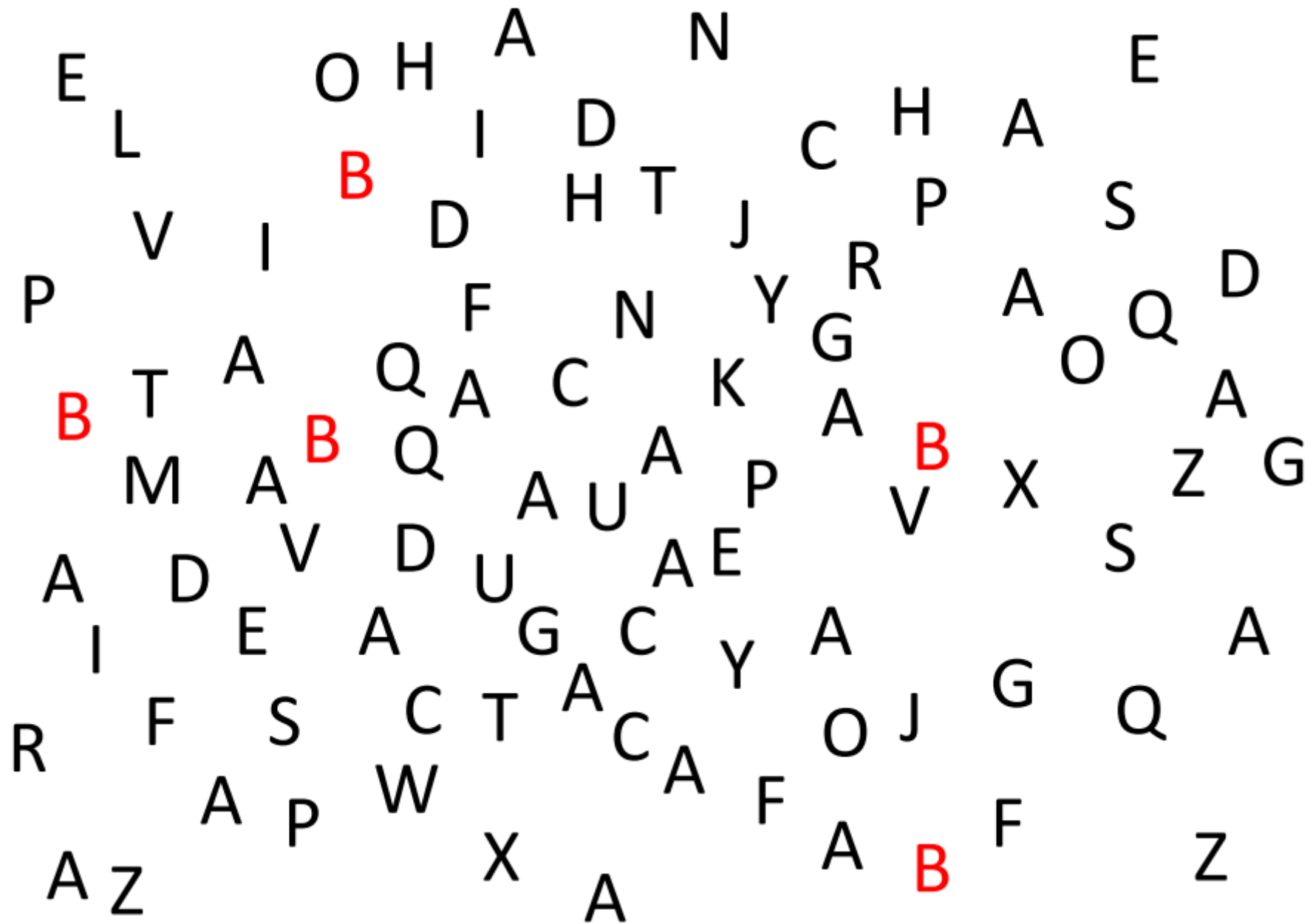
VS.



Visual Information

- How to “visualize” high-dimensional spaces?
- The transformation of results from high-dimensional space $\mathbb{R}^N$ into $\mathbb{R}^2$
- From the complex to the simple (it is superhard to make it as simple as possible!)
- Sampling, modelling, rendering, perception, cognition, decision making ... difficult!
- Trade-off between time and accuracy
- How to model uncertainty
- Integration of visual analytics techniques into the clinical workplace (integrative techniques):
What is not in your **direct workflow** is ignored ...

E L O H A N E
B I D H T C H A
P V I D J Y R P S
B T A Q A C K G A O Q D
M A B Q A U A P B X Z G
A D V D U A E V S
I E A G C Y A A
R F S C T A C A O J G Q
A P W X A F A B F Z
A Z A



How Capacity Limits of Attention Influence Information Visualization Effectiveness

Steve Haroz and David Whitney



Fig. 1. These images each have one colored square that is unique within that image. How long does it take you to find each? How many color categories are there in each panel? Why does grouping make both tasks substantially easier?

Abstract—In this paper, we explore how the capacity limits of attention influence the effectiveness of information visualizations. We conducted a series of experiments to test how visual feature type (color vs. motion), layout, and variety of visual elements impacted user performance. The experiments tested users' abilities to (1) determine if a specified target is on the screen, (2) detect an odd-ball, deviant target, different from the other visible objects, and (3) gain a qualitative overview by judging the number of unique categories on the screen. Our results show that the severe capacity limits of attention strongly modulate the effectiveness of information visualizations, particularly the ability to detect unexpected information. Keeping in mind these capacity limits, we conclude with a set of design guidelines which depend on a visualization's intended use.

Index Terms—Perception, attention, color, motion, user study, nominal axis, layout, goal-oriented design.

Haroz, S. & Whitney, D. 2012. How capacity limits of attention influence information visualization effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 18, (12), 2402-2410.

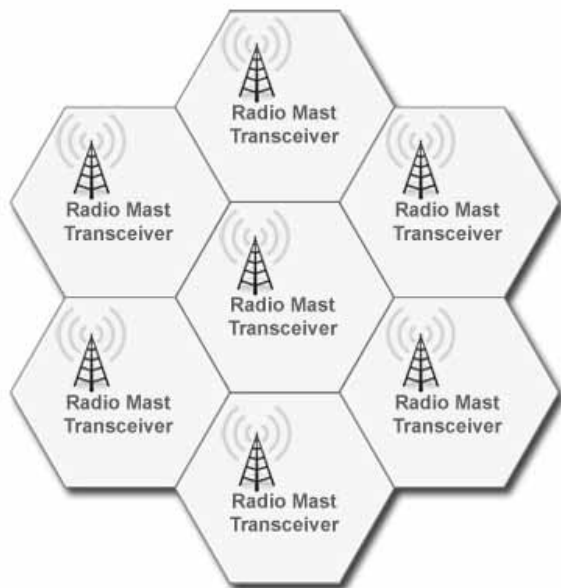
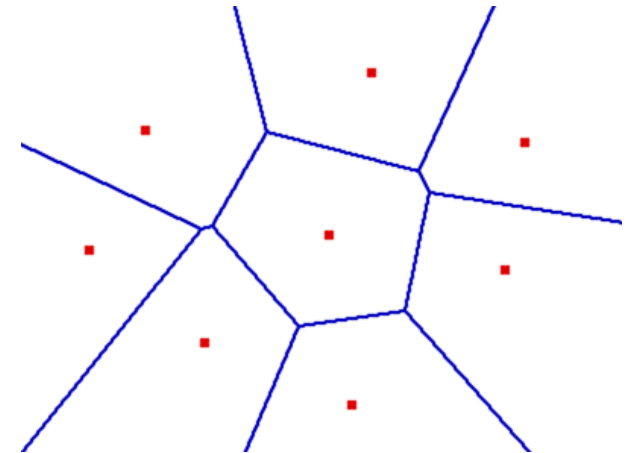


<https://www.youtube.com/watch?v=vJG698U2Mvo>

Simons, D. J. & Chabris, C. F. 1999. Gorillas in our midst: sustained inattention blindness for dynamic events. *Perception*, 28, (9), 1059-1074.

Problem: Context!



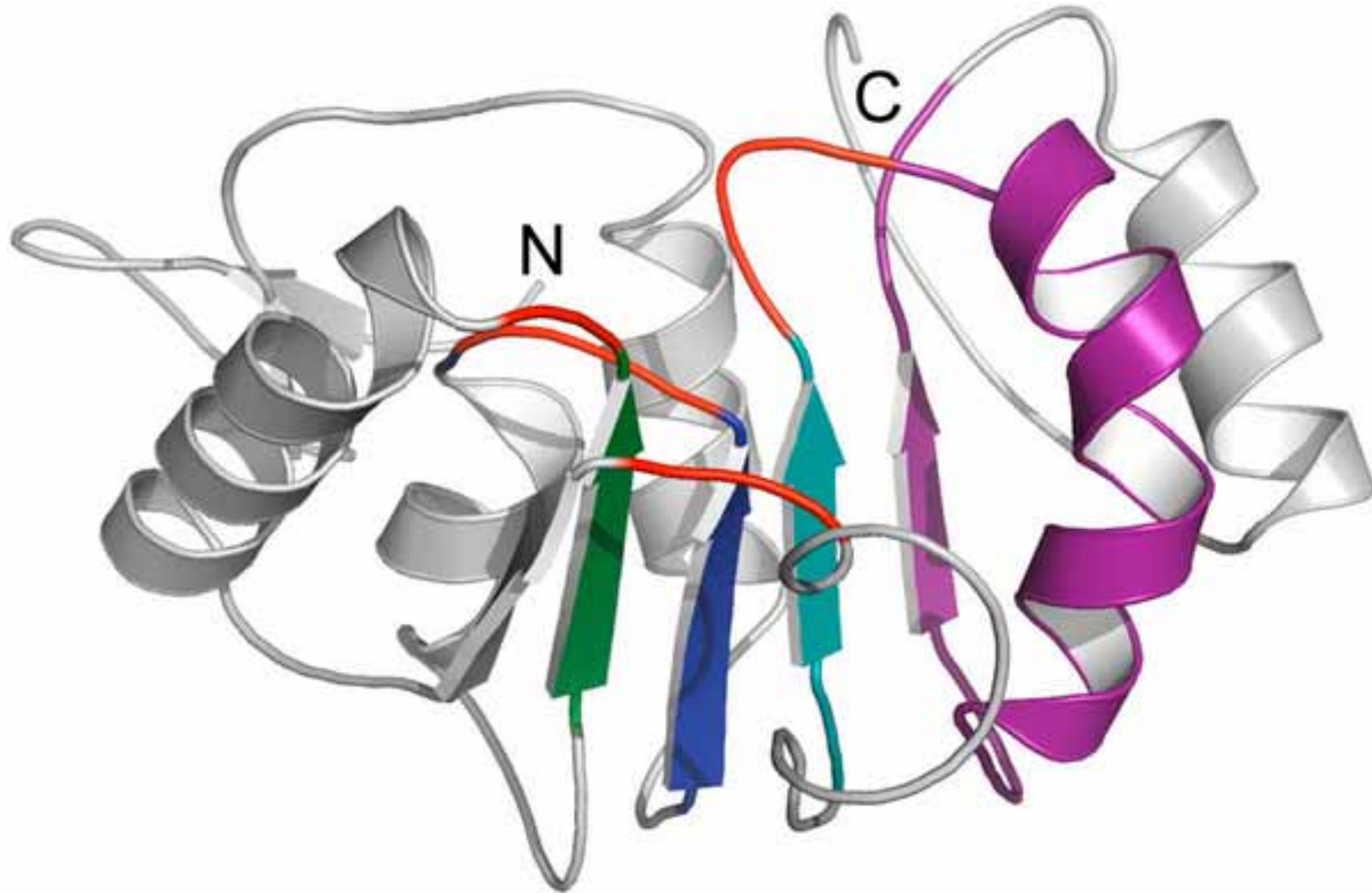


F11 fx =F\$2*\$A11

	A	B	C	D	E	F	G	H	I	J	K
1	Multiplication table - mixed references										
2		1	2	3	4	5	6	7	8	9	10
3	1	1	2	3	4	5	6	7	8	9	10
4	2	2	4	6	8	10	12	14	16	18	20
5	3	3	6	9	12	15	18	21	24	27	30
6	4	4	8	12	16	20	24	28	32	36	40
7	5	5	10	15	20	25	30	35	40	45	50
8	6	6	12	18	24	30	36	42	48	54	60
9	7	7	14	21	28	35	42	49	56	63	70
10	8	8	16	24	32	40	48	56	64	72	80
11	9	9	18	27	36	45	54	63	72	81	90
12	10	10	20	30	40	50	60	70	80	90	100
13	11	11	22	33	44	55	66	77	88	99	110
14	12	12	24	36	48	60	72	84	96	108	120
15											



A picture is worth a thousand words?



Magnani, R., et al. 2010. Calmodulin methyltransferase is an evolutionarily conserved enzyme that trimethylates Lys-115 in calmodulin. *Nature Communications*, 1, 43.



PERSONAL HISTORY: Suicide attempts ☐ Suicide threats ☐ Hobbies, aptitude and skills with firearms, chemicals, etc. ☐ Domestic, premarital or marital conflicts ☐ Financial or business reverses ☐ Social or religious conflicts ☐ Legal difficulties ☐ Criminal record ☐ Unemployment ☐ Fear of disease ☐ Other (specify) _____							
CONDUCT BEFORE DEATH: Efforts to prevent help ☐ Efforts to obtain help ☐ Suicide attempt: Admitted ☐ Denied ☐ Refusal to talk ☐ Written declaration of intended suicide ☐ Accusations against others ☐ Other (specify) _____							
DATE	LAST SEEN ALIVE	INJURY OR ILLNESS	DEATH	DISCOVERY	MEDICAL EXAMINER NOTIFIED	VIEW OF BODY	POLICE NOTIFIED
8-16-77	08:00±		8-16-77	8-16-77	8-16-77	8-16-77	8-16-77
			1530	1400±	1600	1730	1530
INJURY OR ONSET OF ILLNESS		LOCATION		CITY OR COUNTY		TYPE OF PREMISES (HOSPITAL, HOTEL, HIGHWAY, ETC.)	
None		Home		Bm H - Room attempted			
VIEWING OF BODY BY MEDICAL EXAMINER		Bm H					
MEDICAL ATTENTION AND HOSPITAL OR INSTITUTIONAL CARE							
NAME OF PHYSICIAN OR INSTITUTION		ADDRESS		DIAGNOSIS		DATE	
A. N. S. S. S.		Main		HCVB / Cohn P. S.			
CIRCUMSTANCES OF DEATH							
FOUND DEAD BY		NAME		ADDRESS			
Hinges Alden		Hinges Alden		Same			
LAST SEEN ALIVE BY		"		"			
WITNESS TO INJURY OR ILLNESS AND DEATH		None					
NARRATIVE SUMMARY OF CIRCUMSTANCES SURROUNDING DEATH: Found on floor of dressing room by above who had been sleeping in adjoining room. No indication of foul play. Had been ok in early AM. He (?) playing Racketball in early AM. Family consent signed for autopsy - to be performed at Bm H.							
(EDITOR - HERE'S WHAT IT SAYS) FOUND ON FLOOR OF DRESSING ROOM BY ABOVE WHO HAD BEEN SLEEPING IN ADJOINING ROOM. NO INDICATION OF FOUL PLAY. HAD BEEN OK IN EARLY AM. HE (?) PLAYING RACKETBALL IN EARLY AM. FAMILY CONSENT SIGNED FOR AUTOPSY TO BE PERFORMED AT Bm H.							

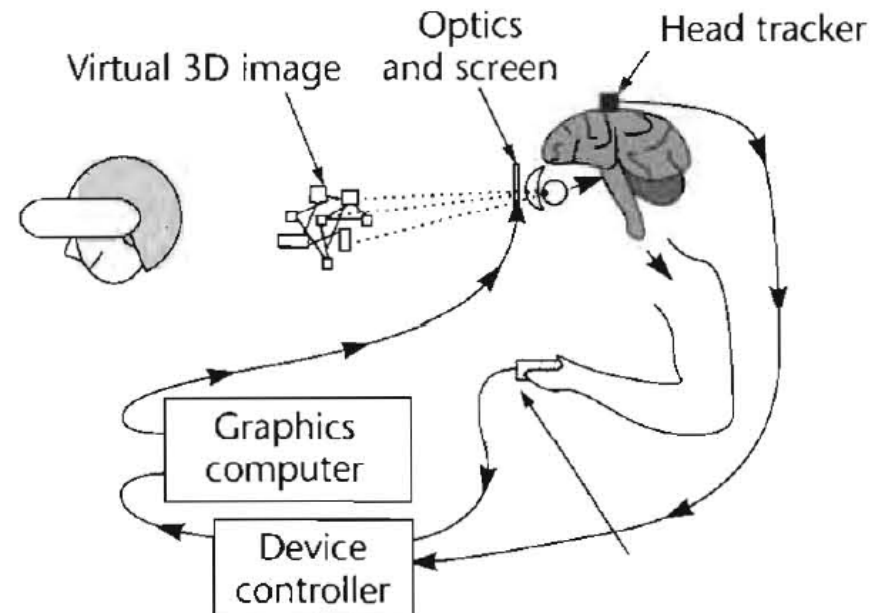
02 Informatics as Semiotics Engineering



Ware, C. (2004) *Information Visualization: Perception for Design (Interactive Technologies) 2nd Edition*. San Francisco, Morgan Kaufmann.

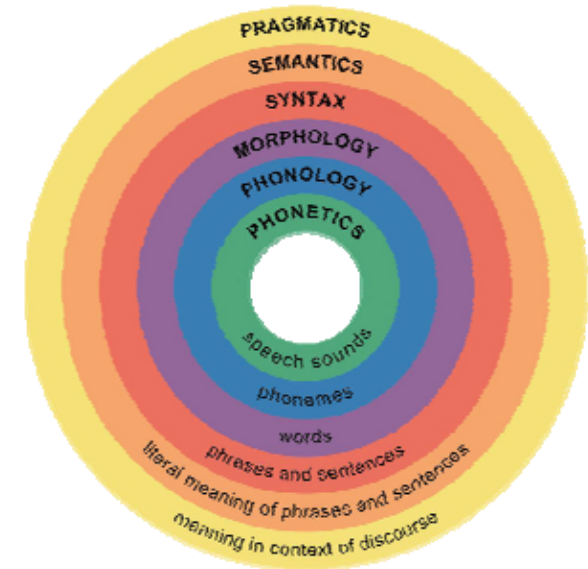
$$W = \{w : w \in T(v)\}$$

$$H = - \sum_{w \in W} p(w) \log_2 p(w)$$



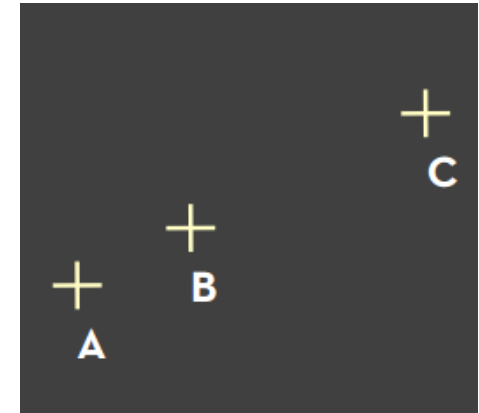
Holzinger, A., Searle, G., Auinger, A. & Ziefle, M. (2011) Informatics as Semiotics Engineering: Lessons learned from Design, Development and Evaluation of Ambient Assisted Living Applications for Elderly People. *Universal Access in Human-Computer Interaction. Context Diversity. Lecture Notes in Computer Science (LNCS 6767)*. Berlin, Heidelberg, New York, Springer, 183-192.

- 1. Physical: is it present?
 - Signals, traces, components, points, ...
- 2. Empirical: can it be seen?
 - Patterns, entropy, codes, ...
- 3. Syntactic: can it be read?
 - Formal structure, logic, deduction, ...
- 4. Semantic: can it be understood?
 - Meaning, proposition, truth, ...
- 5. Pragmatic: is it useful?
 - Intentions, negotiations, communications, ...
- 6. Social: can it be trusted?
 - Beliefs, expectations, culture, ...



Burton-Jones, A., Storey, V. C., Sugumaran, V. & Ahluwalia, P. 2005. A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, 55, (1), 84-102.

- Images are perceived as a set of signs
- Sender **encodes** information in signs
- Receiver **decodes** information from signs
- *“Resemblance, order and proportion are the 3 “signifieds” in graphics”*
- *“With up to three rows, a data table can be constructed directly as a single image ... However, an image has 3 dimensions And this barrier is impassible.”*



Bertin, J. & Barbut, M. 1967. Sémiologie graphique: les diagrammes, les réseaux, les cartes, Mouton Paris.

03 What is Visualization?

- **Visualization** = generally a method of computer science to transform the symbolic into the geometric, to form a mental model and foster unexpected insights;
- **Information visualization** = the interdisciplinary study of the visual representation of large-scale collections of non-numerical data, such as files and software, databases, networks etc., to allow users to see, explore, and understand information at once;
- **Data visualization** = visual representation of complex data, to communicate information clearly and effectively, making data useful and usable;
- **Visual Analytics** = focuses on analytical reasoning of complex data facilitated by **interactive** visual interfaces;
- **Content Analytics** = a general term addressing so-called “unstructured” information – mainly text – by using mixed methods from visual analytics and business intelligence;

Image Formation

object in → image out

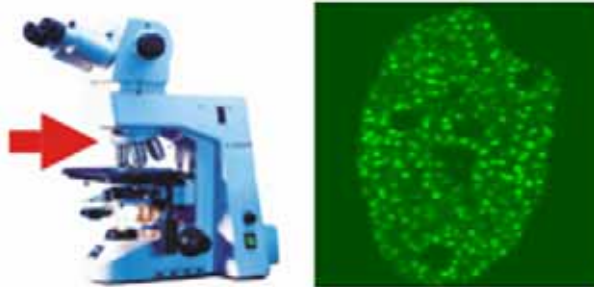
**Image Processing**

image in → image out

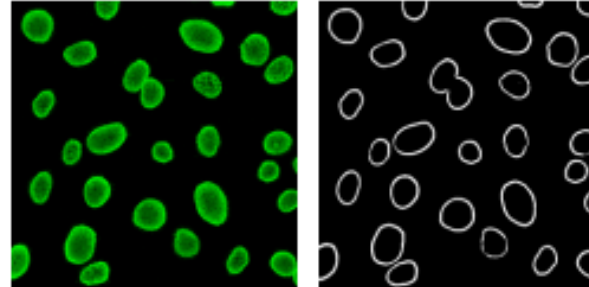
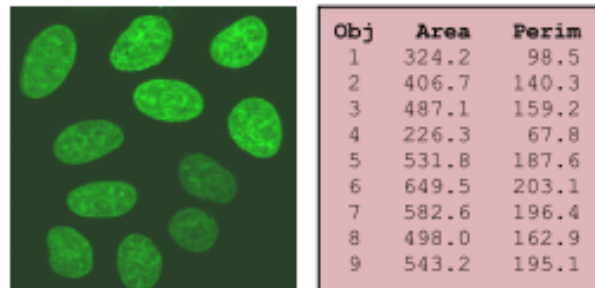
**Image Analysis**

image in → features out

**Computer Graphics**

numbers in → image out

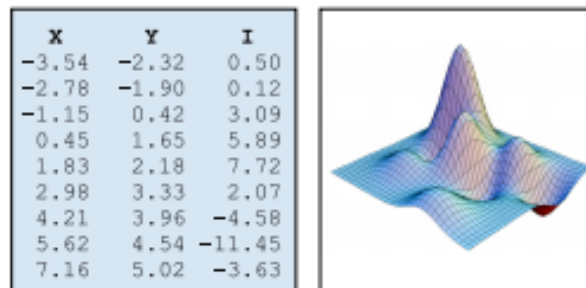
**Computer Vision**

image in → interpretation out

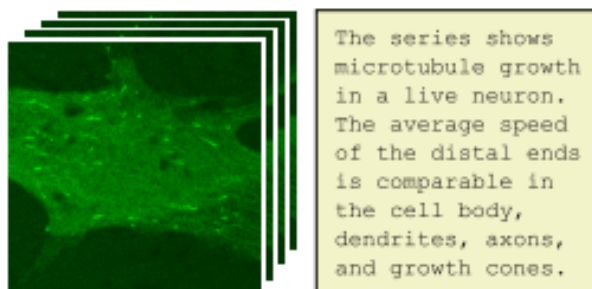
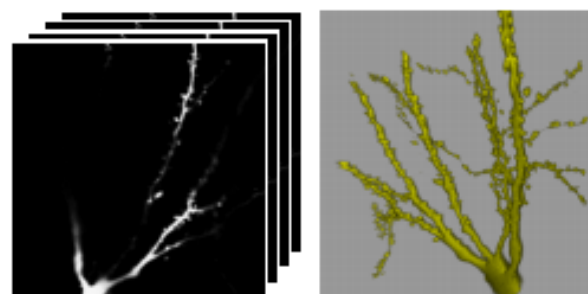
**Visualization**

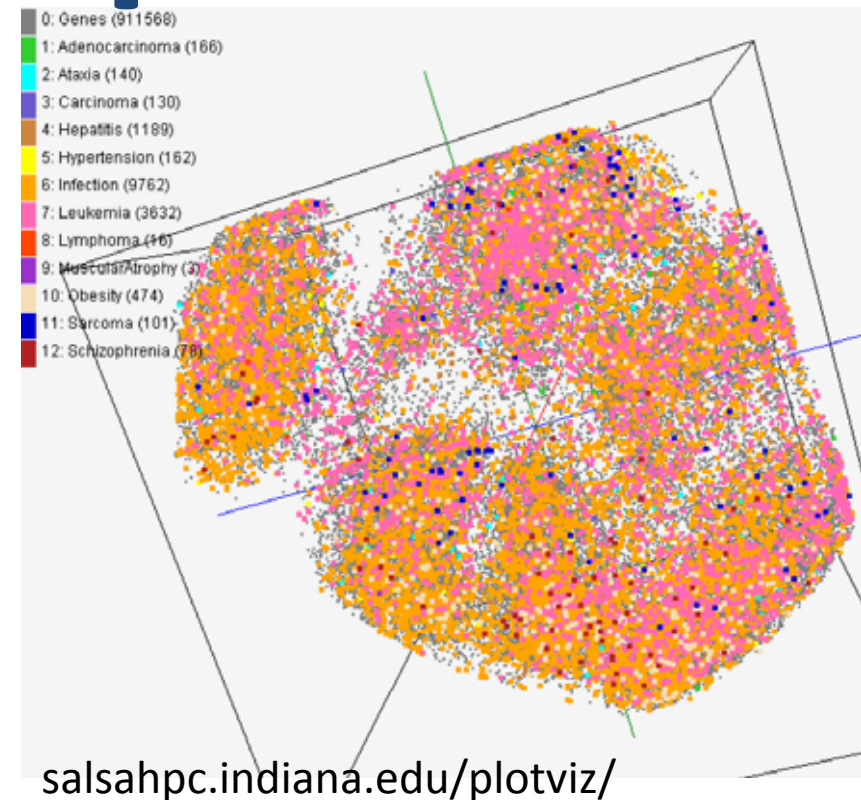
image in → representation out

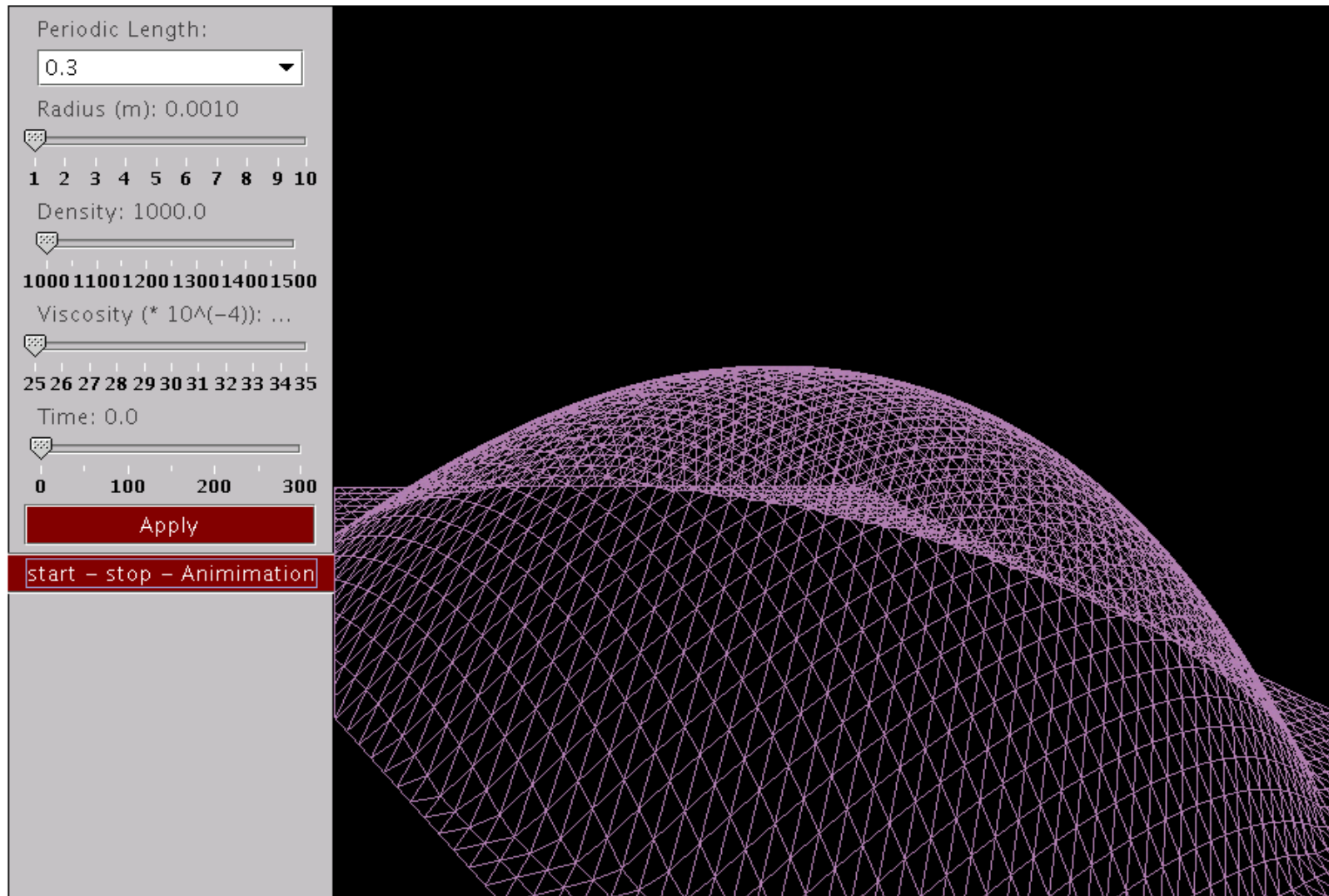


Meijering, Erik &
Cappellen, Gert (2006)
Biological Image Analysis
Primer, available via
<http://www.imagescience.org/meijering/publications/1009/> Erasmus
University Medical Center

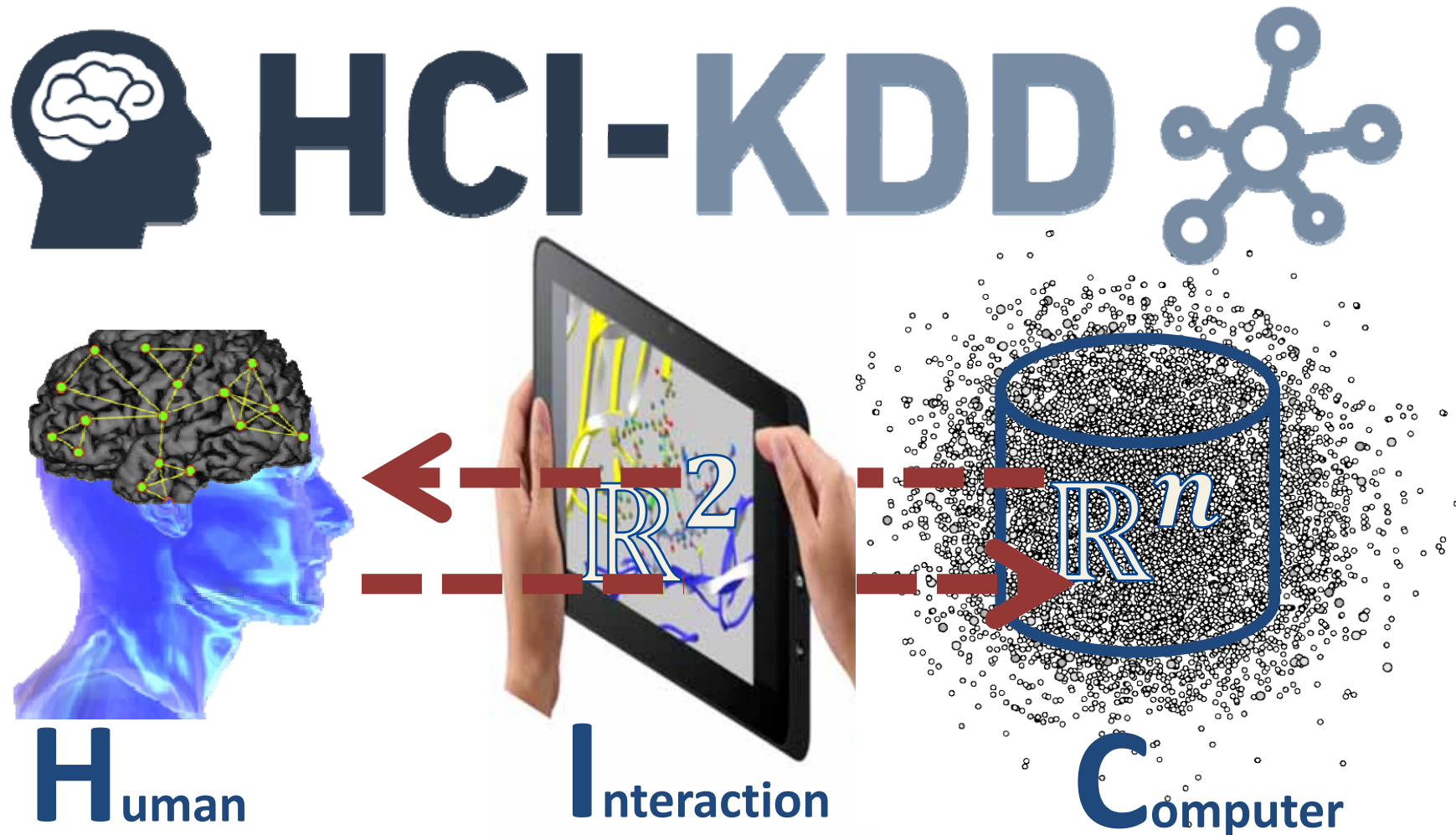
Visualization is a typical HCI topic !

Jong Youl Choi, Seung-Hee Bae, Judy Qiu, Geoffrey Fox, Bin Chen, and David Wild, "Browsing Large Scale Cheminformatics Data with Dimension Reduction," Proceedings of Emerging Computational Methods for the Life Sciences Workshop of ACM HPDC 2010 conference, Chicago, Illinois, June 20-25, 2010.



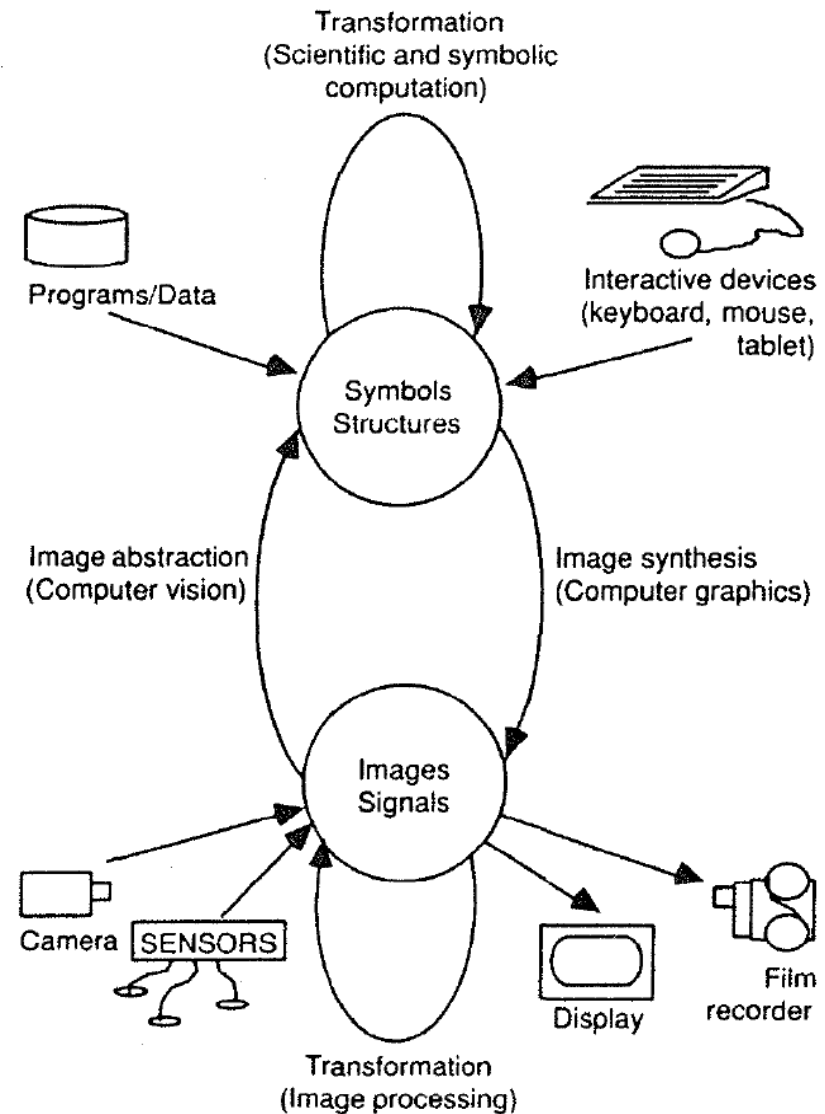


Holzinger, A., Kickmeier-Rust, M. D., Wassertheurer, S. & Hessinger, M. (2009) Learning performance with interactive simulations in medical education: Lessons learned from results of learning complex physiological models with the HAEMOdynamics SIMulator. *Computers & Education*, 52, 2, 292-301.

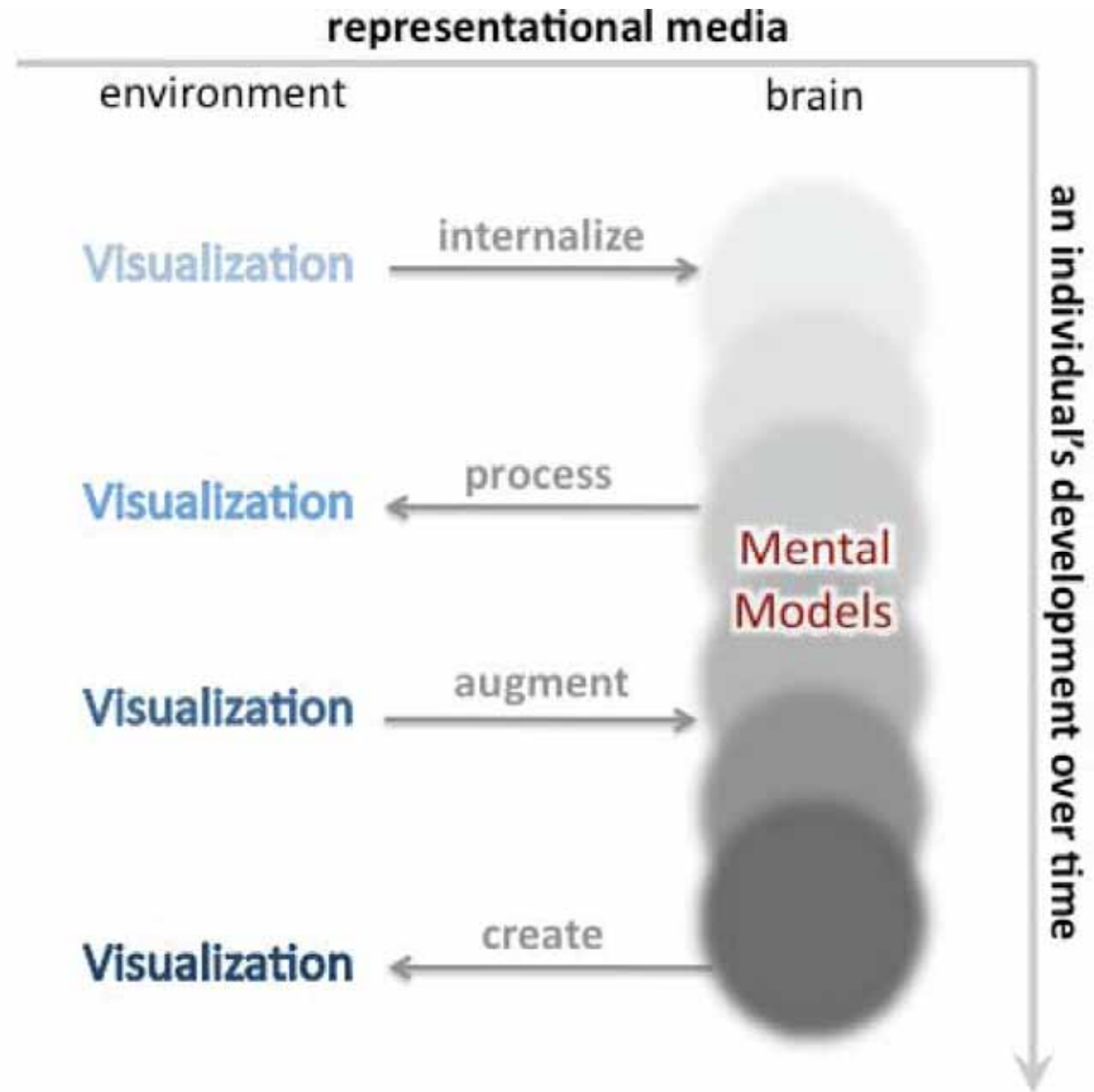


Holzinger, A. 2013. Human–Computer Interaction & Knowledge Discovery (HCI-KDD): What is the benefit of bringing those two fields to work together? *In: Alfredo Cuzzocrea, C. K., Dimitris E. Simos, Edgar Weippl, Lida Xu (ed.) Multidisciplinary Research and Practice for Information Systems, Springer Lecture Notes in Computer Science LNCS 8127. Heidelberg, Berlin, New York: Springer, pp. 319-328.*

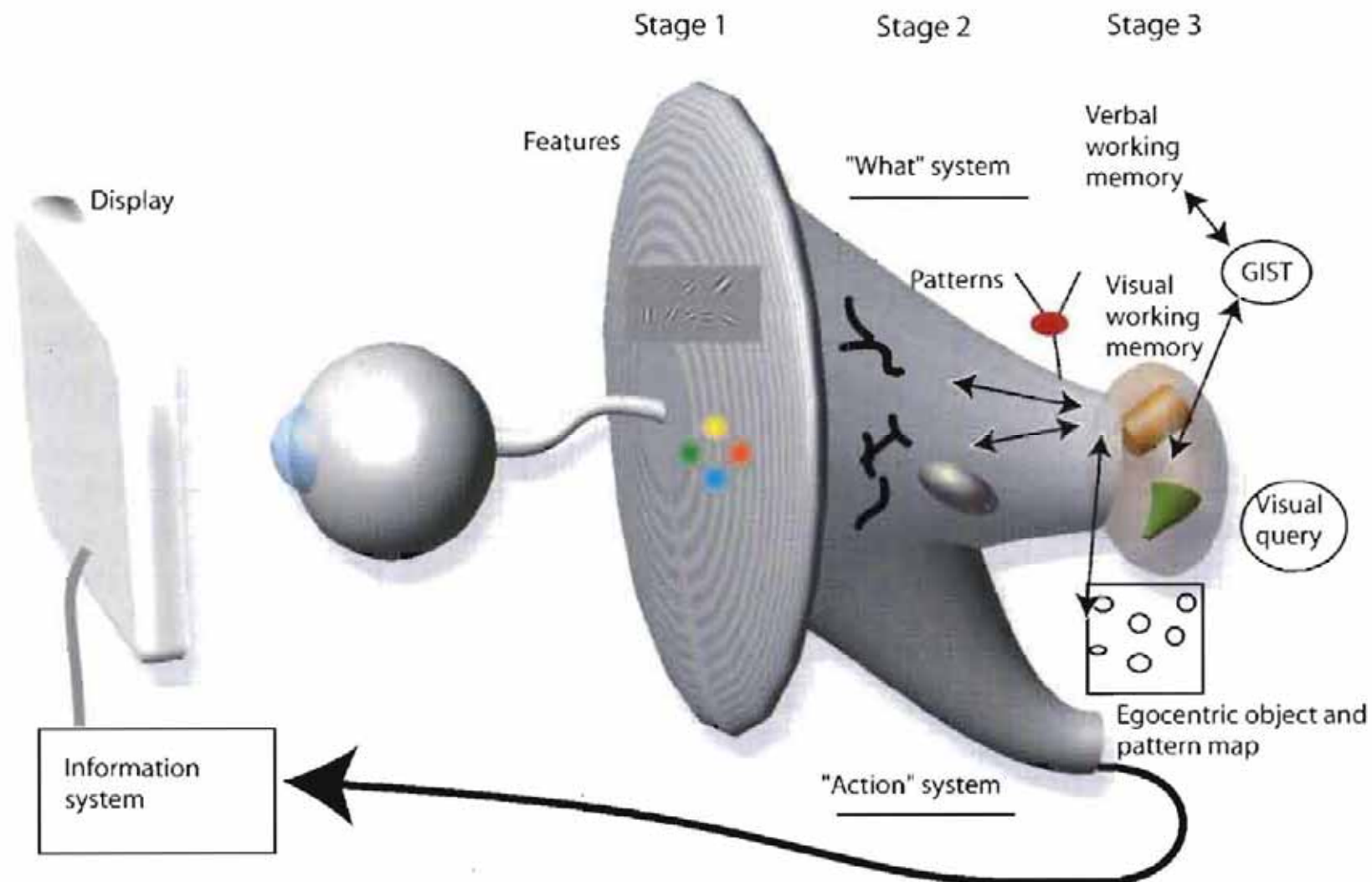
- ... the common denominator of Computational sciences
- ... the transformation of the symbolic into the geometric
- ... the support of human perception
- ... facilitating knowledge discovery in data



McCormick, B. (1987) Scientific and Engineering Research Opportunities. *Computer graphics*, 21, 6.



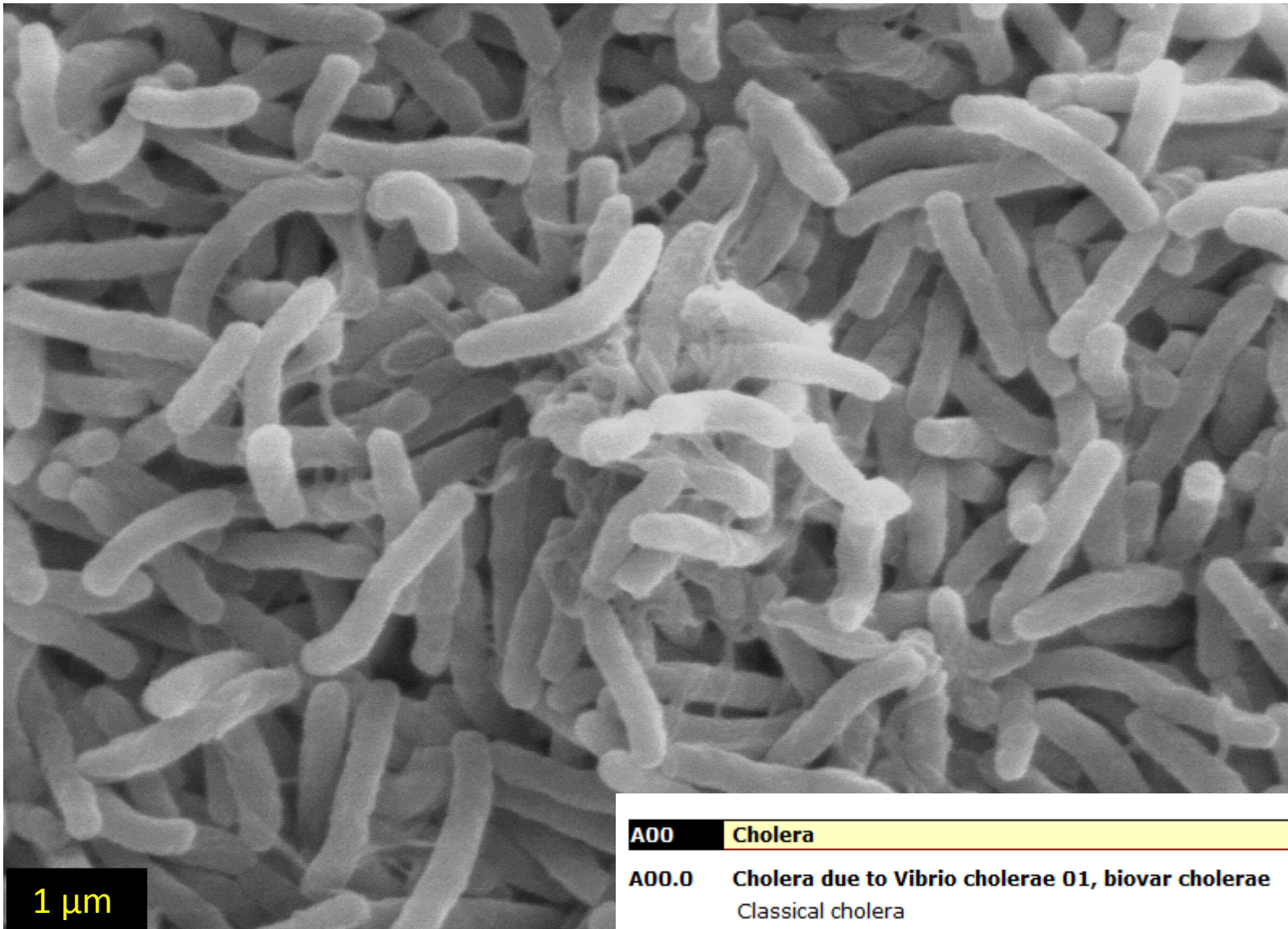
Liu, Z. & Stasko, J. T. (2010)
Mental Models, Visual
Reasoning and Interaction in
Information Visualization: A
Top-down Perspective.
*Visualization and Computer
Graphics, IEEE Transactions
on*, 16, 6, 999-1008.



Ware, C. (2004) *Information Visualization: Perception for Design (Interactive Technologies) 2nd Edition*. San Francisco, Morgan Kaufmann.

04 Usefulness of Visualization Science



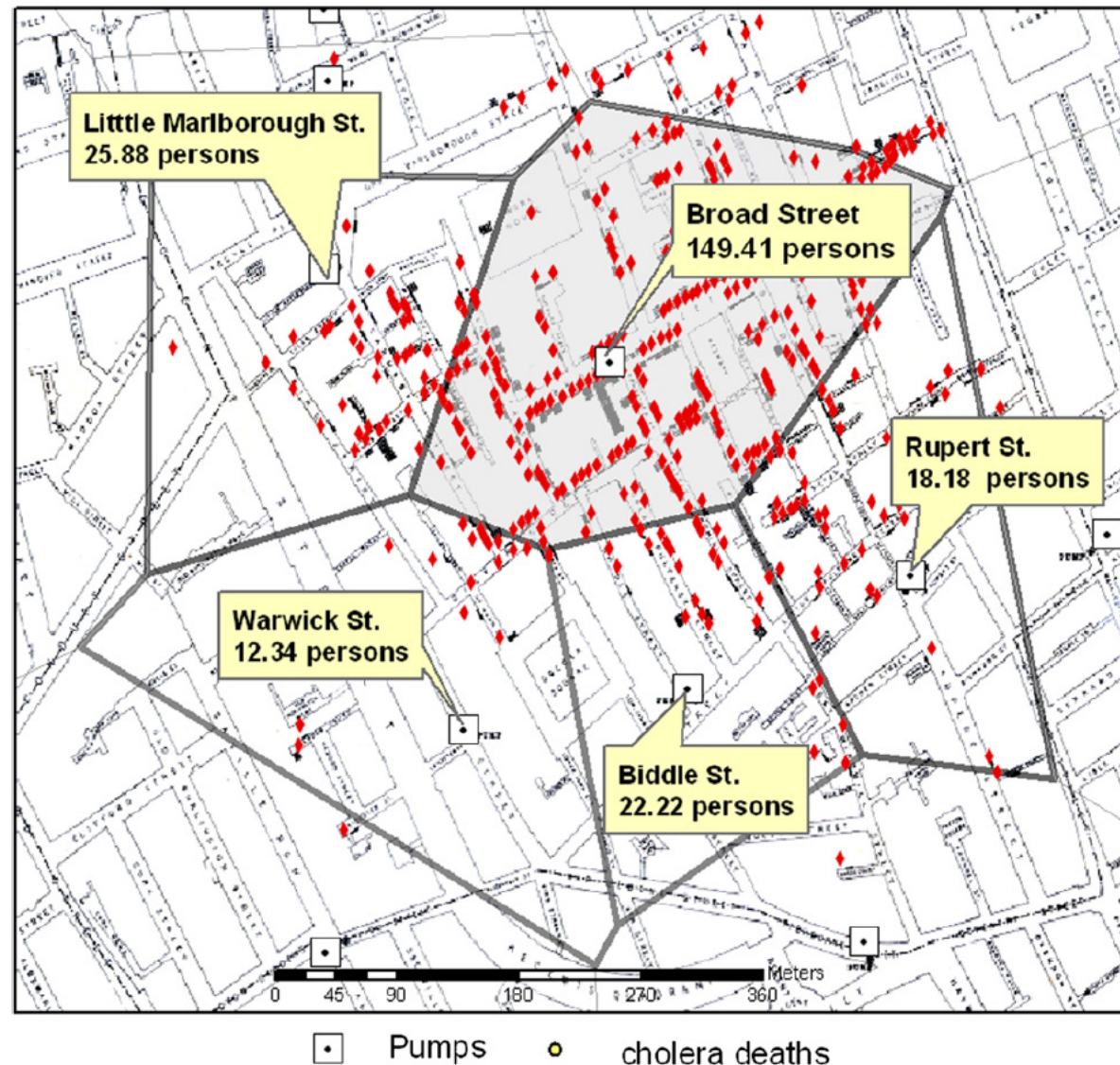


T.J. Kirn, M.J. Lafferty, C.M.P Sandoe and R.K. Taylor (2000) Delineation of pilin domains required for bacterial association into microcolonies and intestinal colonization, *Molecular Microbiology*, Vol. 35, 896-910

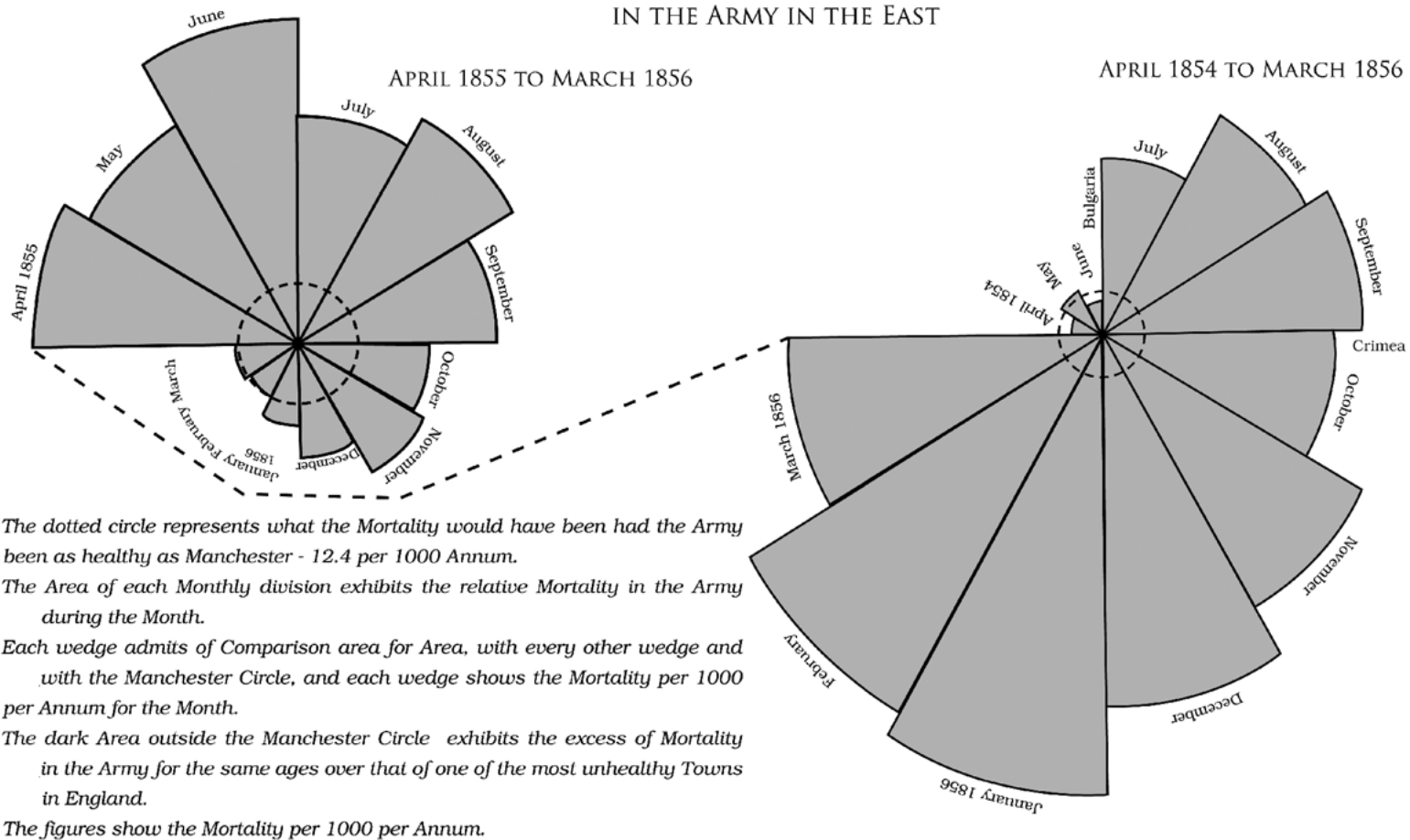


McLeod, K. S. (2000) Our sense of Snow: the myth of John Snow in medical geography. *Social Science & Medicine*, 50, 7-8, 923-935.

Cholera Mortality per 1,000 persons



Koch, T. & Denike, K. (2009) Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854. *Social Science & Medicine*, 69, 8, 1246-1251.

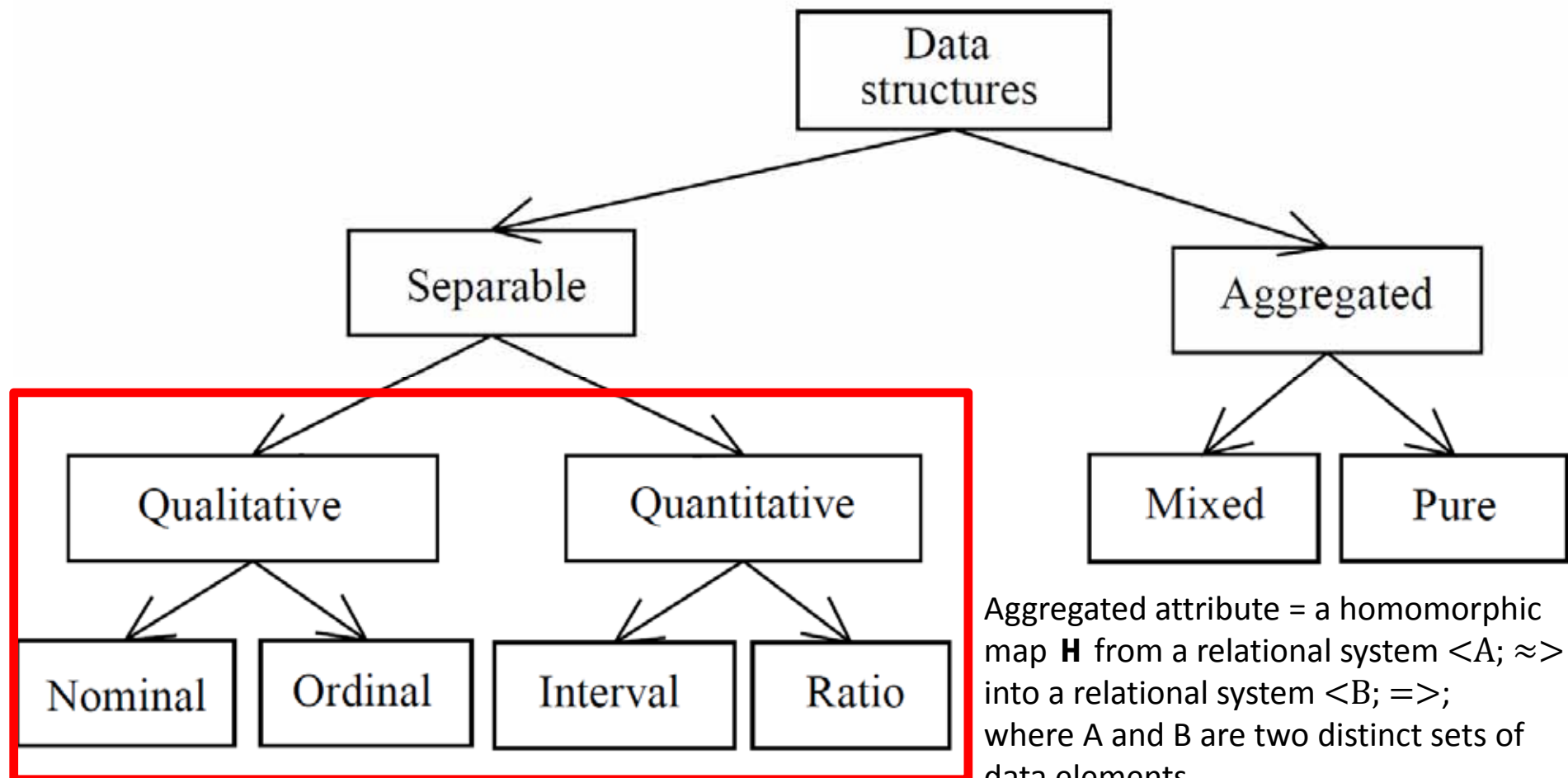
DIAGRAMS OF THE MORTALITY
IN THE ARMY IN THE EAST

Meyer, B. C. & Bishop, D. S. (2007) Florence Nightingale: nineteenth century apostle of quality. *Journal of Management History*, 13, 3, 240-254.

05

Visualization

Basics



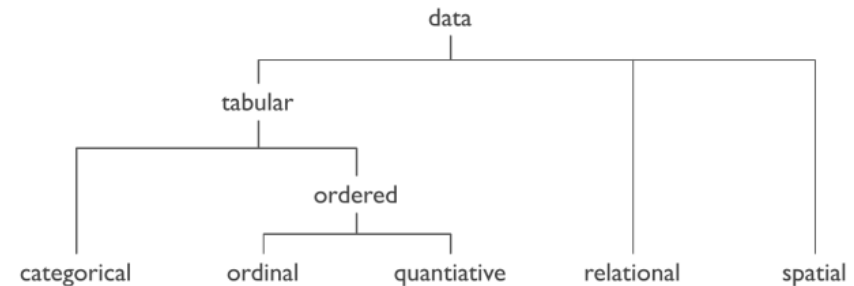
Dastani, M. (2002) The Role of Visual Perception in Data Visualization. *Journal of Visual Languages and Computing*, 13, 601-622.

Aggregated attribute = a homomorphic map H from a relational system $\langle A; \approx \rangle$ into a relational system $\langle B; = \rangle$; where A and B are two distinct sets of data elements.

This is in contrast with other attributes since the set B is the set of data elements instead of atomic values.

Scale	Empirical Operation	Mathem. Group Structure	Transf. in $\mathbb{R}$	Basic Statistics	Mathematical Operations
NOMINAL	Determination of equality	Permutation $x' = f(x)$ $x \dots 1\text{-to-1}$	$x \mapsto f(x)$	Mode, contingency correlation	$=, \neq$
ORDINAL	Determination of more/less	Isotonic $x' = f(x)$ $x \dots \text{mono-tonic incr.}$	$x \mapsto f(x)$	Median, Percentiles	$=, \neq, >, <$
INTERVAL	Determination of equality of intervals or differences	General linear $x' = ax + b$	$x \mapsto rx + s$	Mean, Std.Dev. Rank-Order Corr., Prod.-Moment Corr.	$=, \neq, >, <, -, +$
RATIO	Determination of equality or ratios	Similarity $x' = ax$	$x \mapsto rx$	Coefficient of variation	$=, \neq, >, <, -, +, *, \div$

Stevens, S. S. (1946) On the theory of scales of measurement. *Science*, 103, 677-680.



Marks

Channels

Position

Size

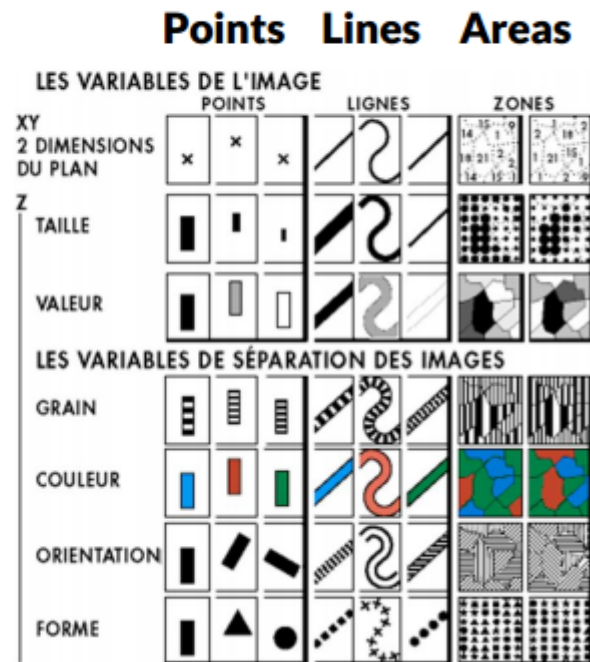
(Grey)Value

Texture

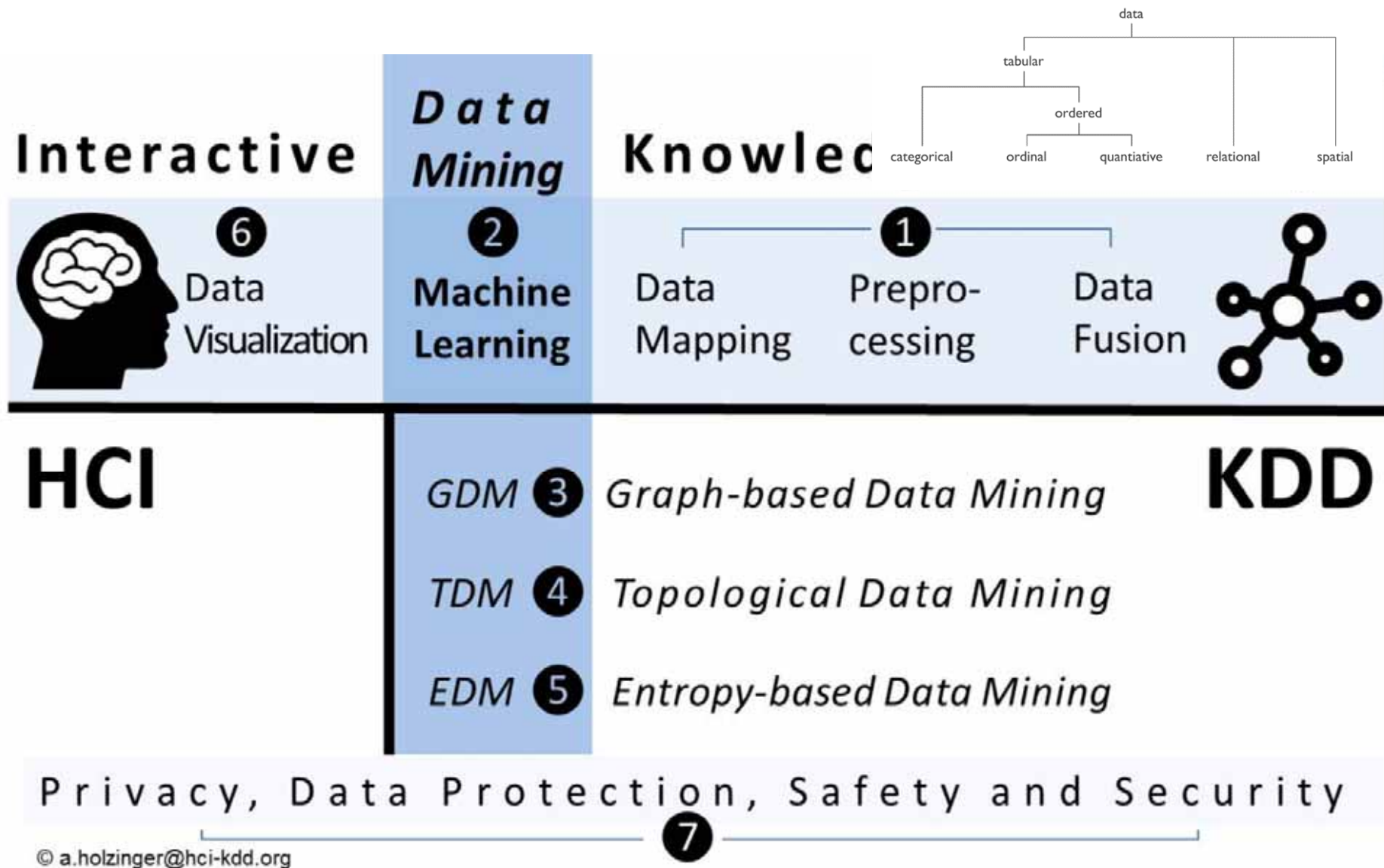
Color

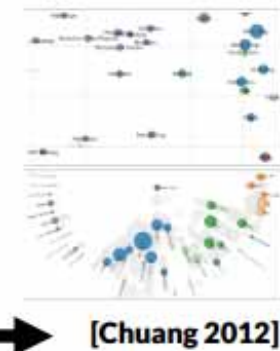
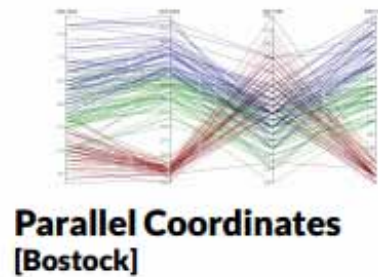
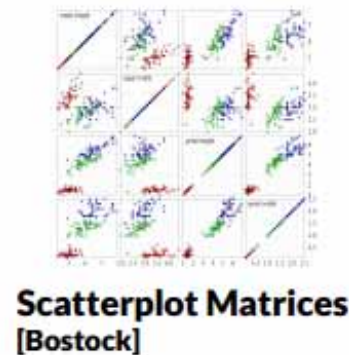
Orientation

Shape



Bertin, J. & Barbut, M. 1967. Sémiologie graphique: les diagrammes, les réseaux, les cartes, Mouton Paris.





no / little analytics

strong analytics
component

Image credit to Alexander Lex, Harvard

Example Chuang (2012) Dissertation Browser:
<http://www-nlp.stanford.edu/projects/dissertations/browser.html>

05

Visualization Methods (Incomplete!)

C continuum	Data Visualization Visual representations of quantitative data in schematic form (either with or without axes) Strategy Visualization The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.														G graphic facilitation		
Tb table	Ca cartesian coordinates	Information Visualization The use of interactive visual representations of data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space. The image can be changed by users as they proceed working with it Metaphor Visualization Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed														Ct cartoon	
Pi pie chart	L line chart	Concept Visualization Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses. Compound Visualization The complementary use of different graphic representation formats in one single schema or frame														Ri rich picture	
B bar chart	Ac area chart	R radar chart cobweb	Pa parallel coordinates	Hy hyperbolic tree	Cy cycle diagram	T timeline	Ve venn. diagram	Mi mindmap	Sq square of oppositions	Cc concentric circles	Ar argument slide	Sw swim lane diagram	Gc gantt chart	Pm perspectives diagram	D dilemma diagram	Pr parameter ruler	Kn knowledge map
Hi histogram	Sc scatterplot	Sa sankey diagram	In information lense	E entity relationship diagram	Pt petri net	Fi flow chart	Cl clustering	Lc layer chart	Py minto pyramid technique	Ce cause-effect chains	Tl toulmin map	Dt decision tree	Cp cpm critical path method	Cf concept fan	Co concept map	Ic iceberg	Lm learning map
Tk tukey box plot	Sp spectrogram	Da data map	Tp treemap	Cn cone tree	Sy system dyn./simulation	Df data flow diagram	Se semantic network	So soft system modeling	Sn synergy map	Fo force field diagram	Ib ibis argumentation map	Pr process event chains	Pe pert chart	Ev evocative knowledge map	V Vee diagram	Hh heaven 'n' hell chart	I infomural

Cy**Process Visualization****Hy****Structure Visualization****Overview****Detail****Detail AND Overview****Divergent thinking****Convergent thinking**

Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

version 1.5

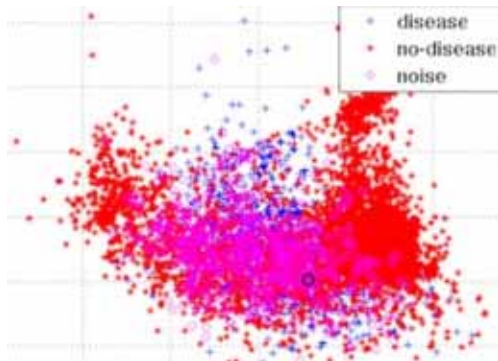
© Ralph Lengler & Martin J. Eppler; www.visual-literacy.org

 Su supply demand curve	 Pc performance charting	 St strategy map	 Oc organisation chart	 Ho house of quality	 Fd feedback diagram	 Ft failure tree	 Mq magic quadrant	 Ld life-cycle diagram	 Po porter's five forces	 S s-cycle	 Sm stakeholder map	 Is ishikawa diagram	 Tc technology roadmap
 Ed edgeworth box	 Pf portfolio diagram	 Sg strategic game board	 Mz mintzberg's organigraph	 Z zwicky's morphological box	 Ad affinity diagram	 De decision discovery diagram	 Bm bcg matrix	 Stc strategy canvas	 Vc value chain	 Hy hype-cycle	 Sr stakeholder rating map	 Ta taps	 Sd spray diagram

Lengler, R. & Eppler, M. J. (2007) Towards a periodic table of visualization methods for management.

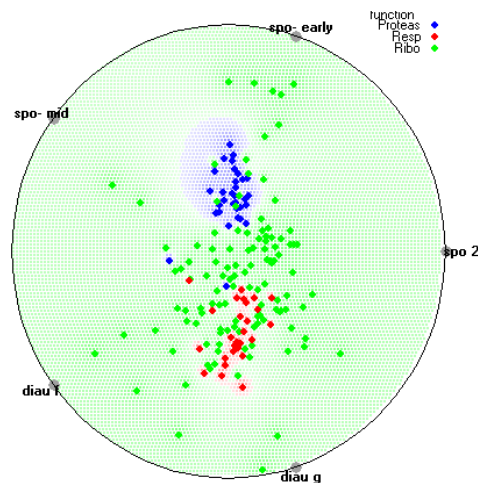
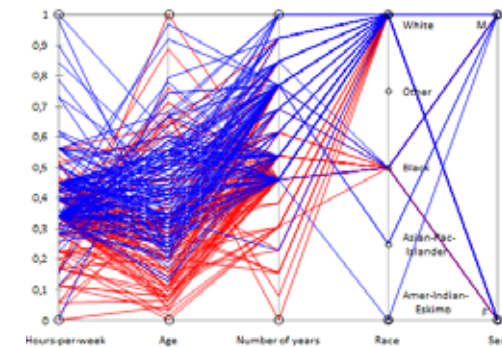
Proceedings of Graphics and Visualization in Engineering (GVE 2007); Online: www.visual-literacy.org

- 1) Data Visualization (Pie Charts, Area Charts or Line Graphs, ...)
- 2) Information Visualization (Semantic networks, tree-maps, radar-chart, ...)
- 3) Concept Visualization (Concept map, Gantt chart, PERT diagram, ...)
- 3) Metaphor Visualization (Metro maps, story template, iceberg, ...)
- 4) Strategy Visualization (Strategy Canvas, roadmap, morpho box,...)
- 5) Compound Visualization

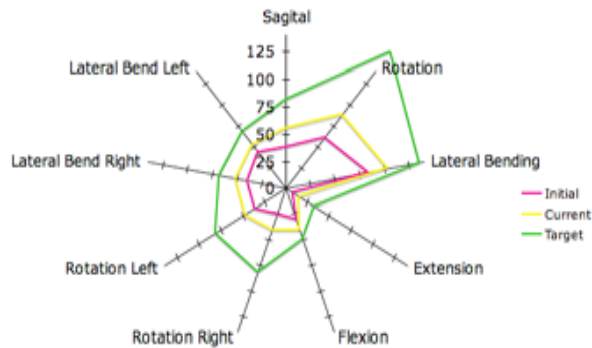


Scatterplot = oldest, point-based technique, projects data from n -dim space to an arbitrary k -dim display space;

Parallel coordinates = (PCP), originally for the study of high-dimensional geometry, data point plotted as polyline;

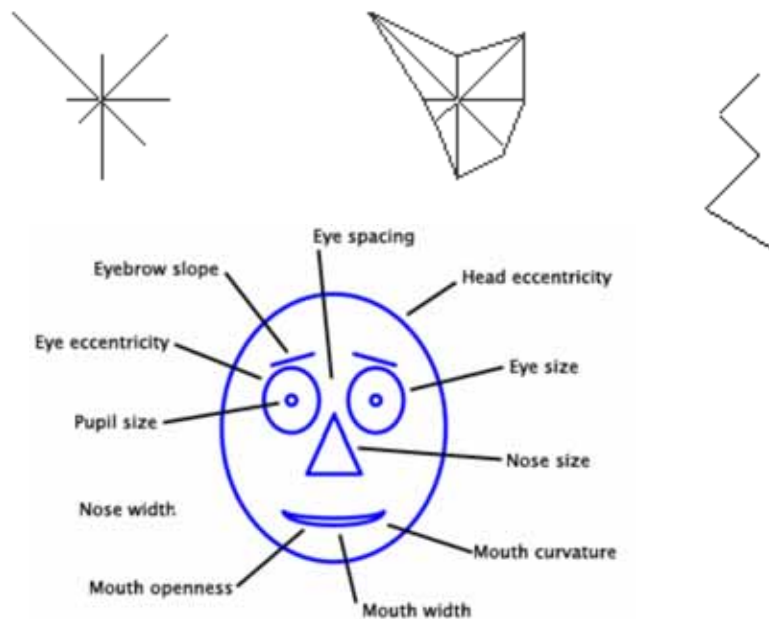
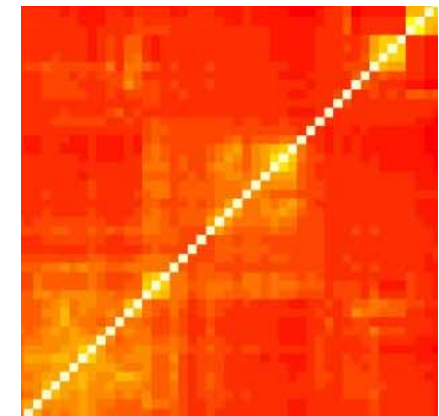


RadViz = Radial Coordinate visualization, is a “force-driven” point layout technique, based on Hooke’s law for equilibrium;



Radar chart (star plot, spider web, polar graph, polygon plot) = radial axis technique;

Heatmap = a tabular display technique using color instead of figures for the entries;



Glyph = a visual representation of the entity, where its attributes are controlled by data attributes;

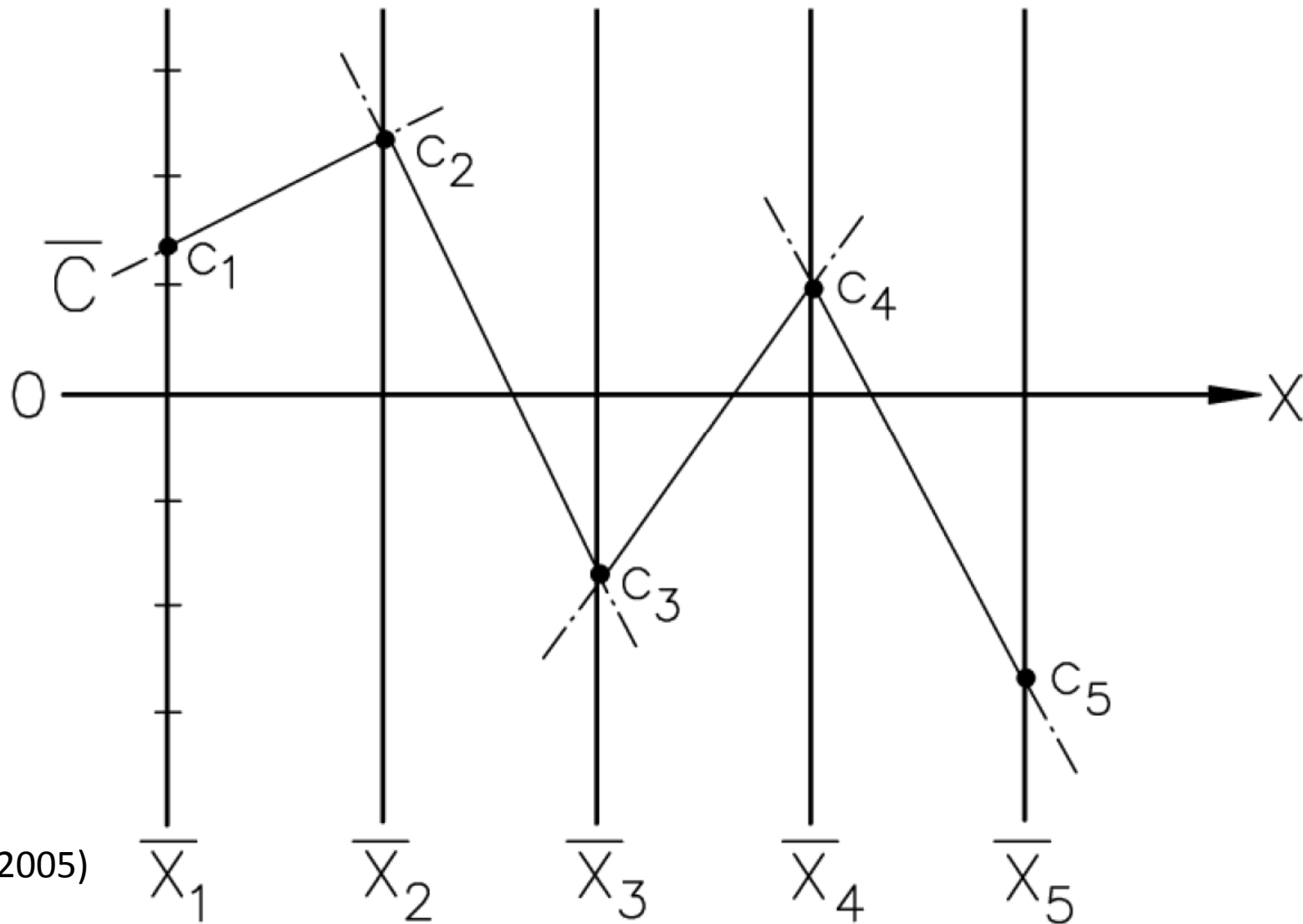
Chernoff face = a face glyph which displays multivariate data in the shape of a human face

- On the plane with Cartesian-coords, a vertical line, labeled $\bar{X}_i$ is placed at each $x = i - 1$ for $i = 1, 2, \dots N$.
- These are the axes of the parallel coordinate system for $\mathbb{R}^N$.
- A point $C = (c_1, c_2, \dots c_N) \in \mathbb{R}^N$ is mapped into the polygonal line $\bar{C}$
- the N -vertices with xy -coords $(i - 1, c_i)$ are now on the parallel axes.
- In $\bar{C}$ the full lines and not only the segments between the axes are included.



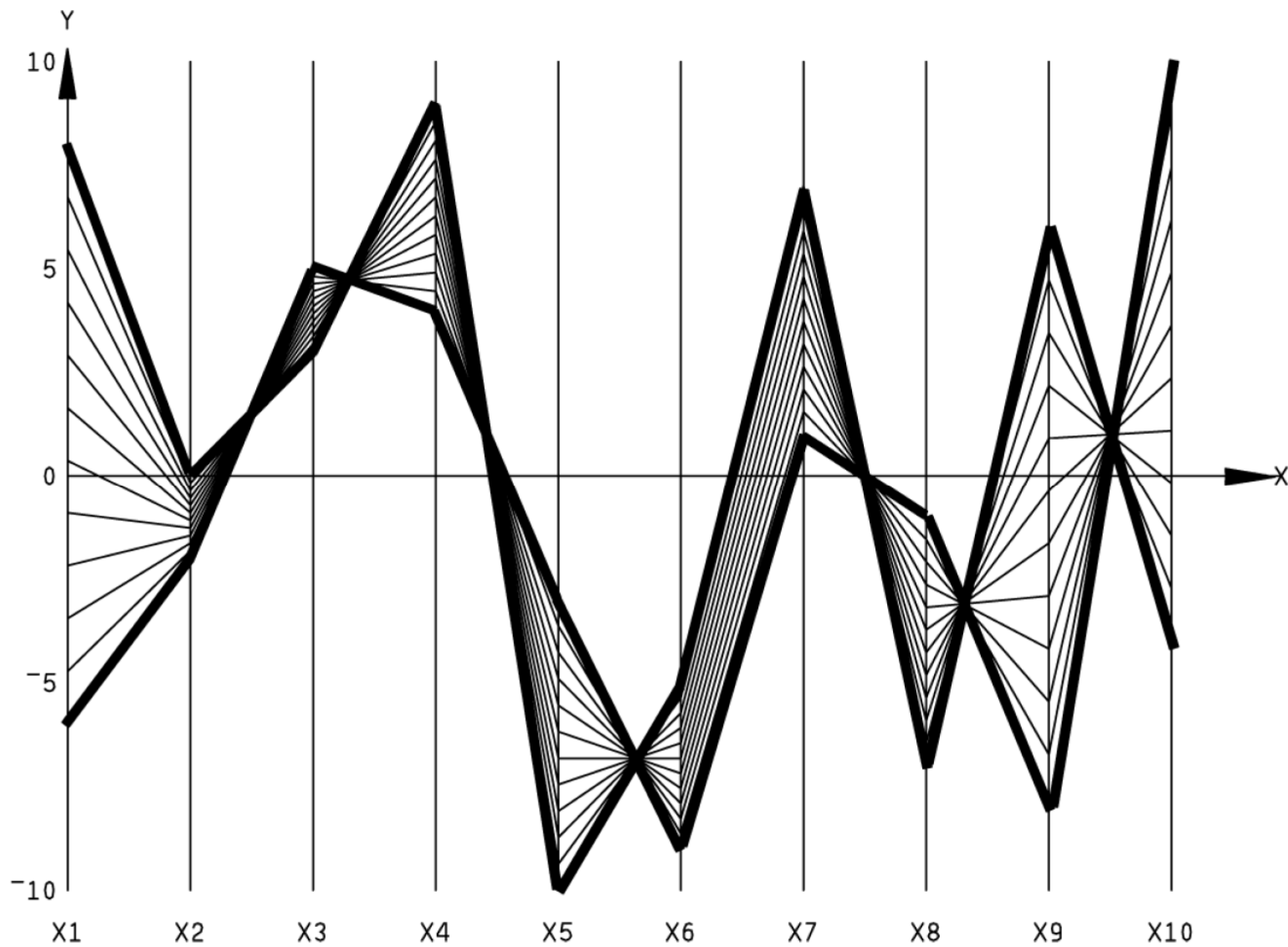
Inselberg, A. (2005) Visualization of concept formation and learning. *Kybernetes: The International Journal of Systems and Cybernetics*, 34, 1/2, 151-166.

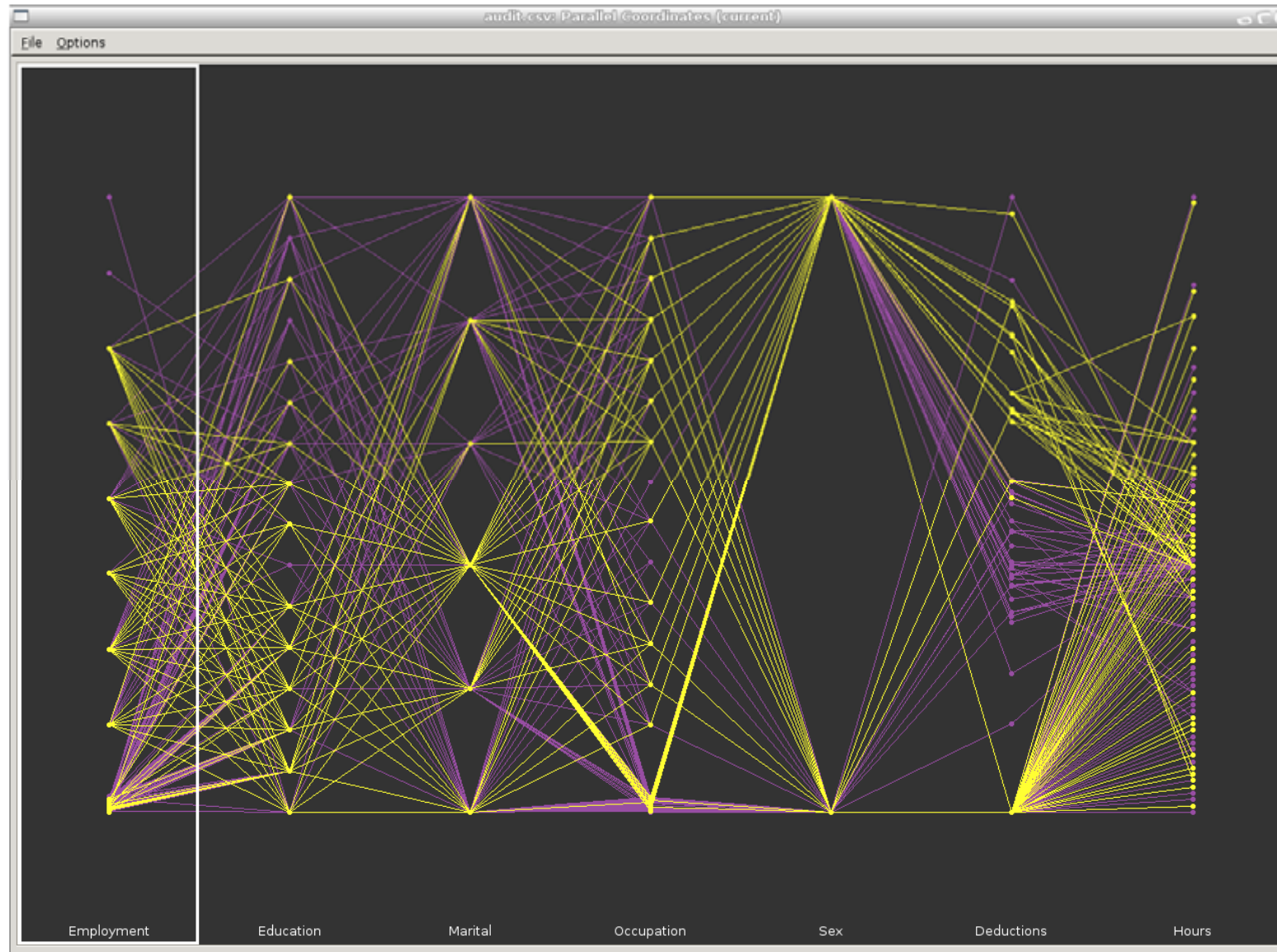
$$\mathbb{R}^5: C = (c_1, c_2, c_3, c_4, c_5)$$



Inselberg (2005)

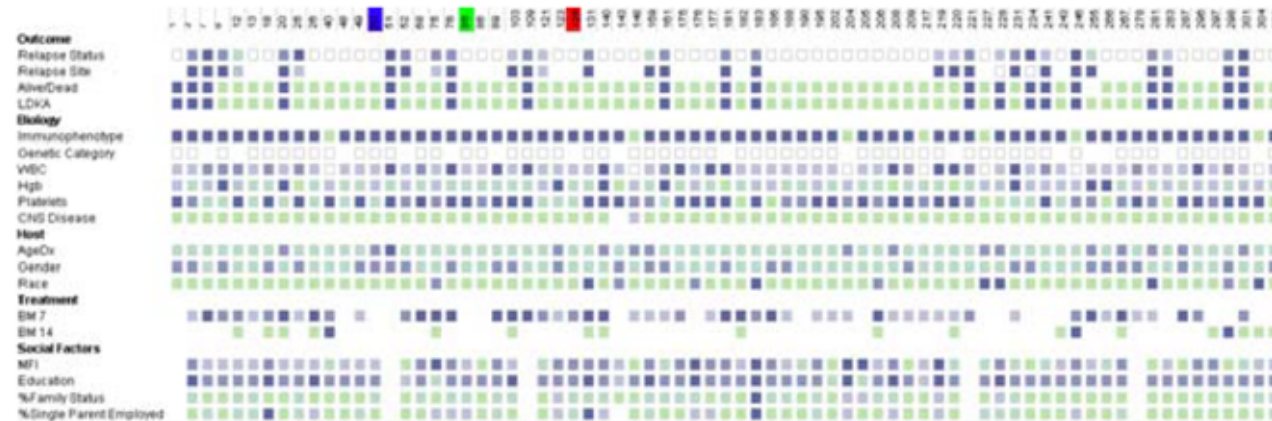
- A polygonal line $\bar{P}$ on the $N - 1$ points represents a point
- $P = (p_1, \dots p_{i-1}, p_i \dots p_N) \in \ell$
- since the pair of values $\dots p_{i-1}, p_i$ marked on the $\bar{X}_{i-1}$ and $\bar{X}_i$ axes.
- In the following slide we see several polygonal lines, intersecting at $\ell_{(i-1),i}$
- representing data points on a line $\ell \subset \mathbb{R}^{10}$.
- Note: The indexing is essential and is important for the visualization of proximity properties such as the minimum distance between a pair of lines.



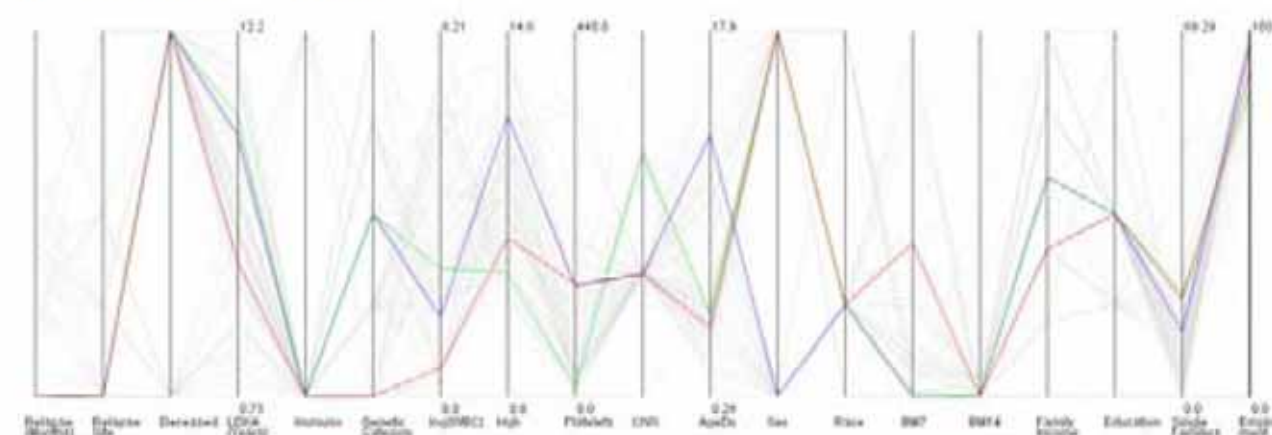


<http://datamining.togaware.com>

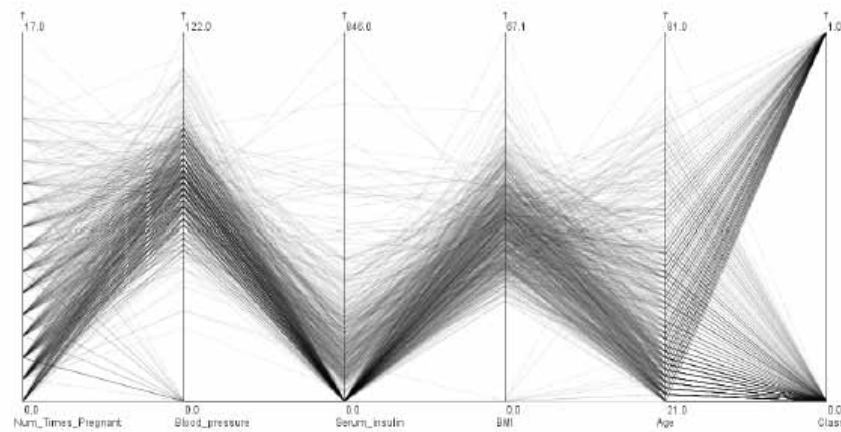
A – Matrix view



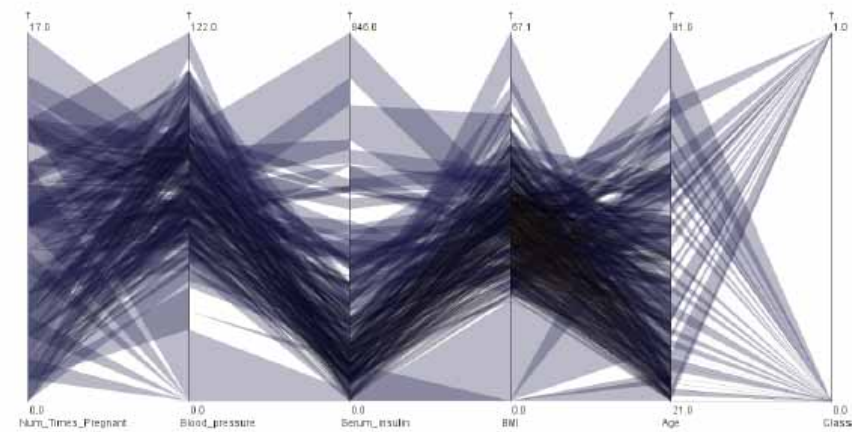
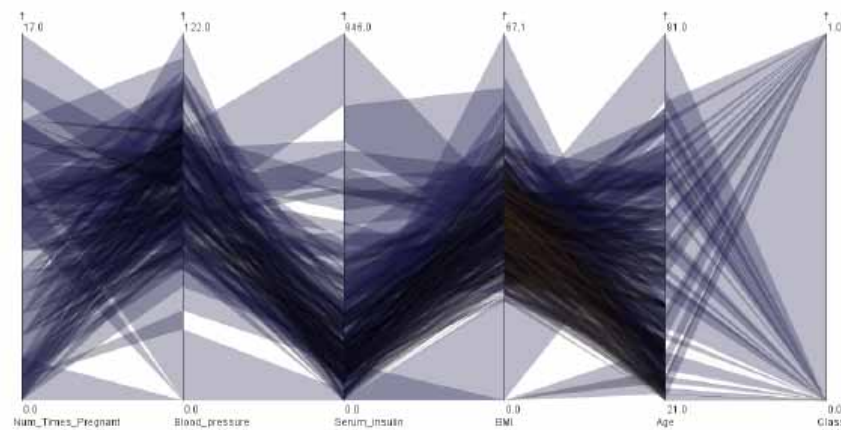
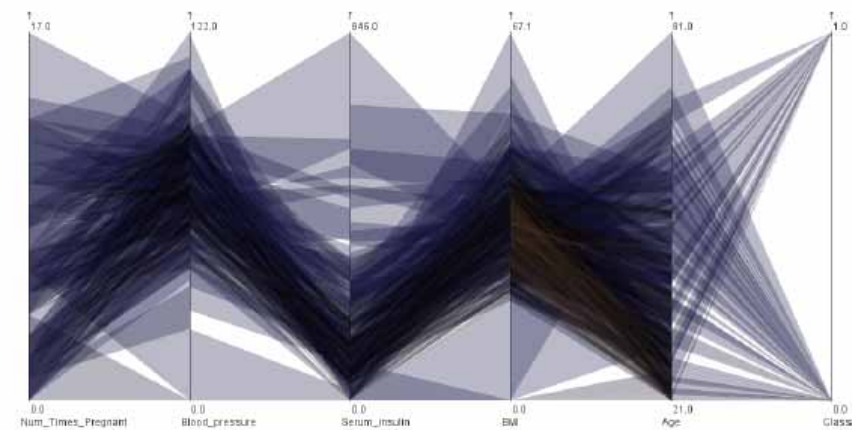
B – Parallel coordinates view



Mane, K. K. & Börner, K. (2007) Computational Diagnostic: A Novel Approach to View Medical Data. Los Alamos National Laboratory.

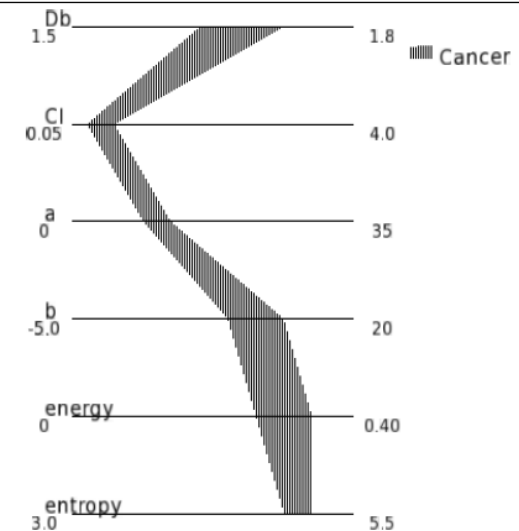
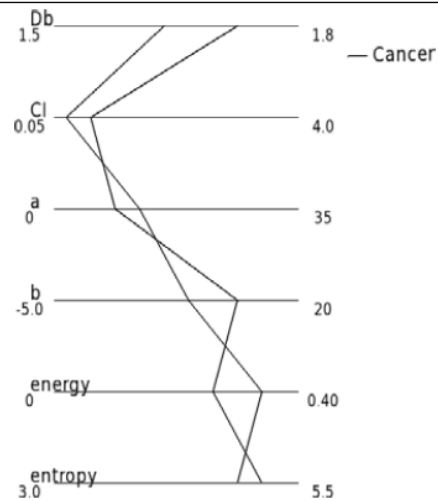
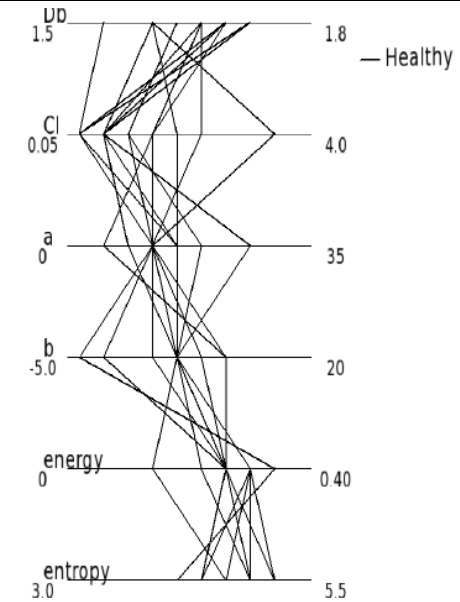
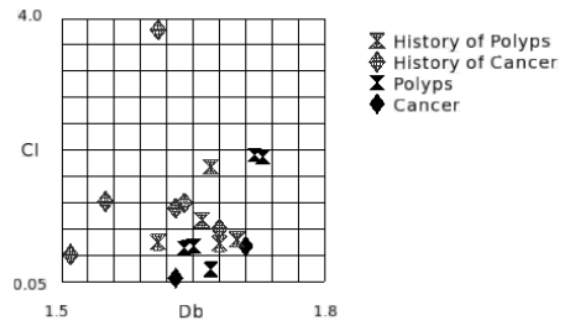
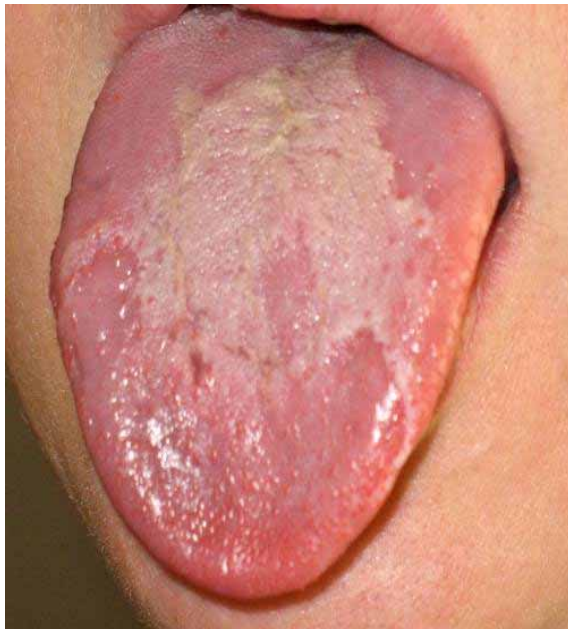


(a) Original View of the raw dataset

(b) Anonymization with $k=2$ (c) Anonymization with $k=3$ (d) Anonymization with $k=4$

Dasgupta, A. & Kosara, R. (2011). *Privacy-preserving data visualization using parallel coordinates*. Visualization and Data Analysis 2011, San Francisco, SPIE.

**Why are such
approaches not used
in enterprise hospital
information systems?**



Pham, B. L. & Cai, Y.
(2004) Visualization
techniques for tongue
analysis in traditional
Chinese medicine.



Source: Stem Cell Institute, Online: <http://www.cellmedicine.com>

April 30, 1968

M. J. FULWYLER
PARTICLE SEPARATOR

3,380,584

Filed June 4, 1965

5 Sheets-She

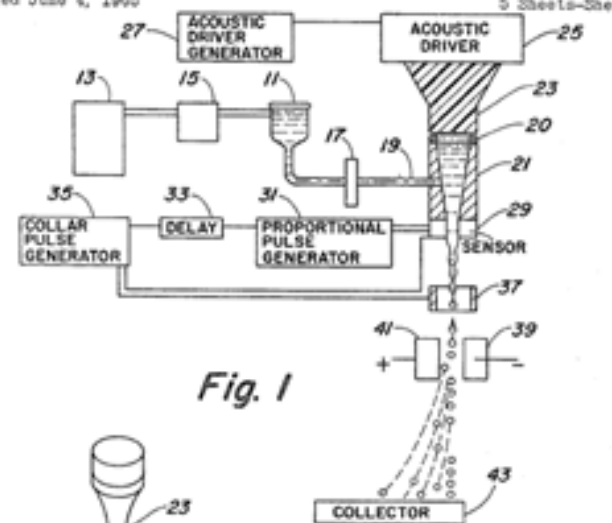


Fig. 1

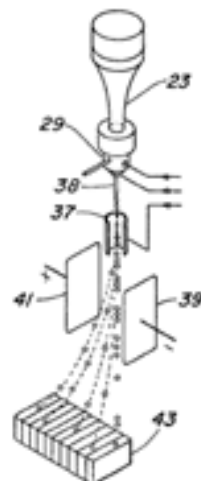


Fig. 2

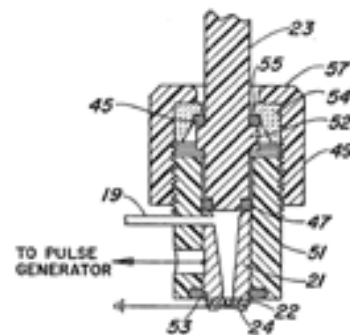
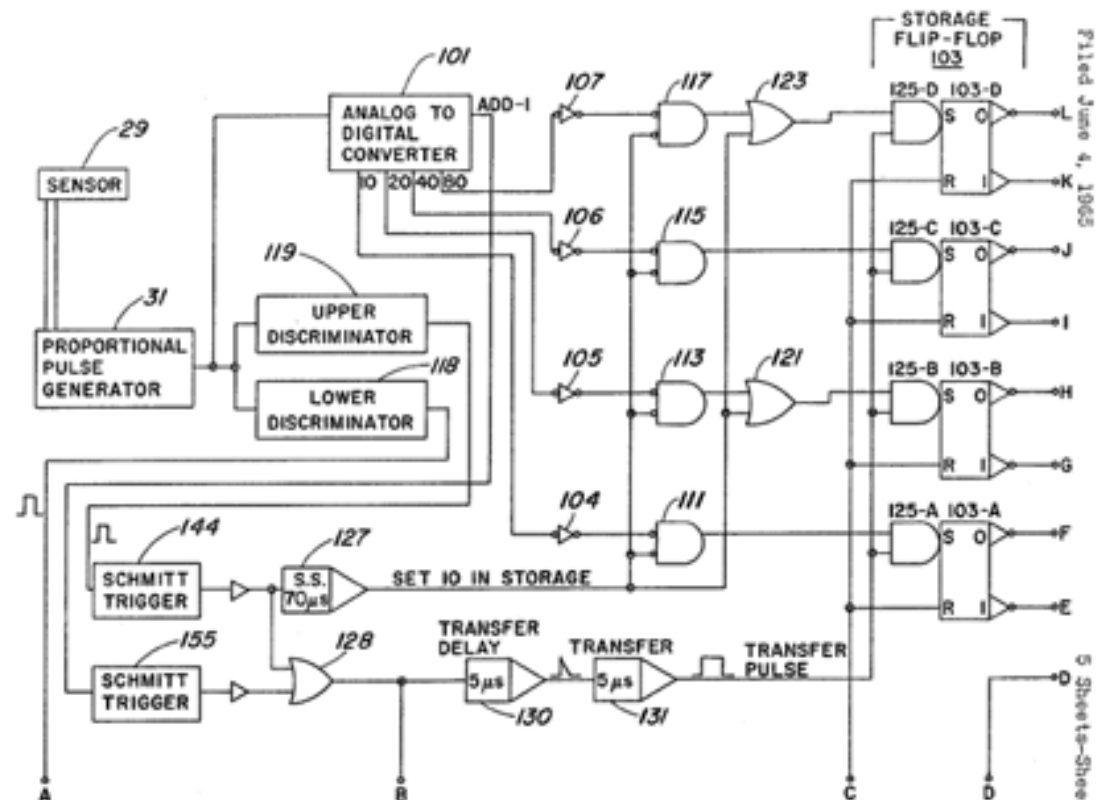
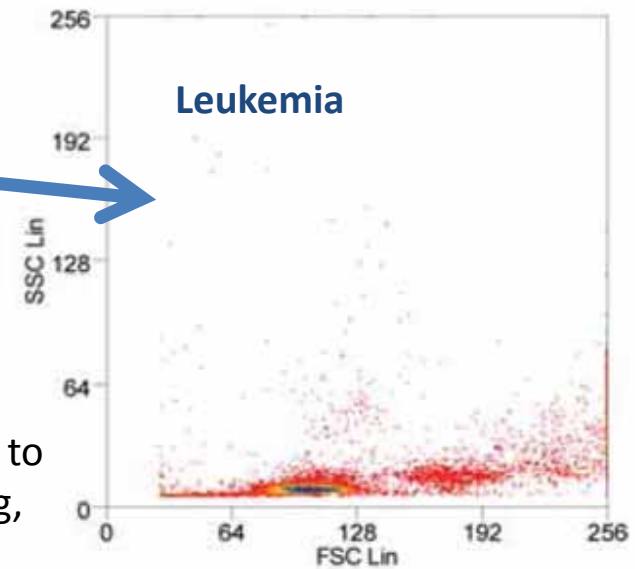
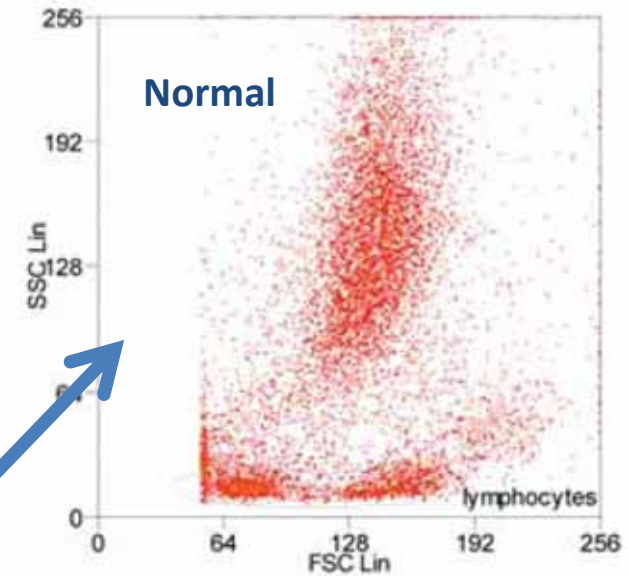
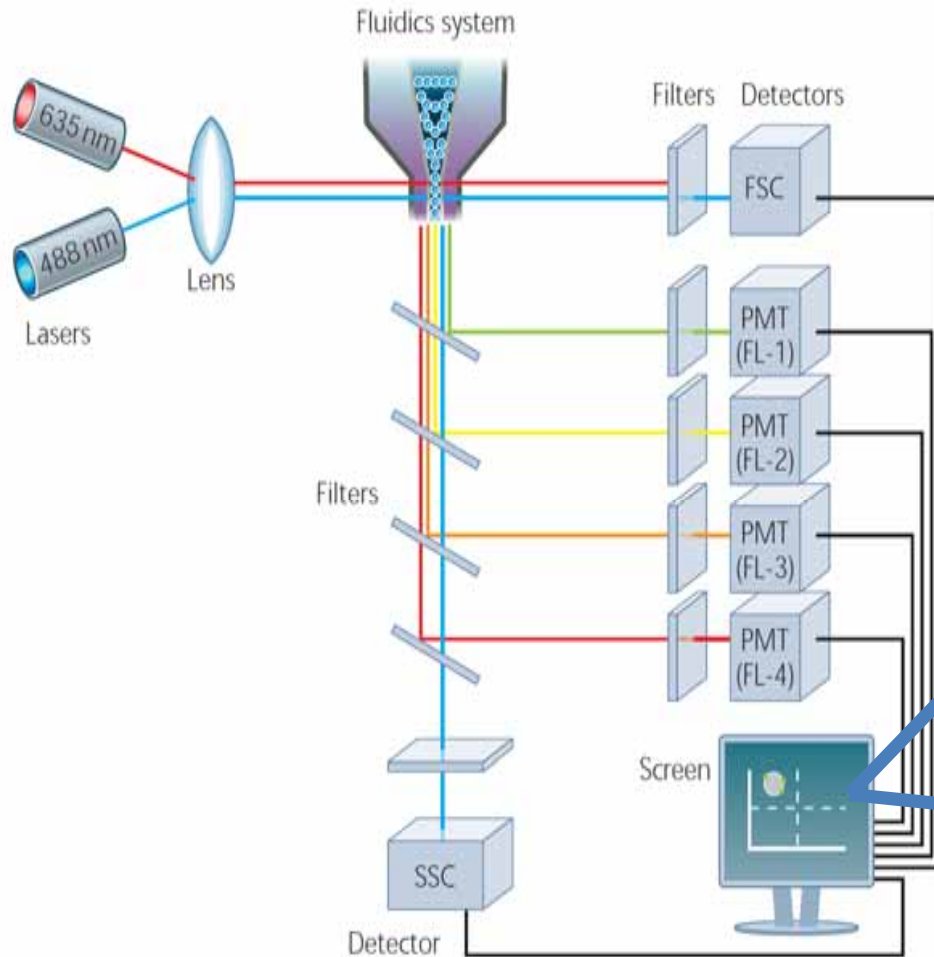


Fig. 3

INVENTOR:
Mack J. FulwylerBY
Richard A. Anderson
Attorney

Fulwyler, M. J. (1968) US Patent 3380584 A
Particle Separator, 1965 applied, 1968 published

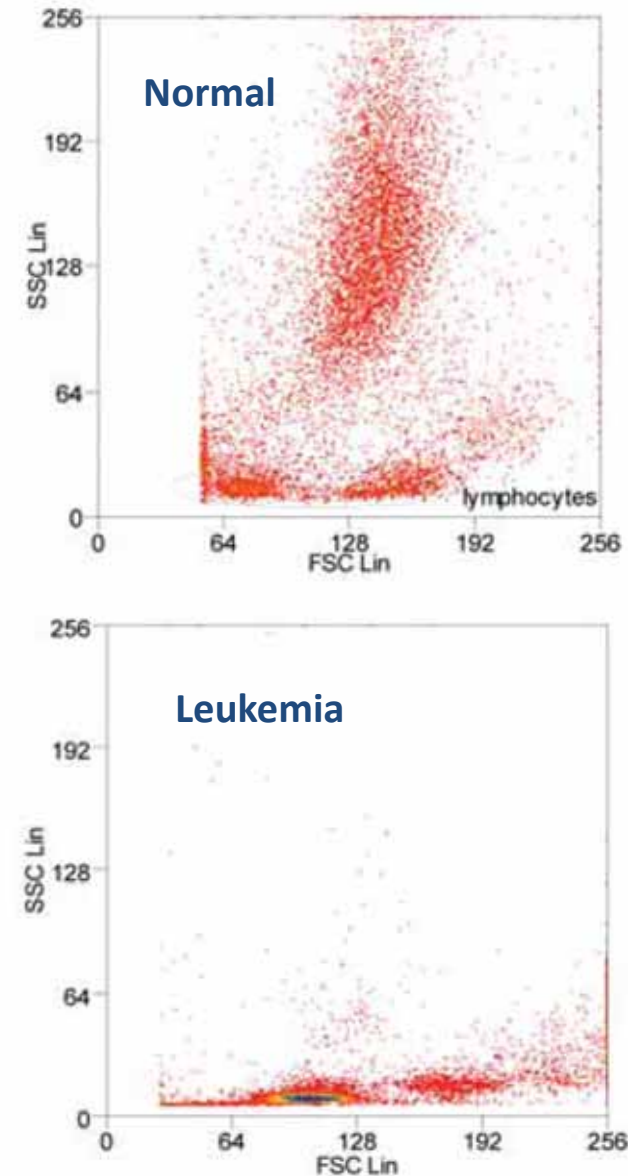
Fulwyler, M. J. (1965) Electronic Separation of Biological Cells by Volume. *Science*, 150, 3698, 910-911.



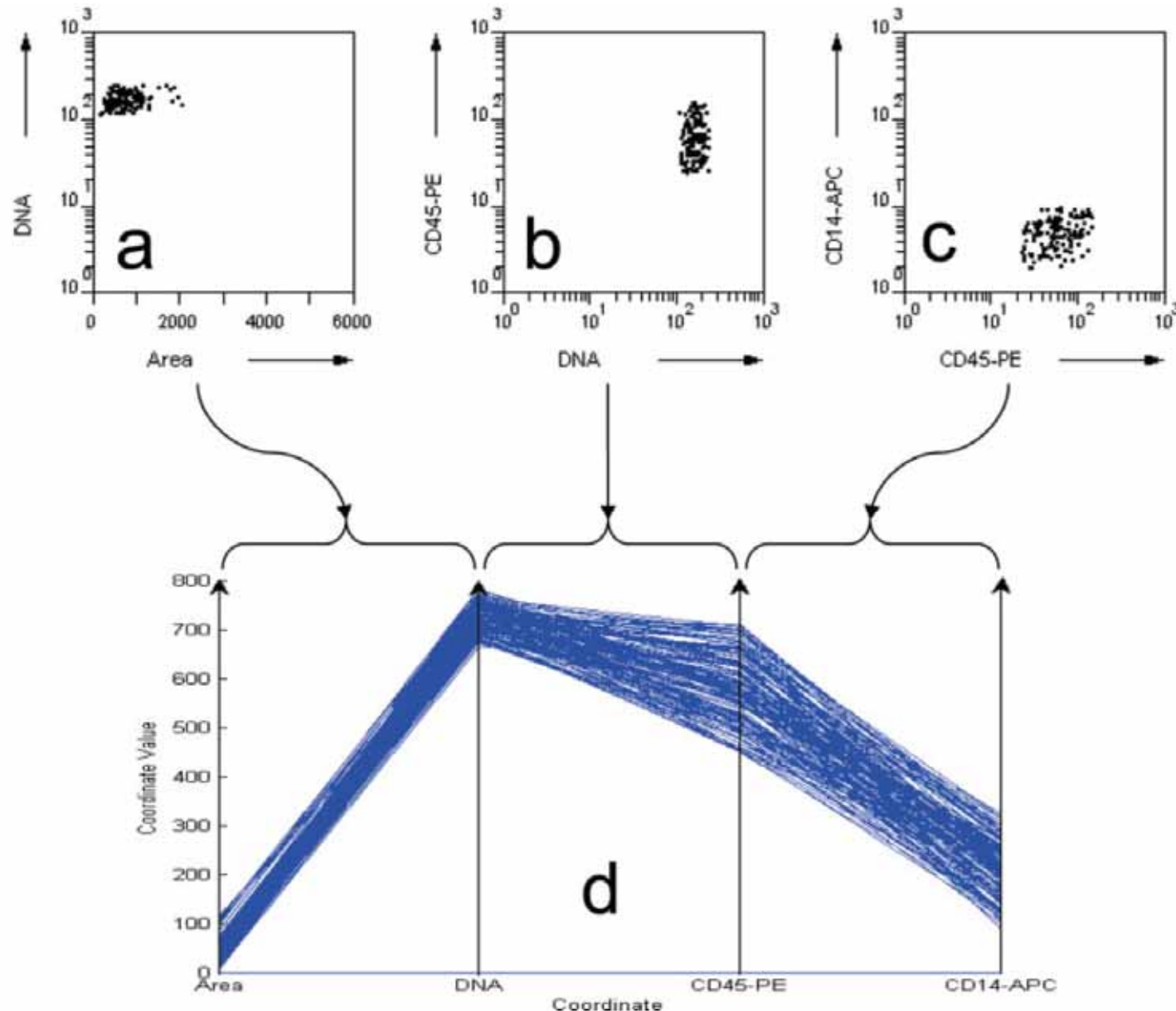
Rahman, M., Lane, A., Swindell, A. & Bartram, S. (2009) Introduction to Flow Cytometry: Principles, Data analysis, Protocols, Troubleshooting, Online available: www.abdserotec.com.

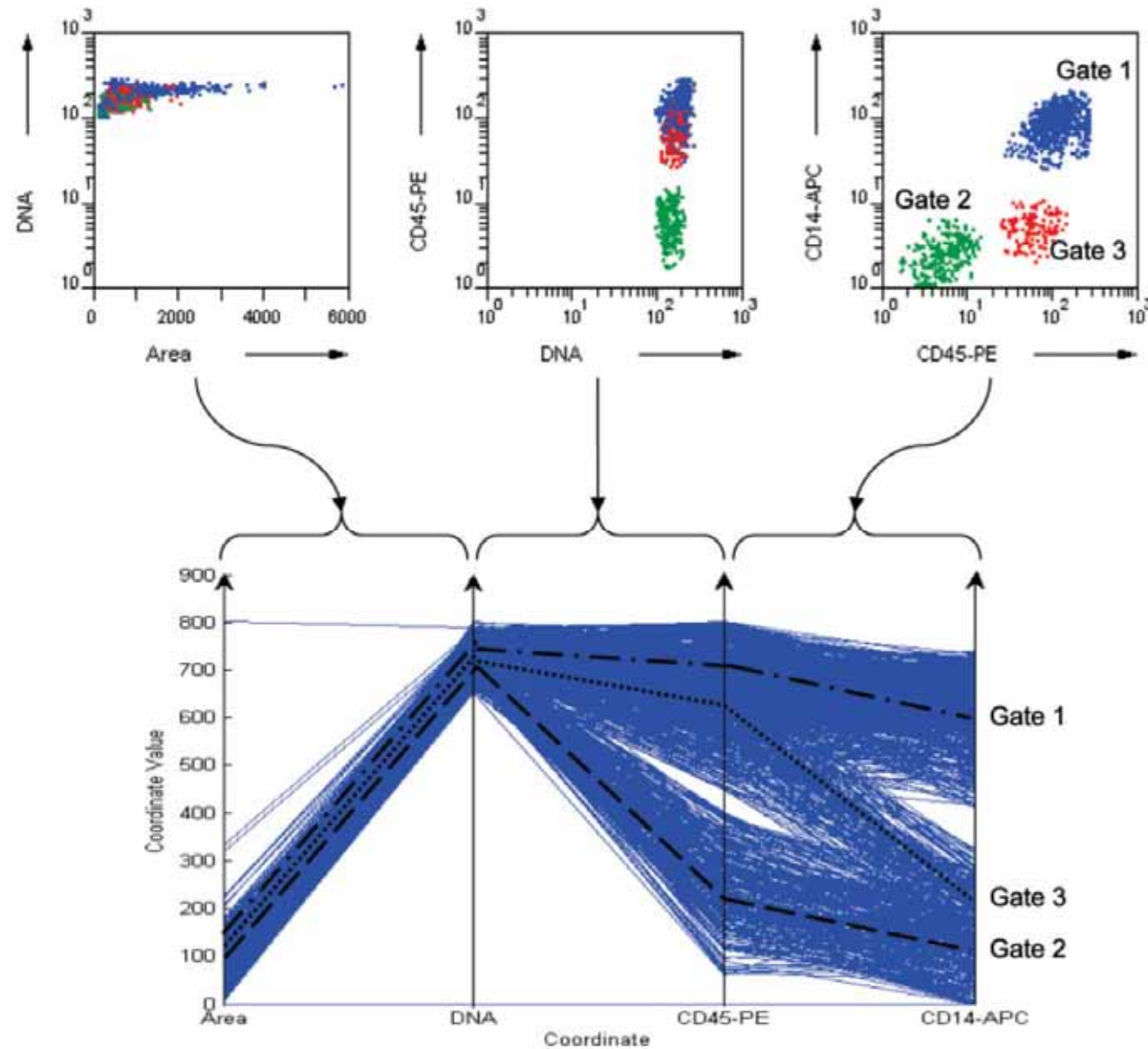
- Forward scatter channel (FSC) intensity equates to the particle's size and can also be used to distinguish between cellular debris and living cells.
- Side scatter channel (SSC) provides information about the granular content within a particle.
- Both FSC and SSC are unique for every particle, and a combination of the two may be used to differentiate different cell types in a heterogeneous sample.

Rahman et al. (2009)

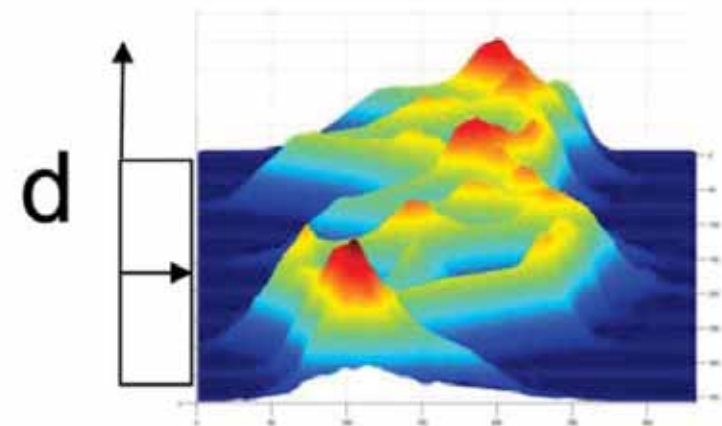
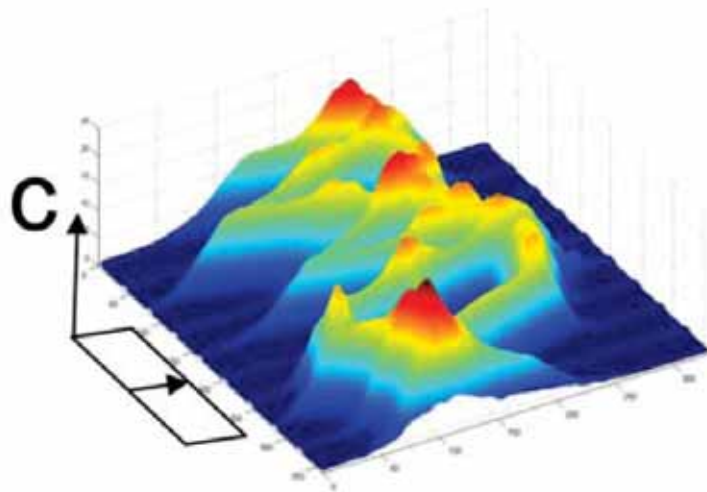
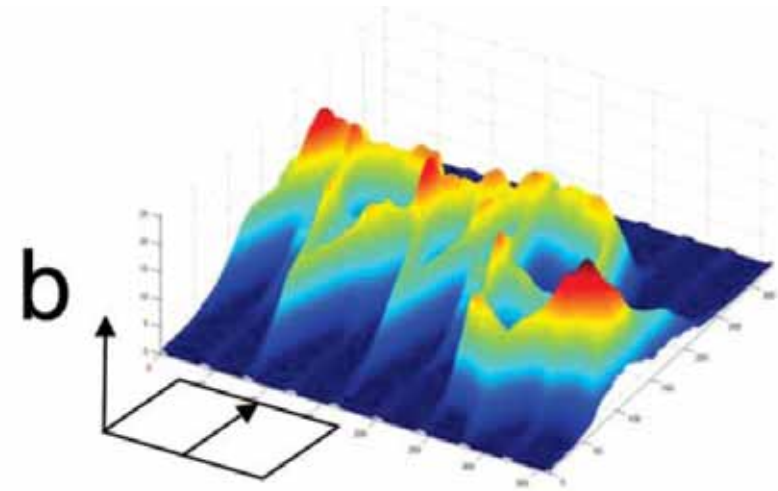
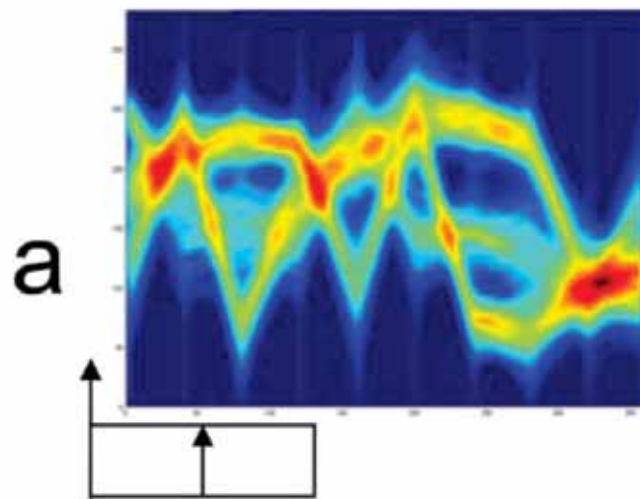


Streit, M., Ecker, R. C., Österreich, K., Steiner, G. E., Bischof, H., Bangert, C., Kopp, T. & Rogojanu, R. (2006) 3D parallel coordinate systems—A new data visualization method in the context of microscopy-based multicolor tissue cytometry. *Cytometry Part A*, 69A, 7, 601-611.



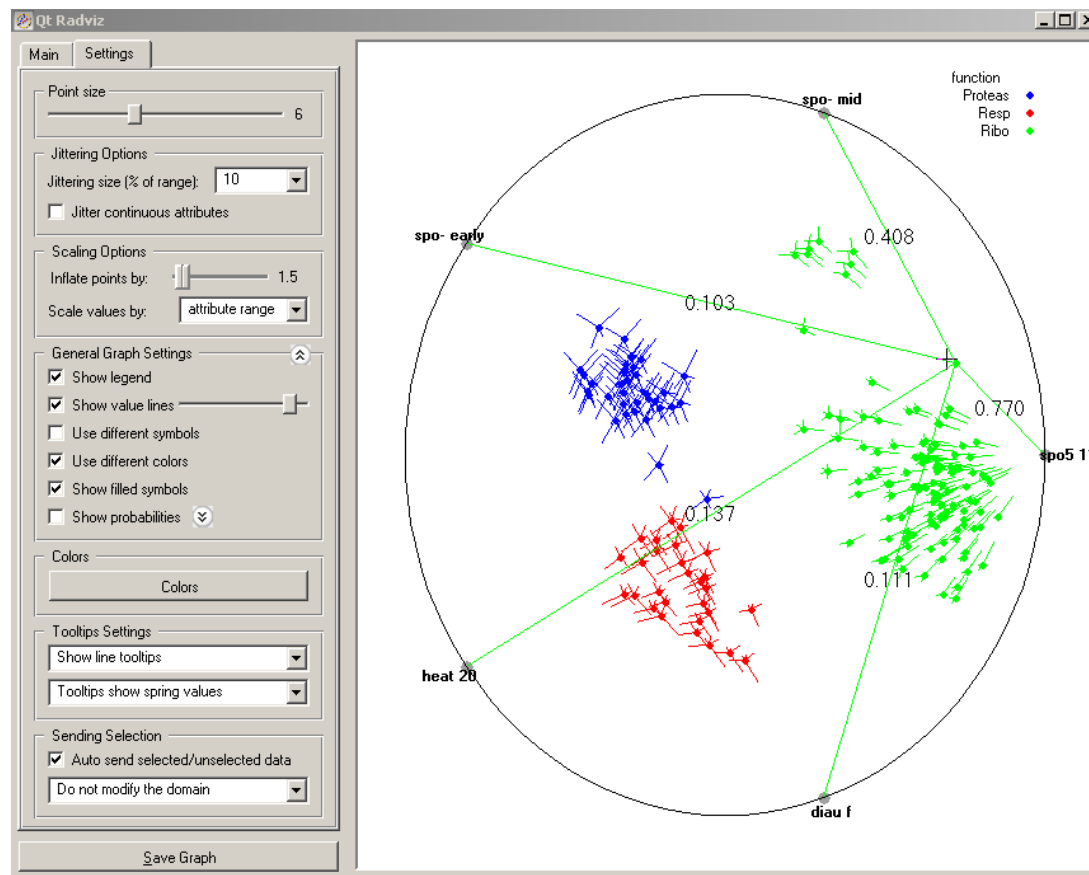


Streit et al. (2006)

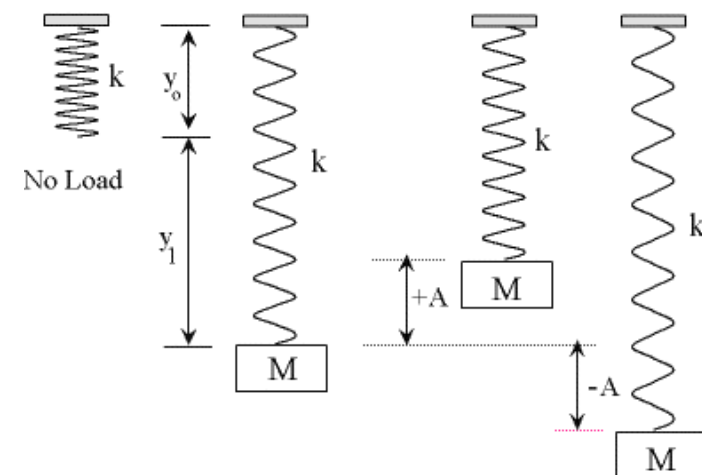
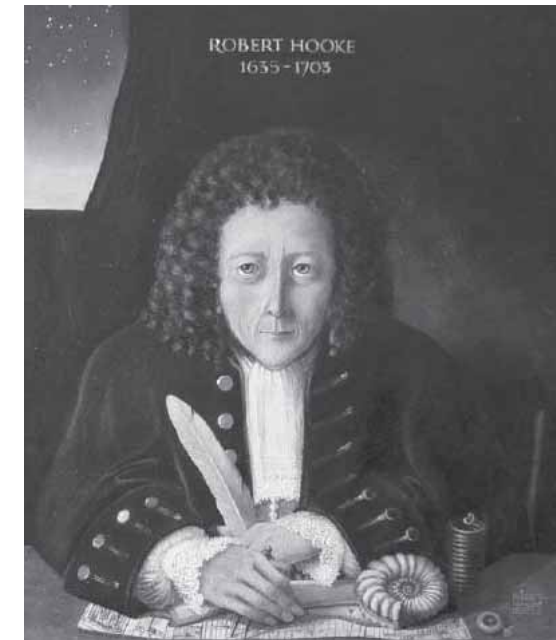


Streit et al. (2006)

Demšar, J., Curk, T., & Erjavec, A. Orange: Data Mining Toolbox in Python; Journal of Machine Learning Research 14:2349–2353, 2013.



Source: <http://orange.biolab.si/>



$$\omega = \sqrt{\frac{k}{M}} \quad \text{where} \quad \omega = 2\pi f.$$

- 1) Let us consider a point $y_i = (y_1, y_2, \dots, y_n)$ from the n -dimensional space
- 2) This point is now mapped into a single point u in the plane of anchors: for each anchor j the stiffness of its spring is set to y_j
- 3) Now the Hooke's law is used to find the point u , where all the spring forces reach equilibrium (means they sum to 0). The position of $u = [u_1, u_2]$ is now derived by:

$$\sum_{j=1}^n (\vec{S}_j - \vec{u}) y_j = 0$$

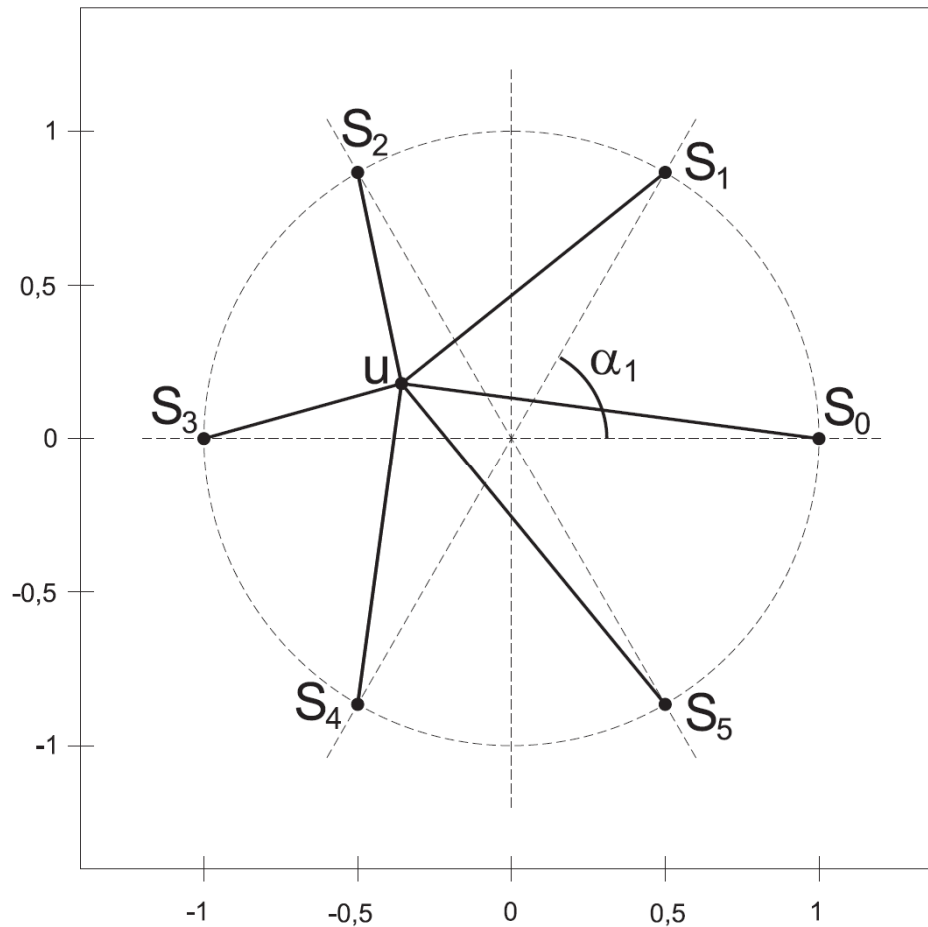
$$\sum_{j=1}^n \vec{S}_j y_j = \vec{u} \sum_{j=1}^n y_j$$

$$\vec{u} = \frac{\sum_{j=1}^n \vec{S}_j y_j}{\sum_{j=1}^n y_j}$$

$$u_1 = \frac{\sum_{j=1}^n y_j \cos(\alpha_j)}{\sum_{j=1}^n y_j}$$

$$u_2 = \frac{\sum_{j=1}^n y_j \sin(\alpha_j)}{\sum_{j=1}^n y_j}$$

Novakova, L. & Stepankova, O. (2009). *RadViz and Identification of Clusters in Multidimensional Data*. 13th International Conference on Information Visualisation, 104-109.



1. Normalize the data to the interval $\langle 0, 1 \rangle$

$$\bar{x}_{ij} = \frac{x_{ij} - \min_j}{\max_j - \min_j}$$

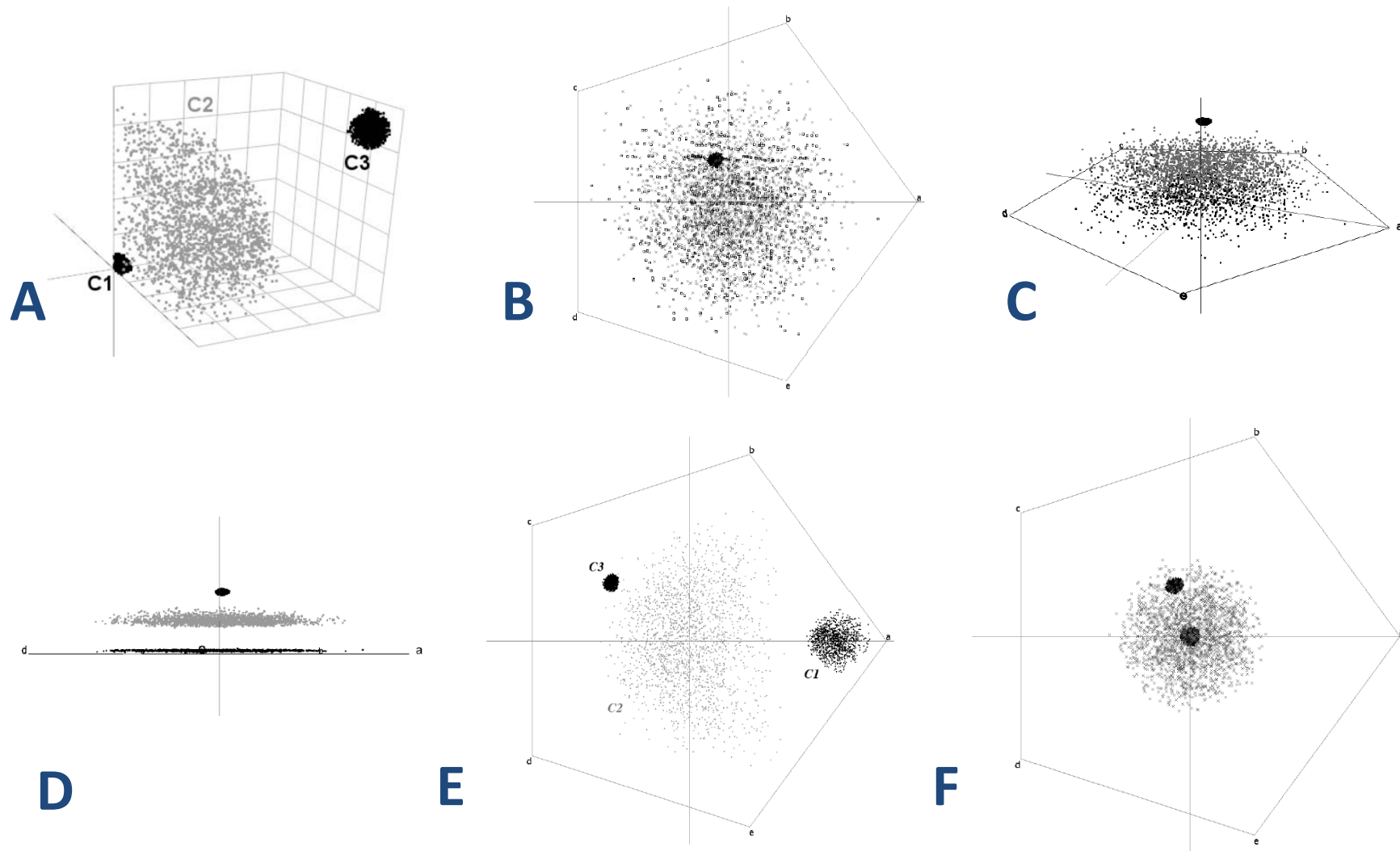
2. Now place the dimensional anchors

3. Now calculate the point to place each record and to draw it:

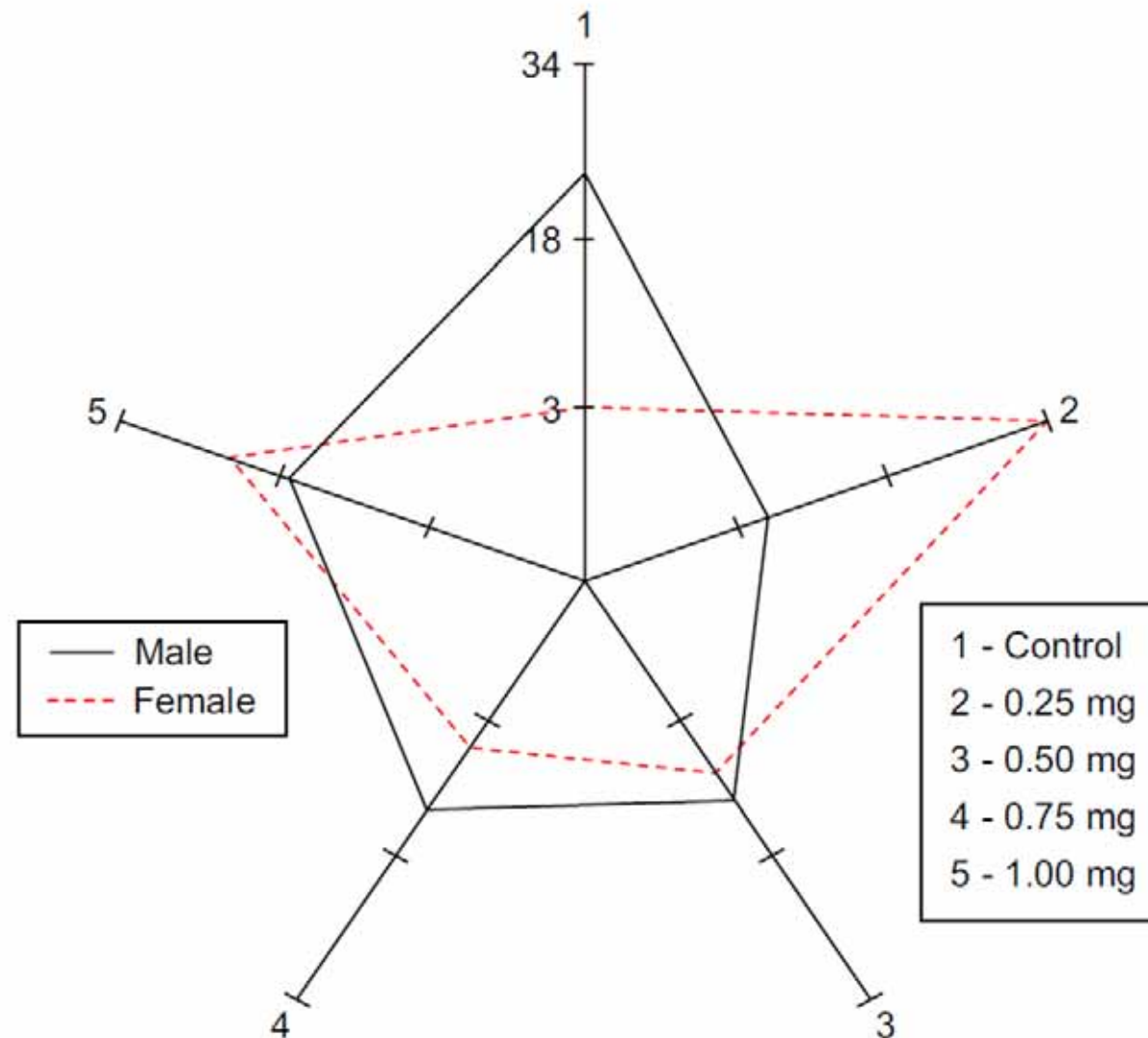
$$y_i = \sum_{j=1}^n \bar{x}_{ij}$$

$$\vec{u}_i = \frac{\sum_{j=1}^n \vec{S}_j \bar{x}_{ij}}{y_i}$$

Novakova, L. & Stepankova, O. (2009). *RadViz and Identification of Clusters in Multidimensional Data*. 13th International Conference on Information Visualisation, 104-109.



Novakova, L. & Stepankova, O. (2009). *RadViz and Identification of Clusters in Multidimensional Data*. 13th International Conference on Information Visualisation, 104-109.



Saary, M. J. (2008) Radar plots: a useful way for presenting multivariate health care data. *Journal Of Clinical Epidemiology*, 61, 4, 311-317.

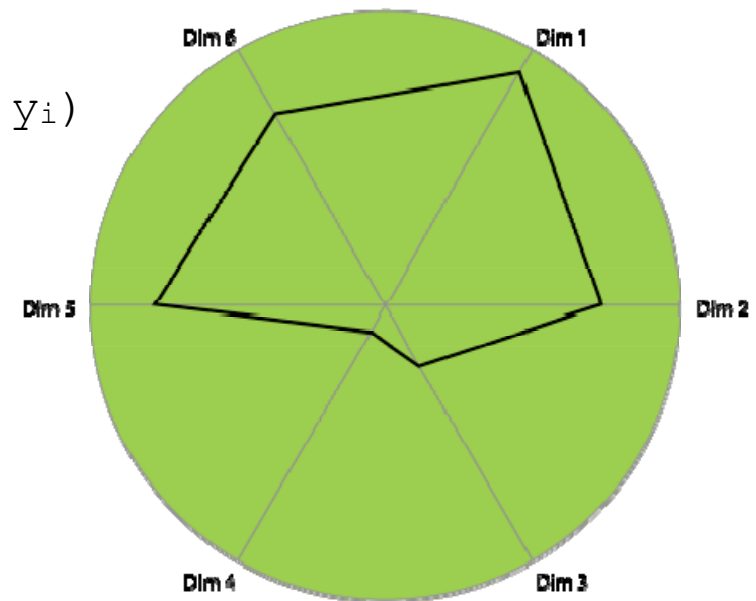
- Arrange N axes on a circle in $\mathbb{R}^2$
- $3 \leq N \leq N_{max}$
Note: An amount of $N_{max} \leq 20$ is just useful, according to Lanzenberger et al. (2005)
- Map coordinate vectors $P \in \mathbb{R}^N$ from $\mathbb{R}^N \rightarrow \mathbb{R}^2$
- $P = \{p_1, p_2, \dots, p_N\} \in \mathbb{R}^N$ where each p_i represents a different attribute with a different physical unit
- Each axis represents one attribute of data
- Each data record, or data point P is visualized by a line along the data points
- A Line is perceived better than points on the axes

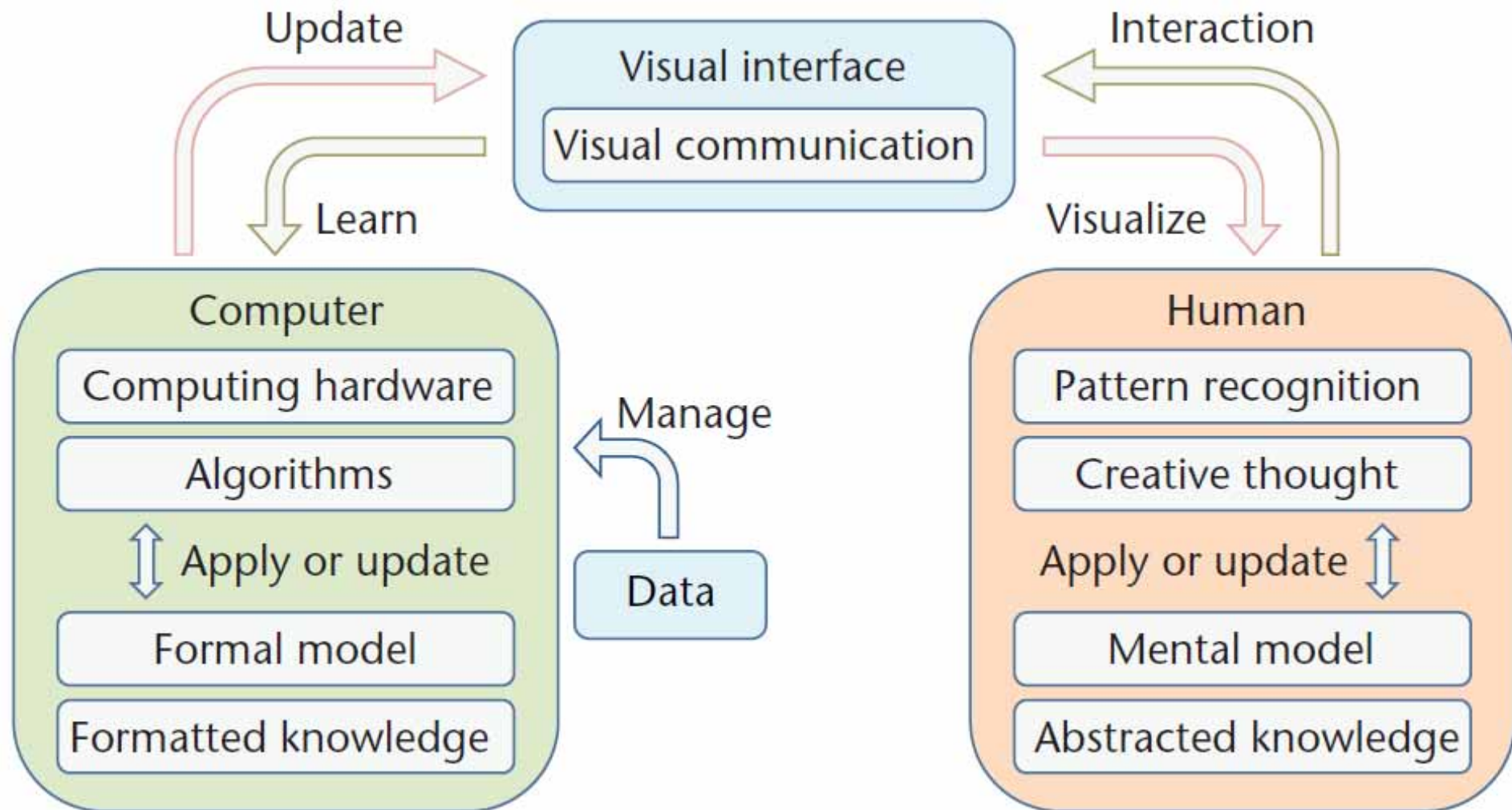
```
angle_sector = 2 * π / N
for each ai from axes[]
{
    anglei = i * angle_sector
    xi = mid.x + r * cos(anglei)
    yi = mid.y + r * sin(anglei)
    DrawLine(midpoint.x, midpoint.y, xi, yi)
```

```
maxi = ai.upperBound()
```

```
scaled_vali = ai.value() * r / maxi
x_vali = mid.x + scaled_vali * cos(anglei)
```

```
y_vali = mid.y + scaled_vali * sin(anglei)
DrawLine(x_vali, y_vali, x_vali-1, y_vali-1)
}
```





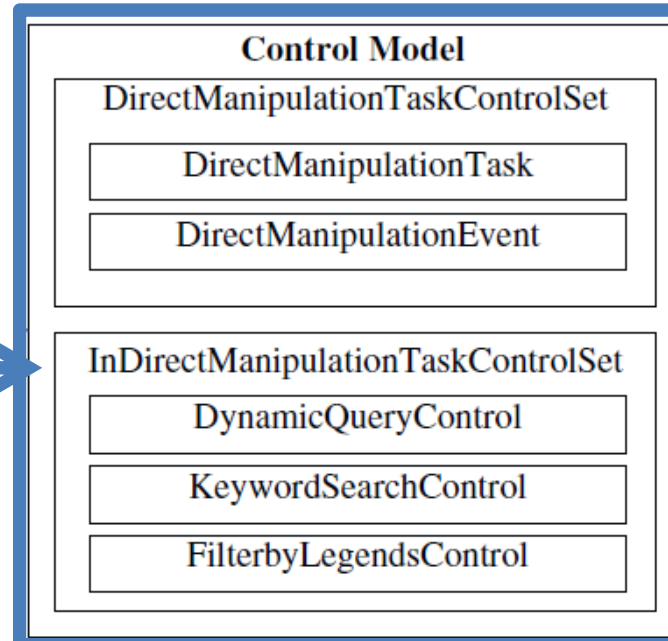
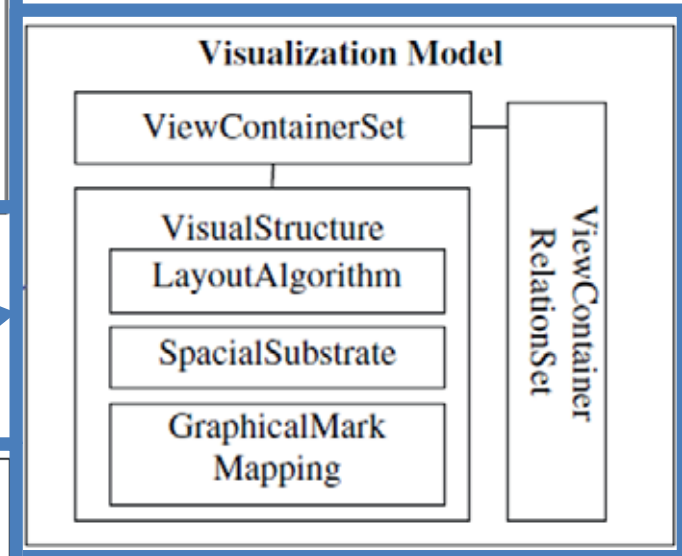
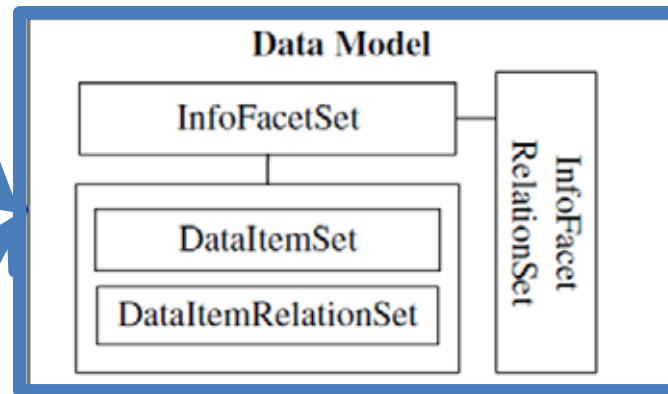
Mueller, K., Garg, S., Nam, J. E., Berg, T. & McDonnell, K. T. (2011) Can Computers Master the Art of Communication?: A Focus on Visual Analytics. *Computer Graphics and Applications, IEEE*, 31, 3, 14-21.

1) What facets of the target information should be visualized?

2) What data source should each facet be linked to and what relationships these facets have?

3) What layout algorithm should be used to visualize each facet?

4) What interactive techniques should be used for each facet and for which infovis tasks?



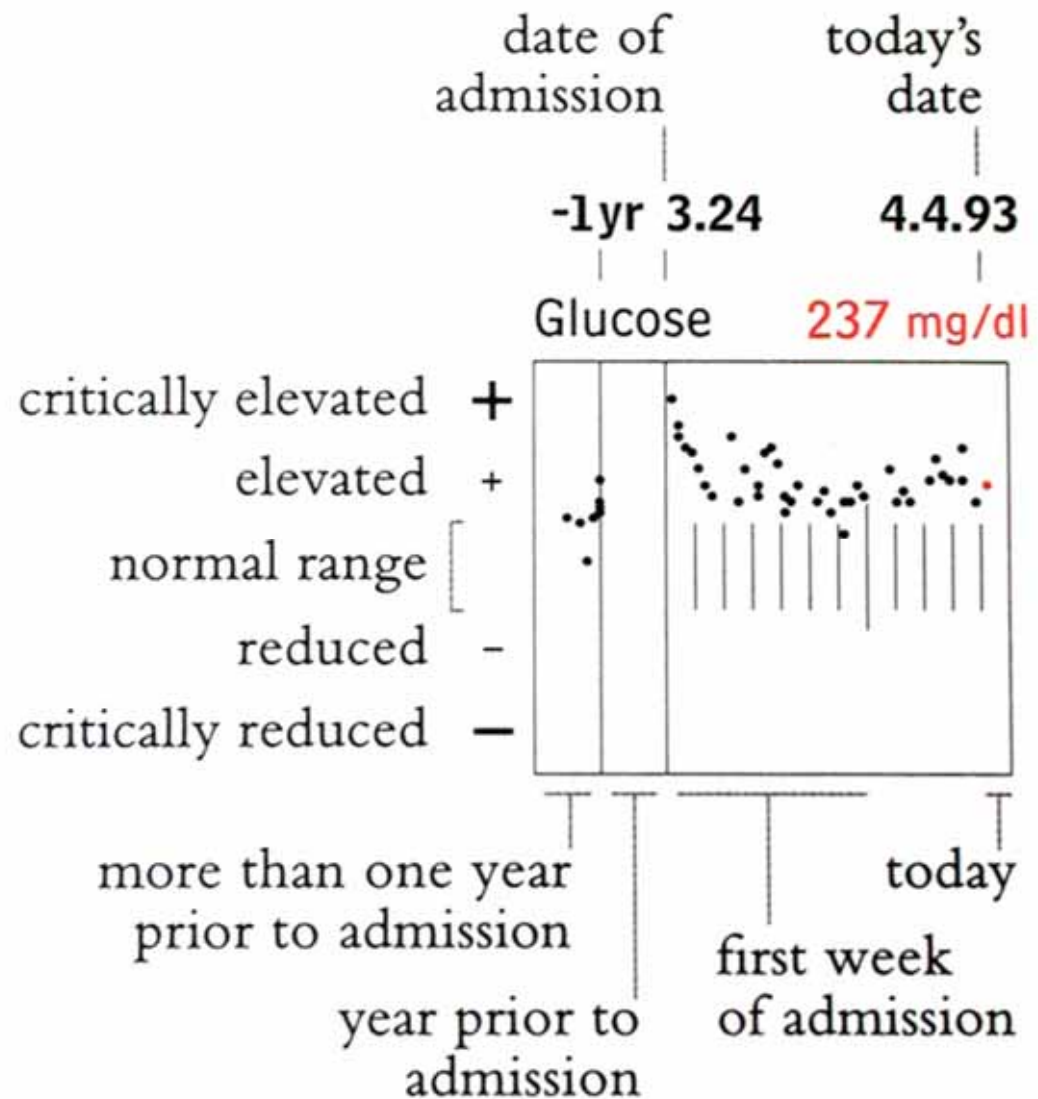
Ren, L., Tian, F., Zhang, X. & Zhang, L. (2010) DaisyViz: A model-based user interface toolkit for interactive information visualization systems. *Journal of Visual Languages & Computing*, 21, 4, 209-229.

- 1) Overview: Gain an overview about the entire data set (know your data!);
- 2) Zoom : Zoom in on items of interest;
- 3) Filter: filter out uninteresting items – get rid of distractors – eliminate irrelevant information;
- 4) Details-on-demand: Select an item or group and provide details when needed;
- 5) Relate: View relationships among items;
- 6) History: Keep a history of actions to support undo, replay, and progressive refinement;
- 7) Extract: Allow extraction of sub-collections and of the query parameters;

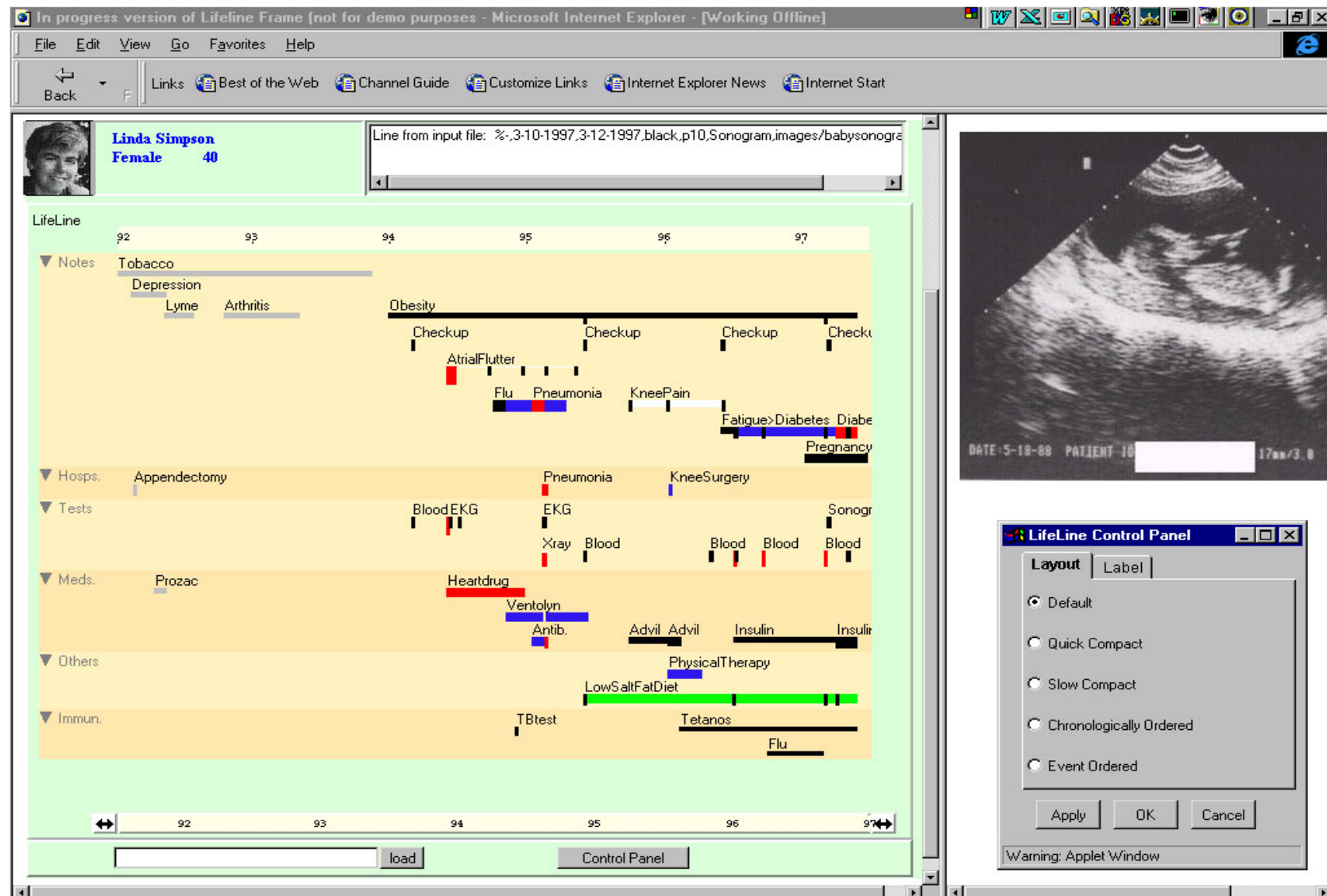
*) Shneiderman, B. (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proceedings of the 1996 IEEE Symposium on Visual Languages*, 336-343.

- Focus Selection = via direct manipulation and selection tools, e.g. multi-touch (in data space a n-dim location might be indicated);
- Extent Selection = specifying extents for an interaction, e.g. via a vector of values (a range for each data dimension or a set of constraints;
- Interaction type selection = e.g. a pair of menus: one to select the space, and the other to specify the general class of the interaction;
- Interaction level selection = e.g. the magnitude of scaling that will occur at the focal point (via a slider, along with a reset button)

Ward, M., Grinstein, G. & Keim, D. (2010) *Interactive Data Visualization: Foundations, Techniques and Applications*. Natick (MA), Peters.



Powsner, S. M. & Tufte, E. R. (1994)
Graphical Summary of Patient
status. *The Lancet*, 344, **8919**, 386-
389.



Plaisant, C., Milash, B., Rose, A., Widoff, S. & Shneiderman, B. (1996). *Life Lines: Visualizing Personal Histories*. ACM CHI '96, Vancouver, BC, Canada, April 13-18, 1996.

What are temporal analysis tasks?



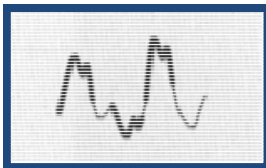
Classification = given a set of classes: the aim is to determine which class the dataset belongs to; a classification is often necessary as pre-processing;



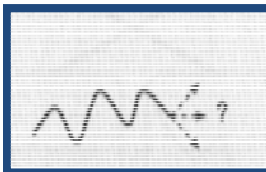
Clustering = grouping data into clusters based on similarity; the similarity measure is the key aspect of the clustering process;



Search/Retrieval = look for a priori specified queries in large data sets (query-by-example), can be exact matched or approximate matched (similarity measures are needed that define the degree of exactness);



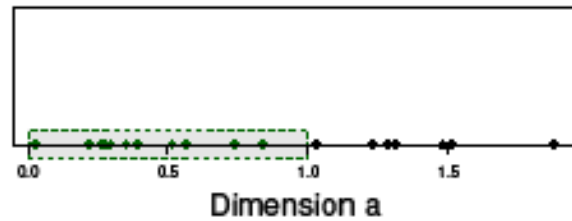
Pattern discovery = automatically discovering relevant patterns in the data, e.g. local structures in the data or combinations thereof;



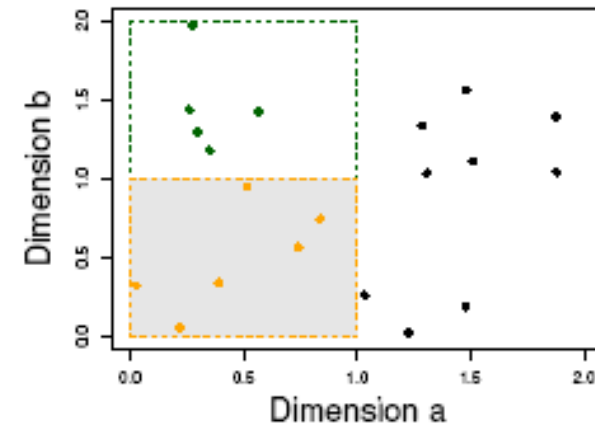
Prediction = foresee likely future behaviour of data – to infer from the data collected in the past and present how the data will evolve in the future (e.g. autoregressive models, rule-based models etc.)

Aigner, W., Miksch, S., Schumann, H. & Tominski, C. (2011) *Visualization of Time-Oriented Data. Human-Computer Interaction Series. London, Springer.*

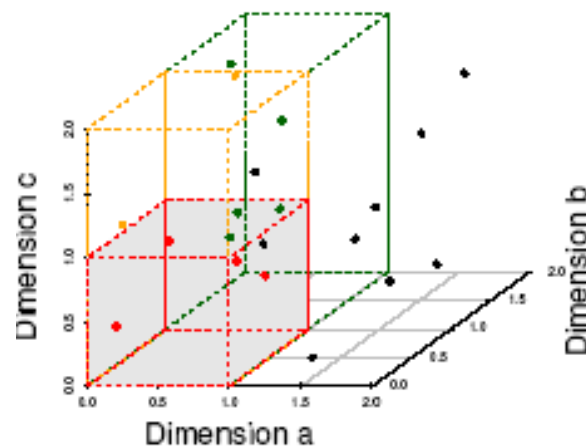
Remember: Subspace Clustering



(a) 11 Objects in One Unit Bin



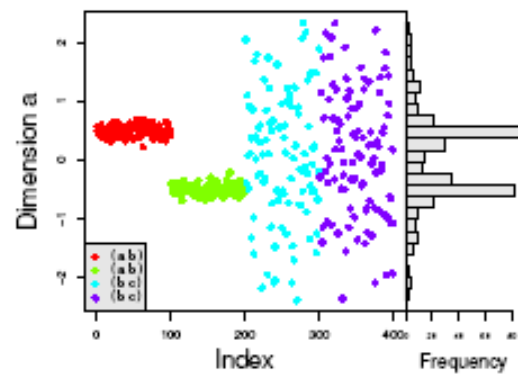
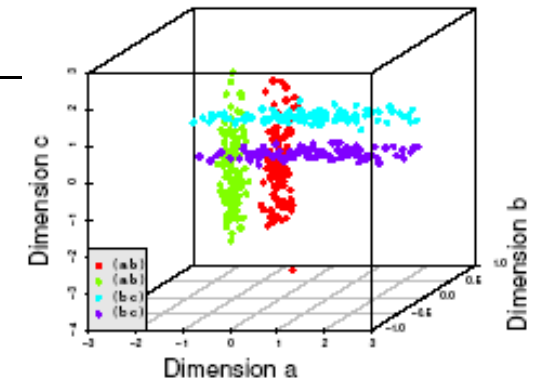
(b) 6 Objects in One Unit Bin



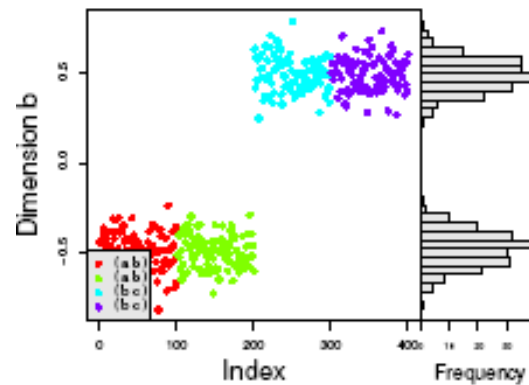
(c) 4 Objects in One Unit Bin

- Dataset - consists of a matrix of data values, rows represent individual instances and columns represent dimensions.
- Instance - refers to a vector of d measurements.
- Cluster - group of instances in a dataset that are more similar to each other than to other instances. Often, similarity is measured using a distance metric over some or all of the dimensions in the dataset.
- Subspace - is a subset of the d dimensions of a given dataset.
- Subspace Clustering – seek to find clusters in a dataset by selecting the most *relevant* dimensions for each cluster separately .
- Feature Selection - process of determining and selecting the dimensions (features) that are most relevant to the data mining task.

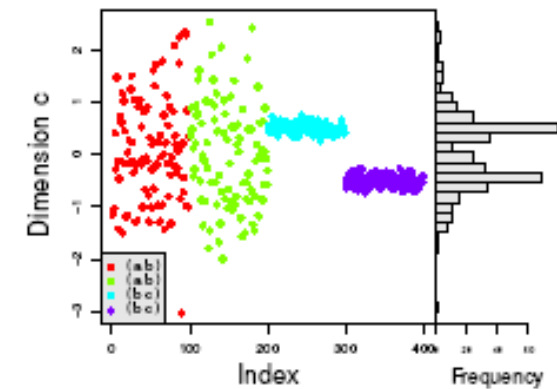
Parsons, L., Haque, E. & Liu, H. 2004. Subspace clustering for high dimensional data: a review. SIGKDD Explorations 6, (1), 90-105.



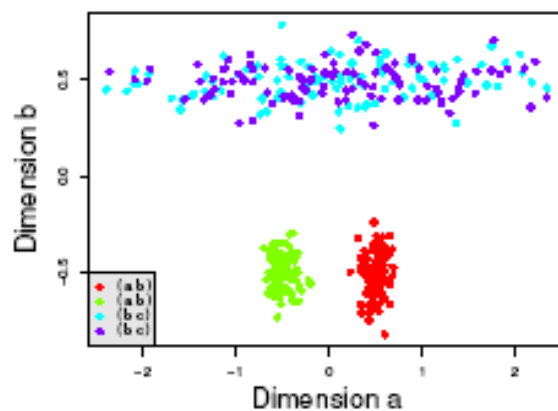
(a) Dimension *a*



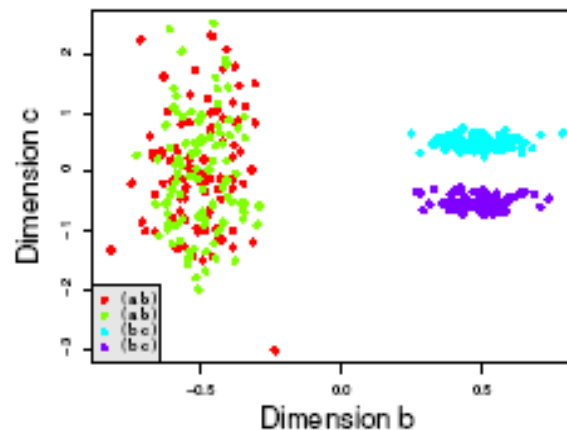
(b) Dimension *b*



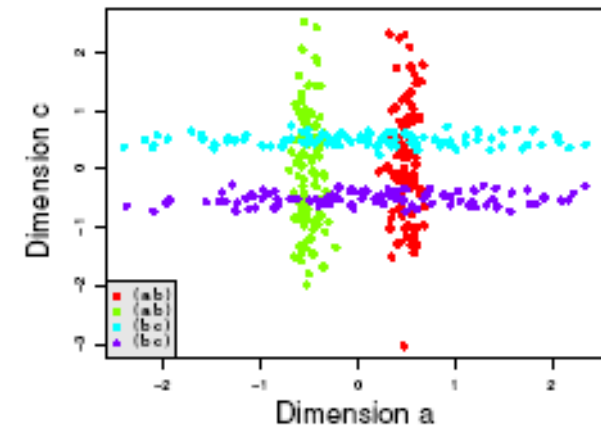
(c) Dimension *c*



(a) Dims *a* & *b*

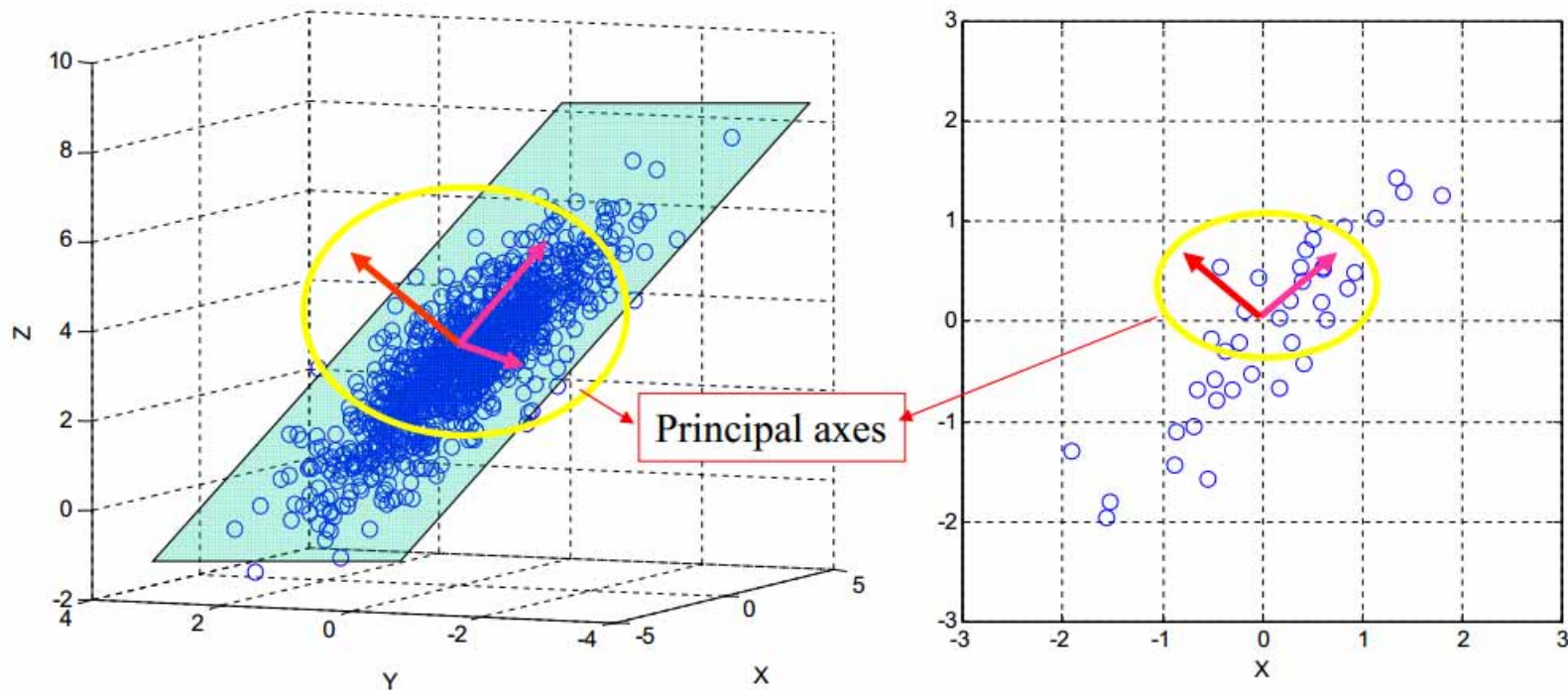


(b) Dims *b* & *c*



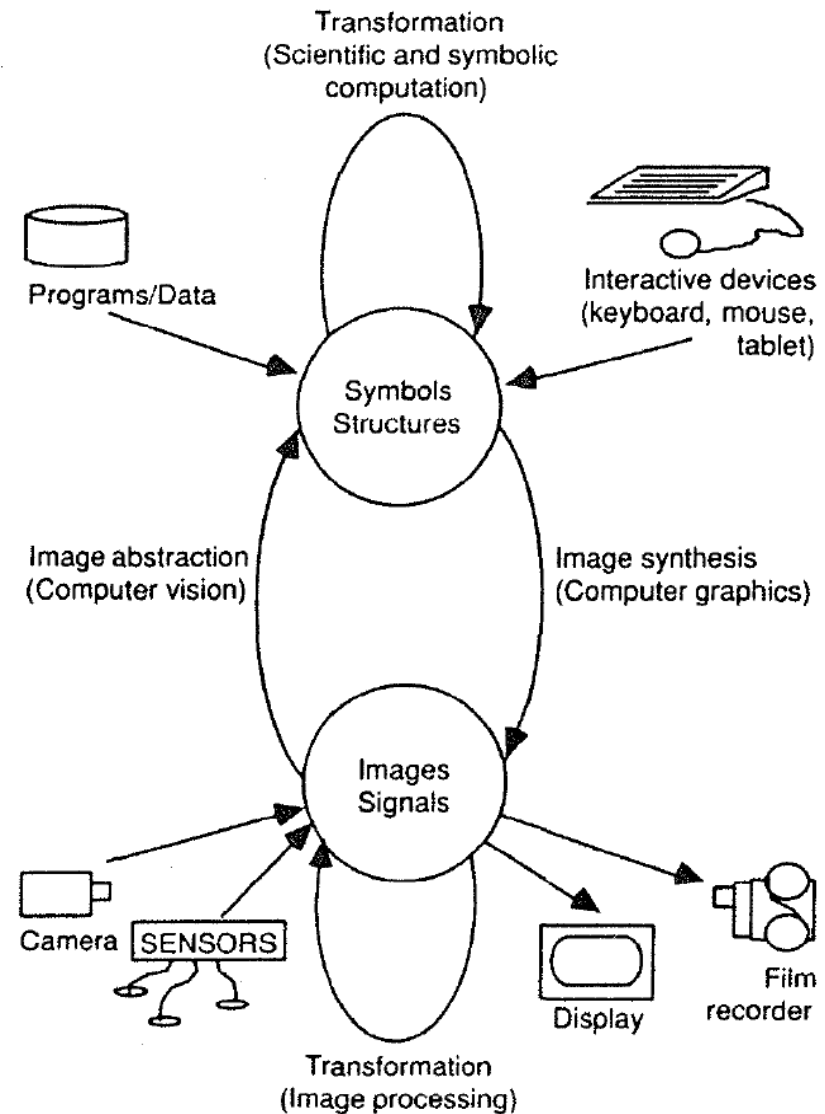
(c) Dims *a* & *c*

- Bellman (1957): The more dimensions, the more sparse the space becomes, and distance measures are less meaningful



05 Conclusion and Future Outlook

- ... the common denominator of Computational sciences
- ... the transformation of the symbolic into the geometric
- ... the support of human perception
- ... facilitating knowledge discovery in data



McCormick, B. (1987) Scientific and Engineering Research Opportunities. *Computer graphics*, 21, 6.

- Time (e.g. entropy) and Space (e.g. topology)
- Knowledge Discovery from “unstructured” ;-)
(Forrester: >80%) data and applications of
structured components as methods to index and
organize data -> Content Analytics
- Open data, Big data, sometimes: small data
- Integration in “real-world” (e.g. Hospital), mobile
- How can we measure the benefits of visual
analysis as compared to traditional methods?
- Can (and how can) we develop powerful visual
analytics tools **for the non-expert end user?**



Thank you!

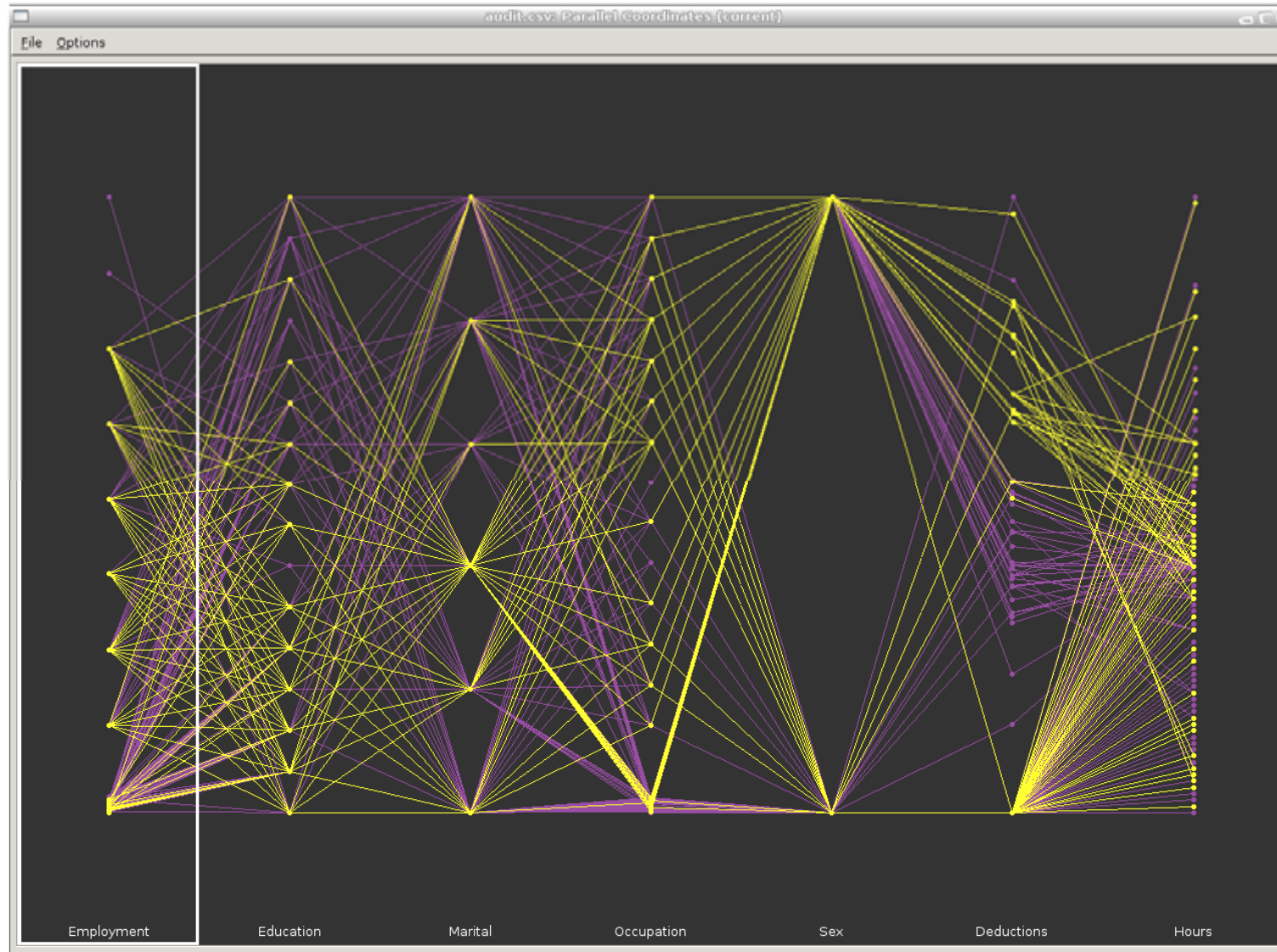
Questions

- What is semiotic engineering?
- Please explain the process of intelligent interactive information visualization!
- What is the difference between visualization and visual analytics?
- Explain the model of perceptual visual processing according to Ware (2004)!
- What was the historical start of systematic visual analytics? Why is this an important example?
- Please describe very shortly 6 of the most important visualization techniques!
- Transform five given data points into parallel coordinates!
- How can you ensure data protection in using parallel coordinates?
- What is the basic idea of RadViz?
- For which problem would you use a star-plot visualization?

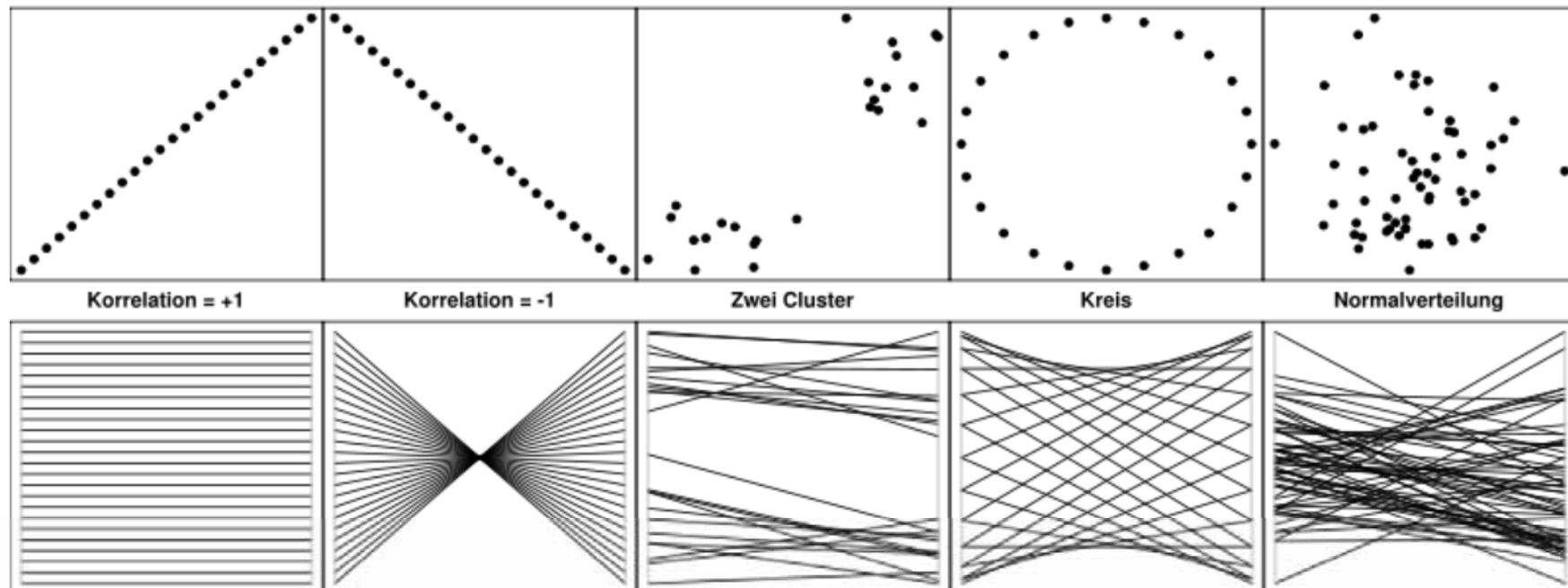
- What are the basic design principles of interactive intelligent visualization?
- What is the visual information seeking mantra of Shneiderman (1996)?
- Which concepts are important to let the end user interactively manipulate the data?
- What is the problem involved in looking at neonatal polysomnographic recordings?
- Why is time very important in medical informatics?
- What was the goal of LifeLines by Plaisant et al (1996)?
- Which temporal analysis tasks can you determine?
- Why is pattern discovery in medical informatics so important?
- What is the aim of foreseeing the future behaviour of medical data?

Appendix

- <http://vis.lbl.gov/Events/SC07/Drosophila/>
(some really cool examples of high-dimensional data)
- <http://people.cs.uchicago.edu/~wiseman/chernoff> (Chernoff Faces in Java)
- <http://lib.stat.emu.edu> (Iris sample data set)
- <http://graphics.stanford.edu/data/voldata> (113-slice MRI data set of CT studies of cadaver heads)



<http://datamining.togaware.com>



Zur Visualisierung von hochdimensionalen Daten in der Statistik müssen drei wichtige Aspekte beachtet werden:

die Anordnung der Achsen

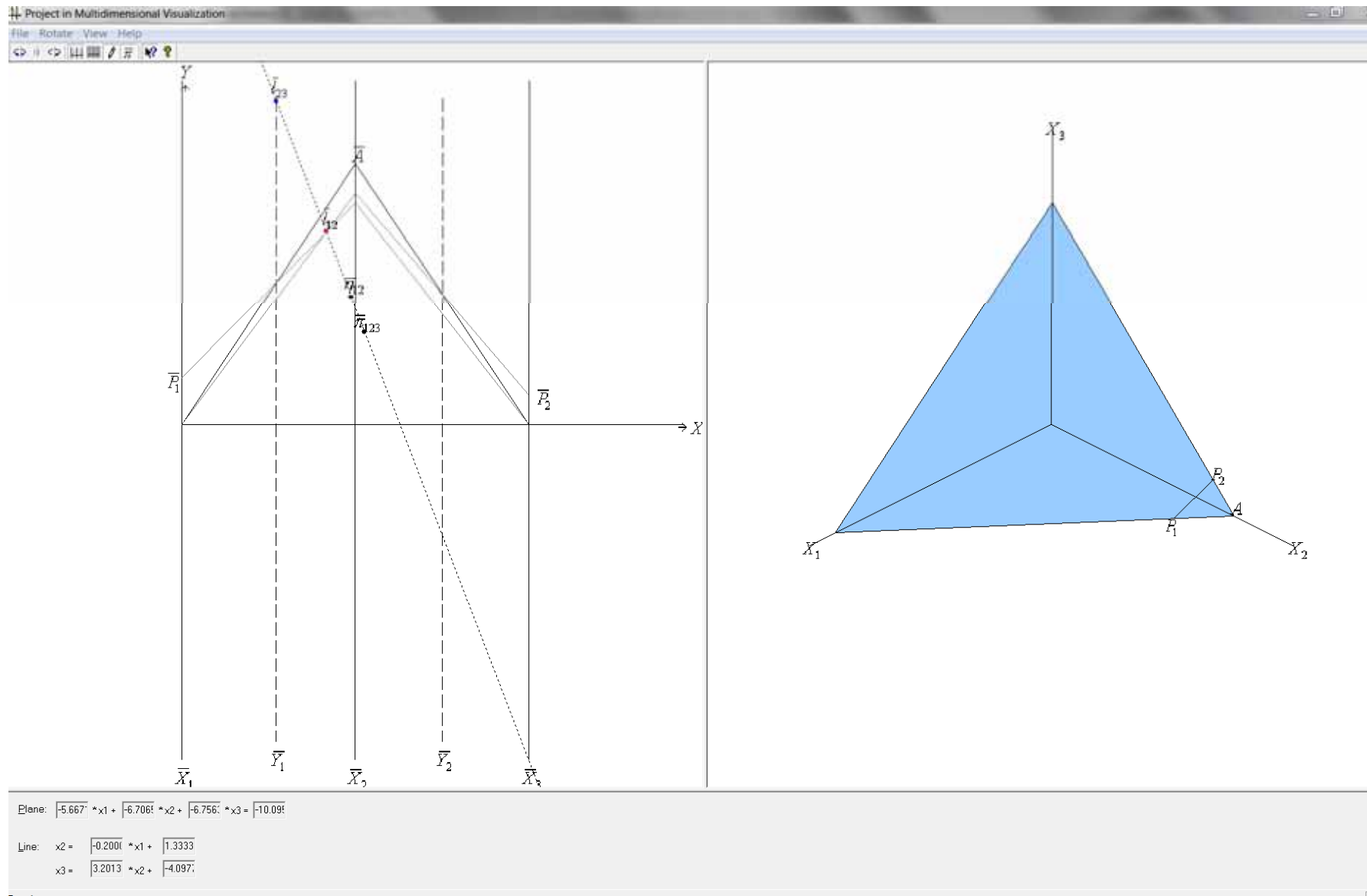
Die Anordnung der Achsen ist entscheidend für die Suche nach Strukturen in den Daten. In einer typischen Datenanalyse werden meist viele Anordnungen ausprobiert. Es wurden Anordnungsheuristiken entwickelt, die Einblicke in interessante Strukturen erlauben.^[7]

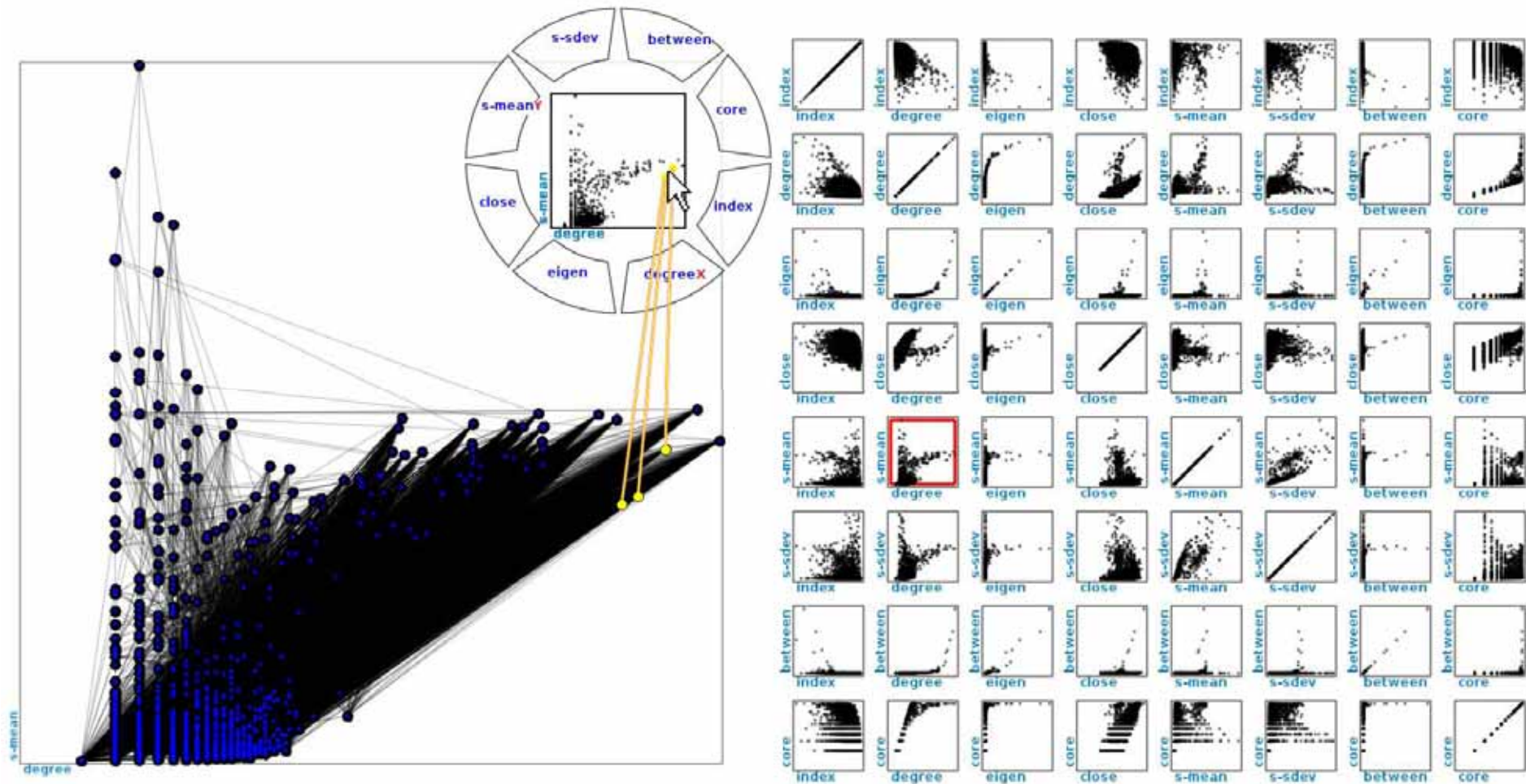
die Rotation der Achsen (Daten)

Da die i -te Koordinate durch die Ecke auf der i -ten Achse bestimmt wird, kann eine Rotation der Achsen (= Rotation der Daten) ein anderes Bild ergeben. Die beiden linken Grafiken können als Rotation der Achsen (oder Daten) um 90 Grad aufgefasst werden. Trotz gleicher Struktur ergeben sich unterschiedliche Strukturen in den parallelen Koordinaten.

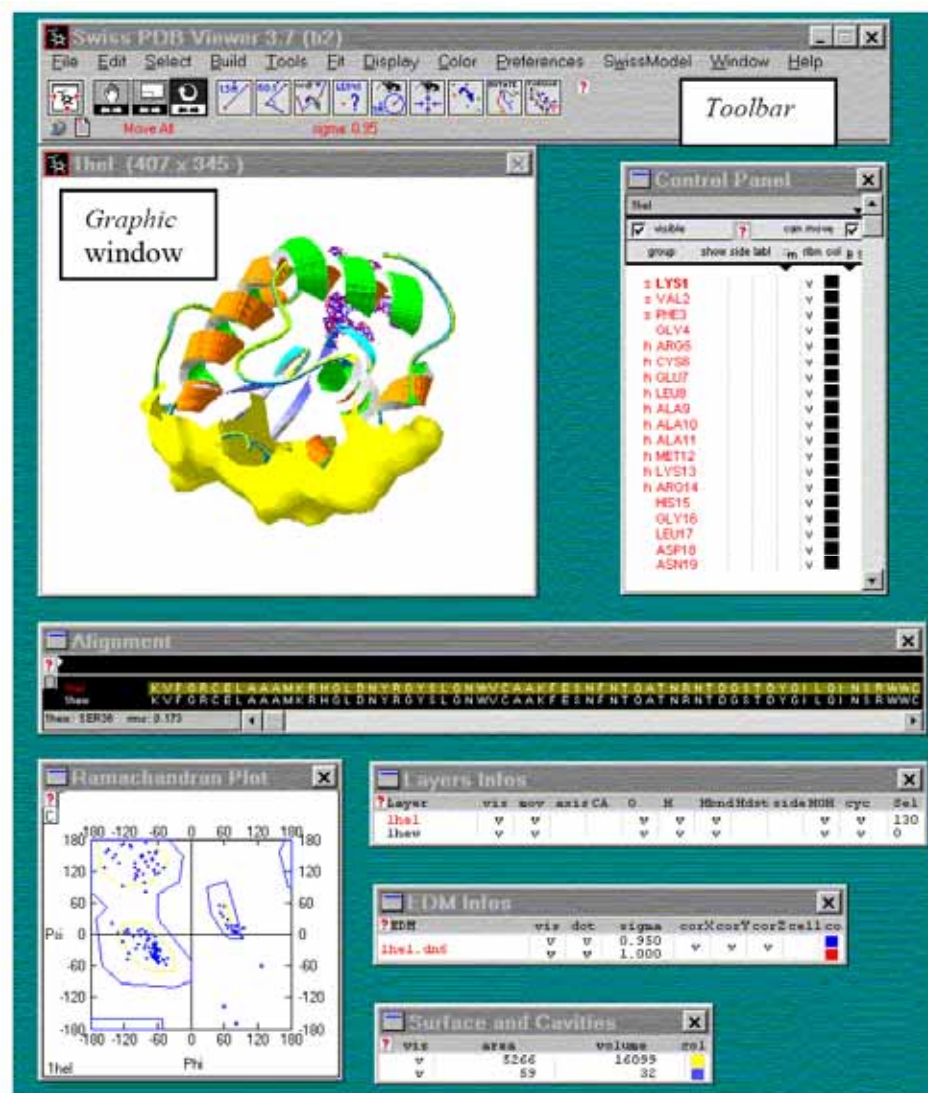
die Skalierung der Achsen

Die parallelen Koordinaten sind im Wesentlichen eine Aneinanderreihung von Linien zwischen Paaren von Koordinatenachsen.^[5] Daher sollten die Variablen auf einen ähnlichen Maßstab skaliert sein. Verschiedene Skalierungen können ebenfalls interessante Einsichten in die Daten geben.





Viau, C., McGuffin, M. J., Chiricota, Y. & Jurisica, I. (2010) The FlowVizMenu and Parallel Scatterplot Matrix: Hybrid Multidimensional Visualizations for Network Exploration. *Visualization and Computer Graphics, IEEE Transactions on*, 16, 6, **1100-1108**.

Main
windowsSpecific
windows<http://www.expasy.org>

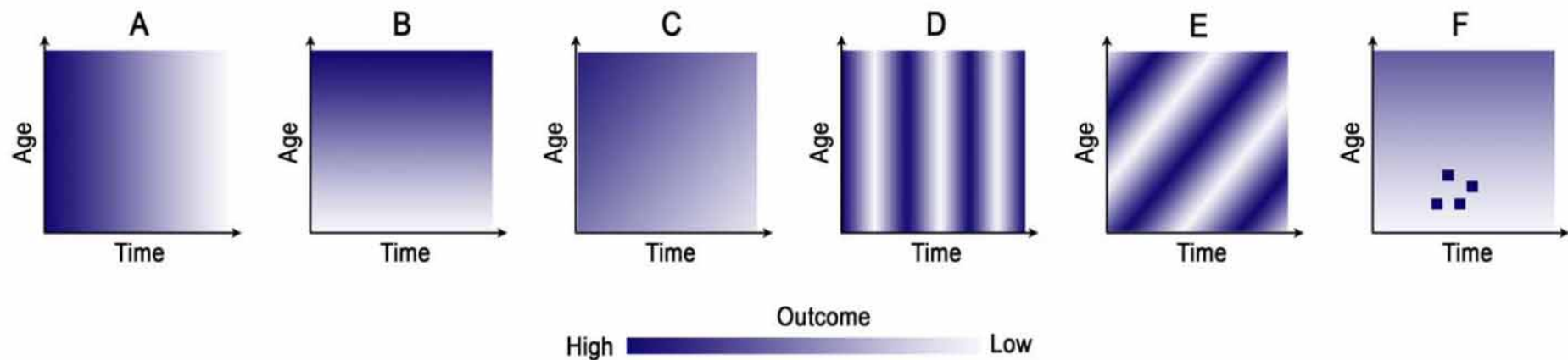
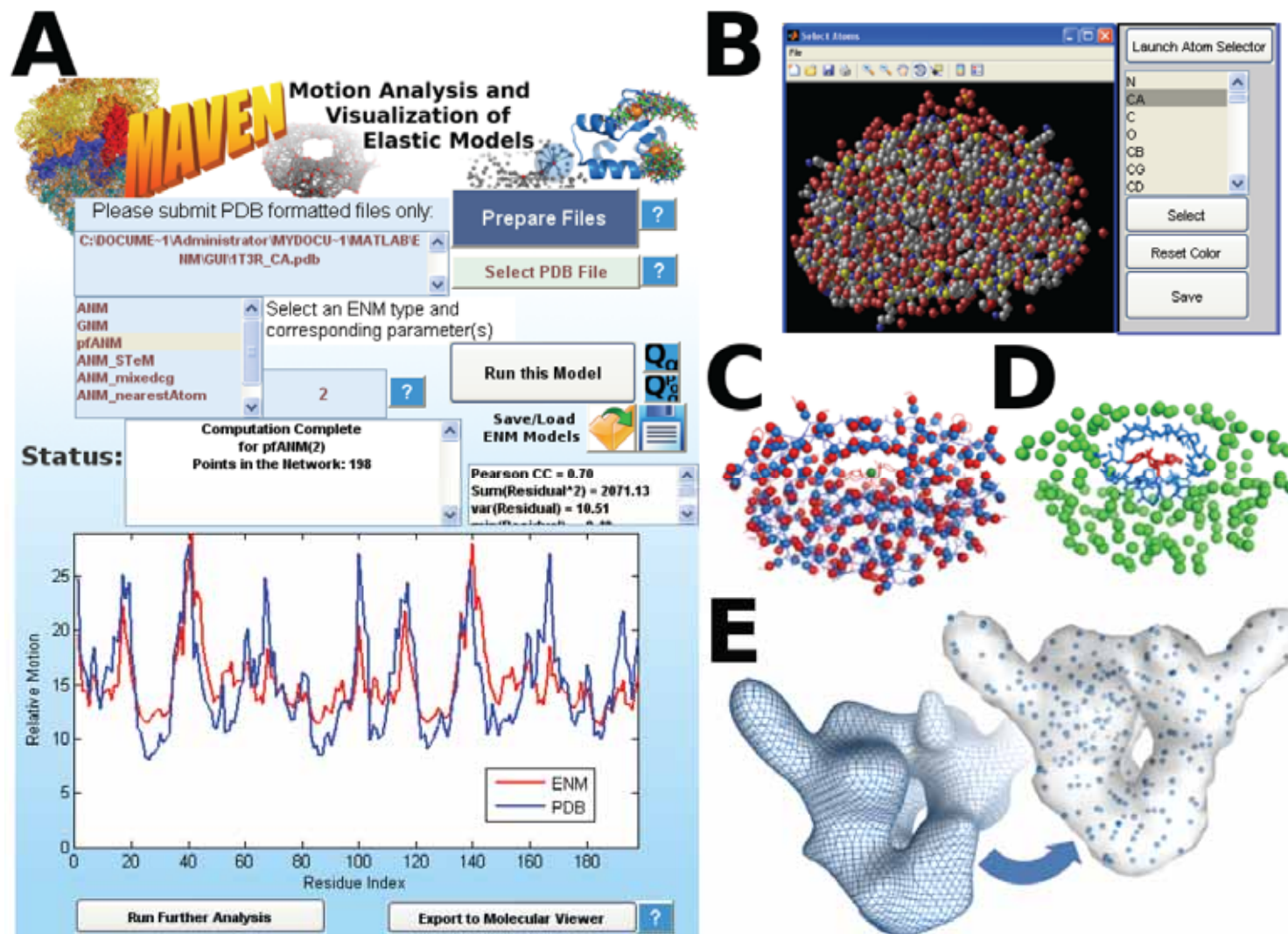
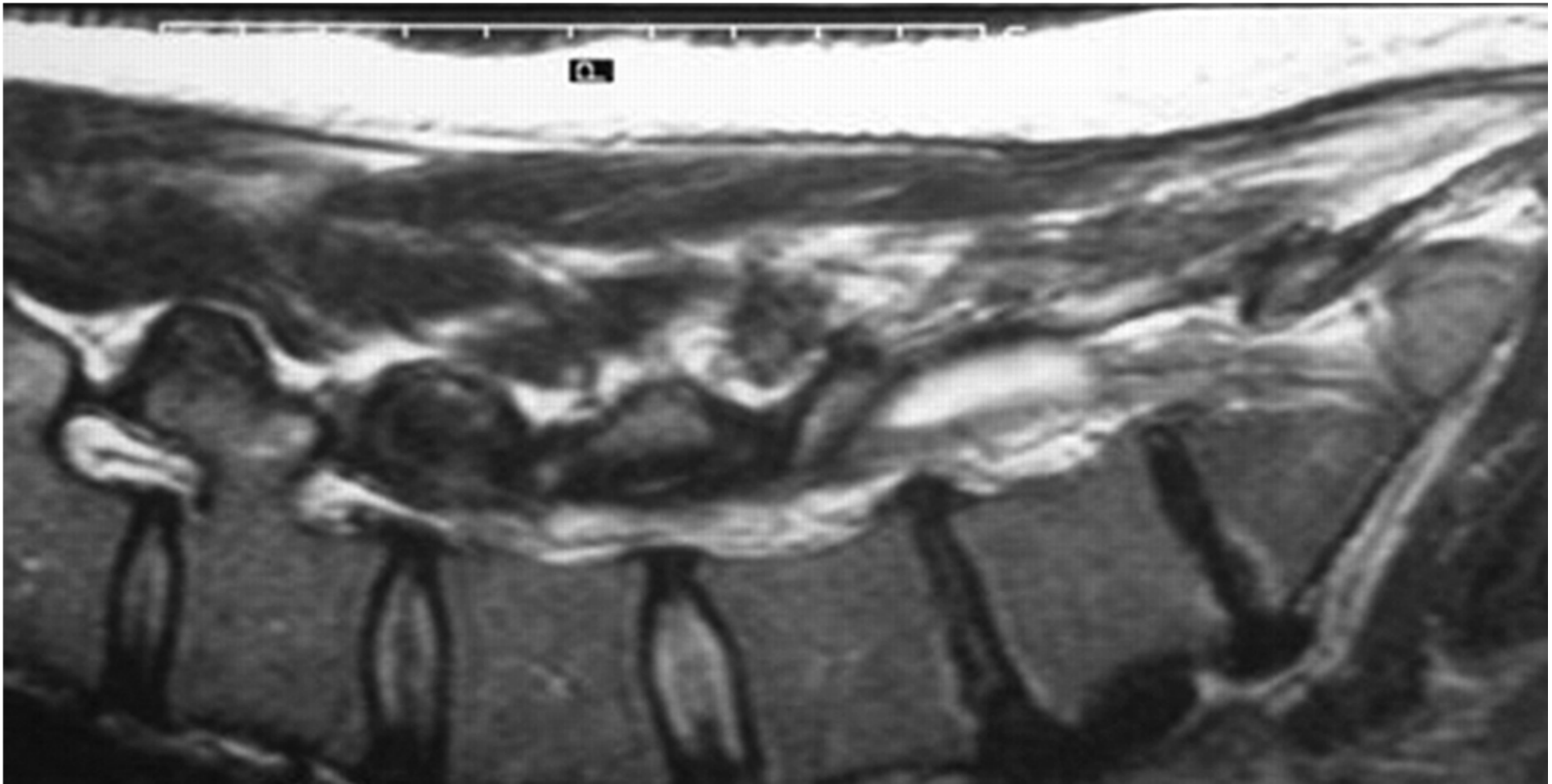


Figure 1. Typical patterns observed in image plots used to study the association between age, time, and the disease of interest.
doi:10.1371/journal.pone.0014683.g001

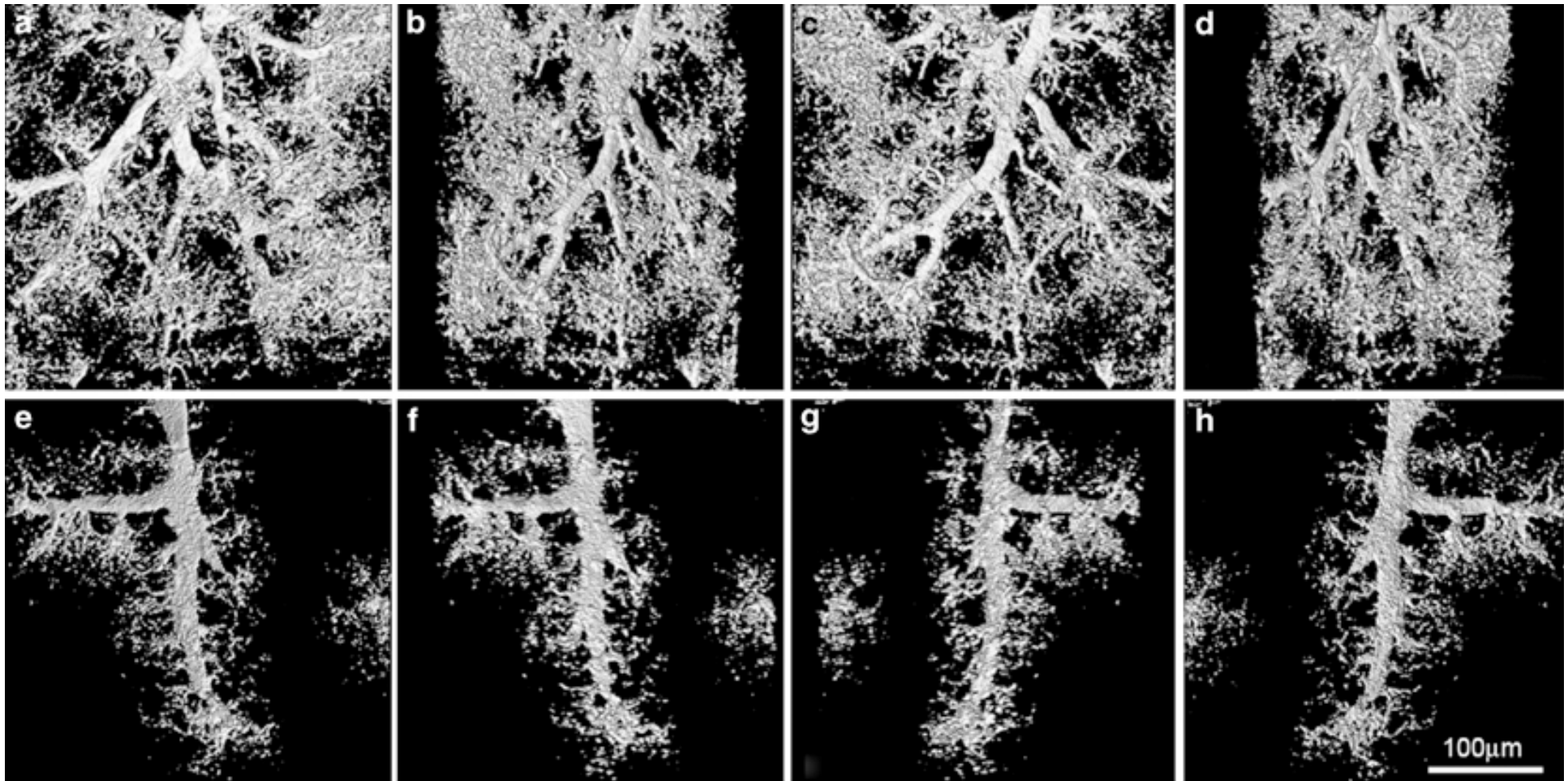
Chui, K. K. H., Wenger, J. B., Cohen, S. A. & Naumova, E. N. (2011) Visual Analytics for Epidemiologists: Understanding the Interactions Between Age, Time, and Disease with Multi-Panel Graphs. *Plos One*, 6, 2.

Zimmermann, M., Kloczkowski, A. & Jernigan, R. (2011) MAVENS: Motion analysis and visualization of elastic networks and structural ensembles. *Bmc Bioinformatics*, 12, 1, 264.

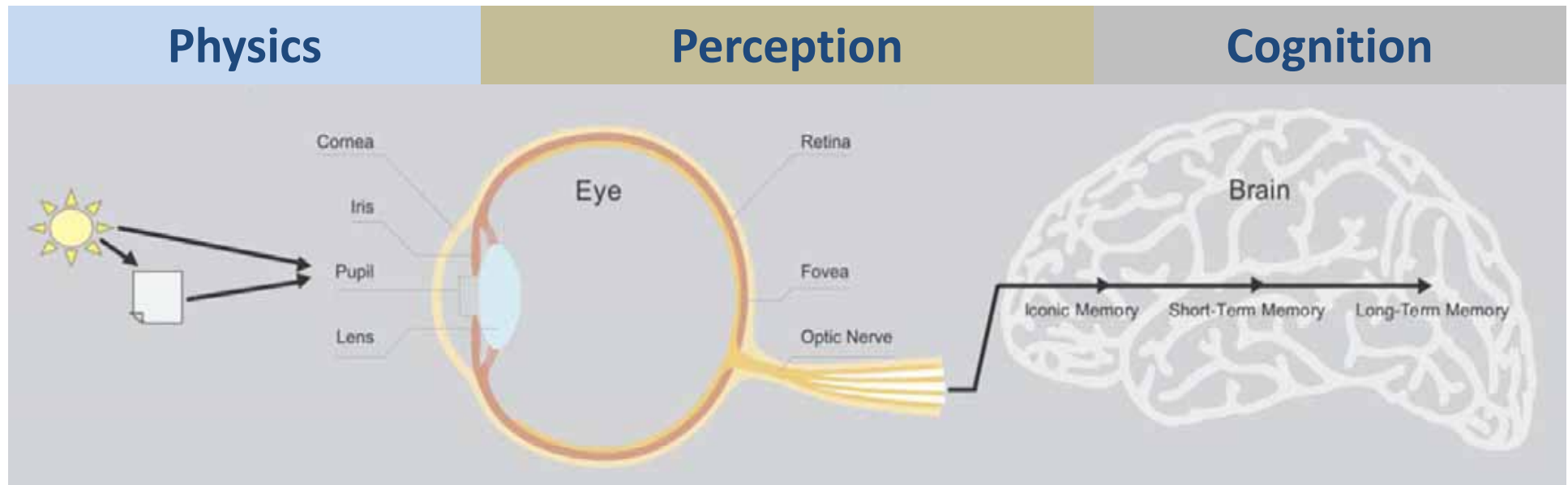




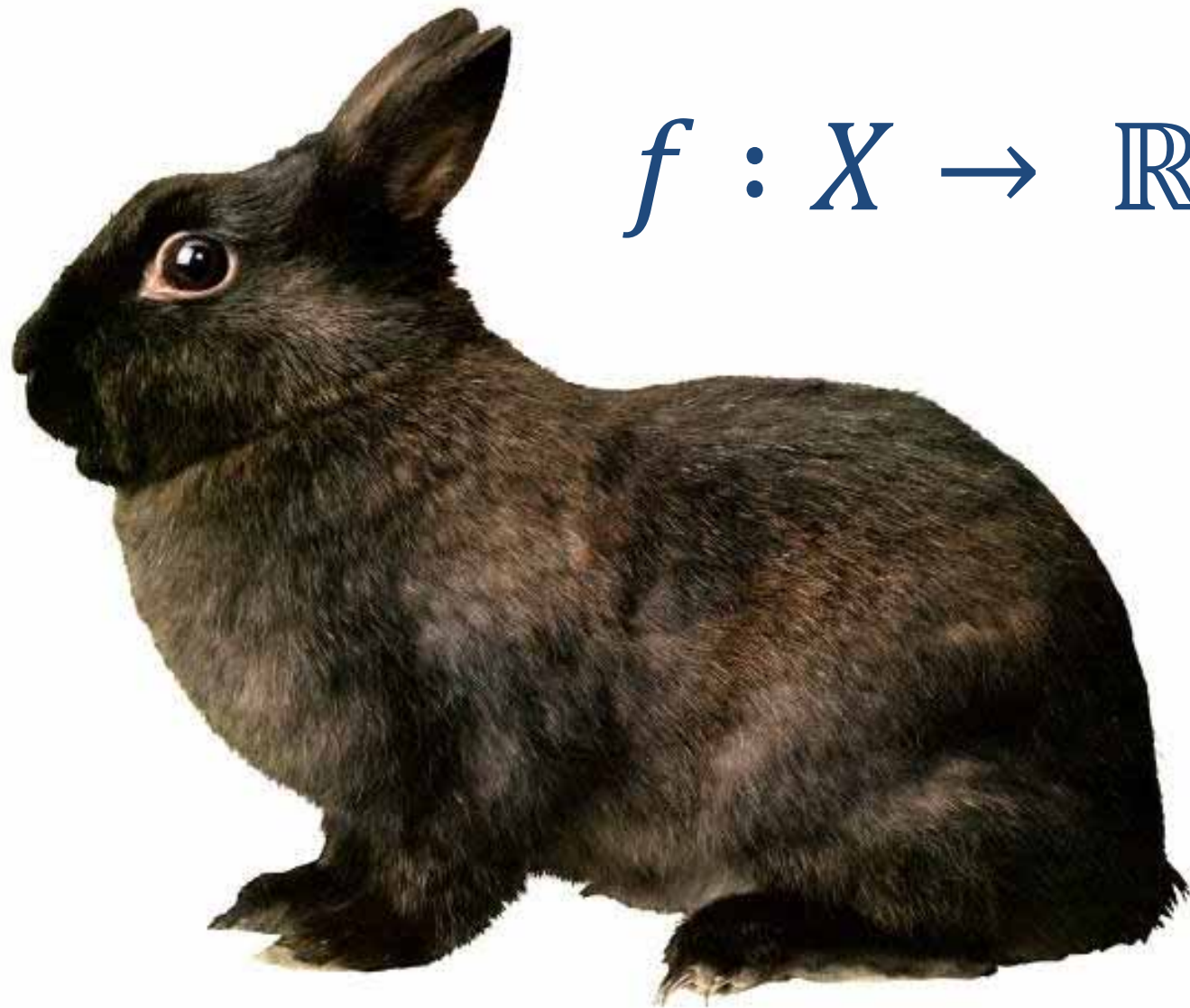
Erginousakis, D. et al. 2011. Comparative Prospective Randomized Study Comparing Conservative Treatment and Percutaneous Disk Decompression for Treatment of Intervertebral Disk Herniation. *Radiology*, 260, (2), 487-493.



Dutly, A. E., Kugathasan, L., Trogadis, J. E., Keshavjee, S. H., Stewart, D. J. & Courtman, D. W. 2006. Fluorescent microangiography (FMA): an improved tool to visualize the pulmonary microvasculature. *Lab Invest*, 86, (4), 409-416.



Few, S. (2006) *Information Dashboard Design*. Sebastopol (CA), O'Reilly.



$$f : X \rightarrow \mathbb{R}$$

Each multivariate observation can be seen as a data point in an n -dimensional vector space

$$x_i = [x_{i1}, \dots, x_{in}]$$

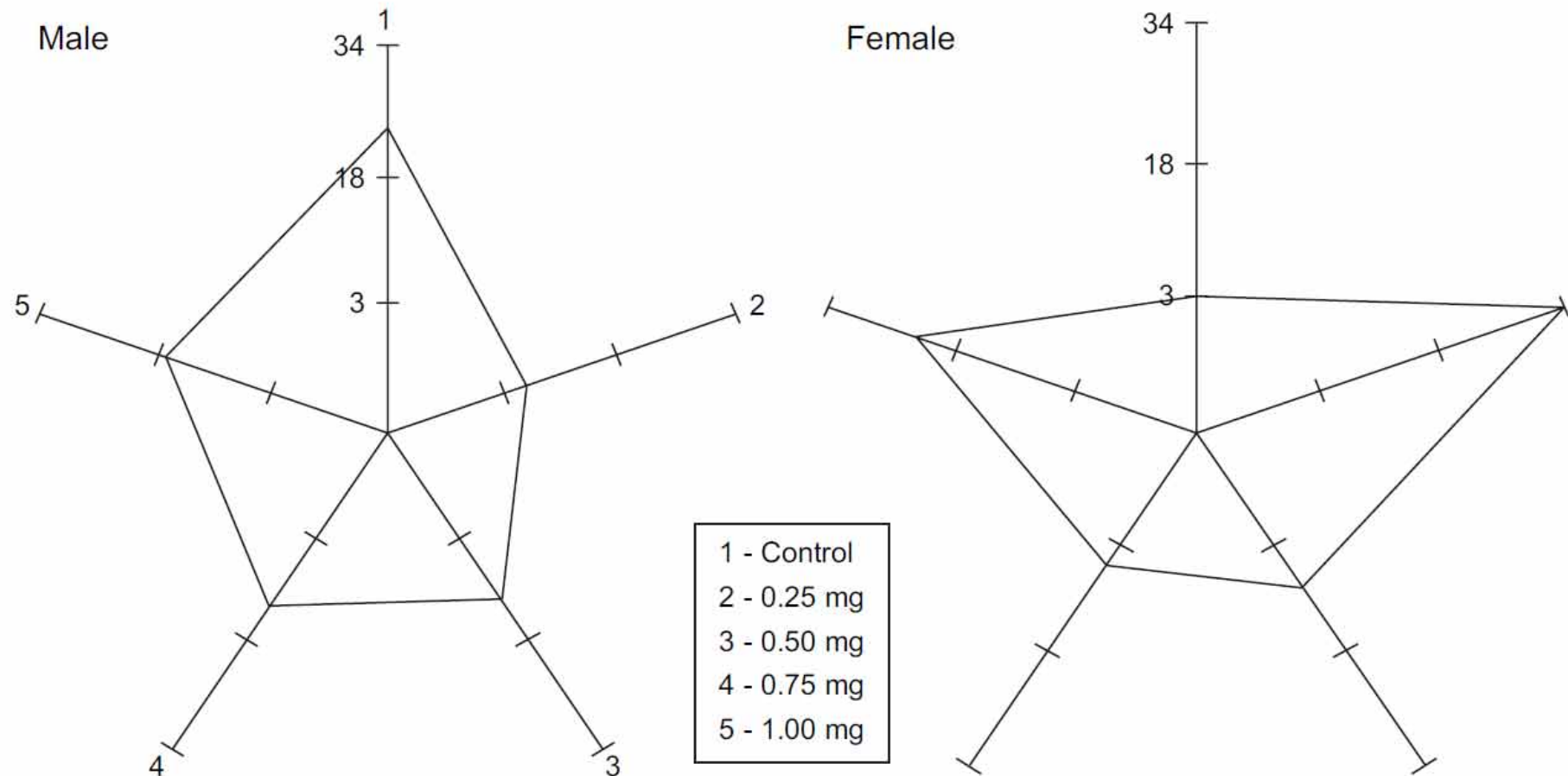
- “Look at your data”
- transfer data into information
- By use of human intelligence ...
- to transfer information into knowledge ($\mathbb{C} \rightarrow \mathbb{P}$)
- Challenge: To reduce the dimensionality of the data ...
- ... it is an information retrieval task!

Remember: The quality can be measured by two measures:

- Recall
- Precision



Holzinger, A., Hoeller, M., Bloice, M. & Urlesberger, B. (2008). *Typical Problems with developing mobile applications for health care: Some lessons learned from developing user-centered mobile applications in a hospital environment. International Conference on E-Business (ICE-B 2008), Porto (PT), IEEE, 235-240.*



Saary, M. J. (2008) Radar plots: a useful way for presenting multivariate health care data. *Journal Of Clinical Epidemiology*, 61, 4, 311-317.

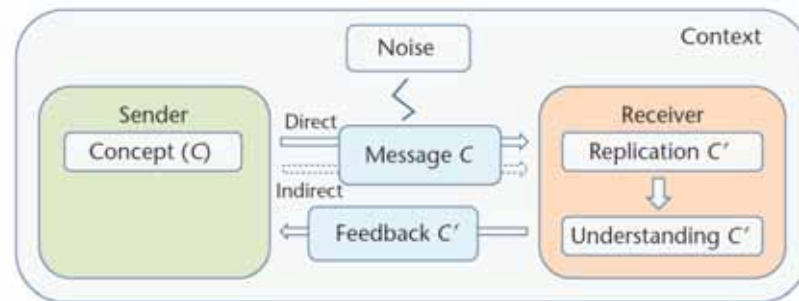
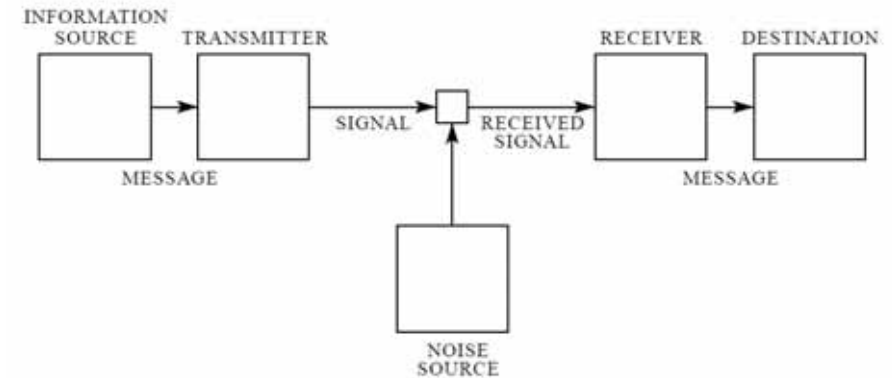


Figure 2. The interpersonal-communication protocol. A sender would like a receiver to comprehend message C, conveyed either straightforwardly or via indirect or subconscious mechanisms. However, noise in the communication channel or the receiver's failure to fully comprehend the message's intended meaning can undermine the sender's objective. An iterative clarification process eventually leads to a mutual understanding of the message.



Shannon, C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423.

- Scatterplot-Select (xDim, yDim, xMin, xMax, yMin, yMax)
- 1 $s \leftarrow \emptyset$ ▷ Initialize the set of records
- 2 **for** each record i ▷ For each record,
- 3 **do** $x \leftarrow \text{NORMALIZE}(i, x\text{Dim})$ ▷ derive the location,
- 4 $y \leftarrow \text{NORMALIZE}(i, y\text{Dim})$
- 5 **if** $x\text{Min} < x < x\text{Max}$ and $y\text{Min} < y < y\text{Max}$
- 6 **do** $s \leftarrow s \cup I$ ▷ select points within rectangle
- 7 return s
- Point-in-Polygon(xs, ys, numPoints, x, y)
- 1 $j \leftarrow \text{numPoints} - 1$
- 2 oddNodes $\leftarrow$ false
- 3 for $i \leftarrow 0$ to numPoints -1
- 4 do if $ys[i] < y$ and $ys[j] \geq y$ or $ys[j] < y$ and $ys[i] \geq y$
- 5 do if $xs[i] + (y - ys[i]) / (ys[j] - ys[i]) * (xs[j] - xs[i]) < x$
- 6 do oddNodes $\leftarrow$ not oddNodes
- 7 $j \leftarrow i$
- 8 return oddNodes

Ward, M., Grinstein, G. & Keim, D. (2010) *Interactive Data Visualization: Foundations, Techniques and Applications*. Natick (MA), Peters.

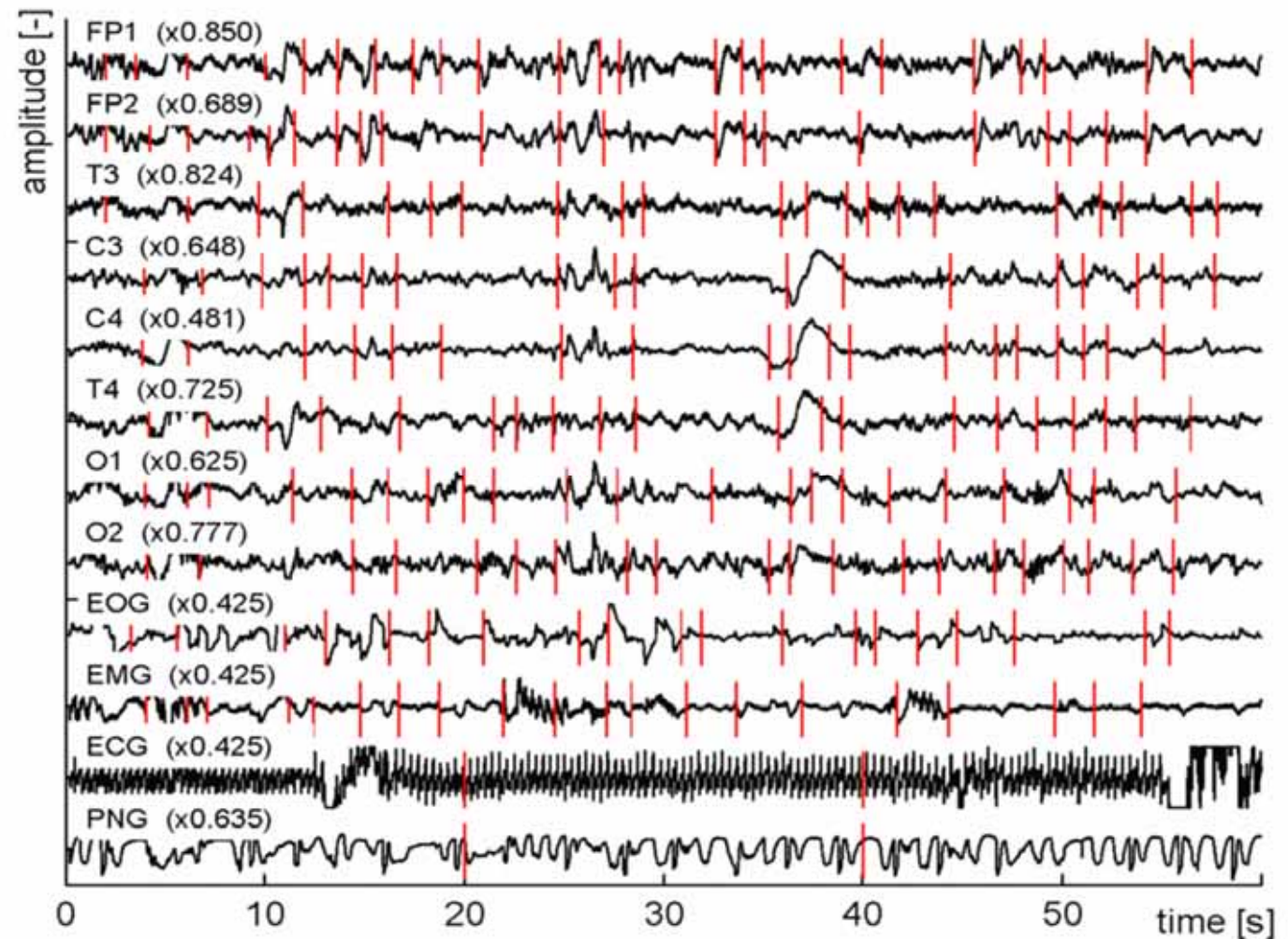
EEG signal from 8 ref.
derivations: FP1, FP2,
T3, T4, C3, C4, O1, O2

EOG =
Electrooculogram

EMG =
Electromyogram

PNG =
Pneumogram

ECG =
Electrocardiogram



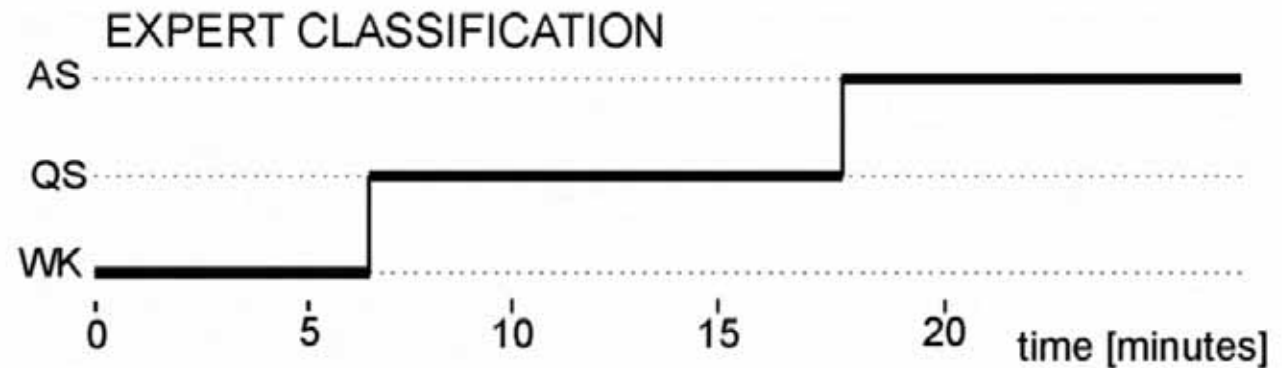
Gerla, V., Djordjevic, V., Lhotska, L. & Krajca, V. (2009). *Visualization methods used for evaluation of neonatal polysomnographic data. ITAB 2009, Information Technology and Applications in Biomedicine, Cyprus, IEEE, 1-4.*

Expert classification :

AS - active sleep,

QS - quiet sleep,

WK - wakefulness)



Representation of
final clusters :
clustering into
9 groups, displayed
channels:

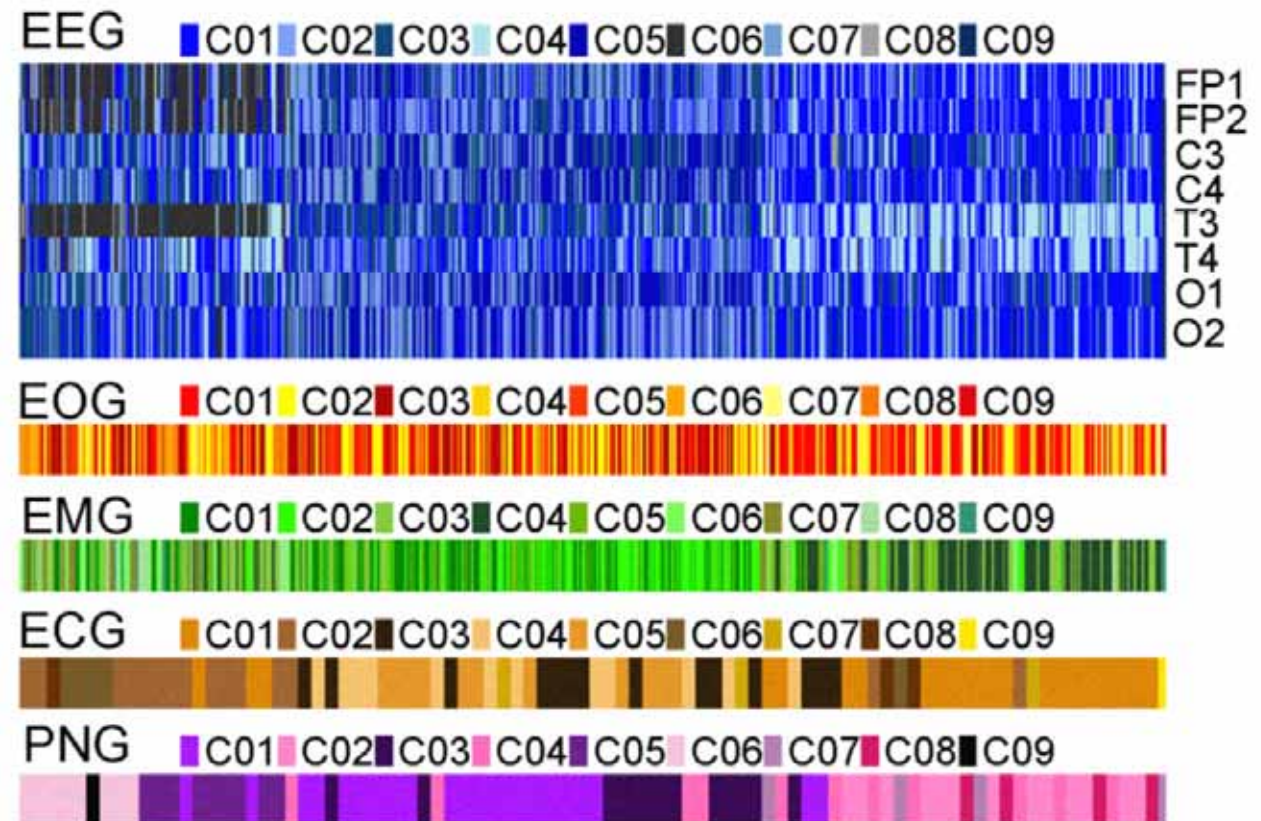
EEG ,

EOG,

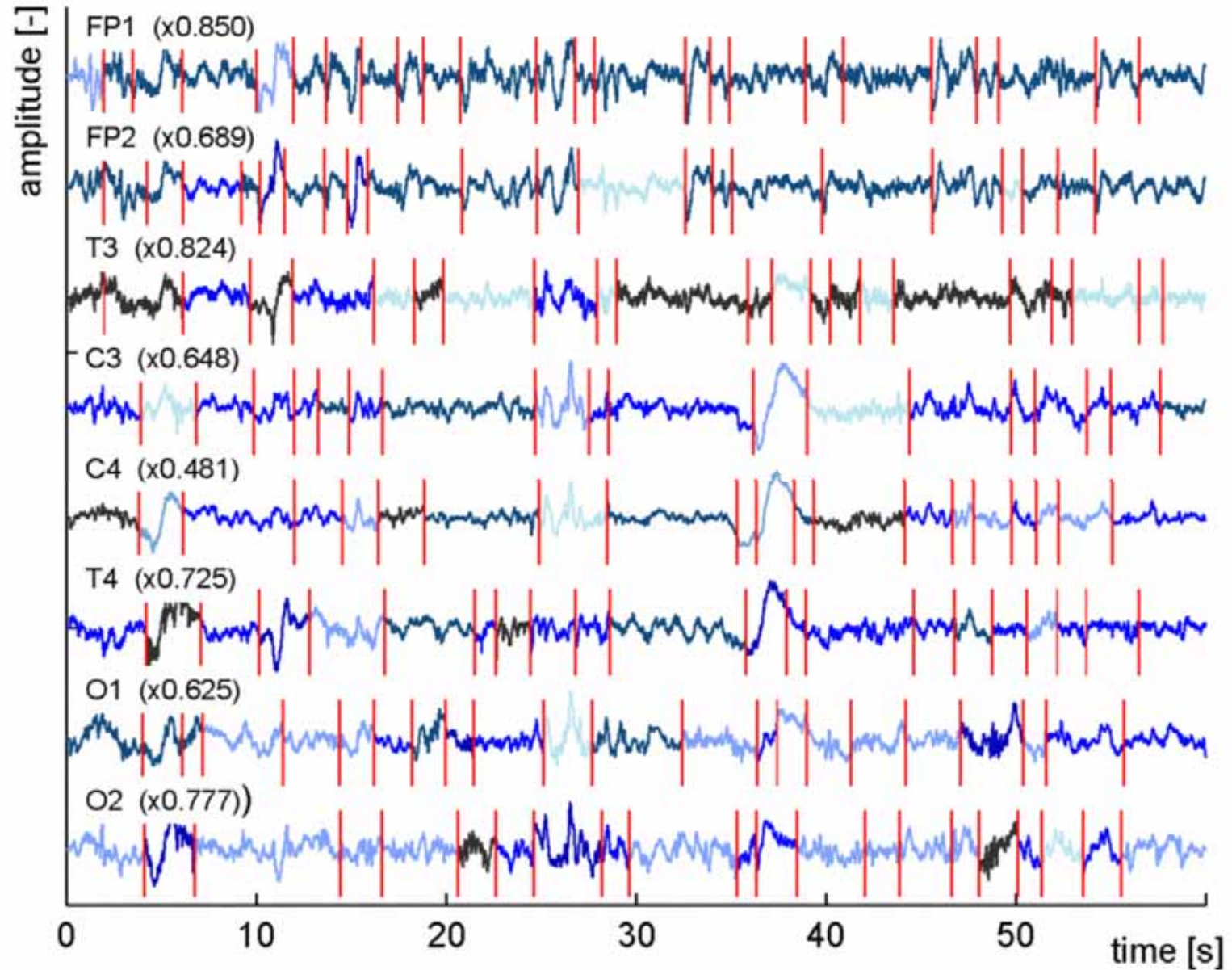
EMG,

ECG and

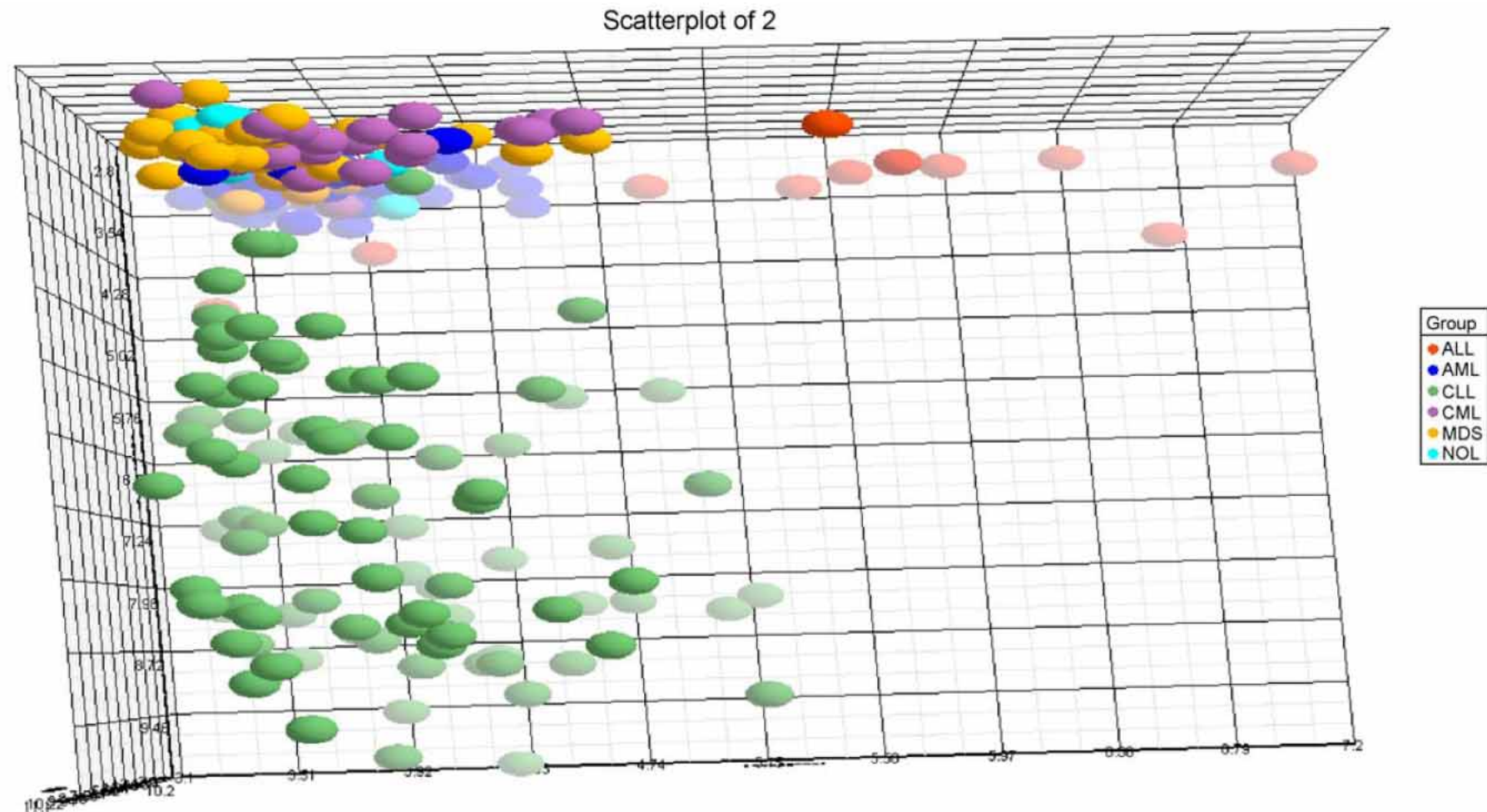
PNG



Gerla et al. (2009)



Classification CLL—ALL. Representation of the probes of the decision tree which classify the CLL and ALL to 1555158_at, 1553279_at and 1552334_at



Corchado et al. (2009)

- The model of Corchado et al. (2009) combines:
- 1) methods to **reduce the dimensionality** of the original data set;
- 2) pre-processing and data filtering techniques;
- 3) a clustering method to classify patients; and
- 4) extraction of knowledge techniques
- The system reflects how human experts work in a lab, but
- 1) **reduces the time** for making predictions;
- 2) **reduces the rate of human error**; and
- 3) **works with high-dimensional data** from exon arrays

- Have a look at
- <http://onesecond.designly.com>
- And then get a feeling on our limits of cognition:

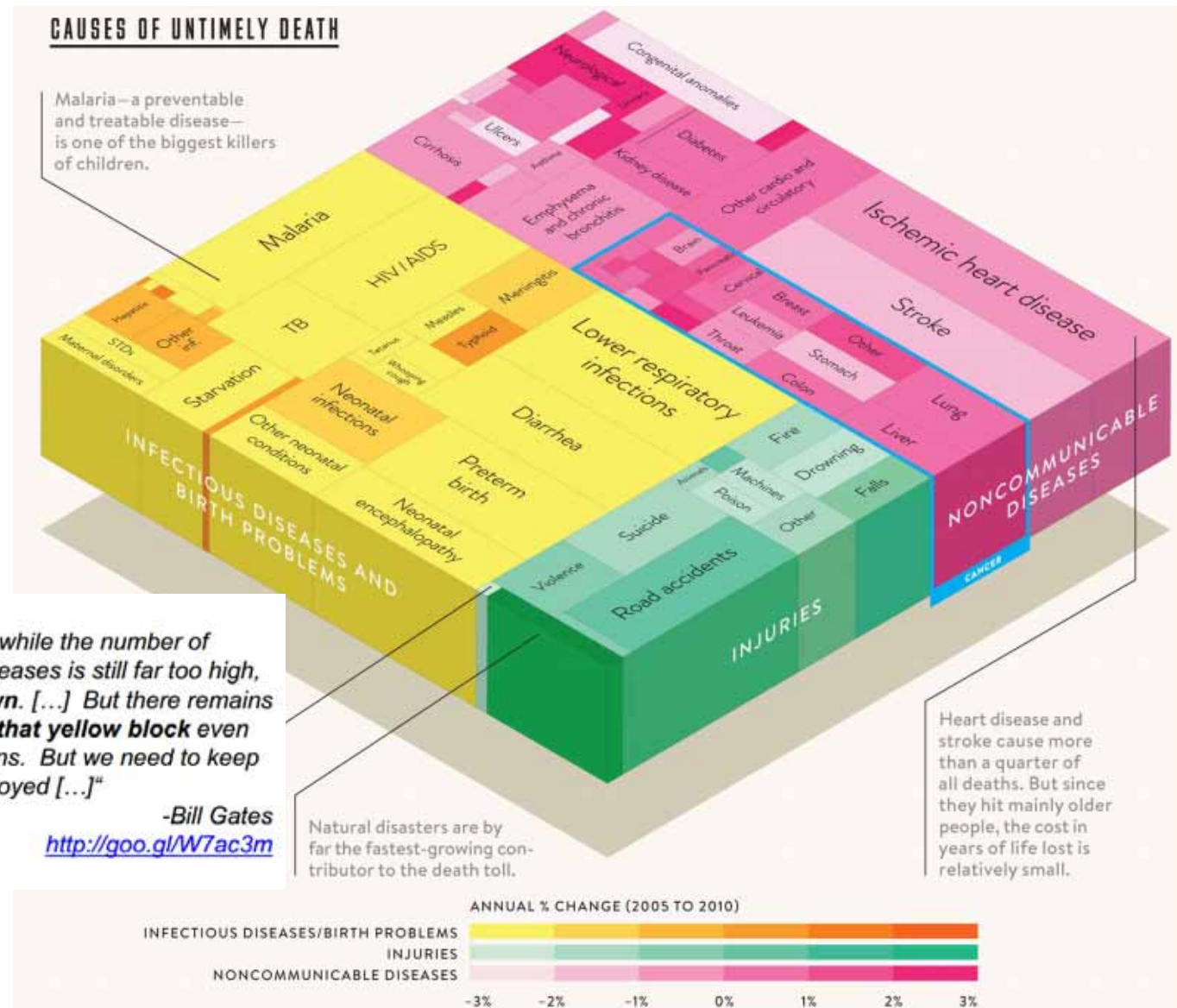
Simons, D. J. & Chabris, C. F. 1999.
Gorillas in our midst: sustained
inattention blindness for dynamic
events. Perception, 28, (9), 1059-
1074.



https://www.youtube.com/watch?v=IGQmdoK_ZfY

CAUSES OF UNTIMELY DEATH

Malaria—a preventable and treatable disease—is one of the biggest killers of children.



"I love this graph because it shows that while the number of people dying from communicable diseases is still far too high, those numbers continue to come down. [...] But there remains much to do to cut down the deaths in that yellow block even more dramatically. We have the solutions. But we need to keep the up support where they're being deployed [...]"

-Bill Gates

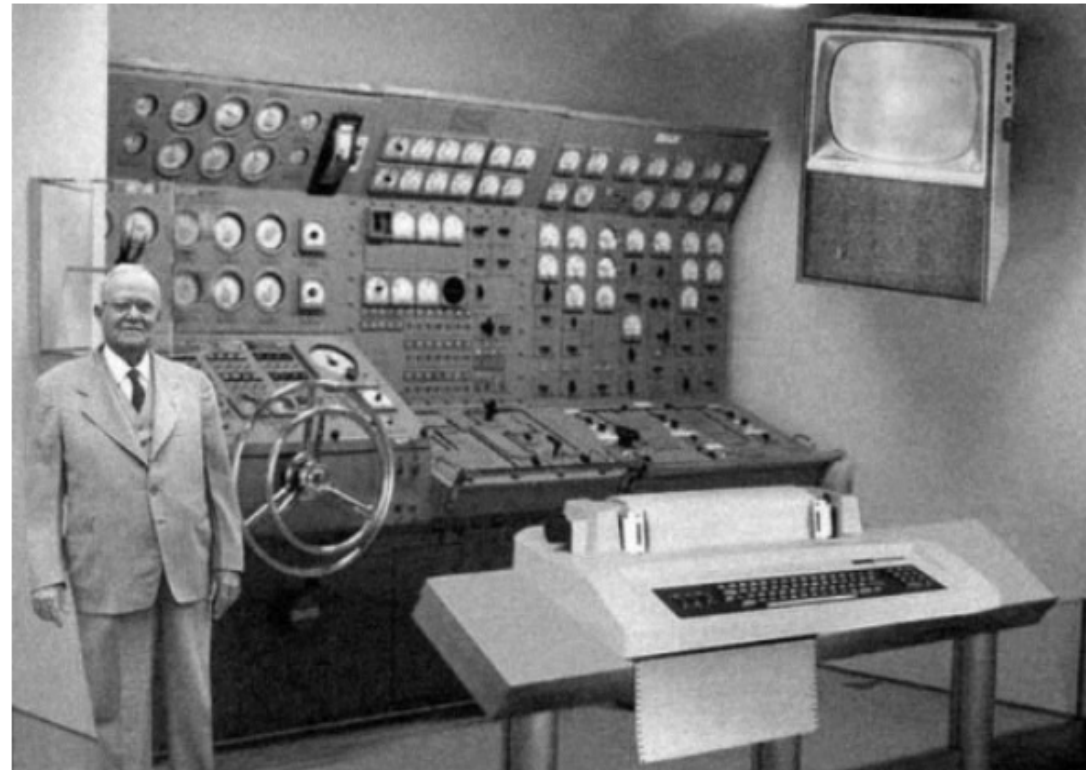
<http://goo.gl/W7ac3m>

Natural disasters are by far the fastest-growing contributor to the death toll.

Heart disease and stroke cause more than a quarter of all deaths. But since they hit mainly older people, the cost in years of life lost is relatively small.



Home Computer of 2004

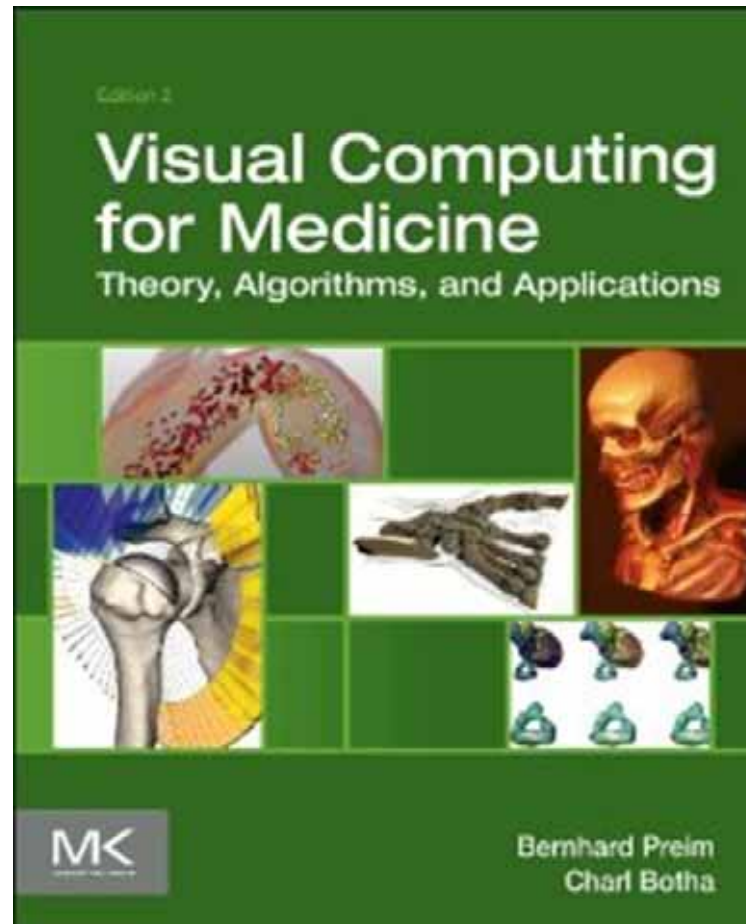


As predicted by the RAND Corporation in 1954: *

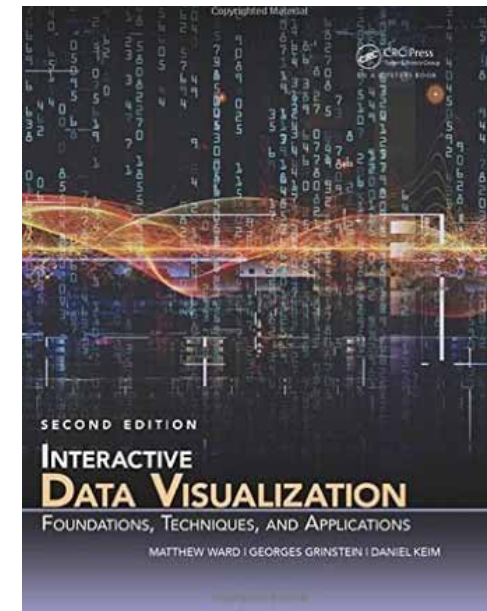
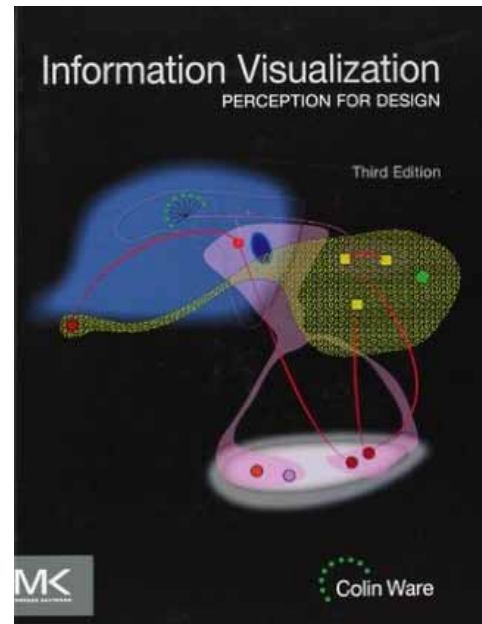
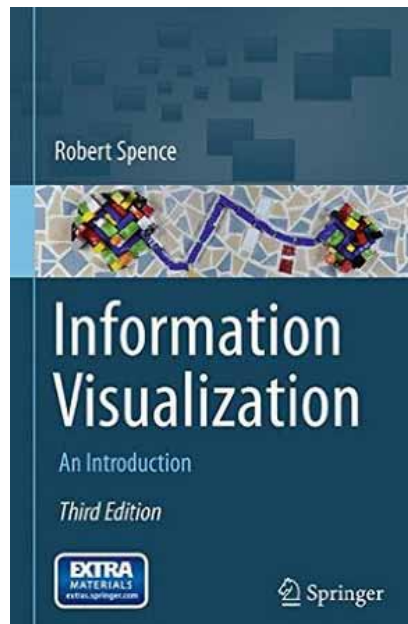
“With teletype interface and the FORTRAN language, it will be easy to use.”

* Actually, this claim is a **hoax**. But BASIC on the IBM PC in 1981 was almost as hard to learn as FORTRAN in 1954, and modern languages such as Java are even harder.

Additional Reading



Preim, B. & Botha, C. P. 2013. Visual Computing for Medicine: Theory, Algorithms, and Applications, Morgan Kaufmann



Parallel Coordinates

