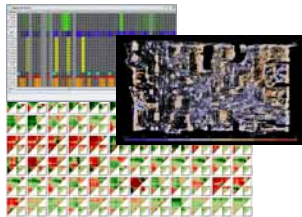


# Quality Metrics for Visual Analytics of High-Dimensional Data

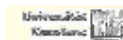


**Daniel A. Keim**  
Data Analysis and Information  
Visualization Group  
University of Konstanz, Germany



Workshop on Visual Analytics and Information Fusion at KDD 2011

August 21, 2011



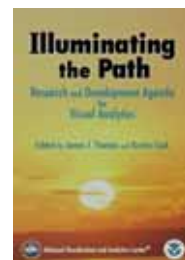
## Definition of Visual Analytics

### Definition:

**Visual Analytics is the science of analytical reasoning facilitated by visual interfaces.**

Visual analytics techniques are needed to

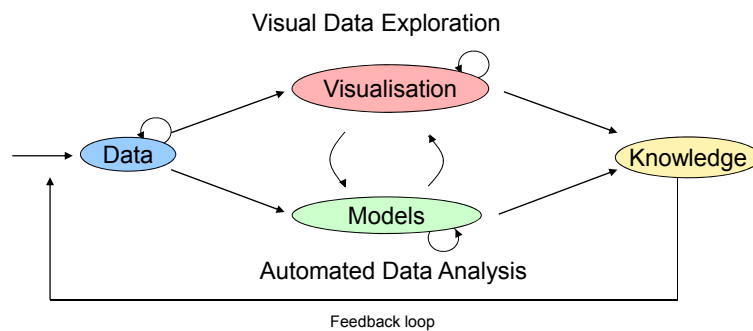
- integrate data from massive, dynamic, ambiguous, and often conflicting sources
- analyze the data to derive new insights
- make critical decisions in real-time.



## Definition of Visual Analytics

### Visual Analytics:

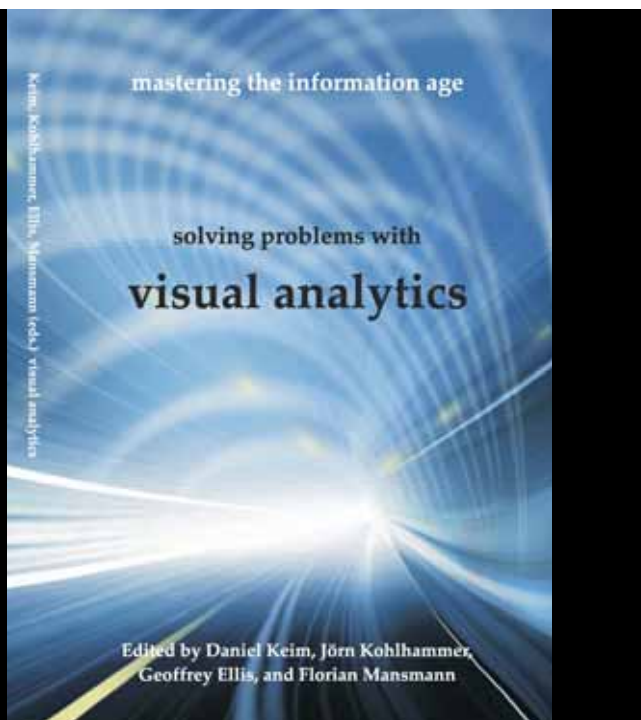
Tight Integration of Visual and Automatic Data Analysis Methods for Information Exploration and Scalable Decision Support



Universität  
Kassel

Roadmap from the  
VisMaster EU Project

[www.visual-analytics.eu](http://www.visual-analytics.eu)

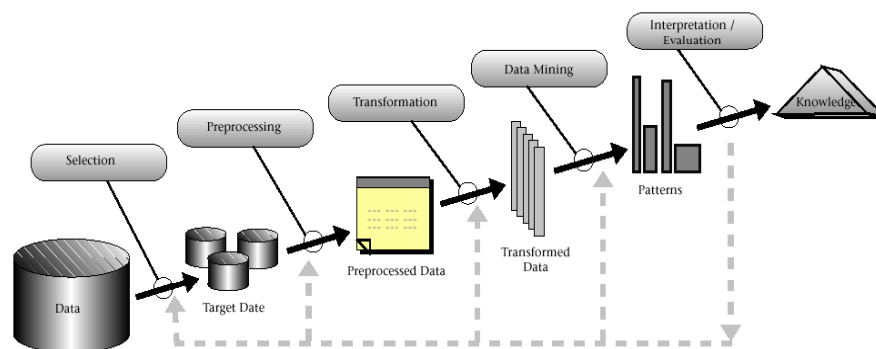


## Technical Challenges

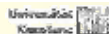
- **When should we use automated techniques versus interactive techniques?**
- **When do we need both?**
- **How to best combine interactive visualizations with automated analysis techniques?**



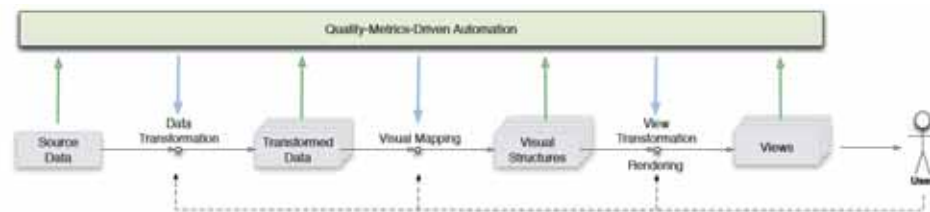
## The KDD Pipeline



Fayyad, U. M. et al. 1996. From data mining to knowledge discovery: an overview. In Fayyad, U. M. et al (Eds.), *Advances in knowledge discovery and data mining*. AAAI Press / The MIT Press.



# Quality-Metrics-Driven Automation

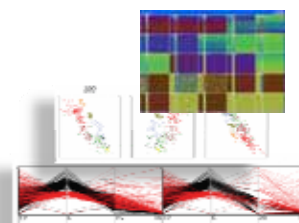


Enrico Bertini, Member, Andrada Tatu, Daniel Keim: Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization, IEEE Conf. on Information Visualization, 2011.



## Outline

- **Visual Analytics**
  - Definition
  - Challenges
  - Quality Metrics
- **Quality Metrics for**
  - Visual Mappings
  - Data Transformations
  - View Transformations
- **Perspectives**



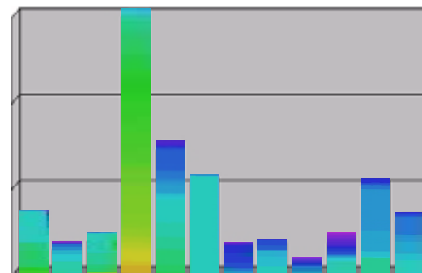
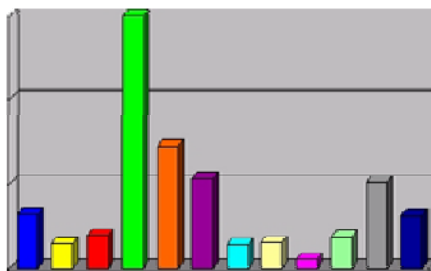
## Quality Metrics for Visual Mapping

### Automated Mapping Support

- for Pixel Bar Charts
- for Scatter Plots
- for Parallel Coordinates



## Pixel Bar Charts

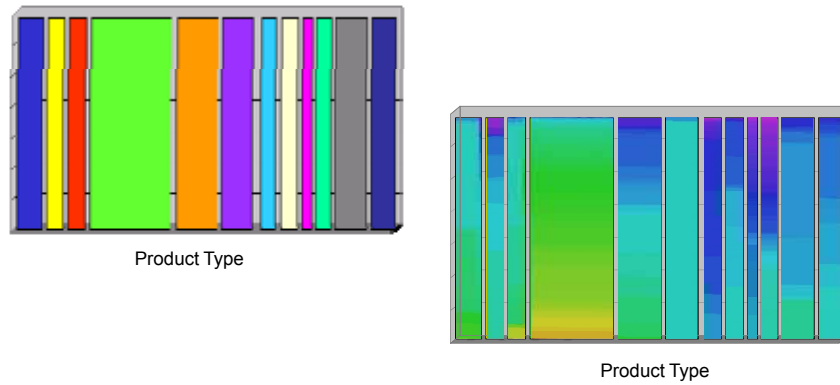


### Equal-Width Pixel Bar Chart



D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.

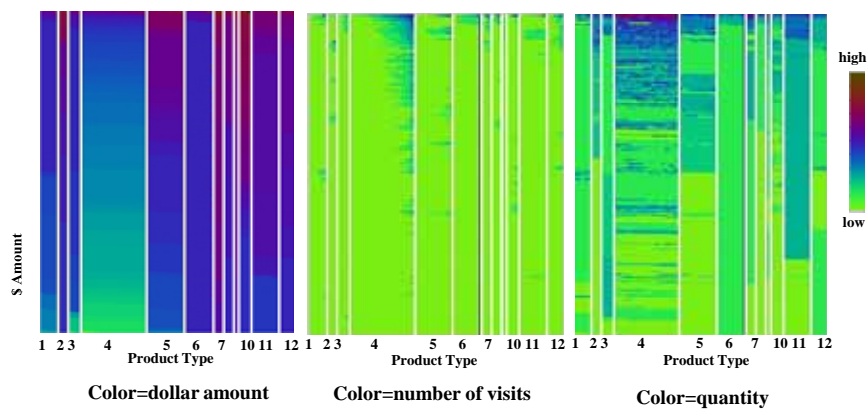
## Pixel Bar Charts



### Equal-Height Pixel Bar Chart

D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.

## Pixel Bar Charts



### Multi Pixel Bar Charts

D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.

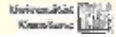
## Pixel Bar Charts

A **pixel bar chart** is defined by a five tuple

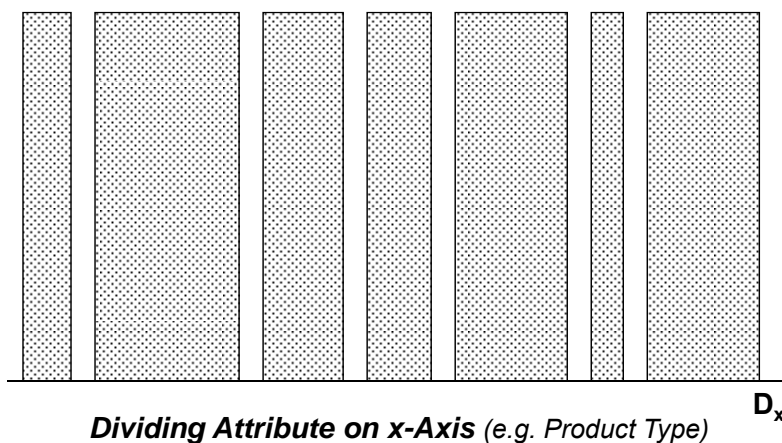
$$\langle D_x, D_y, O_x, O_y, C \rangle$$


where  $D_x, D_y, O_x, O_y, C \in \{A_1, \dots, A_k\}$  and

- $D_x / D_y$  are the *dividing attributes* on the x-/y-axes
- $O_x / O_y$  are the *ordering attributes* on the x-/y-axes
- $C$  defines the *coloring attribute(s)*

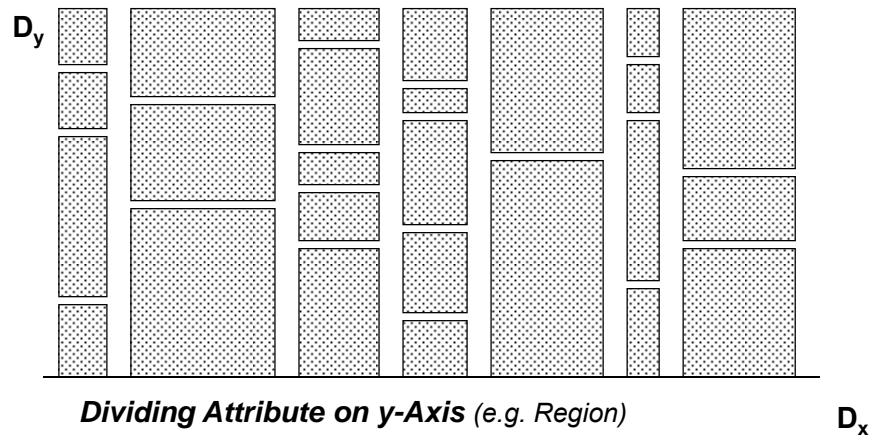
 D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.

## Pixel Bar Charts



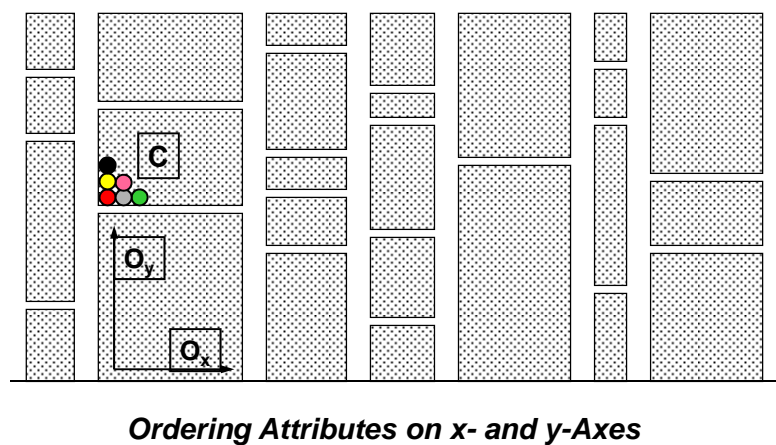
 D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.

## Pixel Bar Charts



D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.

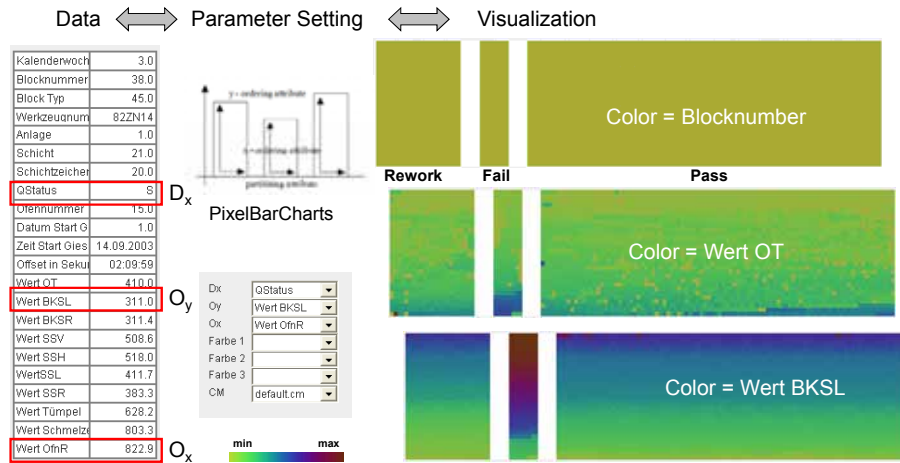
## Pixel Bar Charts



D. A. Keim, M. C. Hao, U. Dayal and M. Hsu. Pixel Bar Charts: A Visualization Technique for Very Large Multi-Attribute Data Sets. Visualization, San Diego 2001, extended version in: Information Visualization Journal, Palgrave, 1 (2), 2002.



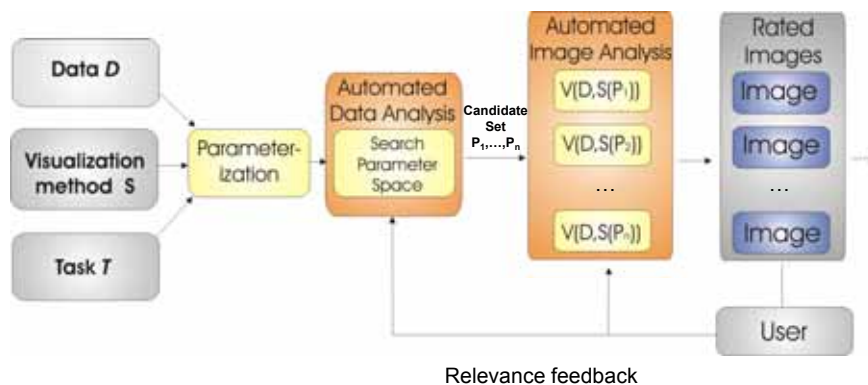
## Automated Mapping Support for Pixel Bar Charts



Parameter Space Very Large → Interactive Exploration impossible

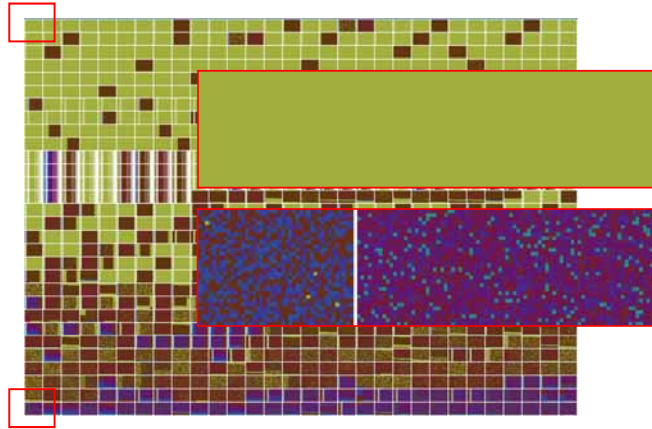


## Automated Mapping Support for Pixel Bar Charts



J. Schneidewind, M. Sips and D. A. Keim. Pixnostics: Towards Measuring the Value of Visualization. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '06), 2006.

## Automated Mapping Support for Pixel Bar Charts

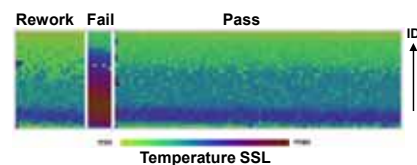
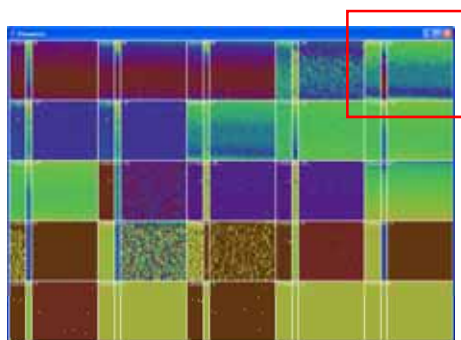


900 Pixel Bar Charts → only few provide insight



J. Schneidewind, M. Sips and D. A. Keim. Pixnostics: Towards Measuring the Value of Visualization. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '06), 2006.

## Automated Mapping Support for Pixel Bar Charts



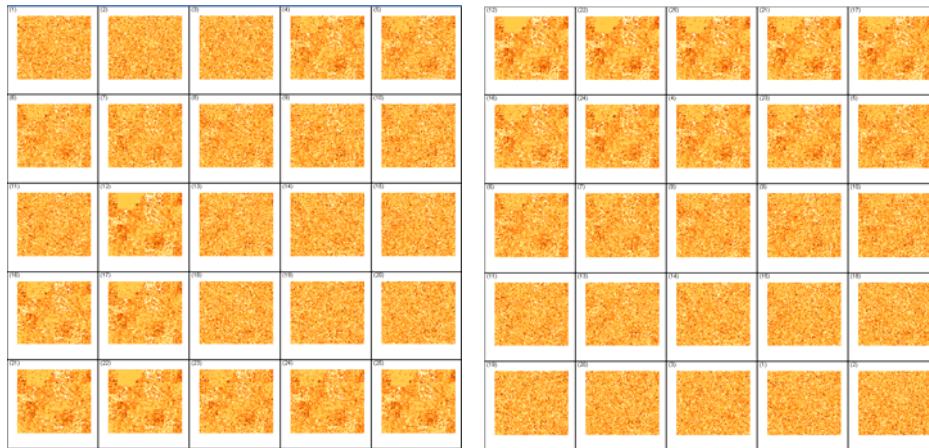
Detailed analysis of attribute combinations:  
The 2nd bar reveals an obvious pattern

25 most relevant Pixel Bar Charts constructed from highly ranked attribute combinations (Correlation Analysis / Classification)

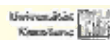


J. Schneidewind, M. Sips and D. A. Keim. Pixnostics: Towards Measuring the Value of Visualization. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '06), 2006.

## Automated Mapping Support for Pixel Bar Charts

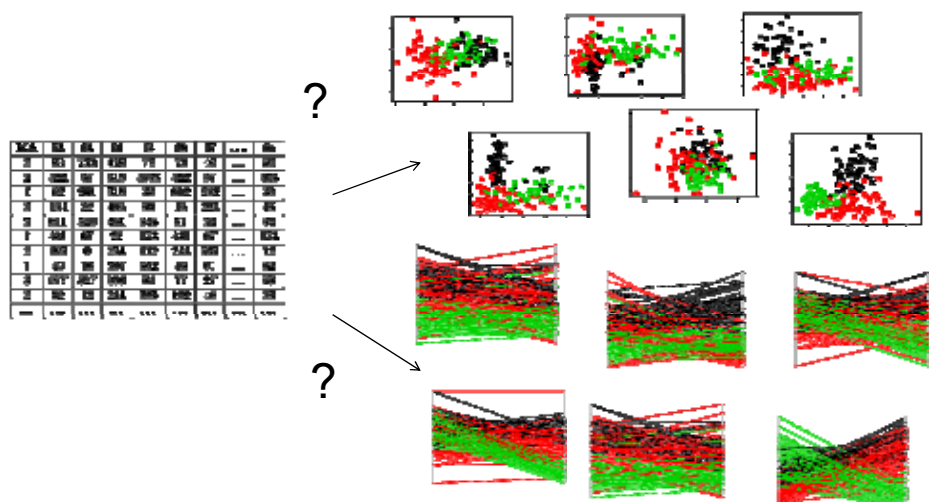


after sorting according to our measure



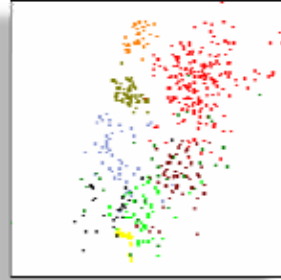
J. Schneidewind, M. Sips and D. A. Keim. Pixnostics: Towards Measuring the Value of Visualization. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '06), 2006.

## Automated Mapping Support



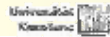
## Automated Mapping Support for Scatter Plots

- Scatterplots with class information
- Each color represents one class
- Evaluate scatterplots according their separation properties
- Algorithm:
  - Separate classes in distinct images
  - Compute a density image for each one
  - Estimate the mutual overlap



$$CDM = \sum_{k=1}^{M-1} \sum_{l=k+1}^M \sum_{i=1}^P \|d_k^i - d_l^i\|$$

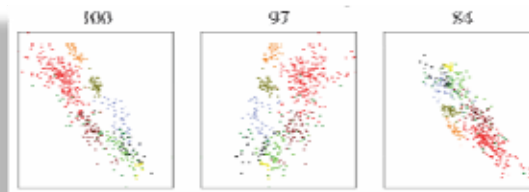
A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.



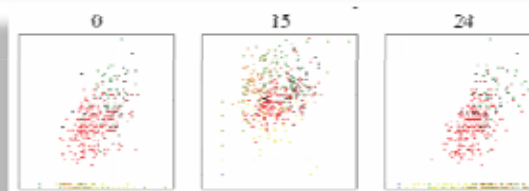
## Automated Mapping Support for Scatter Plots

**Olives dataset** (Class Density Measure)

**Best ranked views**



**Worst ranked views**

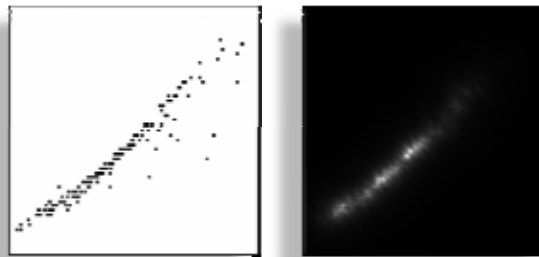


A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.



## Rotating Variance Measure

- Find linear and non-linear correlations between pairwise dimensions
- Continuous density field
  - Local density for each pixel  $p$ :  $\rho = 1/r$
  - $r$  = enclosing sphere of the  $k$ -nearest neighbors of  $p$



A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

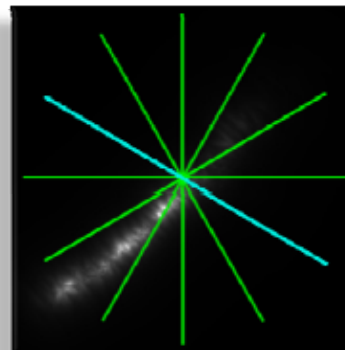


## Rotating Variance Measure

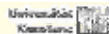
- Take samples along different lines centered at the corresponding pixel.
- Compute the weighted distribution for each pixel  $x^i$

$$v_{\theta}^i = \frac{\sum_{j=1}^3 p_{\theta}^{x_j^i} \|x^i - x^{x_j^i}\|}{\sum_{j=1}^3 p_{\theta}^{x_j^i}}$$

$$v^i = \min_{\theta \in [0, 2\pi]} v_{\theta}^i$$



A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.



## Rotating Variance Measure

- For each column  $y$  in the image we compute the minimum  $v^i$  value and sum up the result

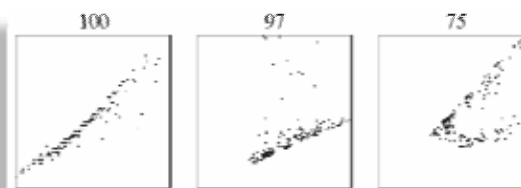
$$RVM = \frac{1}{\sum_x \min_y v(x,y)} \quad , \text{ where } v(x,y) = v^i$$

A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

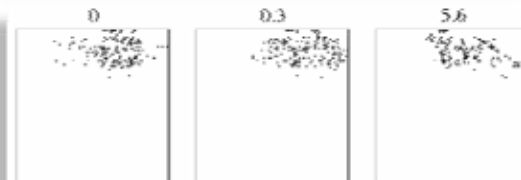
## Automated Mapping Support for Scatter Plots

### Parkinson's disease dataset (Rotating Variance Measure)

#### Best ranked views



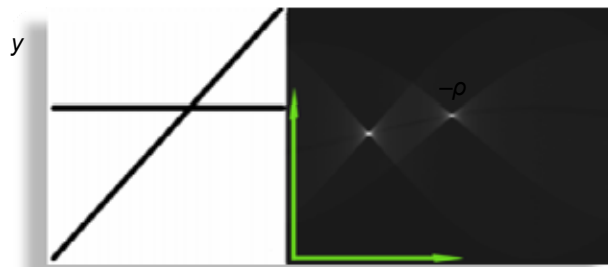
#### Worst ranked views



A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

## Hough Space Measure

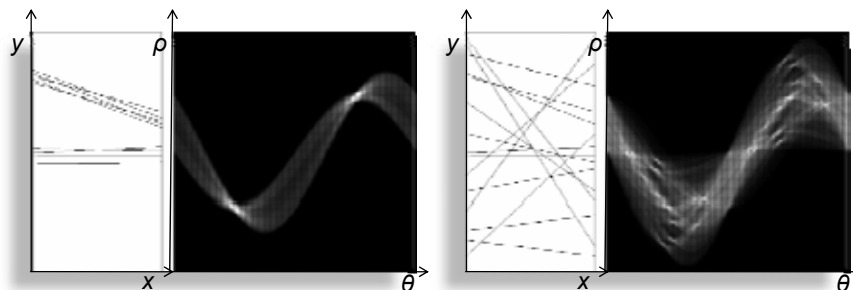
- Interesting patterns are usually clustered lines with similar positions and directions
- Hough Transform
  - $y = ax + b \rightarrow b = -ax + y$
  - *Problem:* the slope of vertical lines is infinite
  - *Normal representation:*  $\rho = x \cos \theta + y \sin \theta$



A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

## Hough Space Measure

- **For each non-background pixel in the visualization, we have a distinct sinusoidal curve in the  $\rho\theta$ -plane**
- **Hough/Accumulator Space**

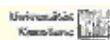


A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

## Hough Space Measure

- Good visualizations must contain fewer well defined clusters, i.e. accumulator cells with high values
- Compute the median value  $m$  as an adaptive threshold

$$s_{i,j} = 1 - \frac{\#highValueCells}{\#Cells}$$



A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

## Hough Space Measure

- Overall quality measure

$$HSM = \sum_{a_i \in I} s_{a_i, a_{i+1}}$$

- Exhaustively computing all n-dimensions combinations
  - Long computation time
  - Unfeasible for a large  $n$
- Solve a Traveling Salesman Problem
  - A\*-Search algorithm



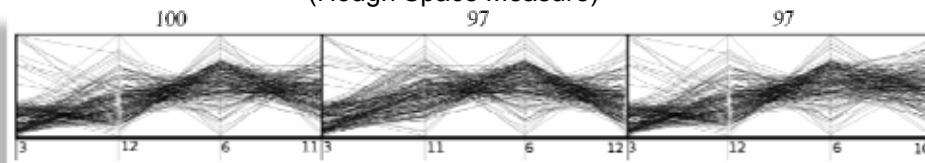
A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.



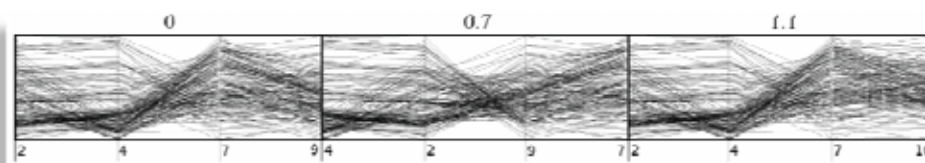
## Automated Mapping Support for Parallel Coordinates

### Parkinson's disease dataset

(Hough Space Measure)



Best ranked views



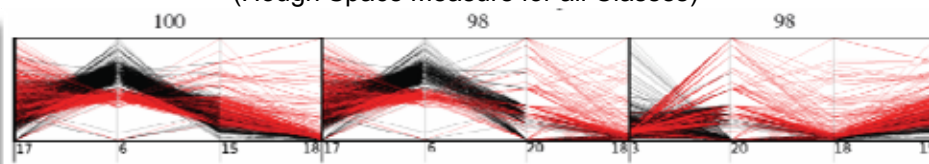
Worst ranked views

A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

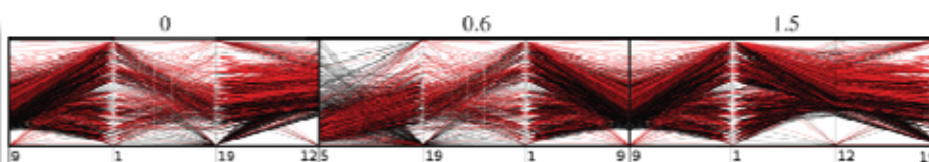
## Automated Mapping Support for Parallel Coordinates

### Cars dataset

(Hough Space Measure for all Classes)



Best ranked views

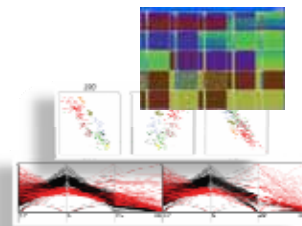


Worst ranked views

A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor and D. A. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '09), 59–66, 2009.

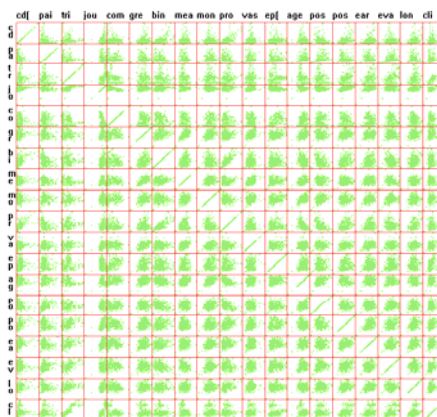
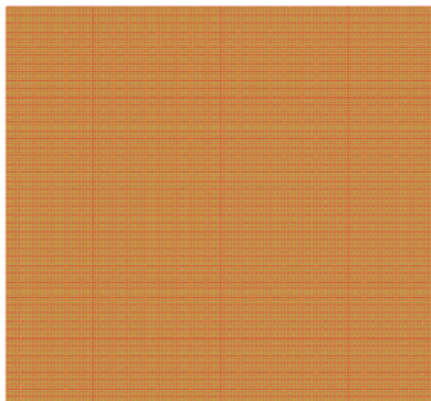
## Outline

- Visual Analytics
  - Definition
  - Challenges
  - Quality Metrics
- Quality Metrics for
  - Visual Mappings
  - Data Transformations
  - View Transformations
- Perspectives



## Quality Metrics for Data Transformation

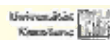
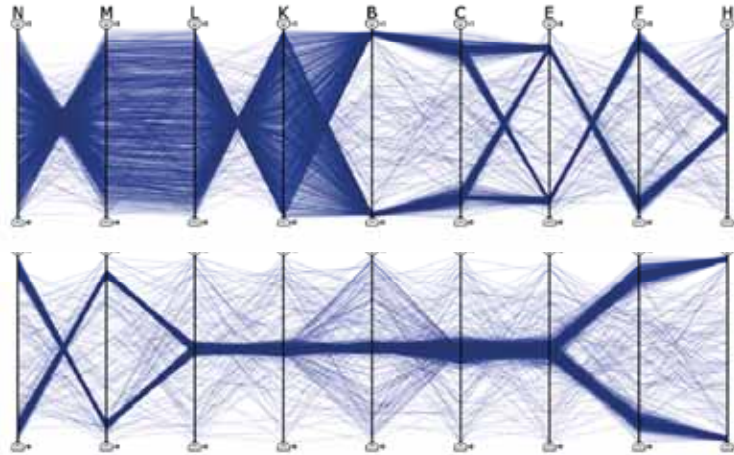
### Quality Metrics for Filtering



Jing Yang, Wei Peng, Matthew O. Ward and Elke A. Rundensteiner: Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets, IEEE Symposium on Information Visualization 2003 (InfoVis 2003), pp 105 - 112, 2003.

# Quality Metrics for Data Transformation

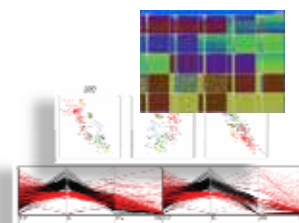
## Quality Metrics for Sorting Dimensions



S. Johansson and J. Johansson: Interactive Dimensionality Reduction Through User-defined Combinations of Quality Metrics, IEEE Trans. On Visualization and Computer Graphics, 2009.

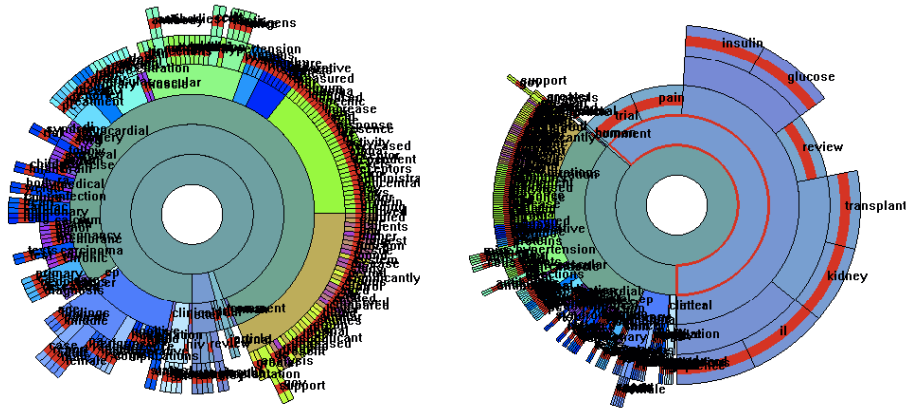
## Outline

- **Visual Analytics**
  - Definition
  - Challenges
  - Quality Metrics
- **Quality Metrics for**
  - Visual Mappings
  - Data Transformations
  - View Transformations
- **Perspectives**



# Quality Metrics for View Transformation

## Metrics for View Distortion

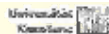
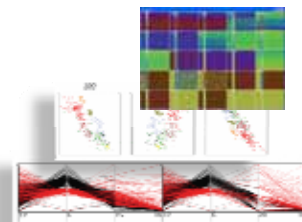


Jing Yang, Wei Peng, Matthew O. Ward and Elke A. Rundensteiner: Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets, IEEE Symposium on Information Visualization 2003 (InfoVis 2003), pp 105 - 112, 2003.



## Outline

- **Visual Analytics**
  - Definition
  - Challenges
  - Quality Metrics
- **Quality Metrics for**
  - Visual Mappings
  - Data Transformations
  - View Transformations
- **Perspectives**



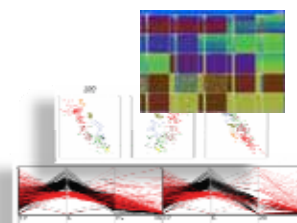
Paper Title	Visualization technique			What is measured						What is to be measured		Purpose					Year
	SP	PC	other	clustering	correlation	outliers	complex patterns	image quality	feature space	data	image	production	analysis	abstraction	visual mapping	view transformation	
Combining automatic analysis and visualization techniques for effective evaluation of high-dimensional data [26]	SP	PC		clustering	correlation		complex patterns		feature space	data	image	production	analysis				
High-Dimensional Visual Analysis: Interactive Exploration Guided by Perceptual Views of Point Distributions [25]	SP	PC		clustering		outliers	complex patterns				image	production	analysis				
Combining visualization and visual approaches for interactive feature selection and evaluation [24]			other		correlation					data		production	analysis				
Exploring High-D Spaces with Multiscale Metrics and Detail Mappings [16]			other							data		production	analysis				
Exploring the Visual Analysis of High-Dimensional Complex Using Quality Measures [23]			other	clustering	correlation	outliers				data	image	production	analysis		visual mapping		
Interactive Dimensionality Reduction Through User-Guided Construction of Quality Metrics [22]	PC			clustering	correlation	outliers				data		production	analysis				5, 7
Interactive Hierarchical Dimension Clustering, Sampling and Filtering for Exploration of High-Dimensional Datasets [20]	PC	other	other	clustering						data		production	analysis			view transformation	5, 7
Perceptual Image Space Metrics for Parallel Coordinates [15]	PC			clustering	correlation			image quality			image	production	analysis				5
A Projection-Retail Approach for Exploratory Data Analysis [12]	SP			clustering						data		production					
A Rank-by-Feature Framework for Comprehensive Multidimensional Data Exploration Using Latent Dimensional Properties [10]	SP	other	other	clustering	correlation	outliers	complex patterns			data		production					6
Filtering and Clustering Features Subsequently for Comprehending High-Dimensional Multivariate Data Using Correlation-Based Metrics [18]	SP	other	other	clustering							image	production					7
Design-Themed, Responsive [14]	SP			clustering		outliers	complex patterns				image	production					
Selecting good views of high-dimensional data using video consistency [13]	SP							image quality	feature space	data		production					7
Cluster Reduction in High-Dimensional Data Visualization Using Dimension Reducing [11]	SP	PC	other		correlation	outliers		image quality		data	image		analysis				
Interactive Clustering of Dimensions for an Exploratory Visualization of Multidimensional Data [9]	PC		other		correlation					data			analysis				
Dimensionality Reduction Quality in Multidimensional Visualizations [8]	SP	PC	other						feature space	data				abstraction			7
Quality Metrics for 3D Scatterplot Techniques: Automatically Feature-Driven Clustering [4]	SP			clustering				feature space	data	image				abstraction			
A Screen Space Quality Metric for Data Abstraction [17]	PC							feature space		image				analysis			
Enabling Automatic Cluster Selection in Parallel Coordinates Plots [19]	PC							image quality		image				analysis			7
Perceptual Trends: Interpreting the Space of Visualization [21]			other		correlation		complex patterns			data	image				visual mapping		

Legend: SP = scatter plot (2D/3D), PC = parallel coordinates, feature space = feature space, correlation = correlation, C = color metric, T = text/metric.

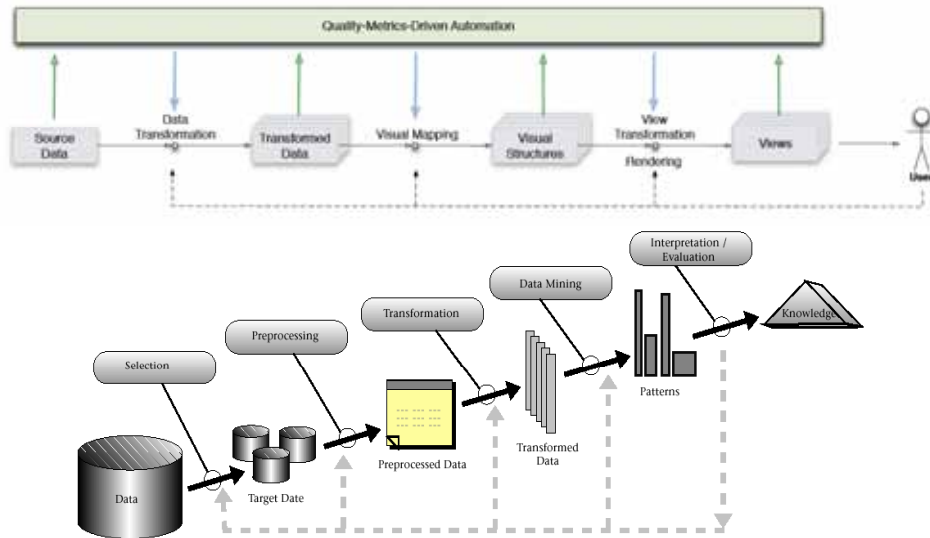
Enrico Bertini, Member, Andrada Tatu, Daniel Keim: Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization, IEEE Conf. on Information Visualization, 2011.

## Outline

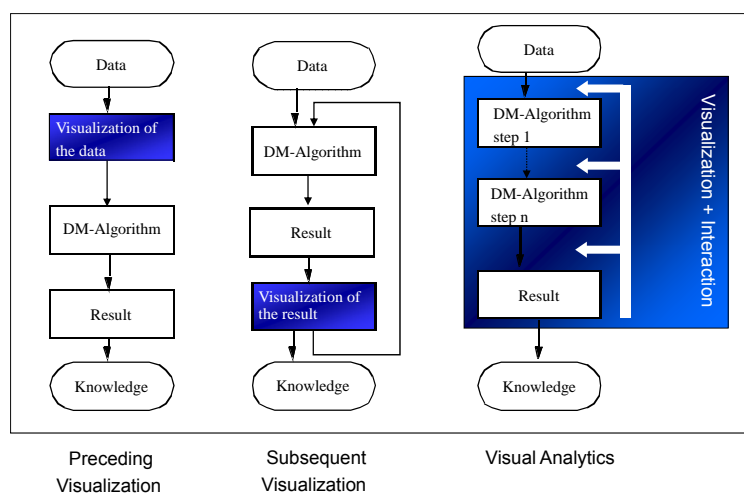
- Visual Analytics
  - Definition
  - Challenges
  - Quality Metrics
- Quality Metrics for
  - Visual Mappings
  - Data Transformations
  - View Transformations
- Perspectives



## Pipelines



## Visual Analytics





## Conclusion

**“All truths are easy to understand once they are discovered; the point is to discover them.”**

**Galileo Galilei (1564-1642)**

## Questions?

*[infovis.uni-konstanz.de](http://infovis.uni-konstanz.de)*

