



# Information Visualization II

## Techniques for High-Dimensional Data



# Outline

- Geometric Transformations
  - Scatterplots *(DONE)*
  - Parallel Coordinates *(today)*
  - Projections *(next time)*
- Pixel-based Techniques *(next time)*



# Schedule InfoVis II (tentative)

Date	Topic
16.04.2019	Organisation & RECAP
23.04.2019	Techniques for High-Dimensional Data (1)
30.04.2019	Techniques for High-Dimensional Data (2)
07.05.2019	Canceled
14.05.2019	Techniques for High-Dimensional Data (3) & Geo-Spatial Data(1)
21.05.2019	Geo-Spatial Data(2)
28.05.2019	Techniques for Temporal Data(1)
04.06.2019	Techniques for Temporal Data(2)
11.06.2019	Techniques for Temporal Data(3) & Techniques for Trees and Networks(1)
18.06.2019	Techniques for Trees and Networks(2)
25.06.2019	Techniques for Glyphs and Iconic Displays
02.07.2019	Techniques for Text Data
09.07.2019	Visualization Systems & Questions
16.07.2019	1st EXAM written or oral

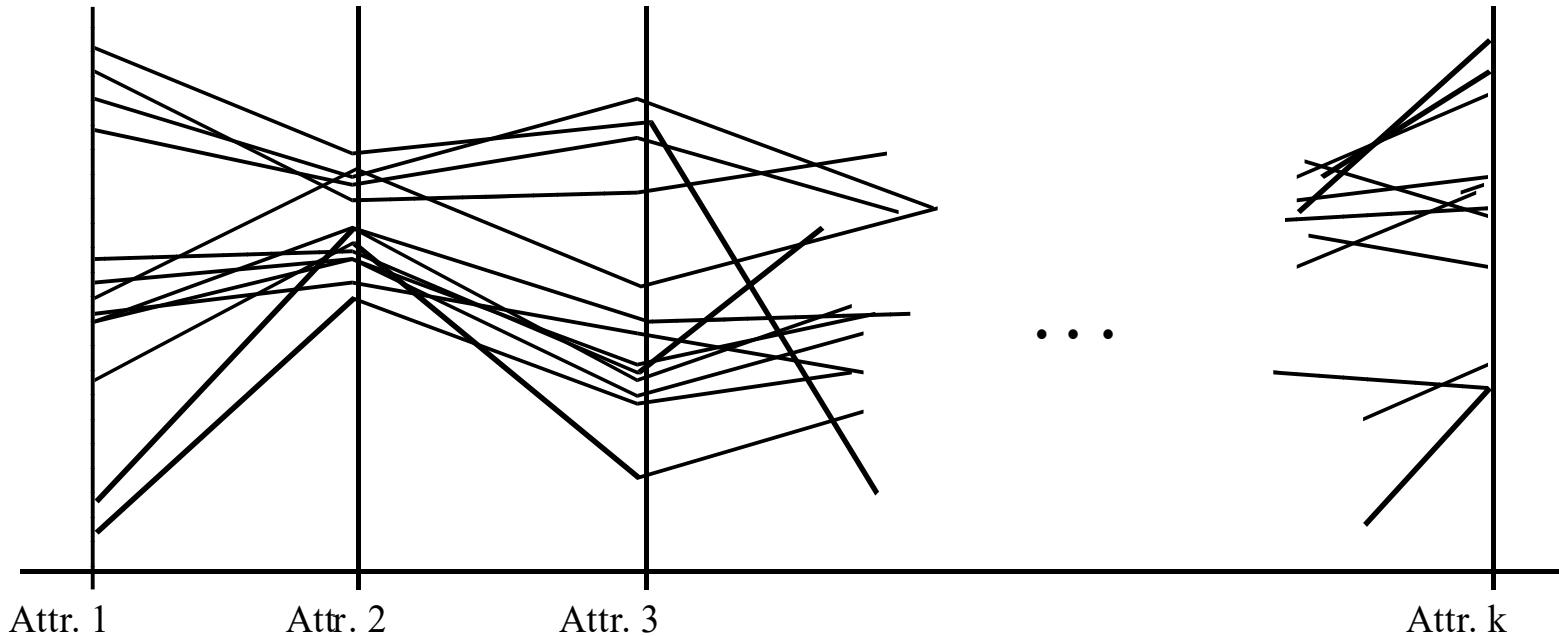


## 2. PARALLEL COORDINATES





# Parallel Coordinates



- $n$  equidistant axes which are parallel to one of the screen axes and correspond to the attributes
- the axes are scaled to the [minimum, maximum] - range of the corresponding attribute
- every data item corresponds to a polygonal line which intersects each of the axes at the point which corresponds to the value for the attribute

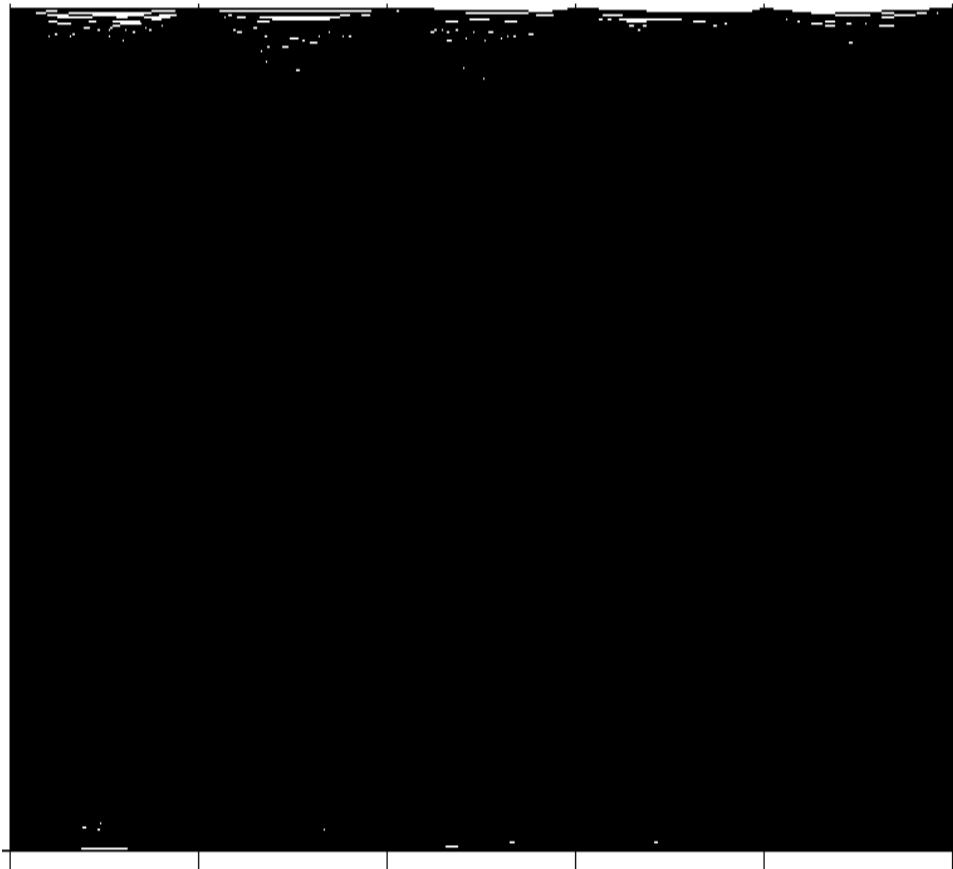


# Parallel Coordinates - Challenges

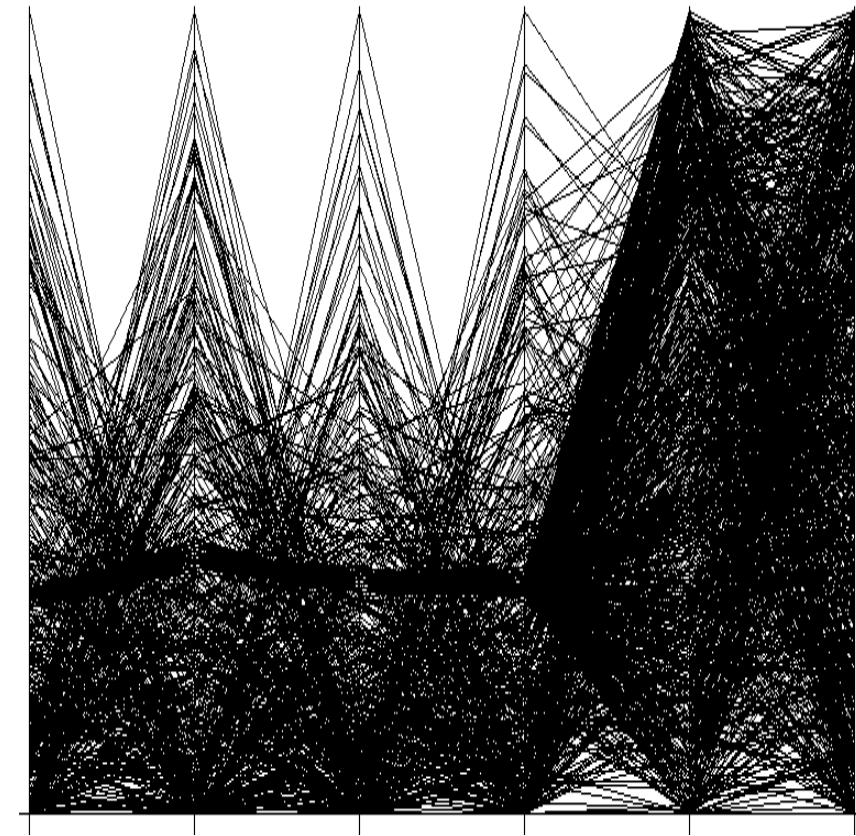
1. Increasing number of data points may result in vanishing patterns due to overplotting.
2. A perceived pattern depends primarily on the ordering of the dimensions.
3. A large number of dimensions decreases the available screen space between axes.
4. (Negative correlations are visible more clearly)



# Parallel Coordinates - Sampling



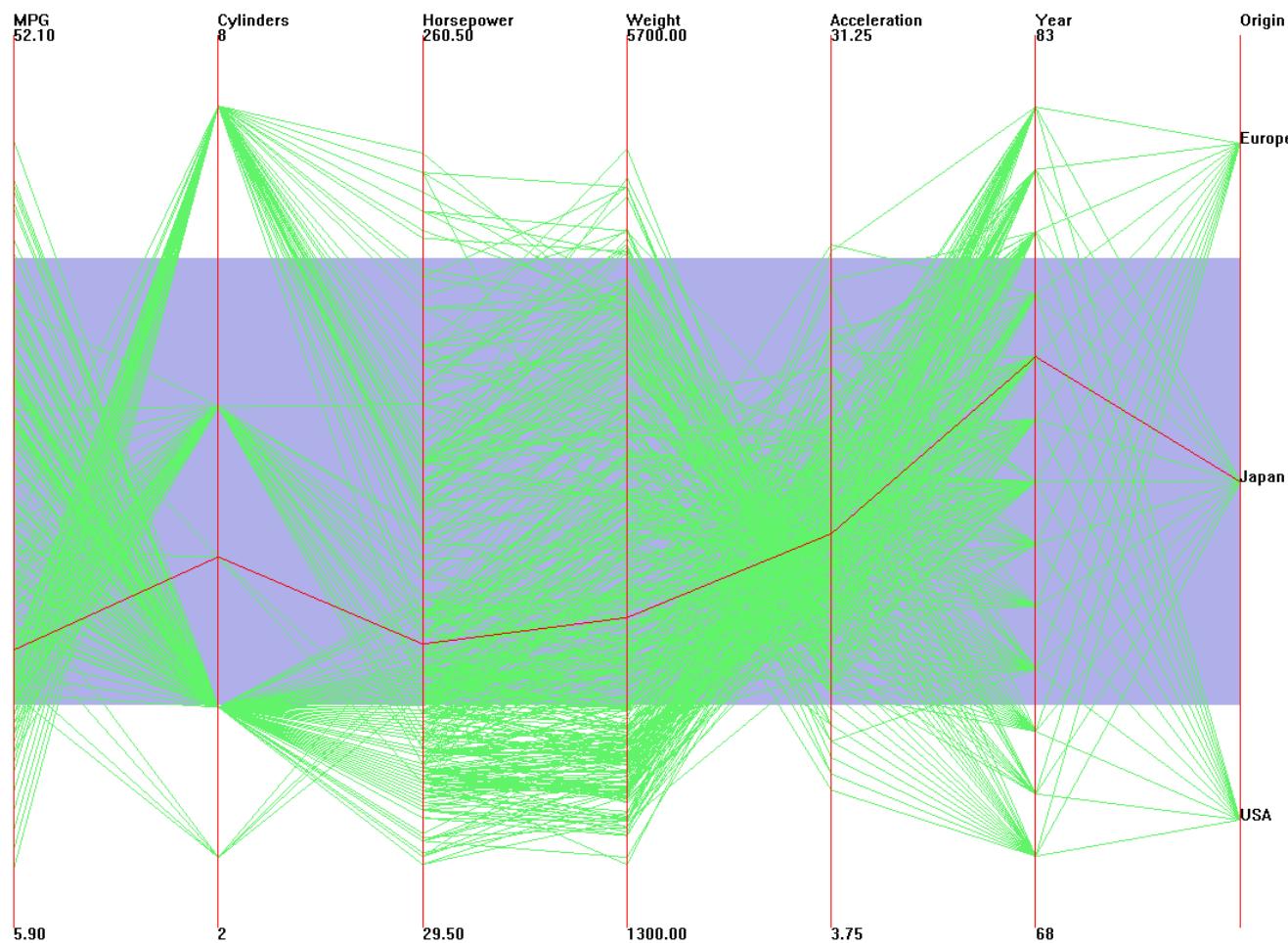
15.000 data items with noise



5 % of the data (750 data items)

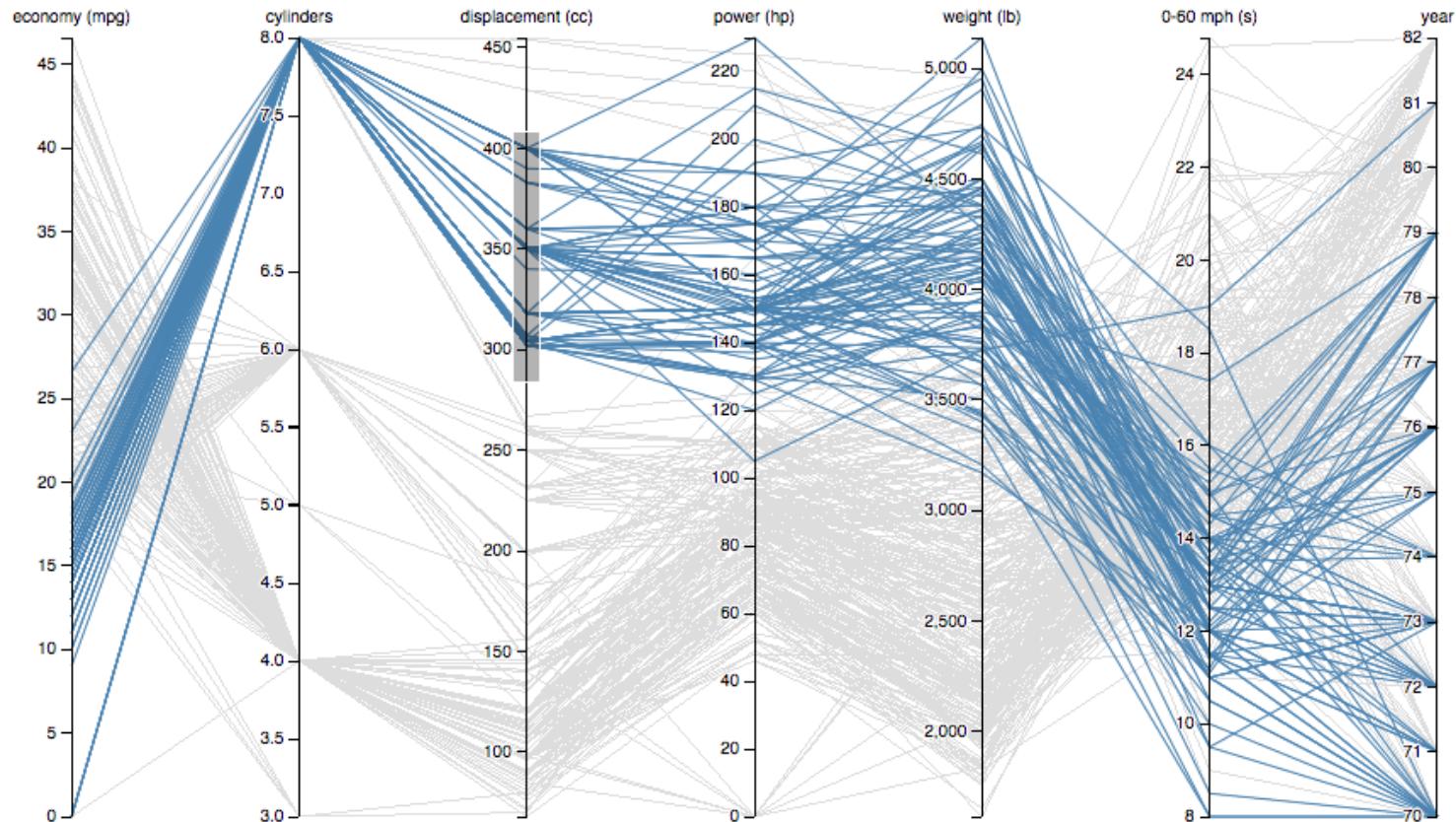


# Parallel Coordinates - Filtering



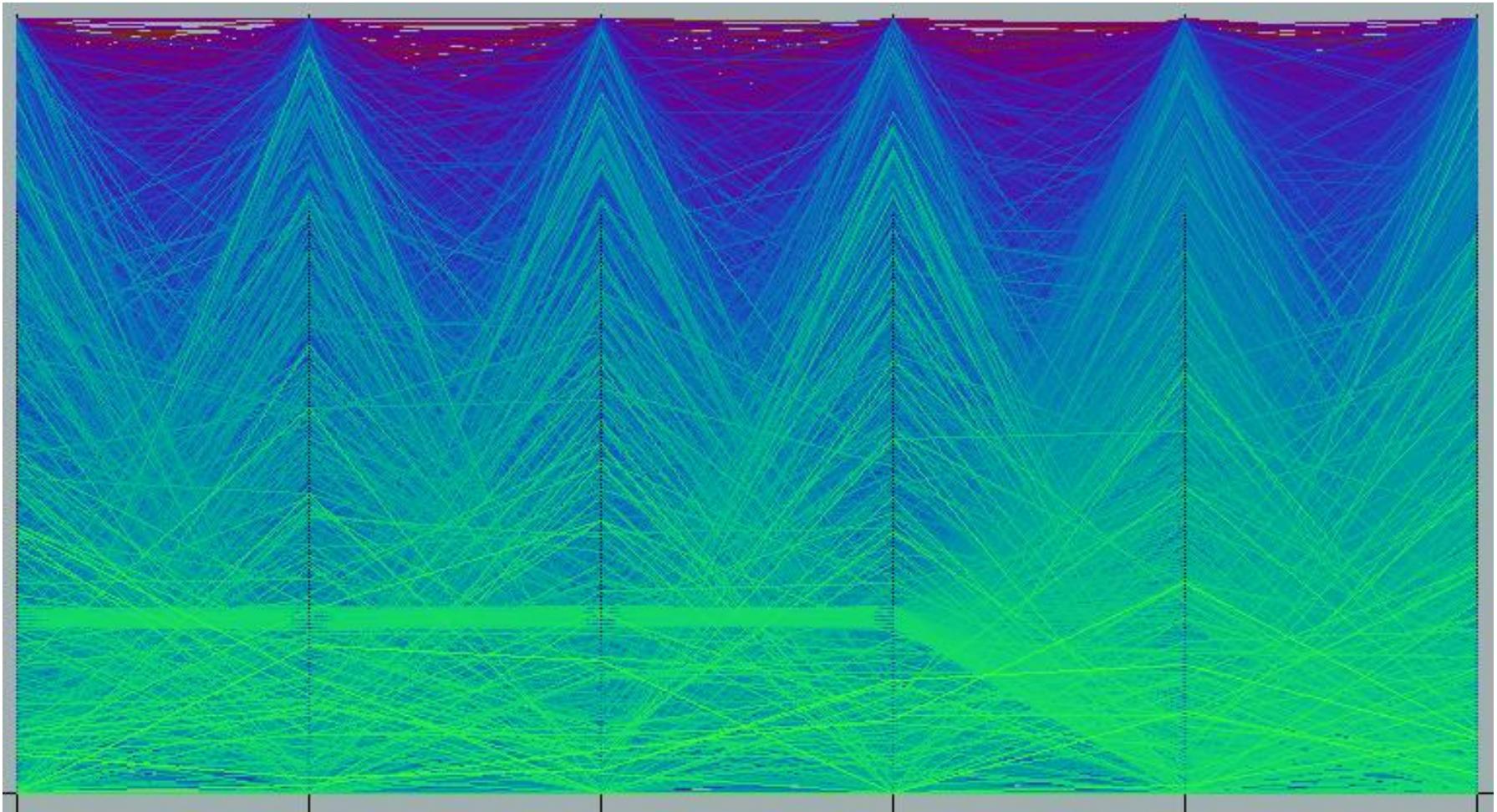


# Parallel Coordinates - Filtering





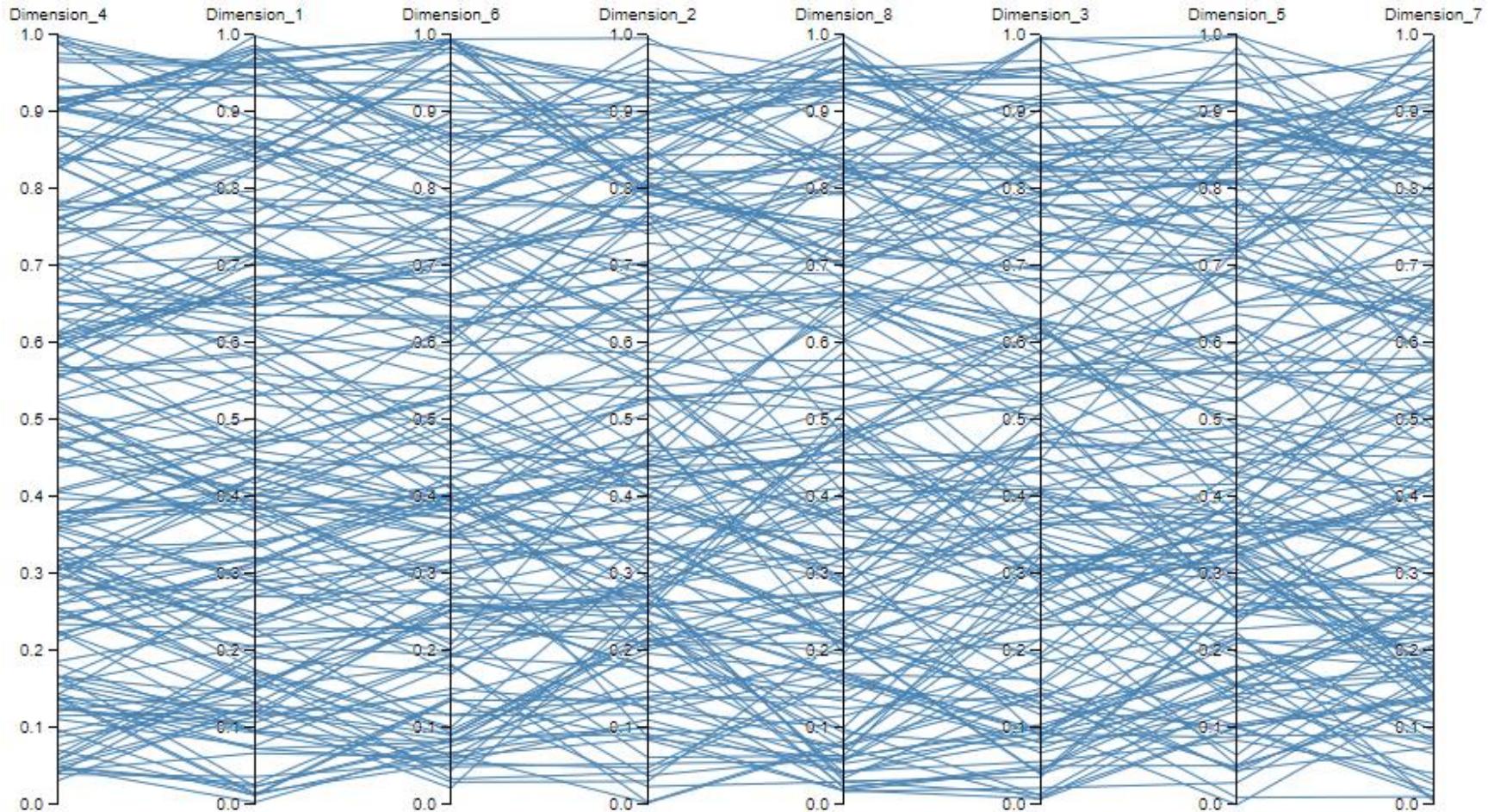
# Parallel Coordinates - Coloring



15.000 data items with a query-dependent coloring

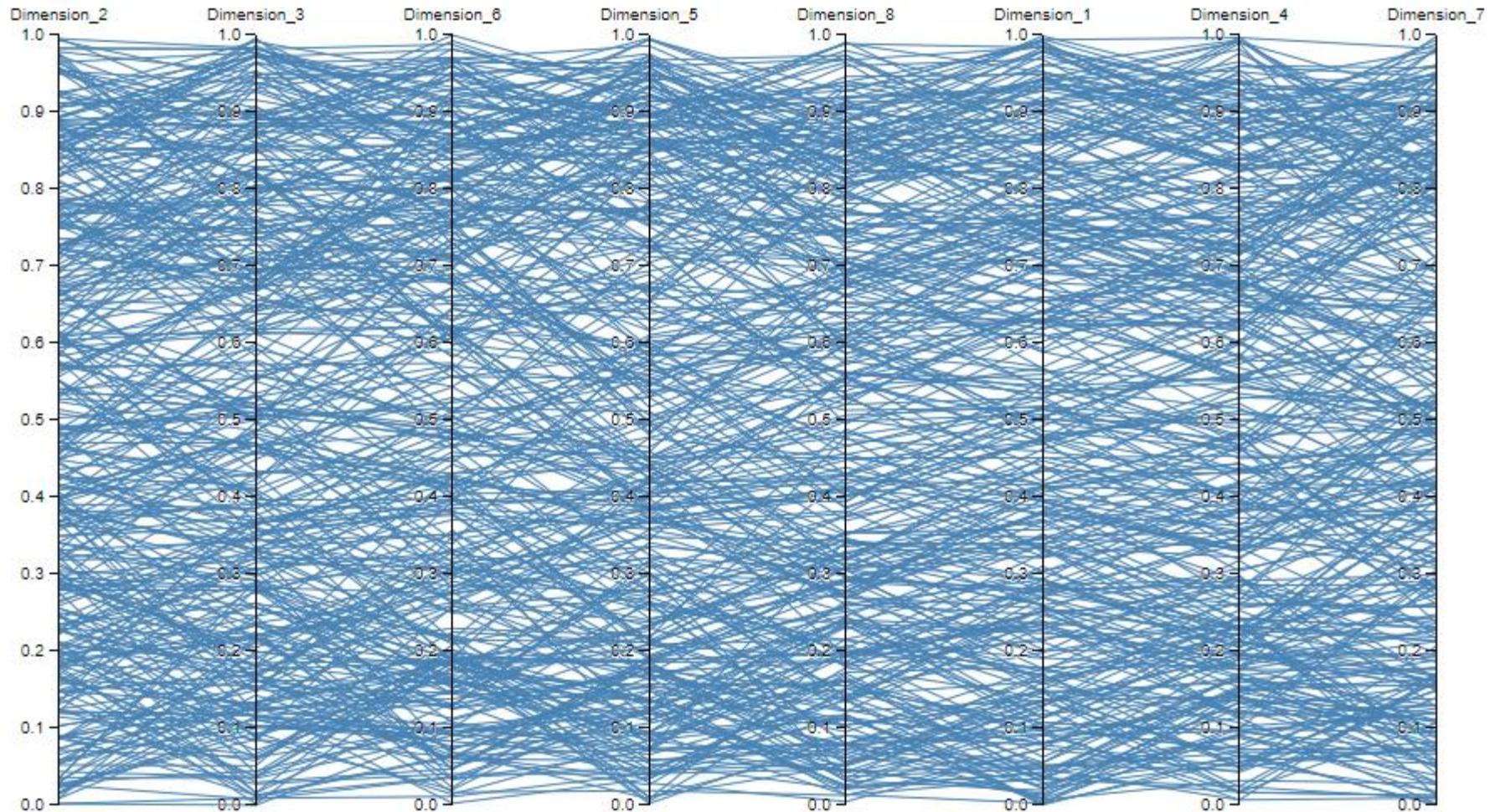


# PCP - Effect of Clutter



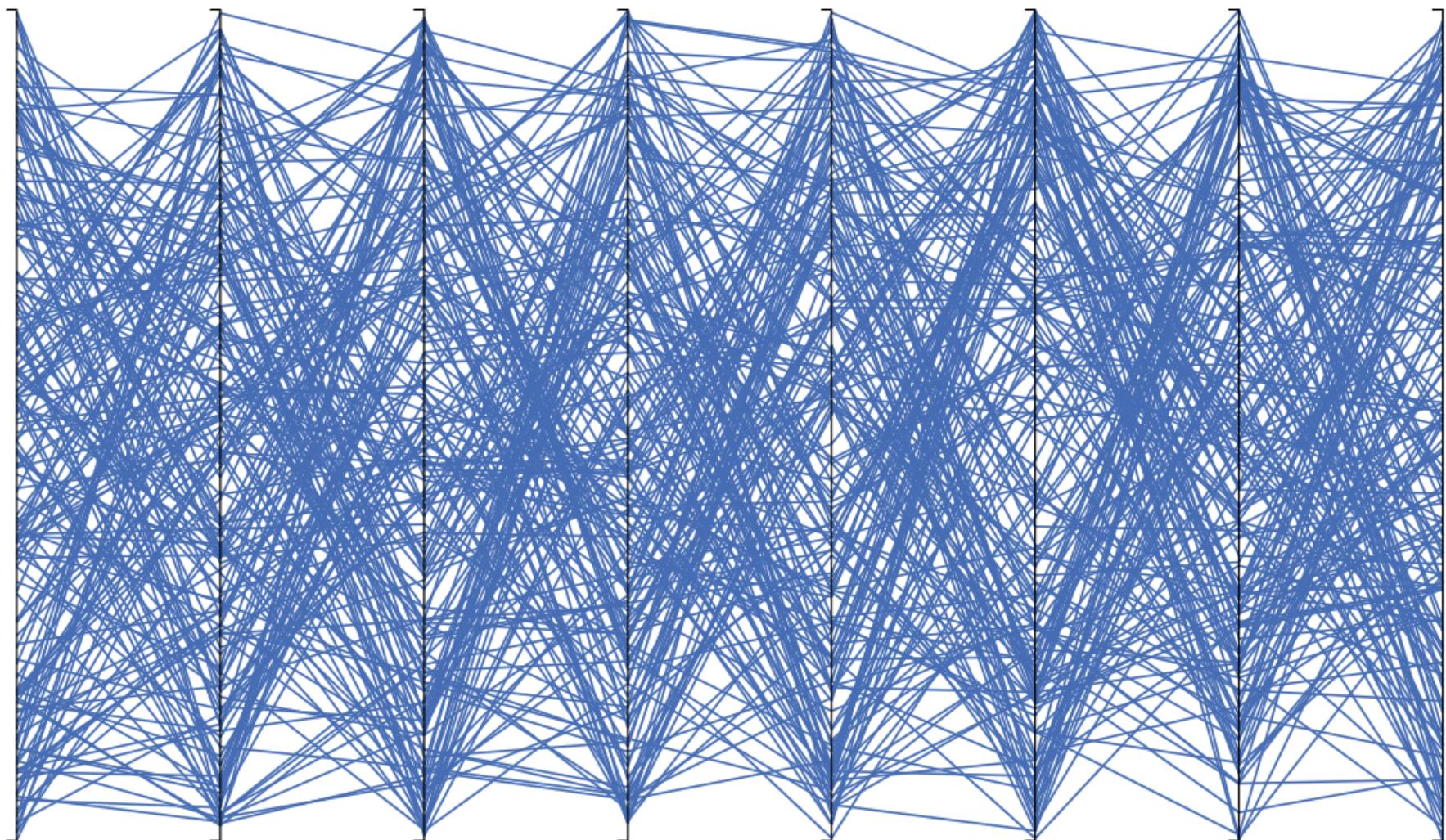


# PCP - Effect of Clutter



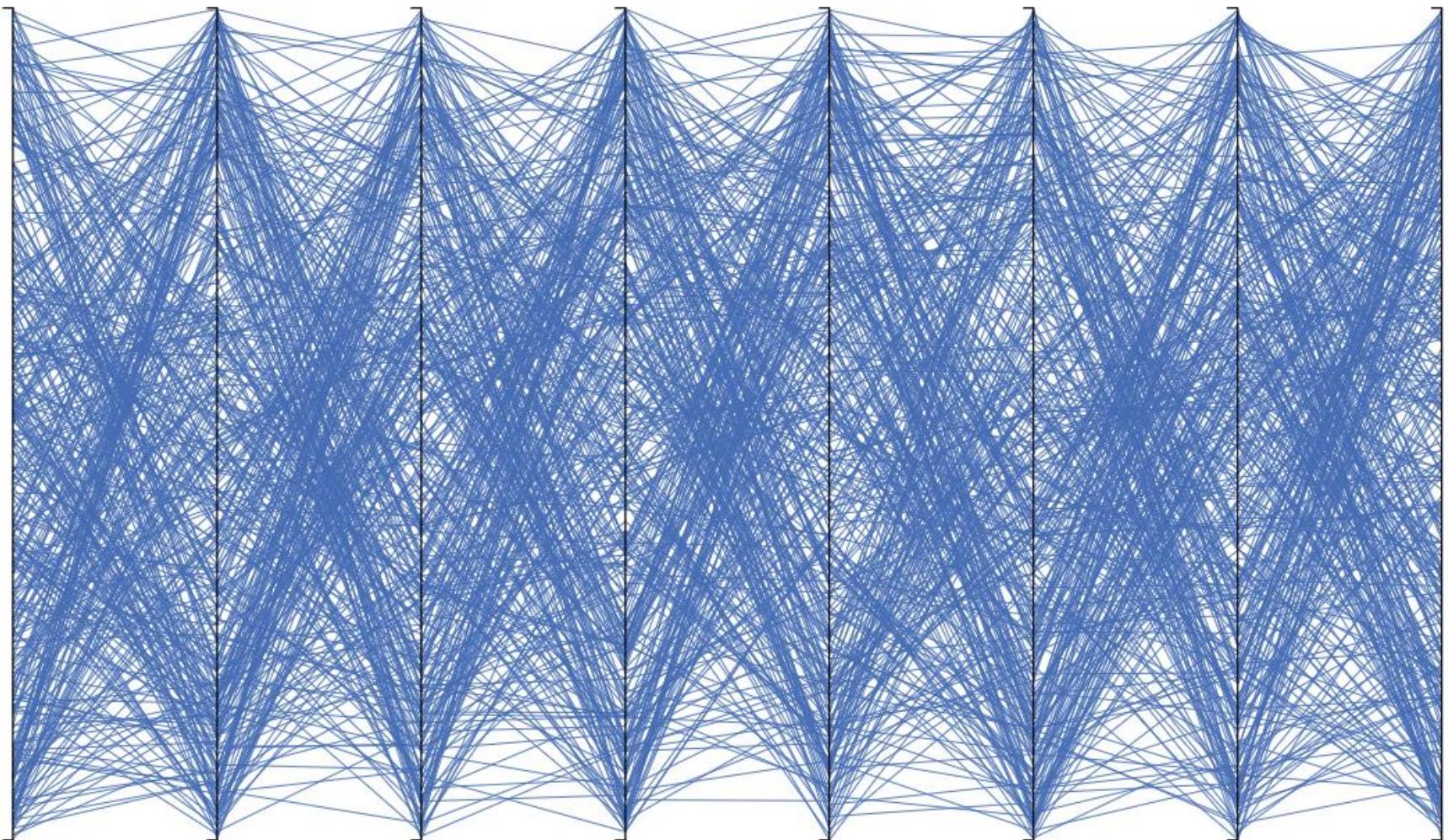


# PCP - Effect of Clutter



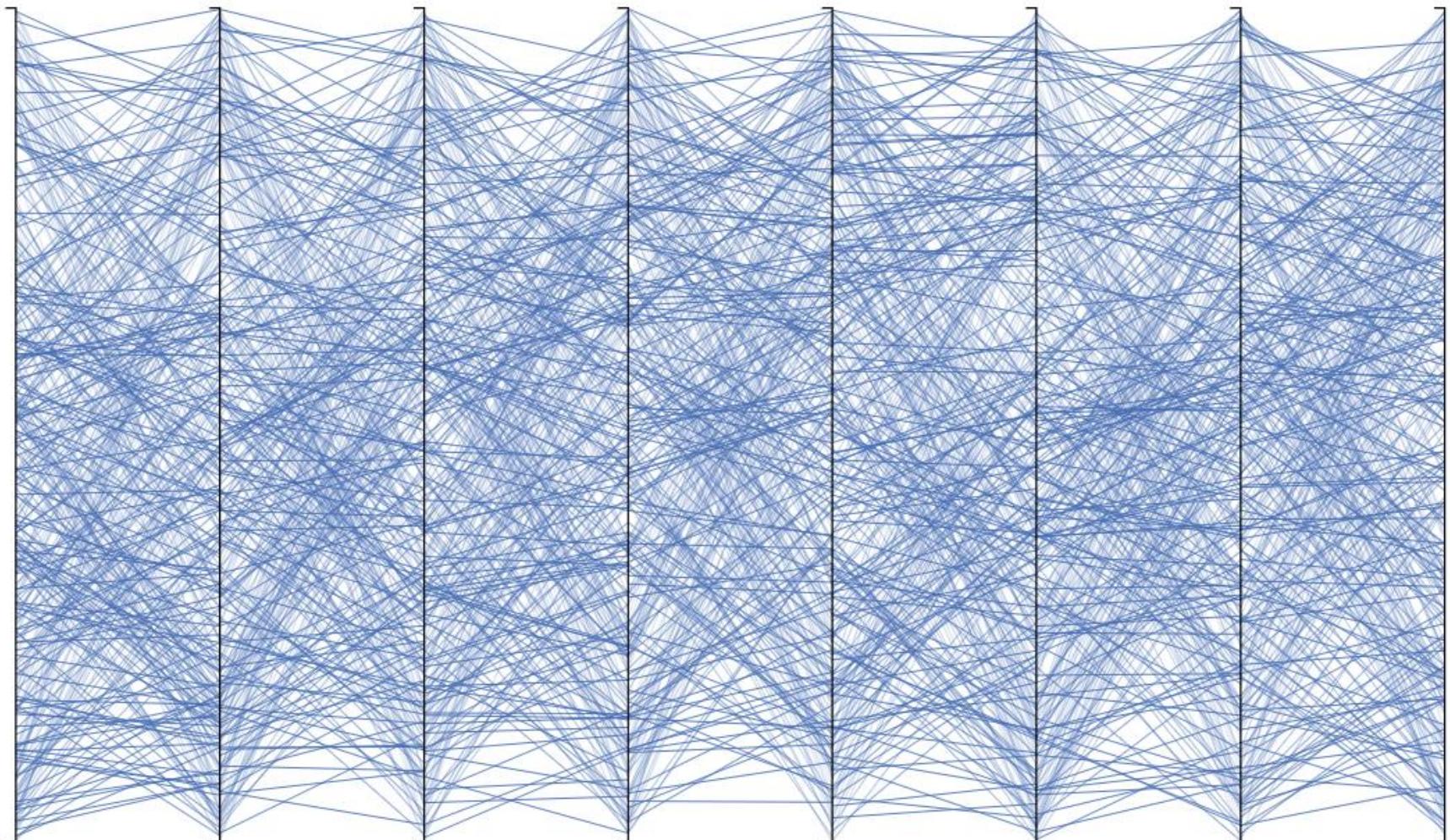


# PCP - Effect of Clutter





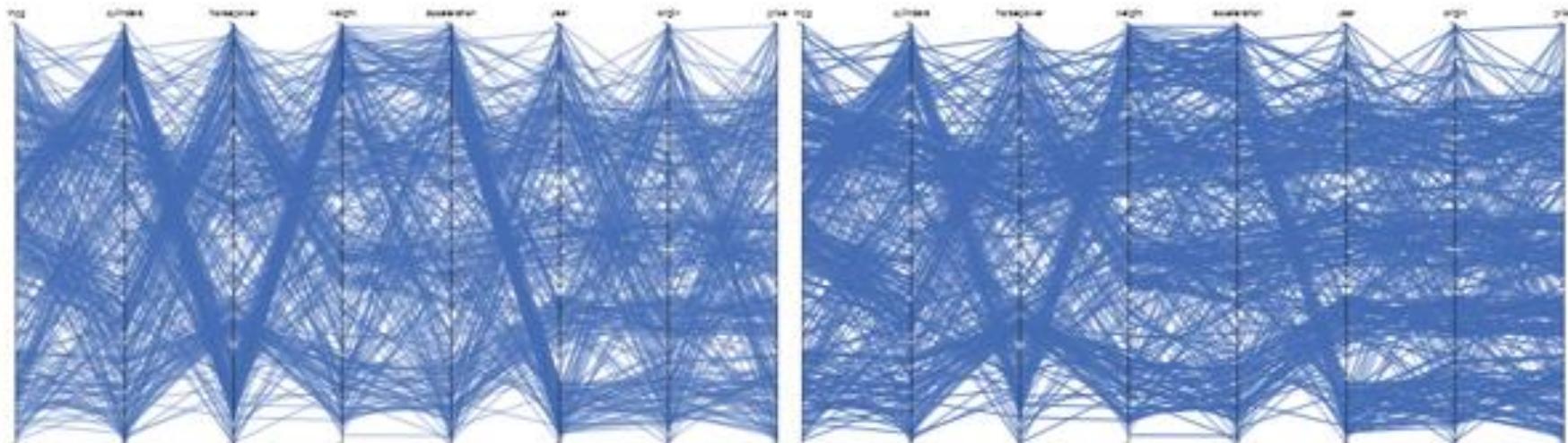
# PCP - Effect of Clutter





# PCP - Effect of Clutter

## Adaptive line width:





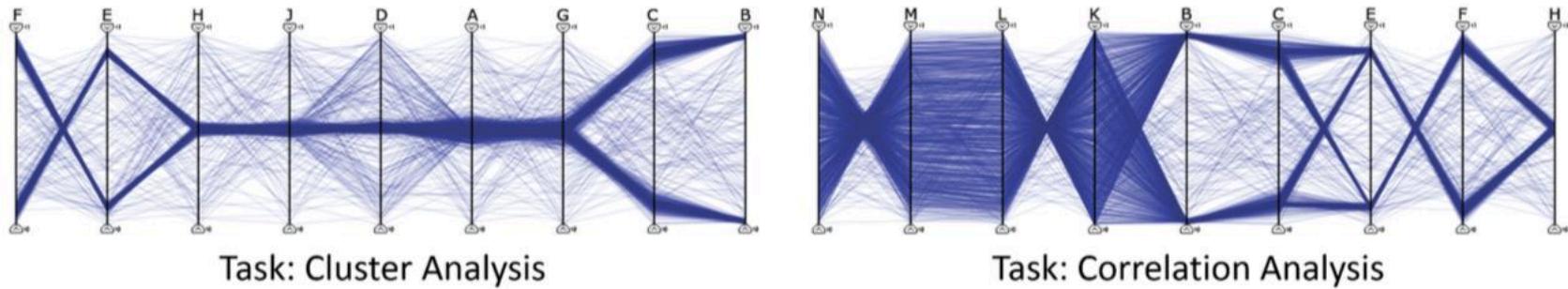
# Parallel Coordinates - Challenges

1. Increasing number of data points may result in vanishing patterns due to overplotting.
2. A perceived pattern depends primarily on the ordering of the dimensions.
3. A large number of dimensions decreases the available screen space between axes.
4. (Negative correlations are visible more clearly)



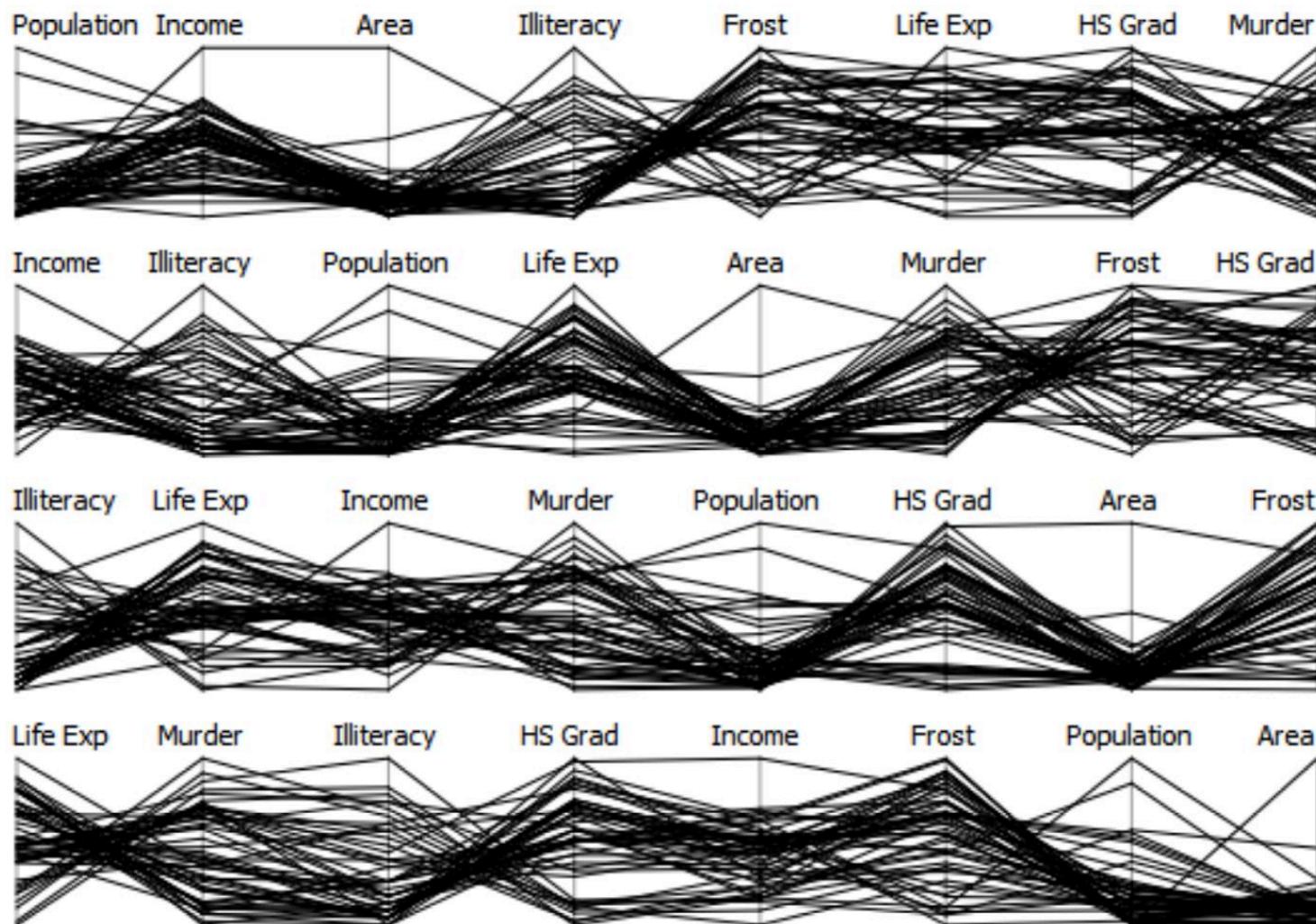
# Parallel Coordinates - Ordering

- Consequences of different ordering strategies:



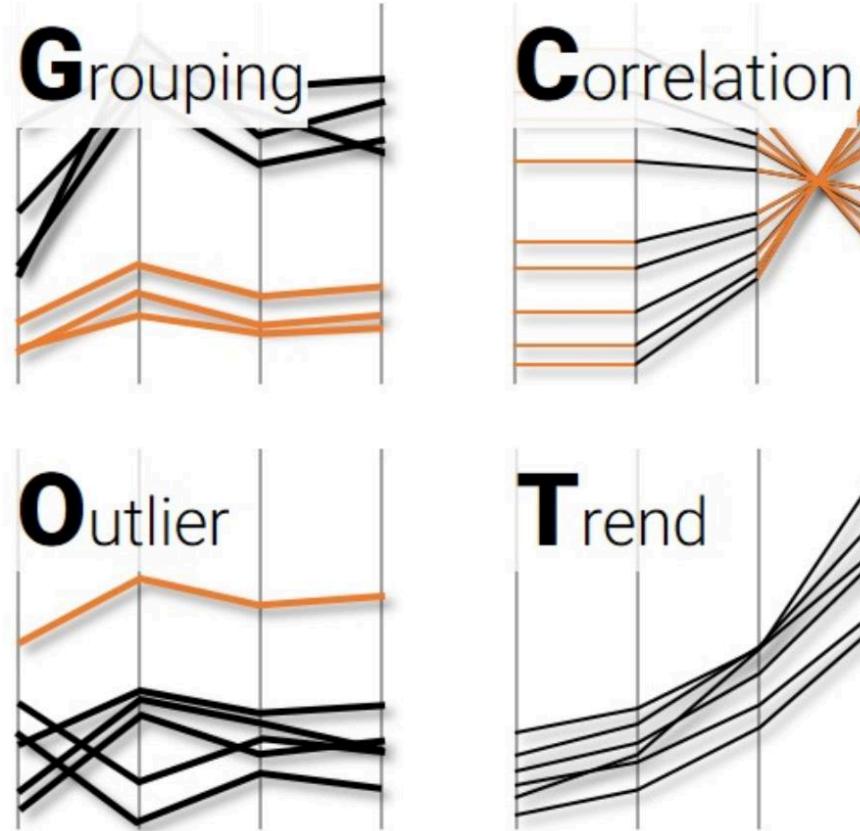
- There are  $n!$  possible orderings.
- Problem of finding optimal ordering is NP-complete

# PCP – Ordering (Example)





# PCP – Patterns/Tasks



Quality metrics for information visualization

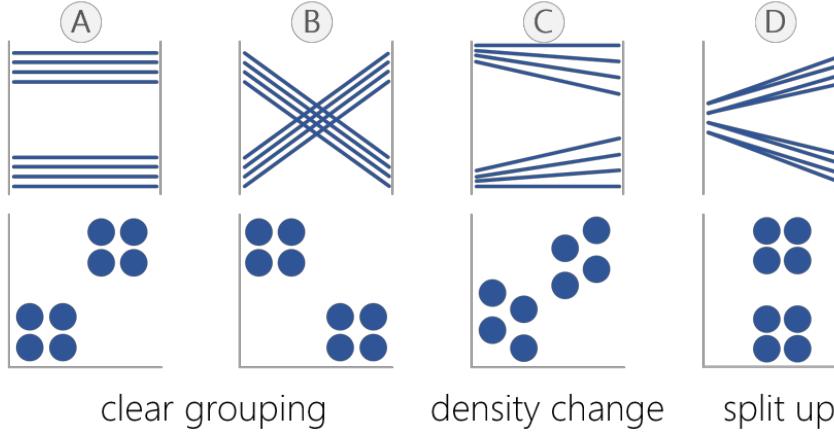
Behrisch, Michael et al.

Computer Graphics Forum: 625–662, Wiley Online Library, 2018



# PCP – Patterns/Tasks

## Clusters and Neighbors



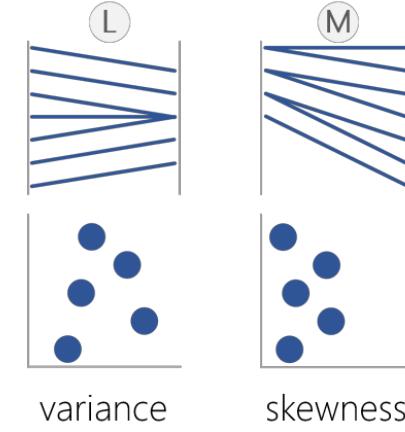
clear grouping

density change

split up

neighborhood

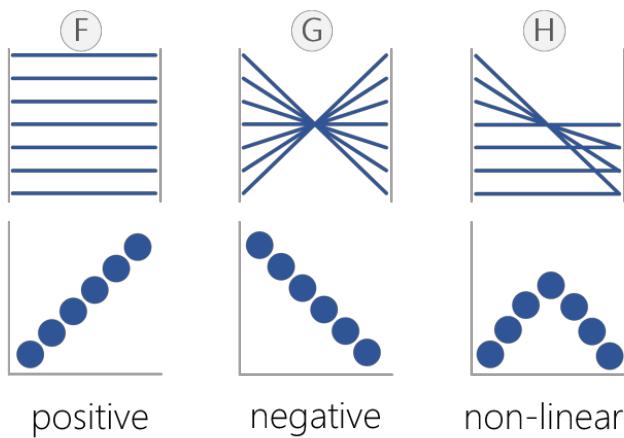
## Dim. Properties



variance

skewness

## Correlation

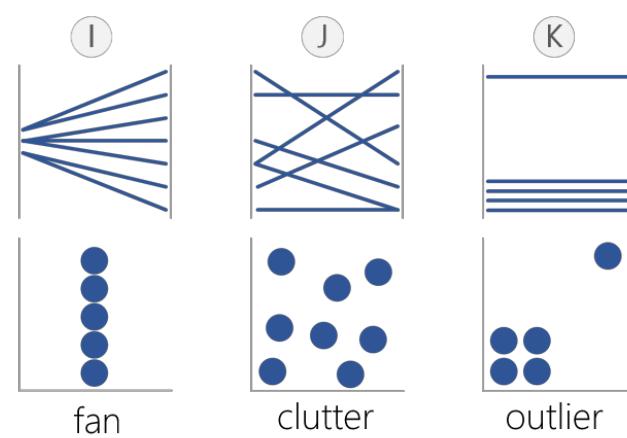


positive

negative

non-linear

## Outlier and Clutter



fan

clutter

outlier

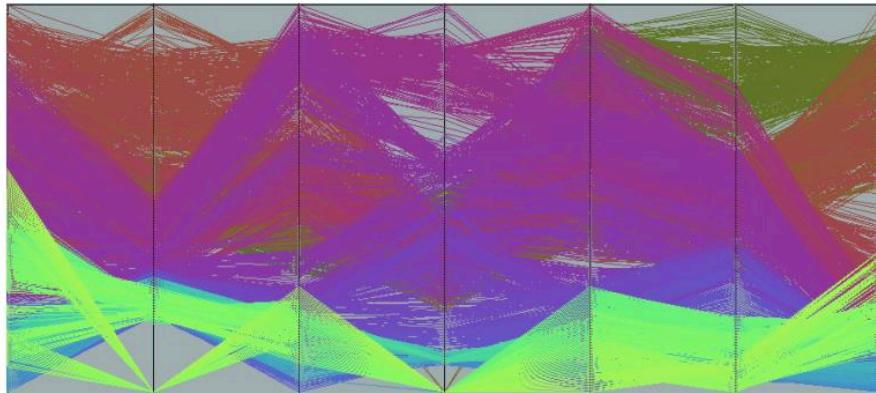
# PCP – Ordering (Overview)

		Clustering & Neighbor	Correlation	Clutter	Dimension	Properties		
						Axes	Dim.	Eval.
Objective Function & Optimization Algorithm	[47] Tatu '11 (no label)	● ● ●		⊗				↔
	[47] Tatu '11 (given label)	● ● ○		⊗				↔
	[54] Zhou '18 (cluster trace)	● ● ○		⊗				E
	[39] Peltonen '17 (neighbor)	● ● ●	●	⊗ ⊗				E
	[9] Dasgupta '10 (overpl.)	● ●	⊗ ⊗					E
	[9] Dasgupta '10 (parall.)	● ⊗	● ⊗	⊗				E
	[3] Ankerst '98 (similar)	● ⊗	● ⊗	⊗				E
	[51] Yang '03 (similarity)	● ⊗	● ⊗	⊗			DR +S	E
	[4] Artero '06 (clutter)	○ ⊗	● ⊗	⊗ ⊗				E
	[3] Ankerst '98 (correl.)	○ ○	● ○	⊗				E
	★ Blumenschein '19 (dis)	⊗ ●	⊗ ●	⊗				∅
	[9] Dasgupta '10 (cross.)	⊗ ○	⊗ ●	○				E
	[9] Dasgupta '10 (angle)	⊗ ○	⊗ ●	○				E
	[9] Dasgupta '10 (mut. i.)		● ● ●	⊗				E
	[9] Dasgupta '10 (diverg.)			⊗ ● ●	● ○			E
	[9] Dasgupta '10 (entropy)			⊗				E
	[40] Peng '04 (outlier)			⊗ ⊗				E
	[32] Makwana '12 (struc.)		●	●				E
Dimension QM Reordering Algorithm	[30] Lu '12 (non-lin. corr.)		○ ○ ○	⊗				E
	[28] Johansson '09 (correl.)	○ ○	● ○	⊗			DR	E
	[4] Artero '06 (similarity)	● ⊗	● ○	⊗			DR	E
	[28] Johansson '09 (cluster)	● ● ● ●		⊗			DR	E
	[13] Ferdosi '11 (subspace)	● ● ● ●		⊗			DR +S	↔
	[21] Huang '11 (set theory)	● ● ●		⊗				E
	[28] Johansson '09 (outlier)		● ○			DR		E
Dimension QM	[31] Lu '16 (svd)			⊗ ● ● ●		DR		E
	[51] Yang '03 (variance)			⊗ ● ● ●		DR +S		E
	[41] Schloerke '18 (skewn.)			⊗ ●				-
	[41] Schloerke '18 (outlier)			● ○				-
	[41] Schloerke '18 (class)	● ● ● ○		⊗				-

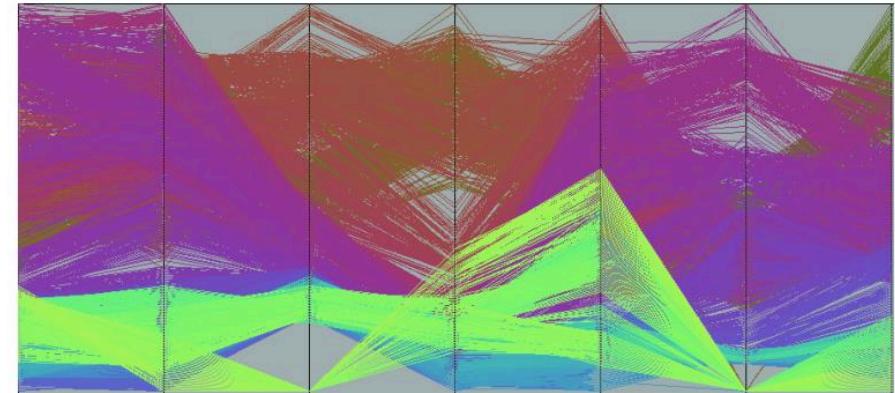


# PCP – Ordering (Similarity)

- Order the dimensions using the Euclidean distance between two dimensions.
- Measure the sum of distances between all dimension pairs.
- Minimize the global distance value (Optimization)



a. Sequential Arrangement



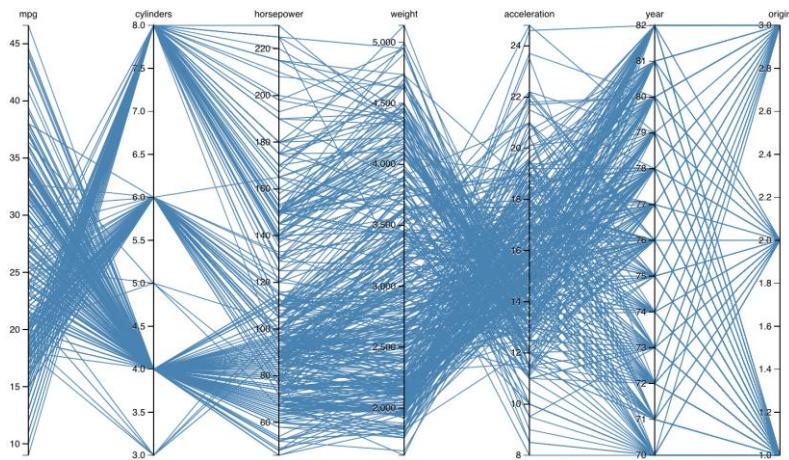
b. Similarity Arrangement

Similarity Clustering of Dimensions for an Enhanced Visualization of Multidimensional Data  
Michael Ankerst, Stefan Berchtold, Daniel A. Keim: Information Visualization: 52-60, IEEE, 1998

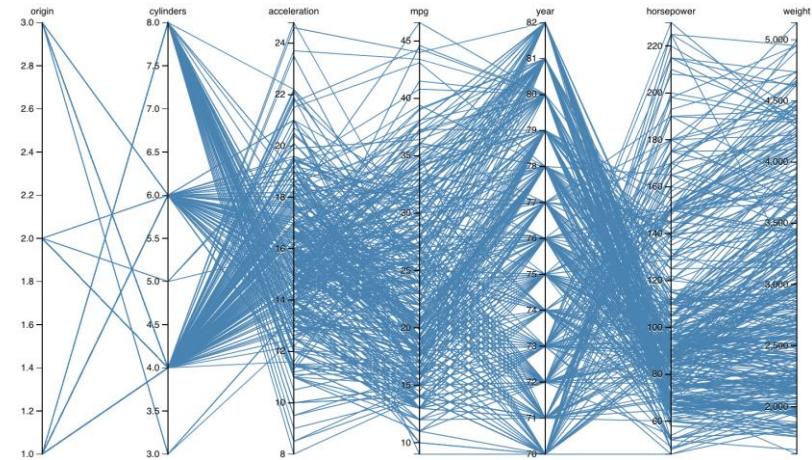


# PCP – Ordering (Similarity)

Default



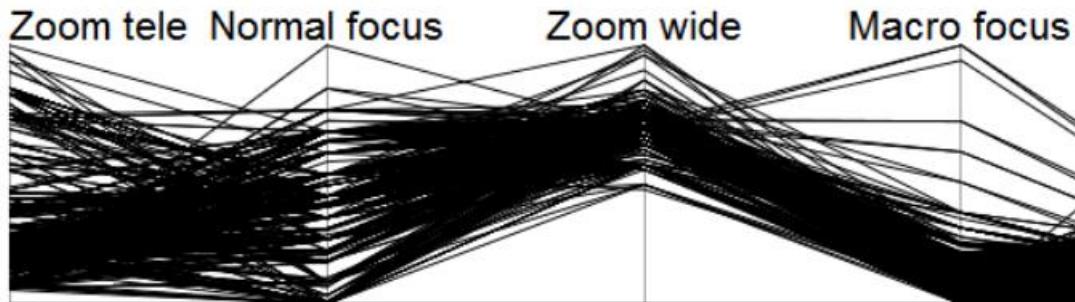
Similarity





# PCP – Ordering (Similarity)

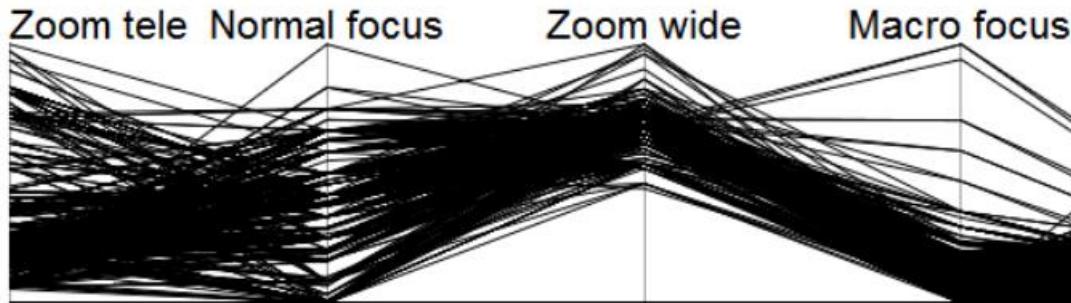
Task: Search for outlier



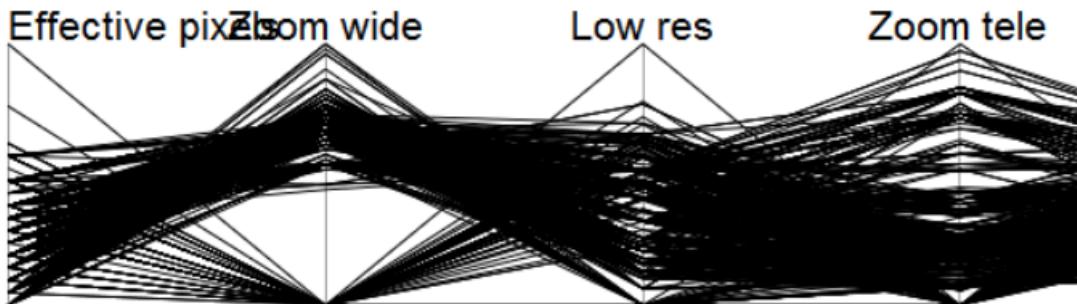


# PCP – Ordering (Similarity)

Task: Search for outlier



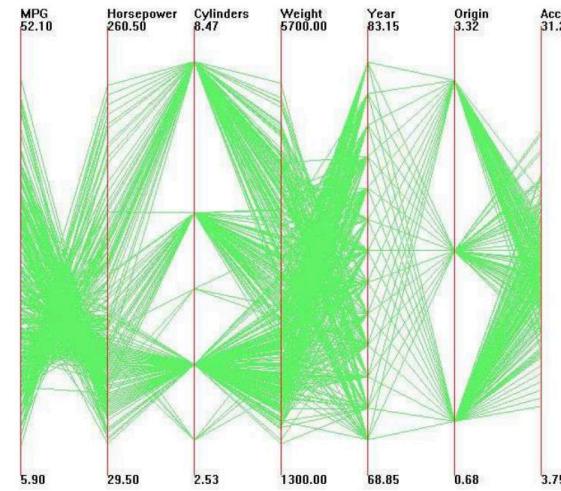
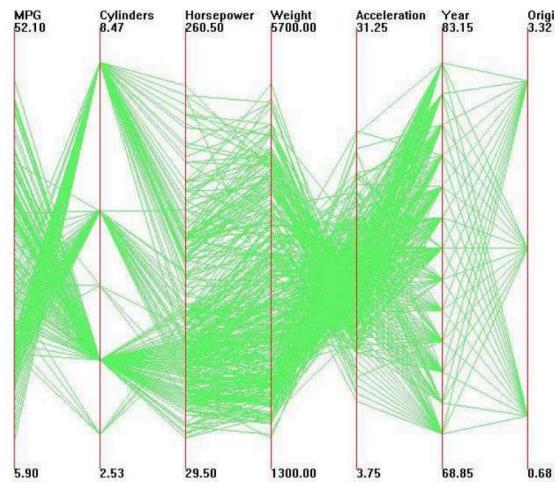
Same data but different ordering strategy





# PCP – Ordering (Similarity)

- Calculate the clutter in PCP
  - Detect outliers between neighboring dimensions using clustering algorithms.
  - Calculate ratio between outliers and data points.
  - Find ordering strategy, which minimizes number of outliers.



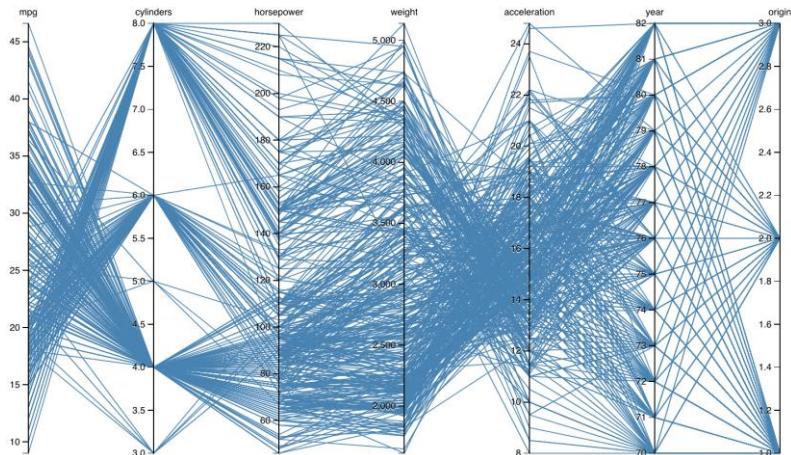
Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering

Wei Peng, Matthew O. Ward and Elke A. Rundensteiner: Information Visualization: 89-96, IEEE, 2004

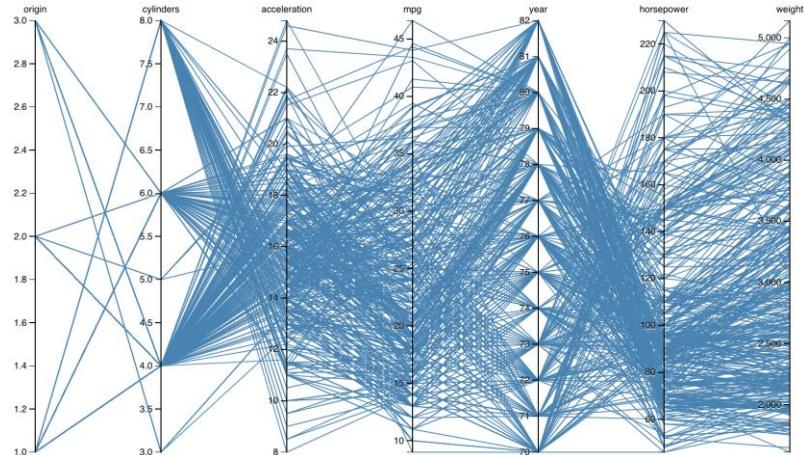


# PCP – Ordering (Similarity)

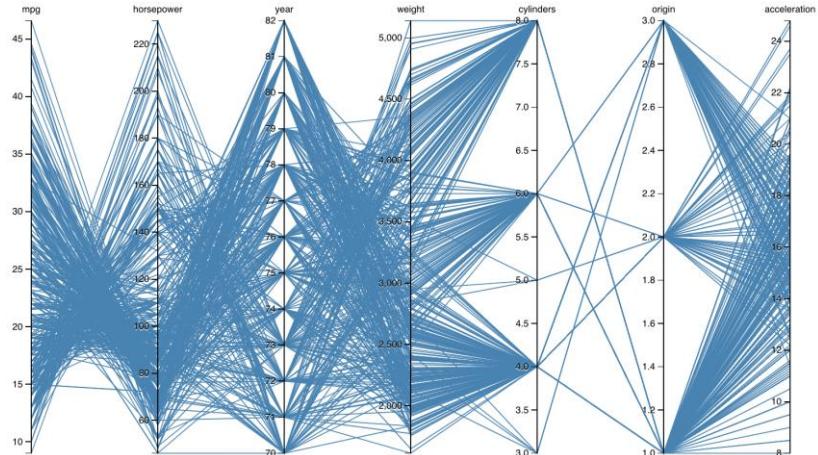
Default



Similarity



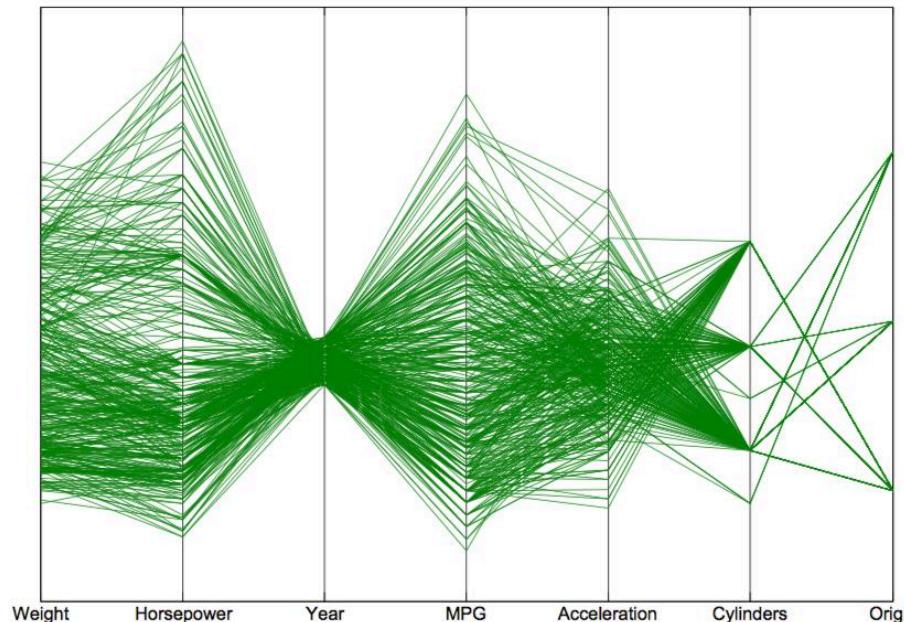
Clutter:





# PCP – Ordering (Importance)

- Researchers argue that the first dimension attracts the most attention of users.
- Order dimensions using singular values.
  - Compute the Singular Value Decomposition to measure the contribution of each dimension.

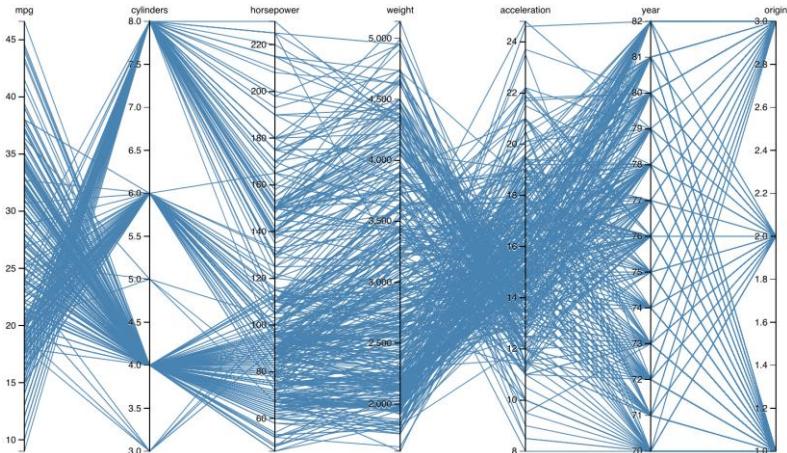


Two axes re-ordering methods in parallel coordinates plots

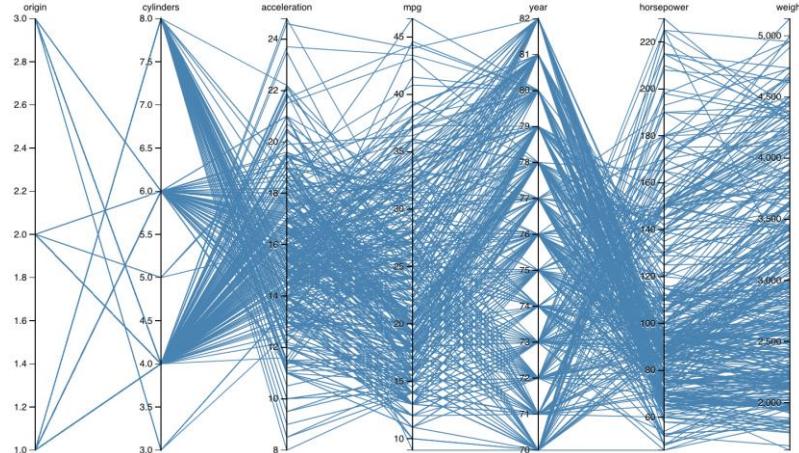
Huang, M and Lu, L and Zhang, J: Journal of Visual Languages, Elsevier, 2016

# PCP – Ordering (Comparisson)

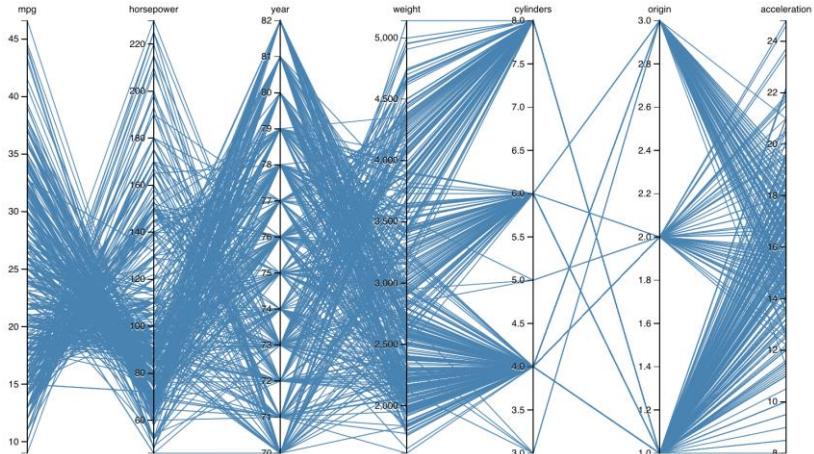
Default



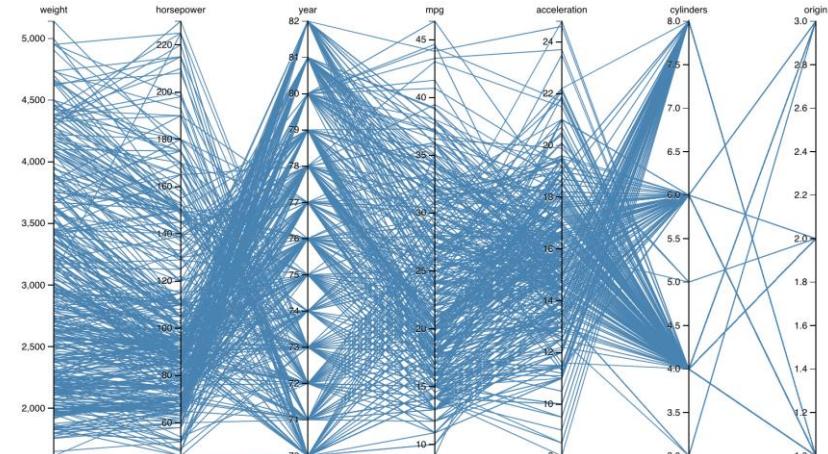
Similarity



Clutter:

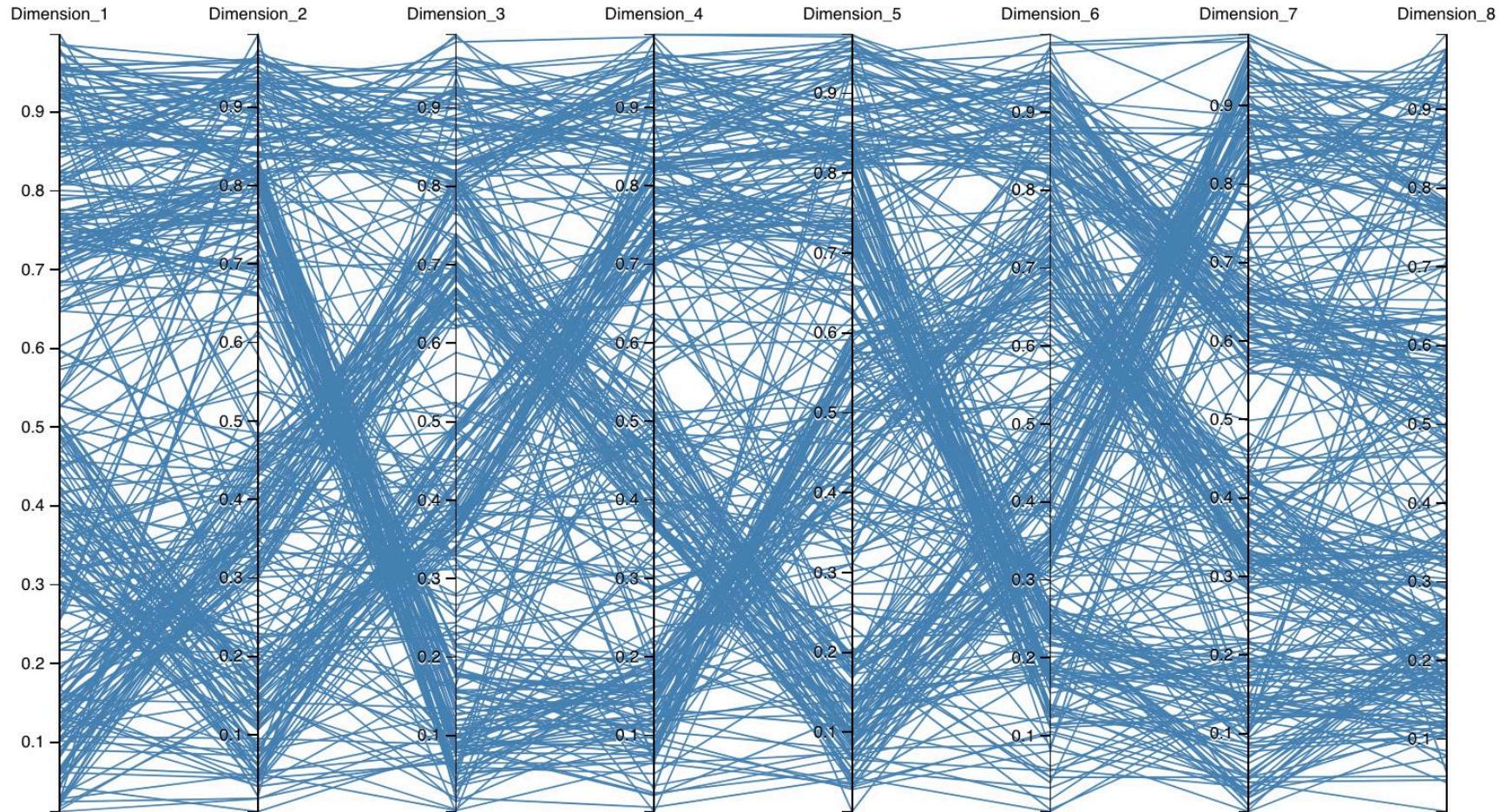


Contribution:



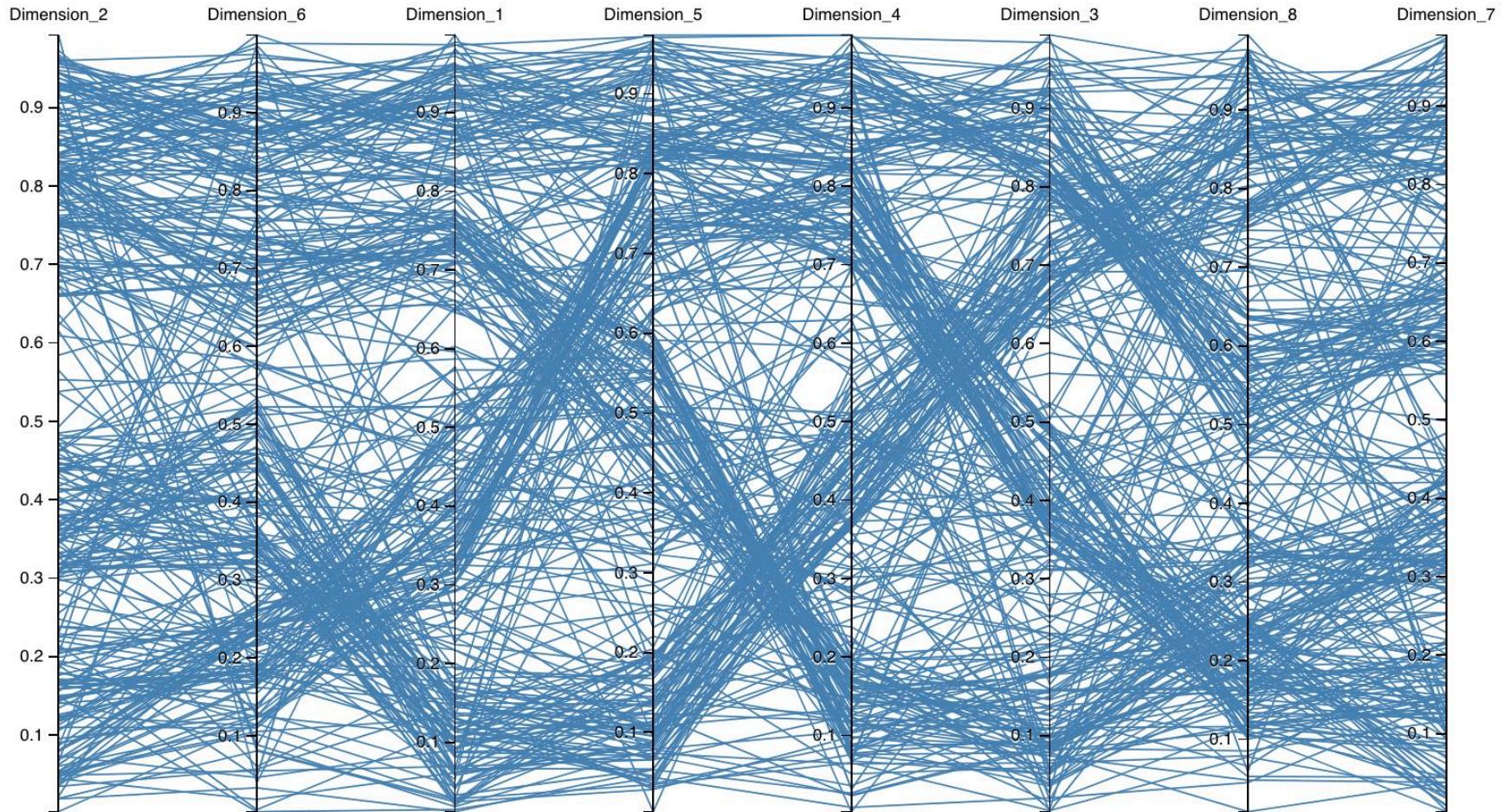


# PCP – Ordering (User Study)



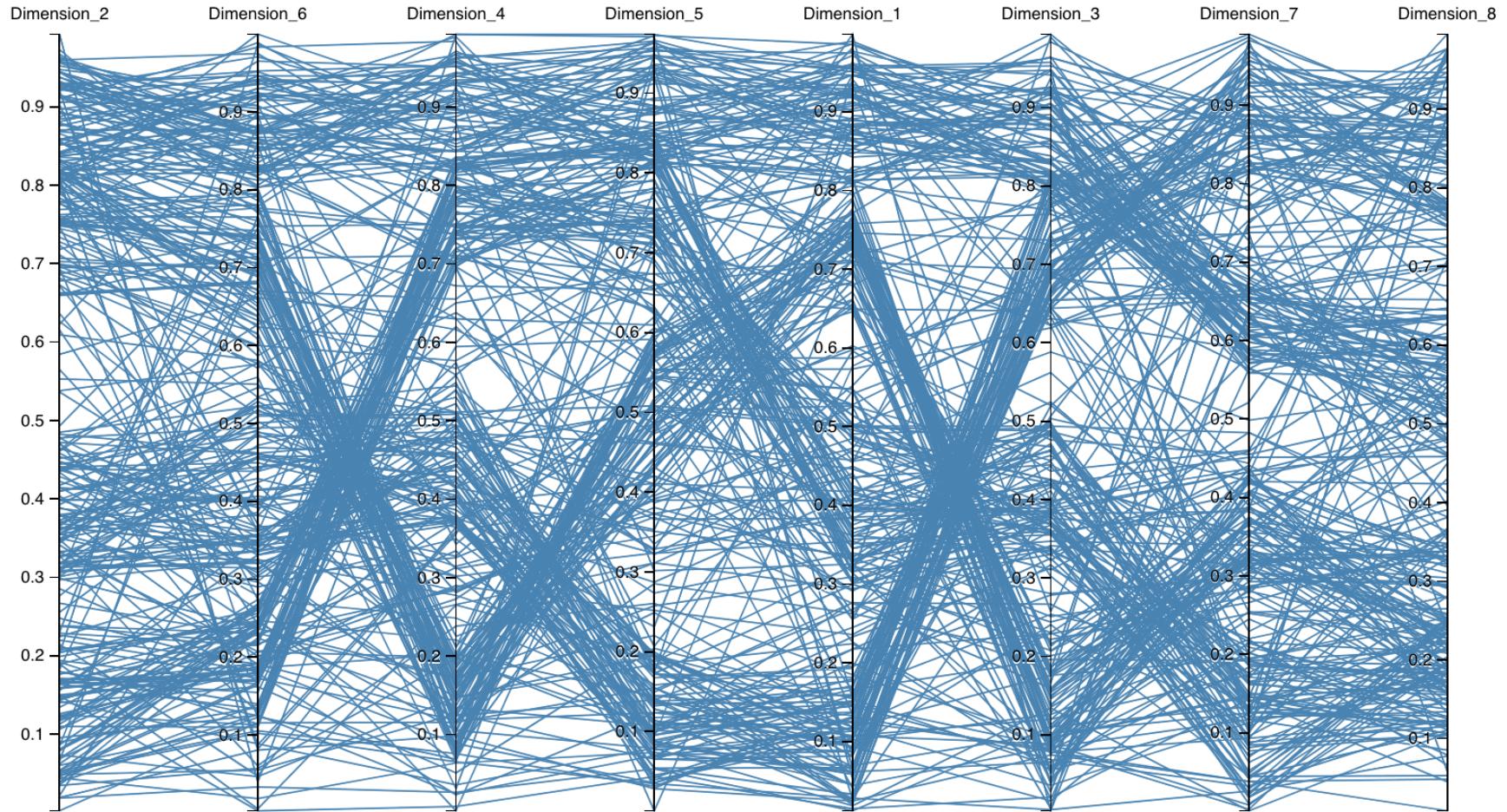


# PCP – Ordering (User Study)



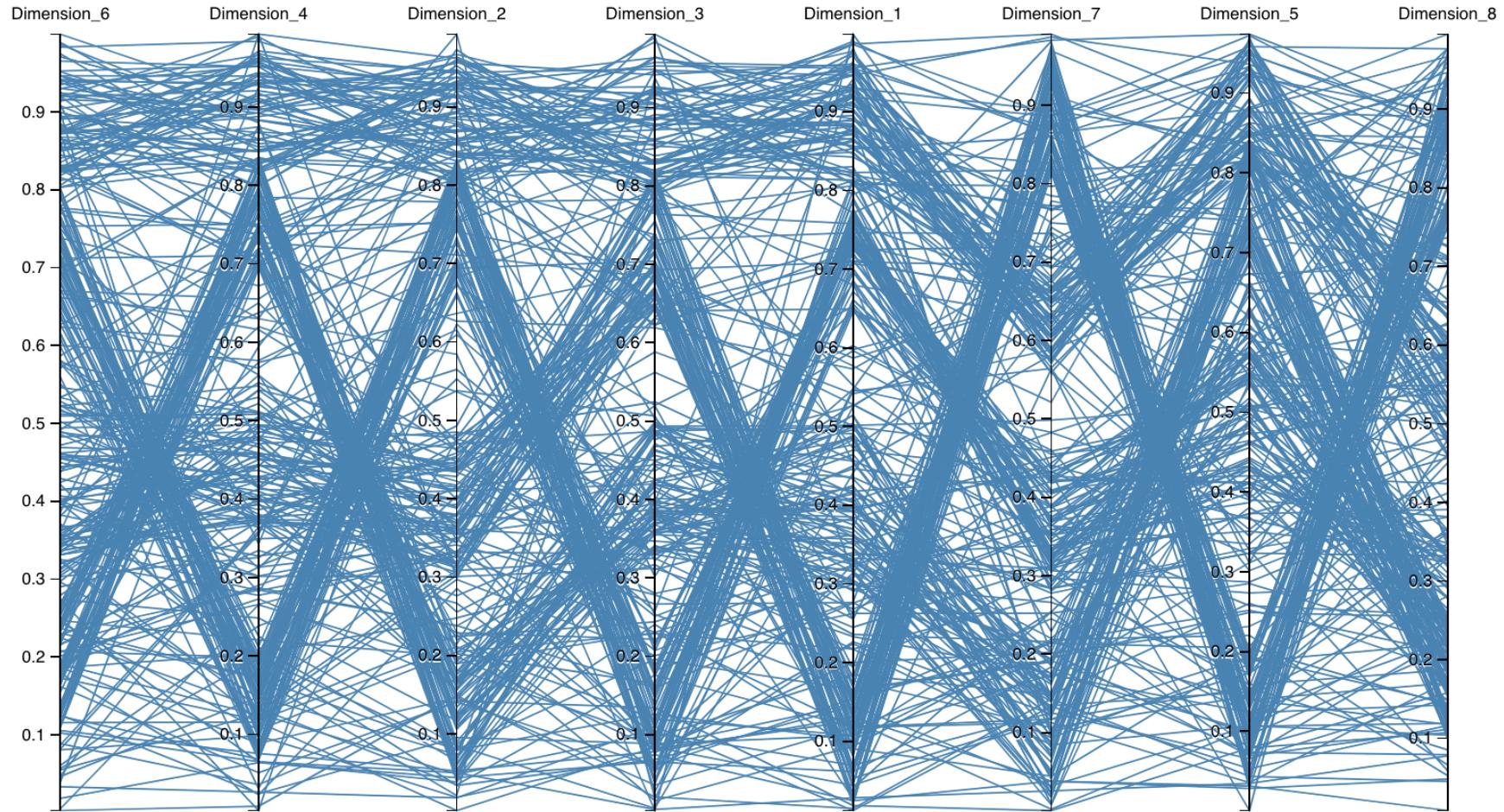


# PCP – Ordering (User Study)





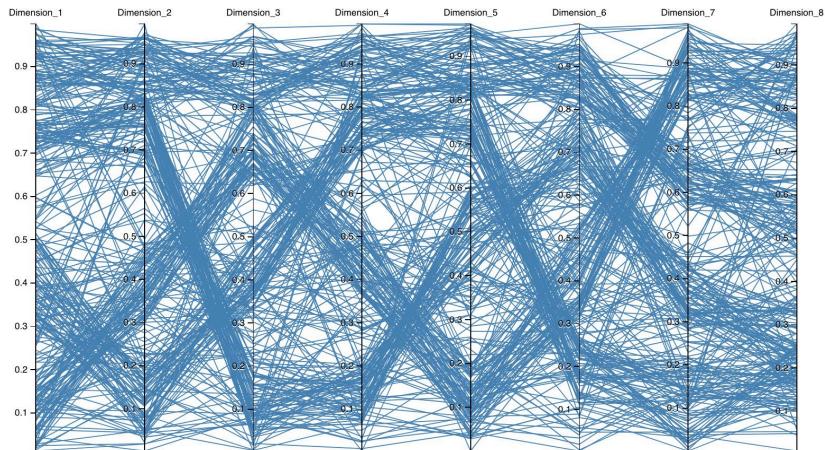
# PCP – Ordering (User Study)



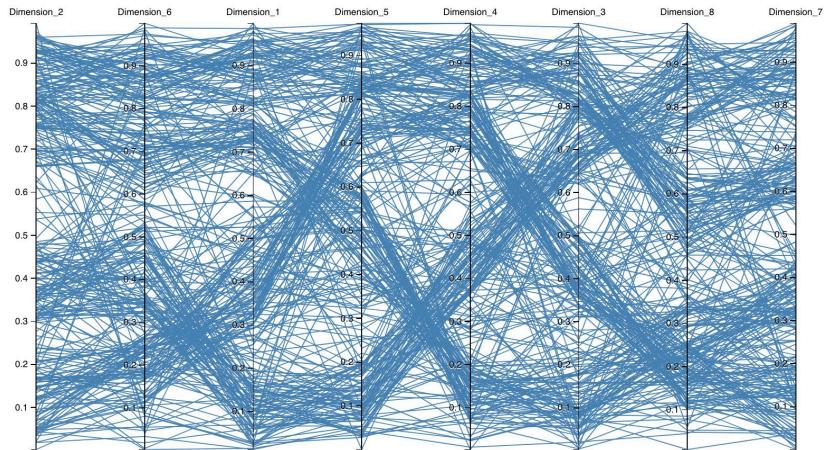


# PCP – Ordering (User Study)

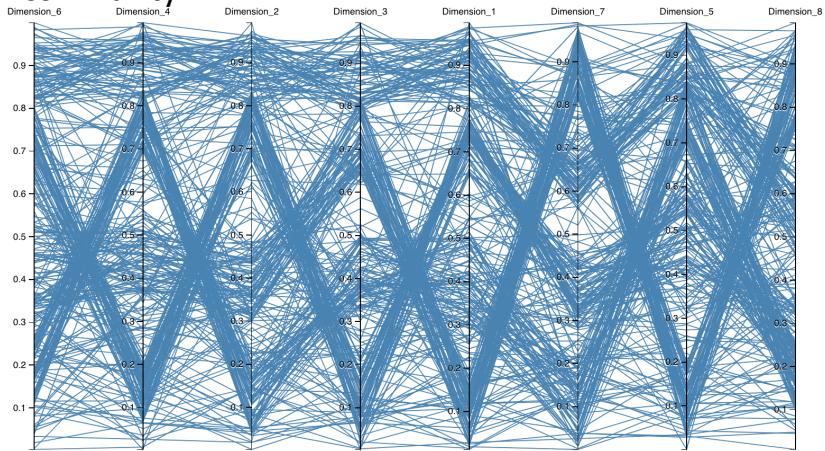
Default



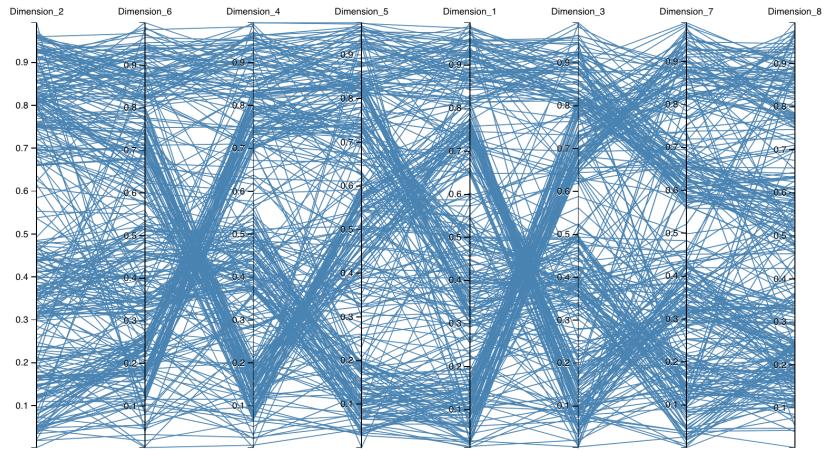
Similarity



Dissimilarity:



Contribution:





# PCP – Quality Metric

## Pargnostics

- Scagnostics for Parallel Coordinates
- Metrics to support different analysis tasks
  - Line crossings
  - Parallelism
  - Convergence-divergence (clusters)
  - ...
- Based on pixel space histograms (1d and 2d)
  - Really efficient
  - Can be weighted

Pargnostics: screen-space metrics for parallel coordinates  
Dasgupta, Aritra and Kosara, Robert: TVCG, 1017-1026, IEEE, 2010



# PCP – Quality Metric

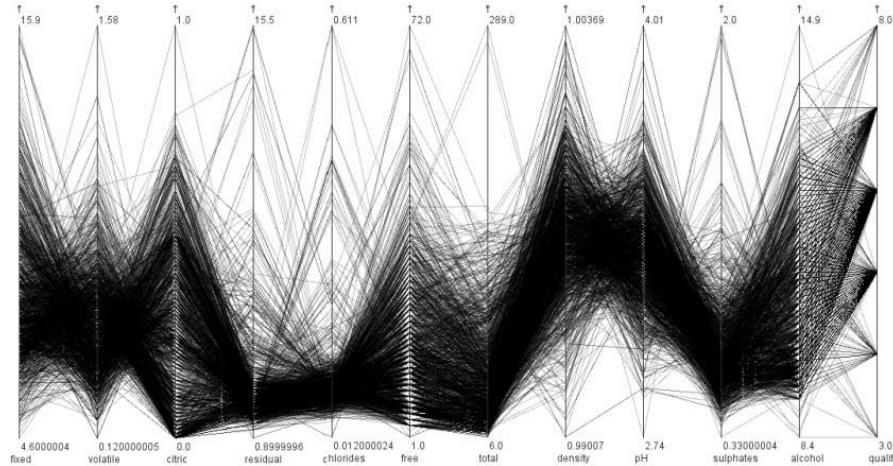
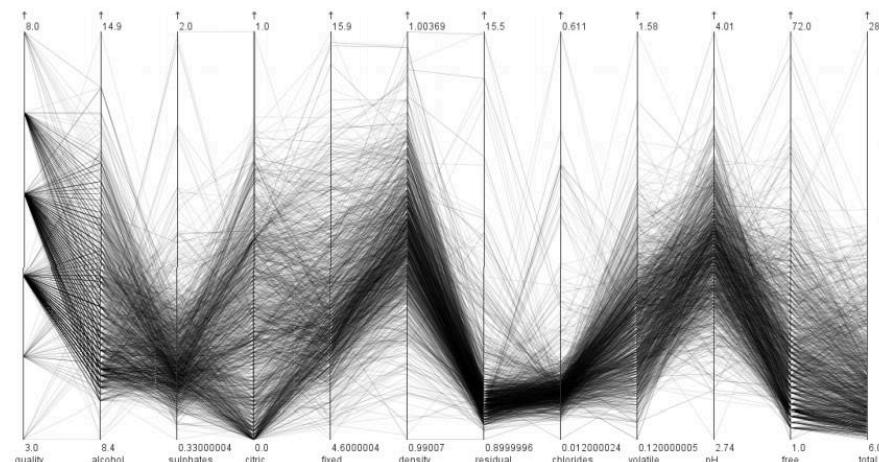
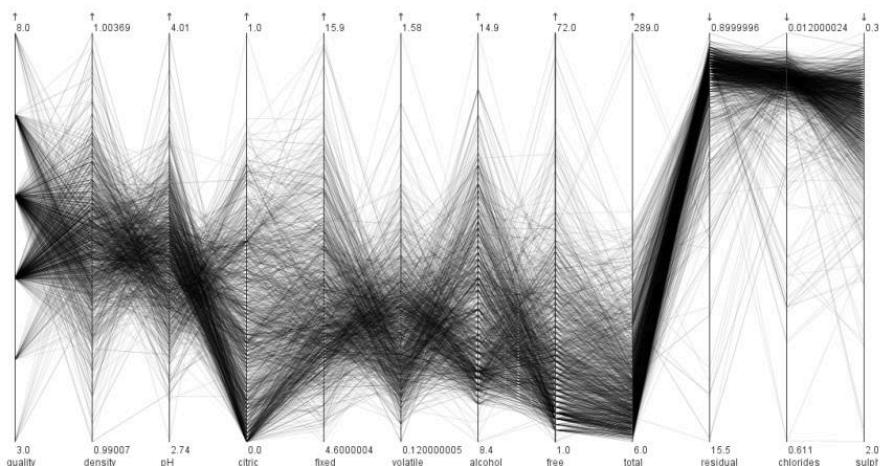
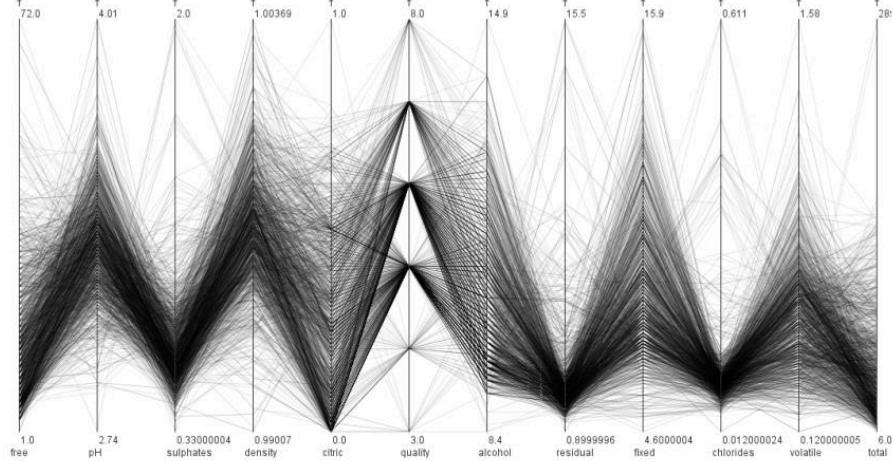


Fig. 12: Initial layout of the wine dataset.



(b) Minimizing the number of crossings.



(c) Minimized angles of crossing and maximum parallelism.

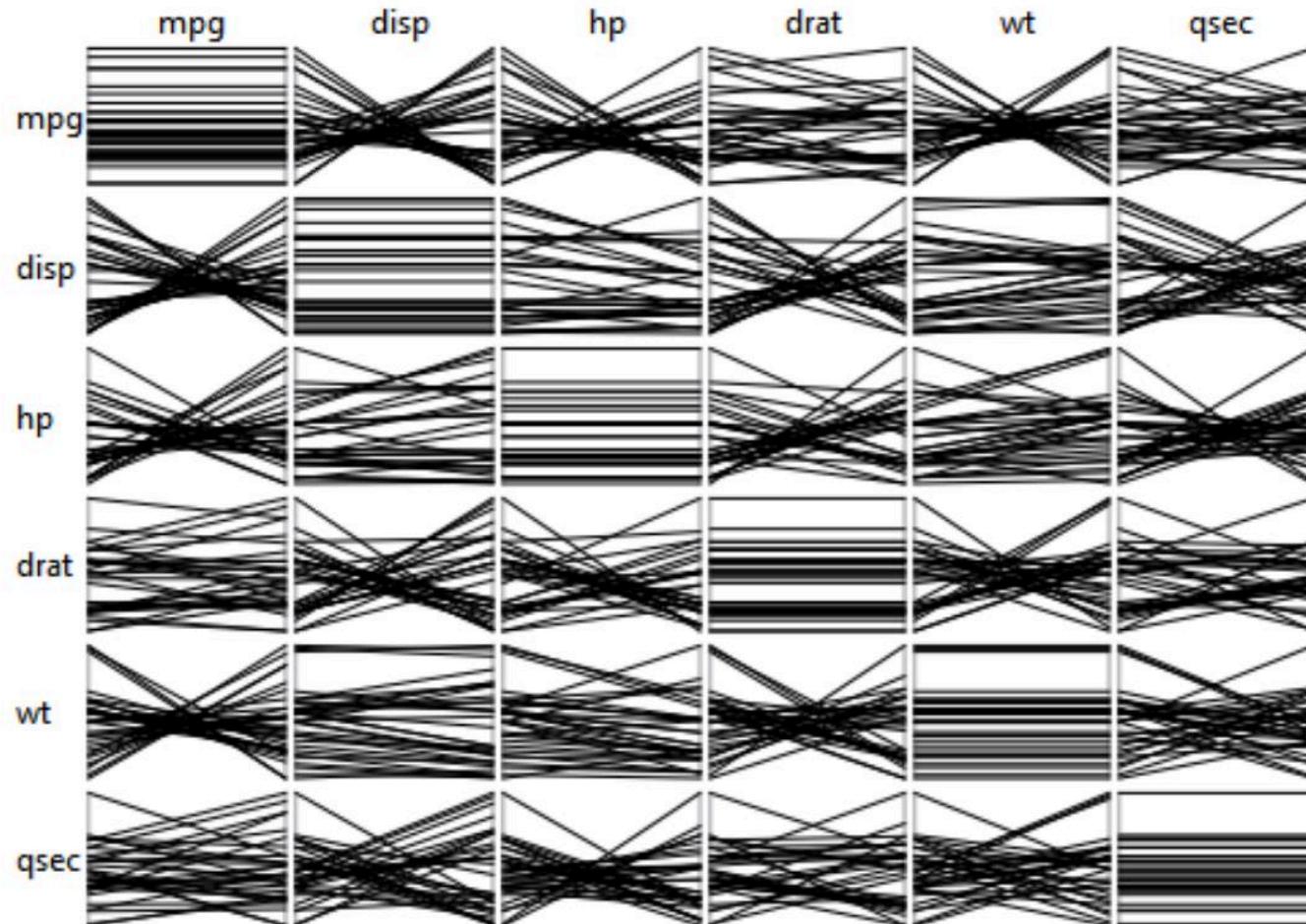


# Parallel Coordinates - Challenges

1. Increasing number of data points may result in vanishing patterns due to overplotting.
2. A perceived pattern depends primarily on the ordering of the dimensions.
3. A large number of dimensions decreases the available screen space between axes.
4. (Negative correlations are visible more clearly)



# PCP – Matrix (Example)

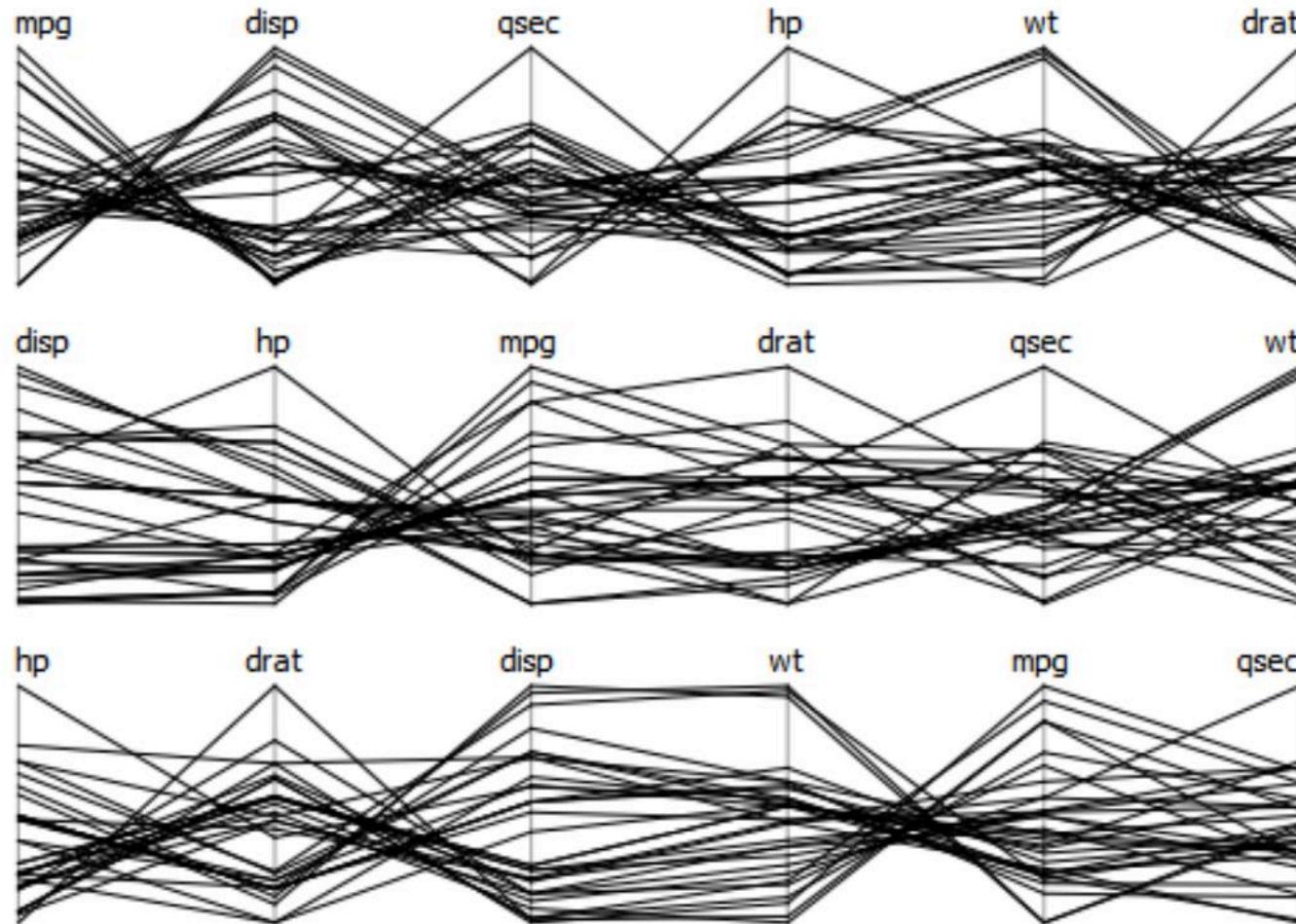


The parallel coordinates matrix

Heinrich, Julian and Stasko, John and Weiskopf, Daniel  
EuroVis--Short Papers. 2012



# PCP – Example



The parallel coordinates matrix

Heinrich, Julian and Stasko, John and Weiskopf, Daniel  
EuroVis--Short Papers. 2012

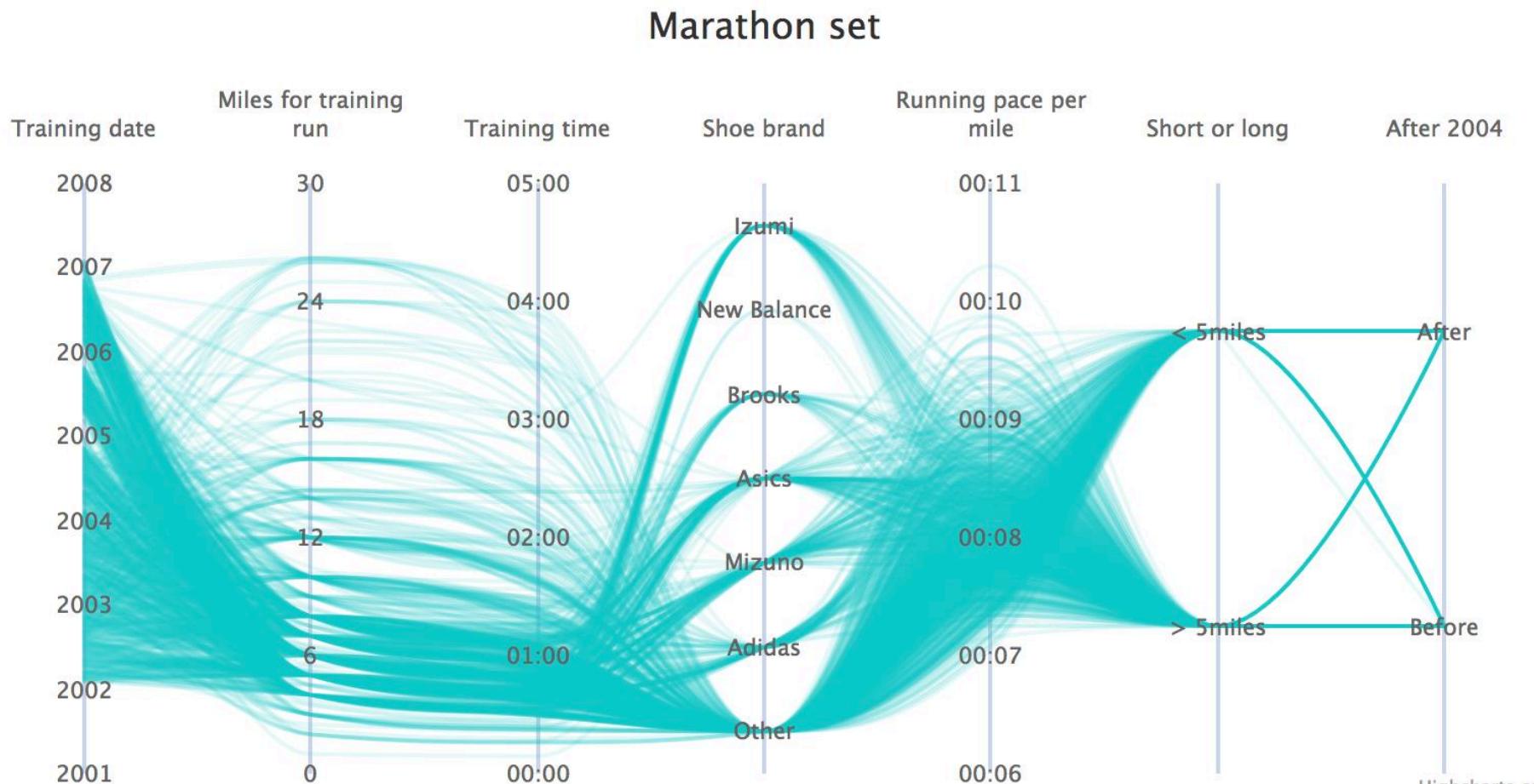


# Parallel Coordinates - Challenges

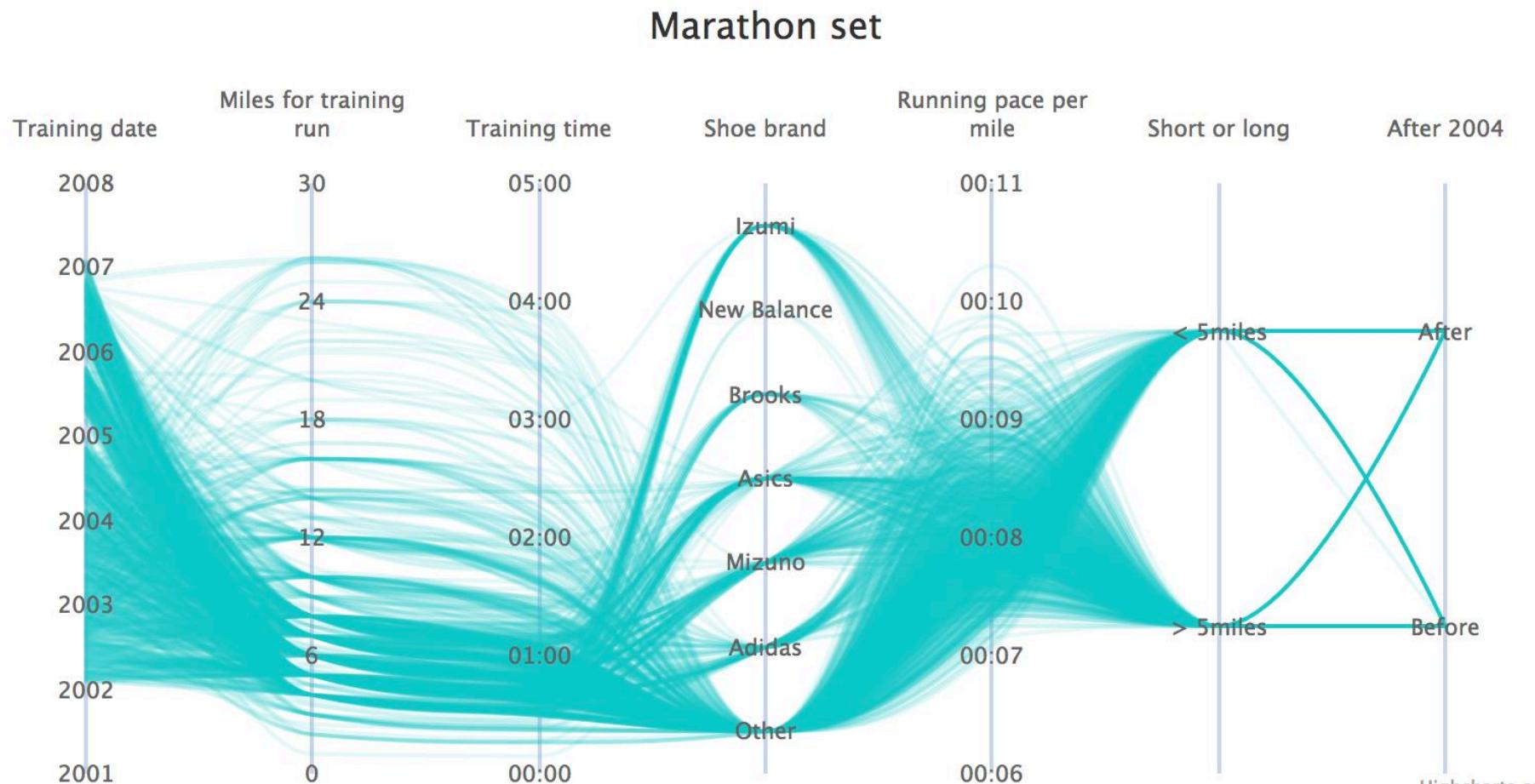
1. Increasing number of data points may result in vanishing patterns due to overplotting.
2. A perceived pattern depends primarily on the ordering of the dimensions.
3. A large number of dimensions decreases the available screen space between axes.
4. **(Negative correlations are visible more clearly)**



# PCP – Design Variations



# PCP – Design Variations

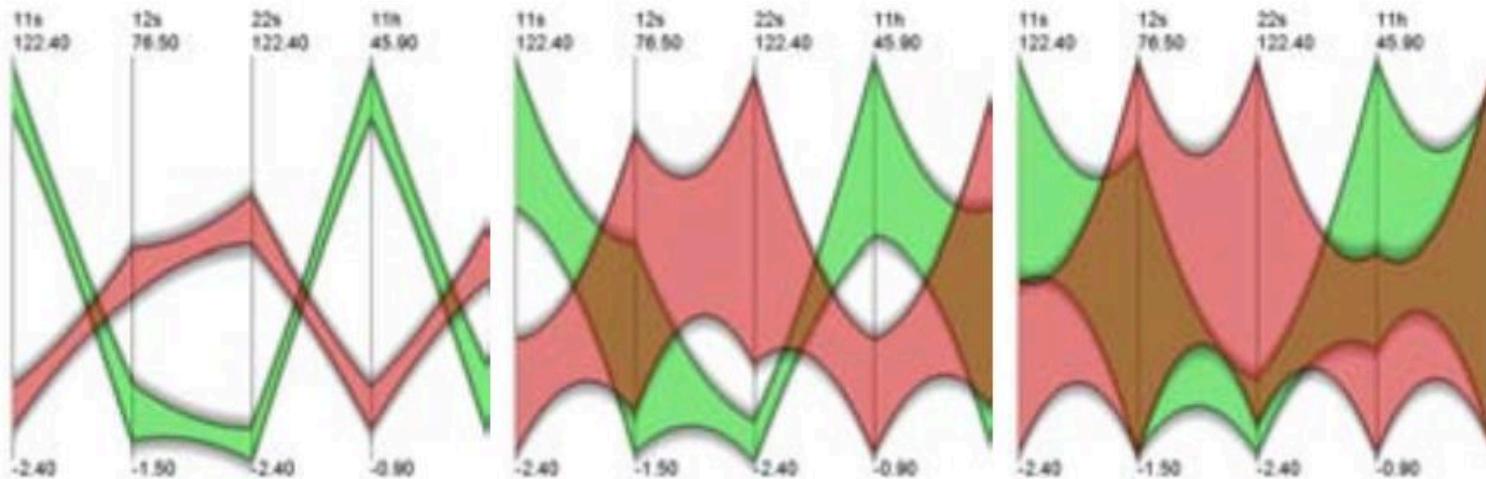


Gestalt Law: good continuation



# PCP – Design Variations

- Representative Polygons for clusters

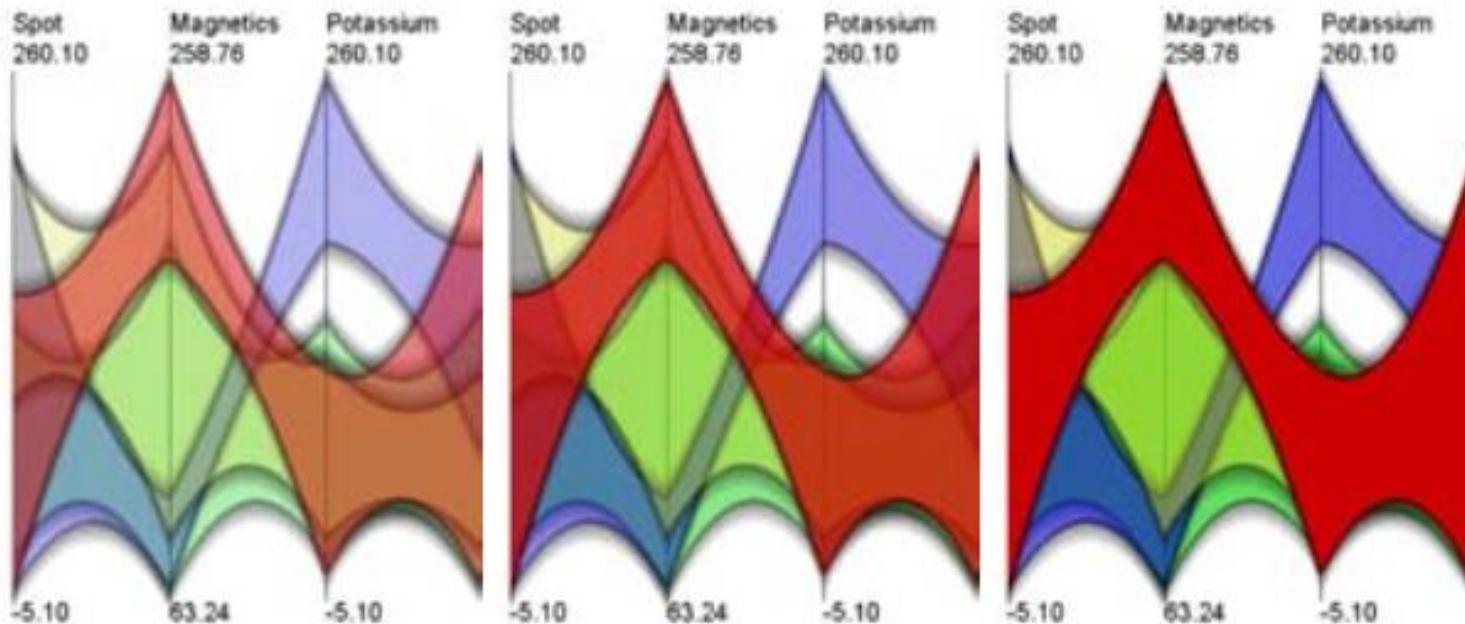


Illustrative parallel coordinates  
McDonnell, Kevin T and Mueller, Klaus  
Computer Graphics Forum 2008



# PCP – Design Variations

- Representative Polygons for clusters

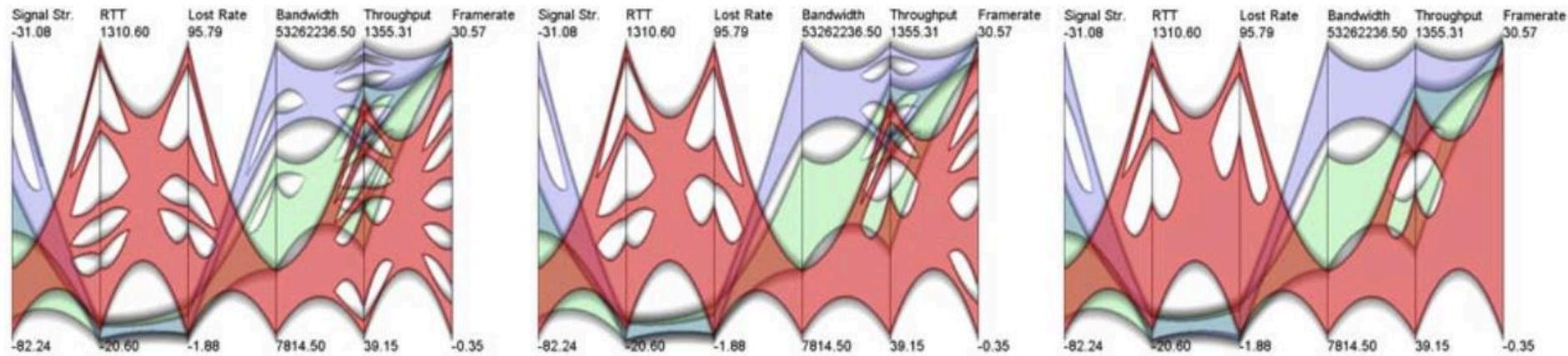


Illustrative parallel coordinates  
McDonnell, Kevin T and Mueller, Klaus  
Computer Graphics Forum 2008



# PCP – Design Variations

- Representative Polygons for clusters

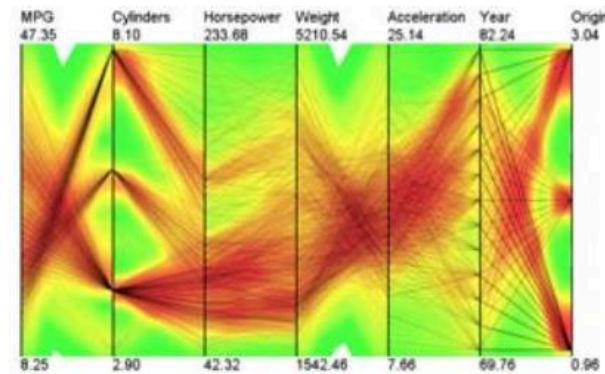
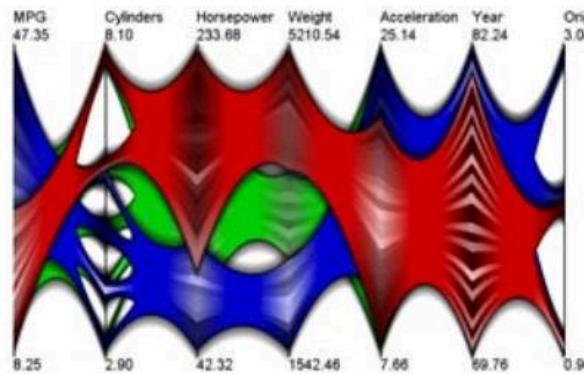
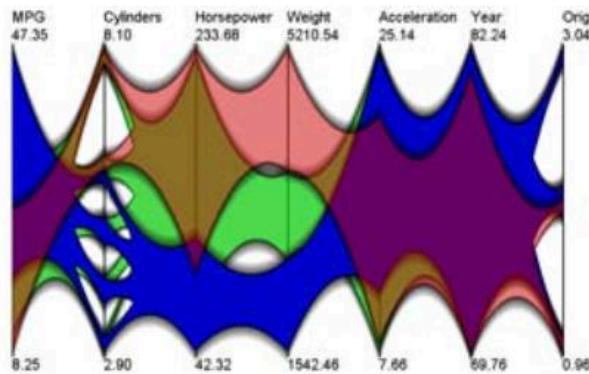


Illustrative parallel coordinates  
McDonnell, Kevin T and Mueller, Klaus  
Computer Graphics Forum 2008



# PCP – Design Variations

- Representative Polygons for clusters

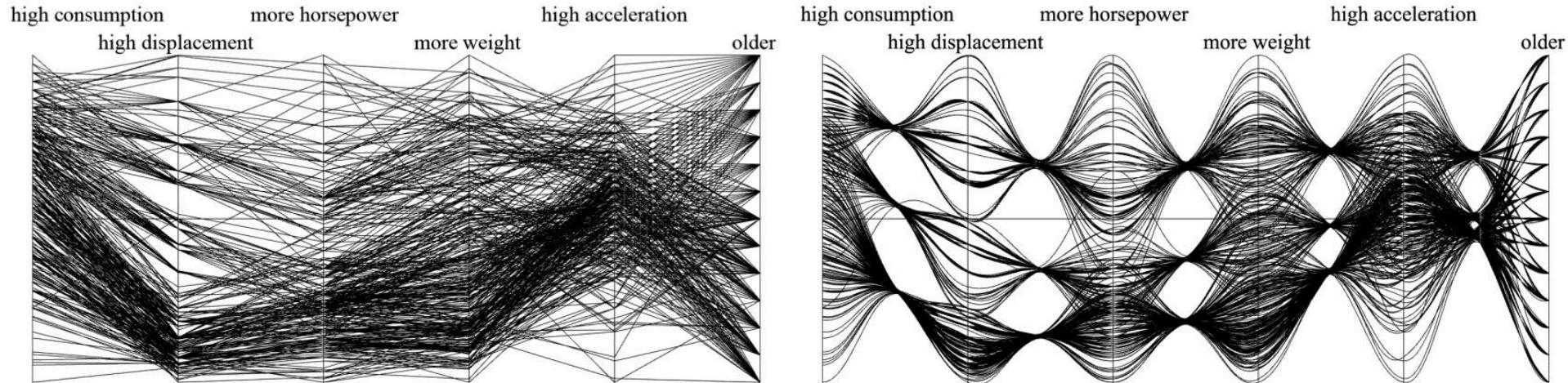


Illustrative parallel coordinates  
McDonnell, Kevin T and Mueller, Klaus  
Computer Graphics Forum 2008



# PCP – Design Variations

- Edge Bundling according to a pre-selected Attribute (e.g., cylinders (4, 6, 8))



Evaluation of a bundling technique for parallel coordinates  
Heinrich, Julian and Luo, Yuan and Kirkpatrick, Arthur E and Zhang, Hao and Weiskopf, Daniel  
Arxiv. 2011



# PCP– Summary

„The most important result is that a fair number of the seemingly valid improvements, [...] do not result in significant performance gains“

... for cluster identification tasks!

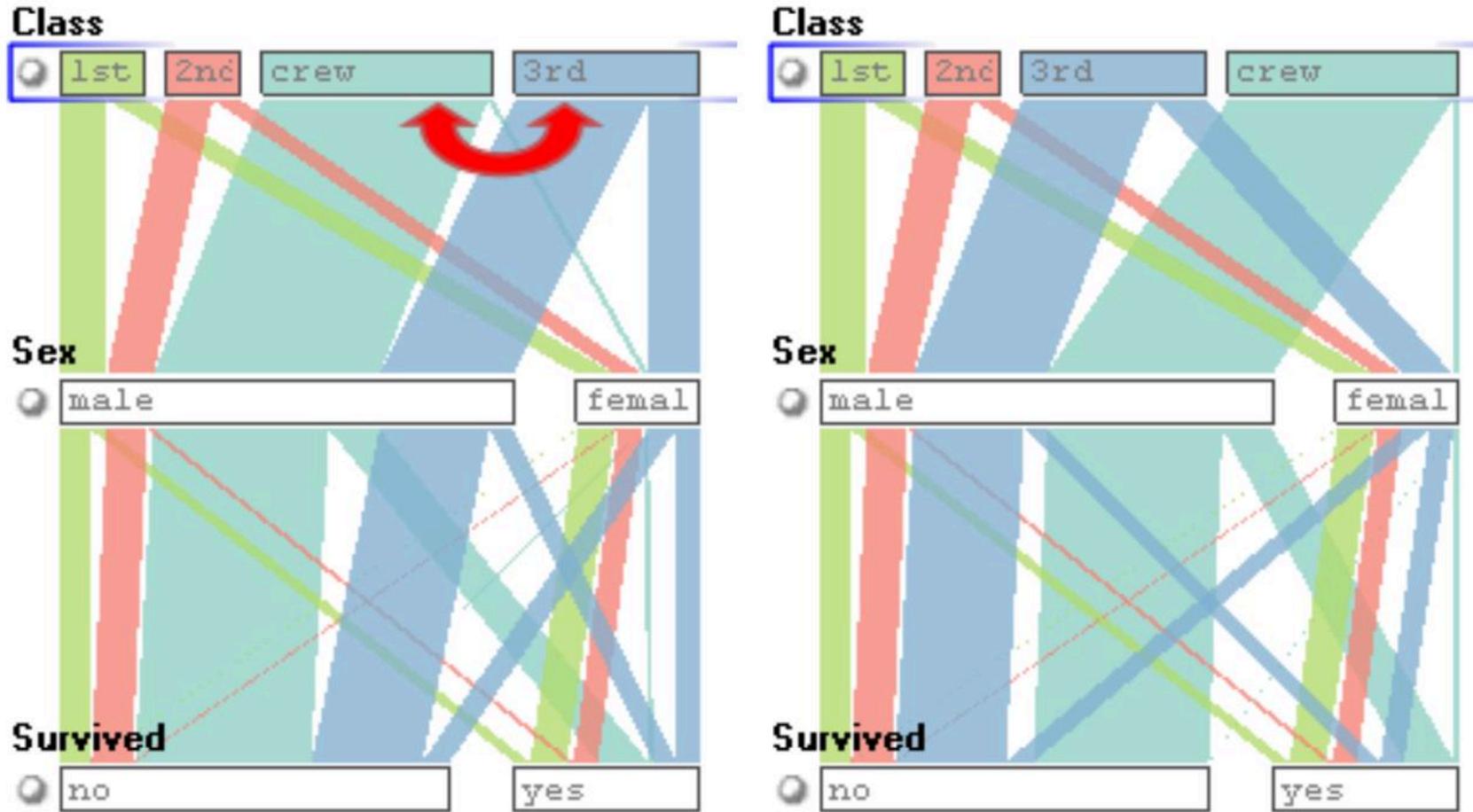
Evaluation of cluster identification performance for different PCP variants  
Holten, Danny and Van Wijk, Jarke J  
Computer Graphics Forum. 2010



# PCP - Problem

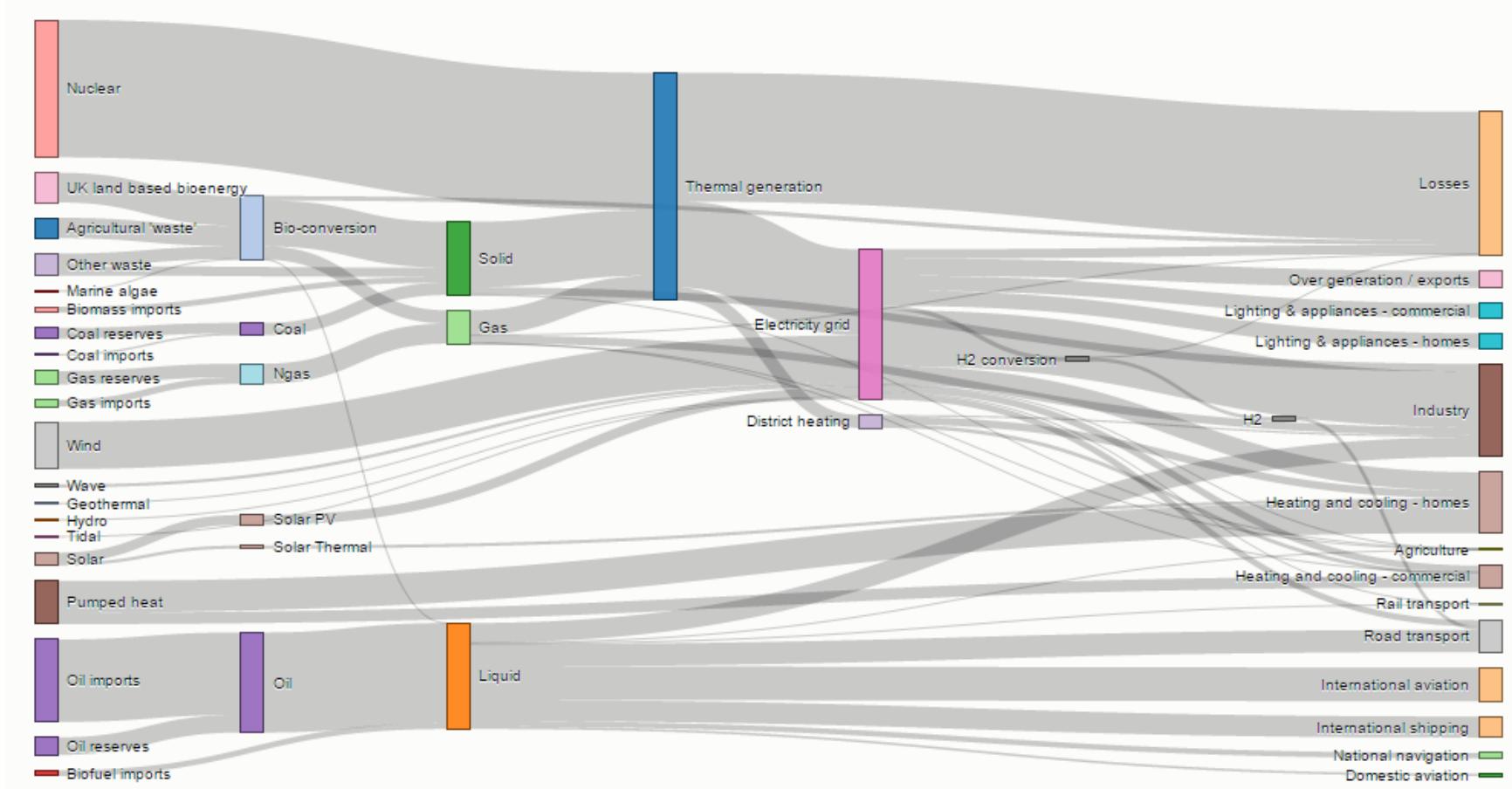
- What happens to a PCP when all (nearly all) dimensions are categorical or discrete?

# Parallel Coordinates – Variant



Parallel sets: Interactive exploration and visual analysis of categorical data  
Kosara, Robert and Bendix, Fabian and Hauser, Helwig: TVCG, 558-568, IEEE, 2006

# Parallel Coordinates – Variant



Sankey Diagrams



# PCP - Example

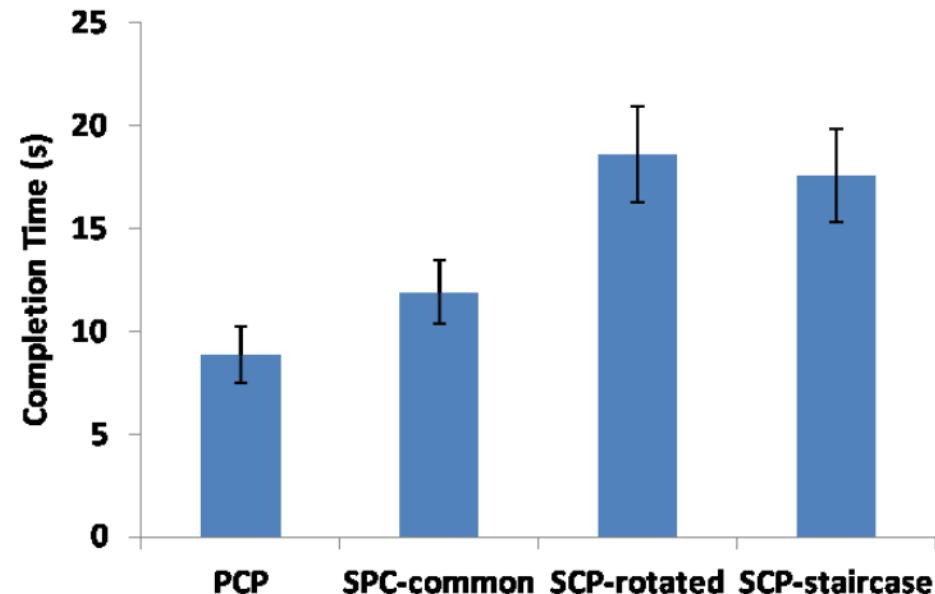
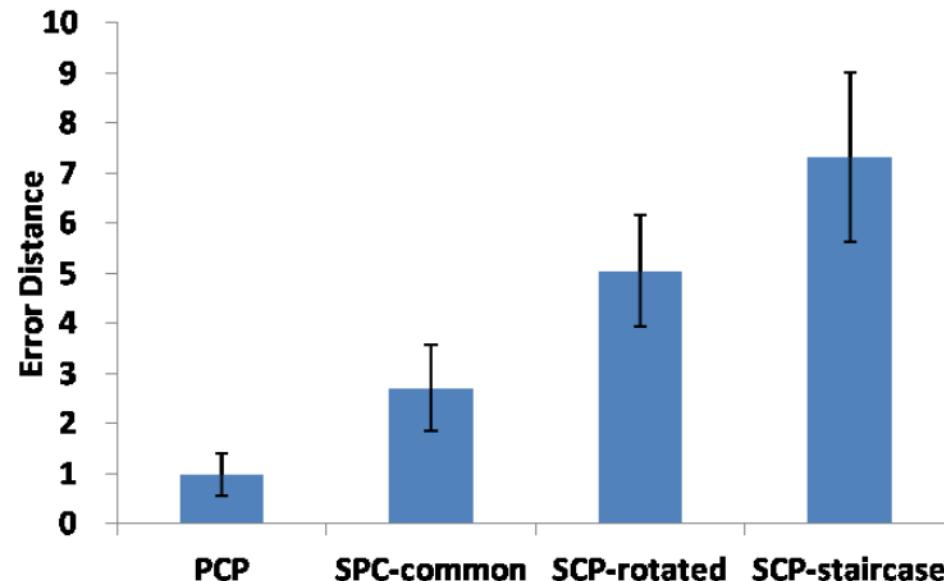
- <https://bl.ocks.org/jasondavies/1341281>
- Name interesting findings:
  - Relation cylinder and horsepower
  - ...



# PCP vs SPLOM

## Value retrieval:

Given multi-dimensional data: detect a specific value for a given data point.



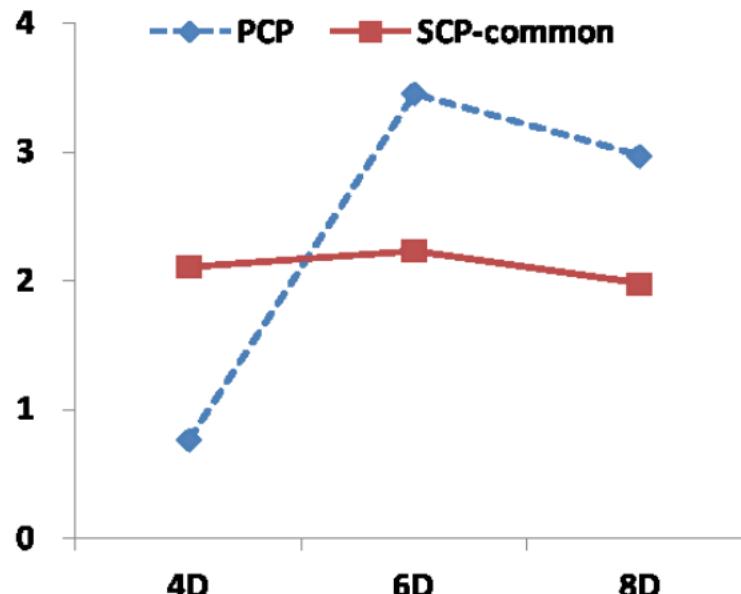
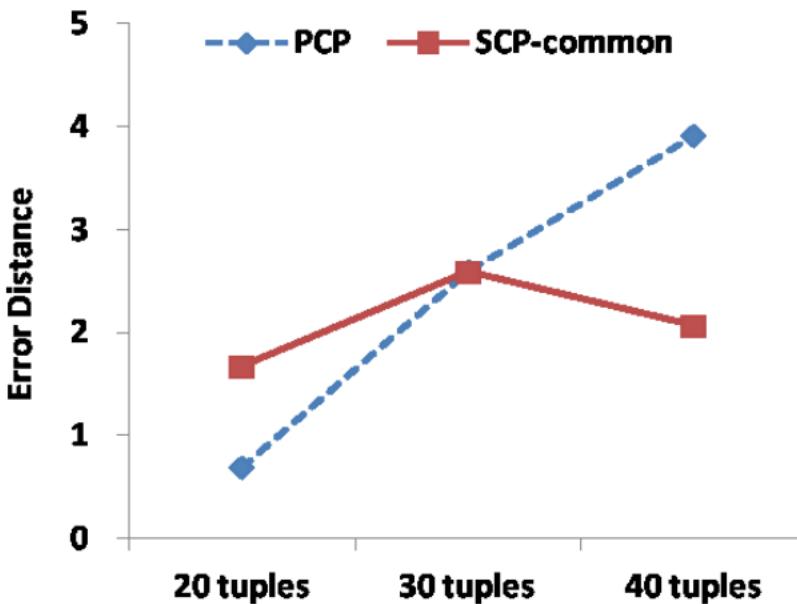
Tracing tuples across dimensions: A comparison of scatterplots and parallel coordinate plots  
Kuang, X and Zhang, H and Zhao, S and McGuffin, Michael J  
Computer Graphics Forum, 2012



# PCP vs SPLOM

## Value retrieval:

Given multi-dimensional data: detect a specific value for a given data point.



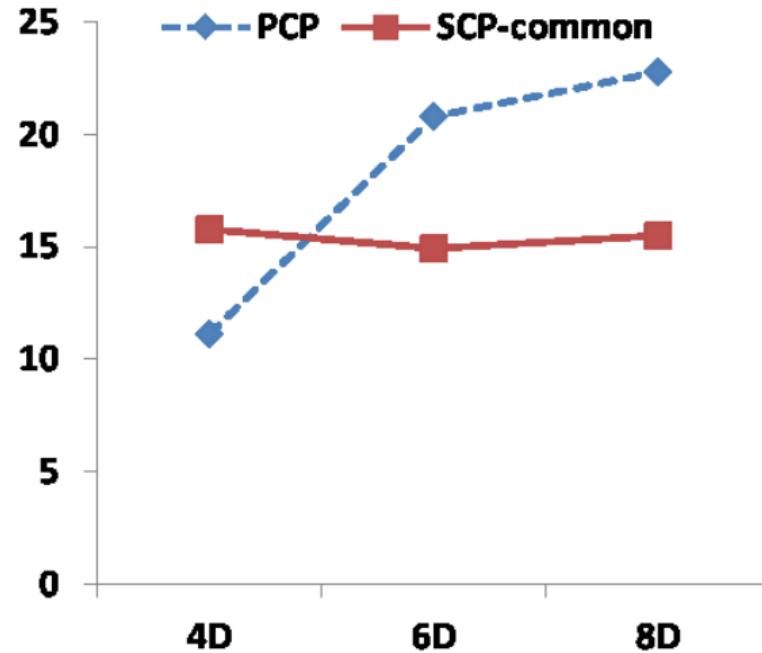
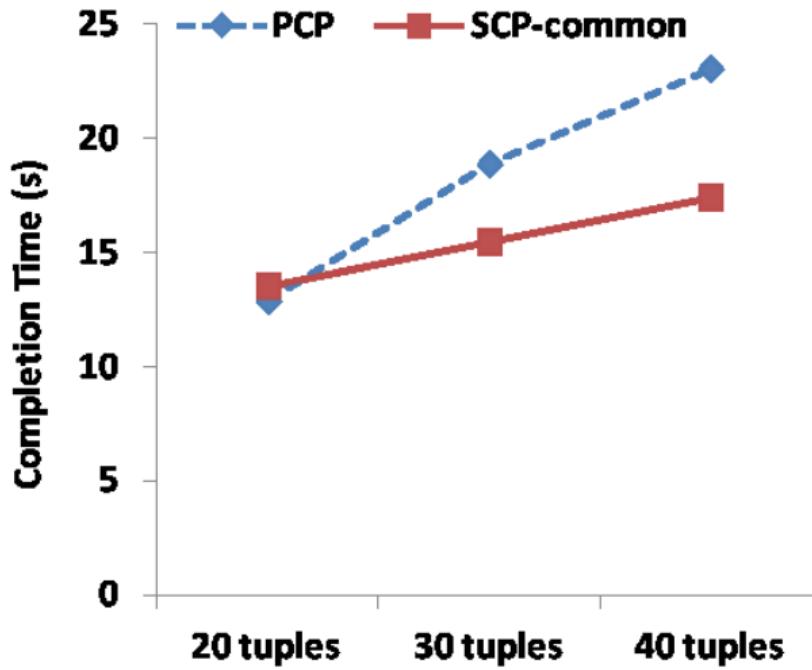
Tracing tuples across dimensions: A comparison of scatterplots and parallel coordinate plots  
Kuang, X and Zhang, H and Zhao, S and McGuffin, Michael J  
Computer Graphics Forum, 2012



# PCP vs SPLOM

## Value retrieval:

Given multi-dimensional data: detect a specific value for a given data point.



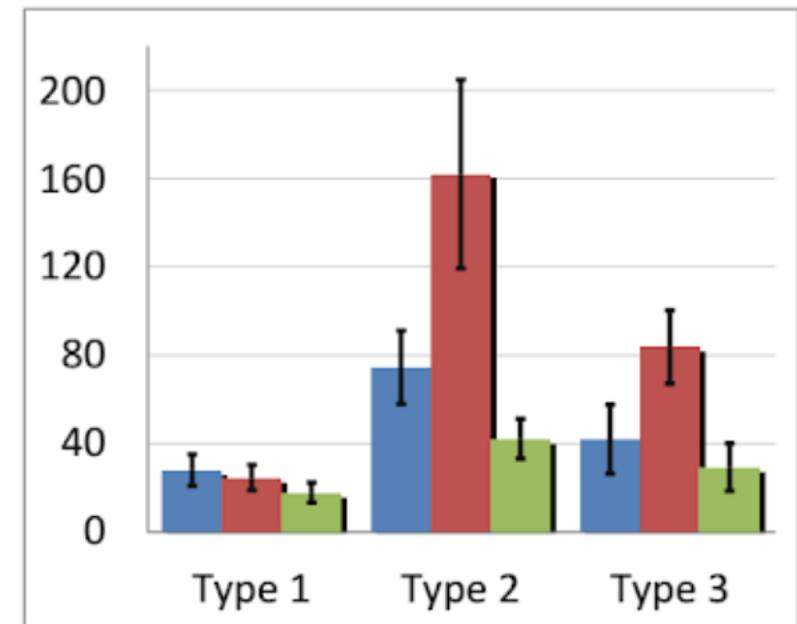
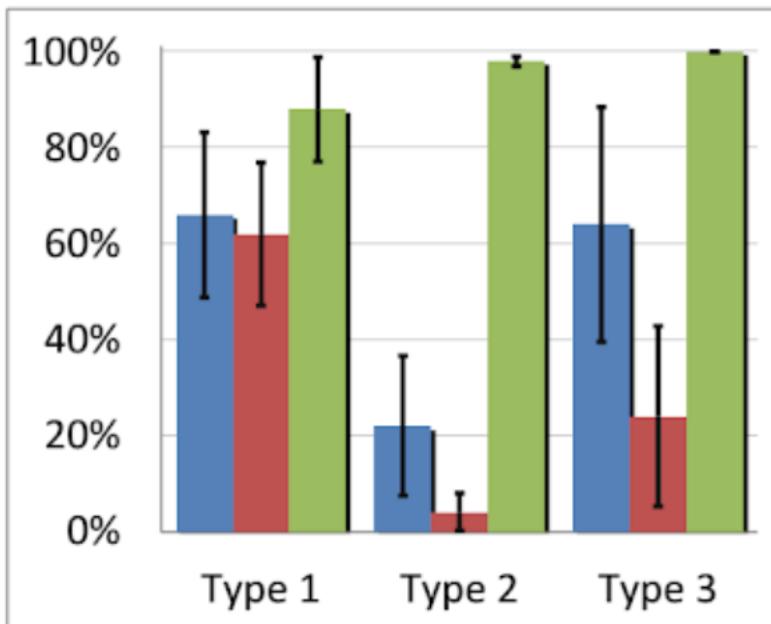
Tracing tuples across dimensions: A comparison of scatterplots and parallel coordinate plots  
Kuang, X and Zhang, H and Zhao, S and McGuffin, Michael J  
Computer Graphics Forum, 2012



# PCP vs SPLOM

## Correlation Analysis:

Detect correlations between any two dimensions



Improved Identification of Data Correlations through Correlation Coordinate Plots  
Nguyen, Hoa and Rosen, Paul  
VISIGRAPP, 2016



# Questions?