

Information Visualization and Visual Analytics (M1522.000500)

Multidimensional Data Visualization

Jinwook Seo, Ph. D.

Professor, Dept. of Computer Science and Engineering
Seoul National University

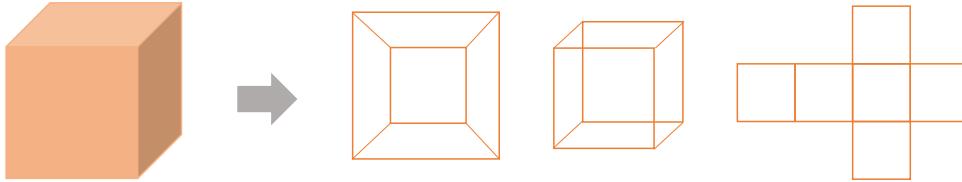
Multidimensional Data Visualizations

Multidimensional Data Visualizations

- Number of variables per data case (item)
 - 1 – Univariate data
 - 2 – Bivariate data
 - 3 – Trivariate data
 - > 3 – Multivariate (Hypervariate) data
- General representation techniques for multivariate (> 3 variables) per data case

Displaying Multidimensional Data in 2D Space

- Fundamentally, we have 2 geometric (position) display dimensions
- For data sets with > 2 variables, we must project data down to 2D
- Need visual mapping that locates each dimension into 2D plane
- Computer Graphics: 3D → 2D representation



Even a 3D cube cannot be exactly projected in 2D space!!

Approaches

- A spreadsheet already does that too
 - Each variable is positioned into a column
 - Data cases in rows
 - This is a projection (mapping)
- What else?

Multidimensional Data Visualizations

Multiple Views

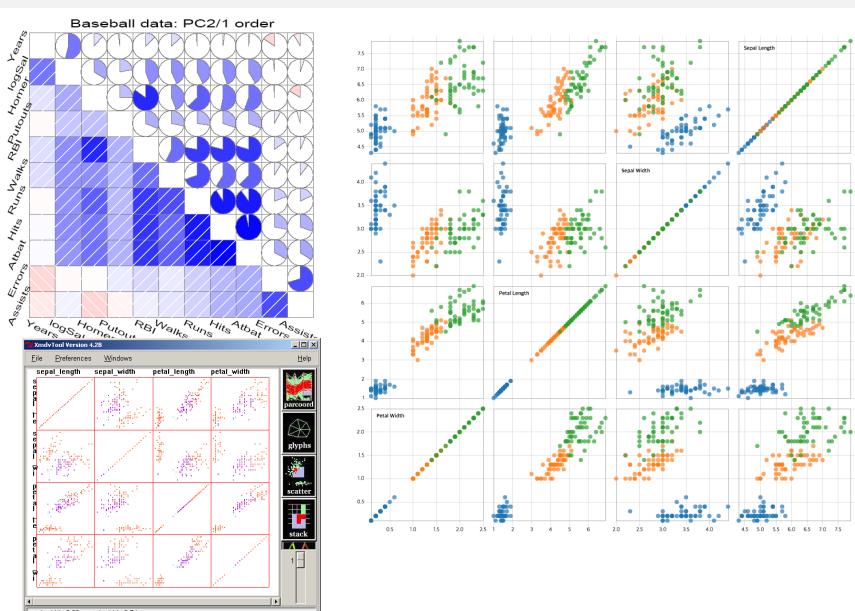
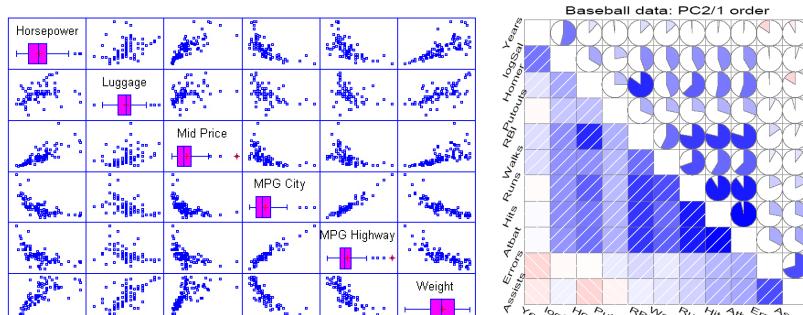


Jinwook Seo/John Stasko

Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional Data Visualizations

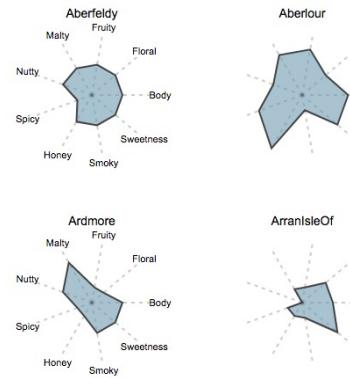
Scatterplot Matrix



Information Visualization and Visual Analytics – Multidimensional data Visualization

Glyphs

- “a graphical object designed to convey multiple data values”
- “composite graphical objects where different geometric and visual attributes are used to encode multidimensional data structures in combination”
- Chernoff Faces*: most prominent example
- Arrows → primitive example
 - length, width, angle, and color

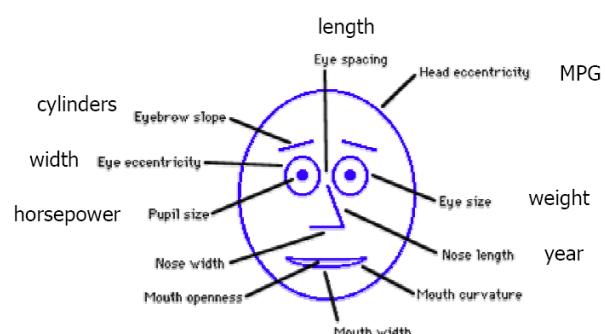
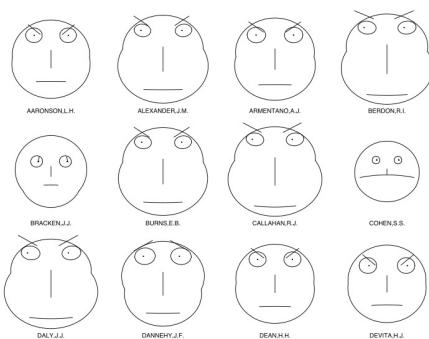


Information Visualization and Visual Analytics – Multidimensional data Visualization

Glyphs

Chernoff Faces

- Encode different variables' values in characteristics of human face
- Represent values of the variables by their shape, size, placement and orientation.

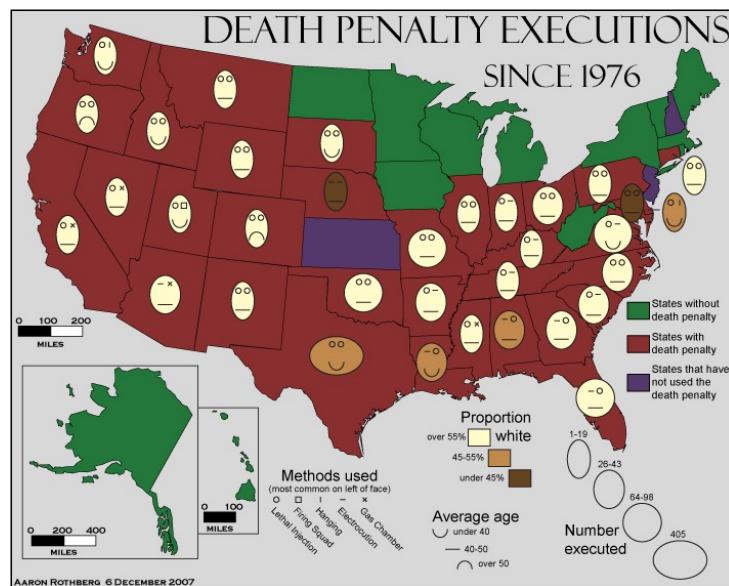


Herman Chernoff (1973). "The Use of Faces to Represent Points in K-Dimensional Space Graphically". Journal of the American Statistical Association 68 (342): 361–368

Information Visualization and Visual Analytics – Multidimensional data Visualization

Glyphs

Chernoff Faces

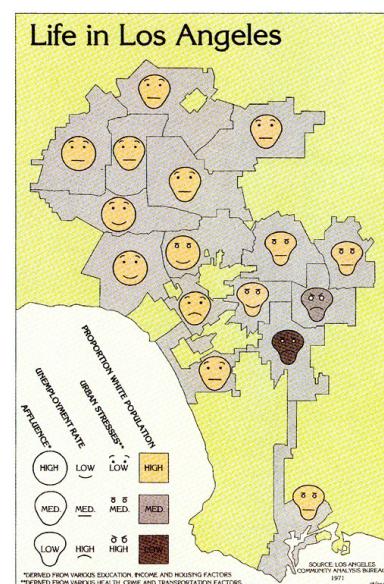


Information Visualization and Visual Analytics – Multidimensional data Visualization

Glyphs

Chernoff Faces

- “Humans easily recognize faces and notice small changes without difficulty” ?
- eye size and eyebrow-slant

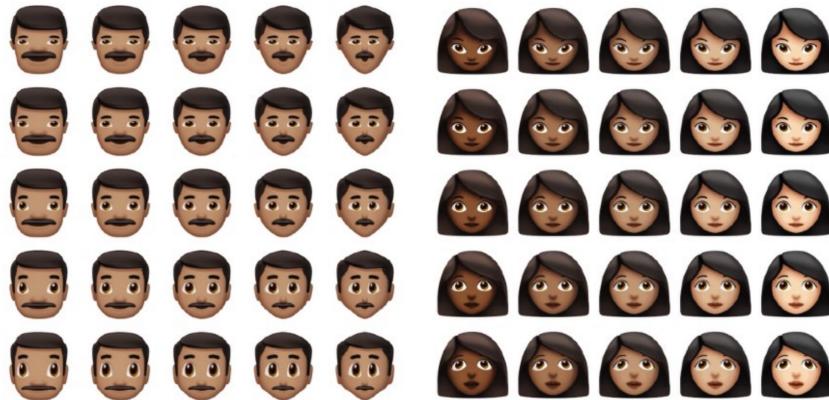


Information Visualization and Visual Analytics – Multidimensional data Visualization

Glyphs

Chernoff Faces

- Combined with Apple Memoji & Generative model



Bisig, D. "2D Generative Faces for Evolutionary Social Simulation."

Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional Data Visualizations

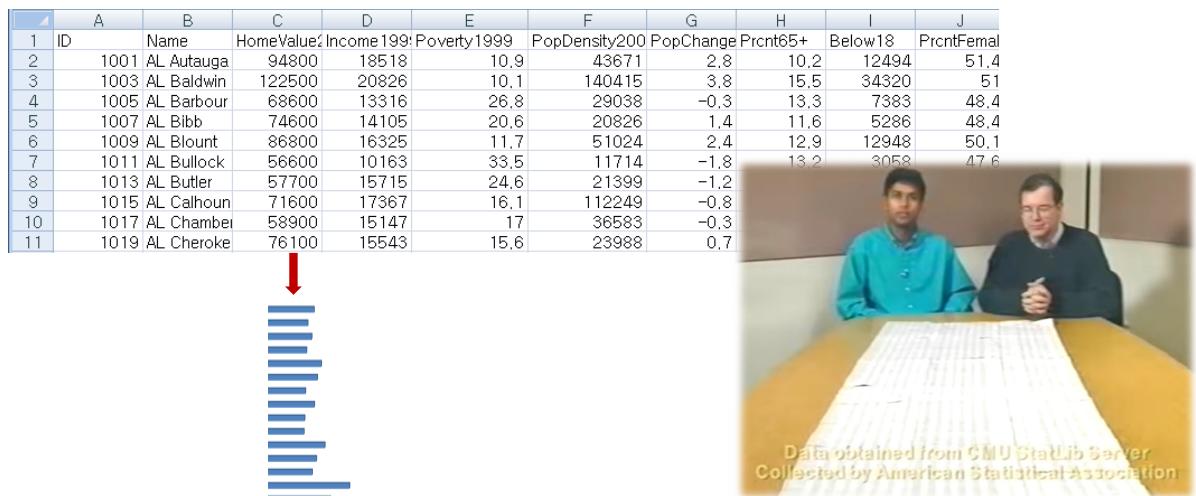
The Tables Lens

- Spreadsheet is certainly a multidimensional data presentation
- Idea: Make the text more **visual** and **symbolic**
- Just leverage basic **bar chart** idea

Information Visualization and Visual Analytics – Multidimensional data Visualization

Encoding

- Change Quantitative values to bars



Information Visualization and Visual Analytics – Multidimensional data Visualization

What do you do for nominal data?

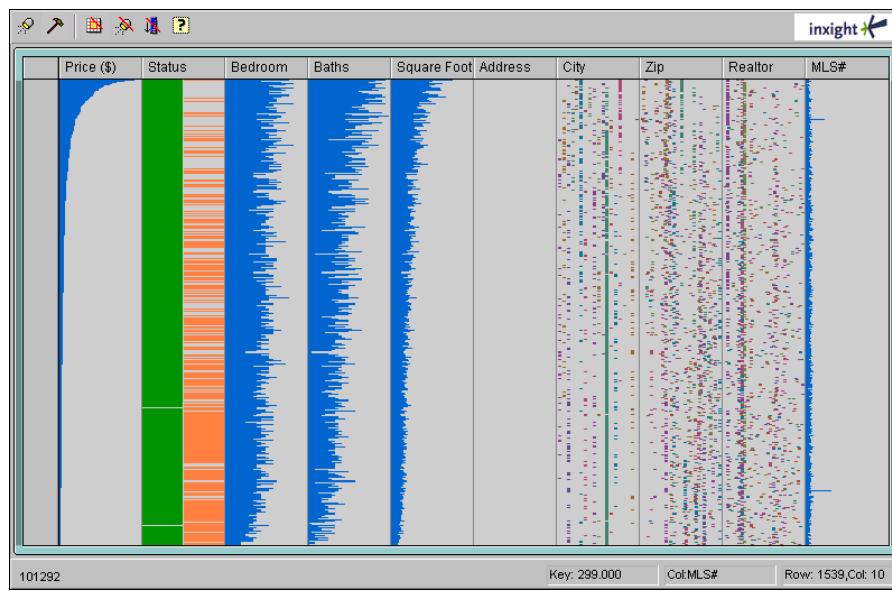
	A	B	C	D	E	F
1	cereal name	manufacture	type	calories	protein(g)	fat(g)
2	100%_Bran	N	C	70	4	1
3	100%_Natural_Bran	Q	C	120	3	5
4	All-Bran	K	C	70	4	1
5	All-Bran_with_Extra_Fiber	K	C	50	4	0
6	Almond_Delight	R	C	110	2	2
7	Apple_Cinnamon_Cheerios	G	C	110	2	2
8	Apple_Jacks	K	C	110	2	0
9	Basic_4	G	C	130	3	2
10	Bran_Chex	R	C	90	2	1
11	Bran_Flakes	P	C	90	3	0



Information Visualization and Visual Analytics – Multidimensional data Visualization

The Table Lens

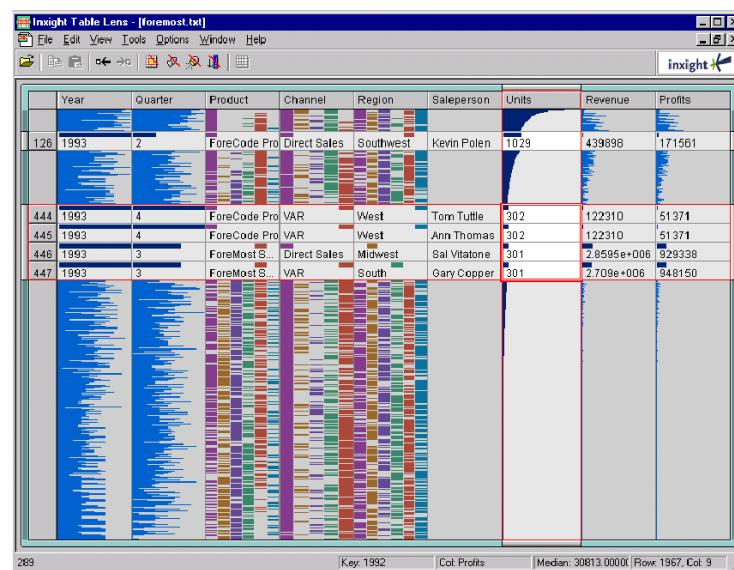
The Table Lens



Information Visualization and Visual Analytics – Multidimensional data Visualization

The Table Lens

The Table Lens



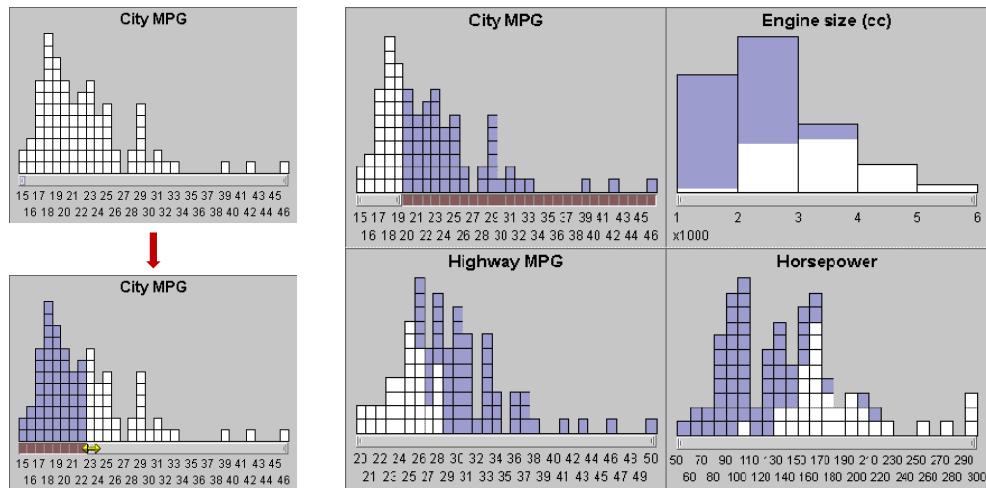
Video at <http://www.open-video.org/details.php?videoid=8304>

Information Visualization and Visual Analytics – Multidimensional data Visualization

Attribute Explorer

Attribute Explorer

- Dynamic Query on Attributes



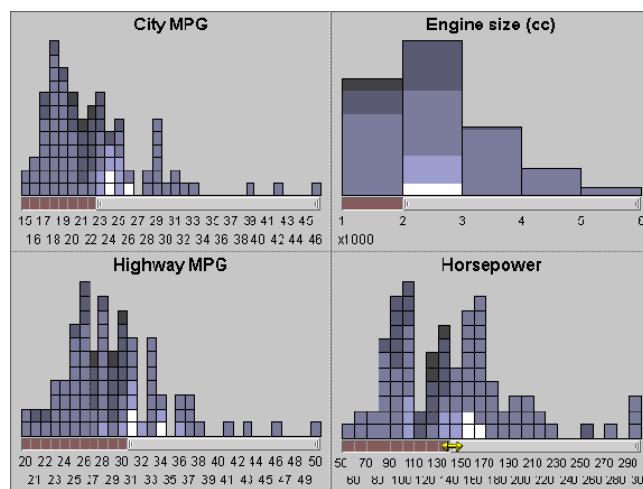
Tweedie, L., Spence, R., Williams, D.M.L., & Bhogal, R. (1994). The Attribute Explorer. ACM, Conference Companion Proceedings CHI '94, pp. 436-436.

Information Visualization and Visual Analytics – Multidimensional data Visualization

Attribute Explorer

Attribute Explorer

- color-coded with a progressively darker shade as they fail more constraints



Information Visualization and Visual Analytics – Multidimensional data Visualization

Characteristics

- Multiple histogram views, one per attribute (like trellis)
- Each data case represented by a square
- Square is positioned relative to that case's value on that attribute
- Selecting a case in a view lights it up in others
- Query sliders for narrowing
- Use shading to indicate level of unsatisfied query match

Jinwook Seo/John Stasko

Video at <http://www.open-video.org/details.php?videoid=8162>

Information Visualization and Visual Analytics – Multidimensional data Visualization

Summary

- Features
 - Attribute histogram
 - Attribute relationship
 - Sensitivity information
 - Especially useful in “zero-hits” situations or when you are not familiar with the data at all

Jinwook Seo/John Stasko

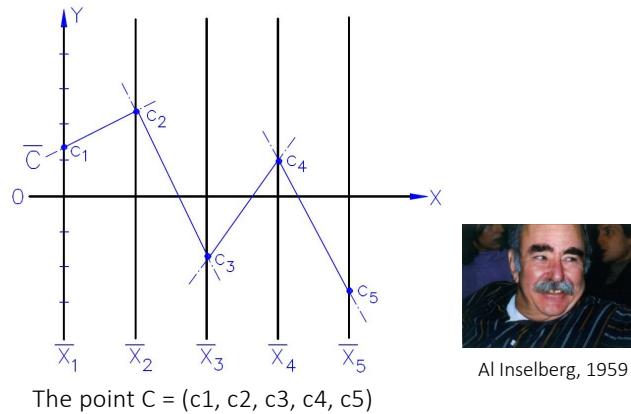
Video at <http://www.open-video.org/details.php?videoid=8162>

Information Visualization and Visual Analytics – Multidimensional data Visualization

Parallel Coordinates Plot and its variants

Parallel Coordinates Plot (PCP), ||- coords

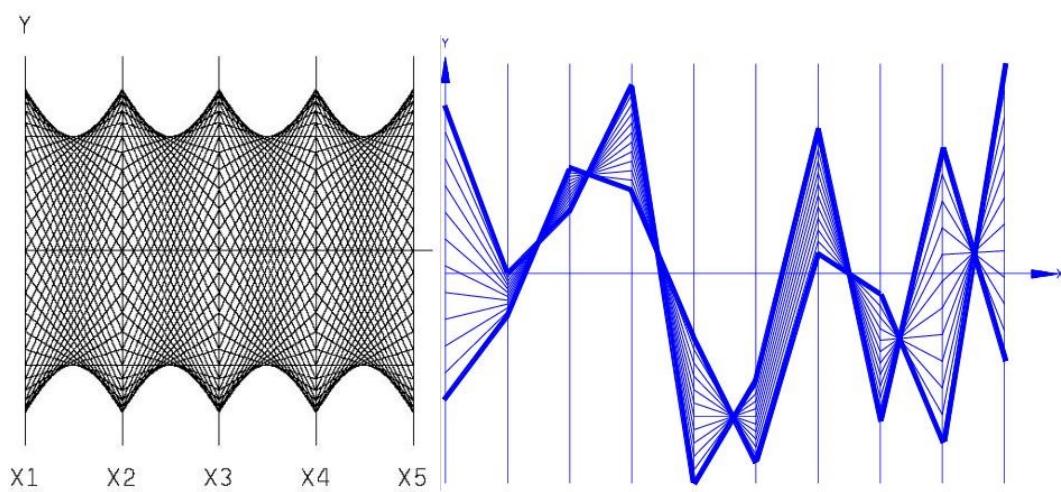
- Each vertical bar represents a variable (and often has its own scale)
- Show how each variables are related to each other



<http://www.math.tau.ac.il/~aiisreal/>

Information Visualization and Visual Analytics – Multidimensional data Visualization

Parallel Coordinates Plot and its variants



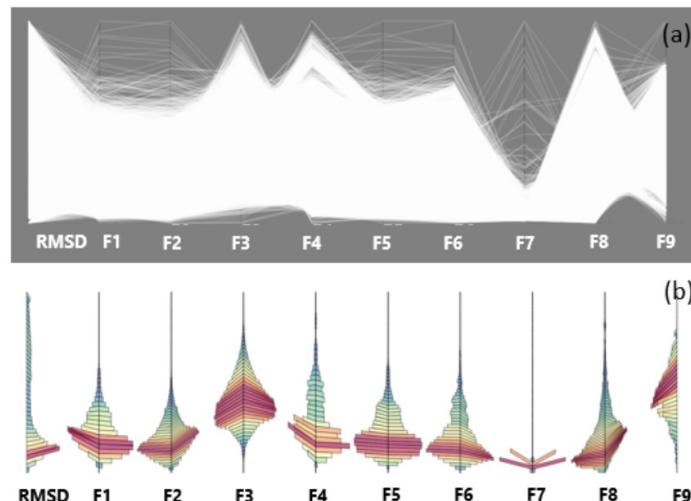
<http://www.math.tau.ac.il/~aiisreal/>

Information Visualization and Visual Analytics – Multidimensional data Visualization

Parallel Coordinates Plot and its variants

Angular Histogram (AH)

- Resolves overplotting problem of PCP

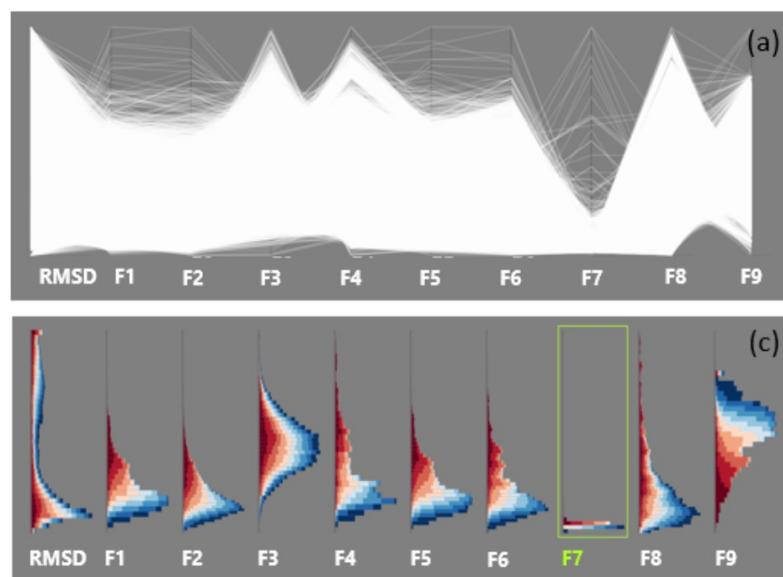


Bok, Jinwook, Bohyoung Kim, and Jinwook Seo. "Augmenting parallel coordinates plots with color-coded stacked histograms. TVCG 2021

Information Visualization and Visual Analytics – Multidimensional data Visualization

Parallel Coordinates Plot and its variants

Parallel Histogram Plots (PHP)

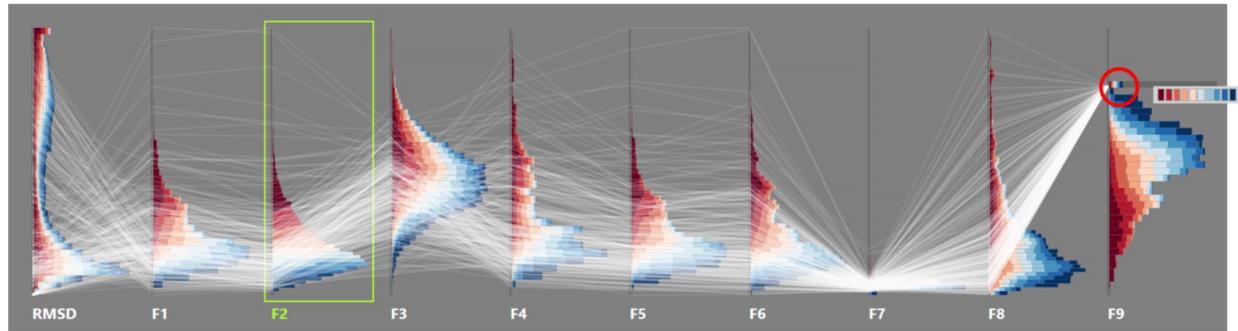


Bok, Jinwook, Bohyoung Kim, and Jinwook Seo. "Augmenting parallel coordinates plots with color-coded stacked histograms. TVCG 2021

Information Visualization and Visual Analytics – Multidimensional data Visualization

Parallel Coordinates Plot and its variants

Parallel Histogram Plots (PHP)



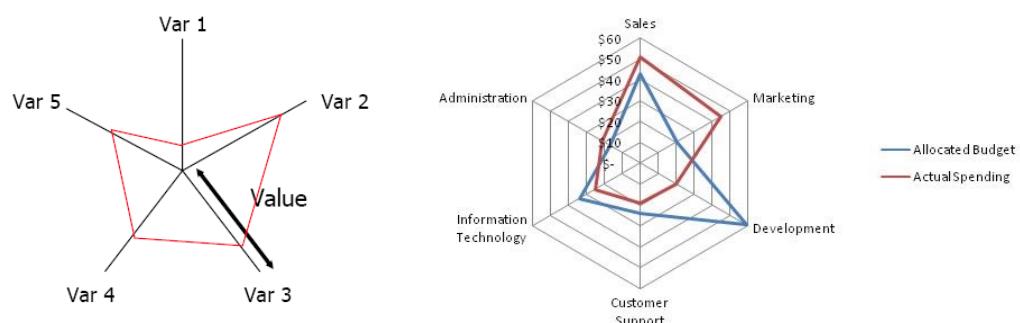
Bok, Jinwook, Bohyoung Kim, and Jinwook Seo. "Augmenting parallel coordinates plots with color-coded stacked histograms. TVCG 2021

Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional Data Visualizations

Star Plot (Radar Chart)

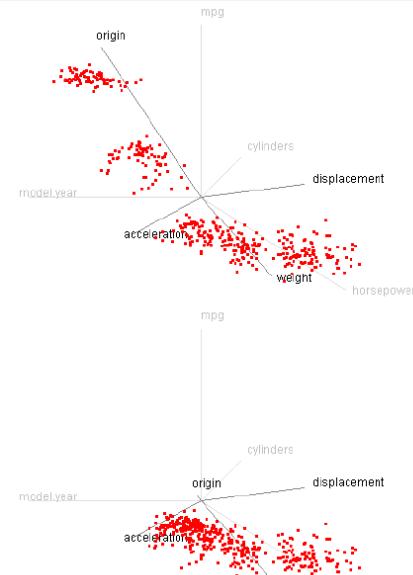
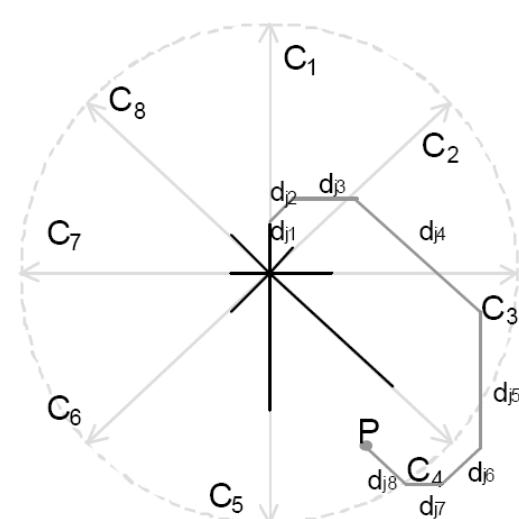
- Space out the n variables at equal angles around a circle
- Each “spoke” encodes a variable’s value
- Data point is now a “shape”



Jinwook Seo/John Stasko
http://en.wikipedia.org/wiki/File:Spider_Chart.jpg

Information Visualization and Visual Analytics – Multidimensional data Visualization

Star Coordinates



E. Kandogan, "Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions", InfoVis 2000
 Jinwook Seo/John Stasko
 Late-breaking Hot Topics, Oct. 2000
 Information Visualization and Visual Analytics – Multidimensional data Visualization

Star Coordinates

- Similar to star plot idiom
- Rather than represent point as polyline, just accumulate values along a *vector parallel to particular axis*
- Data case then becomes a **point**
- Find clusters and/or outliers
- Low dimensional ***projection*** methods

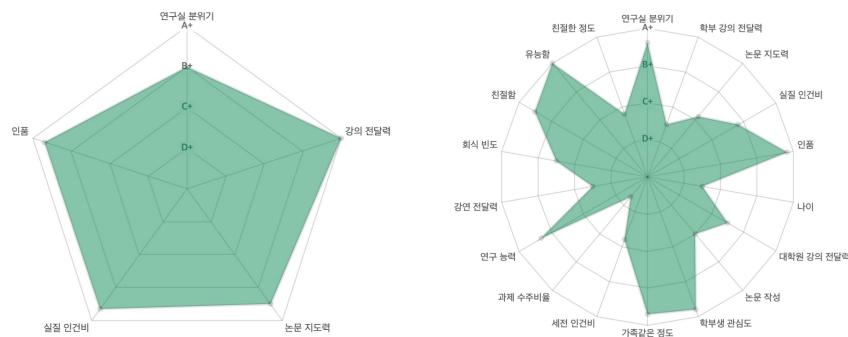
Multidimensional vs. High-dimensional

- No specific boundary
 - Often used interchangeably
- Typically & In this lecture...
- **High-Dimensional data** stands for data that cannot be scalably visualized by conventional multidimensional visualization
- **High-Dimensional visualizations** denotes the idioms that are designed for those data

Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional vs. High-dimensional

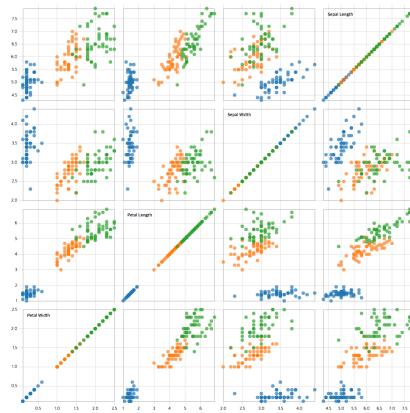
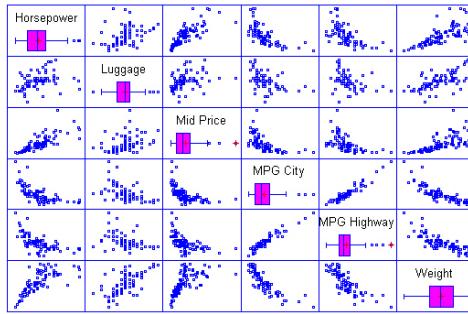
- **High-Dimensional data** stands for data that cannot be scalably visualized by conventional multidimensional visualization
- **High-Dimensional visualizations** denotes the idioms that are designed for those data



Information Visualization and Visual Analytics – Multidimensional data Visualization

Scatterplot Matrix (SPLOM)

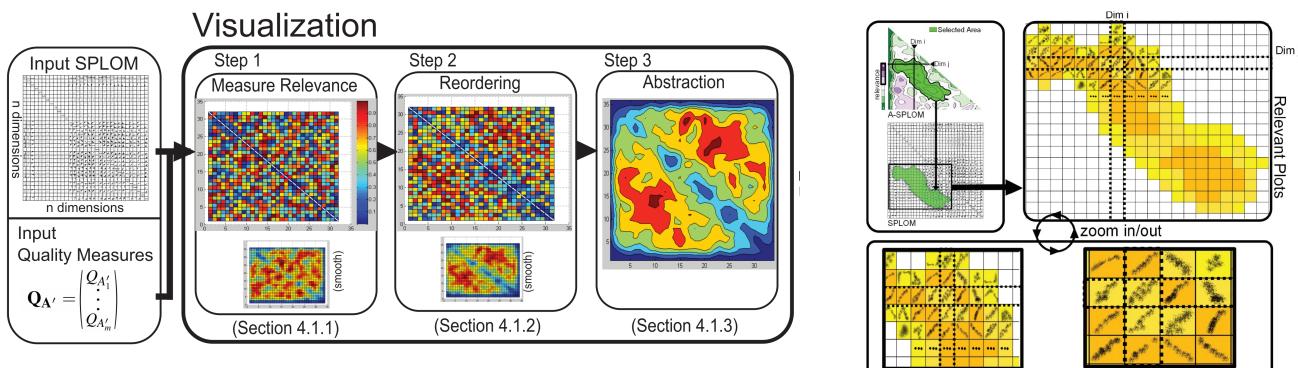
- Number of scatterplot required: $O(n^2)$
 - where n denotes the number of attributes
 - Makes SPLOM unscalable



Information Visualization and Visual Analytics – Multidimensional data Visualization

Large Scatterplot Matrix

- High-dimensional ver. of scatterplot matrix
- Reorder & Abstract scatterplots to find interesting pattern



Dirk J. Lehmann et al. Selecting Coherent and Relevant Plots in Large Scatterplot Matrices, CGF 2012

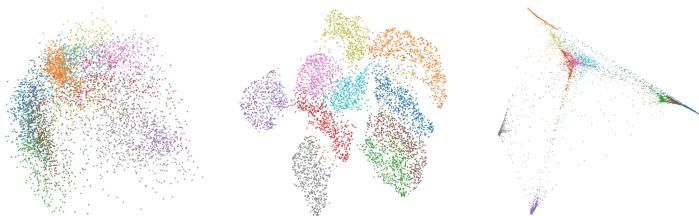
Information Visualization and Visual Analytics – Multidimensional data Visualization

High-dimensional Data Visualizations

Multidimensional Projections (MDP)

- Generate 2D projection of Multidimensional data
- *Try to preserve the characteristics of the original data as most as they can*
 - e.g., cluster structure, outlier, distances between points
- Each Multidimensional data point is transformed to a 2D representation
 - Commonly represented as scatterplot

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4  
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5  
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6  
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7  
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8  
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
```

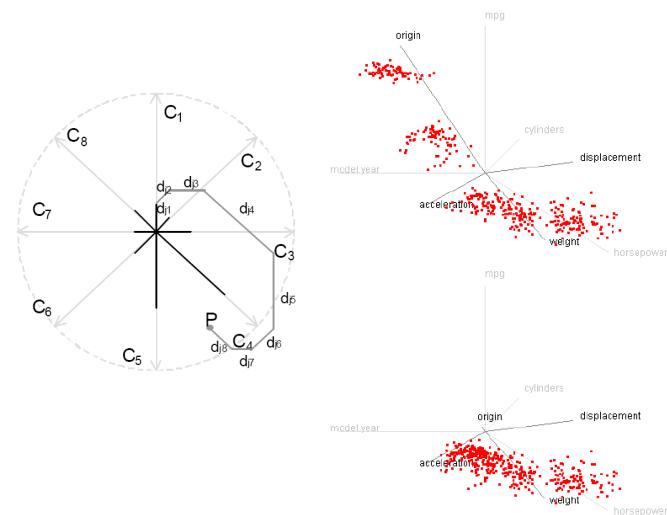


Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional Projections

Star Coordinates

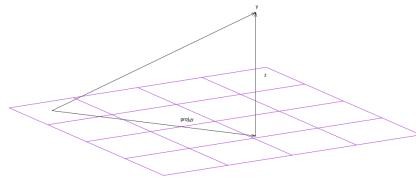
- Star Coordinates are not scalable for high-dimensional data!!



Information Visualization and Visual Analytics – Multidimensional data Visualization

Linear Projections

- Linear transformation of high-dimensional data

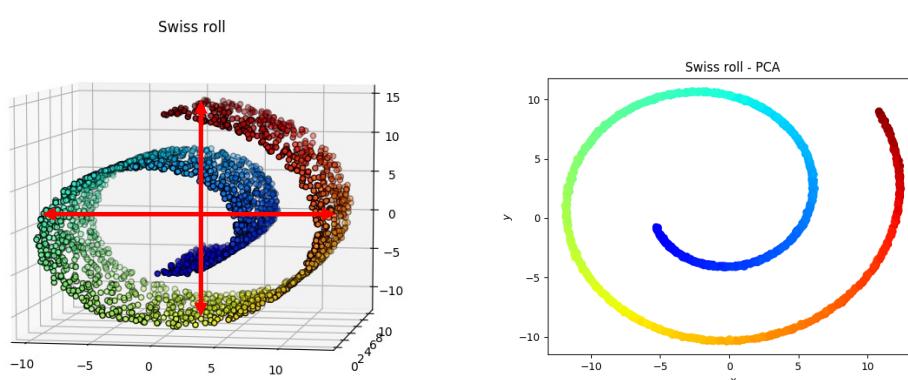


- Principal Component Analysis (PCA)
 - Find linear transformation which best preserves data's variation
- Multidimensional Scaling (MDS)
 - Find linear transformation that best preserves the pairwise distances between points

Information Visualization and Visual Analytics – Multidimensional data Visualization

Linear Projections

- Principal Component Analysis (PCA)
 - Find linear transformation which best preserves data's variation

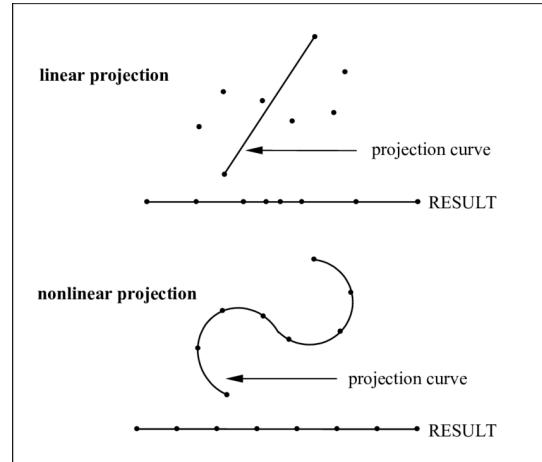
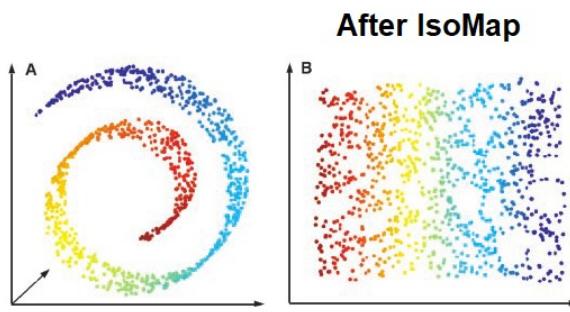


<https://www.thekerneltrip.com/statistics/tsne-vs-pca/>

Information Visualization and Visual Analytics – Multidimensional data Visualization

Linear Projections

- Cannot “unfold” complex manifold in high-dimensional space
- => Nonlinear Projections arise!!



Kavzoglu, Taskin. (2001). An investigation of the design and use of feed-forward artificial neural networks in the classification of remotely sensed images.

Information Visualization and Visual Analytics – Multidimensional data Visualization

Nonlinear Projections

- Anything that's not a linear projection
 - Features (e.g., clusters, manifolds) in high-dimensional space is nonlinear!!
 - Nonlinear projections aims to preserve those features
 - e.g., t-SNE, UMAP, Isomap, LLE
- Popular research topic in Machine learning / Big data / Data mining field
- Let's focus on how *visualization research field* treats projections

Information Visualization and Visual Analytics – Multidimensional data Visualization

Nonlinear Projections

- t-SNE (t-distributed Stochastic Neighbor Embedding)
 - Machine-learning based nonlinear projections
 - Tries to match the data distribution of the original and projected space!!
 - Uses KL-divergence as loss function
 - Well-preserves pairwise similarity between points
 - One of the most widely used projection methods

0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9



Information Visualization and Visual Analytics – Multidimensional data Visualization

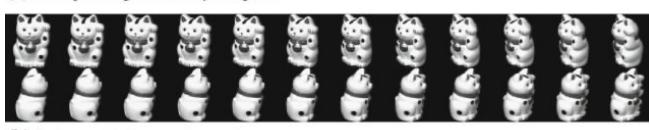
Multidimensional Projections

t-SNE

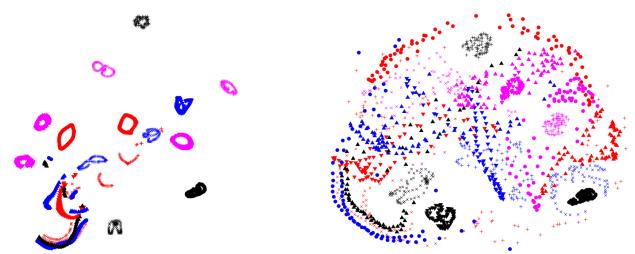
- Coil-20 dataset projections



(a) Example images of twenty categories

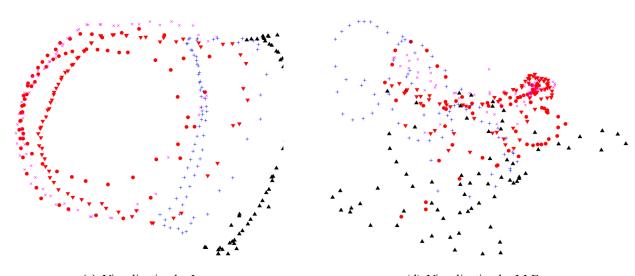


(b) Part example images of one category



(a) Visualization by t-SNE.

(b) Visualization by Sammon mapping.



(c) Visualization by Isomap.

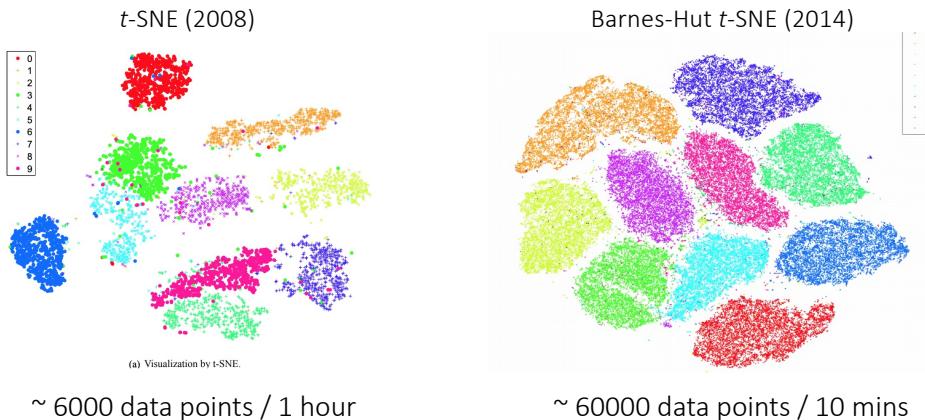
(d) Visualization by LLE.

Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research* 9.11 (2008).

Information Visualization and Visual Analytics – Multidimensional data Visualization

Accelerating t-SNE

- Barnes hut t-SNE (2014): by only focusing on k -nearest neighbors

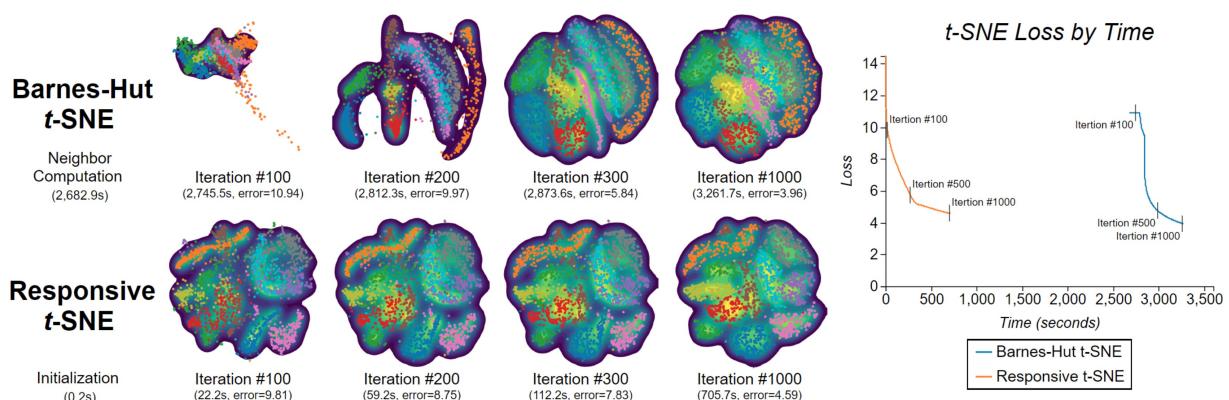


Van Der Maaten, Laurens. "Accelerating t-SNE using tree-based algorithms." *The Journal of Machine Learning Research* 15.1 (2014): 3221-3245.
 Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research* 9.11 (2008).

Information Visualization and Visual Analytics – Multidimensional data Visualization

Accelerating t-SNE

- Responsive t-SNE (2019)
 - Fast initialization / early convergence of loss function => better for visual analytics!!

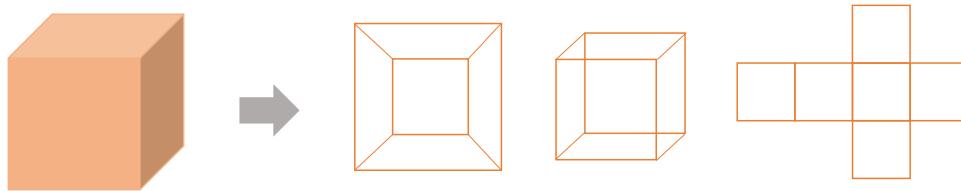


Jaemin Jo et al. PANENE: A Progressive Algorithm for Indexing and Querying Approximate k-Nearest Neighbors, TVCG 2019

Information Visualization and Visual Analytics – Multidimensional data Visualization

Distortions

- Even a 3D cube cannot be exactly projected in 2D space!!

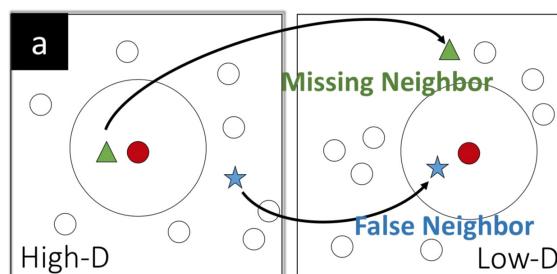


- The features of High-dimensional data are highly distorted in the projection
- Multidimensional Projections suffer due to **distortions***

Information Visualization and Visual Analytics – Multidimensional data Visualization

Distortions

- Distortions in point-point relationships
 - Missing Neighbors
 - Neighbors in the original space are no more neighbors in the projection
 - False Neighbors
 - Neighbors that can be seen in the projection are actually not neighbors in the original space



Hyeon Jeon et al. Measuring and Explaining the Inter-Cluster Reliability of Multidimensional Projections, TCVG 2021

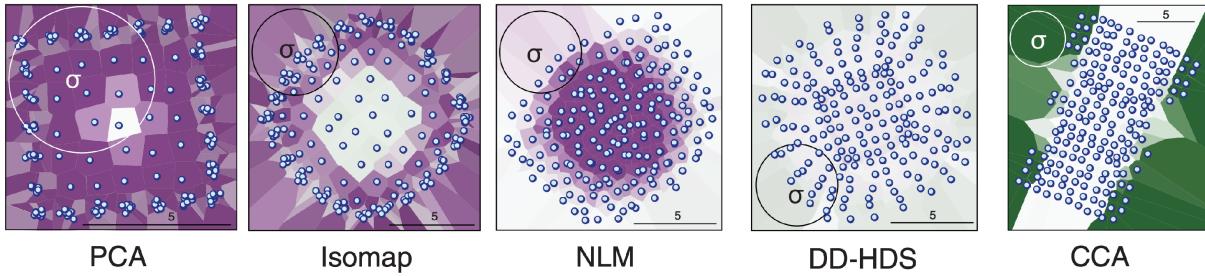
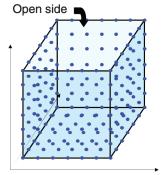
Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional Projections

Distortions

- CheckViz

- Represents how much each point suffers from Missing / False Distortions
- Missing Neighbors: **Green**, False Neighbors: **Purple**



- Goal: Find the distortion patterns occurred in projections

Lespinats, Sylvain, and Michaël Aupetit. "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings." *Computer Graphics Forum*.

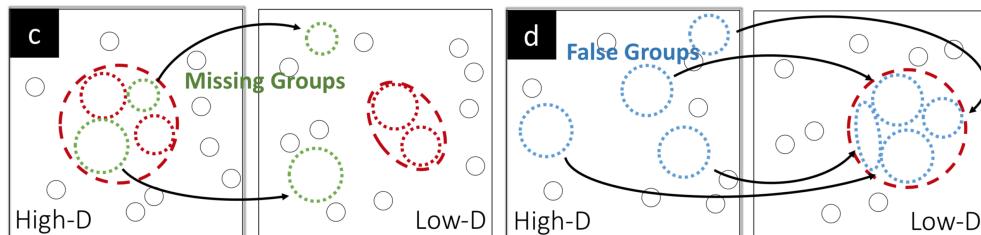
Information Visualization and Visual Analytics – Multidimensional data Visualization

Multidimensional Projections

Distortions

- Missing Groups and False Groups distortions

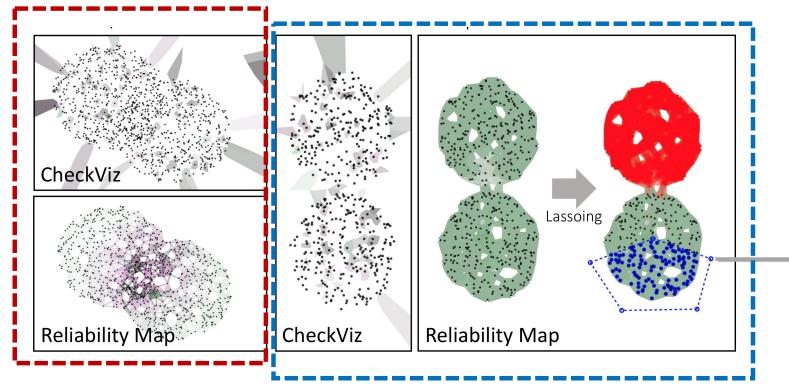
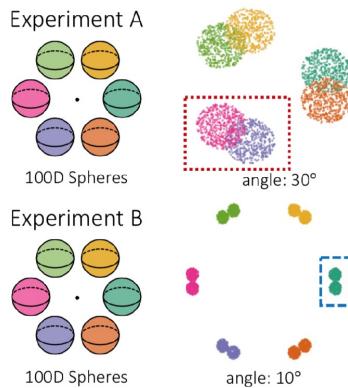
- Represents Missing / False concept in cluster level (cluster-cluster relationship)
- Better matches to visual analytics that focus on cluster analysis



Distortions

- Visualizing Missing / False Groups Distortions (Reliability map)

- Better represents the distortions pattern
- Better explains why distortions have been occurred



Hyeon Jeon et al. Measuring and Explaining the Inter-Cluster Reliability of Multidimensional Projections, TVCG 2021

Information Visualization and Visual Analytics – Multidimensional data Visualization

Note

- Questions?

Information Visualization and Visual Analytics – Multidimensional data Visualization