ARTS1422 Data Visualization

Lecture 7

High-dimensional Data Visualization

Quan Li

Spring 2024
2024. 03.19

INDUSTRIAL REVOLUTION OF DATA

Joe Hellerstein, UC Berkley, 2008

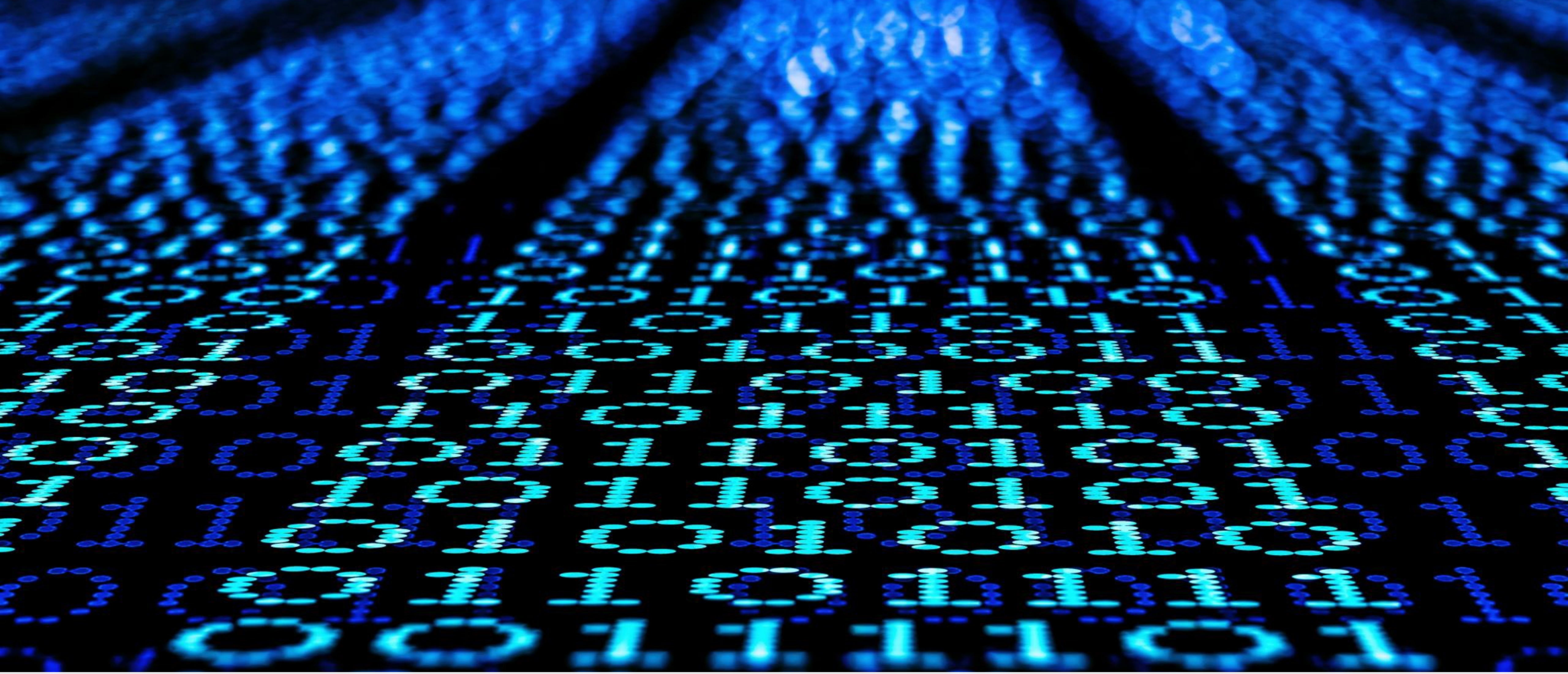
OUTLINE

1 Data Dimension

- 1-D
- 2-D
- 3-D
- High Dimension

2 High-Dimensional Data Visualization

- Dimensionality Reduction
- Scatter-plot Matrix
- Parallel Coordinates
- Glyph-based Methods
- "Small Multiples"
- Interaction: "Dust & Magnet"



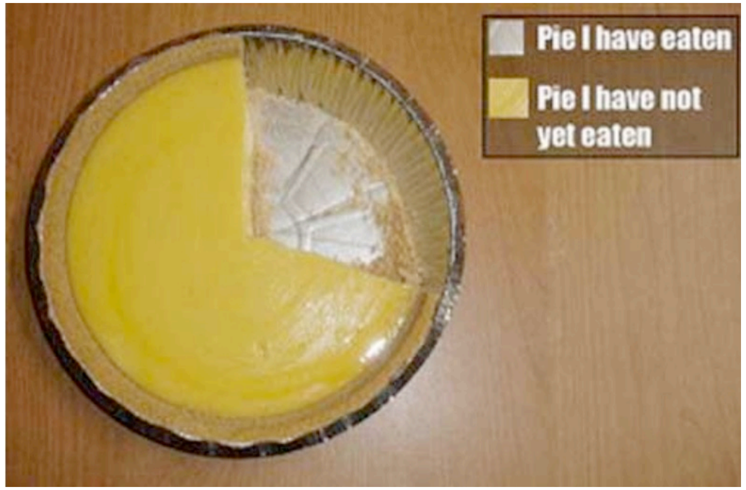
Data Dimension

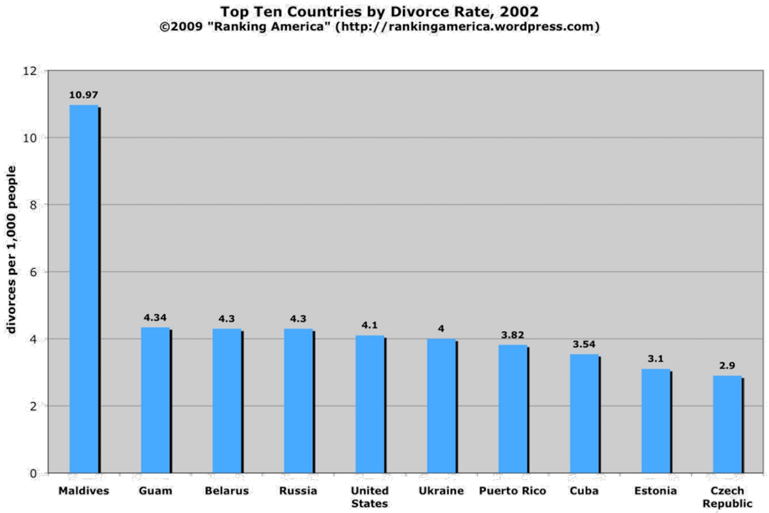
Review: Dimension

- Dimension (Number of attributes):
  - 1-D
  - 2-D
  - 3-D
  - High Dimension

1-D Data

“eaten or not”

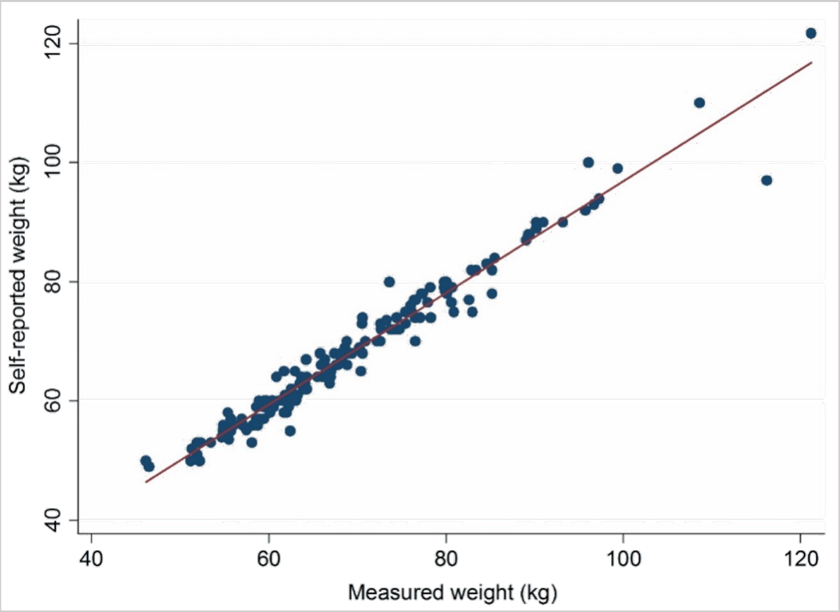


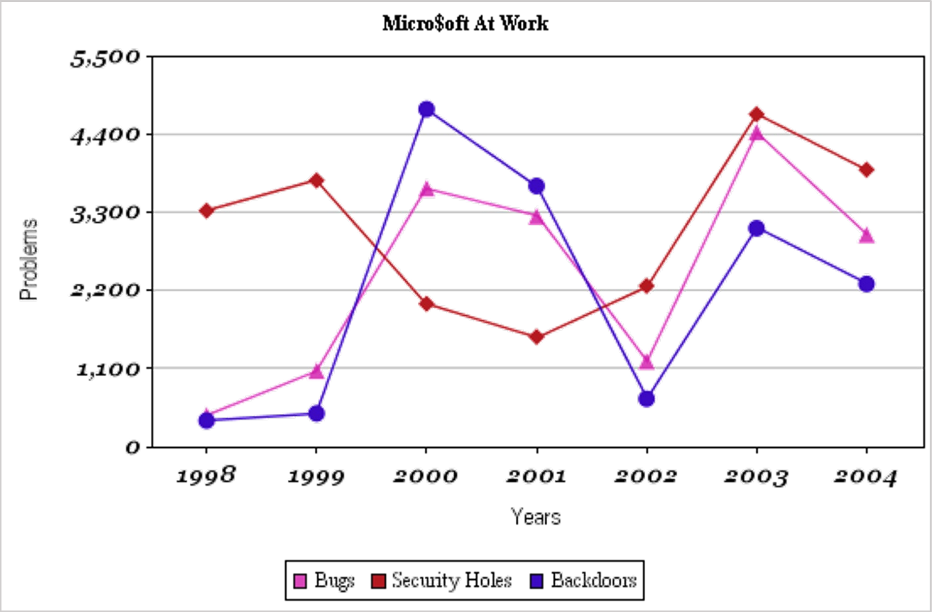


1- D data for each bar

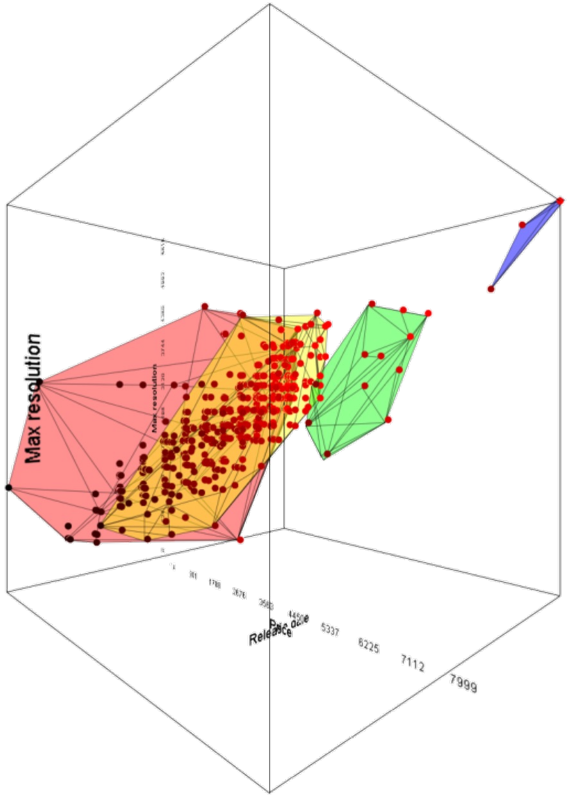
2-D Data

2-D plots





3-D Data



Elmqvist et al. "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation."
IEEE transactions on Visualization and Computer Graphics 14.6 (2008): 1539-1148.

High-Dim Data

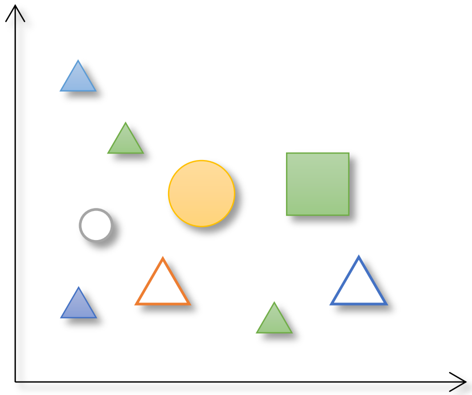
How to visualize high-dimensional data in visual space(2-D or 3-D) ?

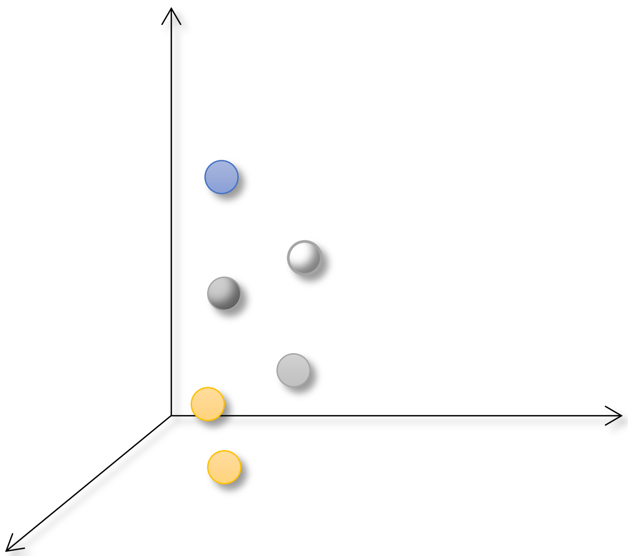
	身高	体重	年龄	性别	教育程度	籍贯
张三	180cm	65kg	23	男	大学	上海
李四	168cm	55kg	18	女	高中	浙江
赵五	175cm	75kg	53	男	初中	广东
…						

Simple Solutions

- Add more channel on 2-D or 3-D plots.

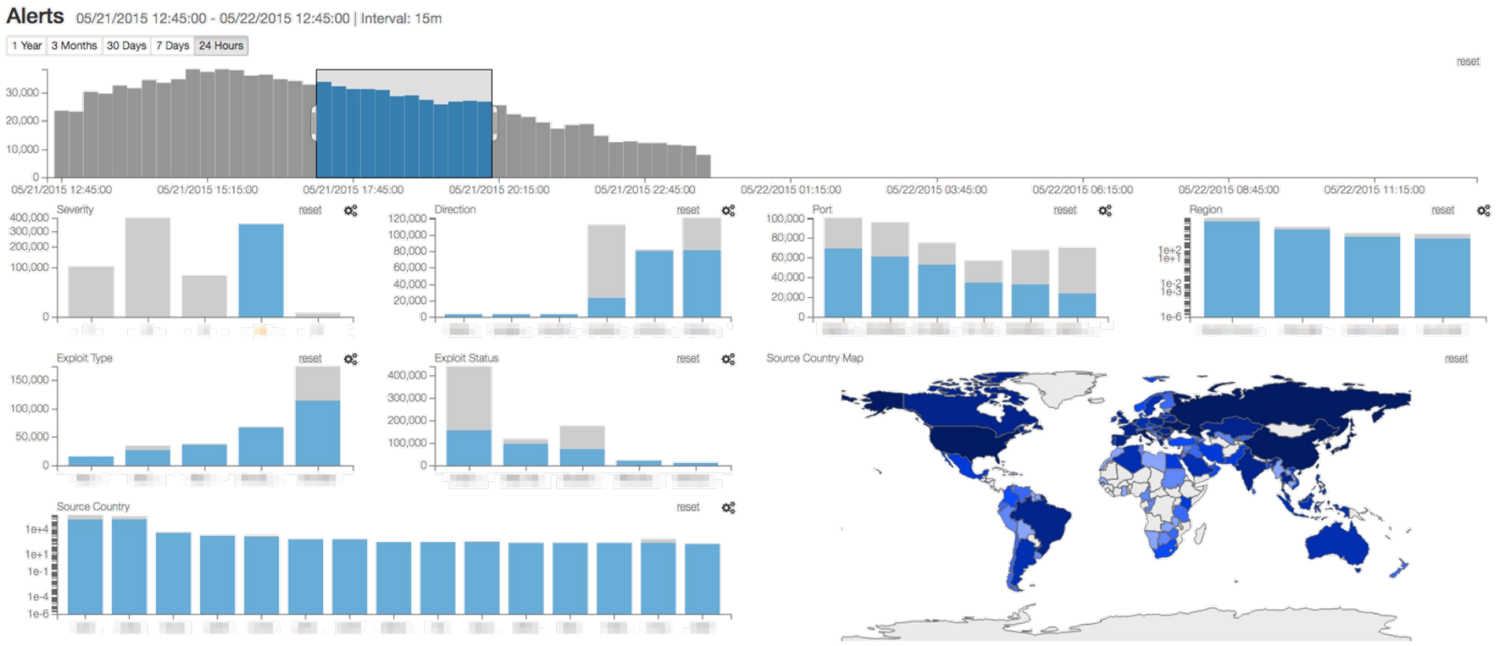
(Shape/fill style/color/size of points)





Simple Solutions

- Multiple coordinated views: present some attributes of objects in a view.



Andrew et al. "Leveraging Interaction History for Intelligent Configuration of Multiple Coordinated Views in Visualization Tools." LIVVIL: Logging Interactive Visualizations & Visualizing Interaction Logs (2016).

More solutions?

High-Dimensional Data Visualization

Approaches

- Dimensionality Reduction.
- Scatter-plot Matrix.
- Parallel Coordinates.
- Glyph-based Methods.
- “Small Multiples”.
- Interaction: “Dust & Magnet”.

Dimensionality Reduction

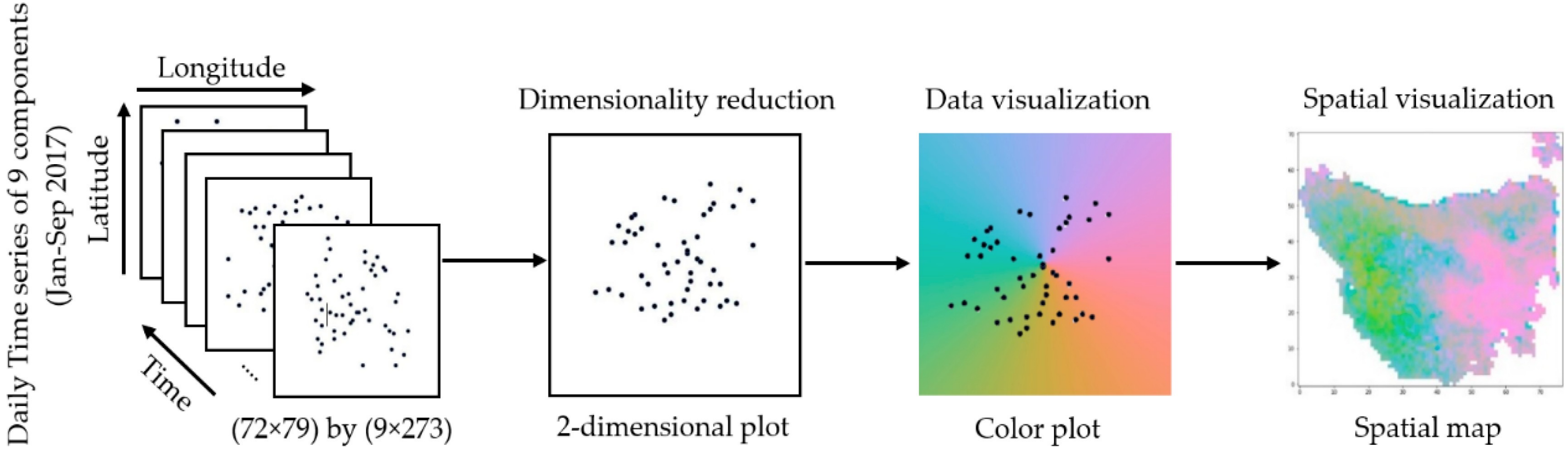
Dimensionality Reduction

- Project the high-dimensional data onto a lower-dimensional subspace using linear or non-linear transformations.
- Projection preserves important relations (e.g., no information loss, data discrimination).

- $x = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_N \end{pmatrix} \rightarrow reduce\ dimensionality \rightarrow \hat{x} = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_K \end{pmatrix} (K \ll N)$

When to Use DR?

- How do you know when you would benefit from DR?
- Consider error for low-dim projection vs. high-dim projection
- No single correct answer; many metrics proposed
- Cumulative variance that is not accounted for
- Strain: match variations in distance (vs. actual distance values)
- Stress: difference between interpoint distances in high and low dimensions



Mazher, A., 2020. Visualization framework for high-dimensional spatio-temporal hydrological gridded datasets using machine-learning techniques. *Water*, *12*(2), p.590.

When to Use DR with Visualization?

Why do people do DR?

- Improve performance of downstream algorithm
  - Avoid curse of dimensionality
- Data analysis
  - If looking at the output: visual data analysis

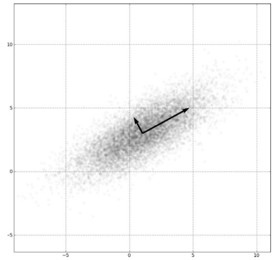
Abstract tasks when visualizing DR data

- Dimension-oriented tasks
- Naming synthesized dimensions, mapping synthesized dimensions to original dimensions

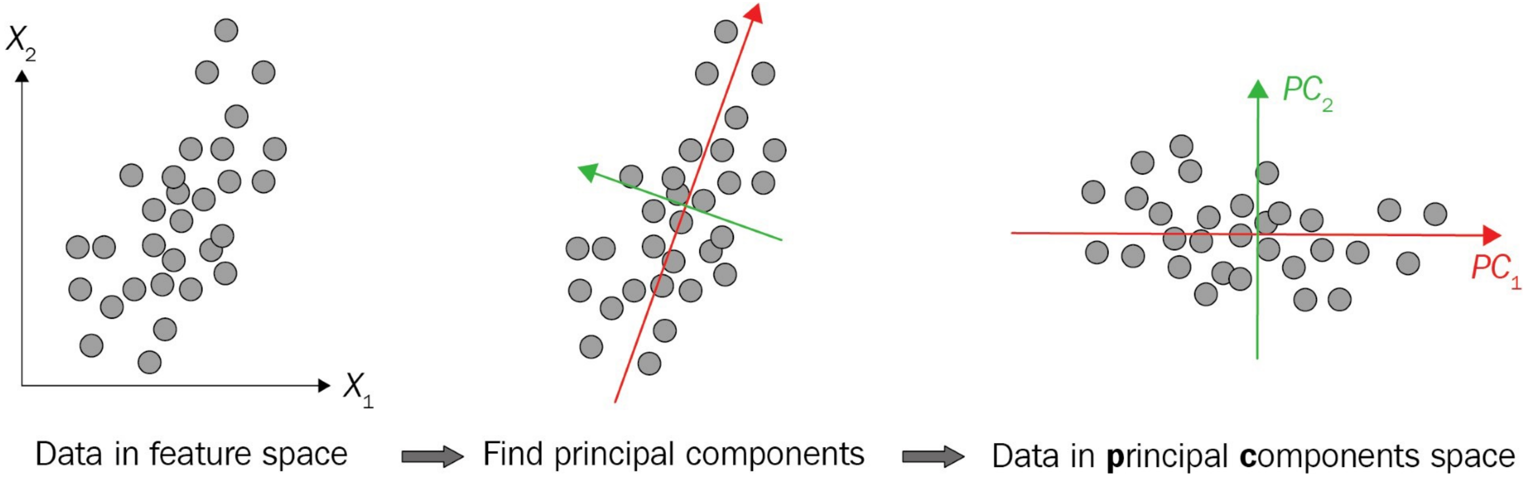
Cluster-oriented tasks

- Verifying clusters, naming clusters, matching clusters and classes

Linear Dimensionality Reduction



- Principal components analysis (PCA)
- Finding axes: first with most variance, second with next most, etc.
- Describe location of each point as linear combination of weights for each axis
- - Mapping synthesized dimensions to original dimensions



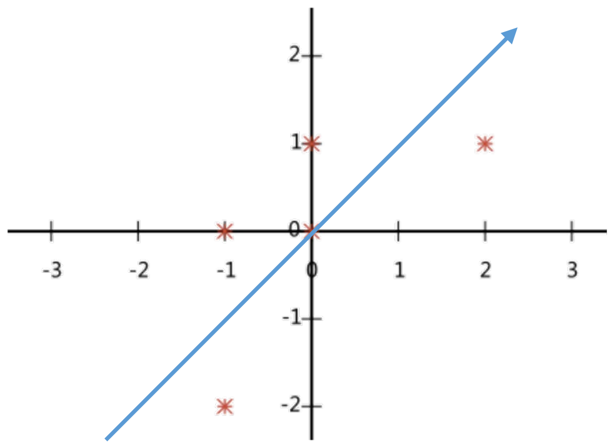
https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781789345070/5/ch05lvl1sec42/dimensionality-reduction

https://en.wikipedia.org/wiki/File:GaussianScatterPCA.png

PCA Motivation I

- Data set has two dimensions.
- Projections should spread as much as possible.

$$\begin{pmatrix} 1 & 1 & 2 & 4 & 2 \\ 1 & 3 & 3 & 4 & 4 \end{pmatrix} \xrightarrow{\text{Subtract the average}} \begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}$$



How can we use one dimension to represent the data with most information preserved?

Variance

- Variance represents the spread of data items.

- Let the average in each dimension be 0, i.e., $\bar{a}$=0.

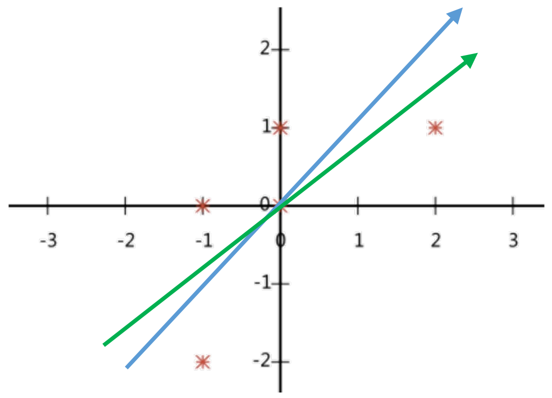
$$\frac{1}{m}\sum_{i=1}^{m}(a_i - \bar{a})^2$$

- Question: How to find one coordinate (projection), such that the data items are projected to the coordinate with the maximum variance?

$$\frac{1}{m}\sum_{i=1}^{m}(a_i)^2$$

PCA Motivation II

- How to choose more coordinates?
  - Shall we consider only the variance?
    - Coordinates may overlap.
- Coordinates should be linearly uncorrelated to preserve more information.
  - Correlations mean two dimensions are dependent.



Covariance

- The correlation between dimensions a and b can be represented by their covariance:
- We make $\bar{a} = 0$, $\bar{b} = 0$, so we have $$\frac{1}{m}\sum_{i=1}^{m}(a_i - \bar{a})(b_i - \bar{b})^T$$
- Covariance $= 0$ means a and b are uncorrelated.
  - The second coordinate must be orthogonal to the first one.
  - The two projection coordinates must be orthogonal.

$$\frac{1}{m}\sum_{i=1}^{m} a_i b_i$$

Covariance Matrix

- Given two dimensions a and b, we have matrix X:
- The covariance matrix can be obtained by

$$X = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{pmatrix}$$

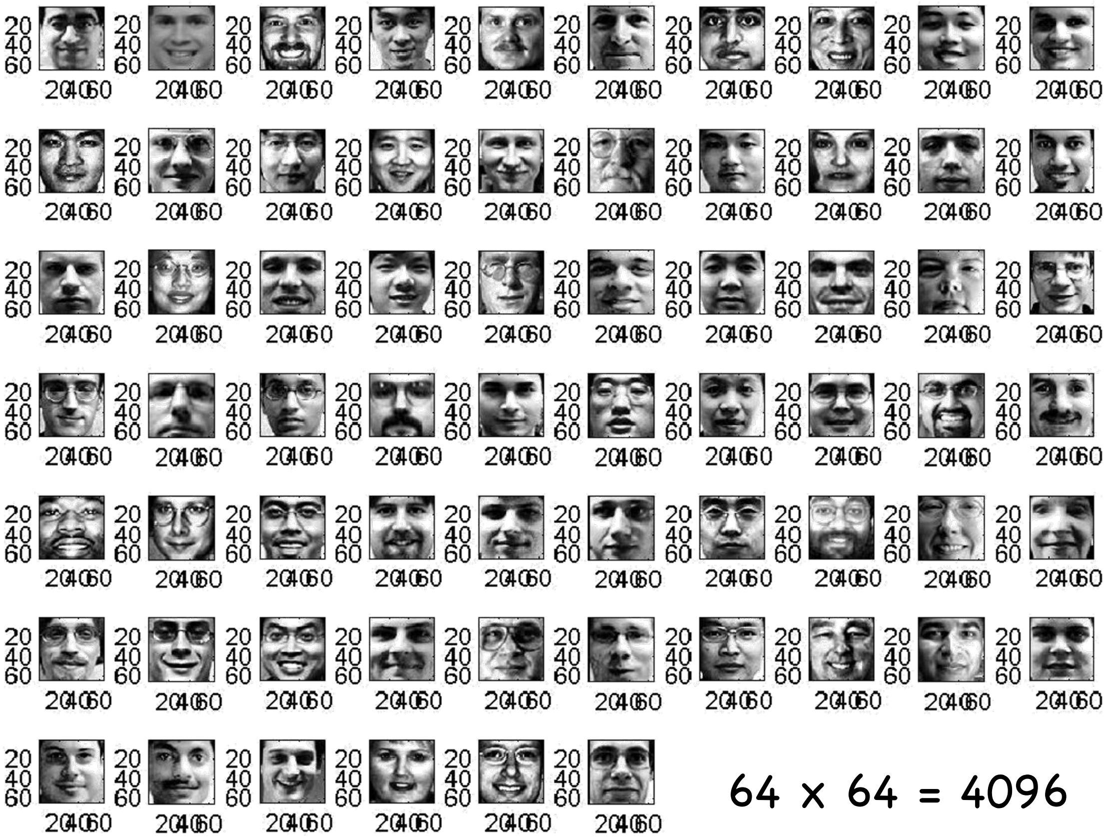
$$S = \frac{1}{m} X X^T = \frac{1}{m} \sum_{i=1}^{m} X_i X_i^T = \begin{pmatrix} \frac{1}{m} \sum_{i=1}^{m} a_i^2 & \frac{1}{m} \sum_{i=1}^{m} a_i b_i \\ \frac{1}{m} \sum_{i=1}^{m} a_i b_i & \frac{1}{m} \sum_{i=1}^{m} b_i^2 \end{pmatrix}$$

PCA Mechanics

Suppose $x_1, x_2, \dots, x_M$ are $H \times 1$ vectors:

1. $\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i$.
2. Subtract the mean $\Phi_i = x_i - \bar{x}$.
3. Form $H \times M$ matrix $A = [\Phi_1 \Phi_2 \cdots \Phi_\mathrm{M}]$.
4. Compute covariance matrix $C = \frac{1}{M}\sum_{i=1}^{M} \Phi_n \Phi_n^T = AA^T$.
5. Compute eigenvalues of C: $\lambda_1 > \lambda_2 > \cdots > \lambda_N$.
6. Compute eigenvectors of C: $u_1, u_2, \cdots, u_N$.

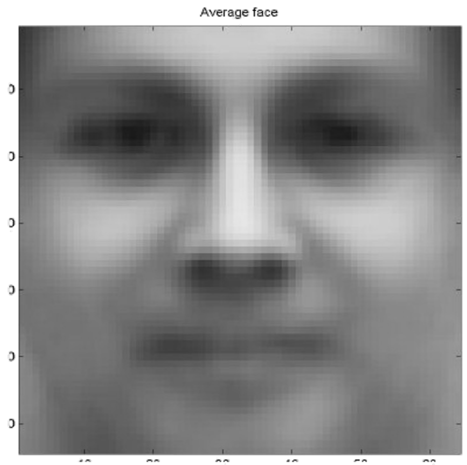
PCA Applied to Faces

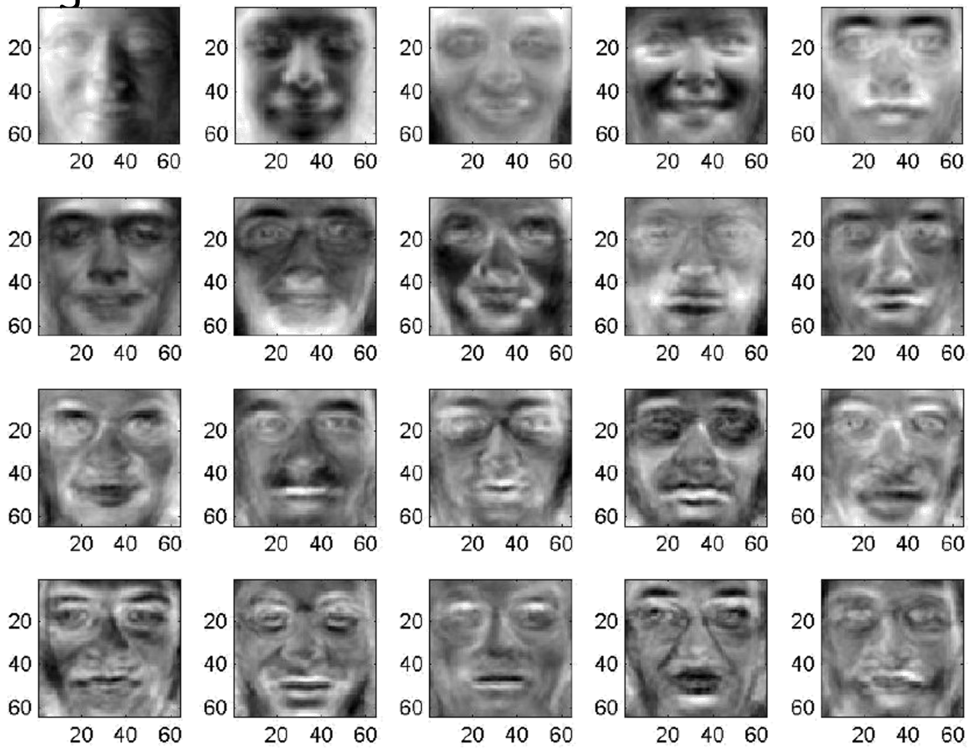


$64 \times 64 = 4096$

PCA Applied to Faces

Reconstruct each face as a linear combination of "basis faces", or Eigenfaces.

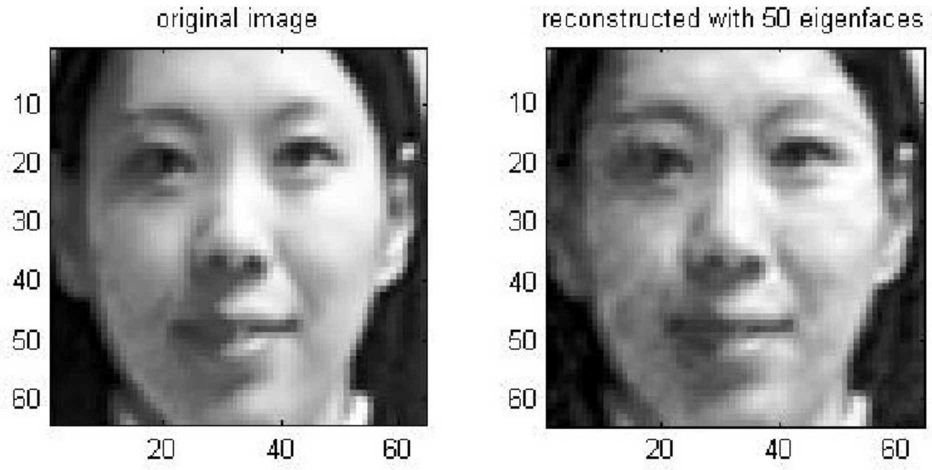


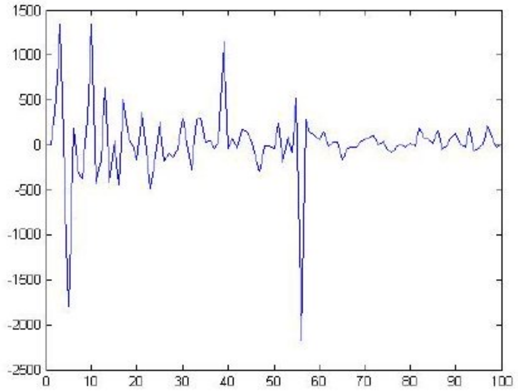


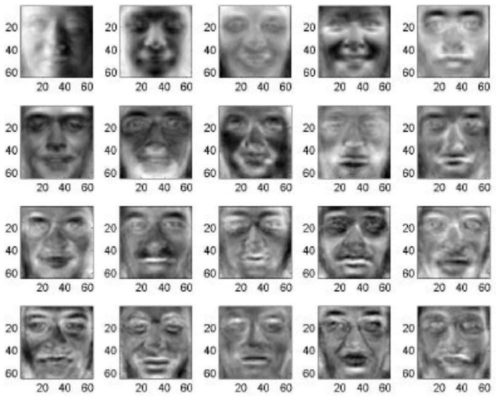
Eigenfaces

Reconstruction

- 90% variance is captured by the first 50 eigenvectors.
- Reconstruct existing faces using only 50 basis images.

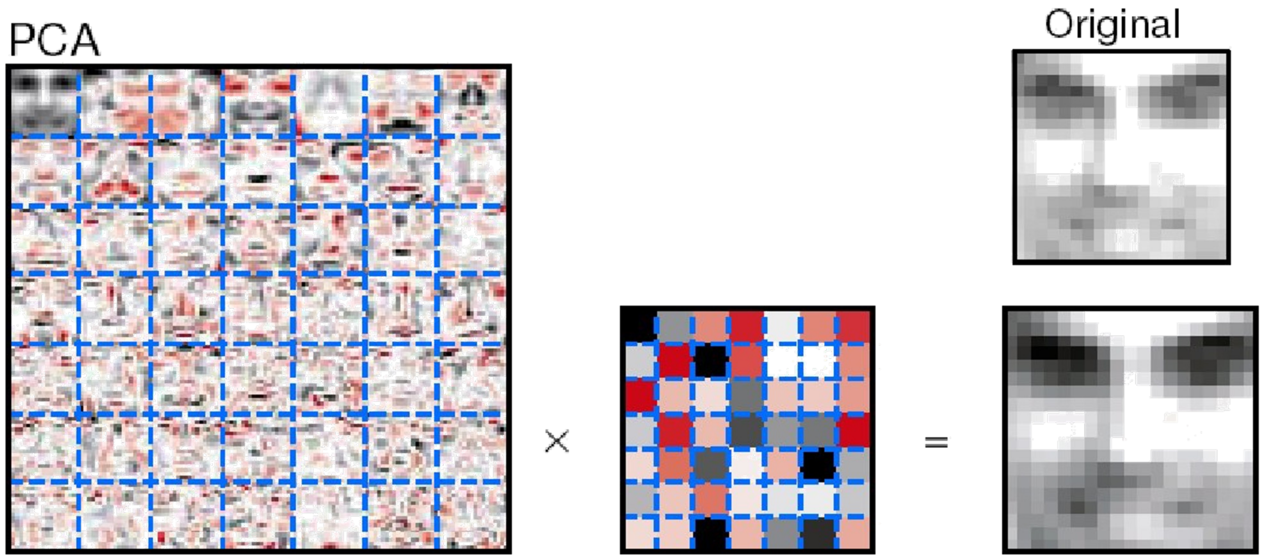

original image
reconstructed with 50 eigenfaces
10
20
30
40
50
60
20
40
60


1500
1000
500
0
-500
-1000
-1500
-2000
-2500
0
10
20
30
40
50
60
70
80
90
100



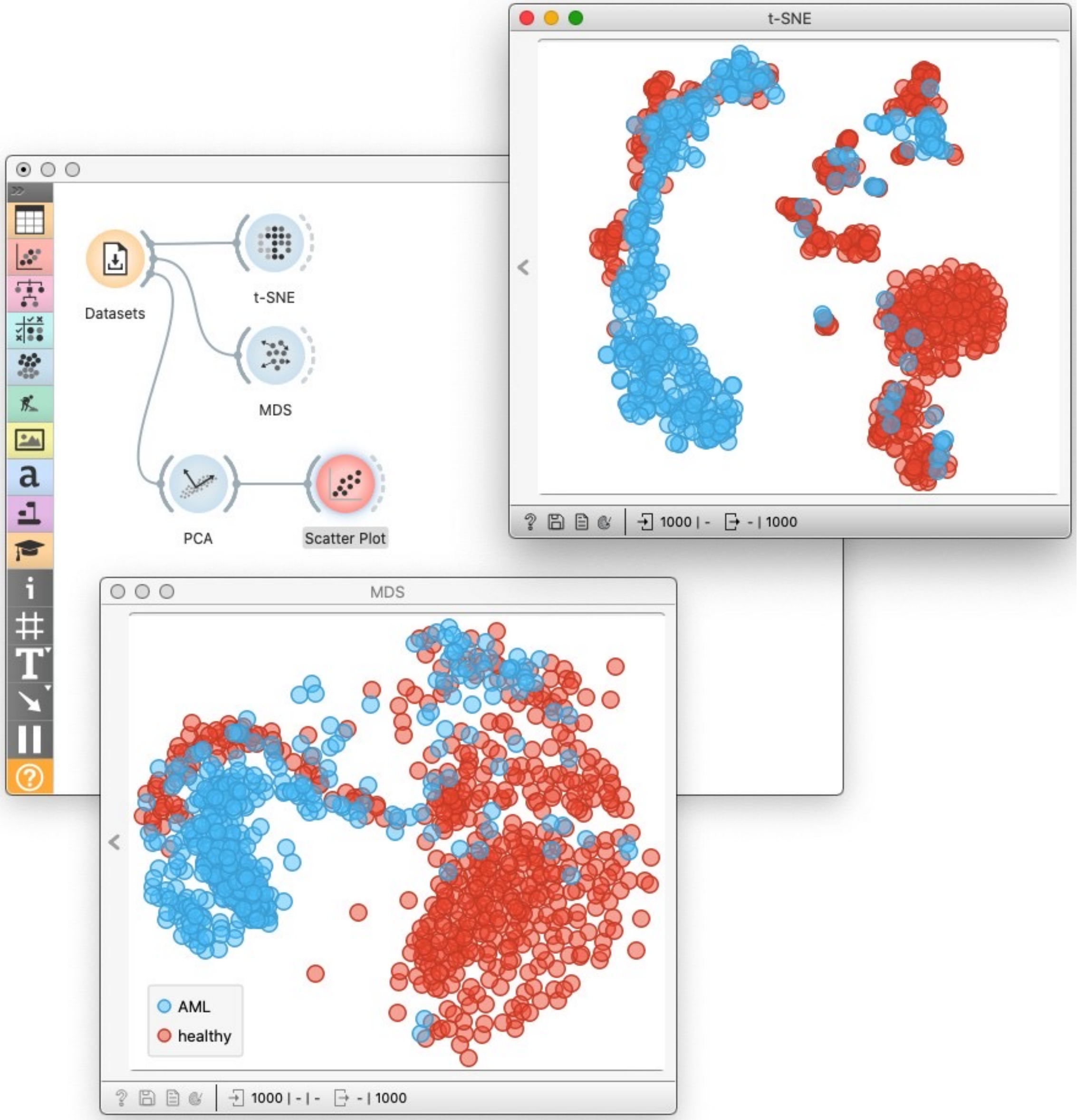
Issues

- PCA involves adding up some basis images and subtracting others.
- The basis images are not physically intuitive.



Nonlinear Dimensionality Reduction

- Pro: can handle curved rather than linear structure
- Con: lose all ties to original dimensions/attributes
  - New dimensions often cannot be easily related to originals
    - Mapping synthesized dims to original dims task is difficult
- Many literatures: visualization, machine learning, optimization, psychology, etc.
- Many techniques proposed: t-SNE, MDS (multidimensional scaling), charting, isomap, LLE, etc.
  - **t-SNE**: excellent for clusters
    - But some trickiness remains
  - **MDS**: confusingly, entire family of techniques, both linear and nonlinear
    - Minimize stress or strain metrics
    - Early formulations equivalent to PCA



[How to Use t-SNE Effectively (href="http://distill.pub/2016/misread-tsne/)]

Multidimensional Scaling (MDS)

- Takes as input a matrix M containing pairwise distances between H-dimensional data points.
- Outputs a projection of data in L-dimensional space where the pairwise distances match the original distances as faithfully as possible.

An Example: US Map

- Suppose you know the distances between a bunch of cities…

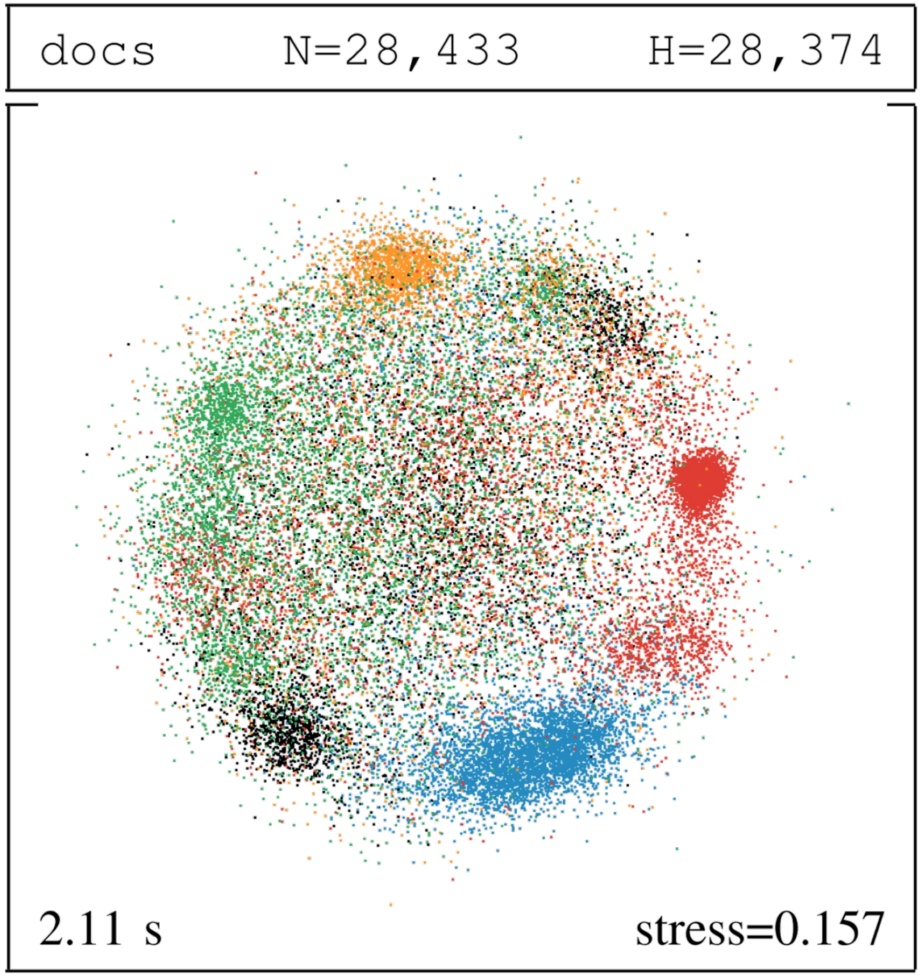
	Chicago	Raleigh	Boston	Seattle	S.F.	Austin	Orlando
Chicago	0						
Raleigh	641	0					
Boston	851	608	0				
Seattle	1733	2363	2488	0			
S.F.	1855	2406	2696	684	0		
Austin	972	1167	1691	1764	1495	0	
Orlando	994	520	1105	2565	2458	1015	0

Result of MDS

Actual Plot of Cities

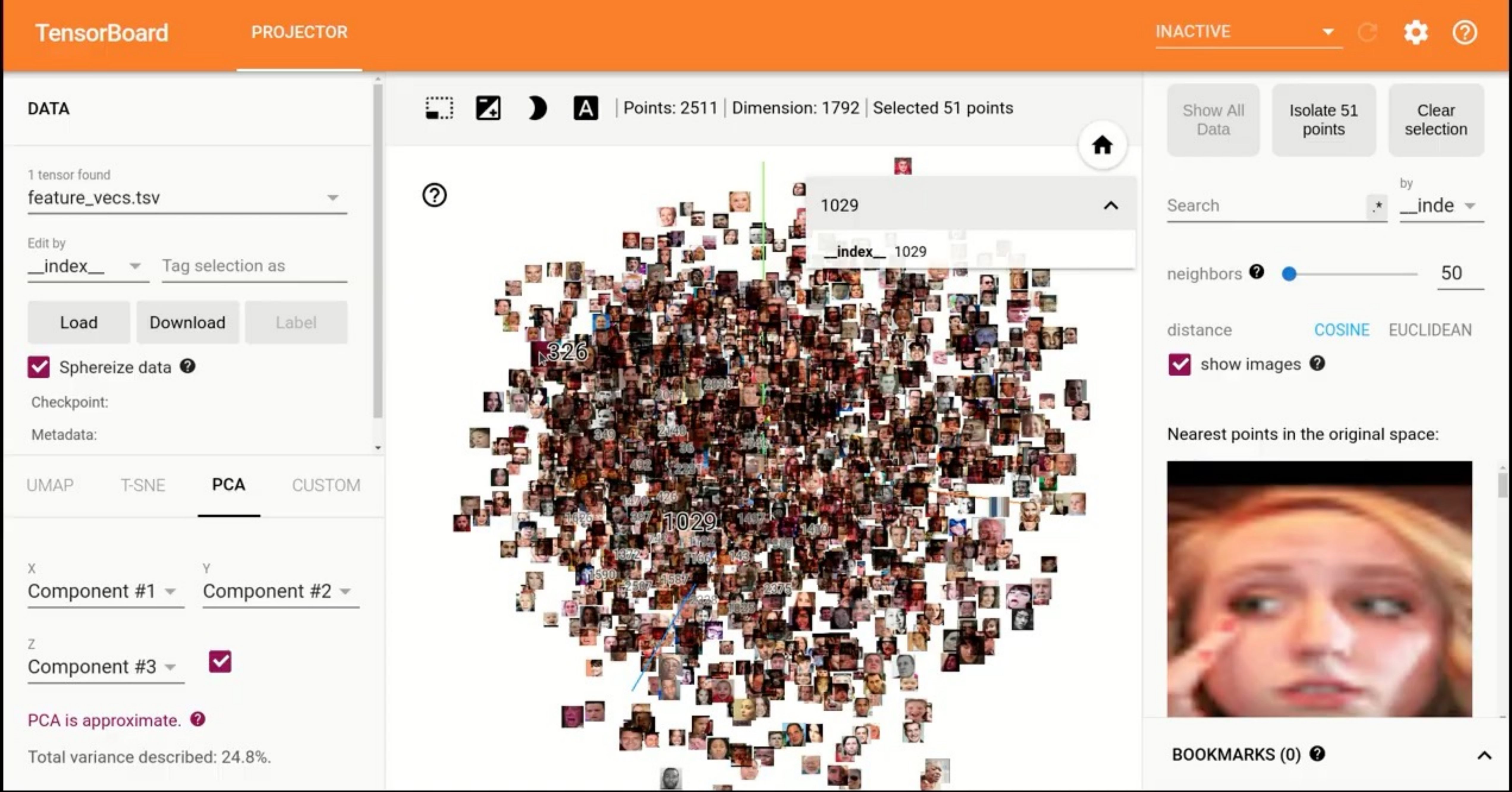
立志 成才 报国 裕民

Docs Dataset



Result of MDS on docs dataset by GlimmerJS

Ingram S, Munzner T, Olano M. Glimmer: Multilevel MDS on the GPU. IEEE Transactions on Visualization and Computer Graphics (2009).


TensorBoard
PROJECTOR
INACTIVE
DATA
1 tensor found
feature_vecs.tsv
Edit by
__index__
Tag selection as
Load
Download
Label
Sphereize data
Checkpoint:
Metadata:
UMAP
T-SNE
PCA
CUSTOM
X
Component #1
Y
Component #2
Z
Component #3
PCA is approximate.
Total variance described: 24.8%.
Points: 2511 | Dimension: 1792 | Selected 51 points
1029
__index__ 1029
Show All Data
Isolate 51 points
Clear selection
Search
by
__inde
neighbors
50
distance
COSINE
EUCLIDEAN
show images
Nearest points in the original space:
BOOKMARKS (0)

VAST PAPER

Visualizing the Hidden Activity of Artificial Neural Networks

Paulo E. Rauber, Samuel G. Fadel, Alexandre X. Falcão, Alexandru C. Telea

VAST PAPER

TPFlow: Progressive Partition and Multidimensional Pattern Extraction for Large-Scale Spatio-Temporal Data Analysis

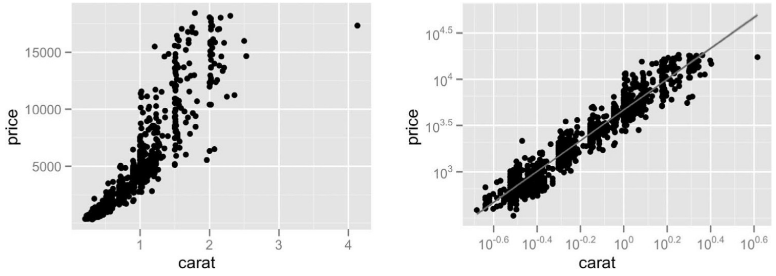
Dongyu Liu, Panpan Xu, Liu Ren

Scatter-plot Matrix

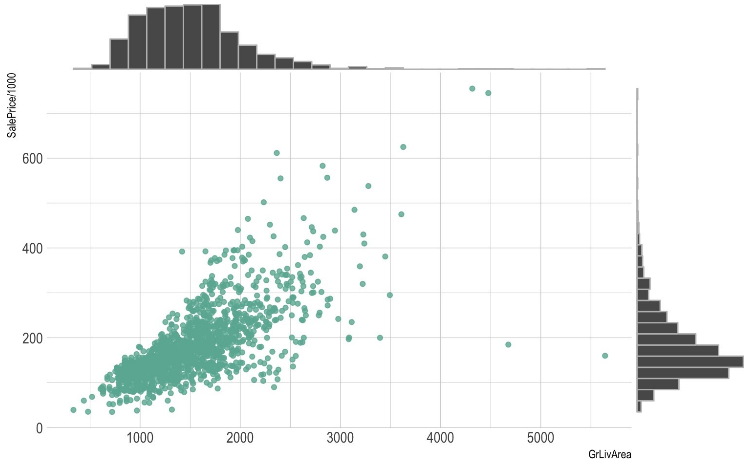
Scatterplot

- Two attributes projected along the x- and y-axis
- Express values
- Quantitative attributes
- No keys, only values
- Data: two quantitative attributes
- Mark: points
- Channels: horizontal and vertical position
- Tasks: identify trends, outliers, distribution, correlation, clusters
- Scalability: hundreds of items

➔ Express Values

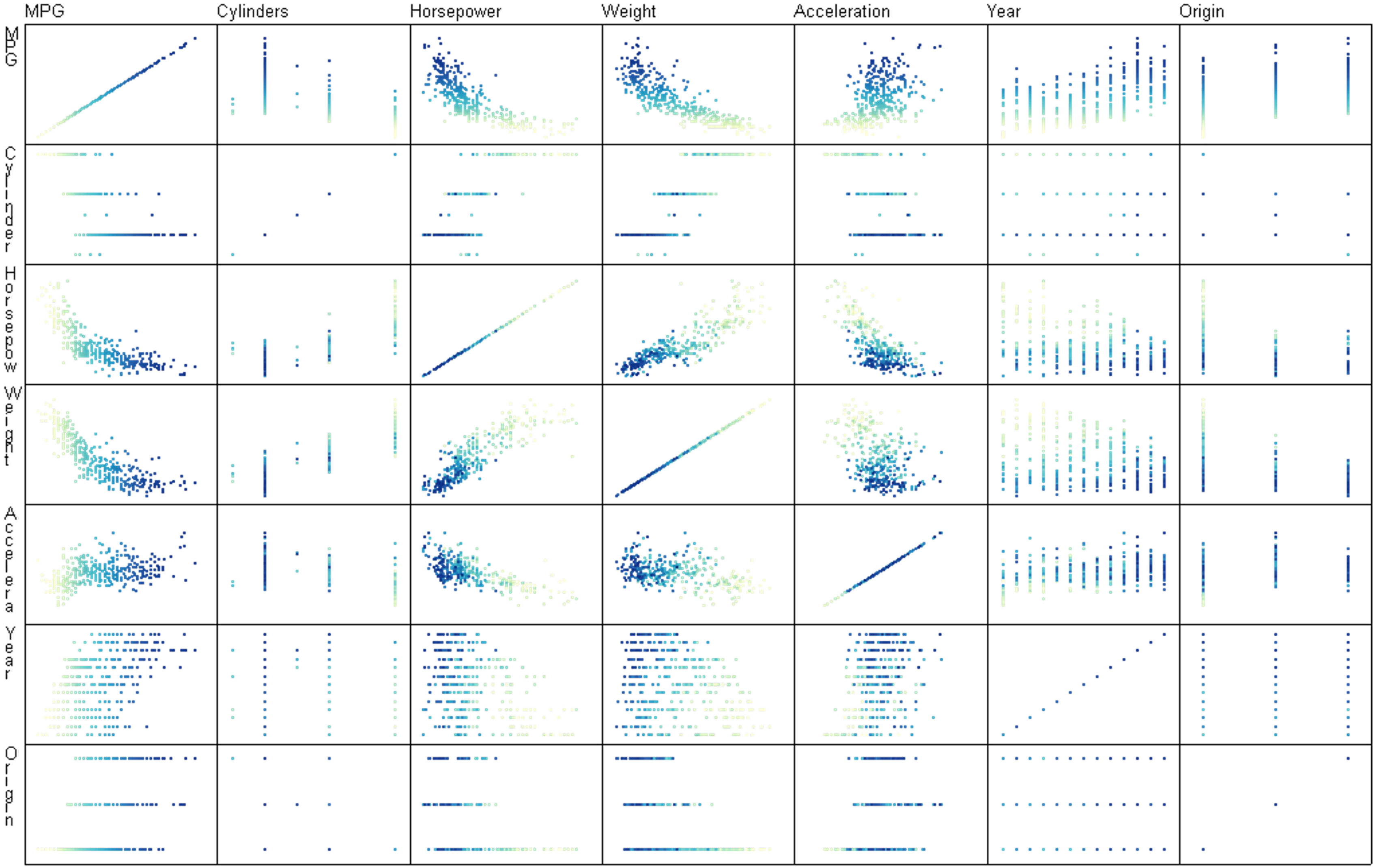


[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]



https://www.data-to-viz.com/graph/scatter.html

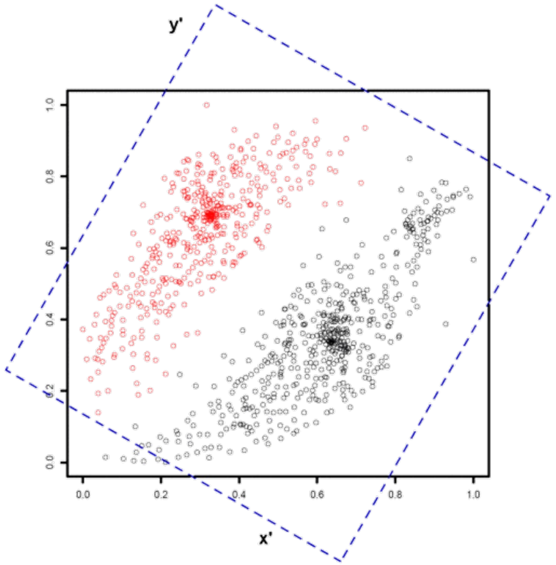
Scatter-plot Matrix



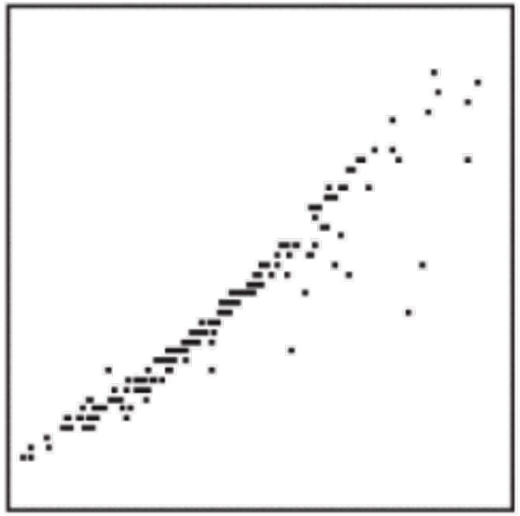
Automatically Exploration

- Recommend scatter-plots with interesting patterns automatically.

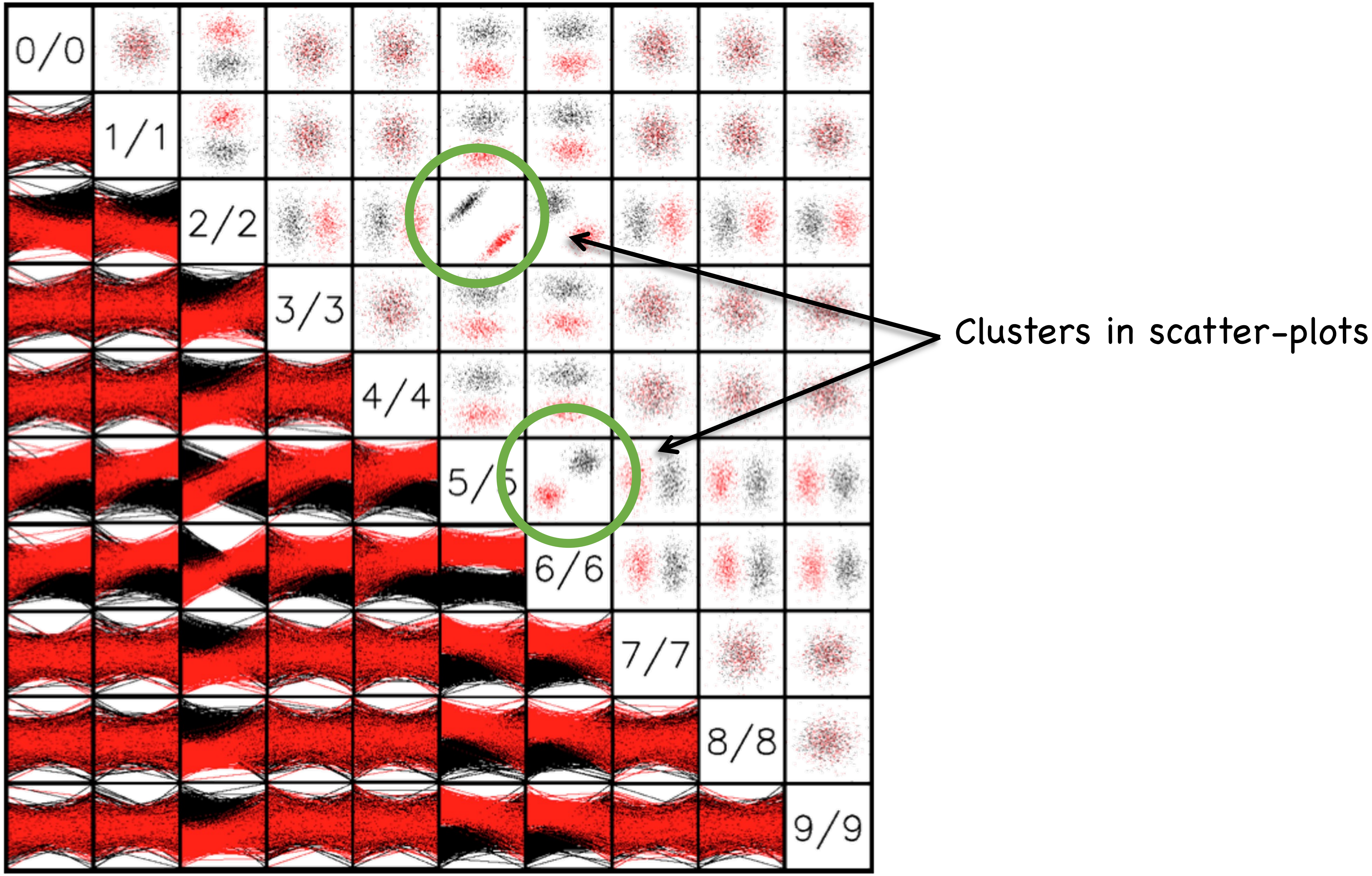
Clusters



Correlations

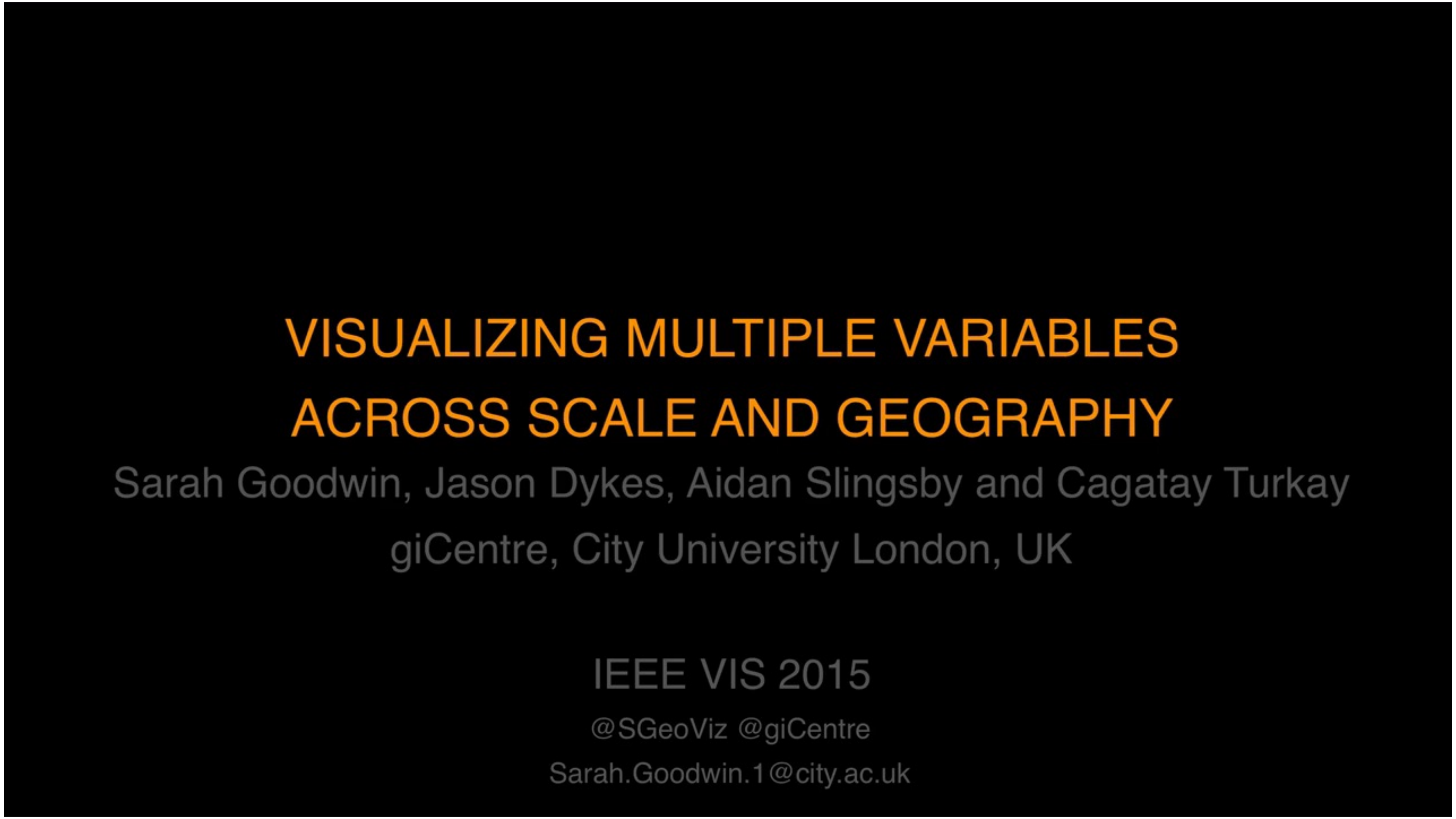


Tatu A, Albuquerque G, Eisemann M. "Automated analytical methods to support visual exploration of high-dimensional data." IEEE Transactions on Visualization and Computer Graphics (2011).


0/0
1/1
2/2
3/3
4/4
5/5
6/6
7/7
8/8
9/9
Clusters in scatter-plots

Across Scale and Geography

上海科技大学
ShanghaiTech University



Goodwin S., Dykes J., Slingsby A. "Visualizing multiple variables across scale and geography."
IEEE Transactions on Visualization and Computer Graphics (2016).

立志成才 报国裕民

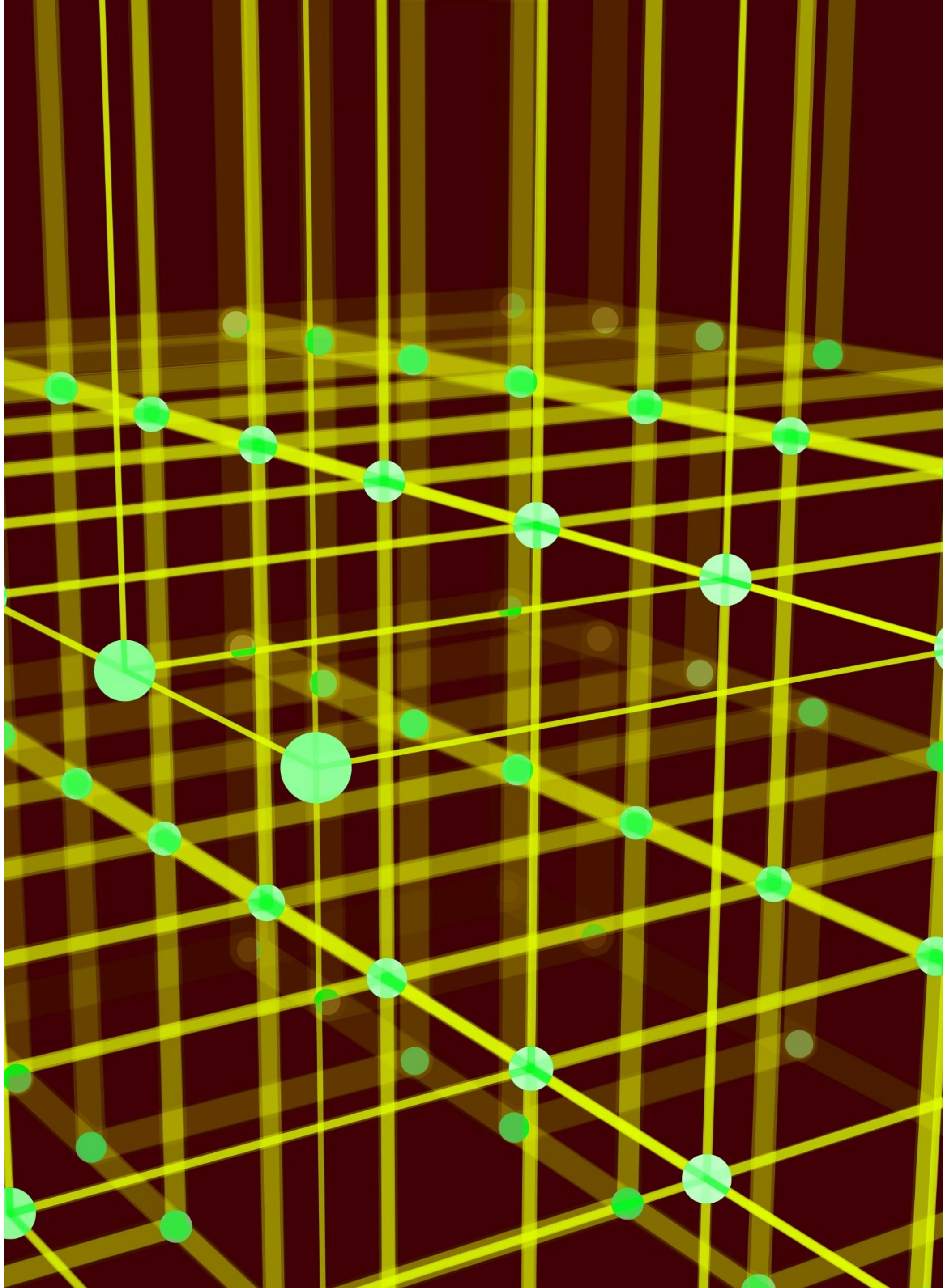
Animated Scatter-plot Matrices

Chen H, Engle S, Joshi A. "Using Animation to Alleviate Overdraw in Multiclass Scatterplot Matrices."
ACM CHI Conference on Human Factors in Computing Systems (2018).

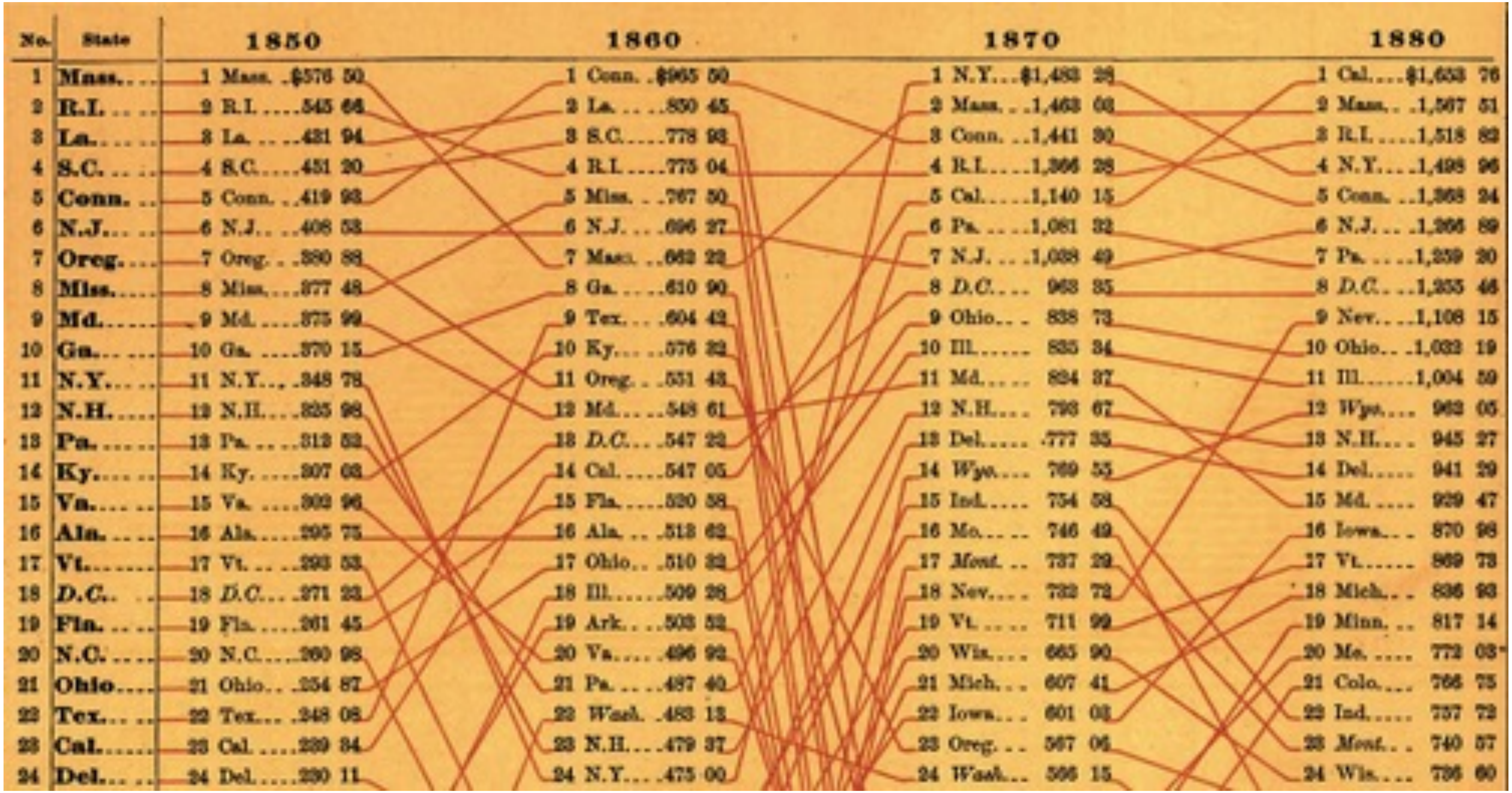
Parallel Coordinates

Parallel Coordinates

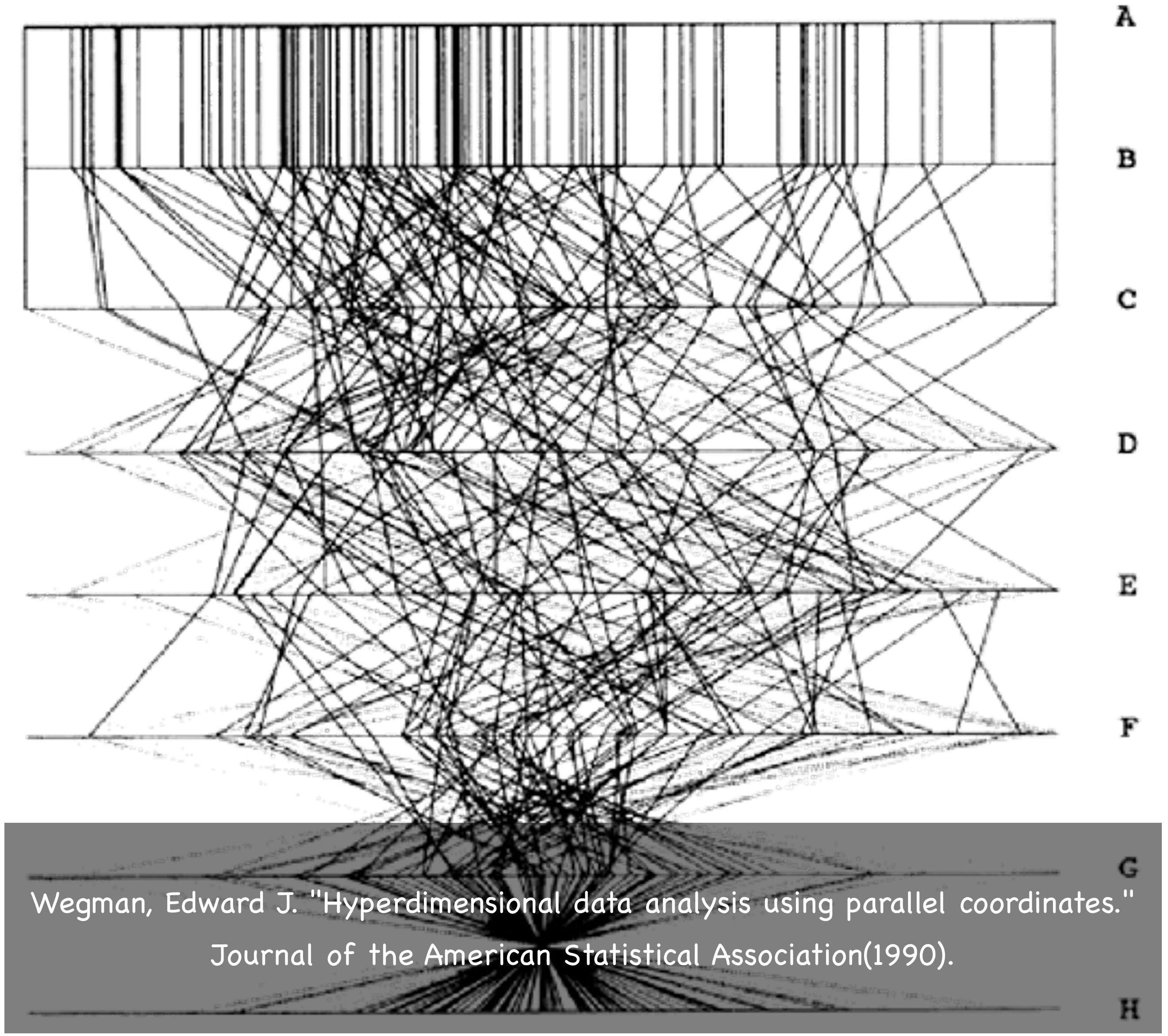
- Presented by Inselberg in 1985 for high-dimensional geometry.
- Parallel axes.
- Data points represented by lines.



Parallel Coordinates in 1880

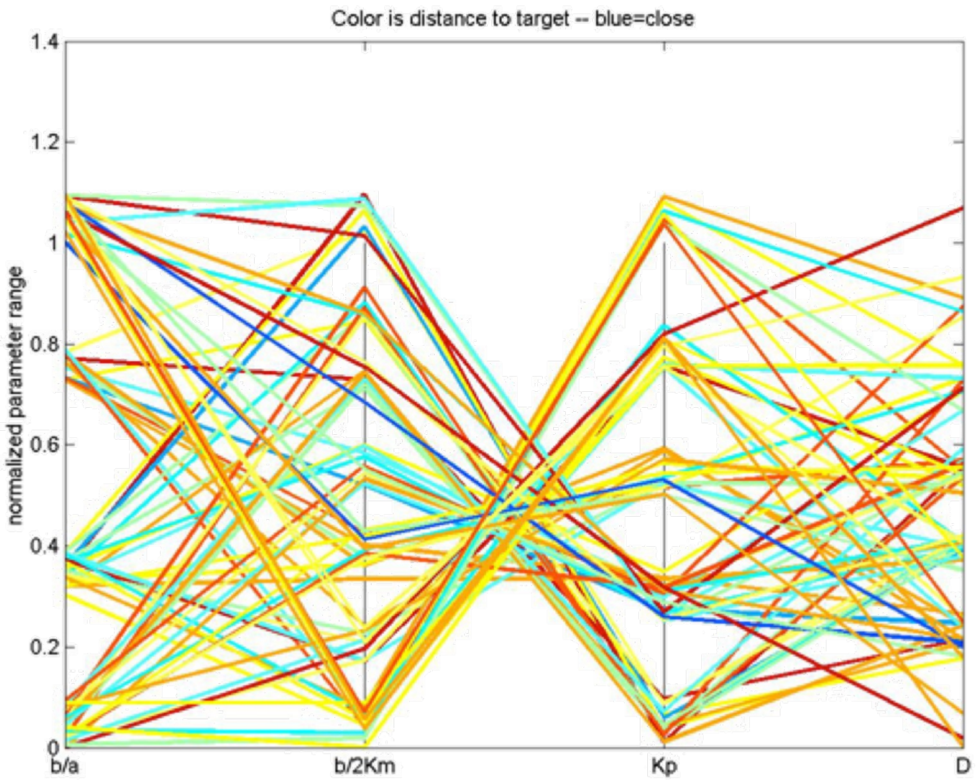


Correlations

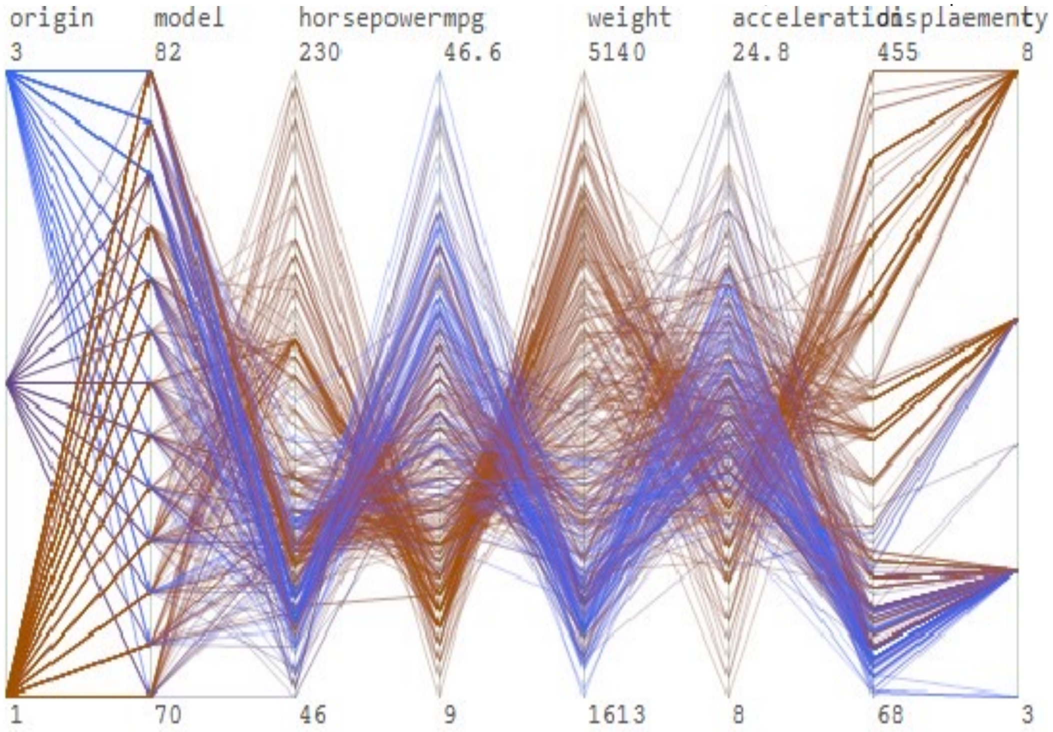


Wegman, Edward J. "Hyperdimensional data analysis using parallel coordinates." Journal of the American Statistical Association(1990).

Color

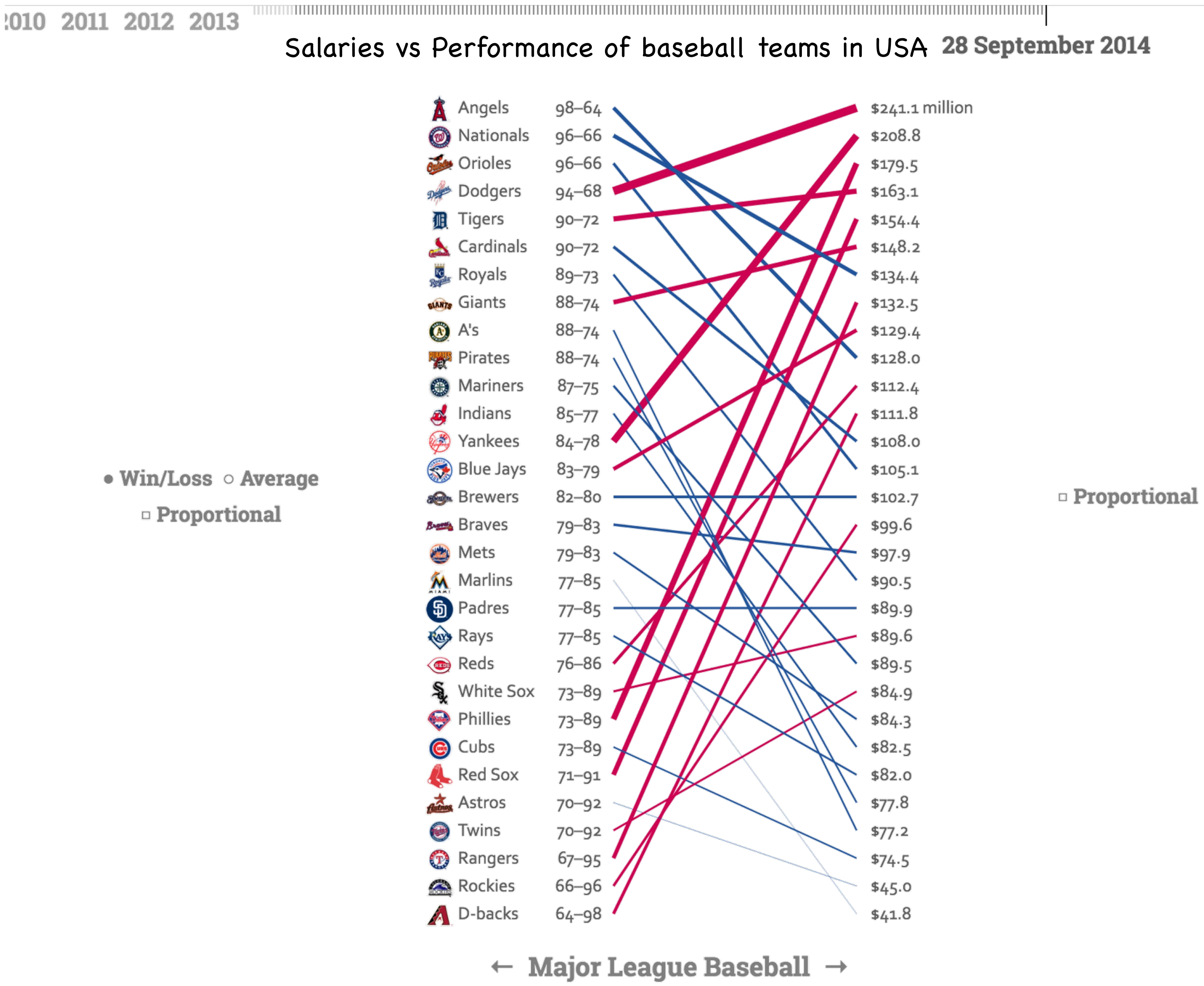


Color is distance to target -- blue=close
normalized parameter range
1.4
1.2
1
0.8
0.6
0.4
0.2
0
b/a
b/2Km
Kp
D

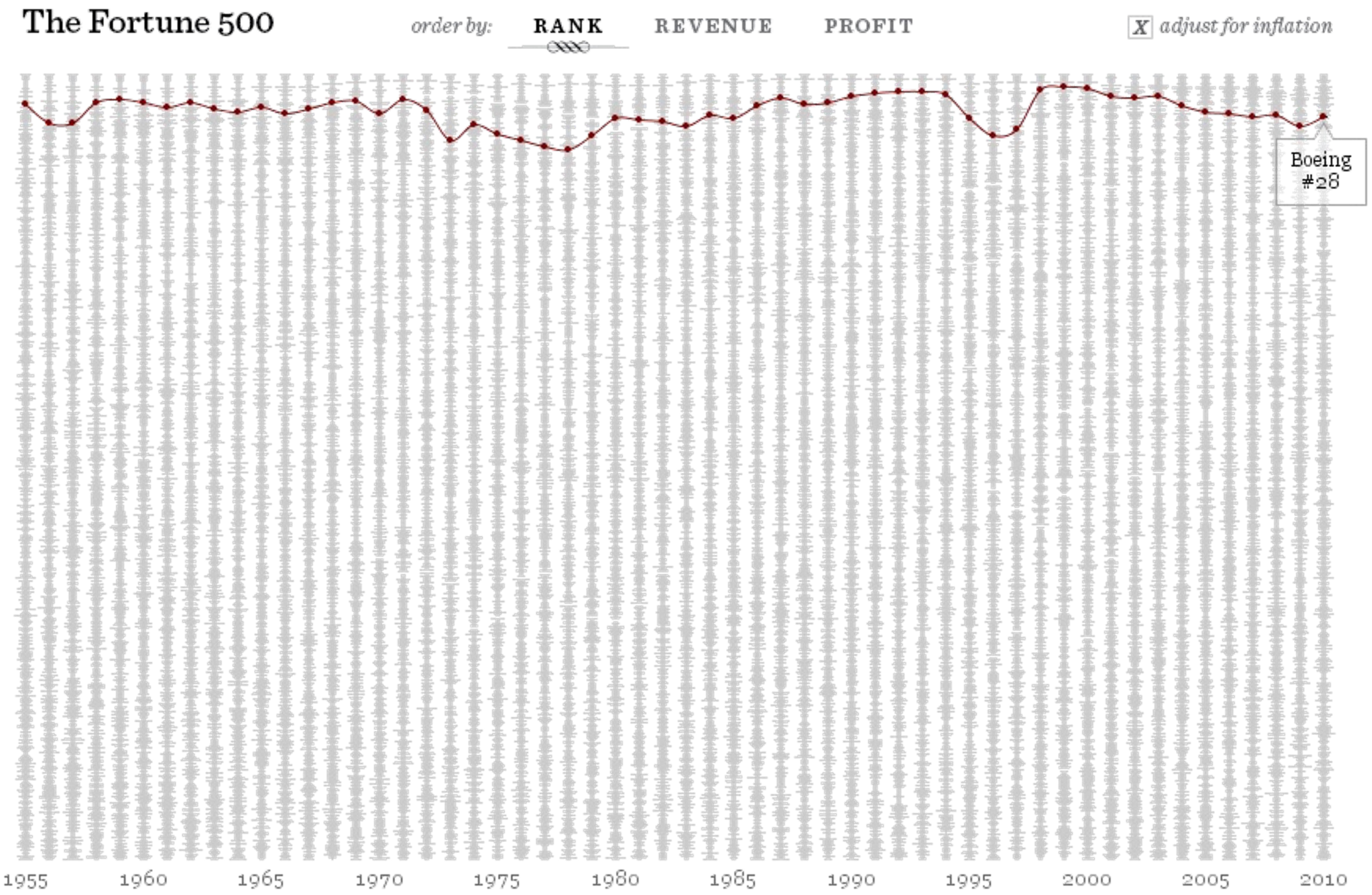


origin
model
horsepower
mpg
weight
acceleration
displacement
cylinders
3
82
230
46.6
5140
24.8
455
8
1
70
46
9
1613
8
68
3

Salaries vs Performance of baseball teams in USA

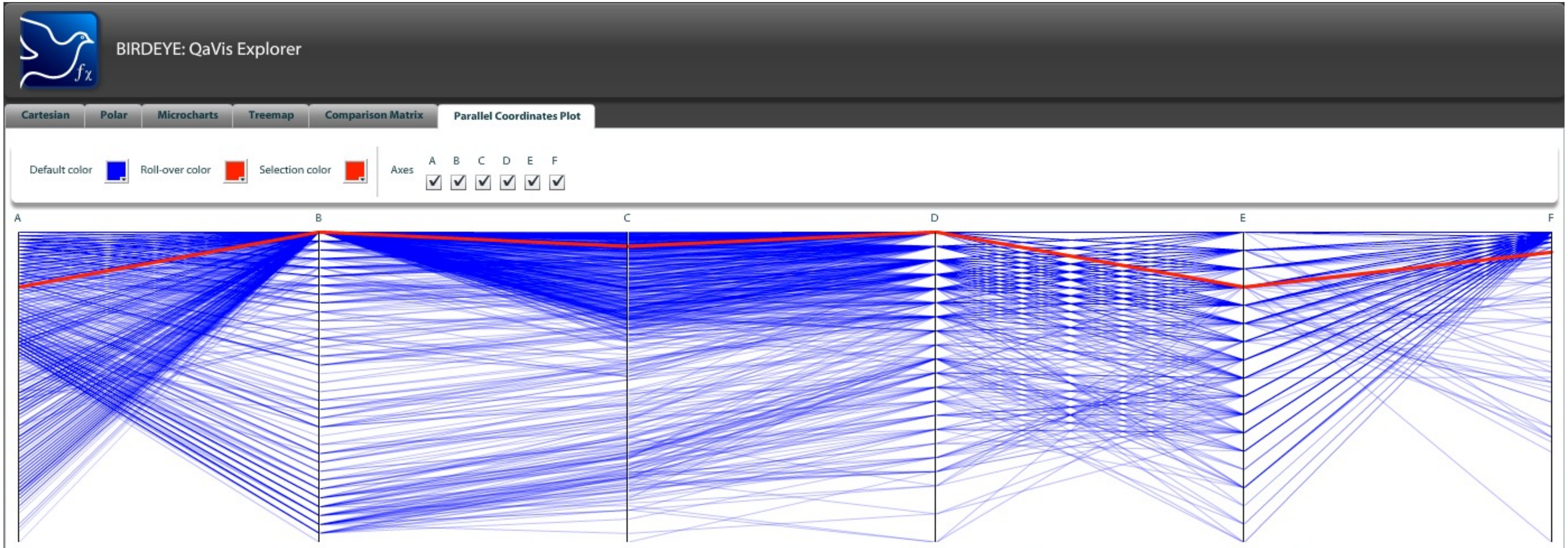


https://fathom.info/salaryper/



World Top 500 Enterprises Ranking

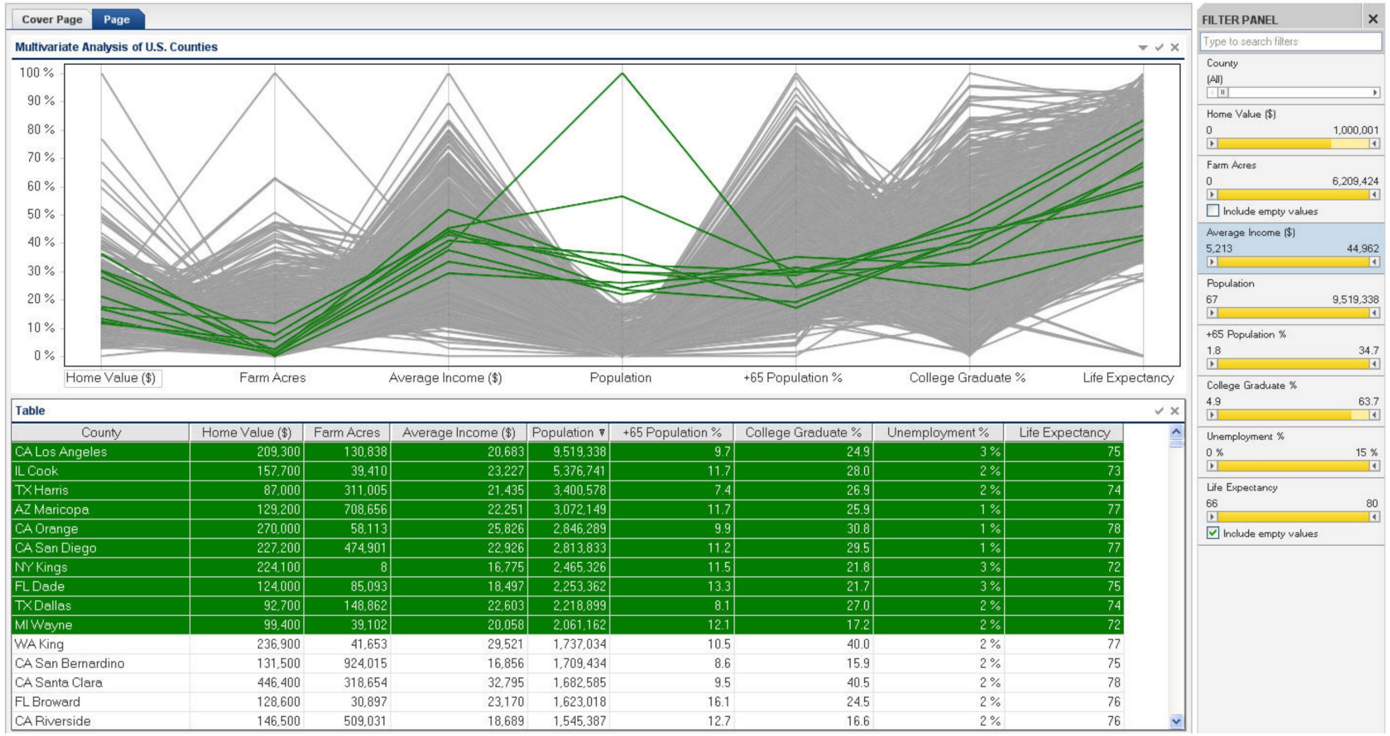
Selection



A	B	C	D	E	F
47	0	48.7	4	2	0
16	0	12.3	0	3	4
6	4	30	1	0	0
46	0	53	3	5	3
54	0	65.3	2	3	0

Brush and Filter

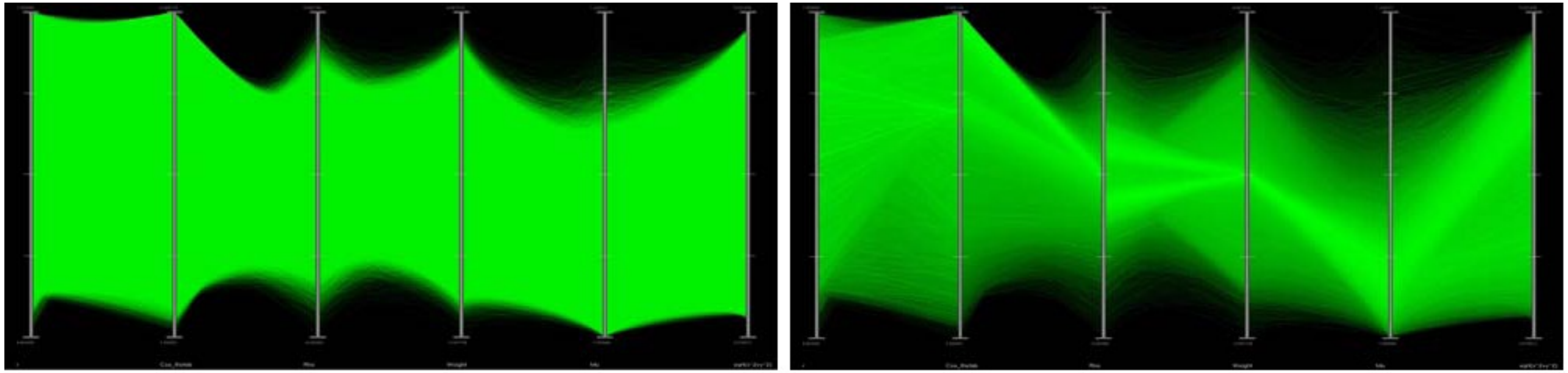
- Multivariate Analysis Using Parallel Coordinates. Stephen Few September 12, 2006
- https://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf
- Created using Spotfire DXP



County	Home Value ($)	Farm Acres	Average Income ($)	Population	+65 Population %	College Graduate %	Unemployment %	Life Expectancy
CA Los Angeles	209,300	130,838	20,683	9,519,338	9.7	24.9	3 %	75
IL Cook	157,700	39,410	23,227	5,376,741	11.7	28.0	2 %	73
TX Harris	87,000	311,005	21,435	3,400,578	7.4	26.9	2 %	74
AZ Maricopa	129,200	708,656	22,251	3,072,149	11.7	25.9	1 %	77
CA Orange	270,000	58,113	25,826	2,846,289	9.9	30.8	1 %	78
CA San Diego	227,200	474,901	22,926	2,813,833	11.2	29.5	1 %	77
NY Kings	224,100	8	16,775	2,465,326	11.5	21.8	3 %	72
FL Dade	124,000	85,093	18,497	2,253,362	13.3	21.7	3 %	75
TX Dallas	92,700	148,862	22,603	2,218,899	8.1	27.0	2 %	74
MI Wayne	99,400	39,102	20,058	2,061,162	12.1	17.2	2 %	72
WA King	236,900	41,653	29,521	1,737,034	10.5	40.0	2 %	77
CA San Bernardino	131,500	924,015	16,856	1,709,434	8.6	15.9	2 %	75
CA Santa Clara	446,400	318,654	32,795	1,682,585	9.5	40.5	2 %	78
FL Broward	128,600	30,897	23,170	1,623,018	16.1	24.5	2 %	76
CA Riverside	146,500	509,031	18,689	1,545,387	12.7	16.6	2 %	76

Transparent Parallel Coordinates

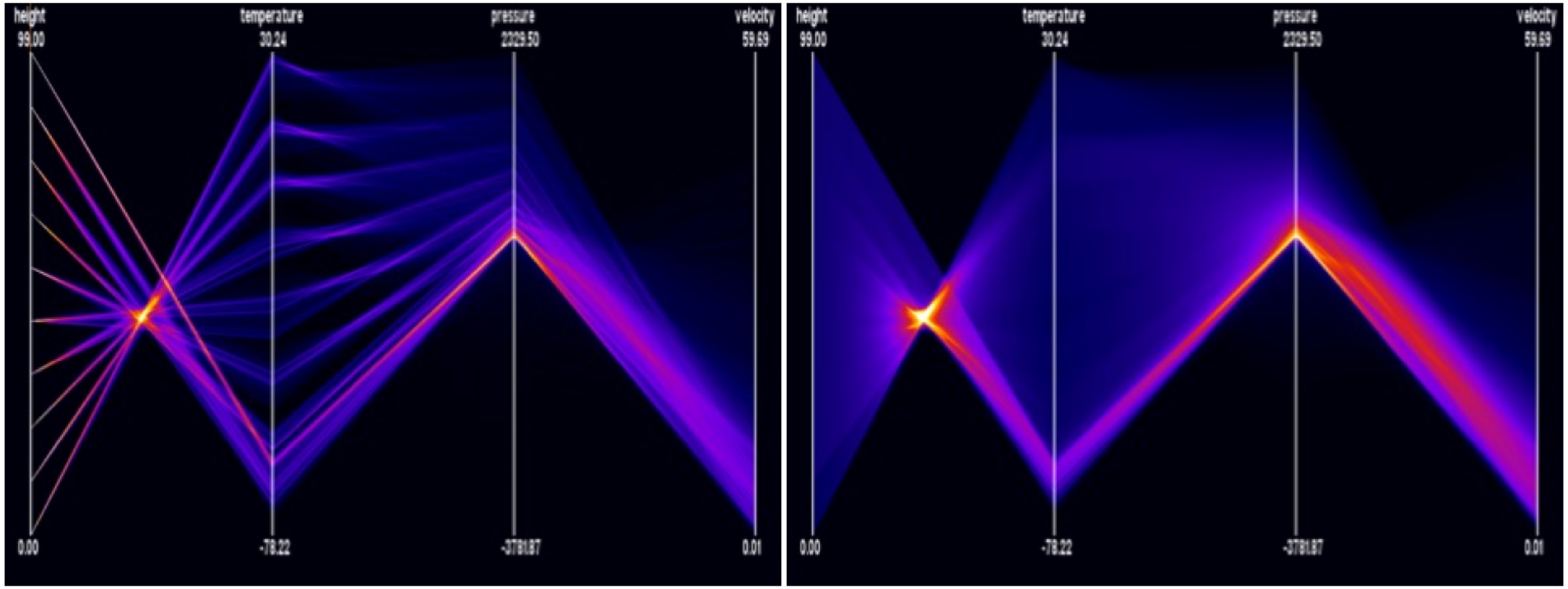
Emphasize main trends



Chad Jones et al. "An Integrated Exploration Approach to Visualizing Multivariate Particle Data."
Computing in Science & Engineering (2008).

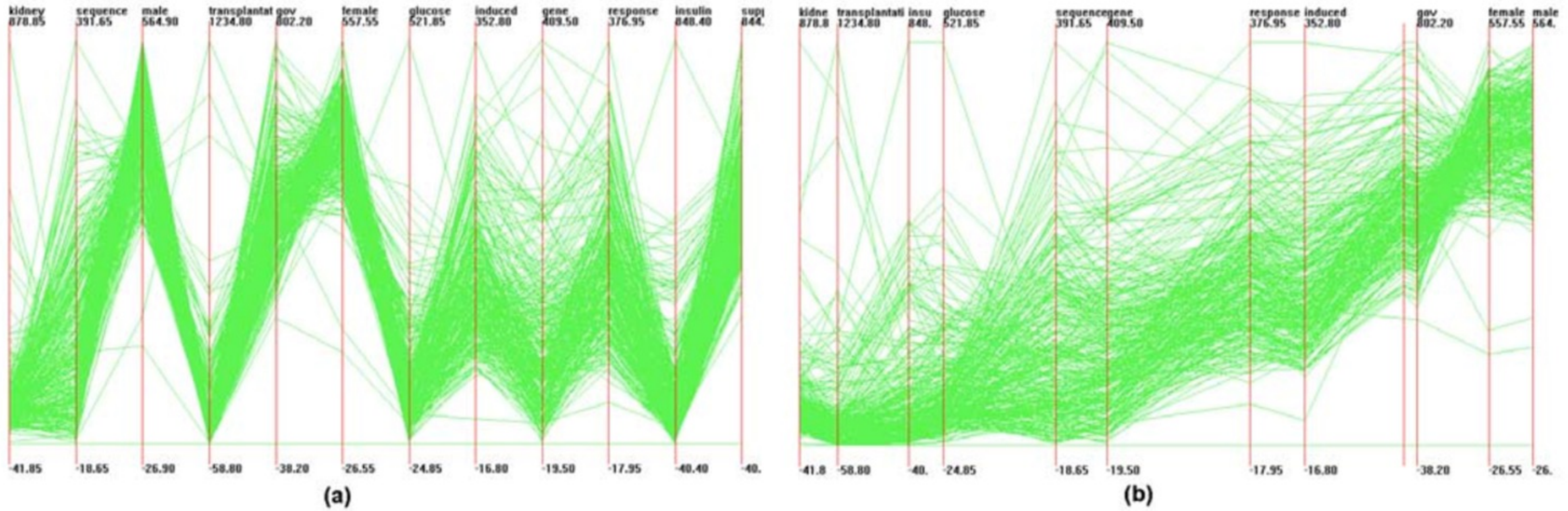
Continuous Parallel Coordinates

Use heatmap to show the trends



J. Heinrich and D. Weiskopf. "Continuous Parallel Coordinates." IEEE Transactions on Visualization and Computer Graphics (2009).

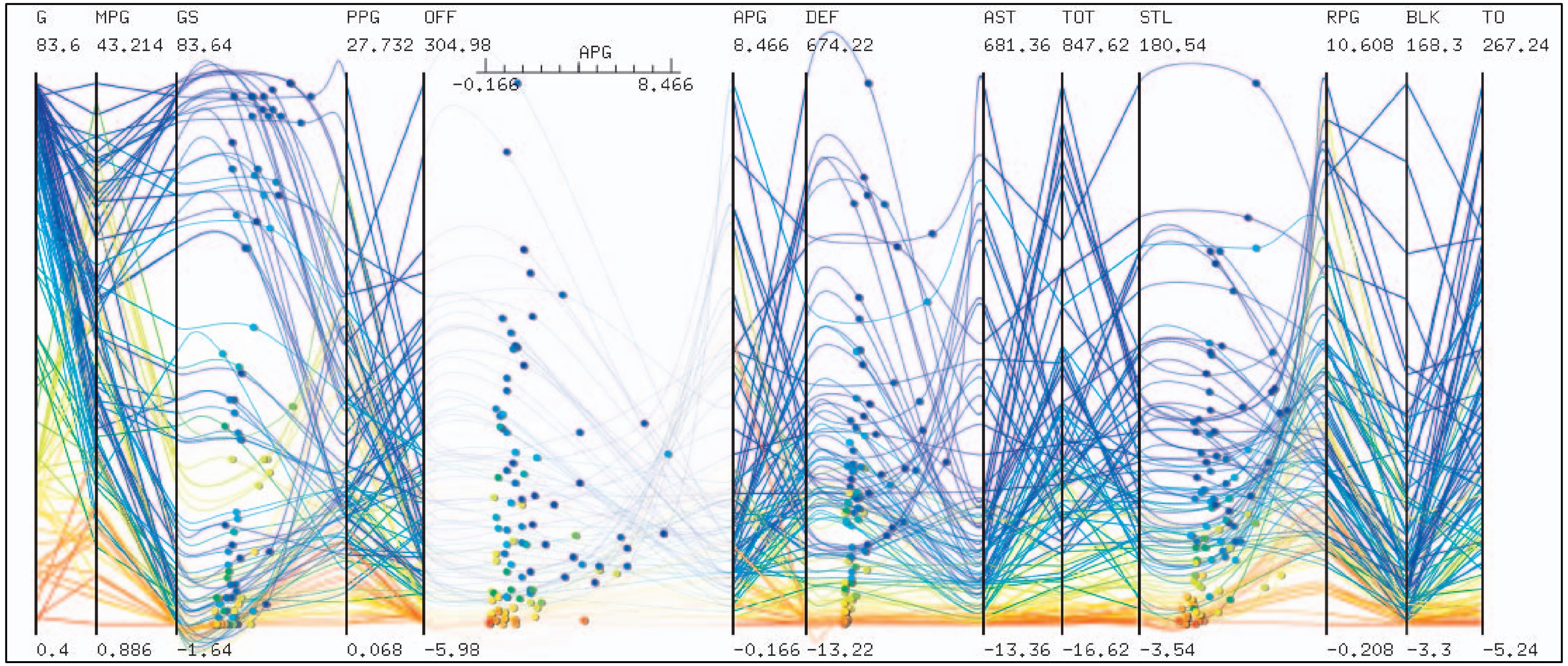
Re-ordering Axes



Peng, Wei , M. O. Ward et al. "Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Re-ordering." IEEE Symposium on Information Visualization (2005).

Parallel Coordinates with Scatter-plots

上海科技大学 ShanghaiTech University



Yuan et al. "Scattering Points in Parallel Coordinates."
IEEE Transactions on Visualization and Computer Graphics (2009).

立志成才 报国裕民

Scattering Points in Parallel Coordinates

Submitted to IEEE Infovis 2009

FLINAView

Flexible LINked Axes for Multivariate Data visualization

Jarry H.T. Claessen
Jarke J. Van Wijk

IEEE InfoVis 2011

上海科技大学
ShanghaiTech University

VIS 2015

VAST • INFOVIS • SCIVIS

25–30 October 2015
CHICAGO, ILLINOIS, USA

Orientation-Enhanced Parallel Coordinate Plots

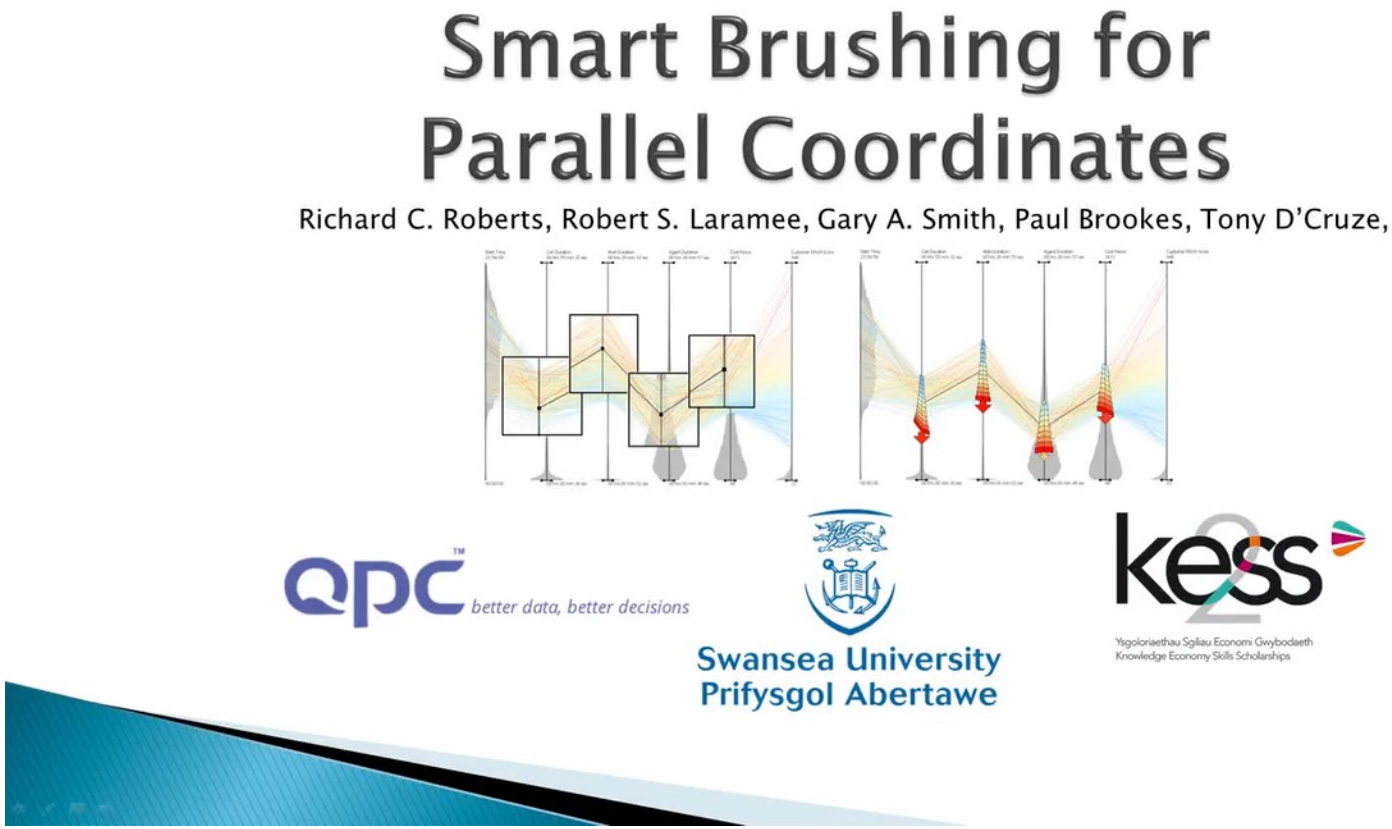
R.G. Raidou, M. Eisemann, M. Breeuwer,
E. Eisemann, A. Vilanova

TU/e

TU Delft

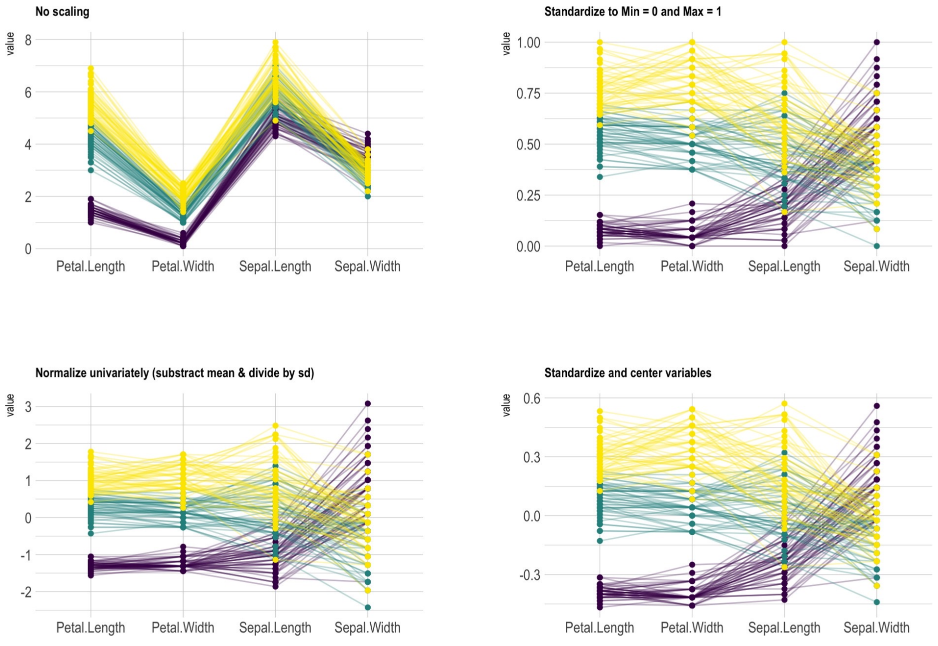
Technology Arts Sciences
TH Köln

Smart Brushing for Parallel Coordinates



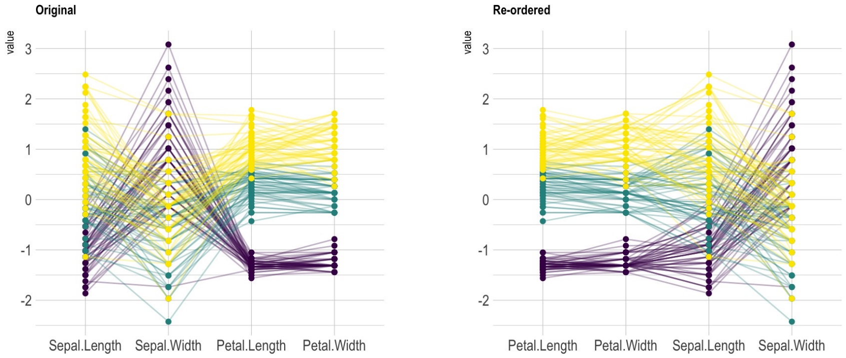
Roberts R, Laramee R S, Smith G A, et al. Smart Brushing for Parallel Coordinates[J]. IEEE Transactions on Visualization and Computer Graphics (2018)

Parallel Coordinates: Improving Readability

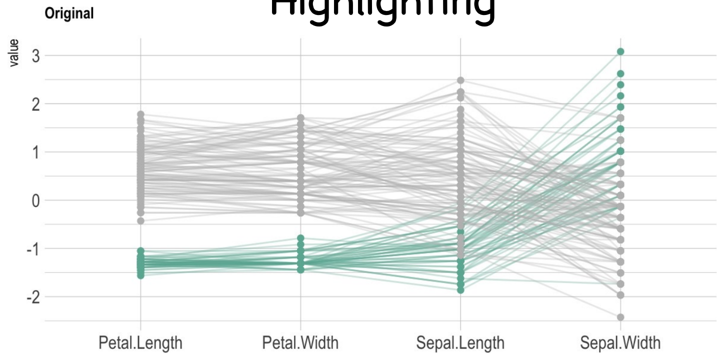


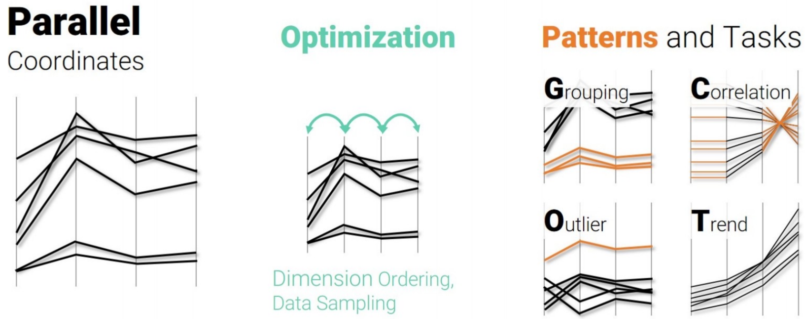
Scaling

Axis Reordering



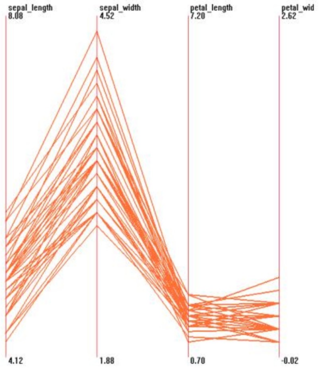
Highlighting



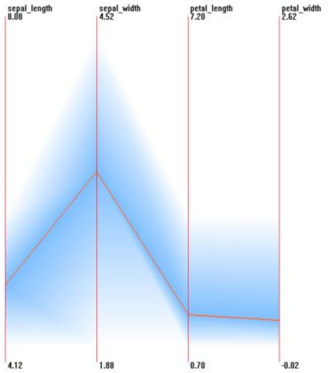


https://www.data-to-viz.com/graph/parallel.html

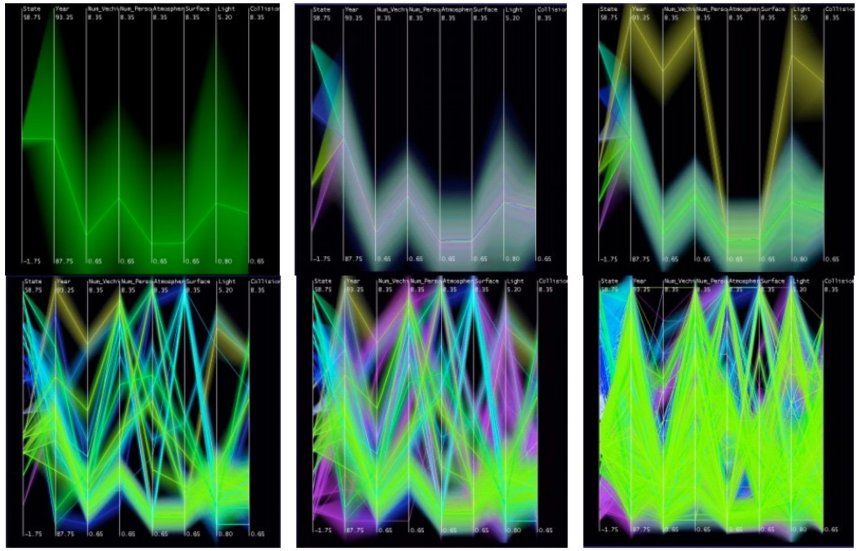
Hierarchical Parallel Coordinates



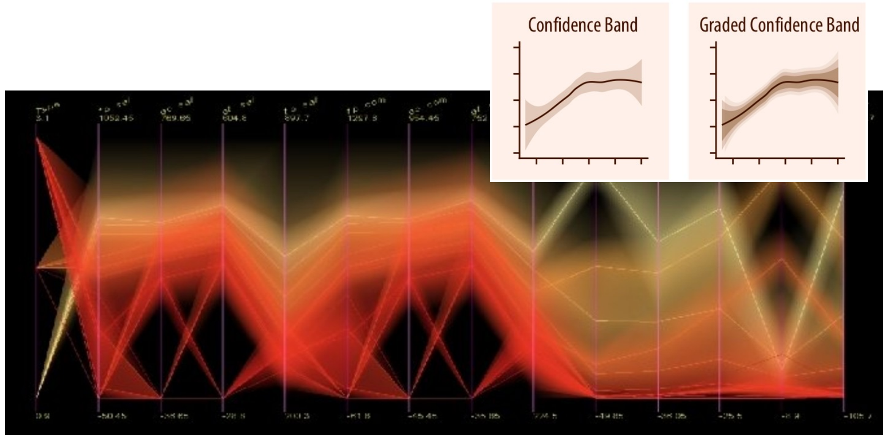
Flat Display



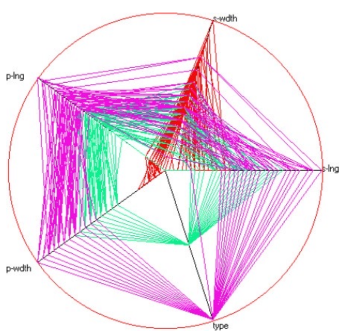
Hierarchical Display

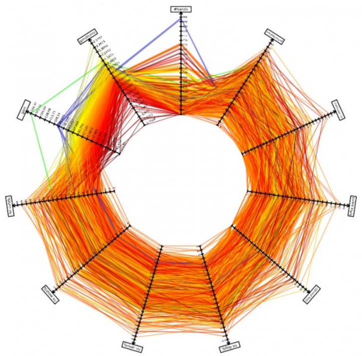


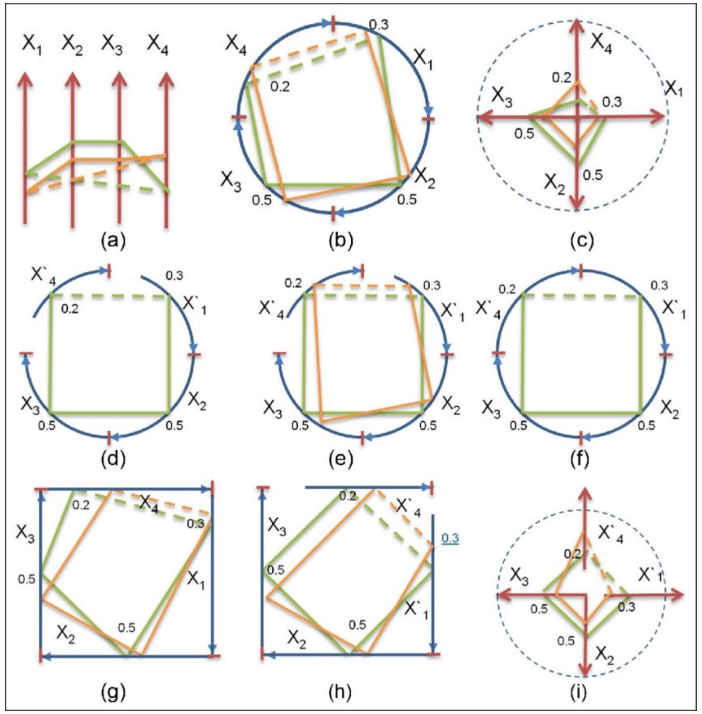
- Statistical aggregation derived from a hierarchical clustering of the data, at different levels of abstraction
  - Bands show cluster extents in each dimension
  - Opacity conveys cluster population
  - Color similarity indicates proximity in hierarchy



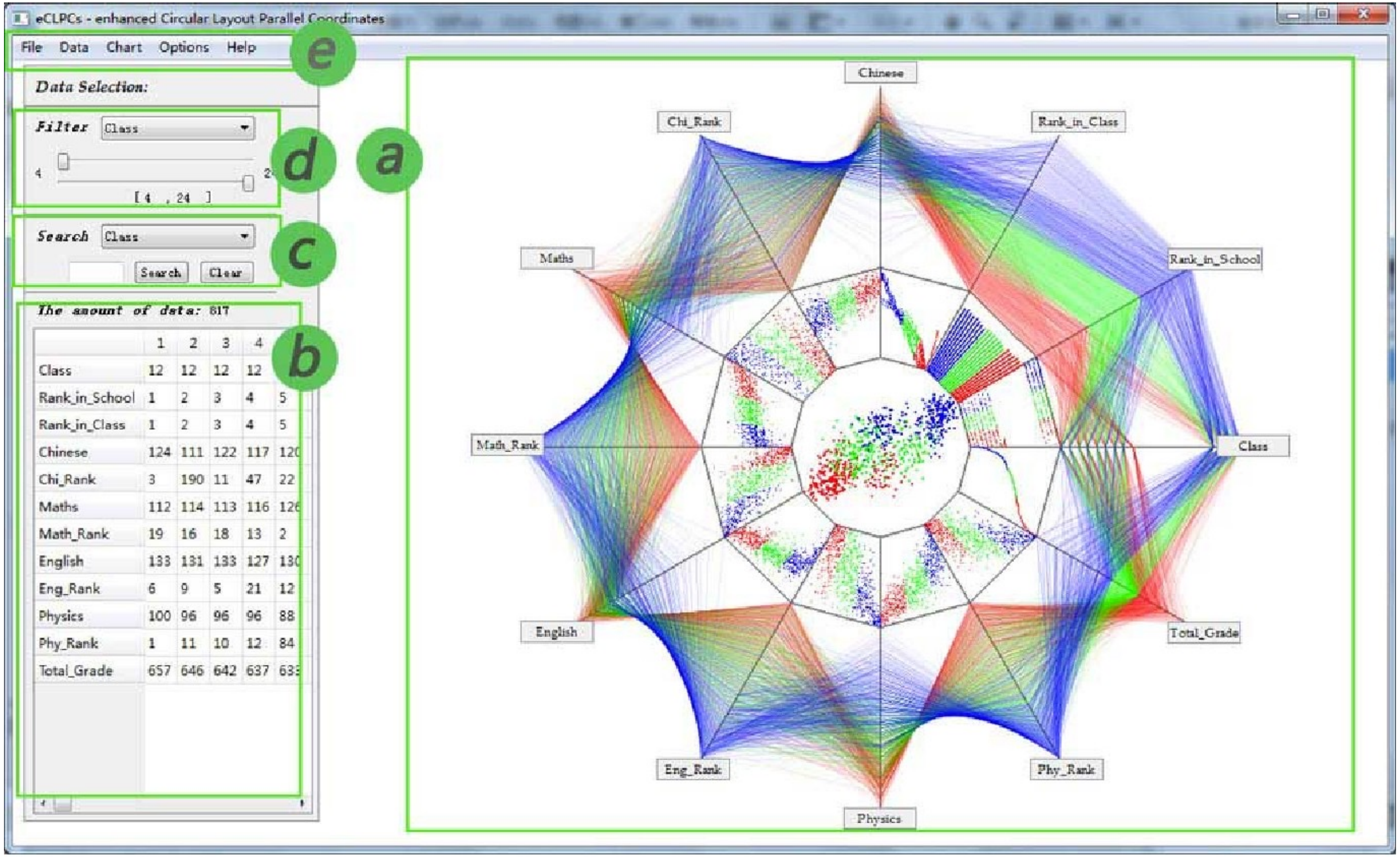
Circular Parallel Coordinates

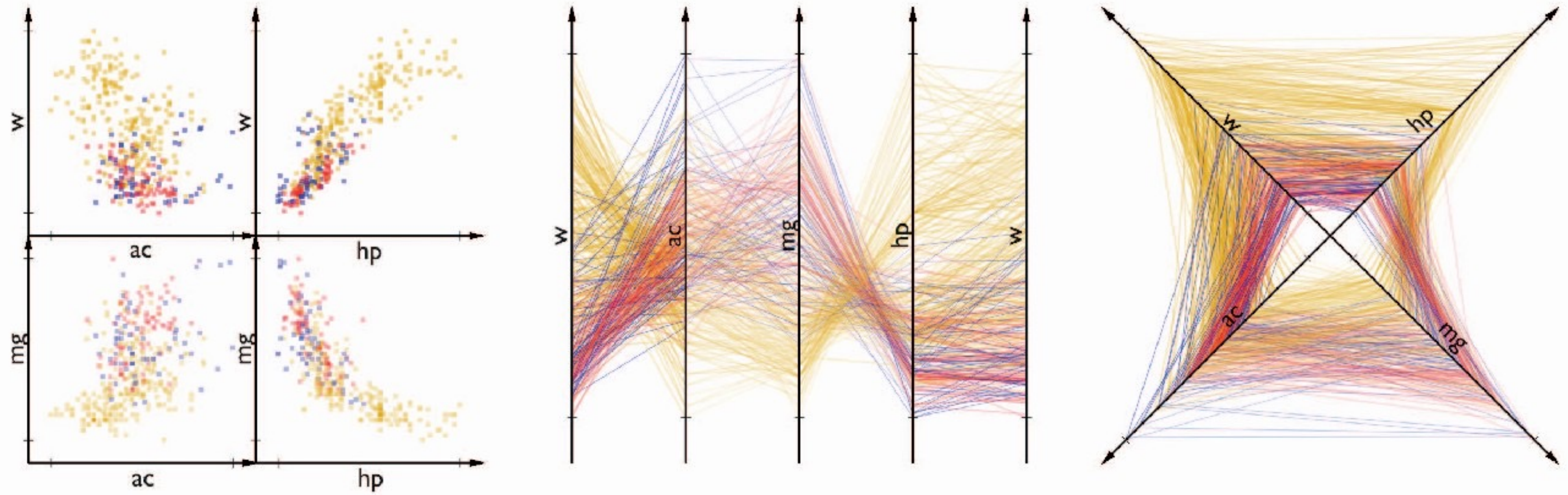




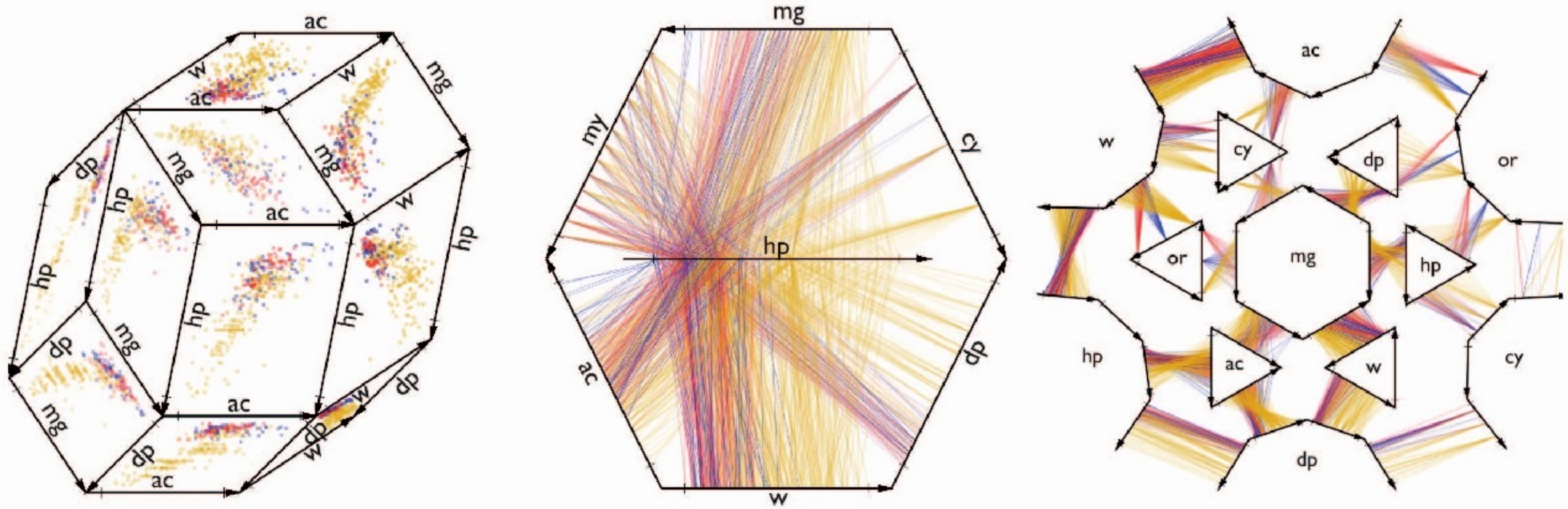


- Possible Arrangement
  - Radial arrangement of axes
  - Wheel arrangement of axes


eCLPCs - enhanced Circular Layout Parallel Coordinates
File Data Chart Options Help
Data Selection:
Filter Class
[4 , 24]
Search Class
Search Clear
The amount of data: 817
Class
Rank_in_School
Rank_in_Class
Chinese
Chi_Rank
Maths
Math_Rank
English
Eng_Rank
Physics
Phy_Rank
Total_Grade
a
b
c
d
e

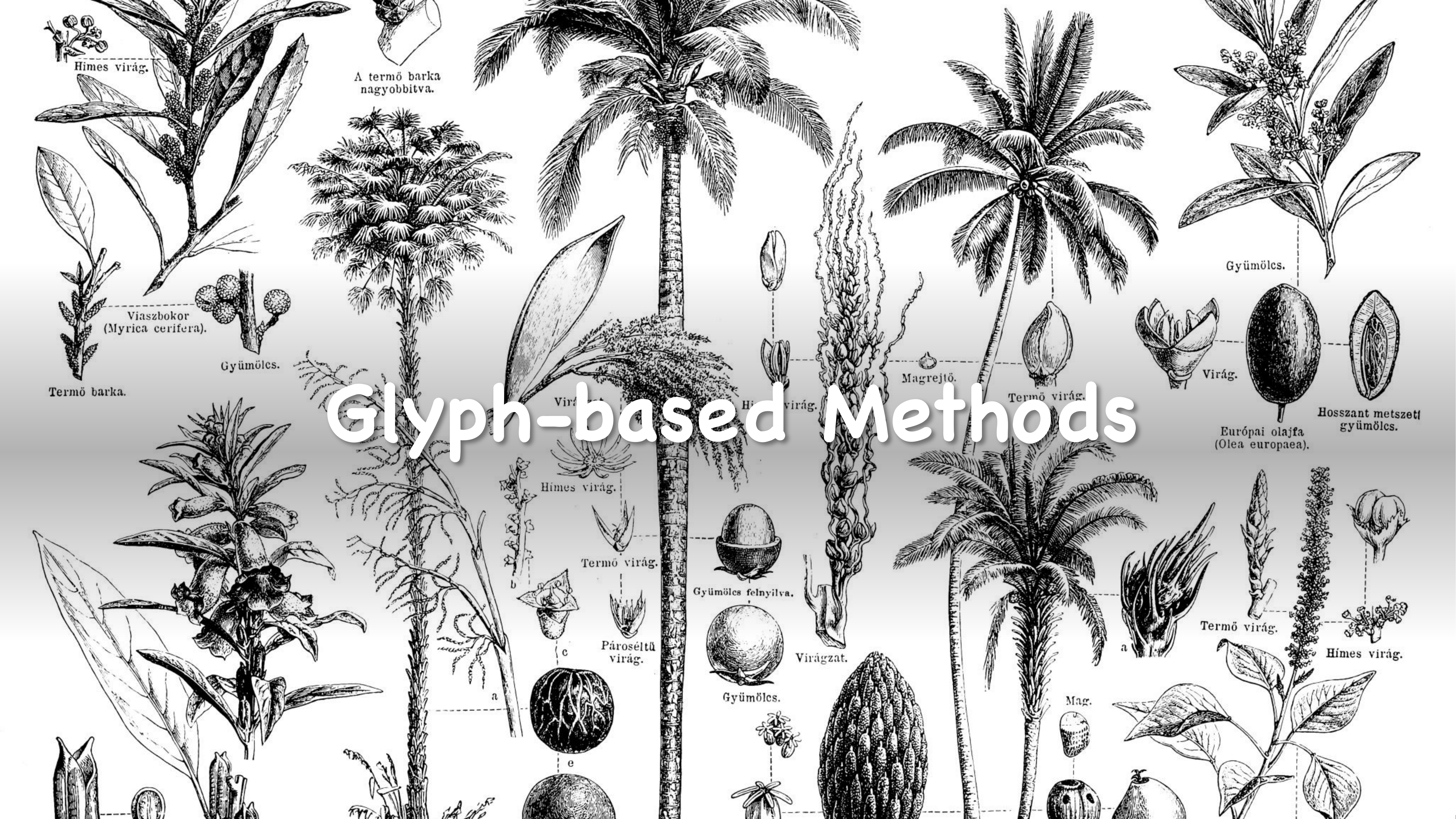


(a) scatterplots (b) Parallel Coordinates Plot (c) radar chart



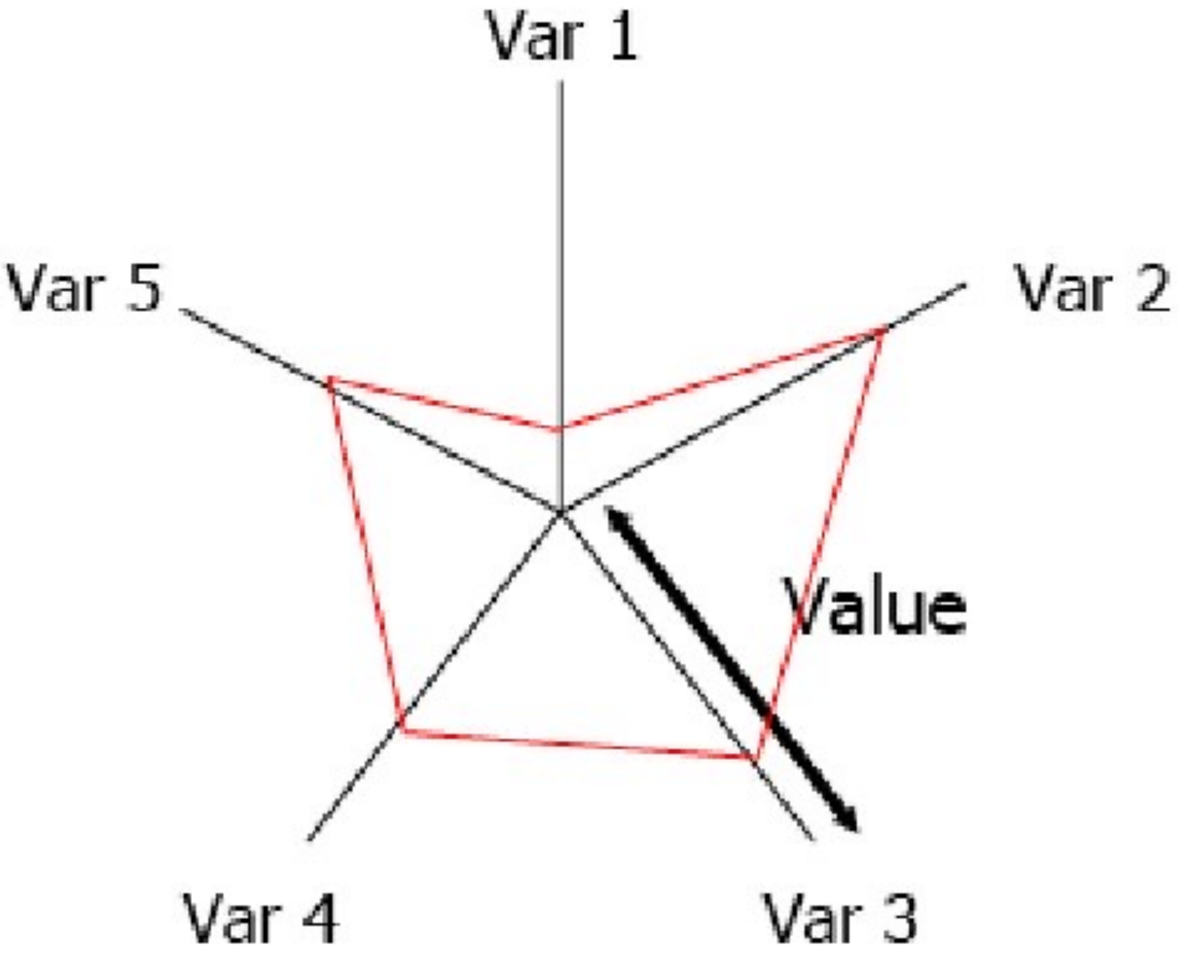
(d) Hyperbox (e) Time Wheel (f) Many-to-many PCP

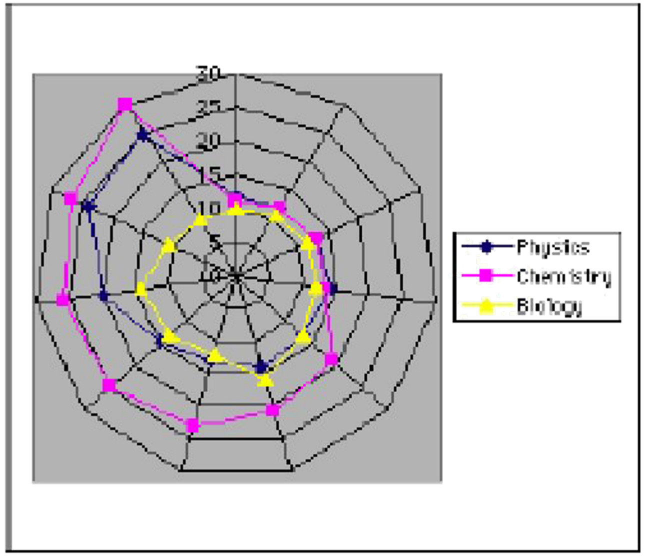
Glyph-based Methods

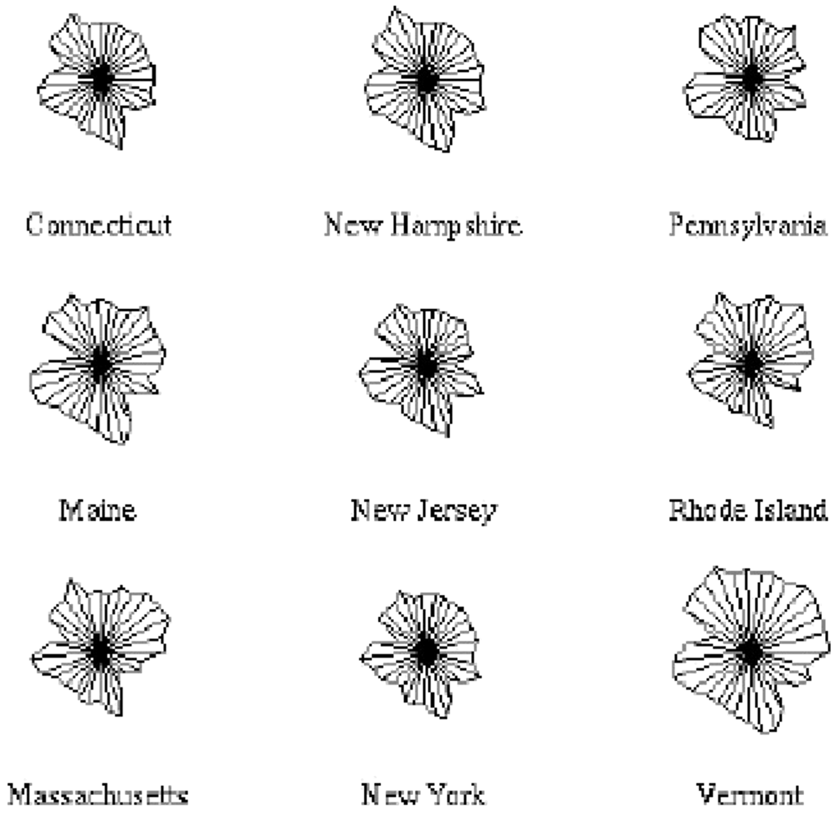


Star Plots

- Space variables around a circle.
- Encode values on "spokes".
- Data point is now a shape.

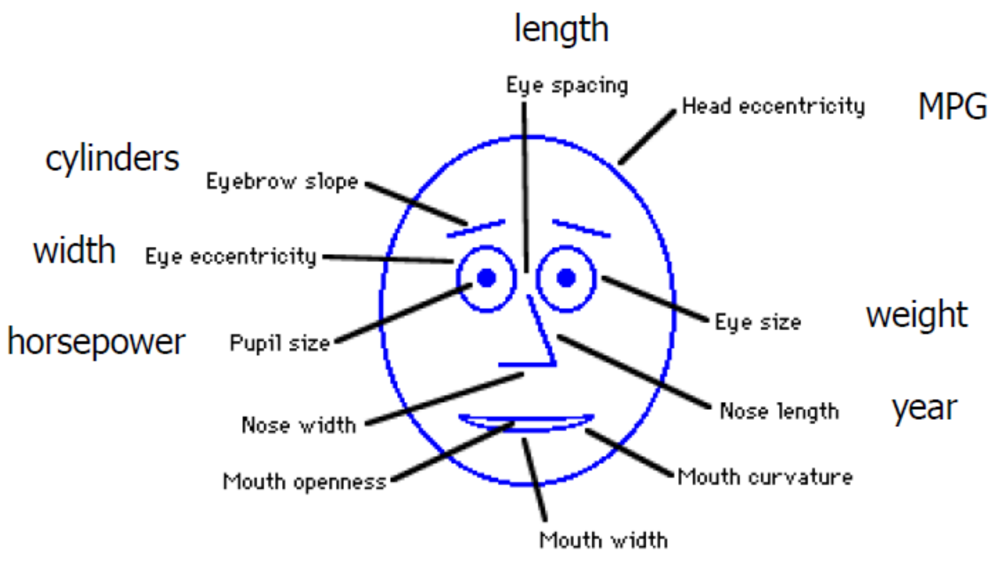




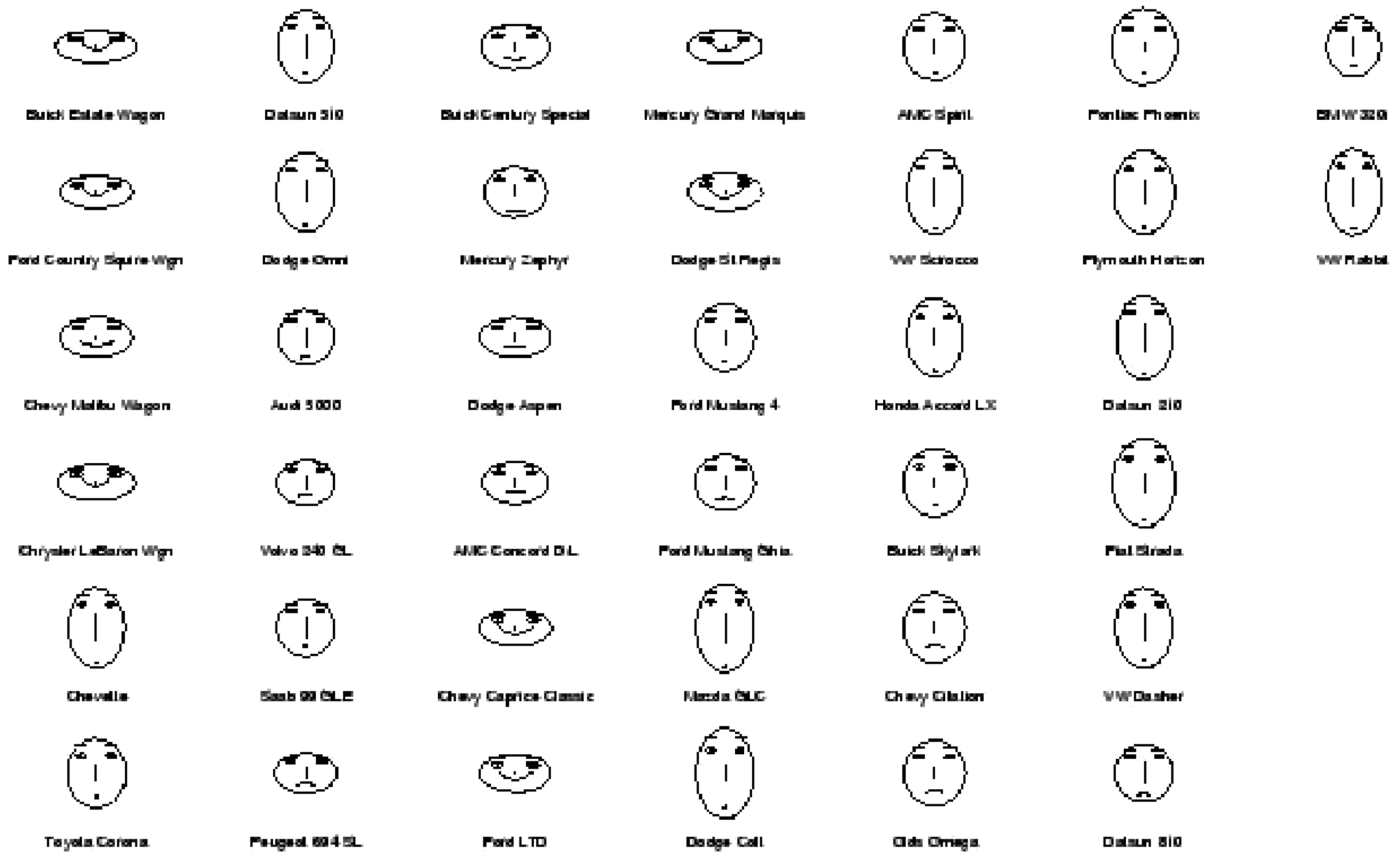


http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/comp1.html

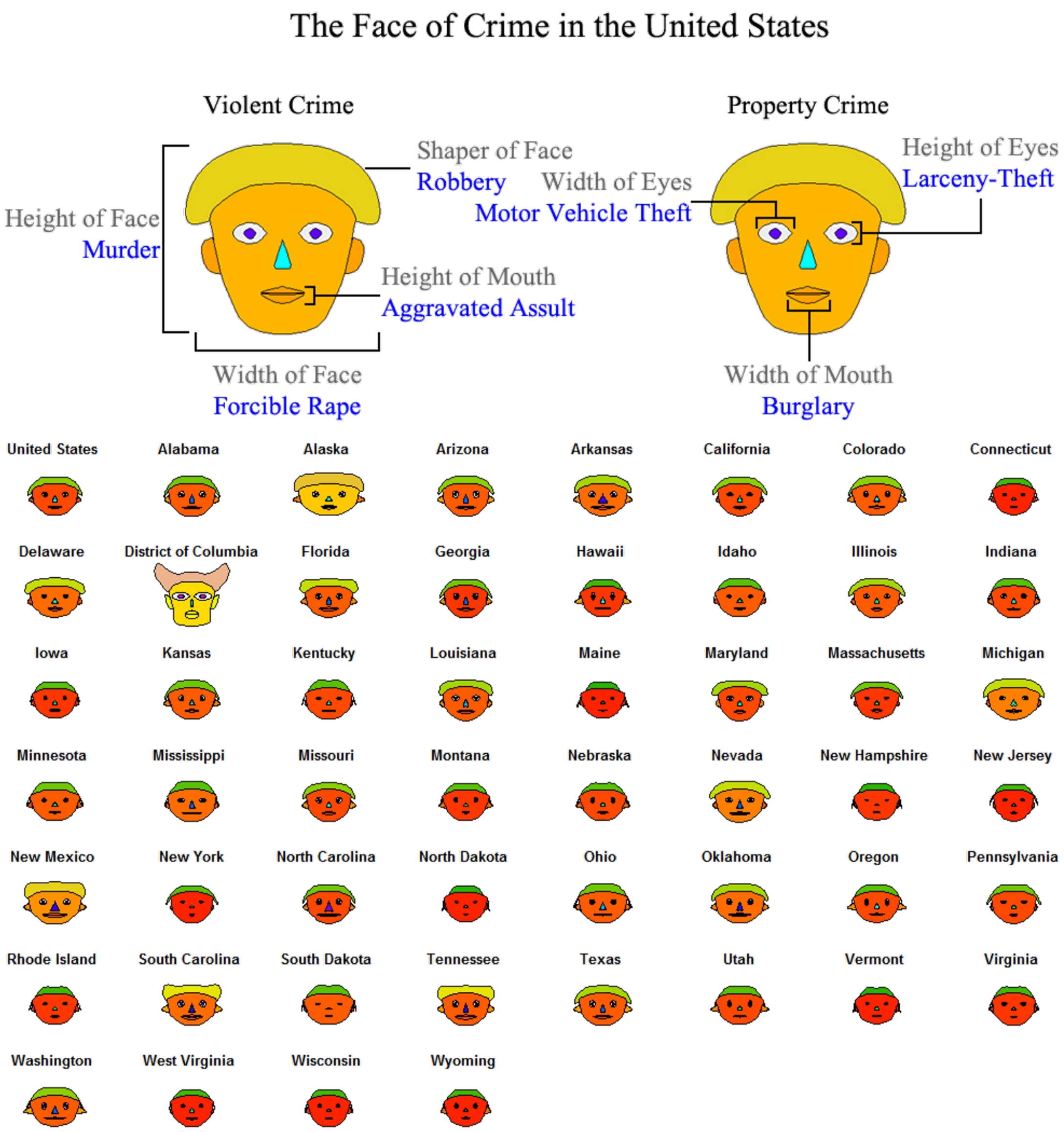
Chernoff Faces



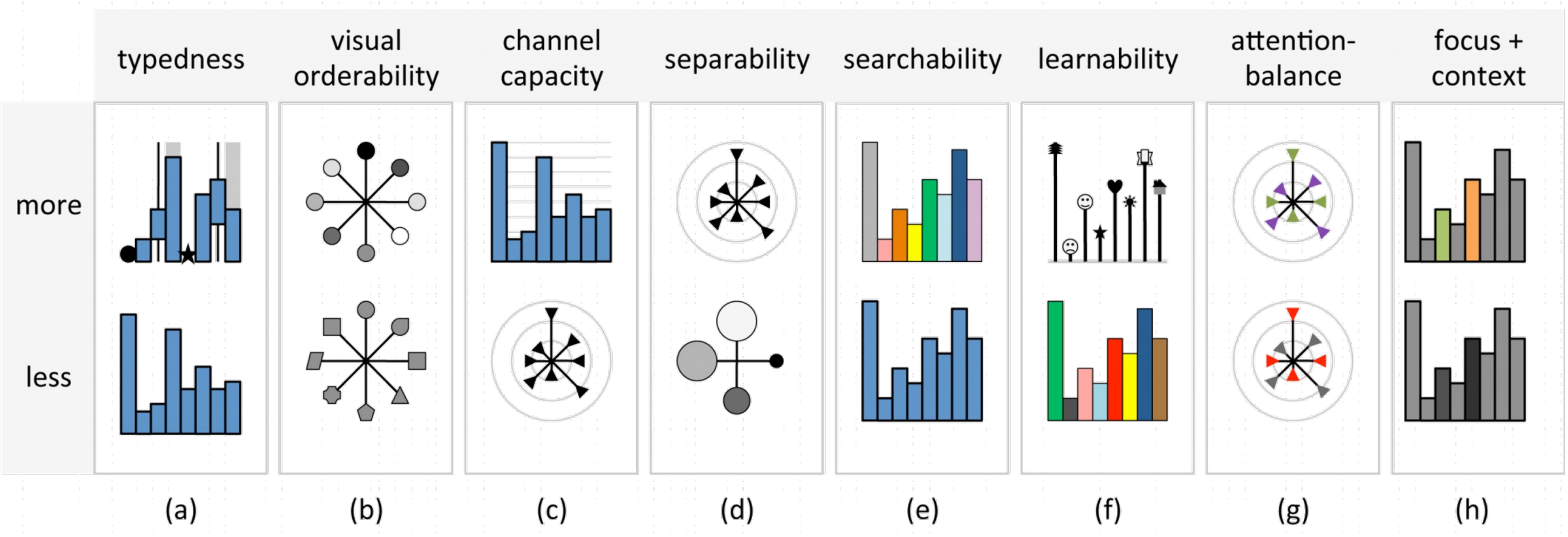
- Use face to encode different attributes.



http://hesketh.com/schampeo/projects/Faces/chernoff.html

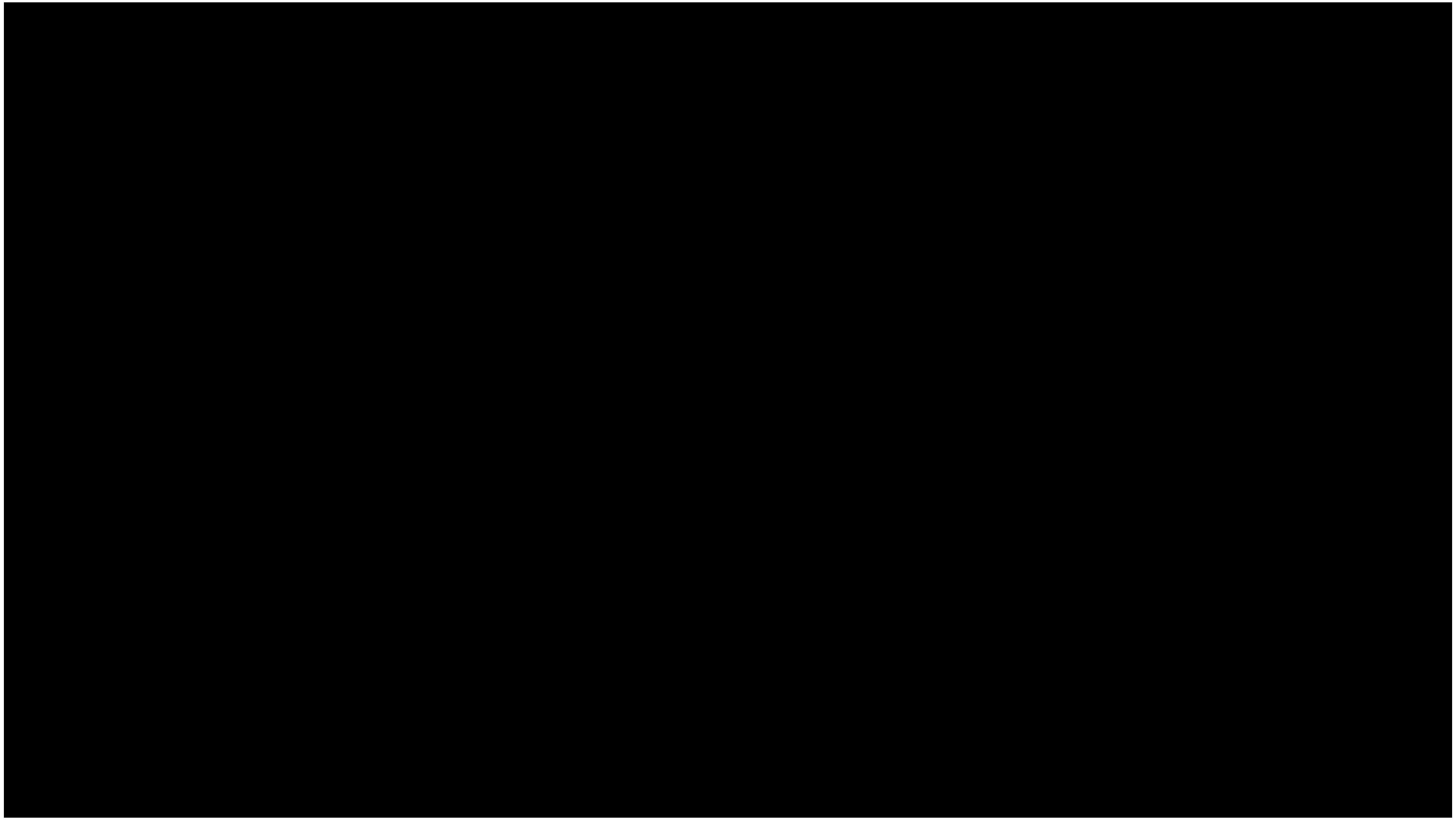

The Face of Crime in the United States
Violent Crime
Property Crime
Shaper of Face
Robbery
Width of Eyes
Motor Vehicle Theft
Height of Eyes
Larceny-Theft
Height of Face
Murder
Height of Mouth
Aggravated Assult
Width of Face
Forcible Rape
Width of Mouth
Burglary
United States
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming

Glyph Design Criteria



Borgo R, Kehrer J et al. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications[C] Eurographics (STARs) (2013)

Glyph Sorting

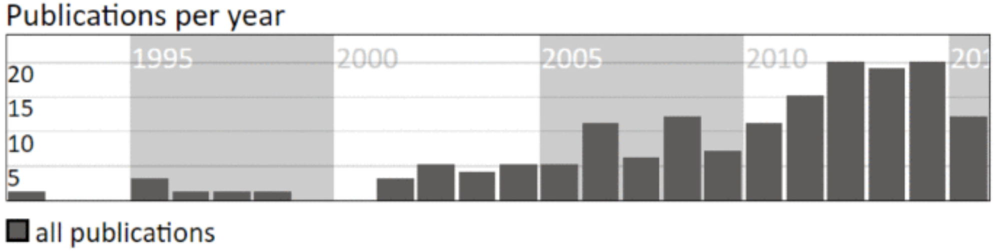


Chung, David HS, et al. "Glyph sorting: Interactive visualization for multi-dimensional data."
InfoVis (2015)

Text + Glyph

A Brief History of Dynamic Graph Visualization

The visualization of dynamic graphs is a growing research area. Starting with **first approaches in the 90s**, the field has been steadily growing to around **20 new publications per year recently**. While early publications mainly introduced new **visualization techniques**, currently a well-balanced mix of **technique**, **application**, and **evaluation** papers is published.



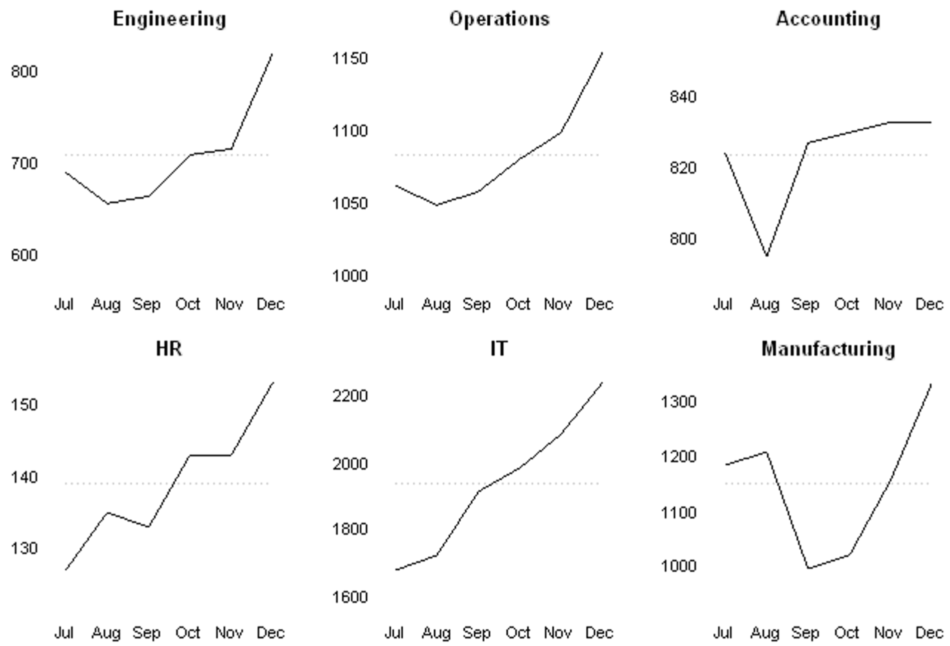
First, animation-based approaches showing graph evolution as **animated changes in node-link diagrams** dominated. **Since 2002**, however, alternative **timeline-based approaches** were suggested that provide an overview of the graph evolution in one view without animation. **By 2010 and later**, these techniques are even dominating the newly proposed approaches, some combining animation and timeline in **hybrid techniques**. Also, researchers **recently** explored using **adjacency matrices** instead of **node-link diagrams** to represent the individual graphs.

Beck F, Weiskopf D. Word-sized graphics for scientific texts[J]. IEEE transactions on visualization and computer graphics (2017).

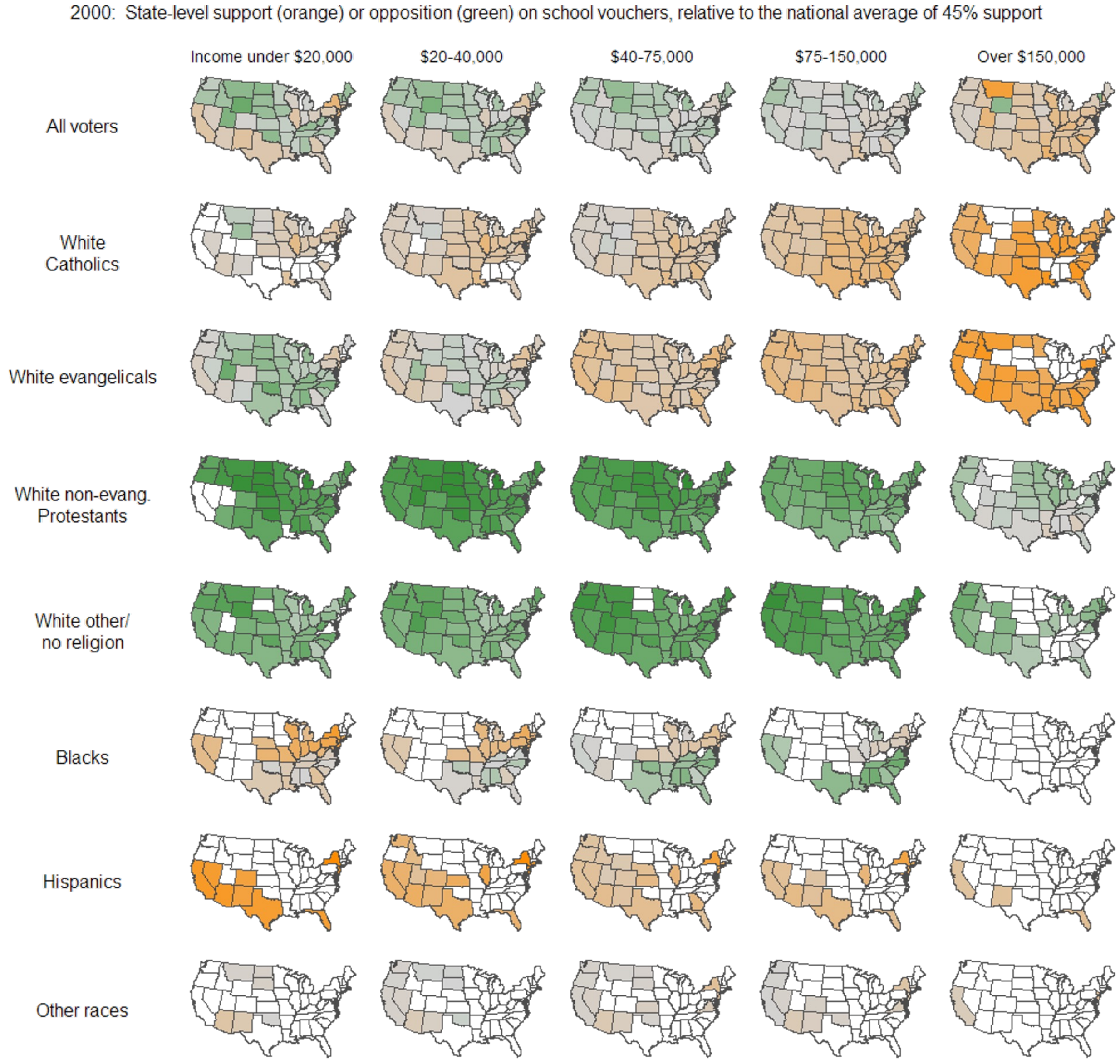
“Small Multiples”

Small Multiples

- Popularized by Edward Tufte.
- A series or grid of small similar graphics or charts for comparison.

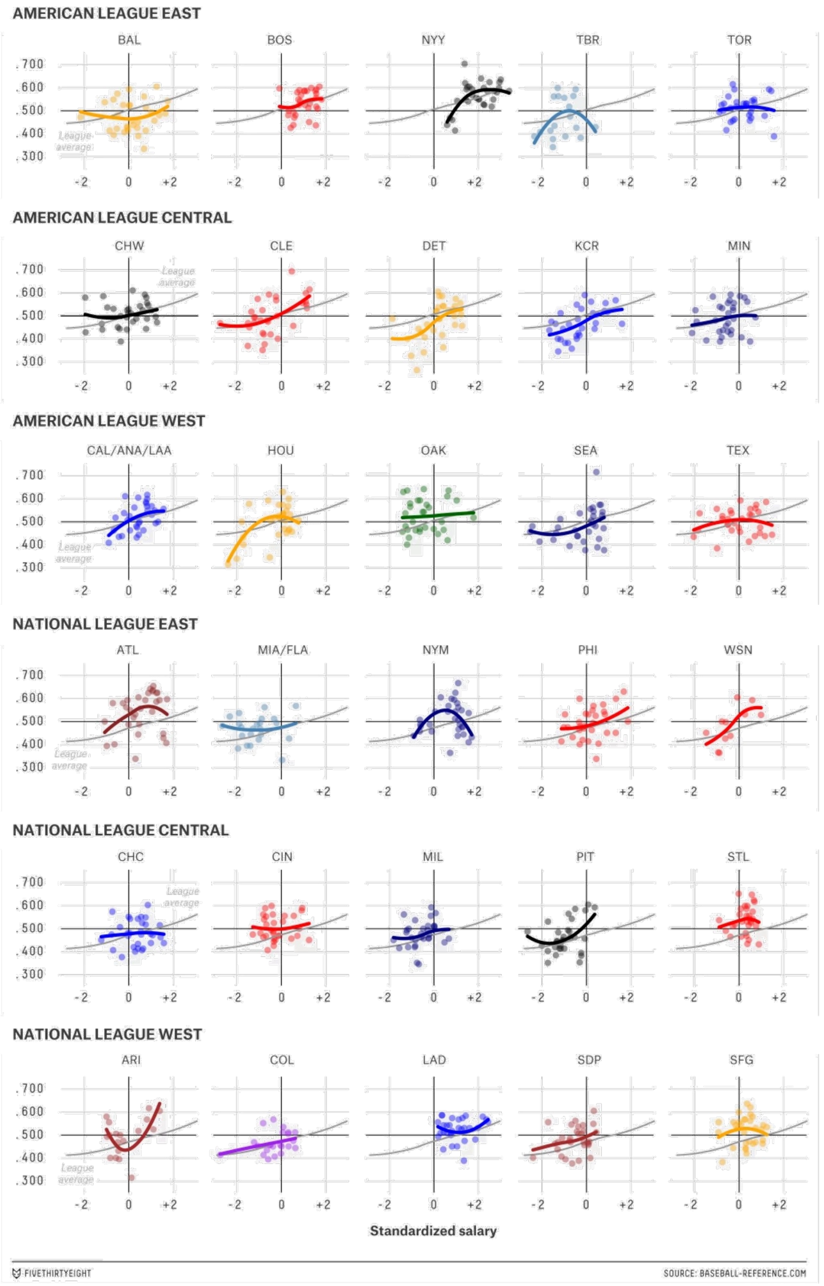


https://en.wikipedia.org/wiki/Small_multiple



Orange and green colors correspond to states where support for vouchers was greater or less than the national average.
The seven ethnic/religious cagetories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.
Where a category represents less than 1% of the voters of a state, the state is left blank.

http://andrewgelman.com/2009/07/hard_sell_for_b/



How your favorite baseball team blows-its-money?

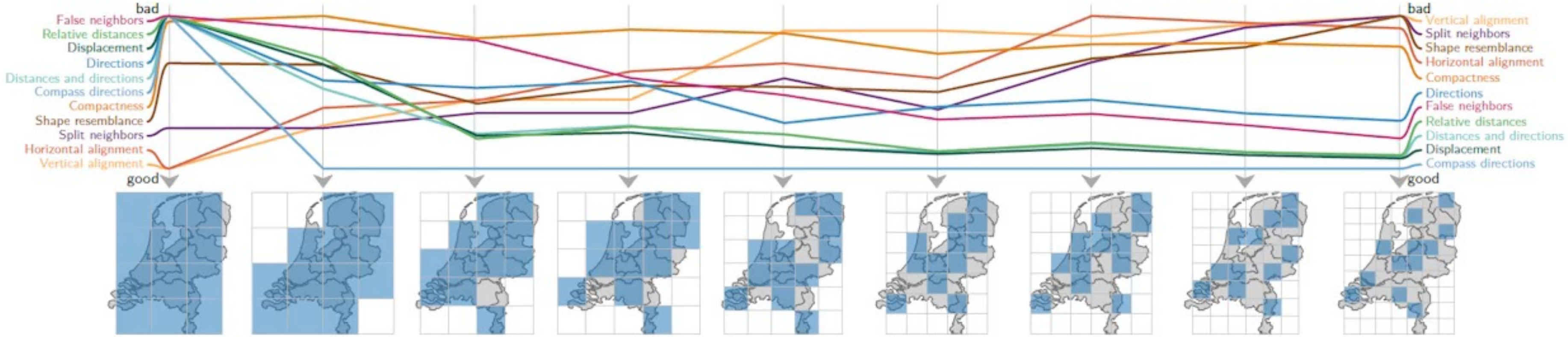
https://fivethirtyeight.com/features/how-your-favorite-baseball-team-blows-its-money/Small_multiple

Small Multiples with Gaps

上海科技大学
ShanghaiTech University

Small Multiples

Wouter Meulemans Jason Dykes Aidan Slingsby
Cagatay Turkay Jo Wood



http://www.gicentre.net/smwg

Meulemans W, Dykes J, Slingsby A, et al. Small multiples with gaps[J]. IEEE transactions on visualization and computer graphics, (2017)

立志成才 报国裕民

Interaction: "Dust & Magnet"

Dust & Magnet

- Metaphor:
  - Dust - Data points.
  - Magnet – Attribute Filters.
  - Dust particles are attracted by magnets while dragging magnet blocks.

Dust & Magnet

Soo Yi J, Melton R, Stasko J, et al. Dust & magnet: multivariate information visualization using a magnet metaphor[J]. Information visualization (2005).

LineUp

Visual Analysis of Multi-Attribute Rankings

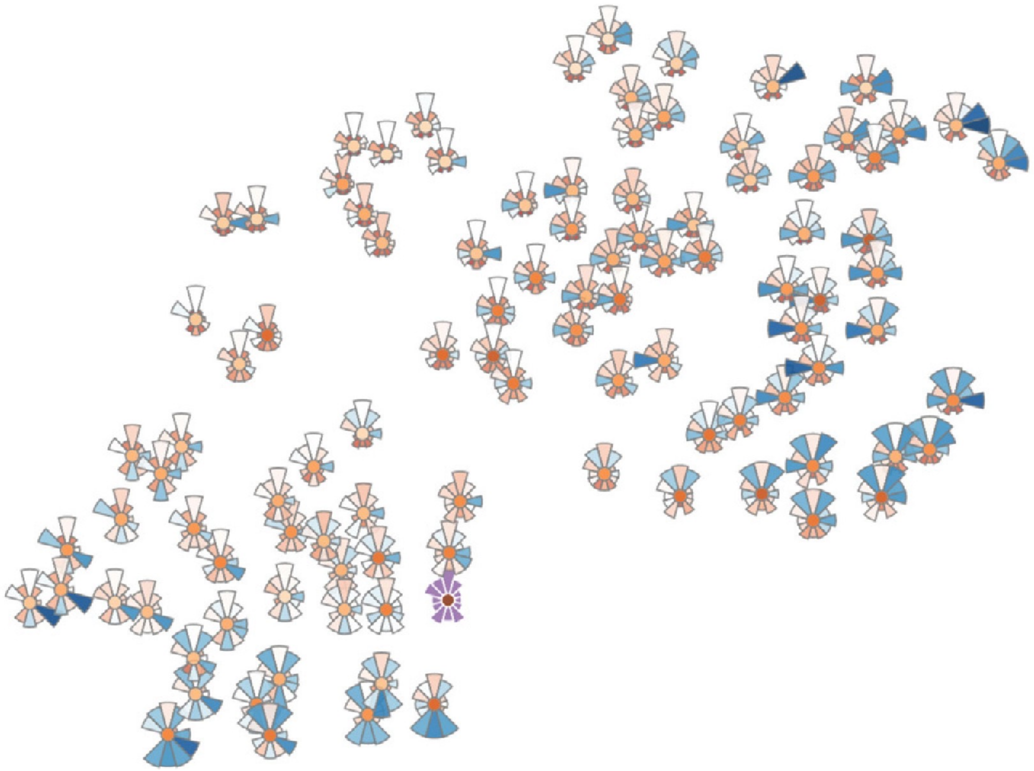
Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister and Marc Streit

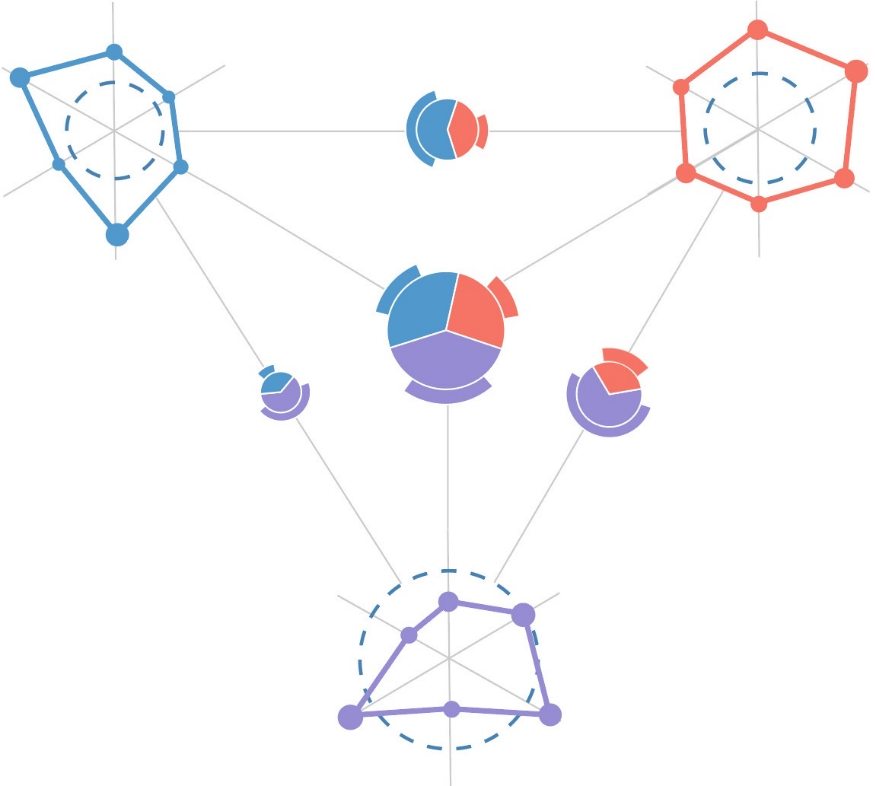
CALEYDO

JKU
JOHANNES KEPLER
UNIVERSITÄT LINZ

HARVARD
School of Engineering
and Applied Sciences

HARVARD
MEDICAL SCHOOL





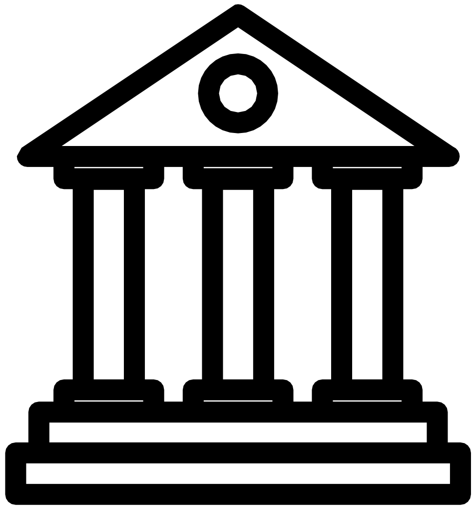
SkyLens: Visual Analysis of Skyline on Multi-dimensional Data

SkyLens: Visual Analysis of Skyline on Multi-dimensional Data
Xun Zhao, Yanhong Wu, Weiwei Cui, Xinnan Du, Yuan Chen, Yong Wang, Dik Lun Lee, Huamin Qu
IEEE Trans. Vis. Comput. Graph.

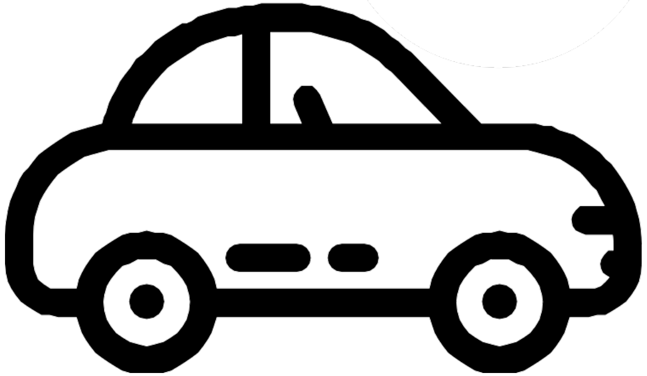
Background

- Multi-criteria decision making

Employee recruitment



University selection



Car comparison

Background

- Suppose you are a college basketball coach, how do you recruit the best players?

PLAYER	TEAM	AGE	GP	W	L	MIN	OFFRTG	DEFRTG	NETRTG	AST%	AST/TO	AST RATIO	OREB%
AJ Hammons	DAL	24	22	4	18	7.4	102.2	102.8	-0.6	3.8	0.40	6.2	4.9
Aaron Brooks	IND	32	65	36	29	13.7	101.5	104.6	-3.0	21.6	1.89	24.6	2.2
Aaron Gordon	ORL	21	80	29	51	28.7	105.4	108.2	-2.8	9.7	1.69	12.5	5.4
Aaron Harrison	CHA	22	5	2	3	3.3	83.3	101.9	-18.6	37.5	0.00	38.1	0.0
Adreian Payne	MIN	26	18	5	13	7.5	102.6	101.8	0.8	8.9	0.88	9.0	6.9
Al Horford	BOS	31	68	46	22	32.3	110.7	105.8	5.0	23.9	2.93	25.7	4.9
Al Jefferson	IND	32	66	33	33	14.1	102.3	108.1	-5.8	11.4	1.73	9.5	8.6
Al-Farouq Aminu	POR	26	61	33	28	29.1	107.7	105.9	1.8	8.2	1.05	13.8	4.9
Alan Anderson	LAC	34	30	20	10	10.3	103.1	114.0	-10.8	5.2	1.57	10.5	1.1
Alan Williams	PHX	24	47	11	36	15.1	105.6	105.8	-0.3	4.9	0.62	6.1	13.8
Alec Burks	UTA	25	42	26	16	15.5	105.0	104.9	0.1	7.4	0.86	8.6	2.9
Alex Abrines	OKC	23	68	37	31	15.5	106.0	108.3	-2.3	5.5	1.21	9.2	1.9
Alex Len	PHX	24	77	21	56	20.3	99.4	110.5	-11.1	4.3	0.43	6.3	10.4

Introduction – Skyline

- Skyline algorithm: automatically select the skyline of the dataset

- In database, skyline algorithm is an important and extensively studied problem

Introduction – Skyline

- Skyline algorithm: automatically select the skyline of the dataset

Created by Jpldesigns from the dreamstime

Introduction – Skyline Definition

- Skyline: a set of superior points that are not dominated by other points in the dataset

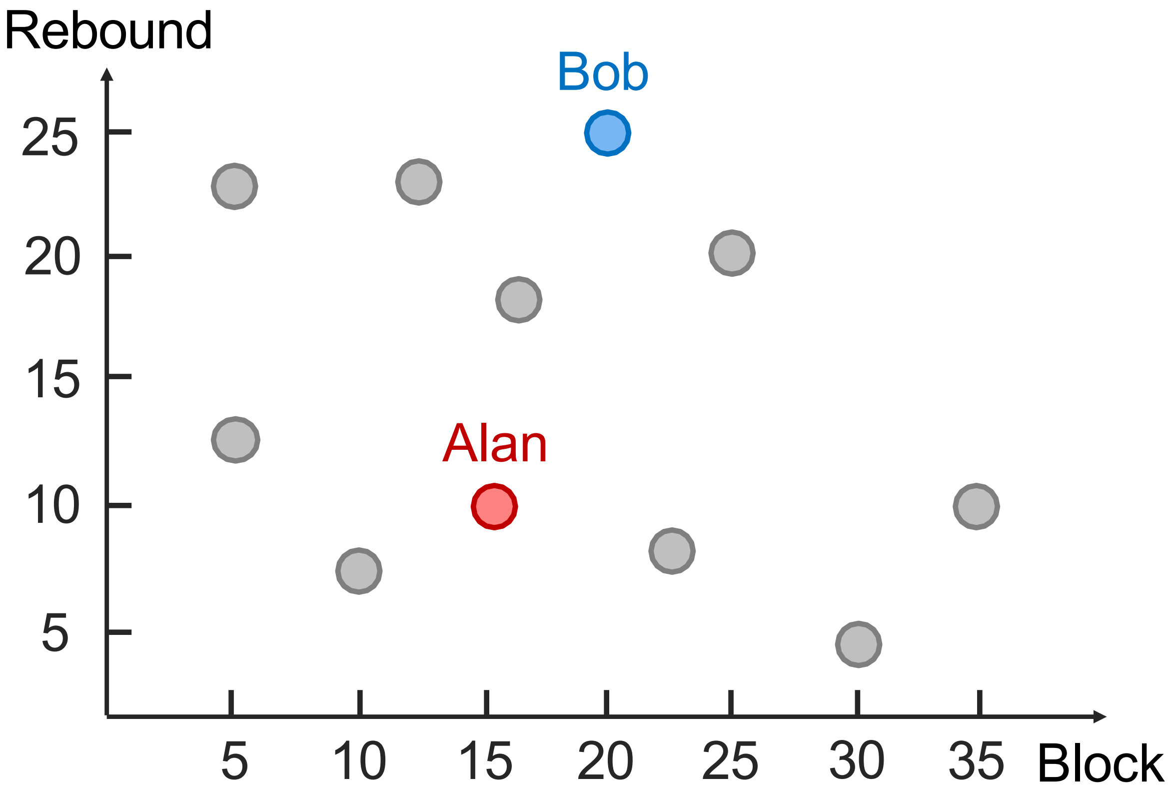
- Dominance:
  - If p dominates q, then:
    - p is not worse than q in all attributes
    - p is at least better than q in one attribute

Introduction – Skyline Example

- Skyline: a set of superior points that are not dominated by other points in the dataset

Players	Block	Rebound
Alan	15	10
Bob	**20**	**25**

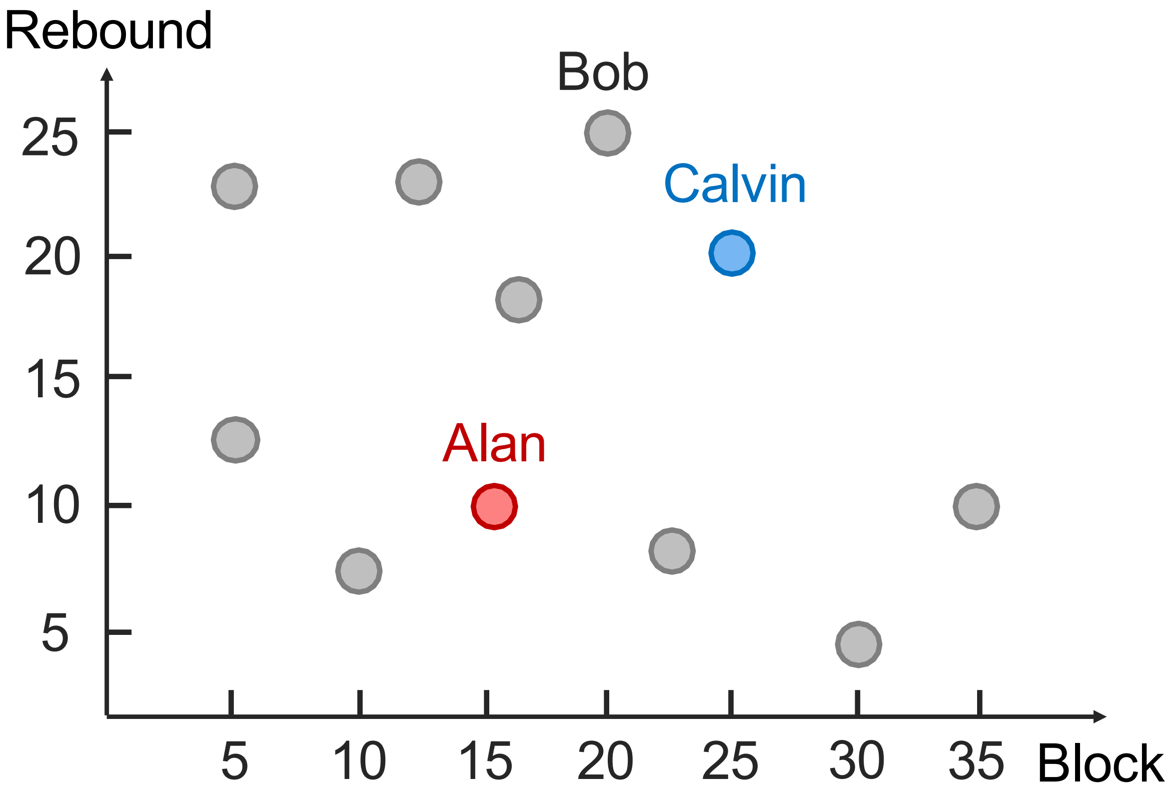
Bob dominates Alan (**block** & **rebound**)



Introduction – Skyline Example

- Skyline: a set of superior points that are not dominated by other points in the dataset

Players	Block	Rebound
Alan	15	10
Bob	20	25
Calvin	**25**	**20**



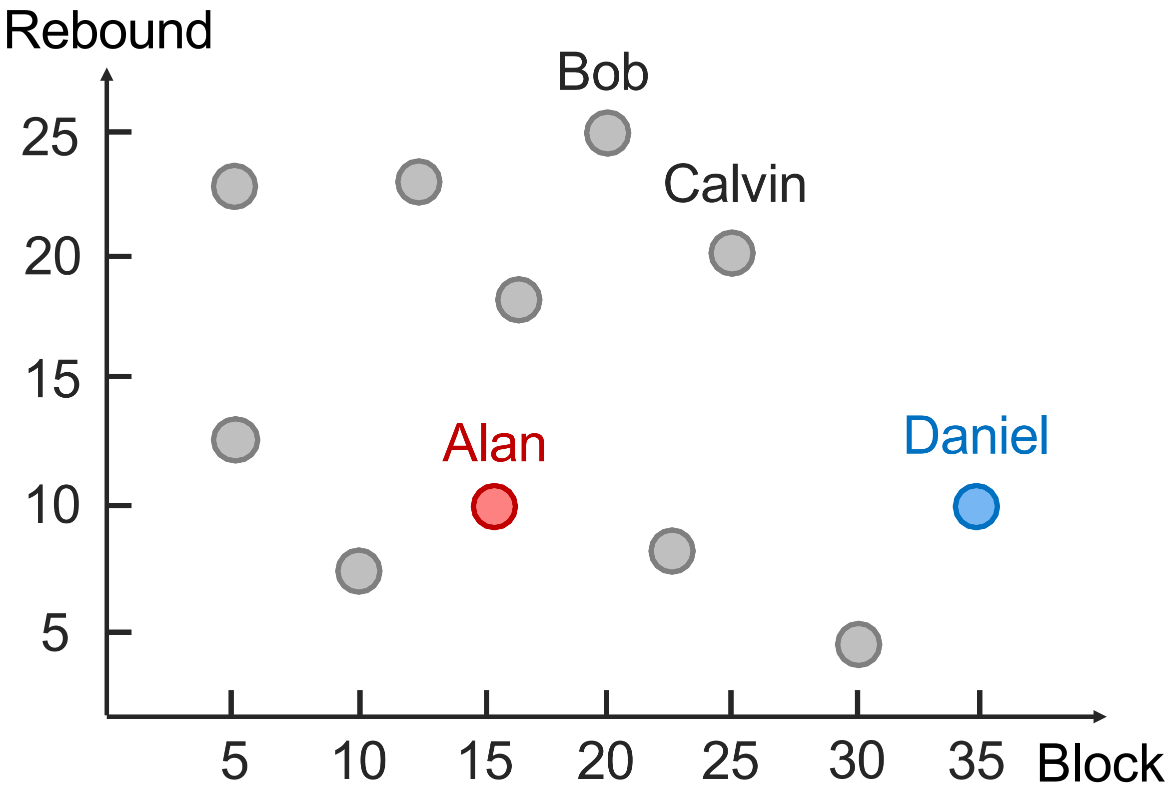
Calvin dominates Alan (**block** & **rebound**)

Introduction – Skyline Example

- Skyline: a set of superior points that are not dominated by other points in the dataset

Players	Block	Rebound
Alan	15	10
Bob	20	25
Calvin	25	20
Daniel	**30**	**10**

Daniel dominates Alan (**block**)

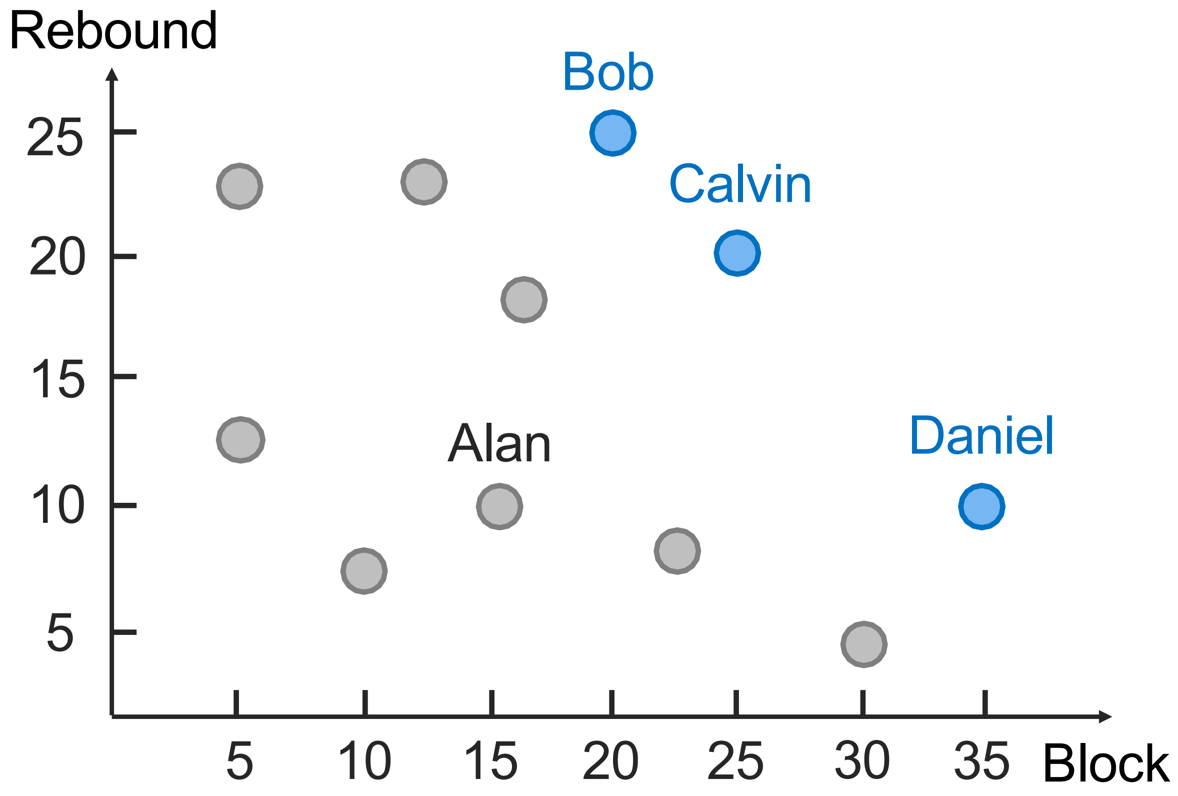


Introduction – Skyline Example

- Skyline: a set of superior points that are not dominated by other points in the dataset

Players	Block	Rebound
Alan	15	10
Bob	**20**	**25**
Calvin	**25**	**20**
Daniel	**30**	**10**

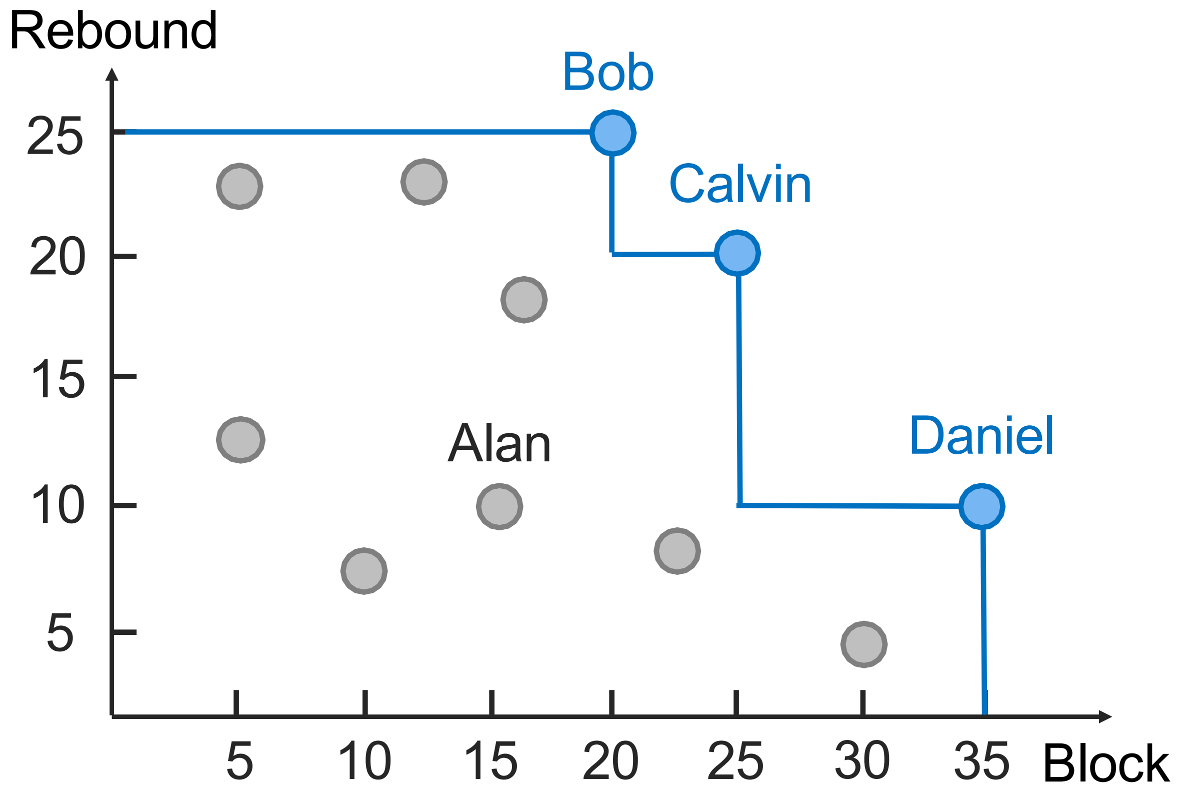
Points: Daniel > Calvin > Bob
Rebound: Bob > Calvin > Daniel



Introduction – Skyline Example

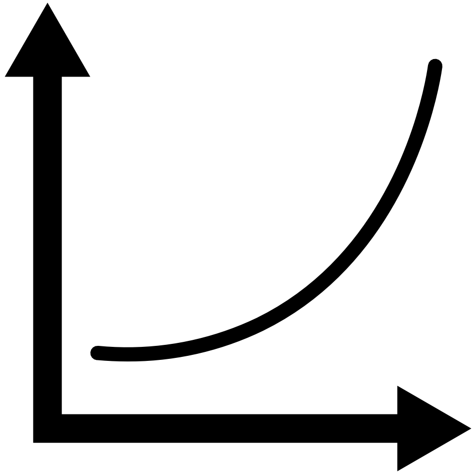
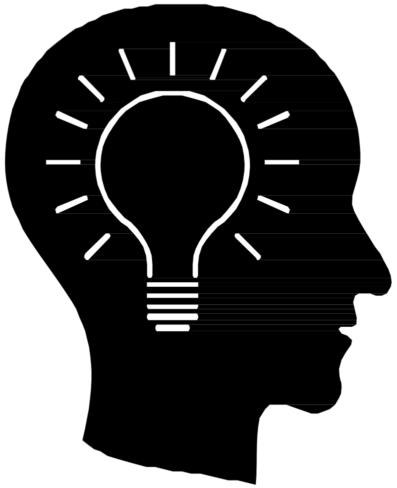
- Skyline: a set of superior points that are not dominated by other points in the dataset

Players	Block	Rebound
Alan	15	10
Bob	**20**	**25**
Calvin	**25**	**20**
Daniel	**30**	**10**



Skyline: Bob, Calvin, Daniel

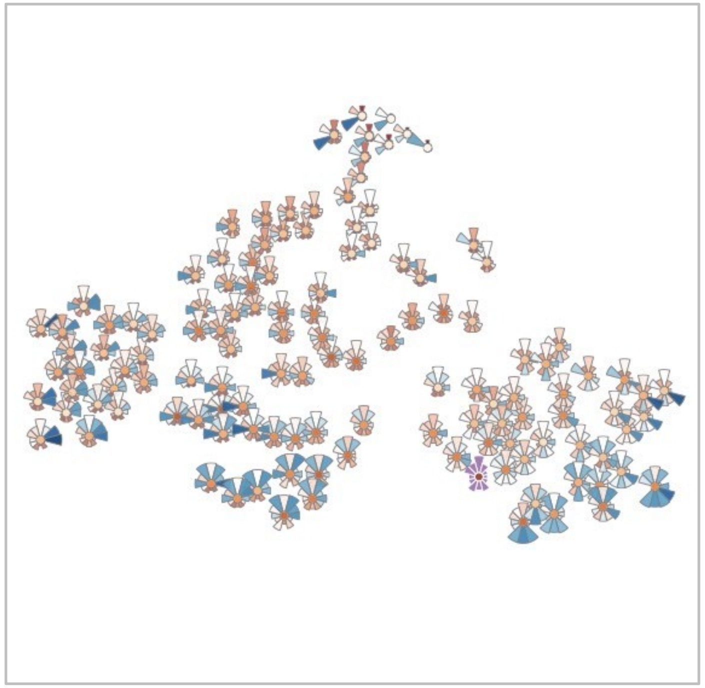
Introduction – Challenges

Scalability	Interpretation	Comparison
		
The size of skyline increases with the number of attributes	The reasons that make a point in skyline is unclear	The strength and weakness of each skyline point is implicit

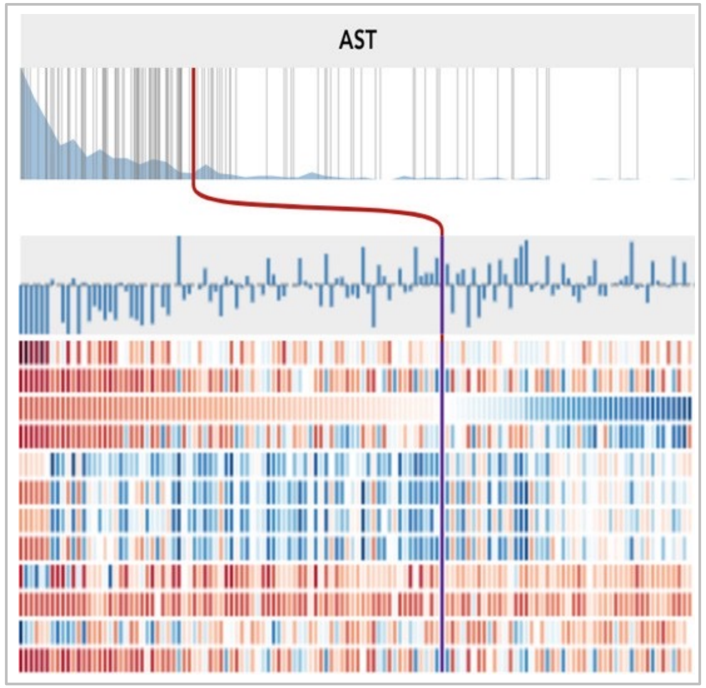
Created by Davo Sime, shashank singh , and Magicon from the Noun Project

SkyLens – Visual Components

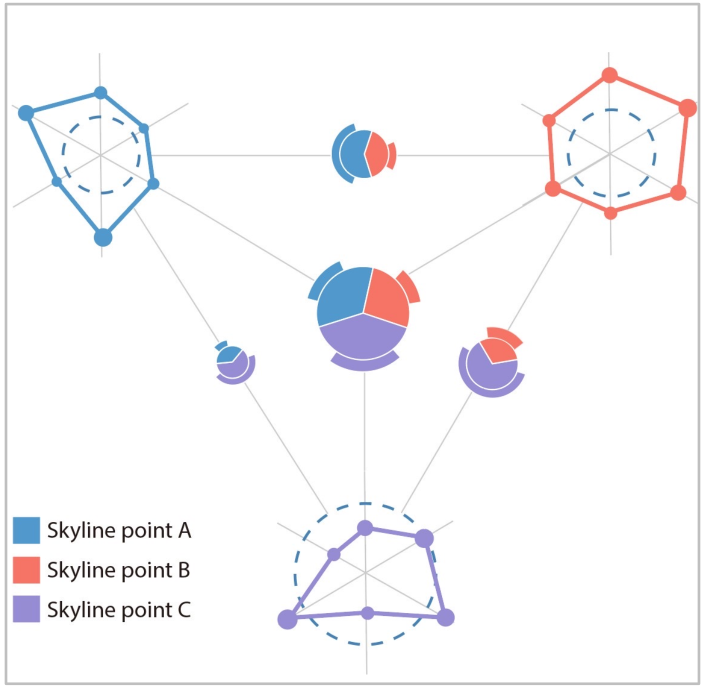
Projection View



Tabular View

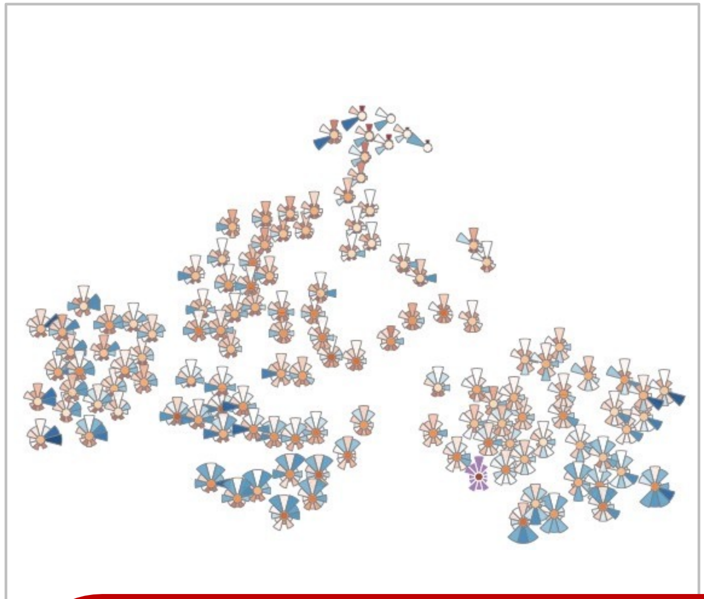


Comparison View



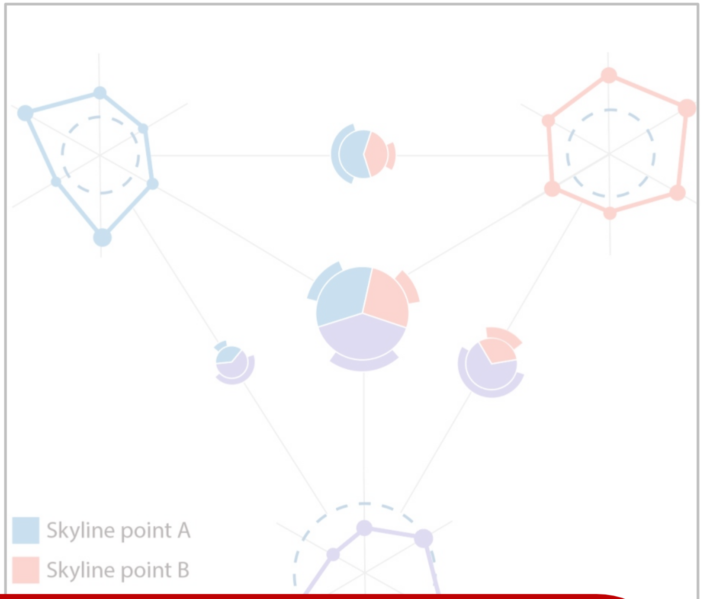
SkyLens – Projection View

Projection View



Tabular View

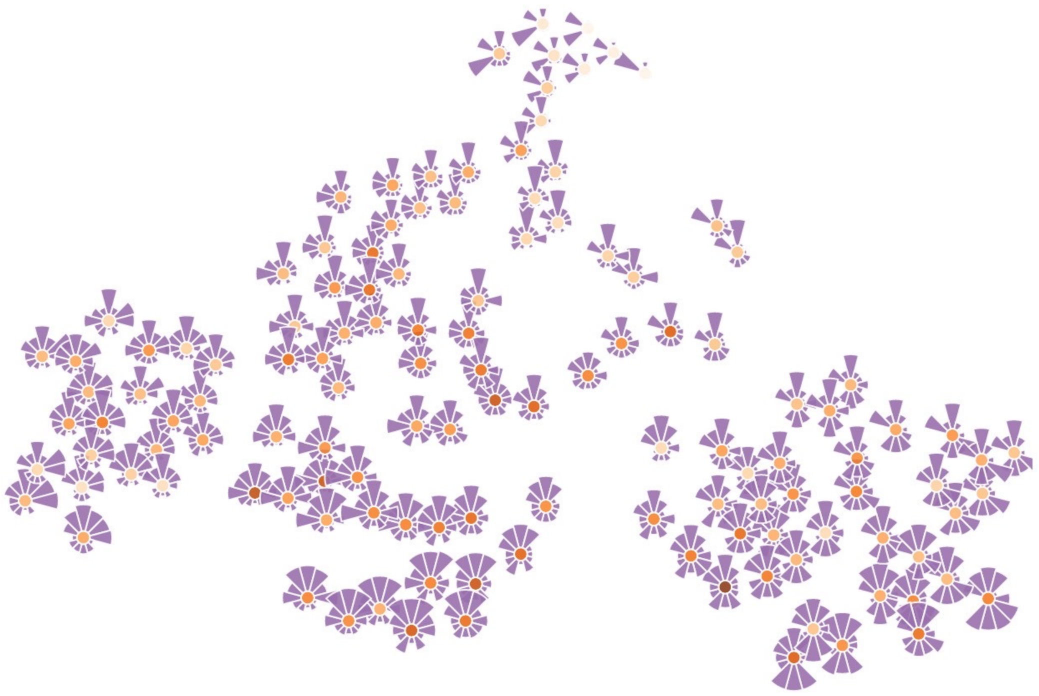
Comparison View



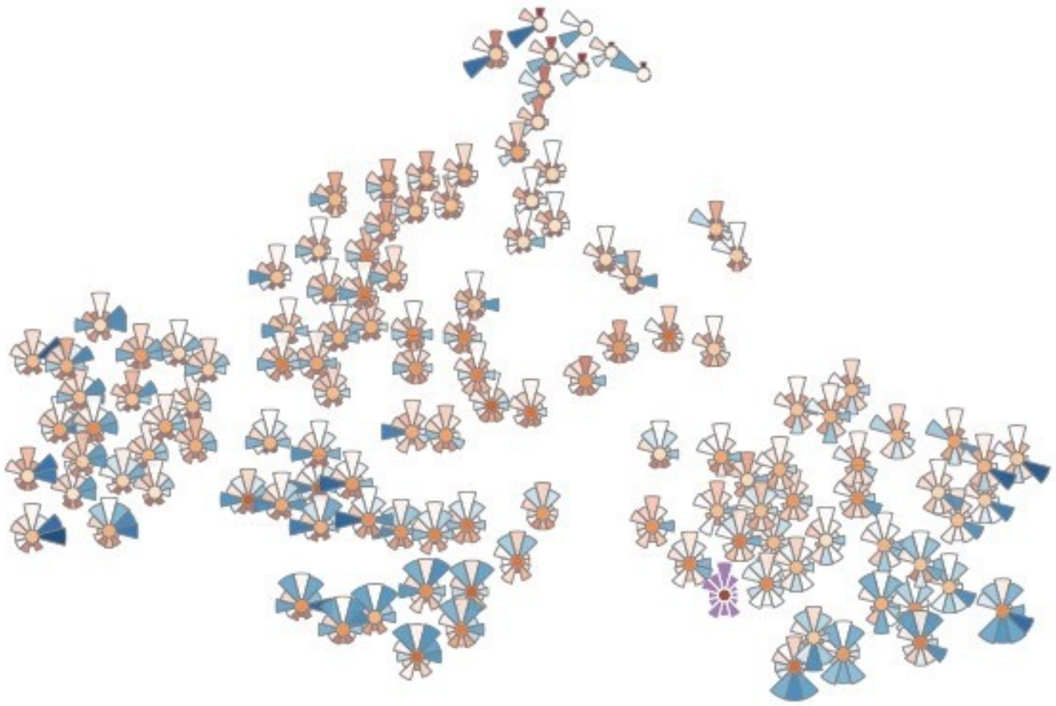
Projection View: provide an overview of skyline (clusters and outliers)

Projection View

- Methods: t-SNE projection and skyline glyphs



Normal mode

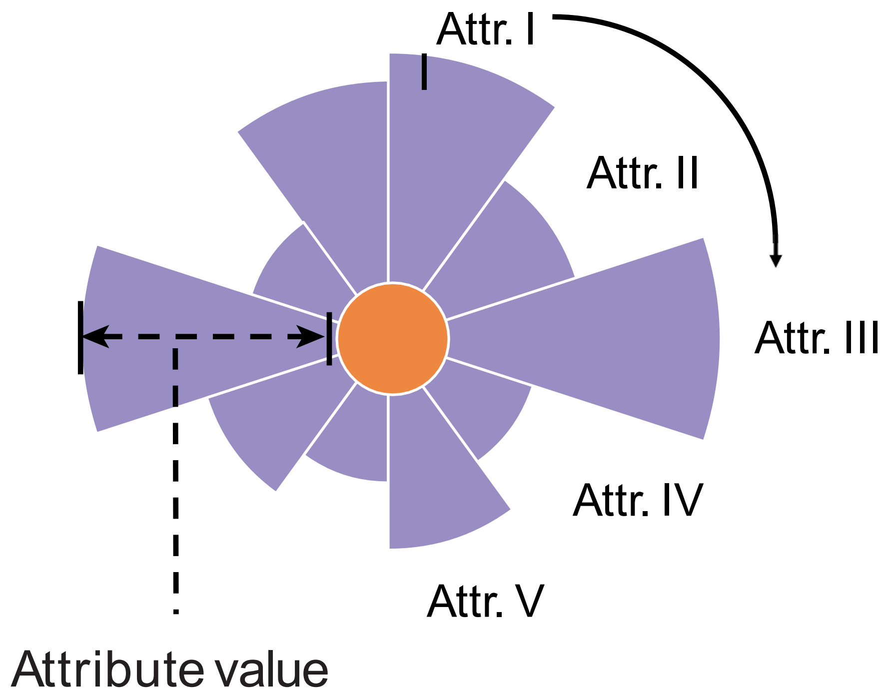


Focus mode

Projection View – Skyline Glyph

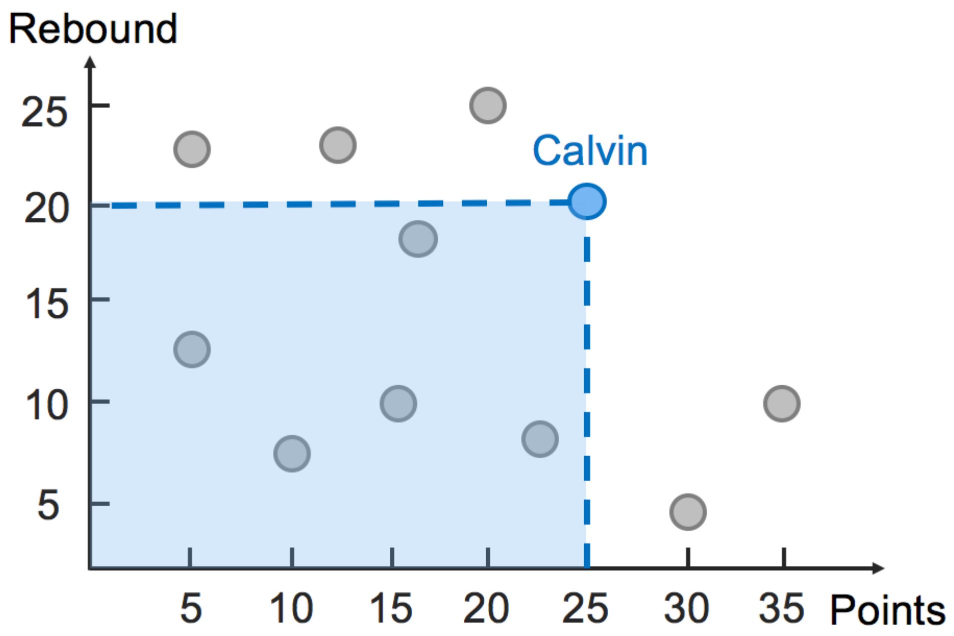
- Normal mode: show the attribute value distribution of skyline

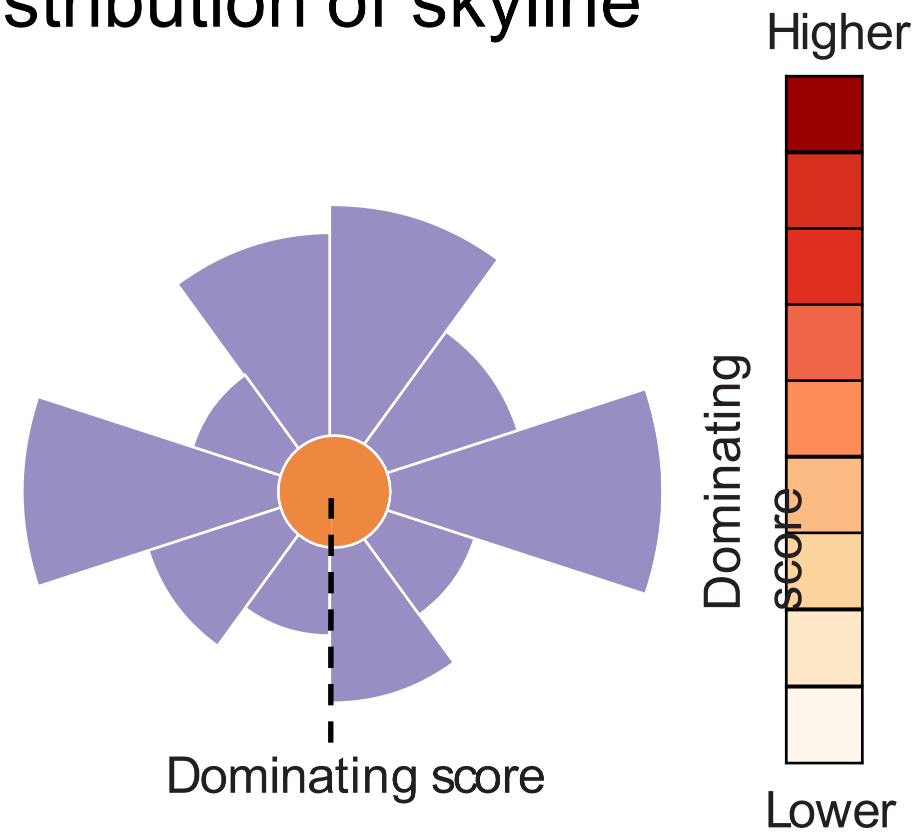
Attribute	Value
Attr. I	5
Attr II	3
Attr. III	7
Attr. IV	1
Attr. V	3
Attr. VI	1
…	



Projection View – Skyline Glyph

- Normal mode: show the attribute value distribution of skyline
- Dominating score (superiority metric):
  - # of points dominated by this point

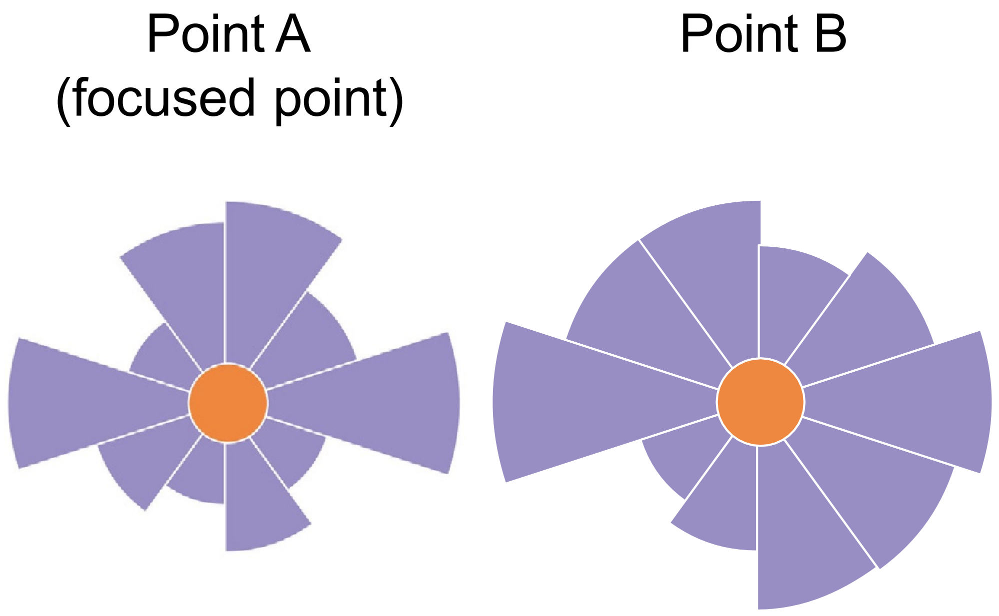




Projection View – Skyline Glyph

- Focus mode: highlight how other points differ from a focused one

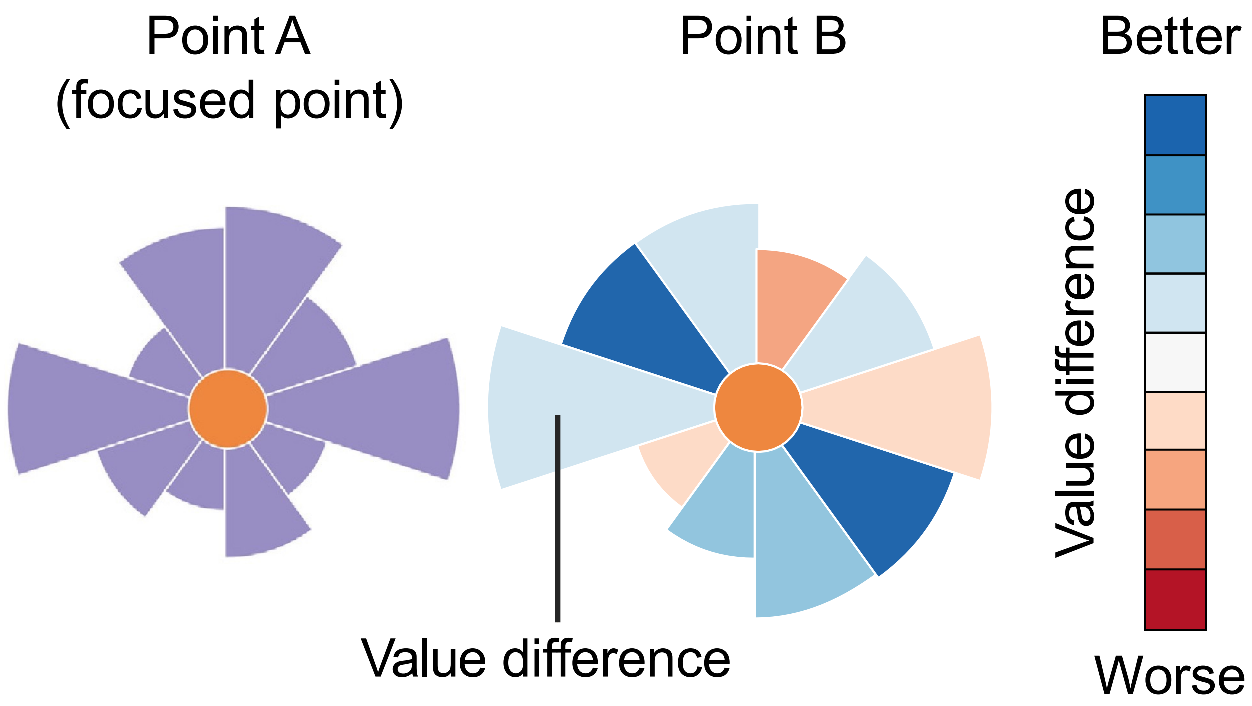
Attribute	Point A	Point B
Attr. I	5	**3**
Attr II	3	**4**
Attr. III	7	**6**
Attr. IV	1	**5**
Attr. V	3	**5**
Attr. VI	1	**3**
…		



Projection View – Skyline Glyph

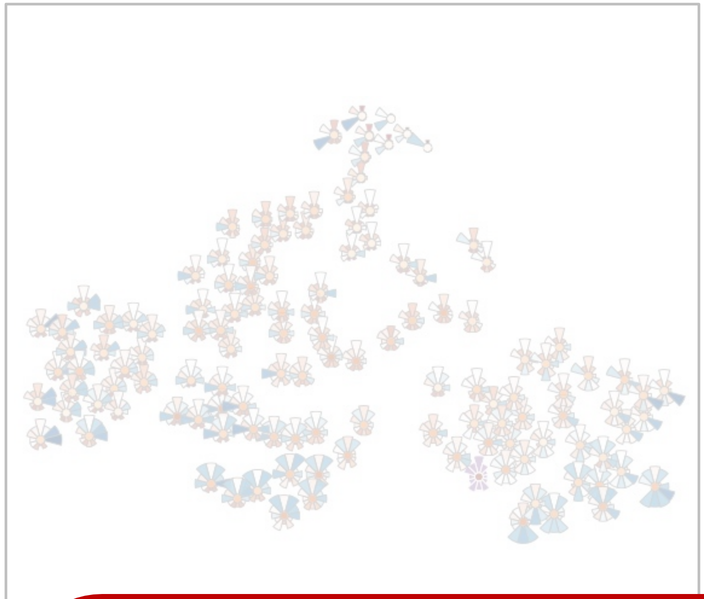
- Focus mode: highlight how other points differ from a focused one using color map

Attribute	Point A	Point B
Attr. I	5	**3 (diff. = -2)**
Attr II	3	**4 (diff. = 1)**
Attr. III	7	**6 (diff. = -1)**
Attr. IV	1	**5 (diff. = 4)**
Attr. V	3	**5 (diff. = 2)**
Attr. VI	1	**3 (diff. = 2)**
	…	

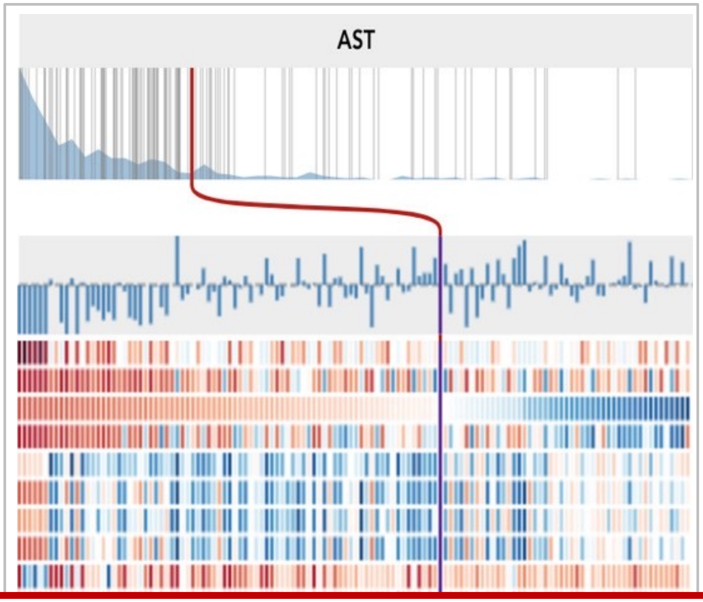


SkyLens – Tabular View

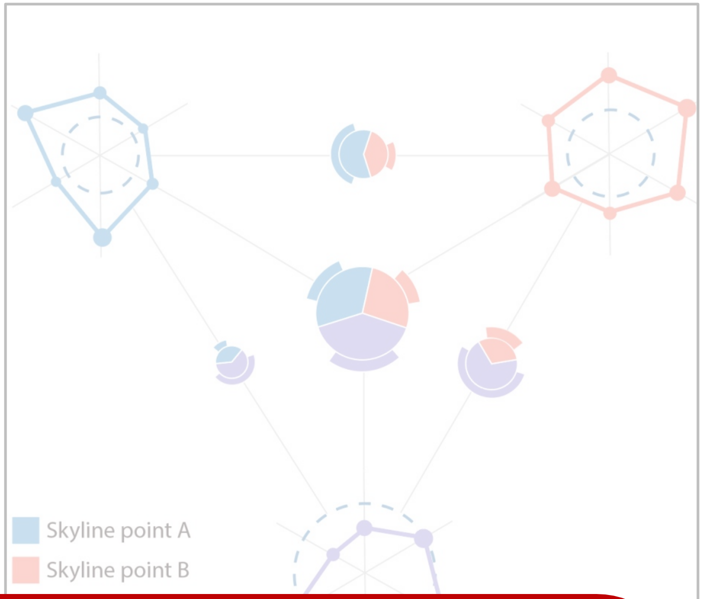
Projection View



Tabular View



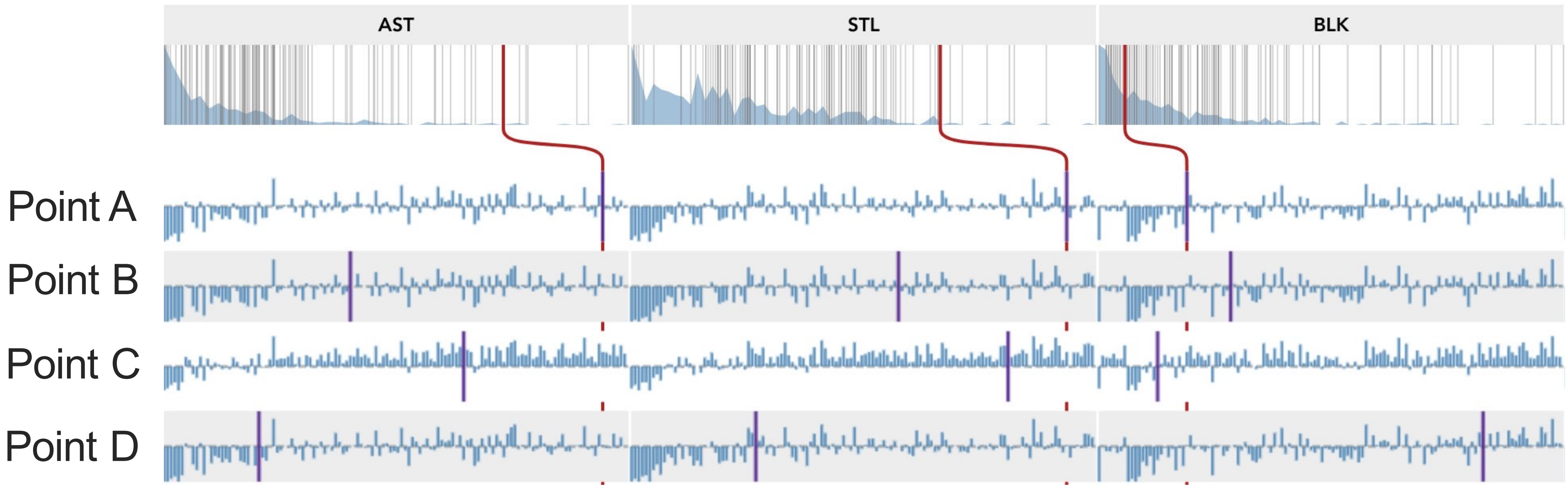
Comparison View



Tabular View: infer the underlying reasons that make a point in skyline

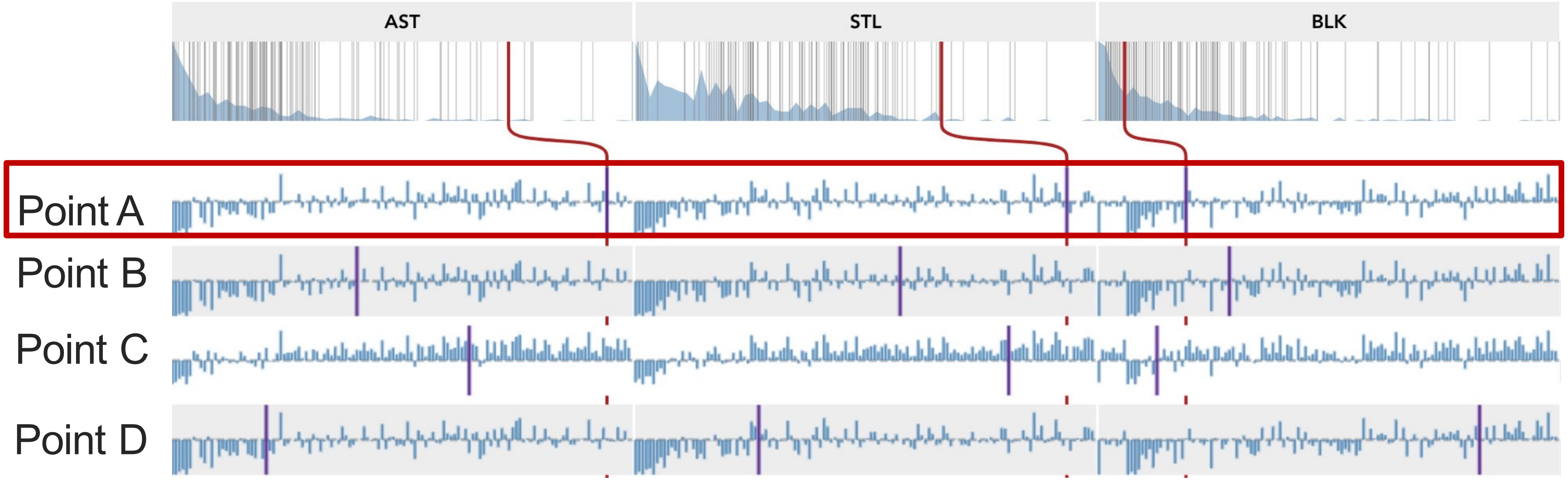
Tabular View

- Methods: matrix representation & in-cell bar chart visualization



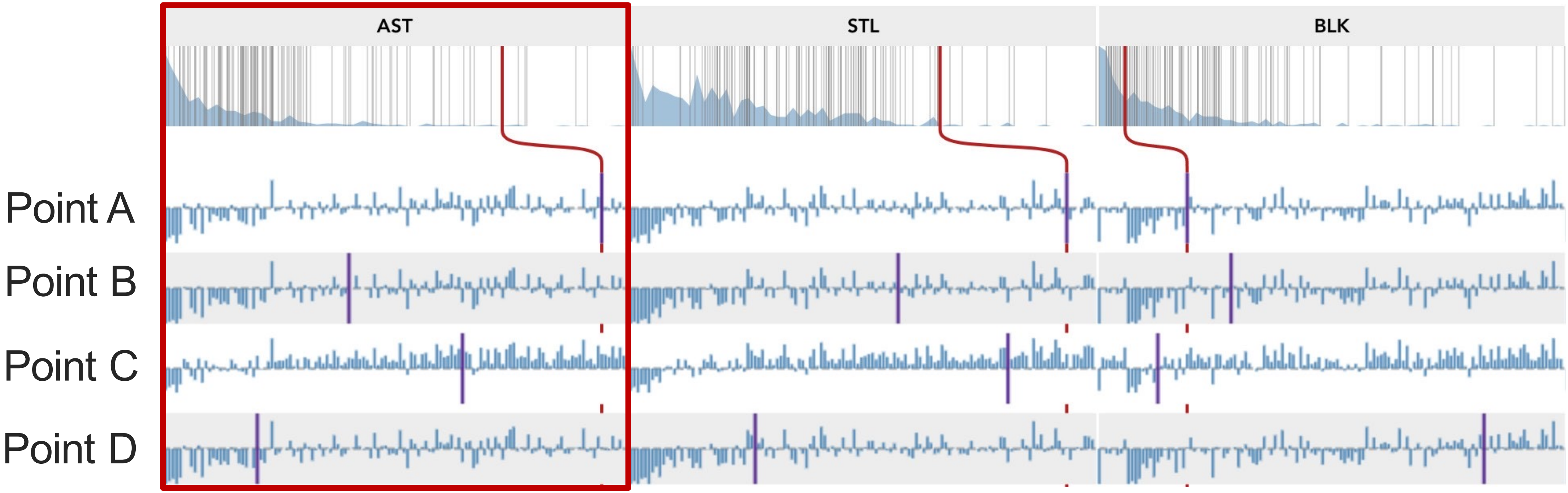
Tabular View

- Methods: matrix representation & in-cell bar chart visualization
  - Each row represents a skyline point



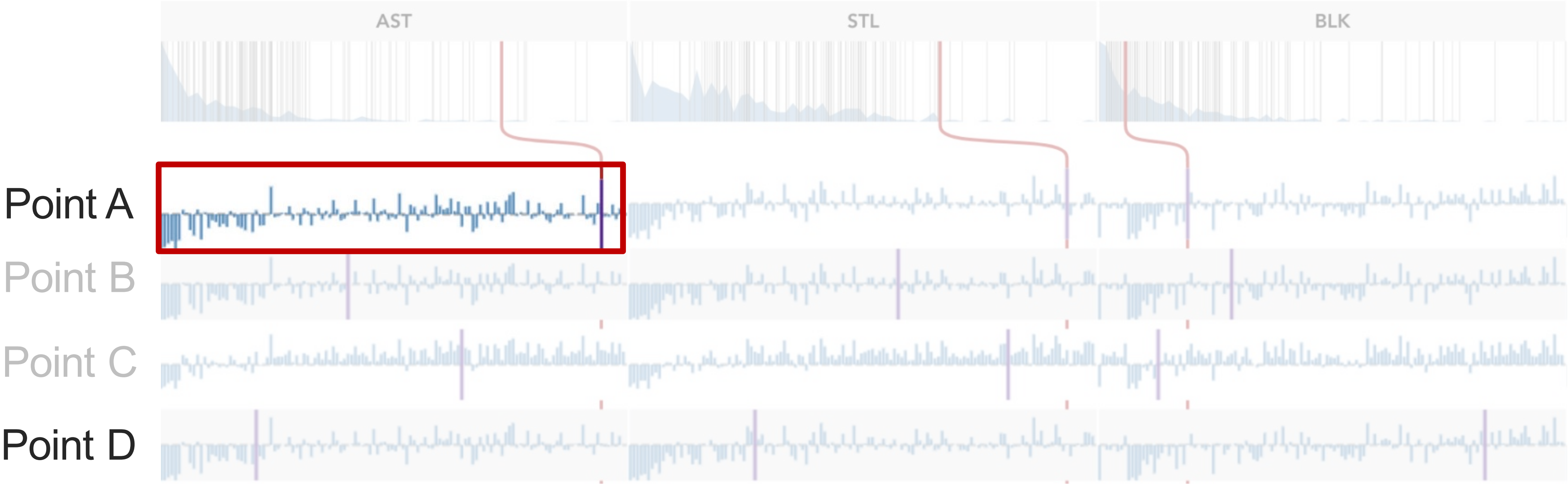
Tabular View

- Methods: matrix representation & in-cell bar chart visualization
  - Each row represents a skyline point
  - Each column represents an attribute



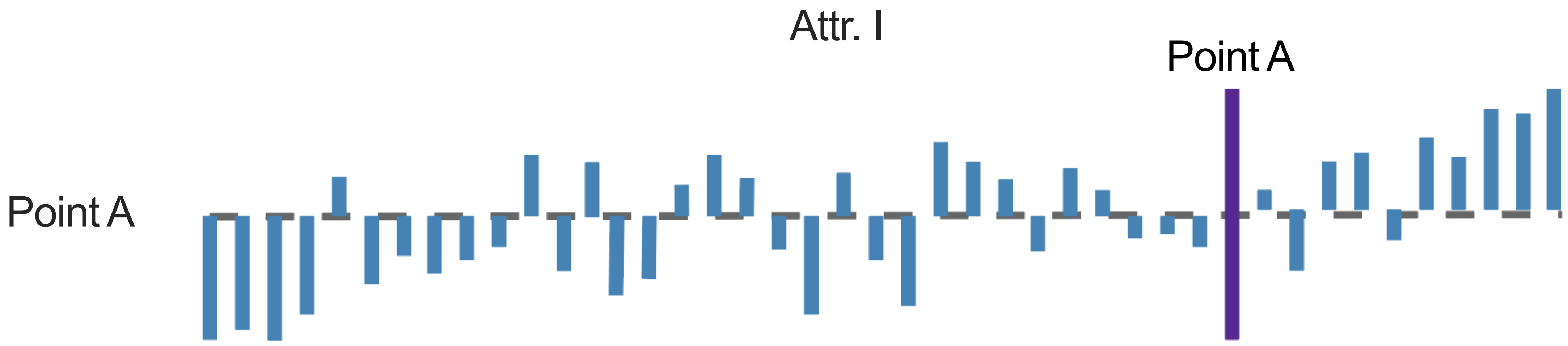
Tabular View

- Table cell – divergent bar chart visualization
  - Goal: summarize the overall differences among skyline points



Tabular View

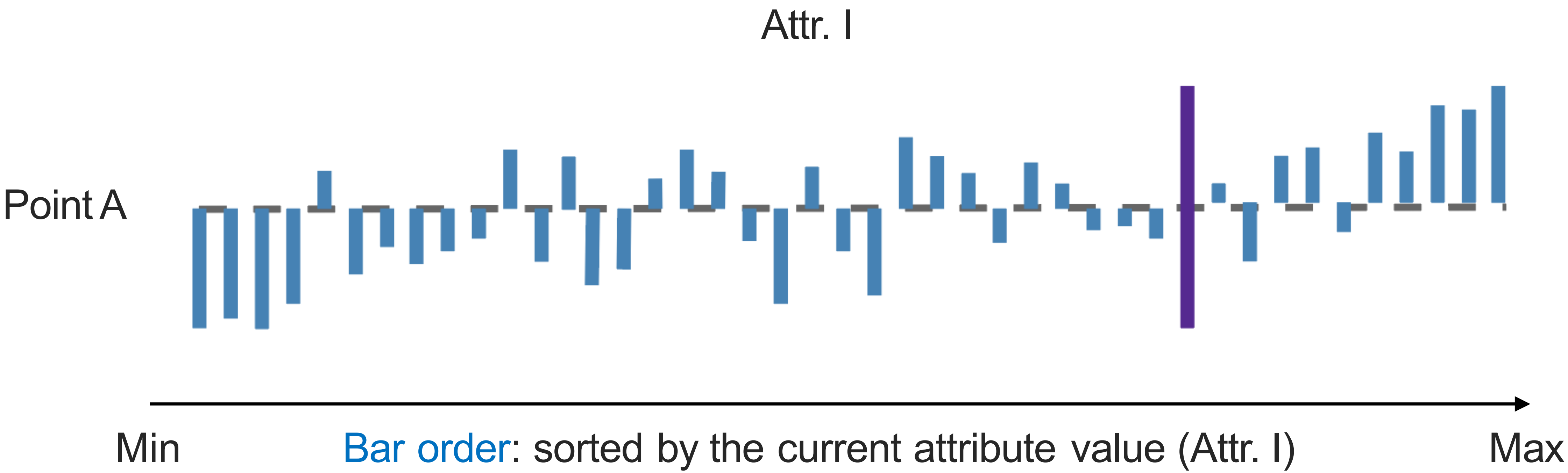
- Table cell – divergent bar chart visualization
  - Goal: summarize the overall differences between skyline points



Each vertical bar represents a skyline point: current point (Point A) other points

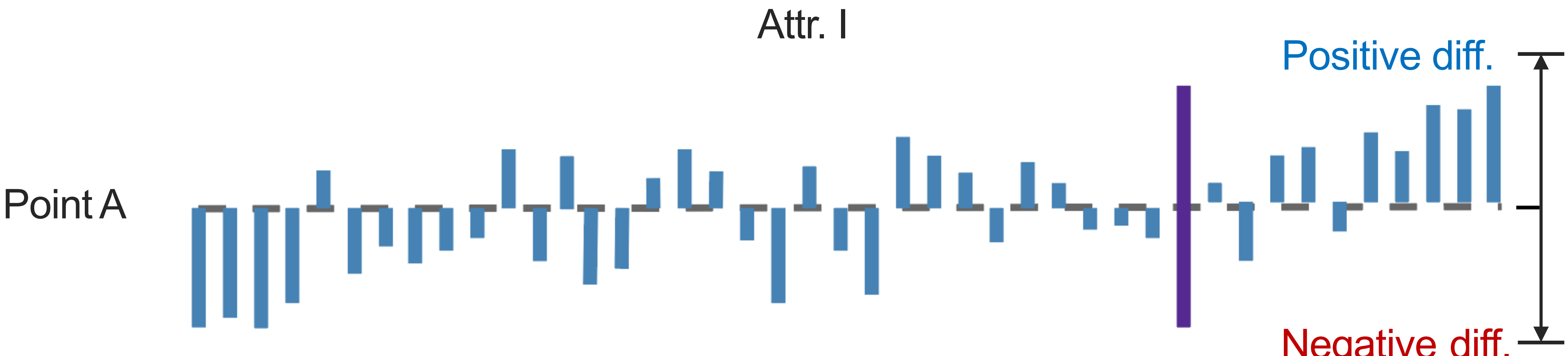
Tabular View

- Table cell – divergent bar chart visualization
  - Goal: summarize the overall differences between skyline points



Tabular View

- Table cell – divergent bar chart visualization
  - Goal: summarize the overall differences between skyline points



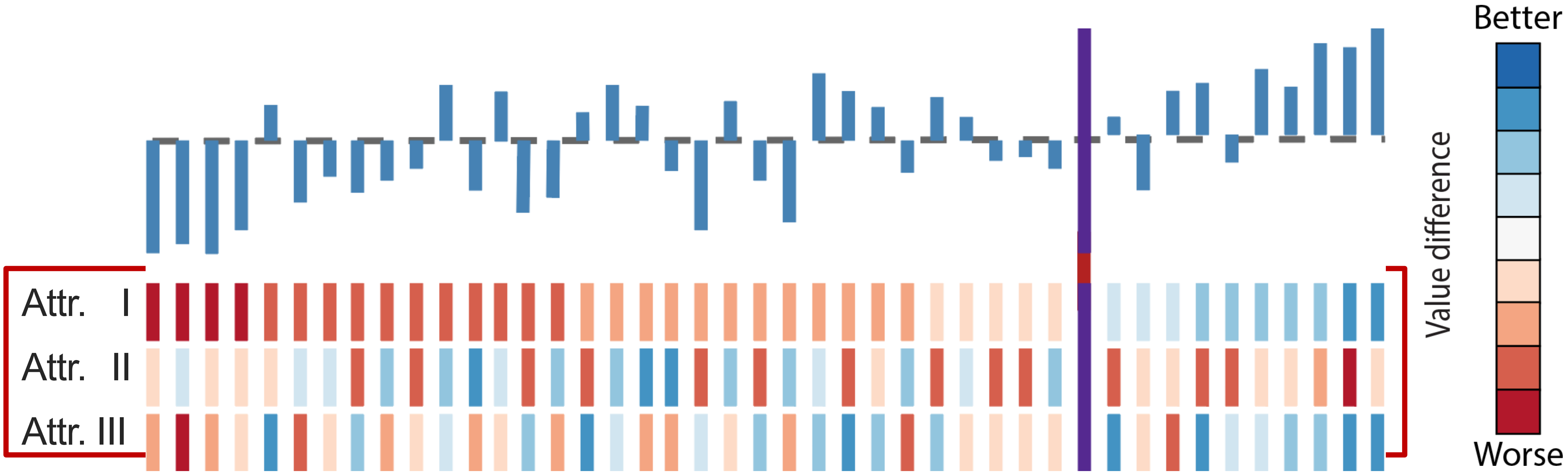
Bar length: other skyline points' average value differences compared with point A

Tabular View

- Table cell interaction: expanding a row for detailed information

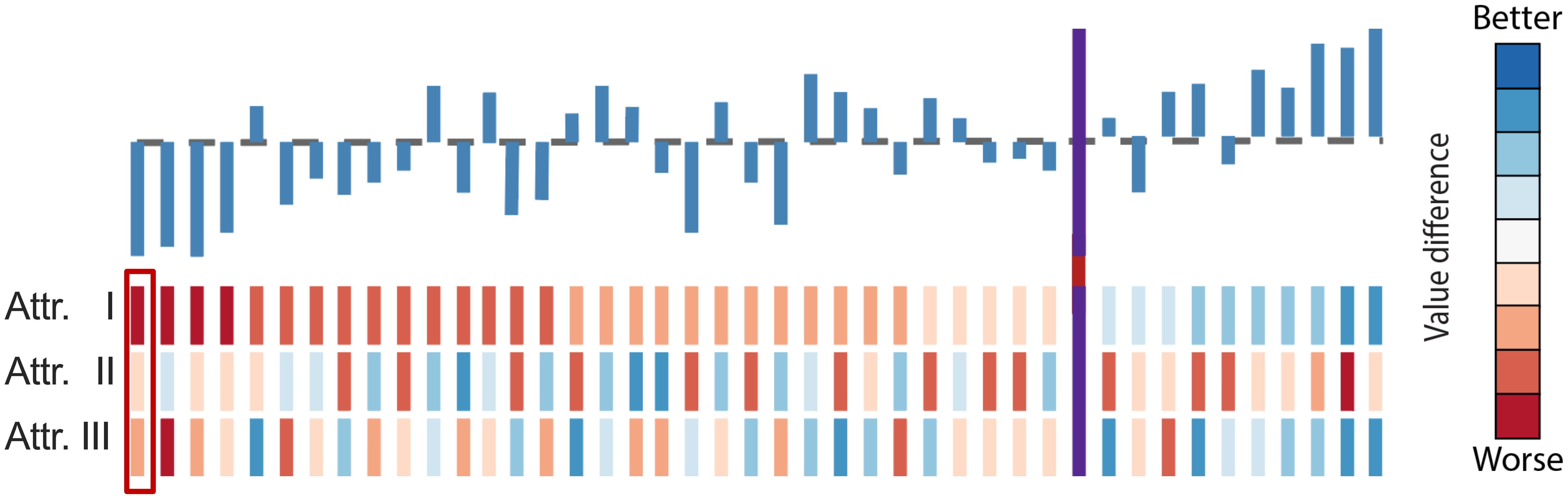
Tabular View

- Table cell interaction: expanding a row for detailed information



Tabular View

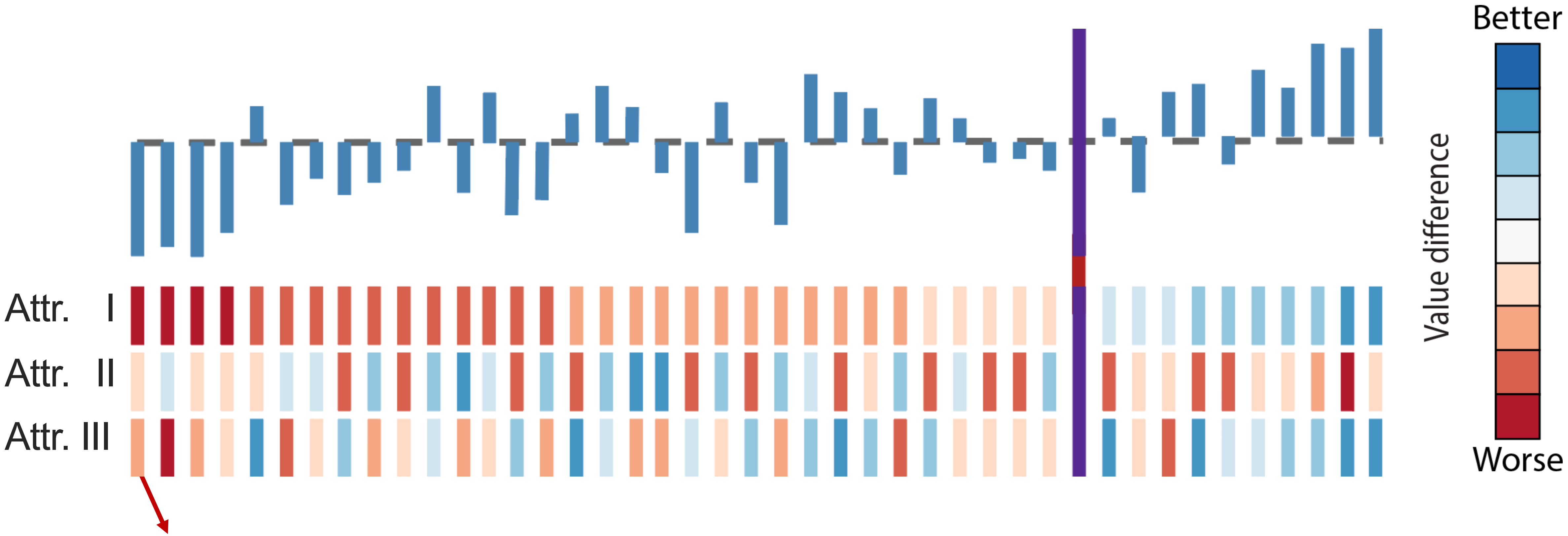
- Table cell interaction: expanding a row for detailed information



Each column is an extension of the corresponding vertical blue bar and represents the same skyline point

Tabular View

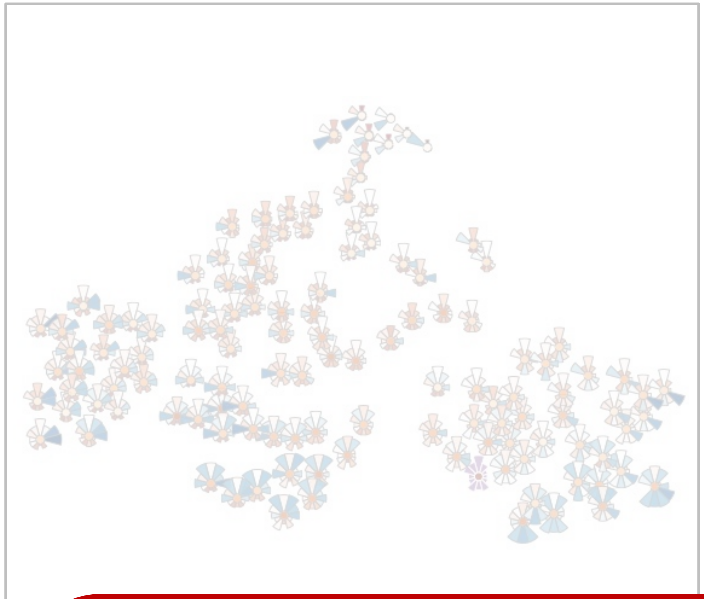
- Table cell interaction: expanding a row for detailed information



Matrix cell color: the value difference of the corresponding attribute (Attr. III)

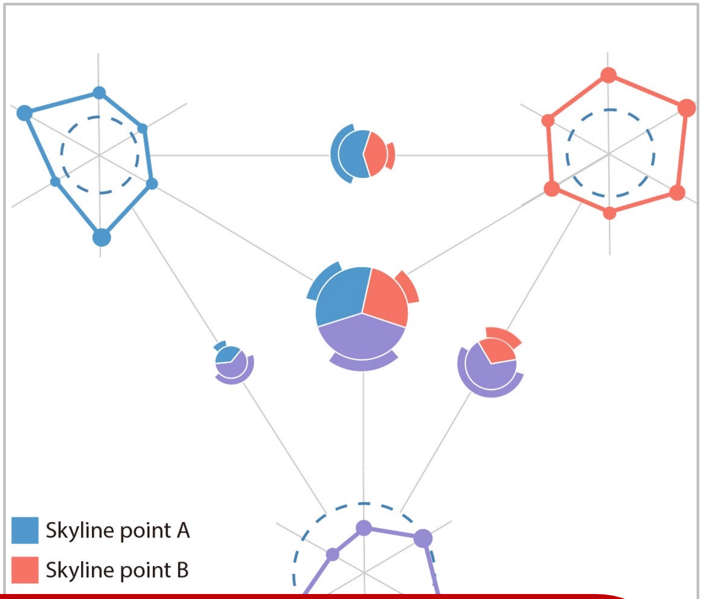
SkyLens – Comparison View

Projection View



Tabular View

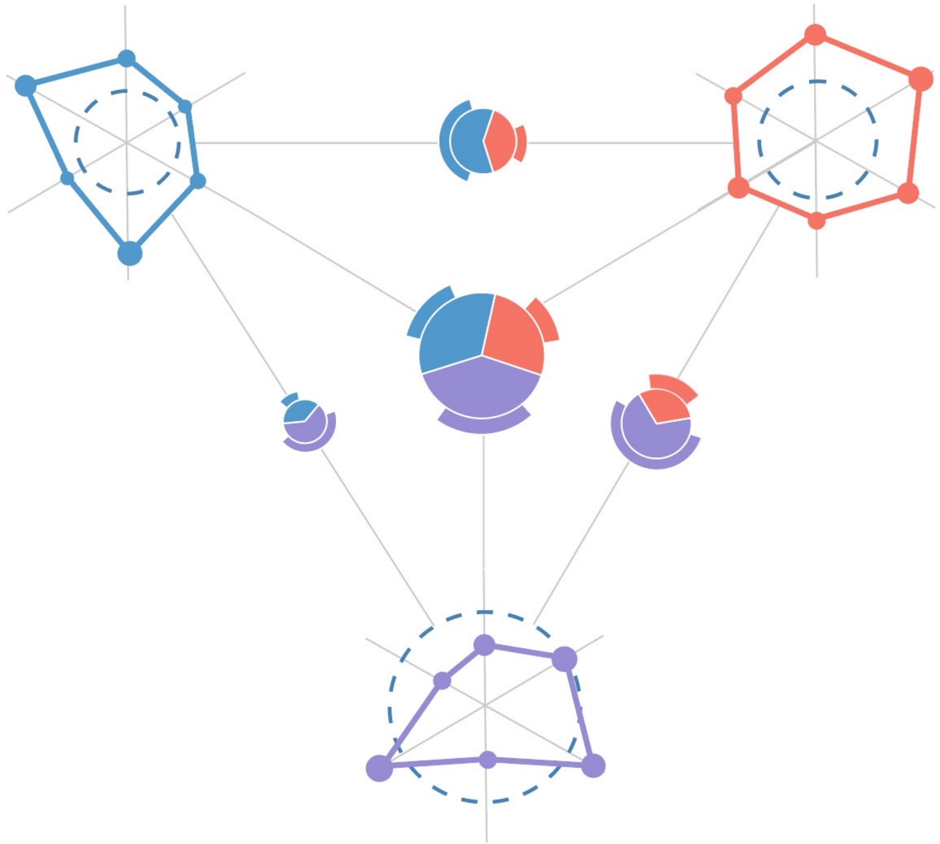
Comparison View



Comparison View: support a thorough comparison between skyline points

Comparison View

- Methods: radar charts & domination glyphs
  - Comparing attribute values
  - Examining dominating scores
  - Investigating dominated points

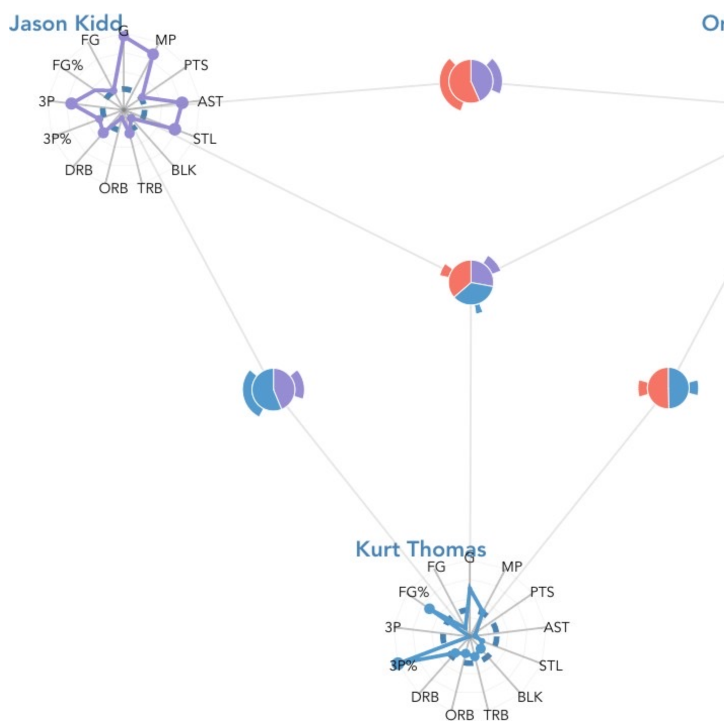


Goal: a thorough comparison on 2 ~ 5 skyline points

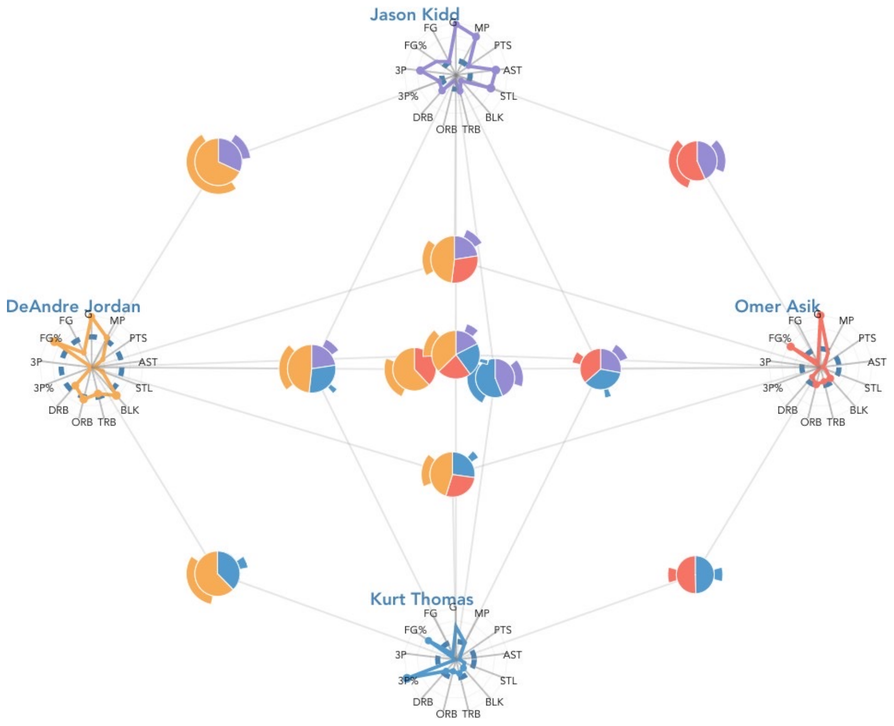
Comparison View

- Radial layout for the radar charts & domination glyphs

2-point comparison

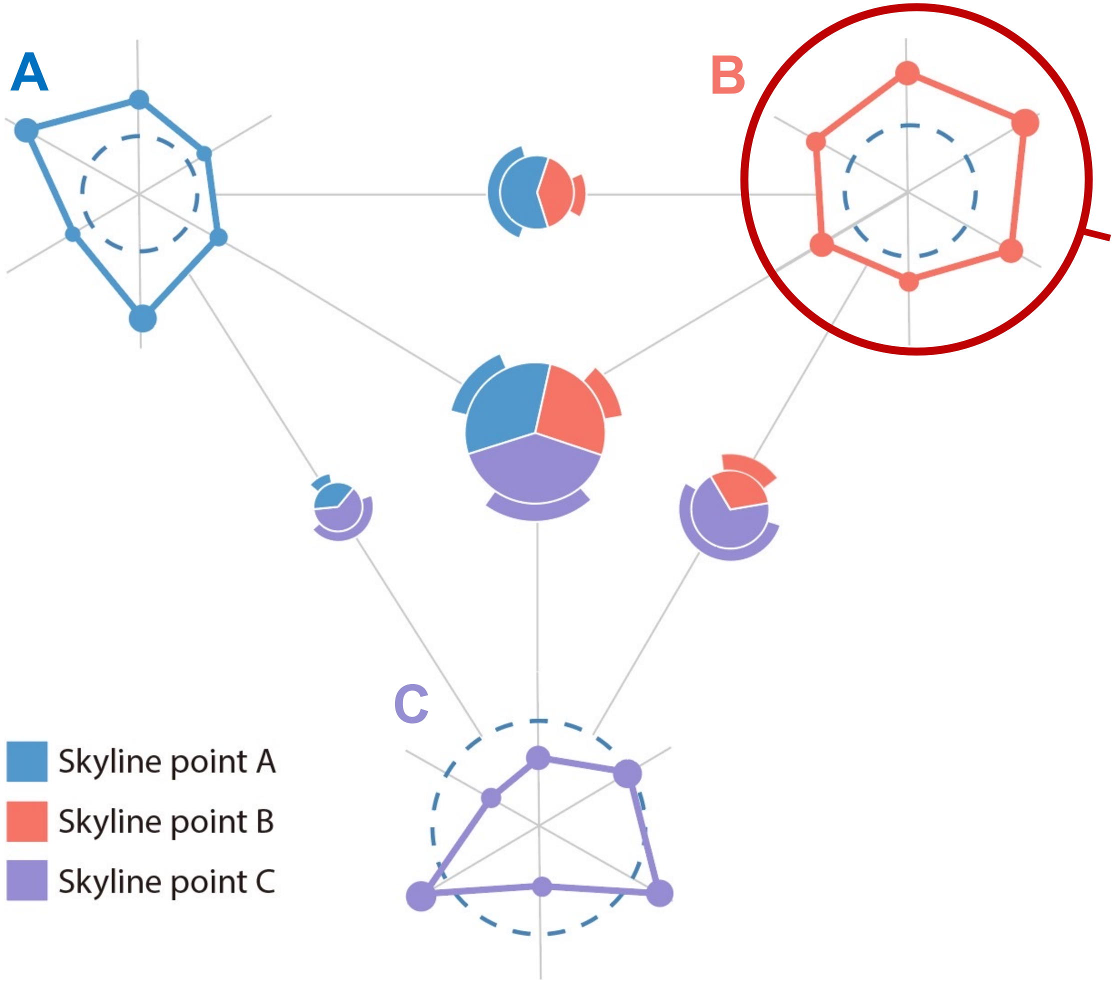


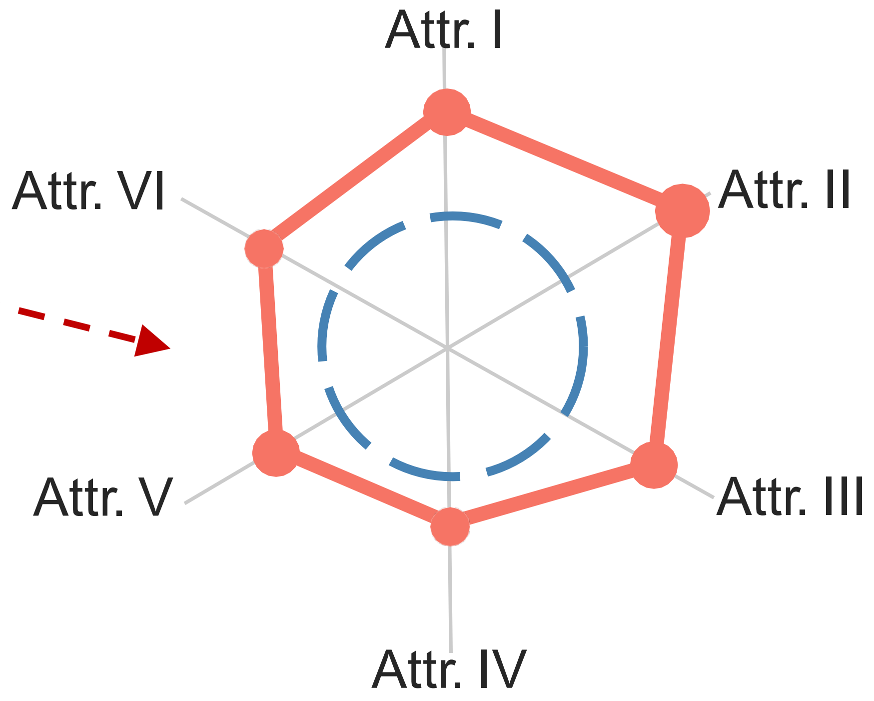
3-point comparison



4-point comparison

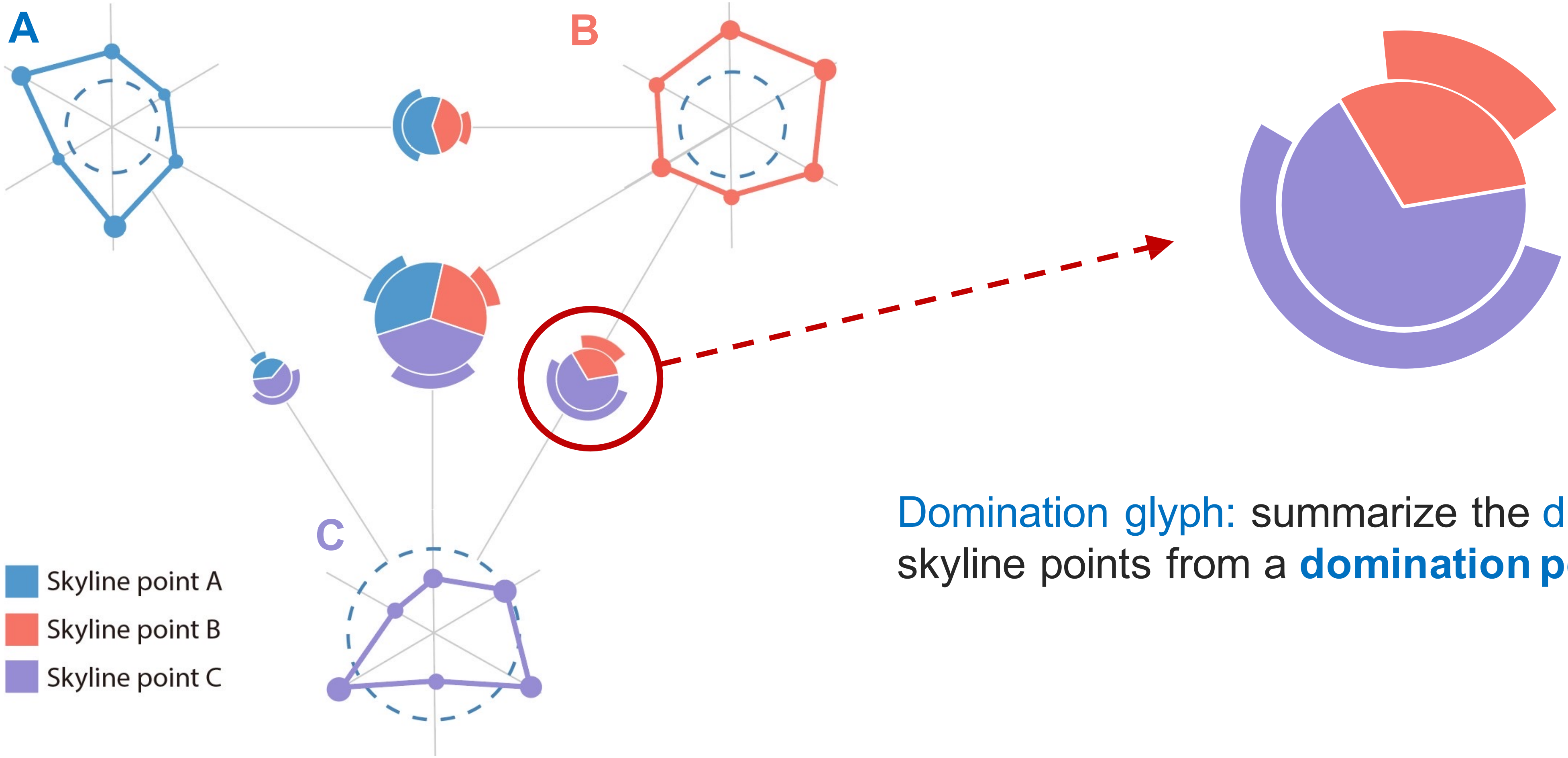
Comparison View – Radar Chart





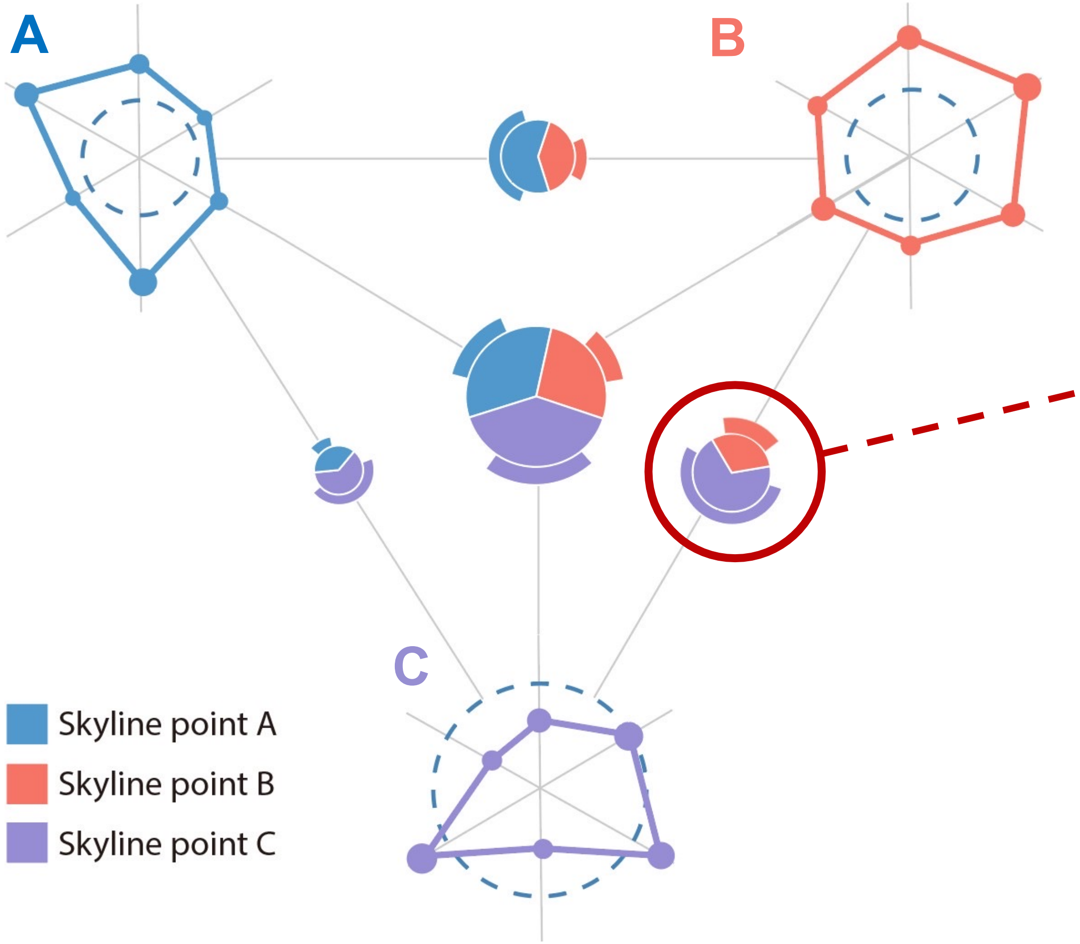
The radar chart encodes:
attribute values and dominating score

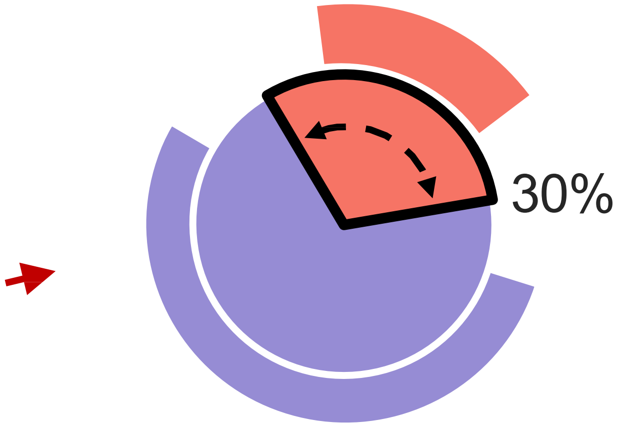
Comparison View – Domination Glyph



Domination glyph: summarize the differences among skyline points from a **domination perspective**.

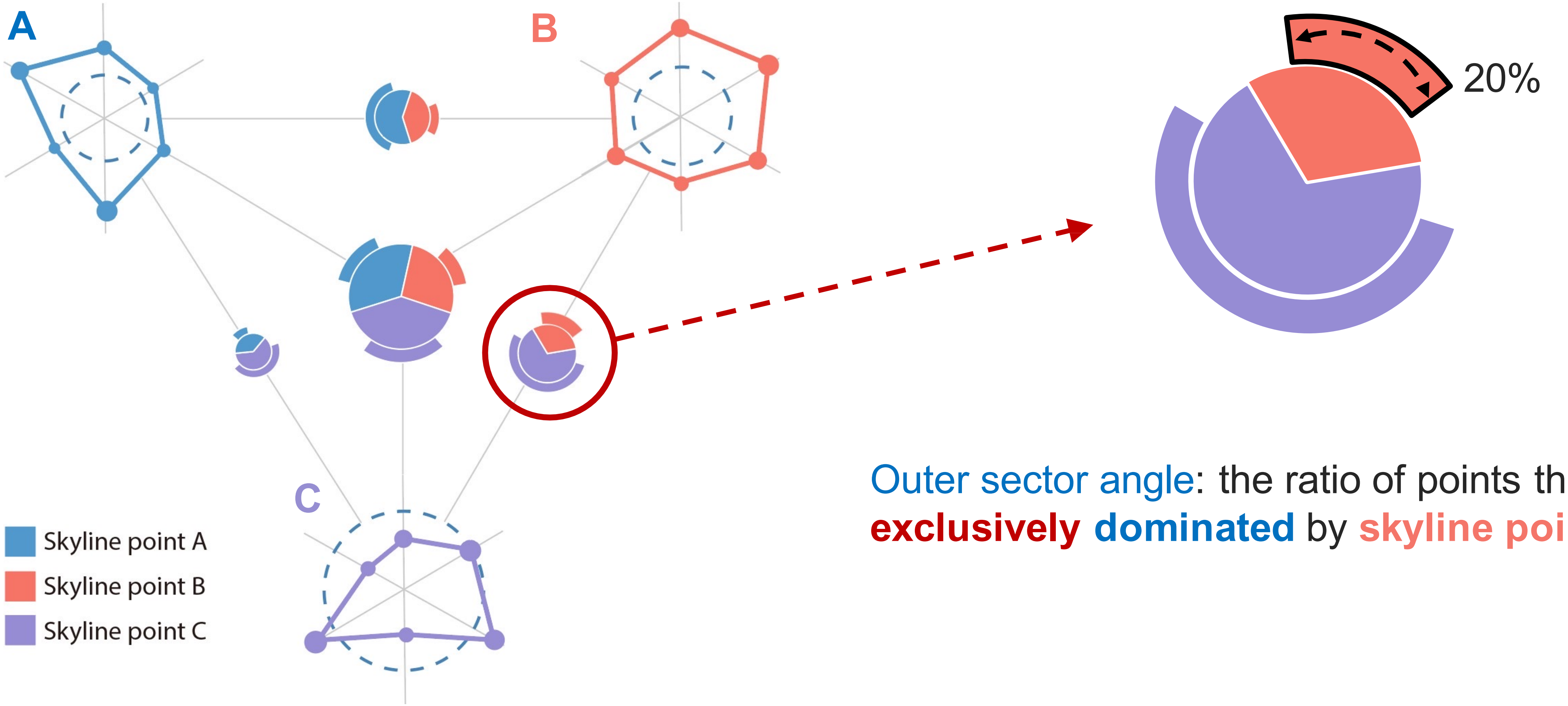
Comparison View – Domination Glyph





Inner sector angle: the ratio of points that are dominated by **skyline point B**

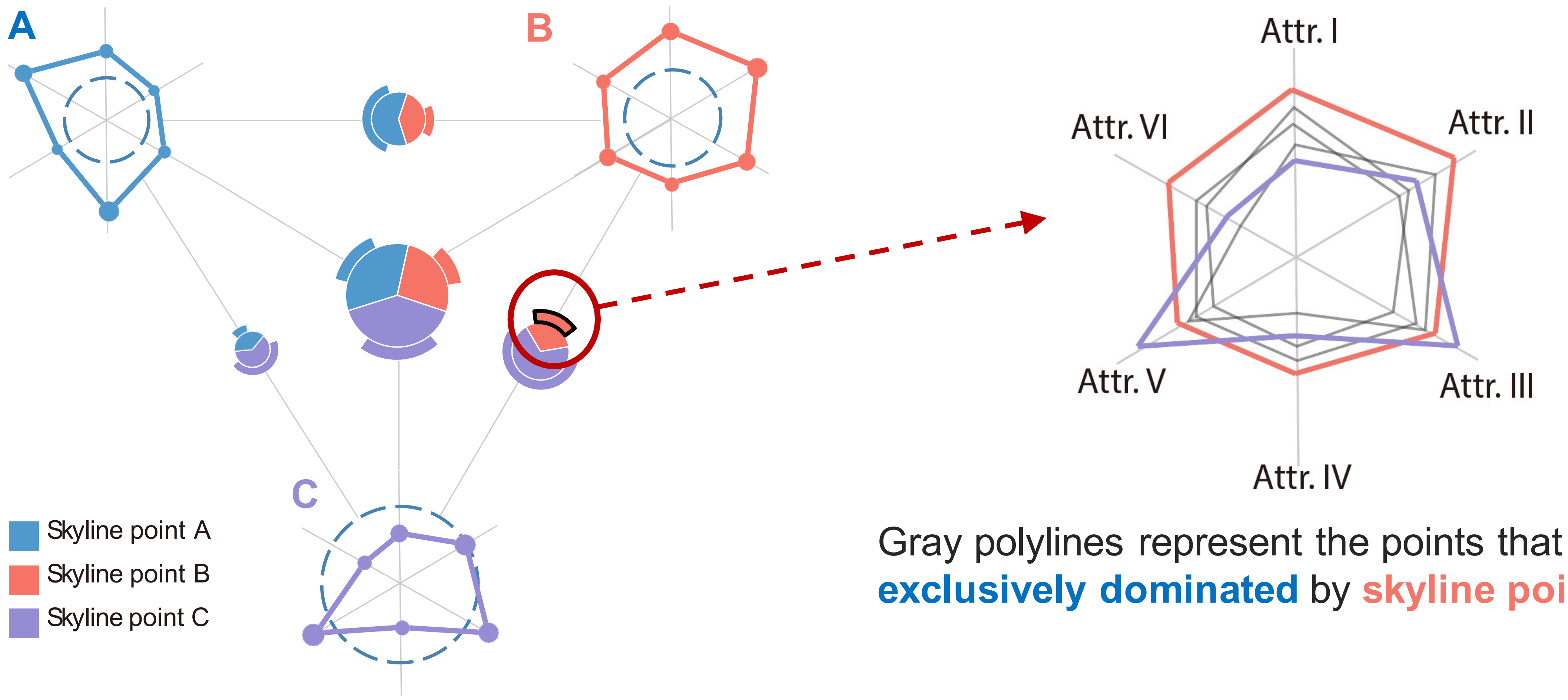
Comparison View – Domination Glyph



Outer sector angle: the ratio of points that are **exclusively dominated** by **skyline point B**

Comparison View – Domination Glyph Interaction

Hovering interaction: pop-up window showing the overlaid radar chart



Gray polylines represent the points that are **exclusively dominated** by **skyline point B**

VAST PAPER

SkyLens: Visual Analysis of Skyline on Multi-dimensional Data

Xun Zhao, Yanhong Wu, Weiwei Cui, Xinnan Du, Yuan Chen, Yong Wang, Dik-Lun Lee, Huamin Qu

Questions?
Thank you ☺

Quan Li

liquan@shanghaitech.edu.cn