Interactive Axis

InterAxis: Steering Scatterplot Axes via Observation-Level Interaction

IEEE VAST 2015

Hannah Kim[1], Jaegul Choo[2], Haesun Park[1], Alex Endert[1]

Georgia Tech[1], Korea University[2]

October 28th 2015

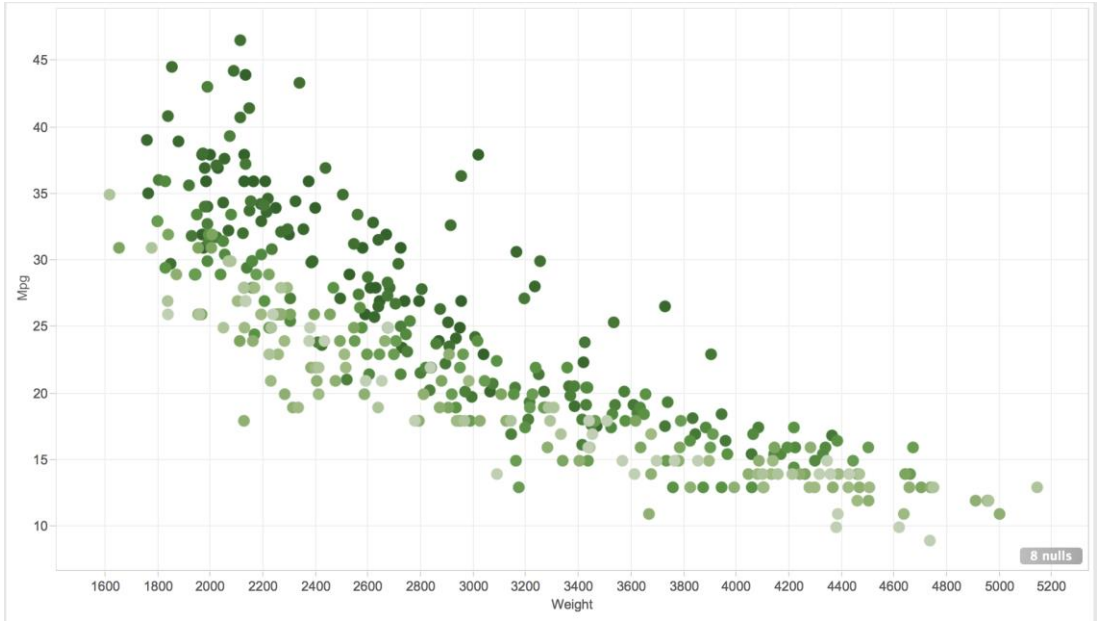
How would you visualize **high-dimensional** data?

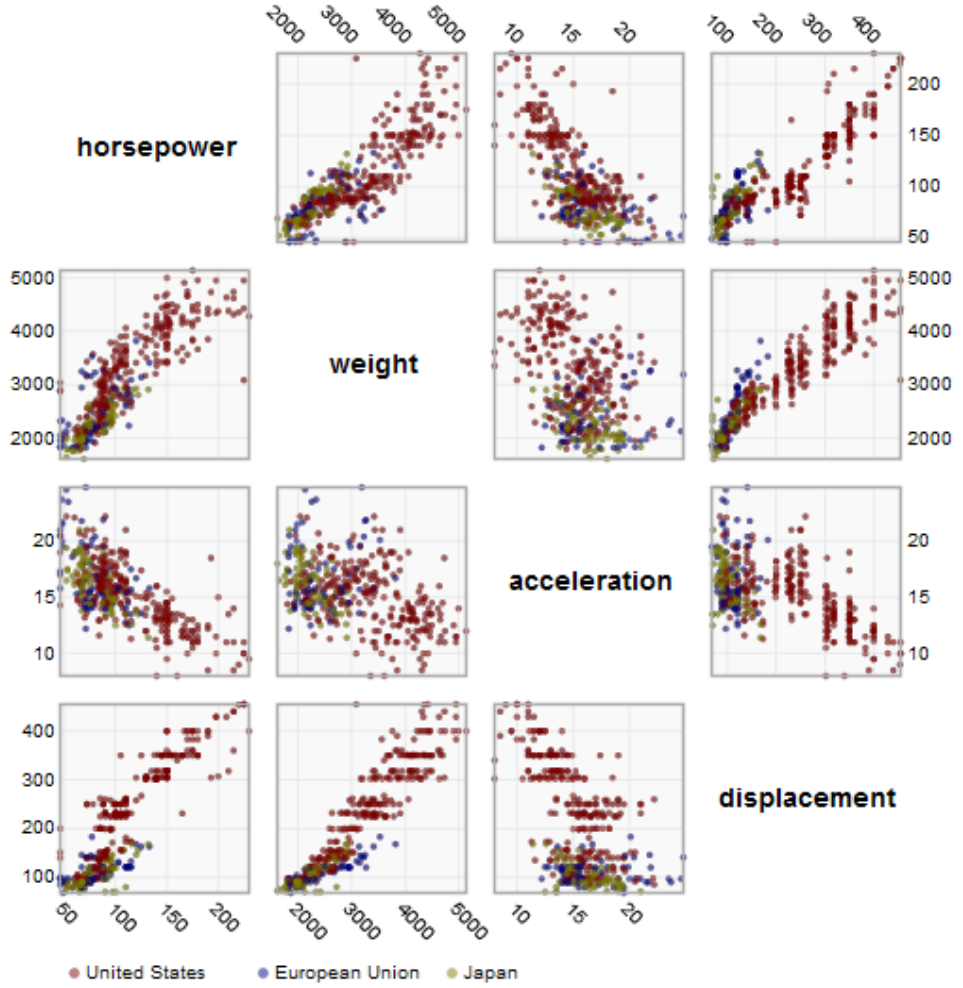
Example - Car Data

Vehicle Name	Retail Price	Dealer Cost	Engine Size (l)	Cyl	HP	City MPG	Hwy MPG	Weight	Wheel Base	Len	Width
Acura 3.5 RL 4dr	43,755	39,014	3.5	6	225	18	24	3,880	115	197	72
Acura 3.5 RL w/Navigation 4dr	46,100	41,100	3.5	6	225	18	24	3,893	115	197	72
Acura MDX	36,945	33,337	3.5	6	265	17	23	4,451	106	189	77
Porsche 911 GT2 2dr	192,465	173,560	3.6	6	477	17	24	3,131	93	175	72
Acura RSX Type S 2dr	23,820	21,761	2	4	200	24	31	2,778	101	172	68
Acura TL 4dr	33,195	30,299	3.2	6	270	20	28	3,575	108	186	72
Acura TSX 4dr	26,990	24,647	2.4	4	200	22	29	3,230	105	183	69
Audi A4 1.8T 4dr	25,940	23,508	1.8	4	170	22	31	3,252	104	179	70
Audi A4 3.0 4dr	31,840	28,846	3	6	220	20	28	3,462	104	179	70
					…						

Source: http://www.idvbook.com/teaching-aid/data-sets/2004-cars-and-trucks-data/

Scatterplot / Scatterplot Matrix

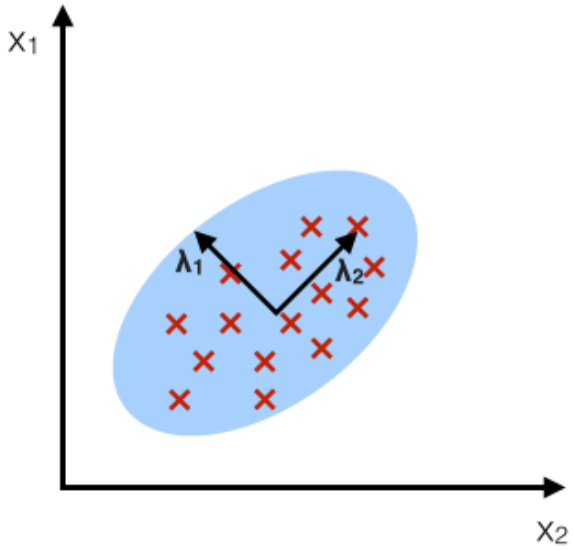


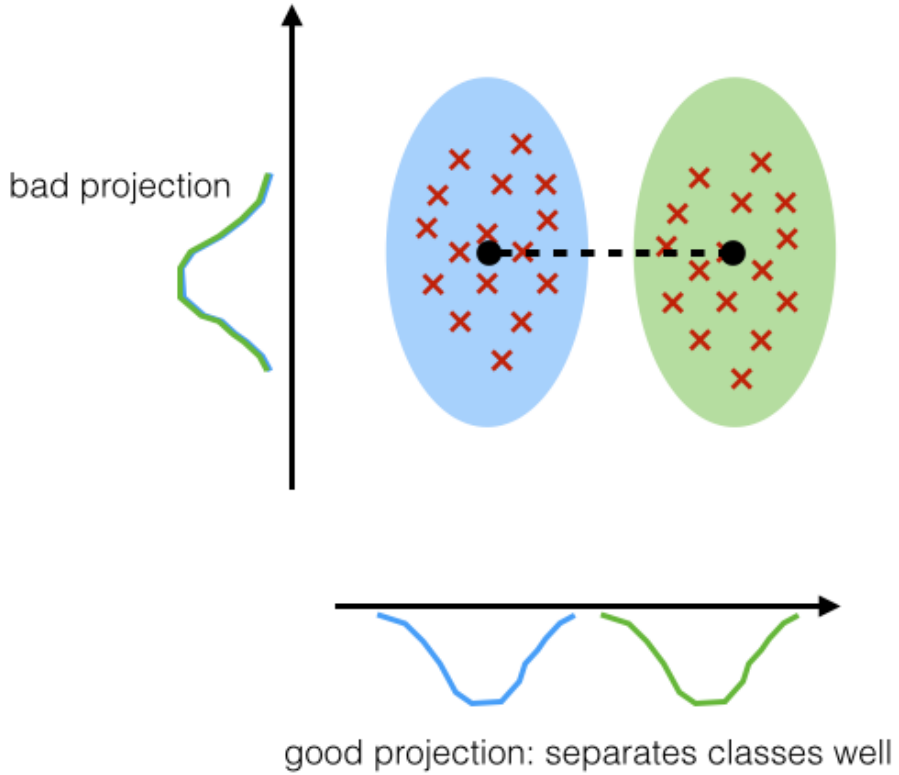


Source: https://datagab.wordpress.com/, http://homes.cs.washington.edu/~jheer/files/zoo/

Dimension Reduction: Linear

- Principal Component Analysis (PCA)
  - Linear combination of attributes that maximizes the variance
- Linear Discriminant Analysis (LDA)
  - Linear combination of attributes that separates classes well

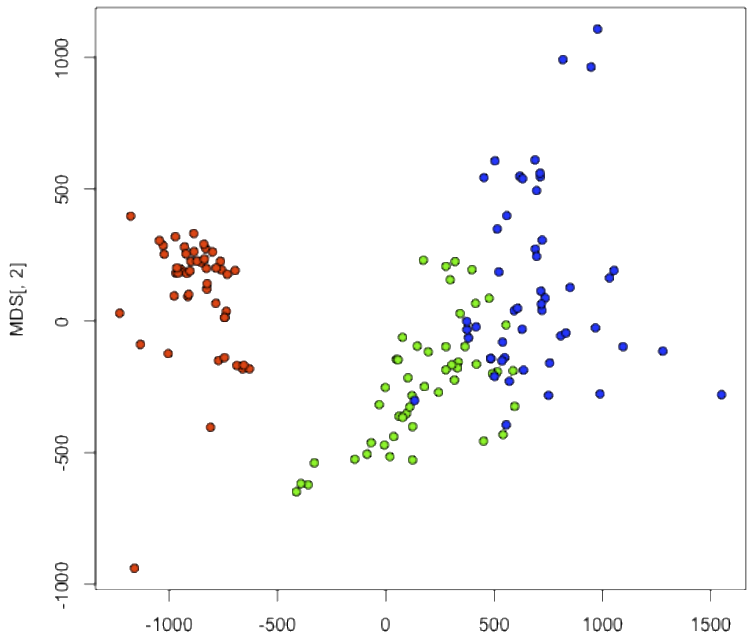


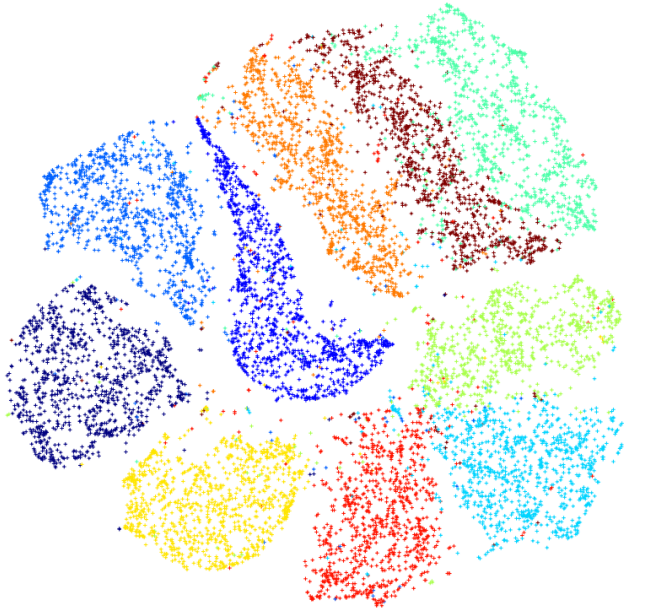


Source: http://sebastianraschka.com/Articles/2014_python_lda.html

Dimension Reduction: Nonlinear

- Multidimensional Scaling (MDS)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Force-Directed Layout



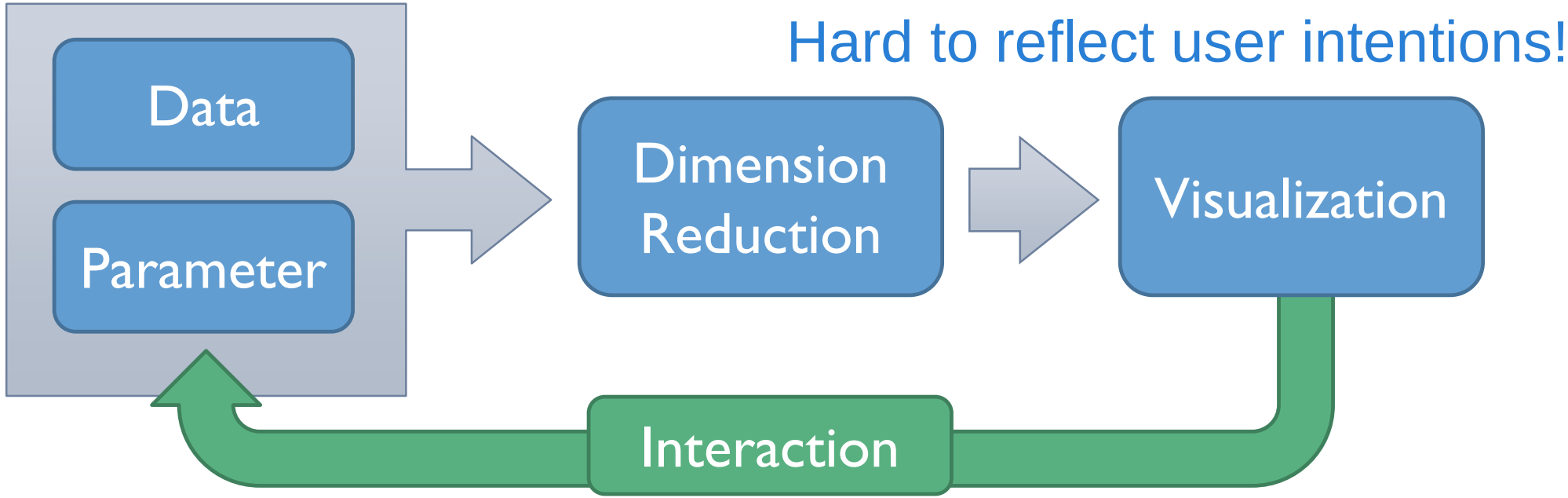


Interpretability Issue -
Axes do not have clear meaning or are not defined at all.

Source: http://statweb.stani

Another Issue – Interactivity

- Dimension reduction techniques are generally automated.
- Interaction with dimension reduction techniques is not easy.
- To make adjustments, one has to try different parameters and check visualization results.



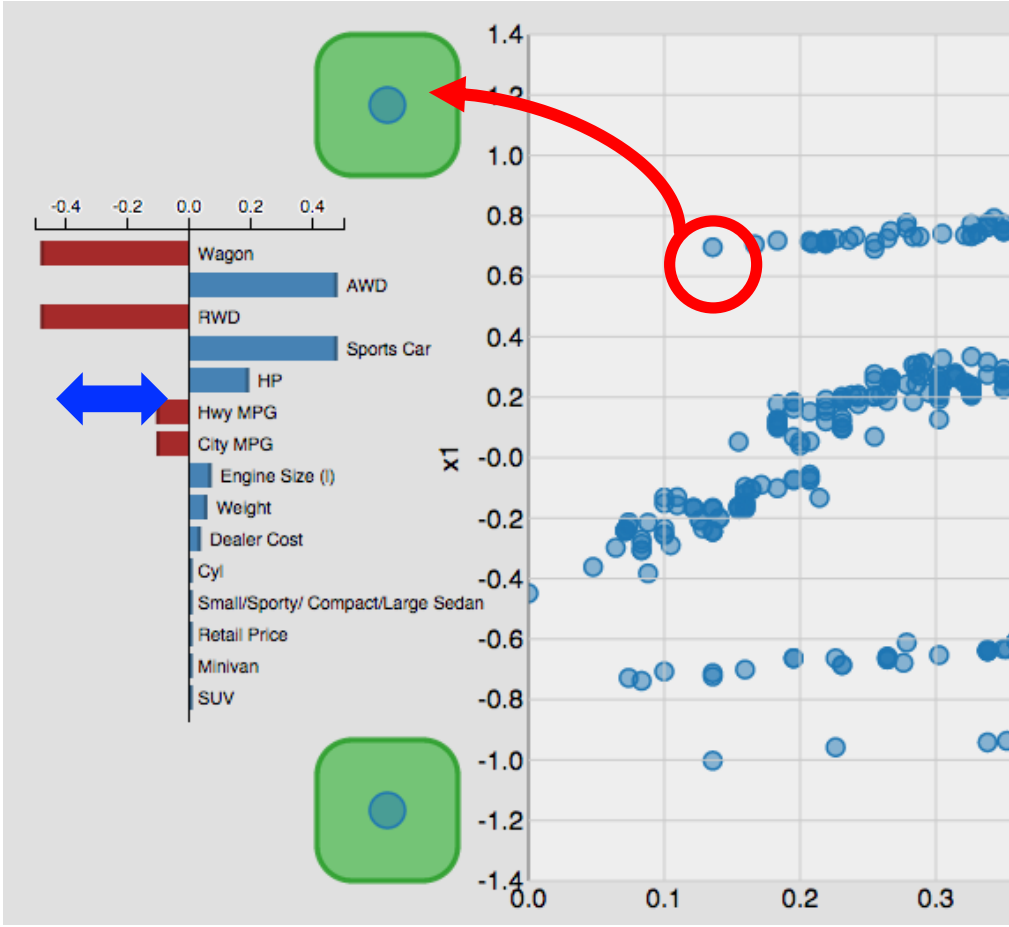
Can scatterplot axes be **interpretable** and **interactive**?

With InterAxis…

- Users pick two (or more) data items with semantic meanings
  - Ones they like vs. ones they don't like
  - Ones they are interested in vs. ones they don't care
  - and so on …
- Then, InterAxis automatically calculates an axis that reflects the semantic meanings.
  - Data items similar to the first group has high values and data items similar to the second group has low values.
  - Each feature's contribution(weight) to the axis is visualized in a bar chart.

Observation-Level Interaction

- Direct interactions with visual objects to reflect user intent
  - **Data-level:** data objects (dots) Find data items that quantify subjective preferences
  - **Feature-level:** features (bars) Directly manipulate contributions/ weights of features that represent an axis

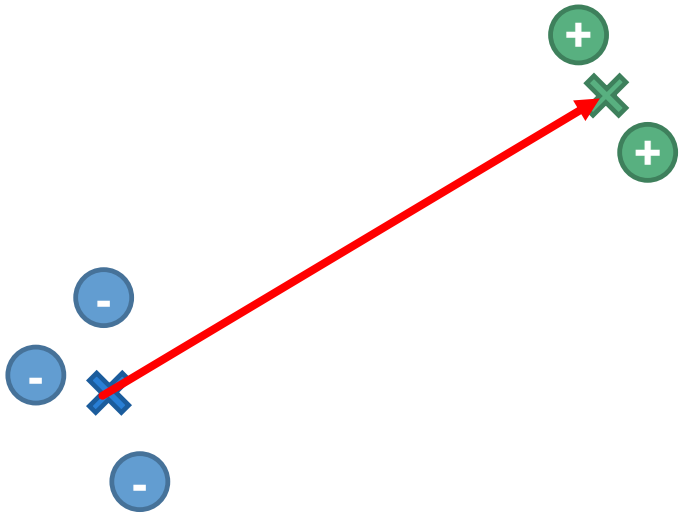


Live Demo – Car Data

http://va.gatech.edu/projects/interaxis/

Behind The Scenes (1)

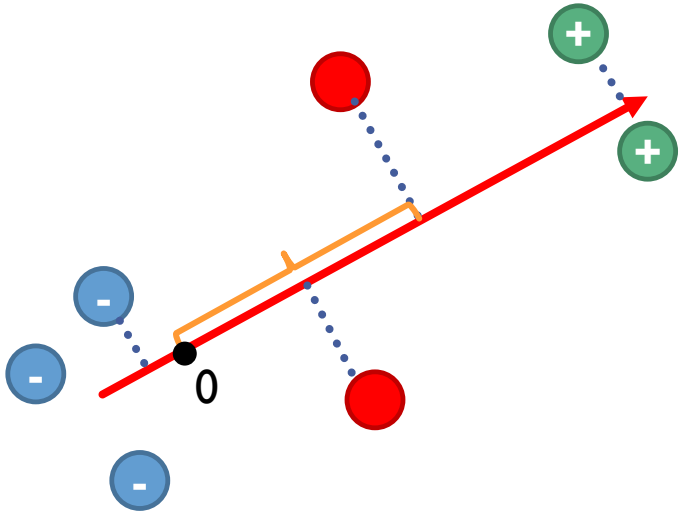
- Select two semantic groups of data items
- Find centroids of each group
- Subtract one from the other to get a projection vector
- Normalize the vector



$$T_x = \frac{1}{n_{x,h}} \sum_{i=1}^{n_{x,h}} a_i^{x,h} - \frac{1}{n_{x,l}} \sum_{i=1}^{n_{x,l}} a_i^{x,l}.$$

Behind The Scenes (2)

- Project origin (0,0) onto the axis vector
- Project data items onto the axis vector
- New coordinate value is the distance from the projected origin to the projected data point.



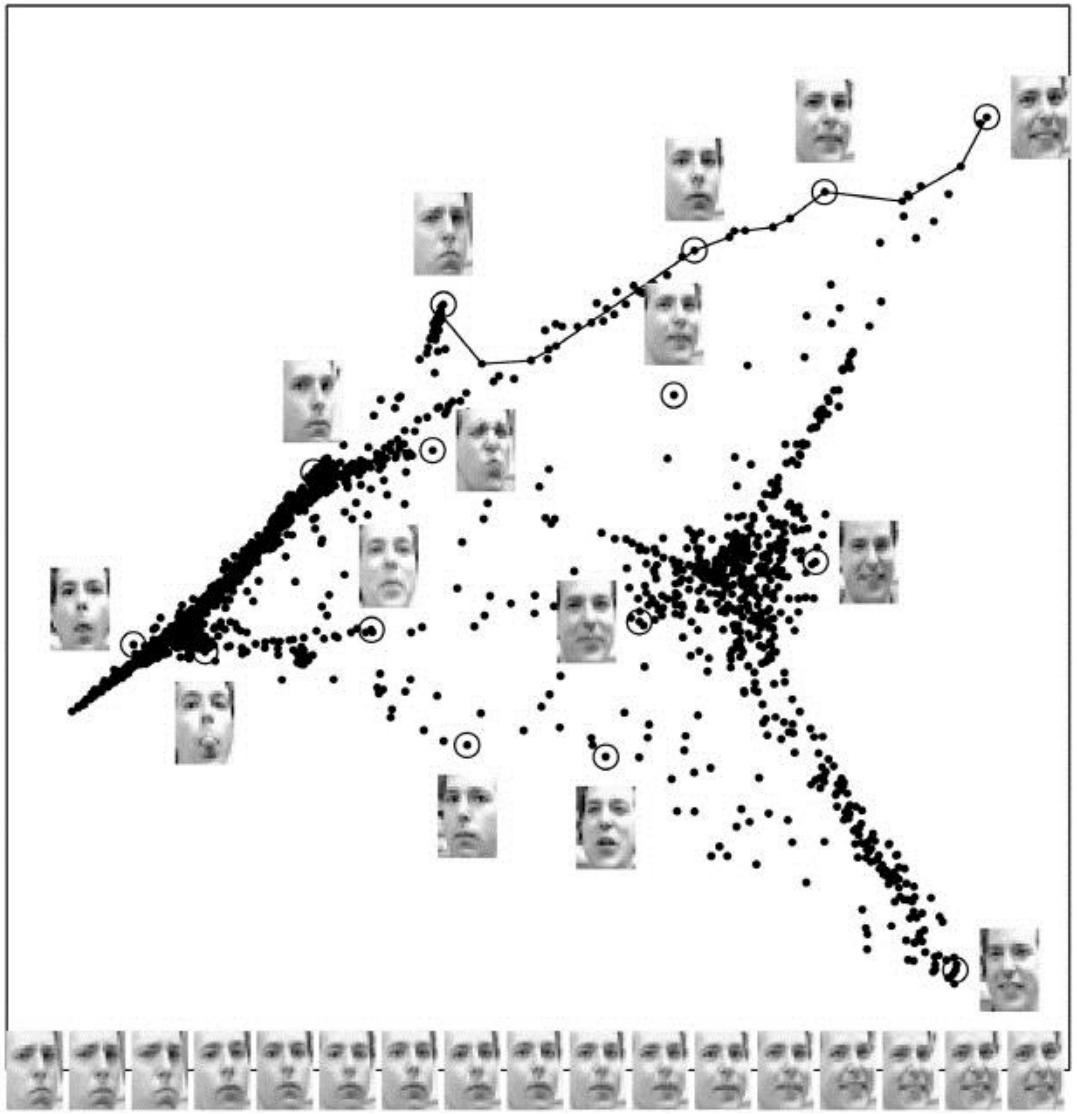
$$\left((T_x)^T a_i, (T_y)^T a_i \right).$$

Discussion: Beyond Linear Models (1)

	Automated	User-Driven
Linear	PCA	InterAxis
Non-linear	Manifold learning (e.g., LLE, Isomap)	?

Discussion: Beyond Linear Models (2)

- InterAxis represents an axis as a weighted linear combination of data attributes.
- However, semantic meanings are not necessarily “linear”.
  - E.g., drawing a curve that means a progression of frowning to smiling

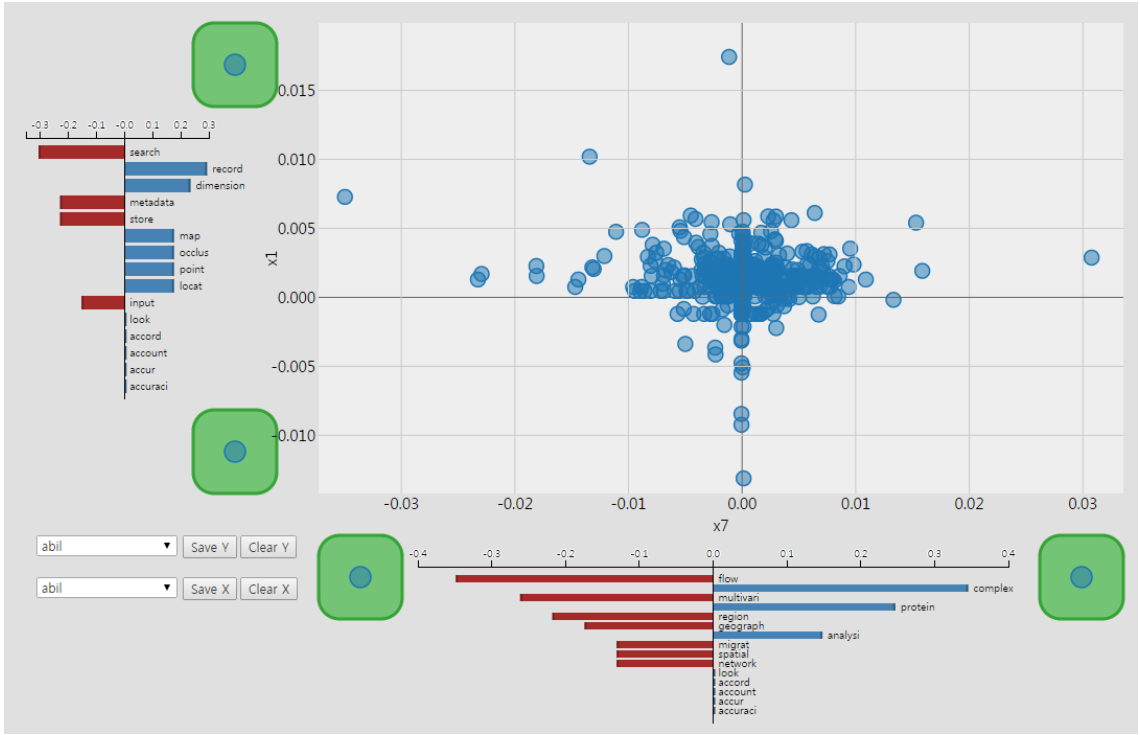


Source: Saul and Roweis, locally linear embedding (LLE) paper (JMLR'13)

Discussion: Handling Sparse Data

- Few non-zero entries for an attribute (or for an item)
- Common in significantly high-dimensional data
  - E.g. text, image, gene expression data

- We have to assign more data items to specify an axis.
- One solution is to aggregate multiple attributes into a group.



Summary

- We introduce **InterAxis**, a visual analytics technique that enables users to
  - Directly define and manipulate axes via observation-level interactions
  - Understand data attributes that quantify subjective preferences