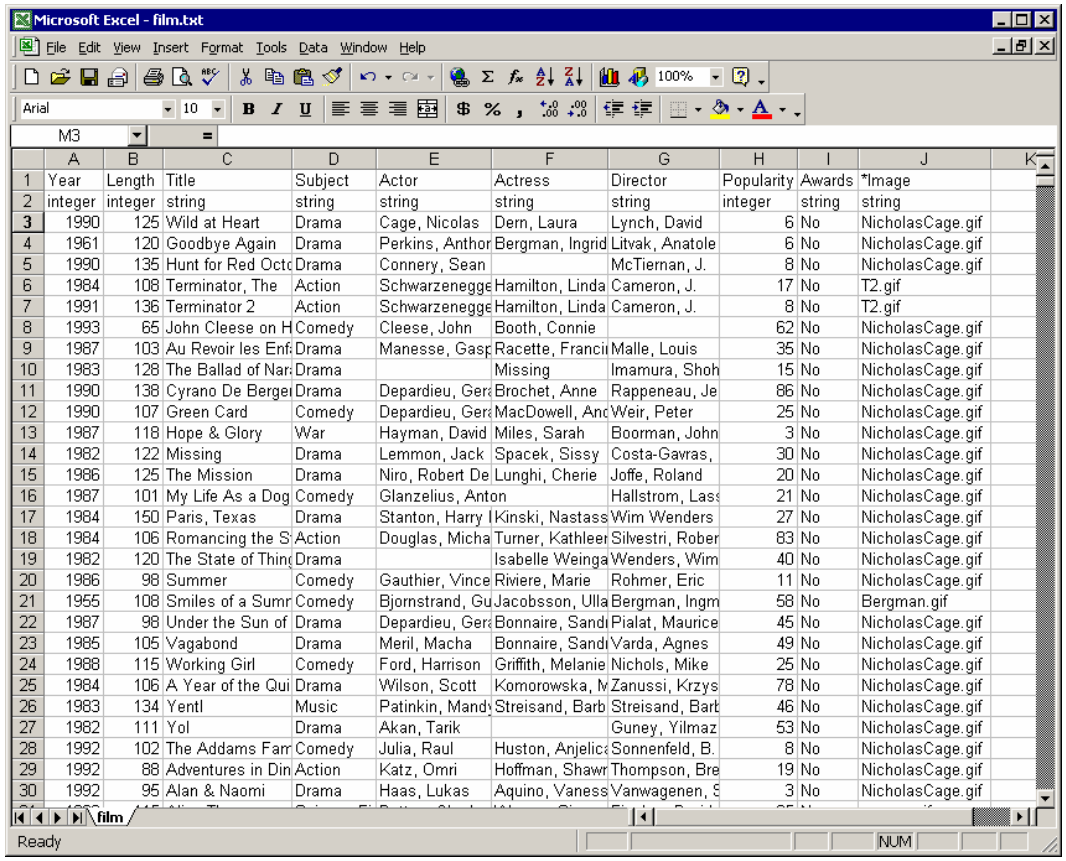
High Dimensional Visualization

By

Mingyue Tan

Mar10, 2004

High Dimensional Data



	A	B	C	D	E	F	G	H	I	J
1	Year	Length	Title	Subject	Actor	Actress	Director	Popularity	Awards	*Image
2	integer	integer	string	string	string	string	string	integer	string	string
3	1990	125	Wild at Heart	Drama	Cage, Nicolas	Dern, Laura	Lynch, David	6	No	NicholasCage.gif
4	1961	120	Goodbye Again	Drama	Perkins, Anthor	Bergman, Ingrid	Litvak, Anatole	6	No	NicholasCage.gif
5	1990	135	Hunt for Red Octo	Drama	Connery, Sean		McTiernan, J.	8	No	NicholasCage.gif
6	1984	108	Terminator, The	Action	Schwarzenegge	Hamilton, Linda	Cameron, J.	17	No	T2.gif
7	1991	136	Terminator 2	Action	Schwarzenegge	Hamilton, Linda	Cameron, J.	8	No	T2.gif
8	1993	65	John Cleese on H	Comedy	Cleese, John	Booth, Connie		62	No	NicholasCage.gif
9	1987	103	Au Revoir les Enf	Drama	Manesse, Gasp	Racette, Francii	Malle, Louis	35	No	NicholasCage.gif
10	1983	128	The Ballad of Nar	Drama		Missing	Imamura, Shoh	15	No	NicholasCage.gif
11	1990	138	Cyrano De Berge	Drama	Depardieu, Ger	Brochet, Anne	Rappeneau, Je	86	No	NicholasCage.gif
12	1990	107	Green Card	Comedy	Depardieu, Ger	MacDowell, And	Weir, Peter	25	No	NicholasCage.gif
13	1987	118	Hope & Glory	War	Hayman, David	Miles, Sarah	Boorman, John	3	No	NicholasCage.gif
14	1982	122	Missing	Drama	Lemmon, Jack	Spacek, Sissy	Costa-Gavras,	30	No	NicholasCage.gif
15	1986	125	The Mission	Drama	Niro, Robert De	Lunghi, Cherie	Joffe, Roland	20	No	NicholasCage.gif
16	1987	101	My Life As a Dog	Comedy	Glanzelius, Anton		Hallstrom, Lass	21	No	NicholasCage.gif
17	1984	150	Paris, Texas	Drama	Stanton, Harry I	Kinski, Nastass	Wim Wenders	27	No	NicholasCage.gif
18	1984	106	Romancing the S	Action	Douglas, Micha	Turner, Kathleer	Silvestri, Rober	83	No	NicholasCage.gif
19	1982	120	The State of Thin	Drama		Isabelle Weinga	Wenders, Wim	40	No	NicholasCage.gif
20	1986	98	Summer	Comedy	Gauthier, Vince	Riviere, Marie	Rohmer, Eric	11	No	NicholasCage.gif
21	1955	108	Smiles of a Sumr	Comedy	Bjornstrand, Gu	Jacobsson, Ulla	Bergman, Ingm	58	No	Bergman.gif
22	1987	98	Under the Sun of	Drama	Depardieu, Ger	Bonnaire, Sandi	Pialat, Maurice	45	No	NicholasCage.gif
23	1985	105	Vagabond	Drama	Meril, Macha	Bonnaire, Sandi	Varda, Agnes	49	No	NicholasCage.gif
24	1988	115	Working Girl	Comedy	Ford, Harrison	Griffith, Melanie	Nichols, Mike	25	No	NicholasCage.gif
25	1984	106	A Year of the Qui	Drama	Wilson, Scott	Komorowska, M	Zanussi, Krzys	78	No	NicholasCage.gif
26	1983	134	Yentl	Music	Patinkin, Mandy	Streisand, Barb	Streisand, Barb	46	No	NicholasCage.gif
27	1982	111	Yol	Drama	Akan, Tarik		Guney, Yilmaz	53	No	NicholasCage.gif
28	1992	102	The Addams Fam	Comedy	Julia, Raul	Huston, Anjelica	Sonnenfeld, B.	8	No	NicholasCage.gif
29	1992	88	Adventures in Din	Action	Katz, Omri	Hoffman, Shawr	Thompson, Bre	19	No	NicholasCage.gif
30	1992	95	Alan & Naomi	Drama	Haas, Lukas	Aquino, Vaness	Vanwagenen, S	3	No	NicholasCage.gif

High-D data:

- ungraspable to a human's mind

What does a 10-D space look like?

We need effective multi-D visualization techniques

Paper Reviewed

- Dimensional Anchors: a Graphic Primitive for Multidimensional Multivariate Information Visualizations, P. Hoffman, G. Grinstein, & D. Prinkney, Proc. Workshop on New Paradigms in Information Visualization and Manipulation, Nov. 1999, pp. 9-16.
- Visualizing Multi-dimensional Clusters, Trends, and Outliers using Star Coordinates, Eser Kandogan, Proc. KDD 2001
- StarClass: Interactive Visual Classification Using Star Coordinates , S. Teoh & K. Ma, Proc. SIAM 2003

Dataset

- Car

  - contains car specs (eg. mpg, cylinders, weight, acceleration, displacement, type(origin), horsepower, year, etc)

  - type: American, Japanese, & European

Dimensional Anchors (DA)

Dimensional Anchor:

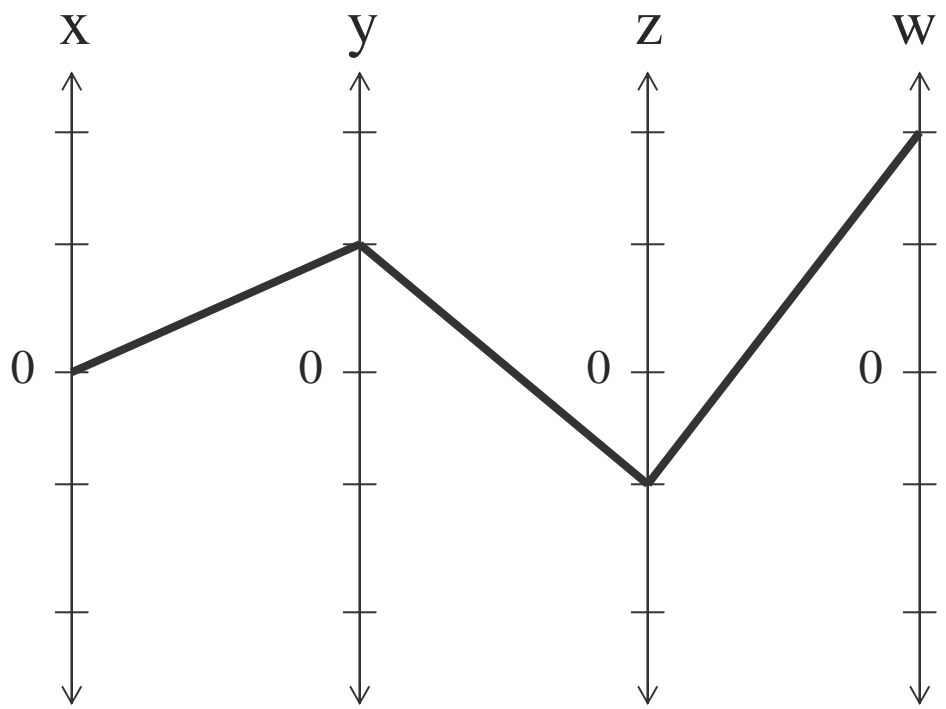
- Attempt to unify many different multi-var visualizations
- Uses of 9 DA parameters

Base Visualizations

- Scatter Plot
- Parallel Coordinates
- Survey Plot
- Radviz spring visualization

Parallel Coordinates

- Point -> line
- (0,1,-1,2)=



Base Visualizations

- Scatter Plot
- Parallel Coordinates
- Survey Plot
- Radviz spring visualization

Parameters of DA

- Nine parameters are selected to describe the graphics properties of each DA:

  p1: size of the scatter plot points

  p2: length of the perpendicular lines extending from individual anchorpoints in a scatter plot

  p3: length of the lines connecting scatter plot points that are associated with the same data point

  p4: width of the rectangle in a survey plot

  p5: length of the parallel coordinate lines

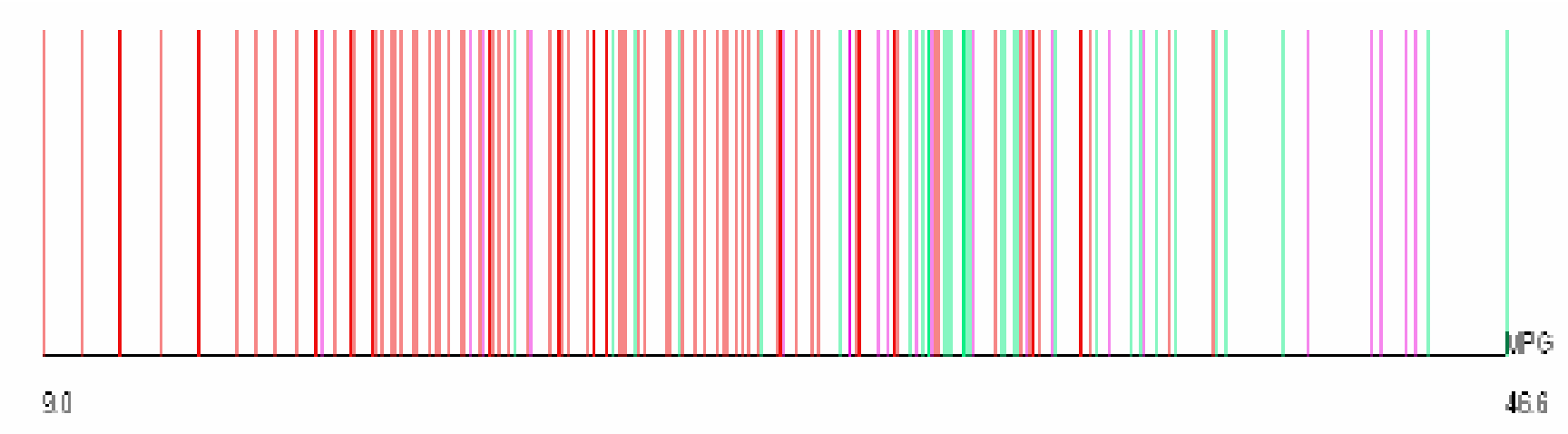
  p6: blocking factor for the parallel coordinate lines

  p7: size of the radviz plot point

  p8: length of the “spring” lines extending from individual anchorpoints of a radviz plot

  p9: the zoom factor for the “spring” constant K

Basic Single DA



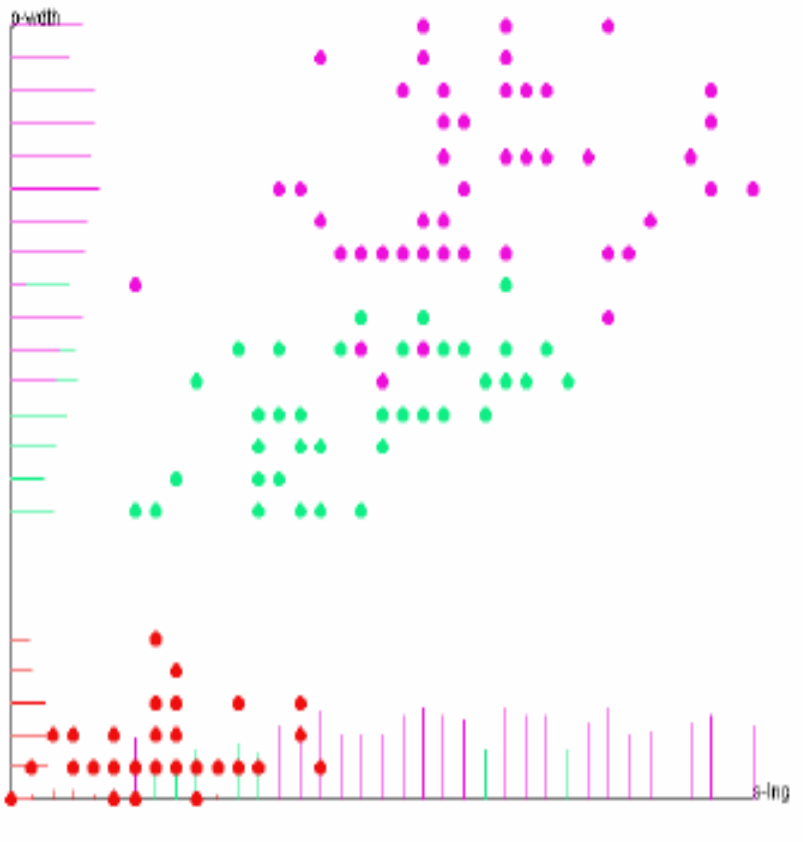
- Dimension – miles per gallon
- Data values are mapped to the axis
- Mapped data points - *anchorpoints*, represent the coord values(points along a DA)
- Lines extended from anchorpoints
- Color – type of car (American – red, Japanese – green, and European – purple)

Two-DA scatter plot

DA scatter plot using two DAs

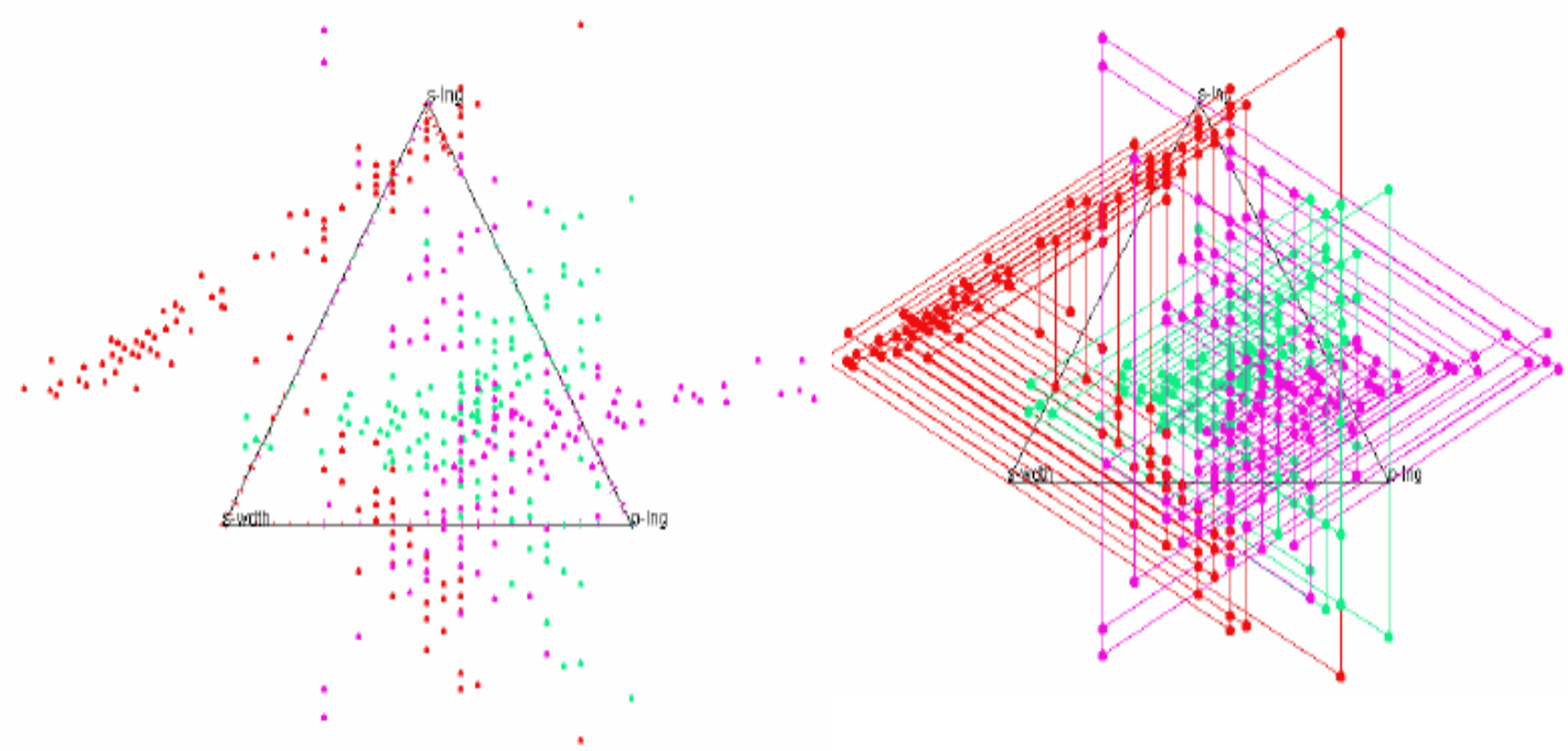
- Perpendicular lines extending outward from the anchor points
- If they meet, plot the point at the intersection

- p1: size of the scatter plot points
- p2: length of the perpendicular lines extending from individual anchor points in a scatter plot
- p3: length of the lines connecting scatter plot points that are associated with the same data point



P = (0.8, .2, 0, 0, 0, 0, 0, 0, 0)

Three DAs



P = (0.6, 0, 0, 0, 0, 0, 0, 0, 0)

P = (.6, 0, 1.0, 0, 0, 0, 0, 0, 0)

P3: length of lines connecting all displayed points associated with one real data point(record)

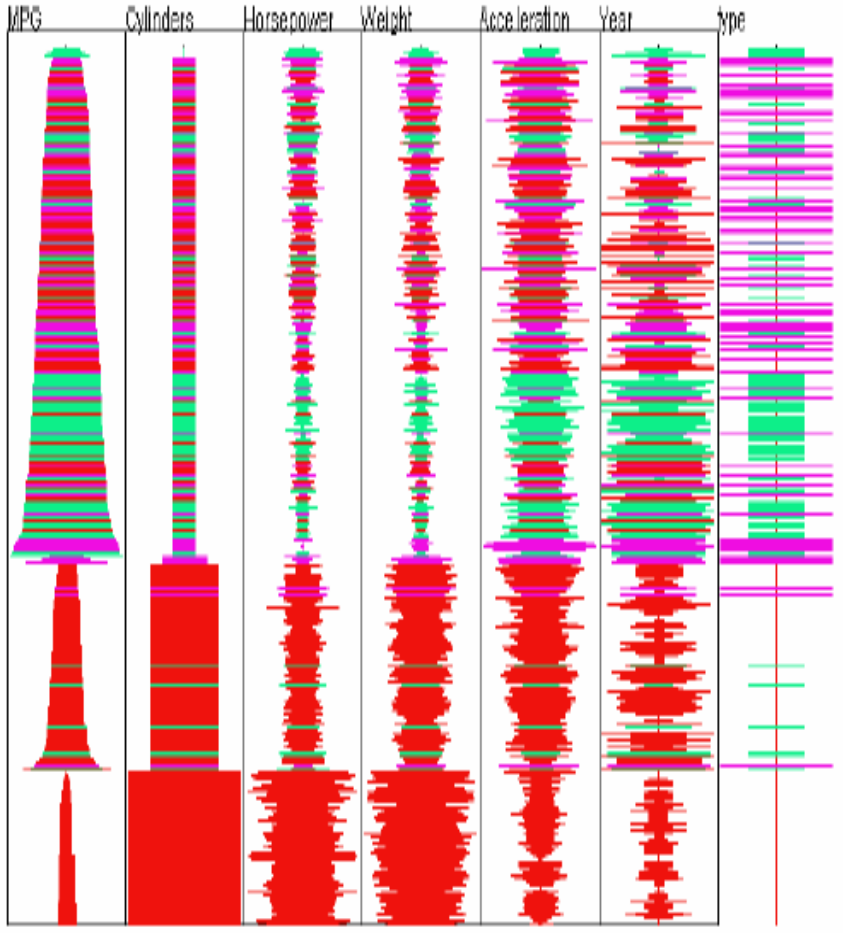
Seven DA Survey Plot

- 7 vertical DAs in a row
- Rectangle extending from an anchor point

  - size is based on the dimensional value

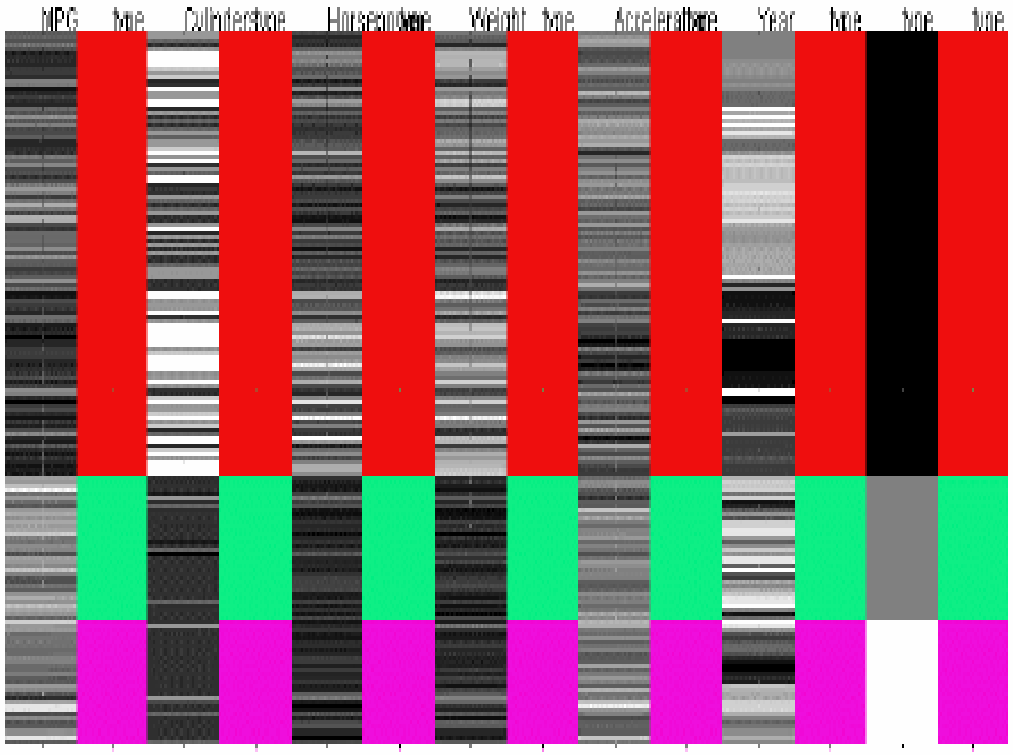
  - eg. Type- discrete value

    red < green < purple



CCCViz – Color Correlated Column

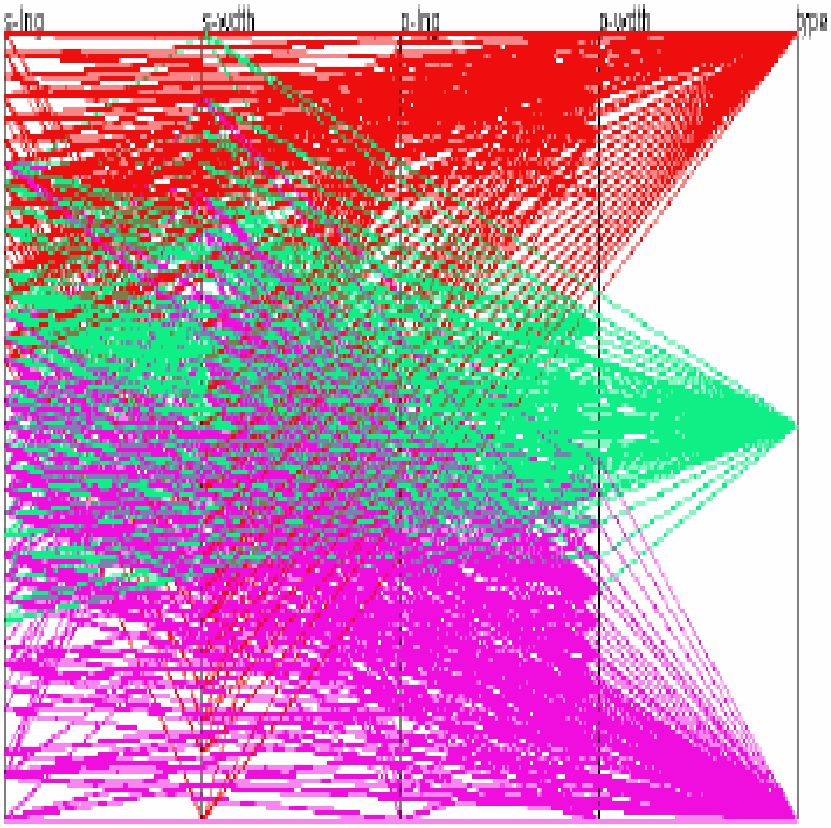
- Does a dimension (gray scales) correlate with a particular classification dimension(color scale) ?
- Correlation is seen in mpg, cylinders etc.
- p4: width of the rectangle in a survey plot



CCCViz DAs with P = (0, 0, 0, 1.0, 0, 0, 0, 0, 0)

DAs in PC configuration

- Line from one DA anchorpoint is drawn to another
  - length of these connecting lines is controlled by p5.
  - p5 = 1.0, fully connected, every anchorpoint connects to all the other (N-1) anchorpoints
- P6 controls how many DAs a p5 connecting line can cross
  - p6 = 0, traditional PC



P = (0, 0, 0, 0, 1.0, 1.0, 0, 0, 0)

DAs in Regular Polygon

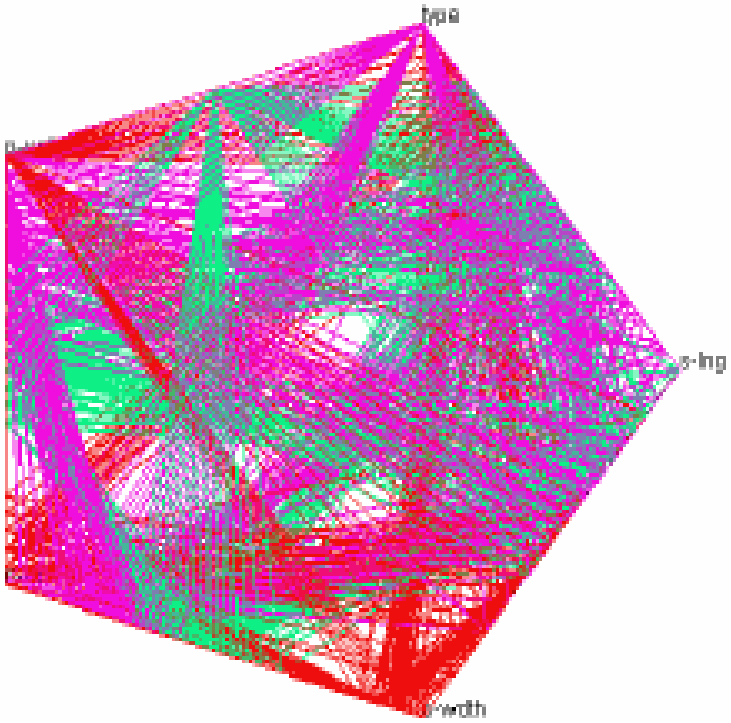
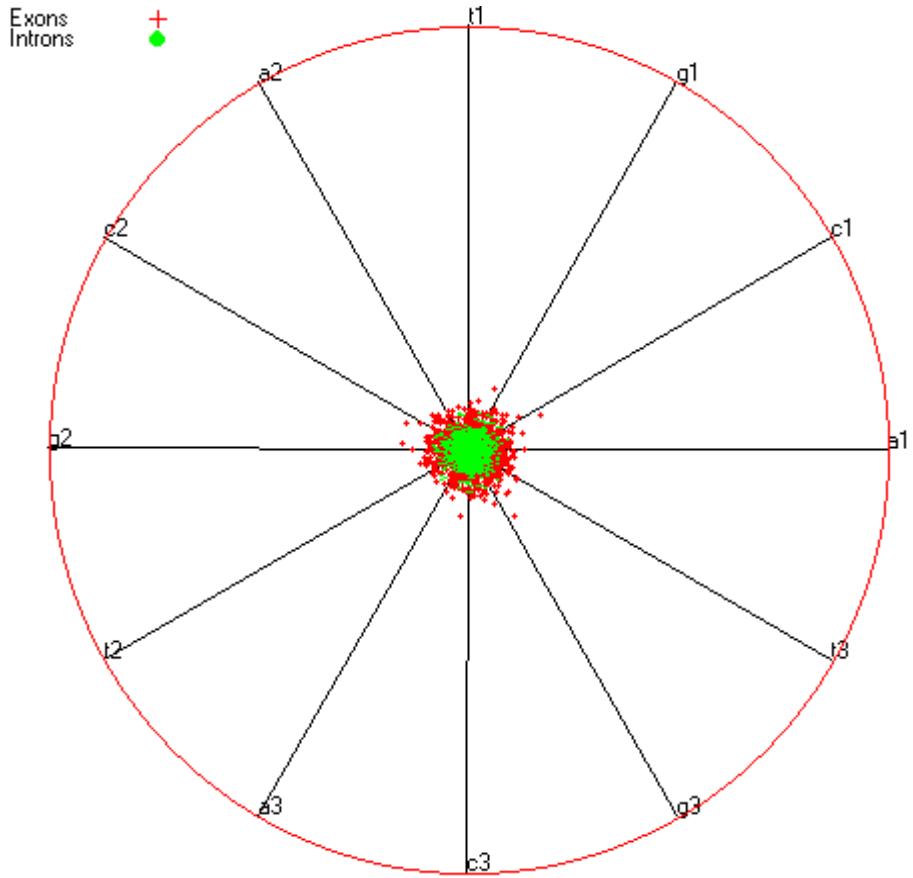


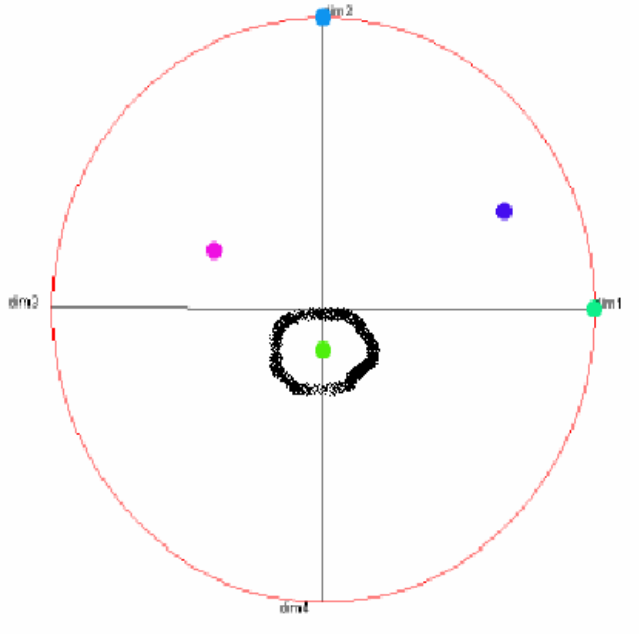
Figure 15 DAs in a regular polygon P = (0, 0, 0, 0, 1.0, 0, 0, 0, 0) Mesh Plot

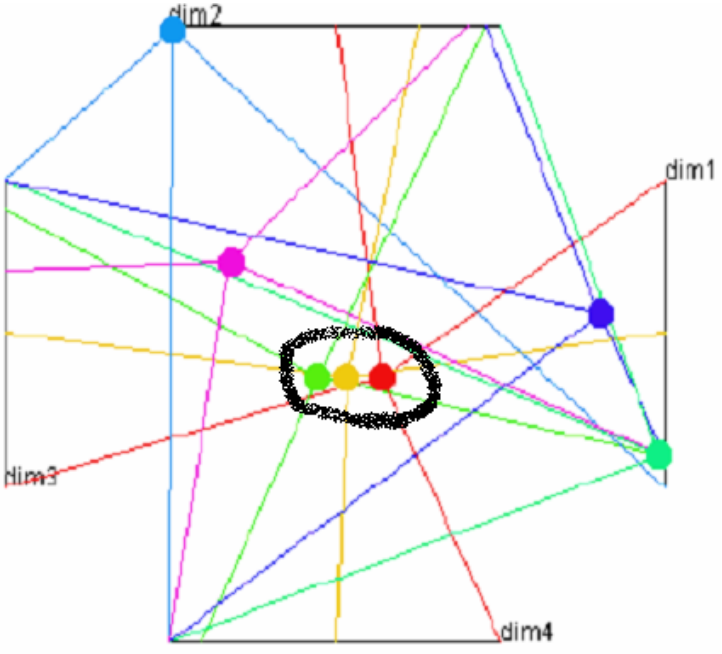
Intro. to RadViz Spring Force

- a radial visualization
- One spring for each dimension.
- One end attached to perimeter point. The other end attached to a data point.
- Each data point is displayed where the sum of the spring forces equals 0.



DAs RadViz





Original Radviz – 3 overlapping points

Limitation: data points with different values can overlap

DAs spread polygon
P = (0, 0, 0, 0, 0, 0, .5, 1.0, .5)

DA layout

- ✓ Parameters – Done !
- ☐ Layout
  - DAs can be arranged with any arbitrary size, shape or position
  - Permits a large variety of visualization designs

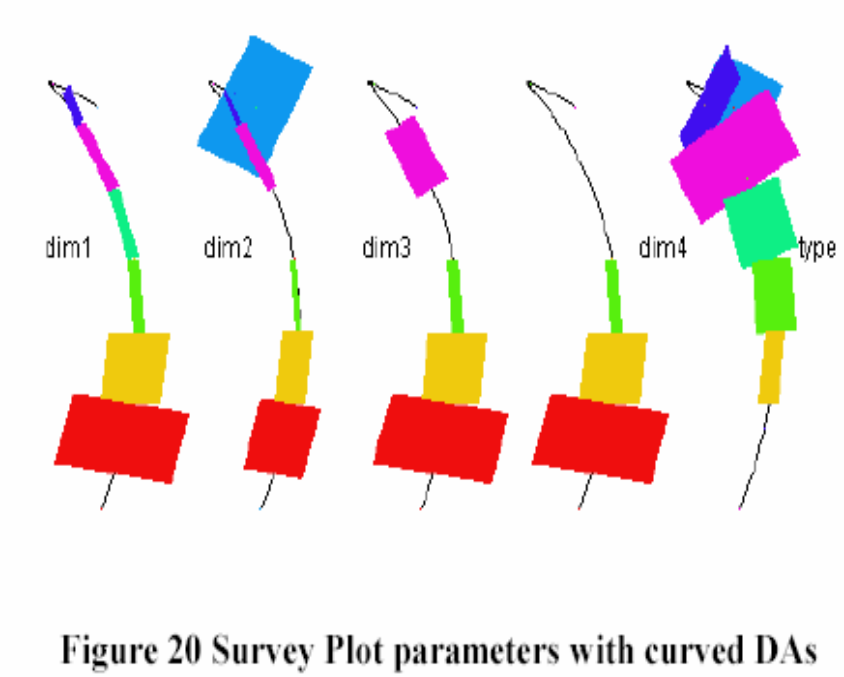
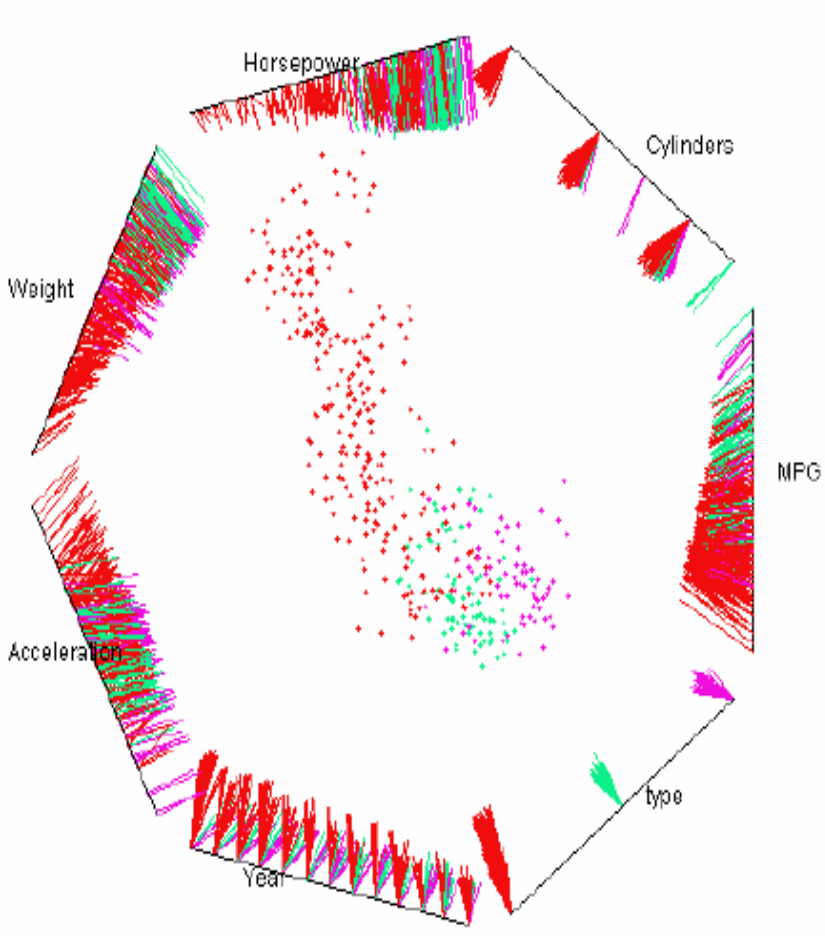


Figure 20 Survey Plot parameters with curved DAs

Combinations of Visualizations

- Can we combine features of two (or more) visualizations?
  - Combination of Parallel Coordinates and Radviz



Visualization Space

- Nine parameters define the size of our visualization space as R^9
- Include the geometry of the DAs, assuming 3 parameters are used to define the geometry
- The size of our visualization space is R^{12}
- "Grand Tour" through visualization space is possible
- New visualizations can be created during a tour

Evaluation

Strong Points

- ☺ Idea
- ☺ Many examples of visualizations with real data

Weak Points

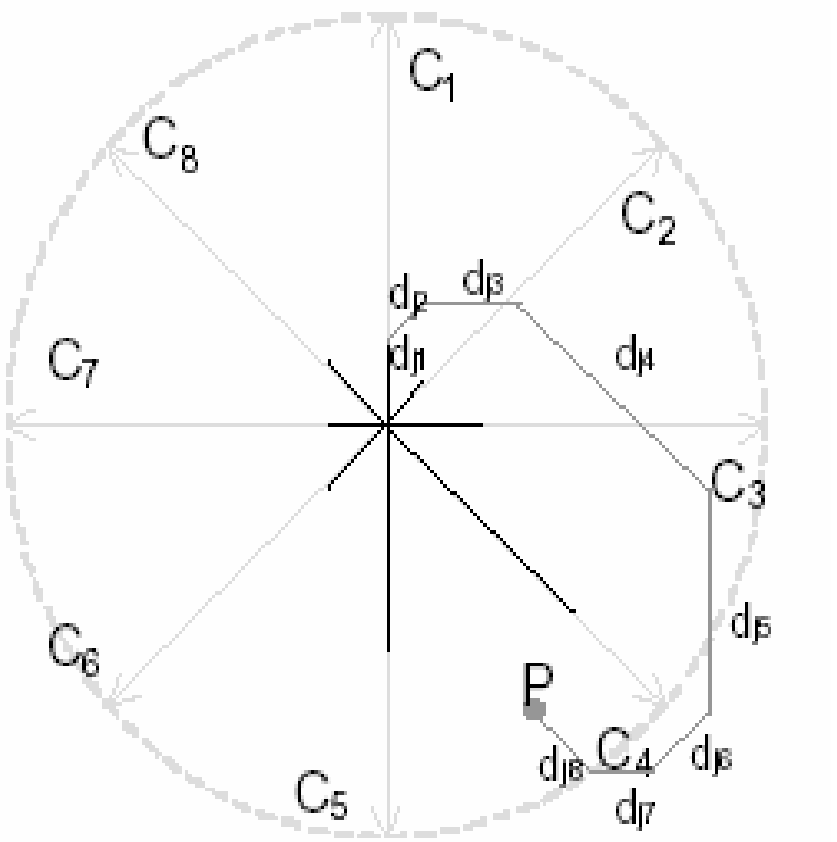
- ☹ Not accessible
- ☹ Short explanation of examples
- ☹ Lack of examples for some statement
- ☹ No implementation details

Where are we

- ✓ Dimensional Anchors
- ➢ Star Coordinates
  - a new interactive multidimensional technique
  - helpful in visualizing multi-dimensional clusters, trends, and outliers
- ❑ StarClass – Interactive Visual Classification Using Star Coordinates

Star Coordinates

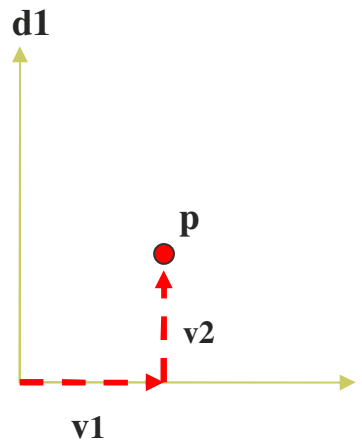
- Each dimension shown as an axis
- Data value in each dimension is represented as a vector.
- Data points are scaled to the length of the axis
  - min mapping to origin
  - max mapping to the end



Star Coordinates Contd

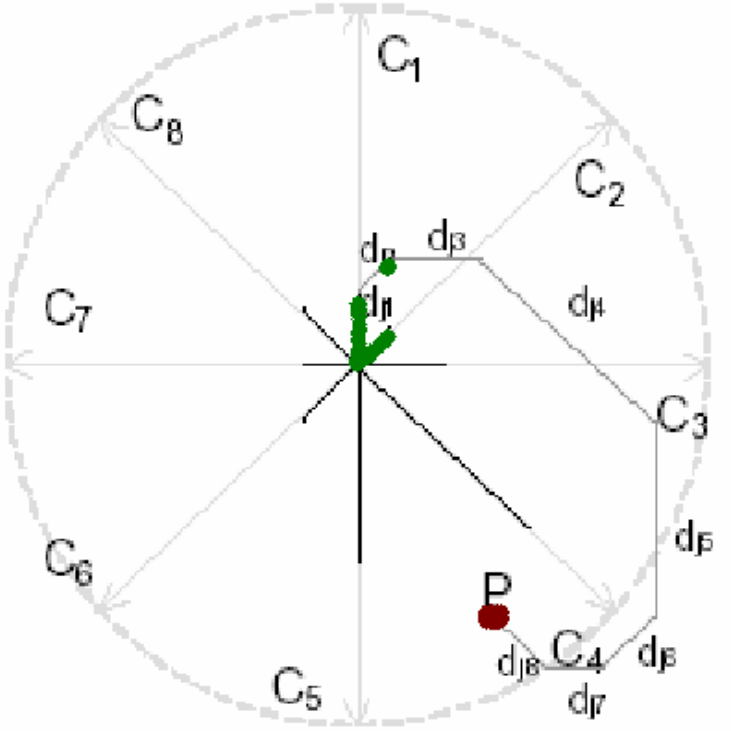
Cartesian

P=(v1, v2)



Star Coordinates

P=(v1,v2,v3,v4,v5,v6,v7,v8)



Mapping:

- Items → dots
- Σ attribute vectors → position

Interaction Features

- Scaling
  - allows user to change the length of an axis
  - increases or decrease the contribution of a data column
- Rotation
  - changes the direction of the unit vector of an axis
  - makes a particular data column more or less correlated with the other columns
- Marking
  - selects individual points or all points within a rectangular area and paints them in color
  - makes points easy to follow in the subsequent transformations

Interaction Features

- Range Selection

  - select value ranges on one or more axes, mark and paint them

  - allows users to understand the distribution of particular data value ranges in current layout

- Histogram

  - provides data distribution for each dimension

- Footprints

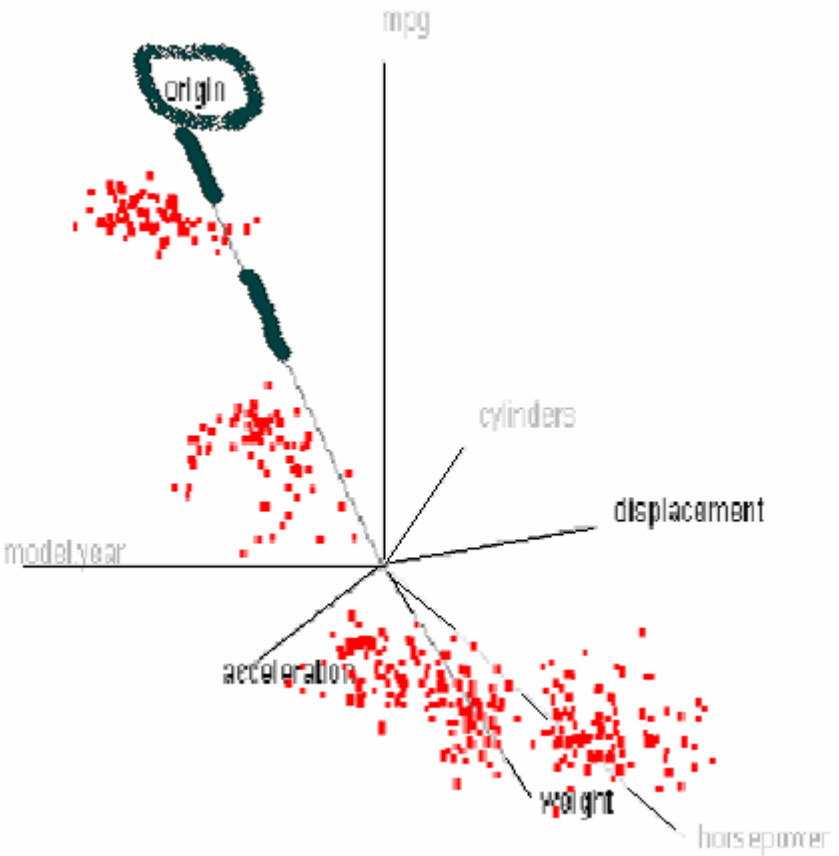
  - leave marks of data points on the trail for recent transformations

Applications – Cluster Analysis

- Playing with the "cars" dataset

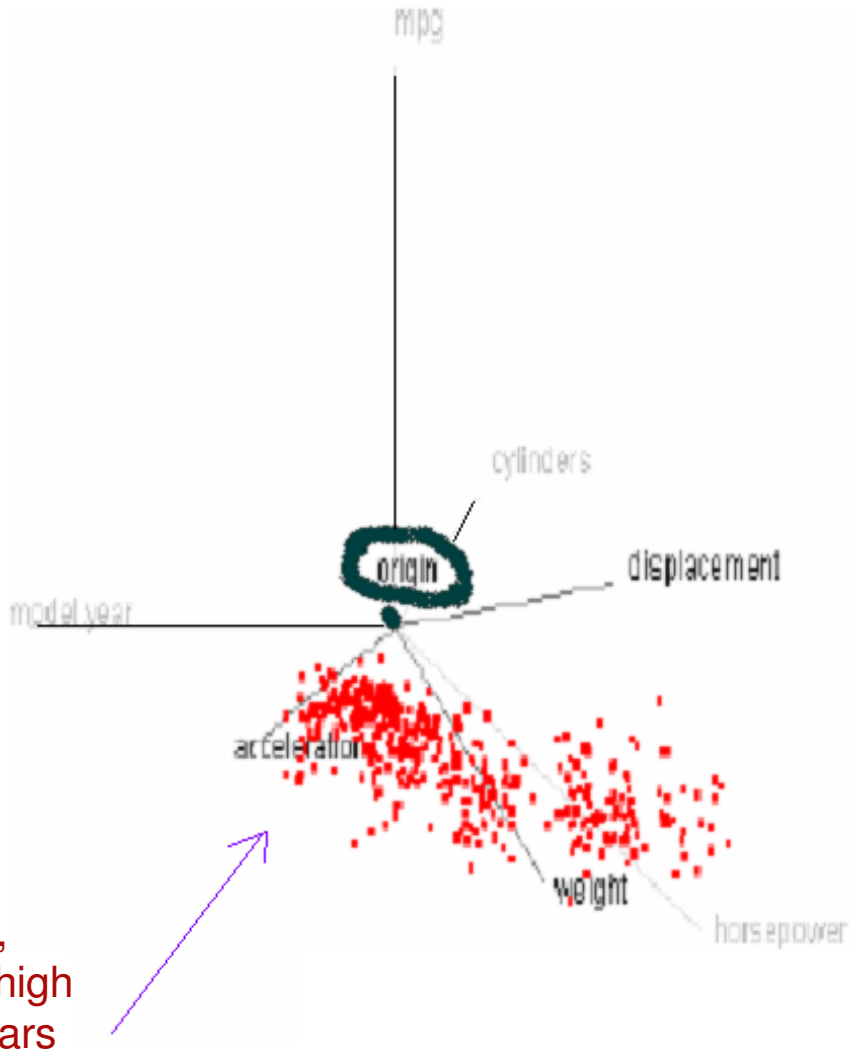
  - scaling, rotating, & turning off some coordinates

- Four major clusters in the data discovered



Applications – Cluster Analysis

- Scaling the "origin" coordinate moves only the top two clusters
  - (JP & Euro)
- Down-scaling the origin
  - these two clusters join one of the other clusters(American-made cars of similar specs)
- Result: two clusters



Low weight, displacement, high acceleration cars

SC – useful in visualizing clusters

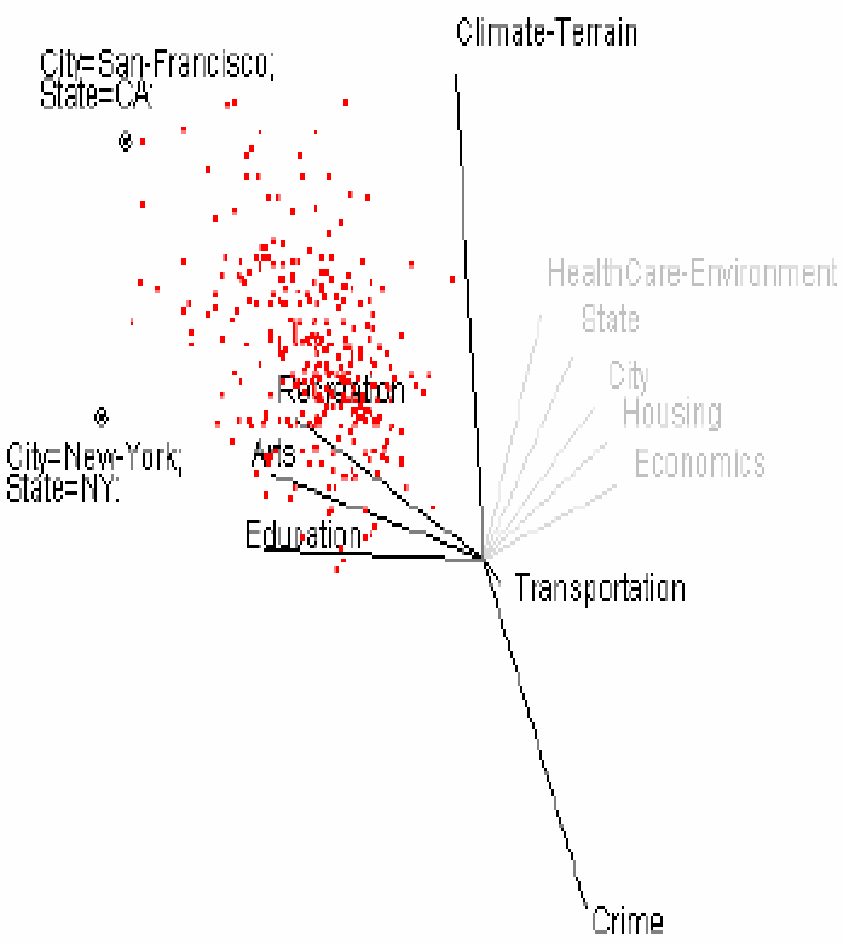
- Within few minutes users can identify how the data is clustered
- Gain an understanding of the basic characteristics of these clusters

Multi-factor Analysis

- Dataset – "Places"

  – ratings wrt climate, transportation, housing, education, arts, recreation, crime, health-care, and economics

- Important desirable factors pulled together in one direction and neg. undesirable factors in the opposite

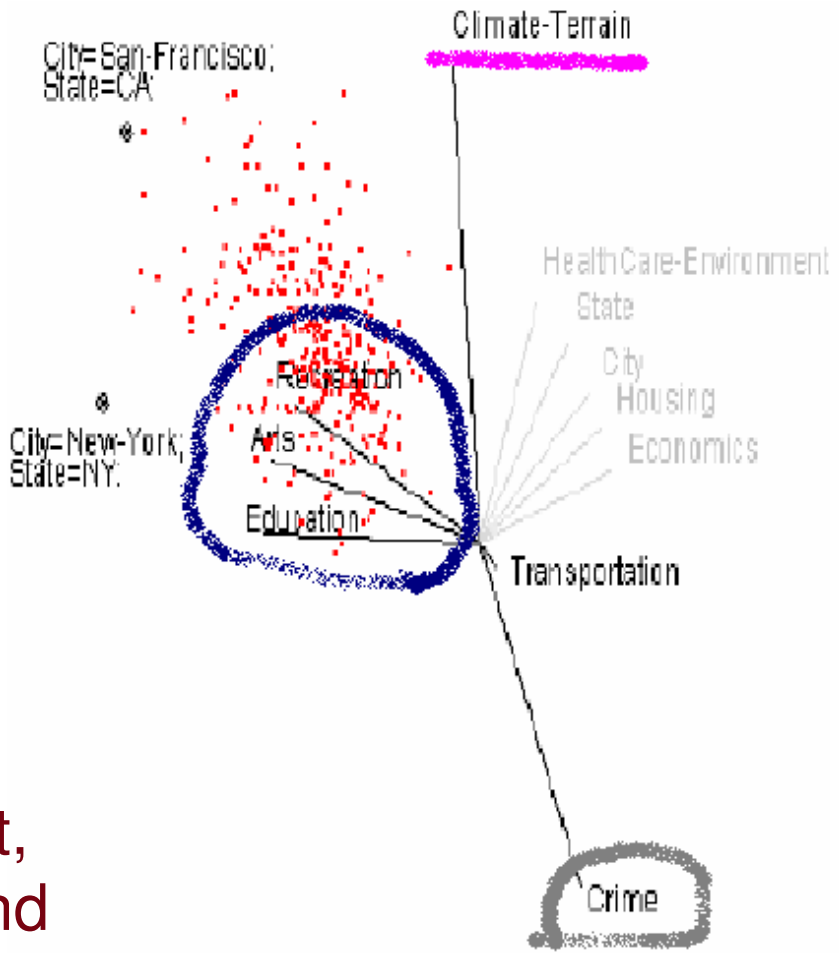


Mutli-factor Analysis con't

- Desirable factors:
  - recreation, art, & education
  - climate (most)
- Undesirable factor:
  - crime

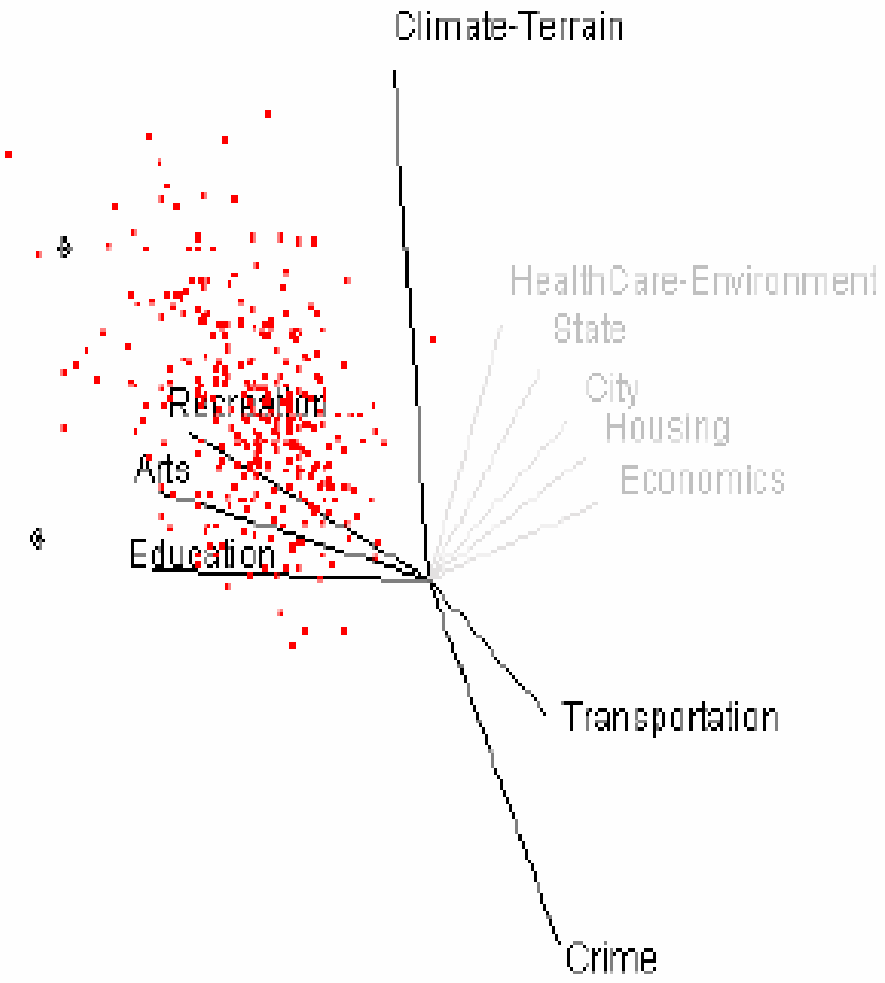
What can you conclude about NY and SF?

- NY – outlier
- SF – comparable arts, ect, but better climate and lower crime



Multi-factor Analysis contd

- Scale up transportation
  - other cities beat SF in the combined measure



Evaluation of SC in Multi-factor Analysis

- ☹ Exact individual contributions of these factors are not immediately clear
- ☺ The visualization provides users with an overview of how a number of factors affect the overall decision making

Evaluation

Strong Points

☺ idea

☺ many concrete examples with full explanations

Weak points

☹ ugly figures (undistinguishable)

Where we are

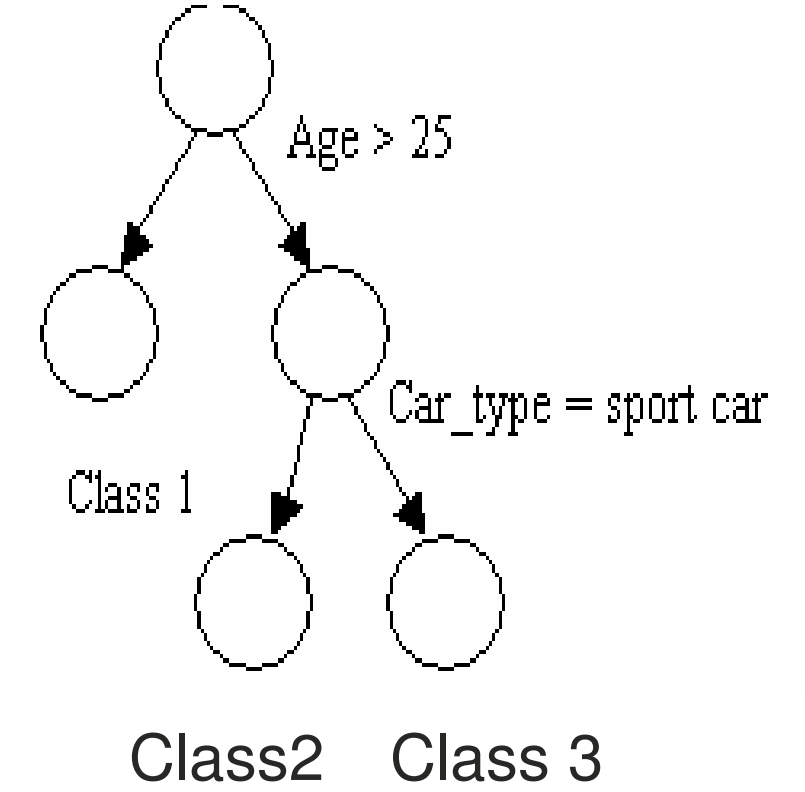
- ✓ Dimensional Anchors
- ✓ Star Coordinates
  - a new interactive multi-D visualization tech.
- ➢ StarClass – Interactive Visual Classification Using Star Coordinates

Classification

- Each object in a dataset belongs to exactly one class among a set of classes.
- Training set data : labeled (class known)
- Build model based on training set
- Classification: use the model to assign a class to each object in the testing set.

Classification Method

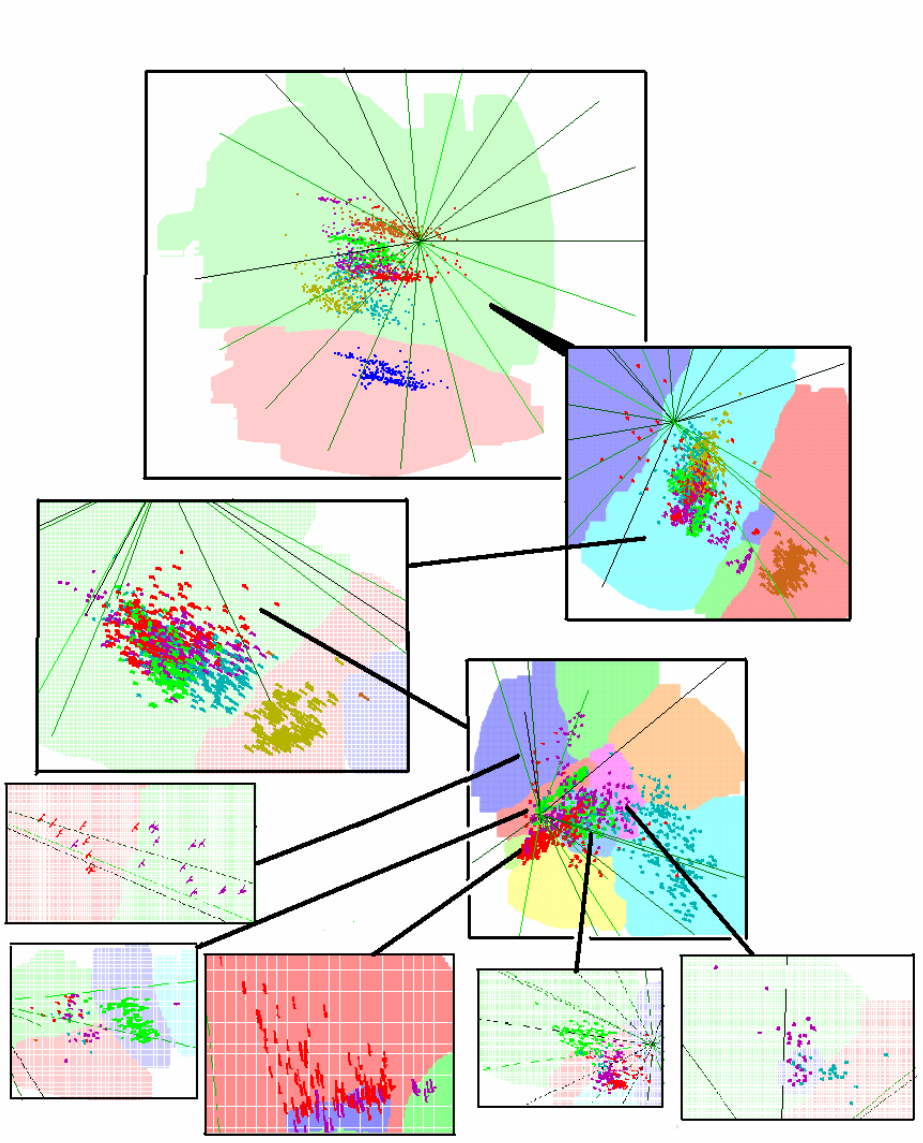
- Decision trees



Visual-base DT Construction

- Visual Classification
  - projecting
  - painting
  - region can be re-projected
  - recursively define a decision tree.
  - each project correspond to a node in decision tree
  - Majority class at leaf node determines class assignment

  (the class with the most number of objects mapping to a terminal region is the "expected class")



Evaluation of the system

Good

- ☺ Makes use of human judgment and guides the classification process
- ☺ Good accuracy
- ☺ Increase in user's understanding of the data

Bad

- ☹ expertise required?

Evaluation of the Paper

Good

☺ Ideas

☺ Accessible

☺ Concrete examples

Bad

☹ No implementation discussed

Summary

- Dimensional Anchor
  - unify visualization techniques
- Star Coordinate
  - new interactive visualization techniques
  - Visualizing clusters and outliers
- StarClass
  - interactive classification using star coordinate

Reference

- Dimensional Anchors: a Graphic Primitive for Multidimensional Multivariate Information Visualizations, P. Hoffman, G. Grinstein, & D. Prinkney, Proc. Workshop on New Paradigms in Information Visualization and Manipulation, Nov. 1999, pp. 9-16.
- Visualizing Multi-dimensional Clusters, Trends, and Outliers using Star Coordinates, Eser Kandogan, Proc. KDD 2001
- StarClass: Interactive Visual Classification Using Star Coordinates , S. Teoh & K. Ma, Proc. SIAM 2003
- http://graphics.cs.ucdavis.edu/~steoh/research/classification/SDM03.ppt