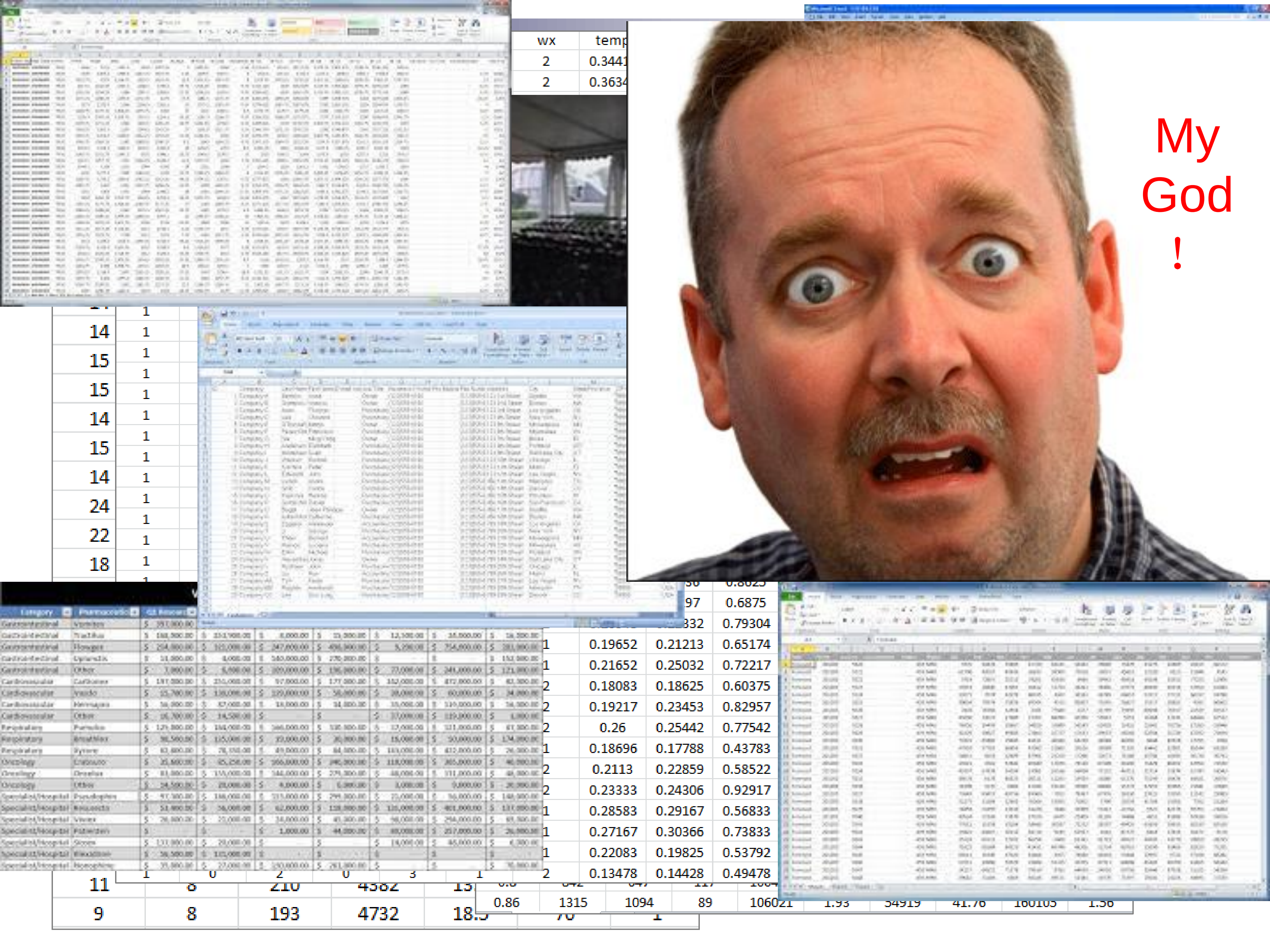
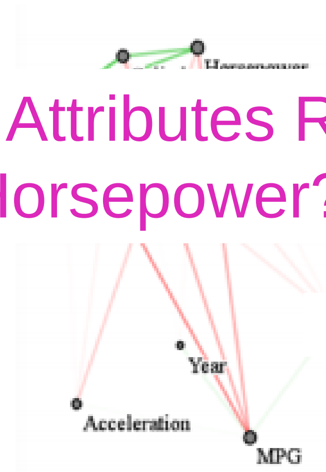
Improving the Fidelity of Contextual Data Layouts Using a Generalized Barycentric Coordinates Framework

Shenghui Cheng and Klaus Mueller

Visual Analytics and Imaging Lab,
Computer Science Department,
Stony Brook University and SUNY Korea


My
God
!

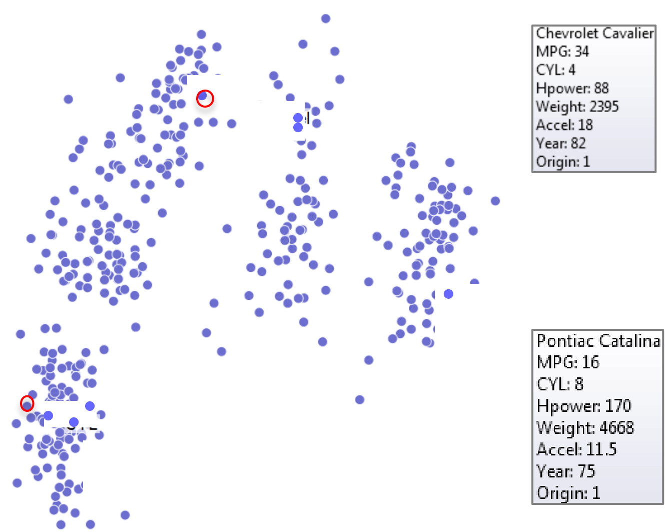
- Attributes (Variables) to Attributes Relation (e.g Horsepower and weight? Positive or negative?)



How to get the context, Data to Attributes Relation ? (which cars have high Horsepower?)

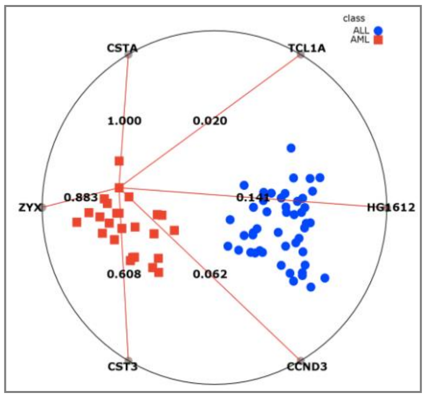
MPG	CYL	Hpower	Weight	Accel	Year	Origin
18	8	130	3504	12	70	1
15	8	165	3693	11.5	70	1
18						
16						
17						
15						
14						
14	8	215	4312	8.5	70	1
14	8	225	4425	10	70	1
15	8	190	3850	8.5	70	1
15	8	170	3563	10	70	1
14	8	160	3609	8	70	1
15	8	150	3761	9.5	70	1
14	8	225	3086	10	70	1
24	4	95	2372	15	70	3
22	6	95	2833	15.5	70	1
18	6	97	2774	15.5	70	1
21	6	85	2587	16	70	1
27	4	88	2130	14.5	70	3
26	4	46	1835	20.5	70	2
25	4	87	2672	17.5	70	2
24	4	90	2430	14.5	70	2
25	4	95	2375	17.5	70	2
26	4	113	2234	12.5	70	2
21	6	90	2648	15	70	1
10	8	215	4615	14	70	1
10	8	200	4376	15	70	1
11	8	210	4382	13.5	70	1
9	8	193	4732	18.5	70	1

- Data to Data Relation (which cars are similar? Cluster? Outliers?)

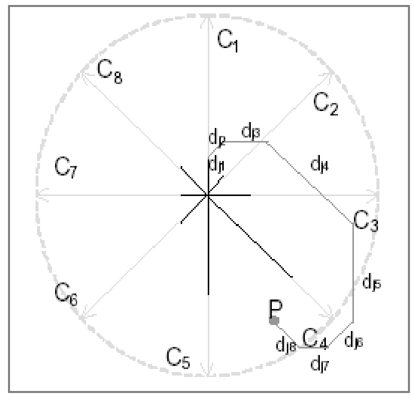


Contextual layouts

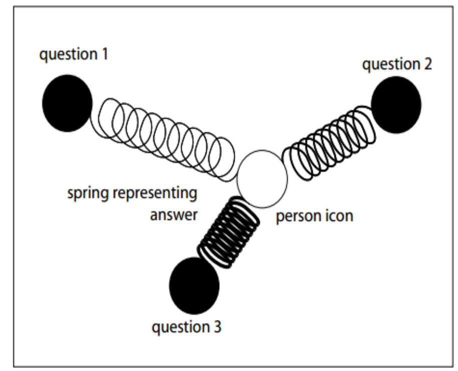
Radviz



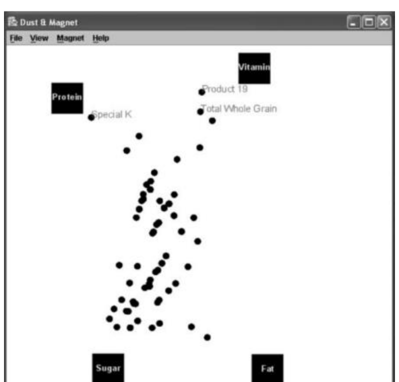
Star Coordinates



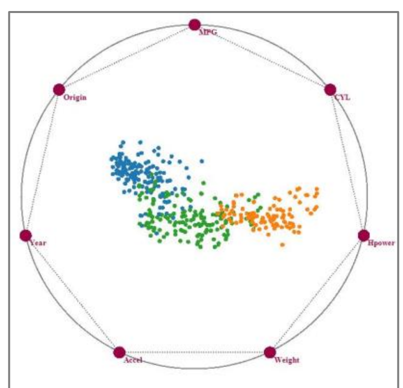
Gravi++



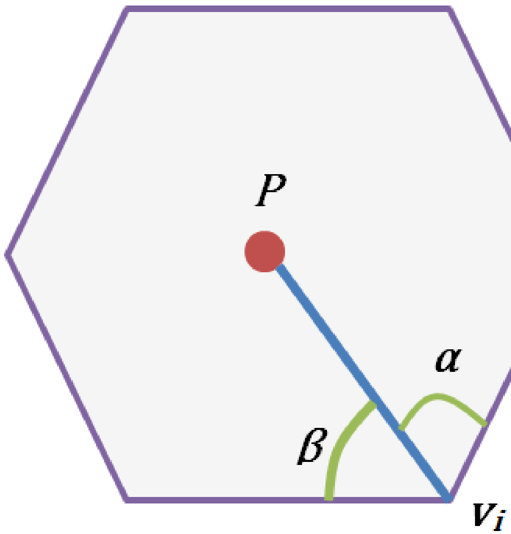
Dust & Magnet



Generalized Barycentric Coordinates



Generalized Barycentric Coordinates (GBC)



GBC interpolation

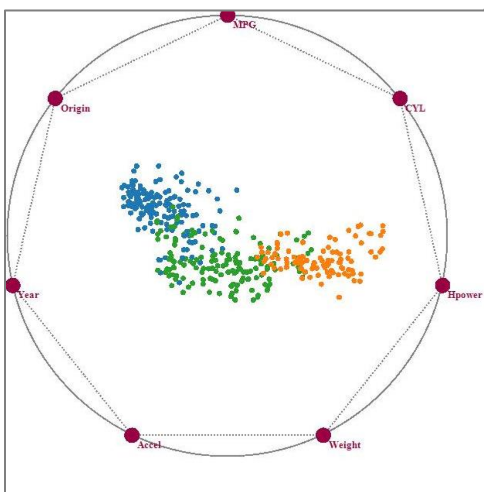
the interpolation weight w_i of vertex v_i for P is

$$w_i = \frac{\cot(\alpha)+\cot(\beta)}{\|P-v_i\|^2}$$

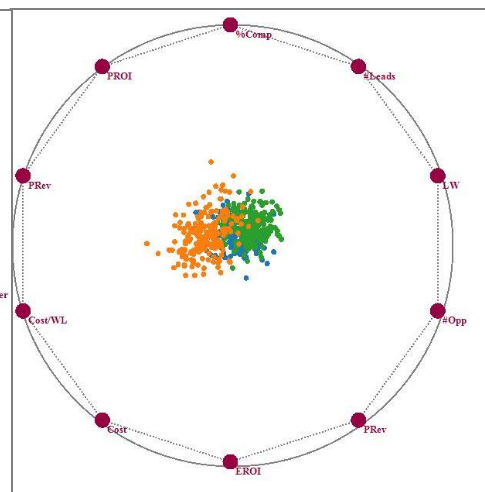
The interpolated value Pv at P is

$$Pv = \sum_{i=1}^{n} w_i v_i \text{ where } a_i = w_i / \sum_{k=1}^{n} w_k \text{ and } \sum_{i=1}^{n} a_i = 1$$

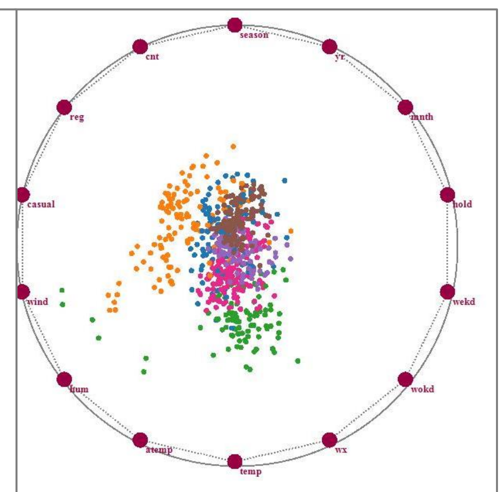
Car



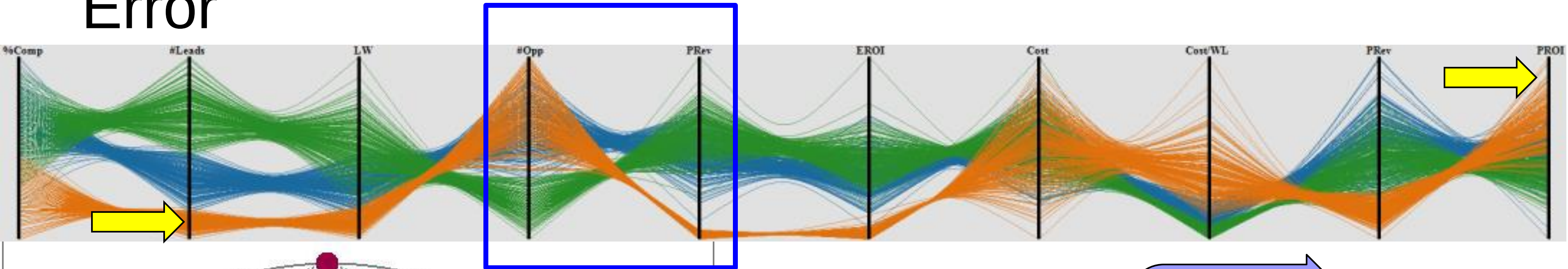
Sales Campaign

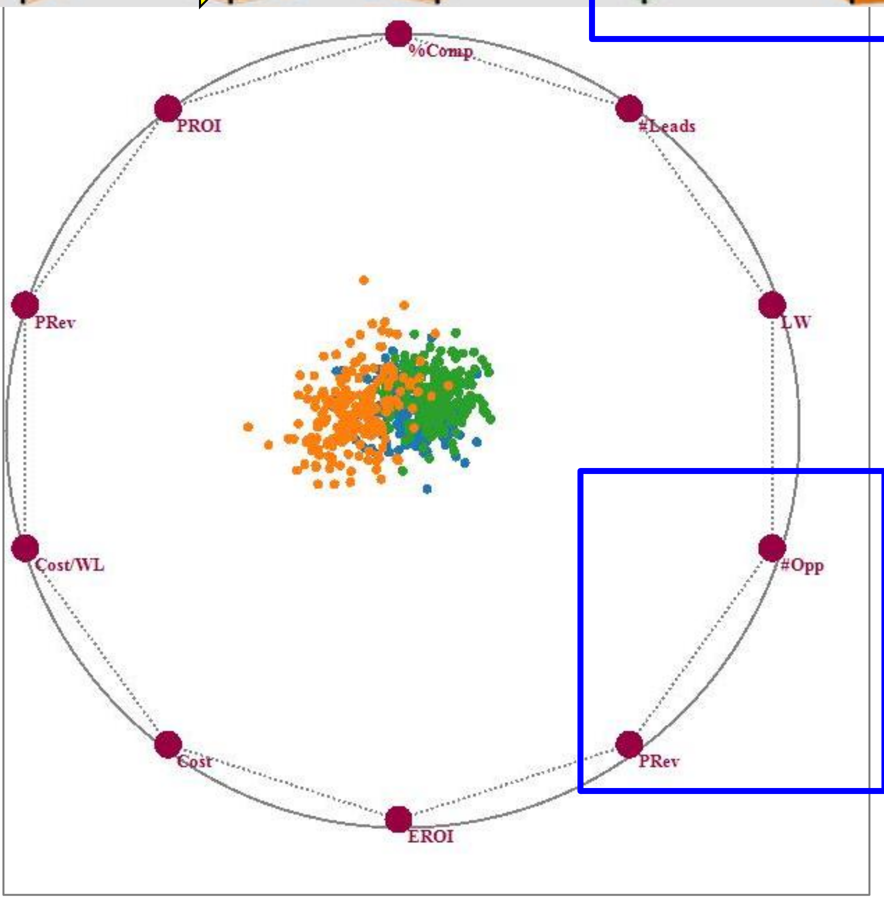


Bike



Error





Mapped

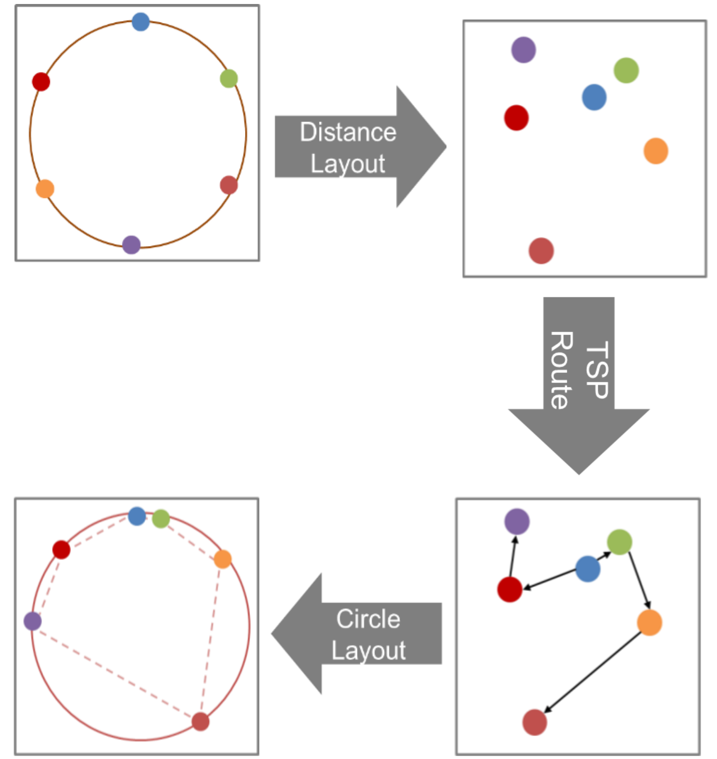
$$stress(L, C) = \sqrt{\frac{\sum_{ij}(L_{ij} - C_{ij})^2}{\sum_{ij} C_{ij}^2}}$$

Original

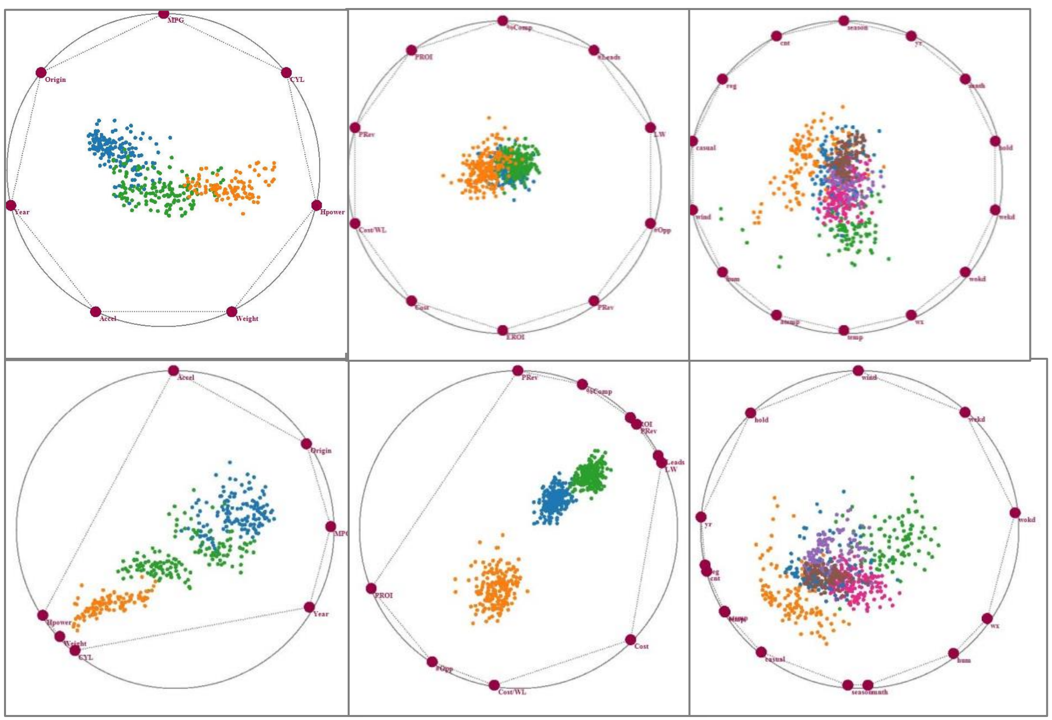
- Variables to Variables Error
- Data to Variables Error
- Data to Data Error

Variables to Variables Error - distance spaced layout

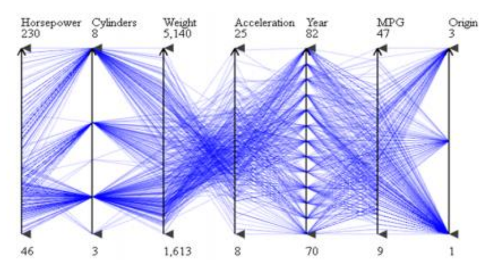
- Linear ordering of the vertices
  - Correlation matrix
  - Approximate Traveling Salesman Problem (TSP) for ordering
- Circle layout
  - Arrange on the circle, spaced with correlation



Car Sales Campaign Bike

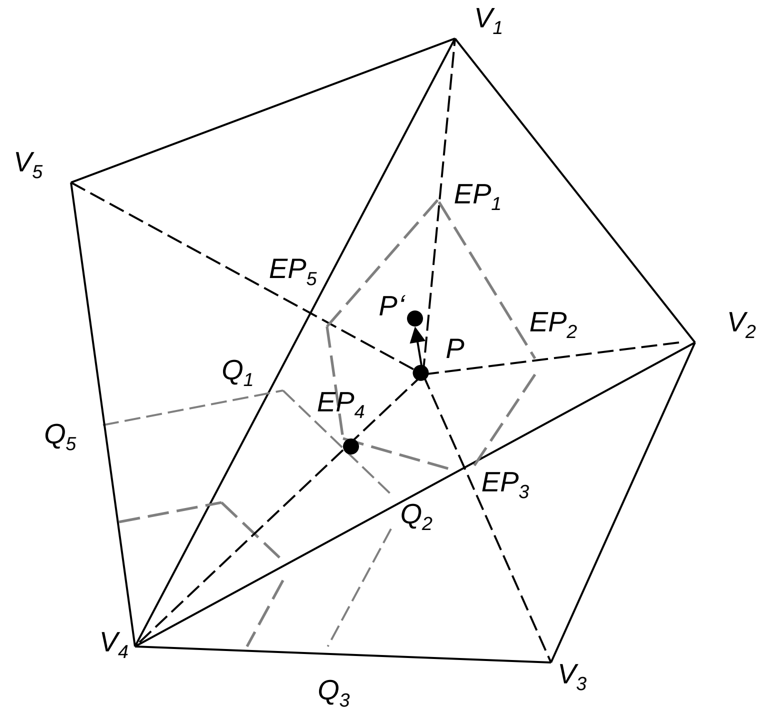


49% 85% 44%

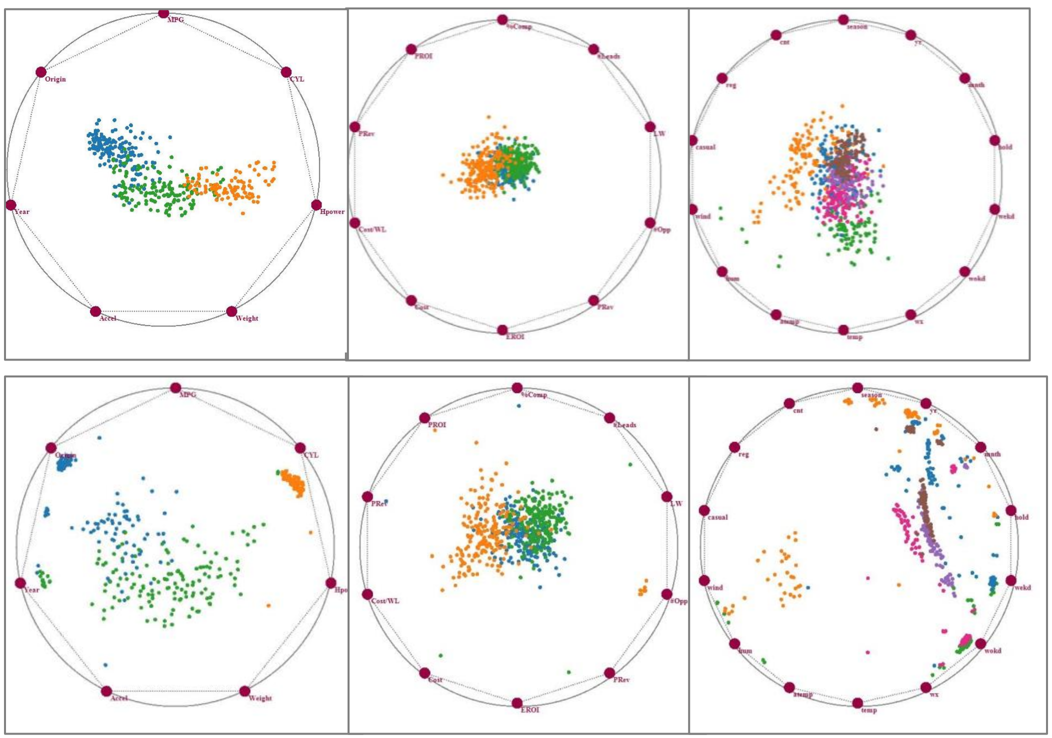


Data to Variables Error - iterative error reduction

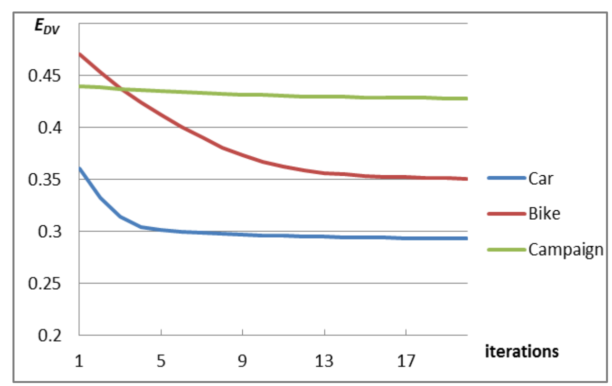
- Construct the iso-contours.
- Compute the error of each dimension
- Construct the error polygon
- Move to the center of polygon iteratively



Car Sales Campaign Bike

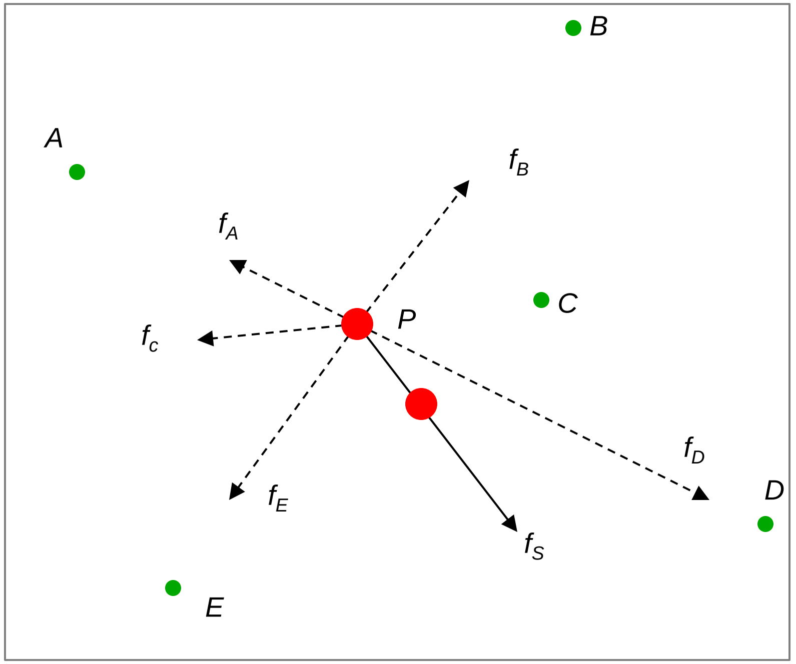


25% 2% 29%

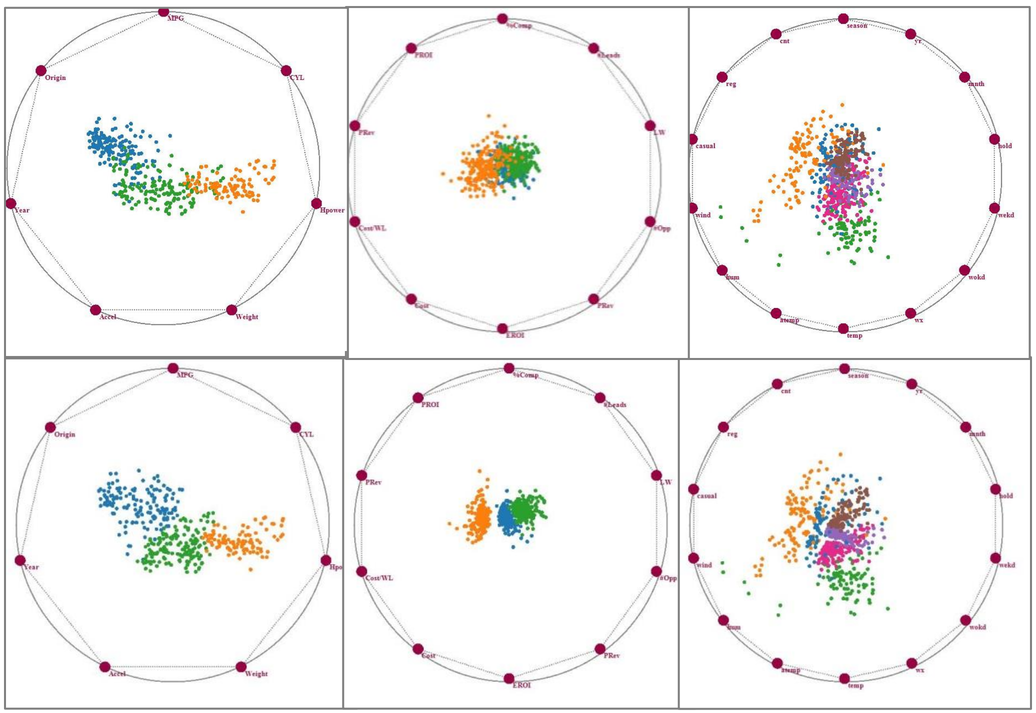


Data to Data Error - force directed adjustment

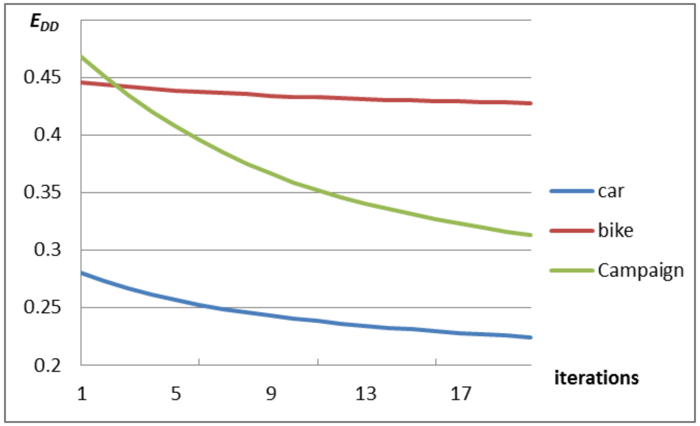
- Construct the Network
  vertices: data points
  edges: springs
  force: error *real distance – mapped distance*
- Drag or push the points in turn.



Car Sales Campaign Bike



28% 62% 5%



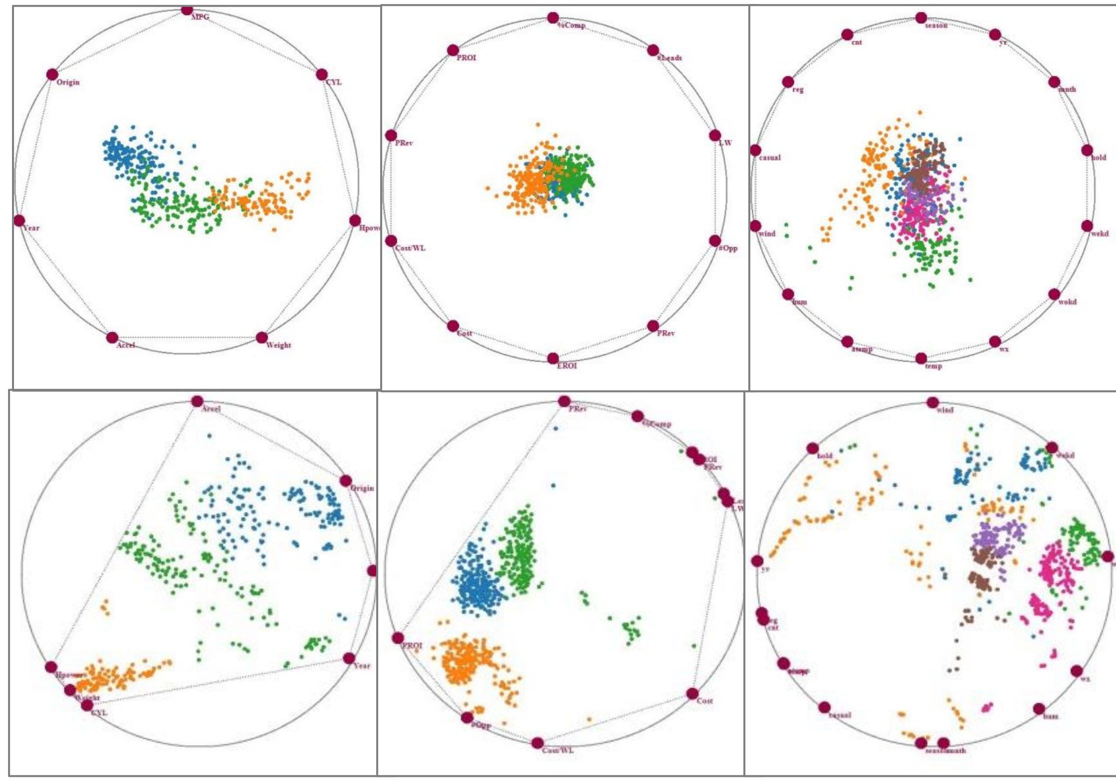
Finally – Combine together

Variables to Variables:
Distance spaced layout

Data to Variables:
Iterative error reduction

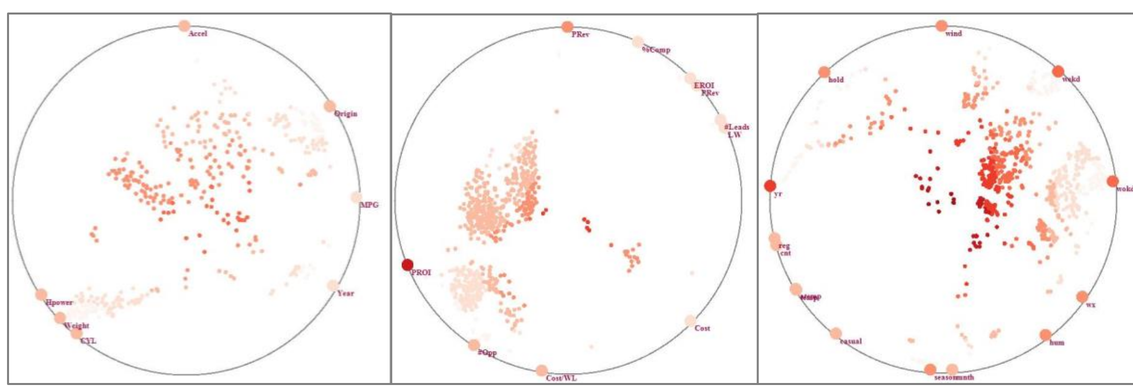
Data to Data:
Force directed adjustment

Car | Sales Campaign | Bike



44% | 78% | 41%

Error distribution

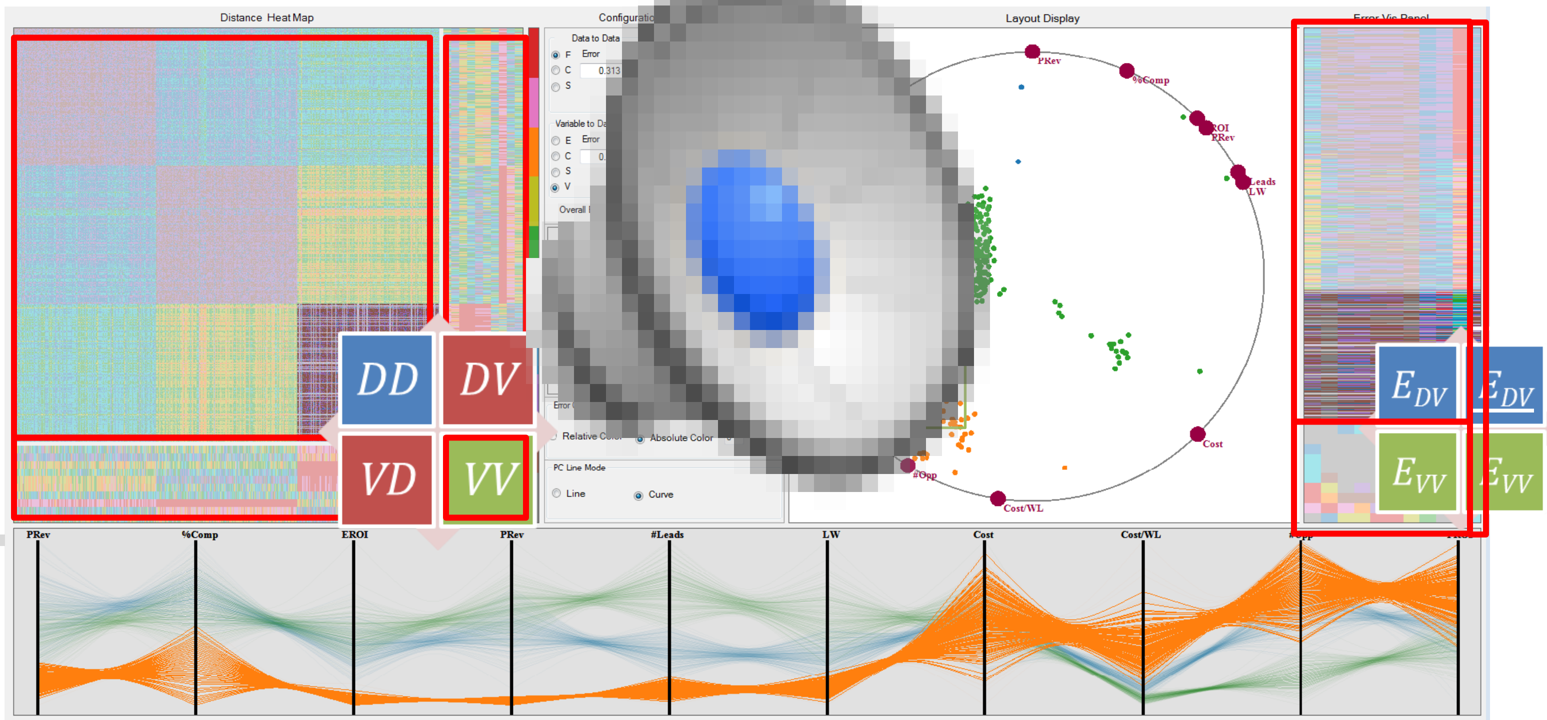


More Error ? – Data Overlap

[0.1, 0.2, 0.3],   [0.2, 0.4, 0.6] – same location ??

GBC Error Explorer - combining different visualization methods into a interface

Distance heatmap | Configuration Control Panel | Layout Display | Error Vis Panel



Parallel Coordinates Display

Conclusion and Future Work

Conclusion

- We unified the different contextual layouts.
- We proposed three algorithms – distance spaced layout, iterative error reduction and force directed adjustment – to reduce the error.
- We developed an interface by which users can explore the error by combining the different visualization schemes with interactions.

Future Work

- Attributes (variables) are arranged at the periphery of the data points. Better optimizations might be achievable by allowing the attribute points to mingle with the data points.

Reference

J. Chambers, W. Cleveland, P. Tukey, *Graphical Methods for Data Analysis,* Duxbury Press, 1983.

J. Hartigan, “Printer Graphics for Clustering,” *J. Statistical Computation and Simulation,* 4(3): 187-213, 1975.

L.Maaten, G.Hinton, “Visualizing data using t-SNE”, *J. Machine Learning Research*, 9: 2579-2605, 2008

J. Nam, K. Mueller, "TripAdvisorN-D: A Tourism-Inspired High-Dimensional Space Exploration Framework with Overview and Detail," *IEEE Trans. Visualization and Computer Graphics*, 19(2):291-305, 2013.

M. Meyer, A. Barr, H. Lee, M. Desbrun, “Generalized Barycentric Coordinates on Irregular Polygons,” *J. Graphics Tools*, 7(1):13-22, 2002.

P. Hoffman, G. Grinstein, K. Marx, I. Grosse, E. Stanley, "DNA Visual and Analytic Data Mining", *Proc. IEEE Vis*, pp. 437-441, 1997.

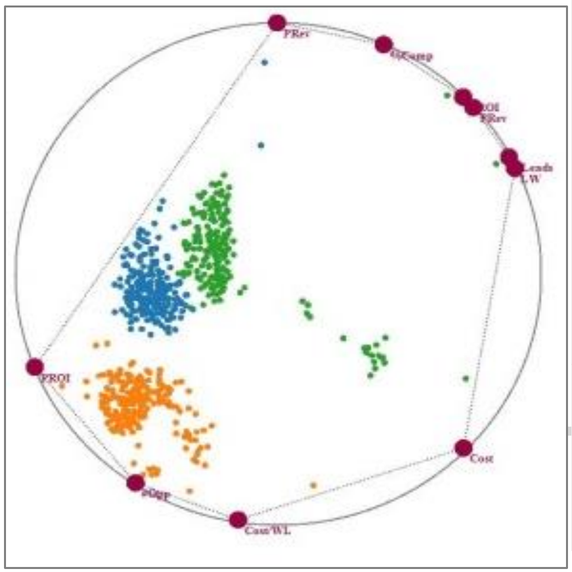
K. Hinum, S. Miksch, W. Aigner, S. Ohmann, C. Popow, M. Pohl, M. Rester, “Gravi++: Interactive Information Visualization to Explore Highly Structured Temporal Data,” *J. Universal Computer Science* 11(11):1792-1805, 2005.

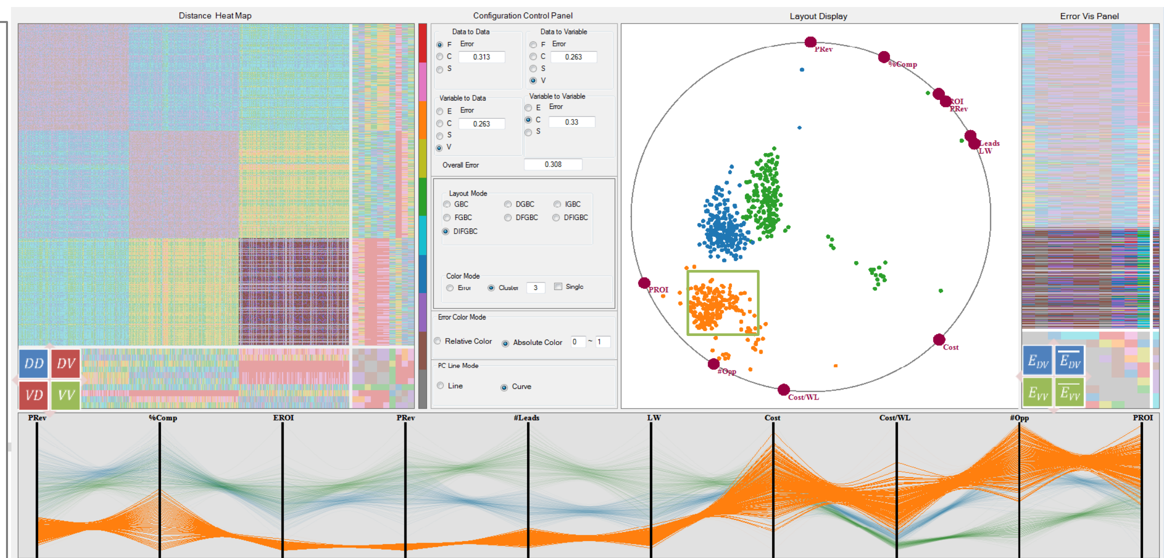
J. Yi, R. Melton, J. Stasko, J. Jacko, "Dust & Magnet: Multivariate Information Visualization using a Magnet Metaphor," *Information Visualization,* 4(4) : 239-256, 2005.

Z. Zhang, K. McDonnell, K. Mueller, "A Network-Based Interface for the Exploration of High-Dimensional Data Spaces," *Proc. IEEE Pacific Vis*, Songdo, Korea, pp. 17-24, 2012.

Questions ?

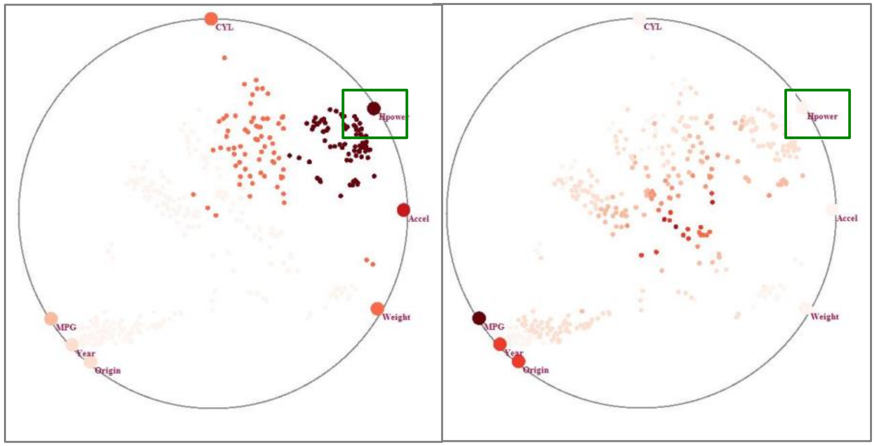
This research was partially supported by NSF grant IIS 1117132 and the MSIP, Korea, under the "IT Consilience Creative Program (ITCCP)" (NIPA-2013-H0203-13-1001) supervised by NIPA.

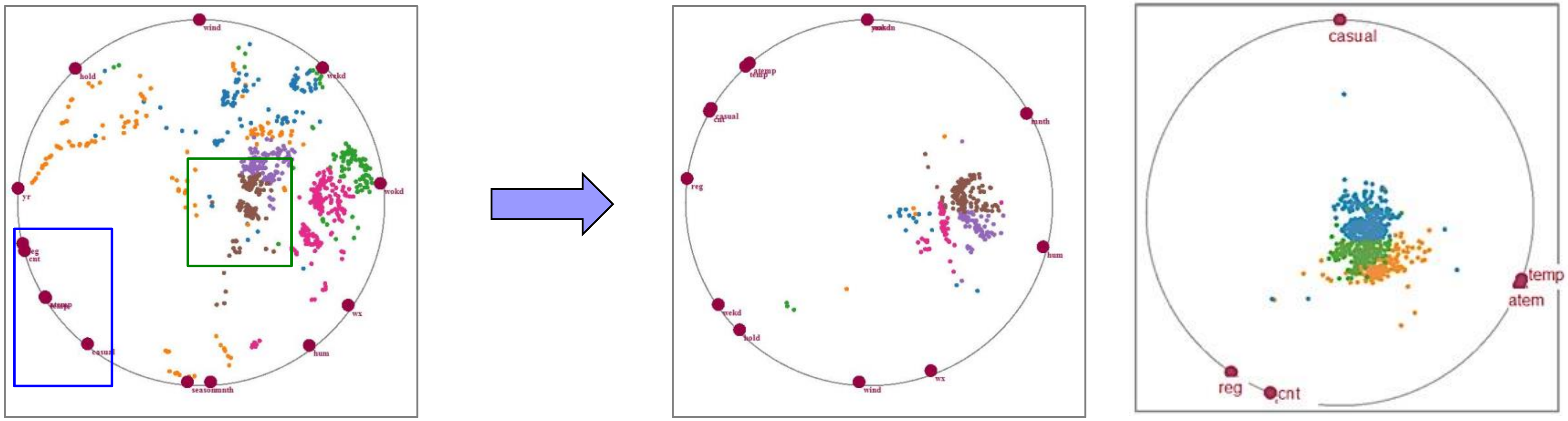




Interactions

- Verification coloring:
  Distance color and Error color
- Linked displays
- Local layout refinement
  Data-centric refinement and
  Variable-centric refinement

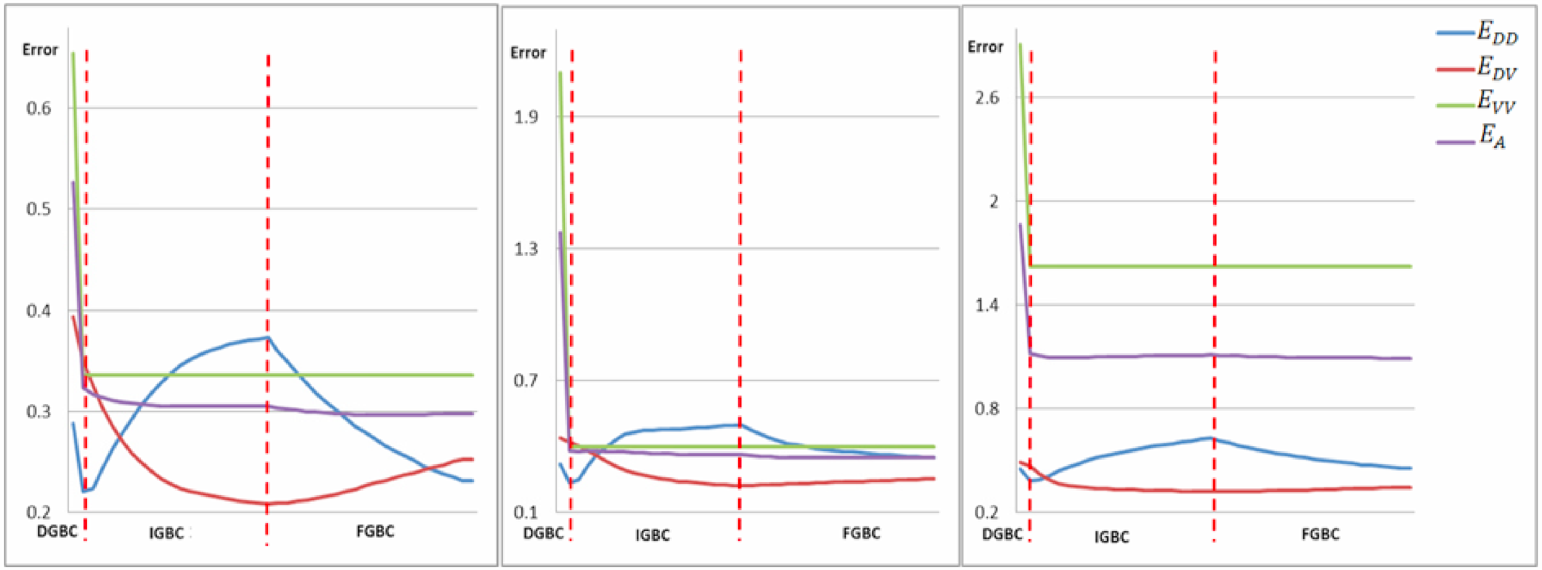




Unified Definition

Table 1. The features of different layout methods

Method	VF	MF (p_i)
Radviz	$v_i = \left(r \cdot \cos\frac{i}{2\pi}, r \cdot \sin\frac{i}{2\pi}\right)$	$\sum_{j=1}^{n} \frac{x_{ij}}{\sum_{k=1}^{n} x_{ik}} \cdot v_j$
Star Coordinates	$v_i = \left(r \cdot \cos\frac{\theta_i}{2\pi}, r \cdot \sin\frac{\theta_i}{2\pi}\right)$ Or other	$\sum_{j=1}^{n} x_{ij} v_j$
Gravi++	$v_i = \left(r \cdot \cos\frac{\theta_i}{2\pi}, r \cdot \sin\frac{\theta_i}{2\pi}\right)$ Or other free layout	$\sum_{j=1}^{n} \frac{s_j x_{ij}}{\sum_{k=1}^{n} s_k x_{ik}} \cdot v_j$
Dust & Magnet	$v_i = \left(r \cdot \cos\frac{\theta_i}{2\pi}, r \cdot \sin\frac{\theta_i}{2\pi}\right)$ Or other free layout	$\sum_{j=1}^{n} a_{ij} x_{ij} \cdot v_j$
GBC	$v_i = \left(r \cdot \cos\frac{\theta_i}{2\pi}, r \cdot \sin\frac{\theta_i}{2\pi}\right)$ Or other convex polygon	$\sum_{j=1}^{n} \frac{x_{ij}}{\sum_{k=1}^{n} x_{ik}} \cdot v_j$
Remarks	$\theta_1 + \sum_{i=2}^{n}(\theta_j - \theta_{j-1}) = 2\pi$. s_j stands for the strength multiplicator of v_j. a_{ij} is the attraction between dust *i* and magnet j. r is the circle radius.	



Error
0.6
0.5
0.4
0.3
0.2
DGBC
IGBC
FGBC
Error
1.9
1.3
0.7
0.1
DGBC
IGBC
FGBC
Error
2.6
2
1.4
0.8
0.2
DGBC
IGBC
FGBC
E_{DD}
E_{DV}
E_{VV}
E_A