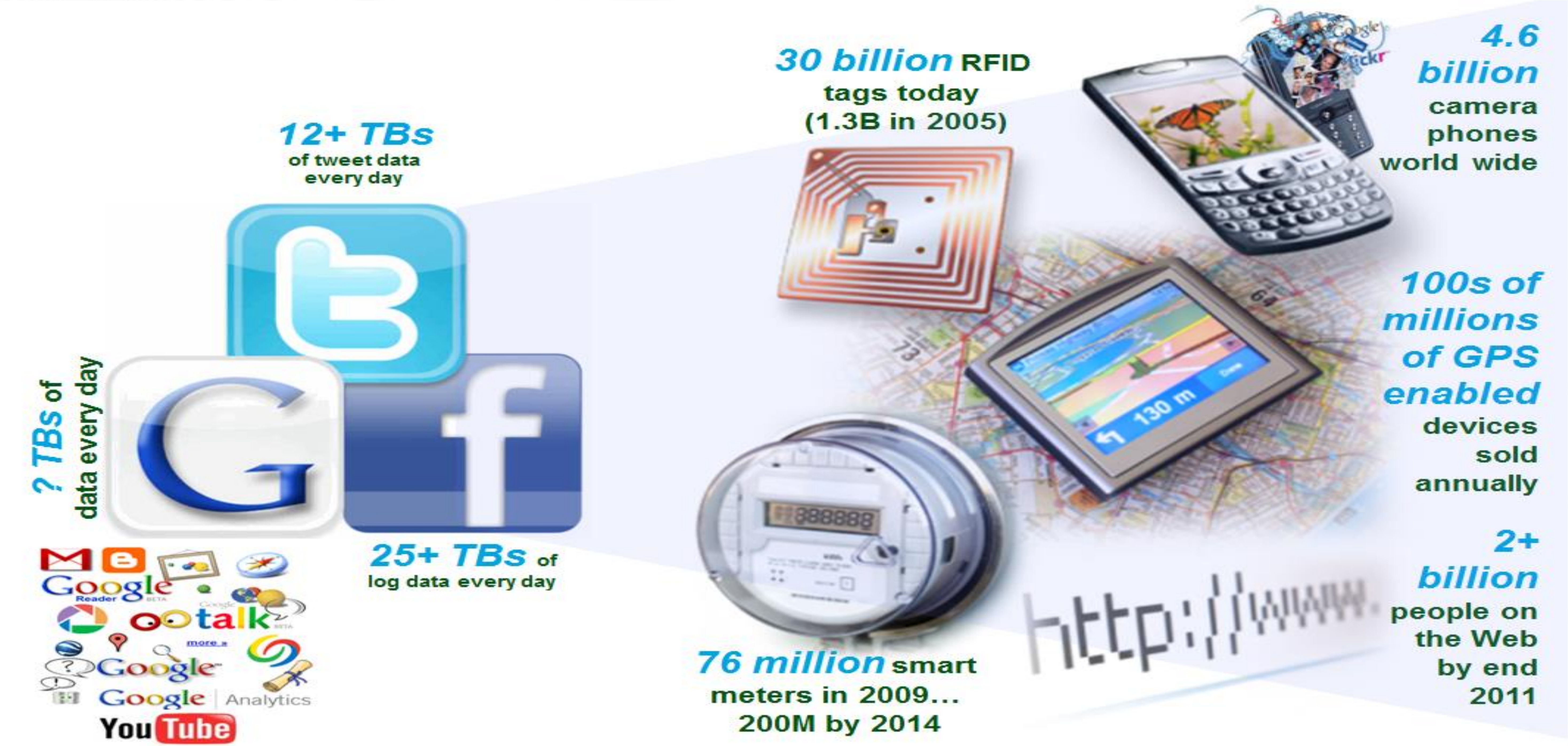
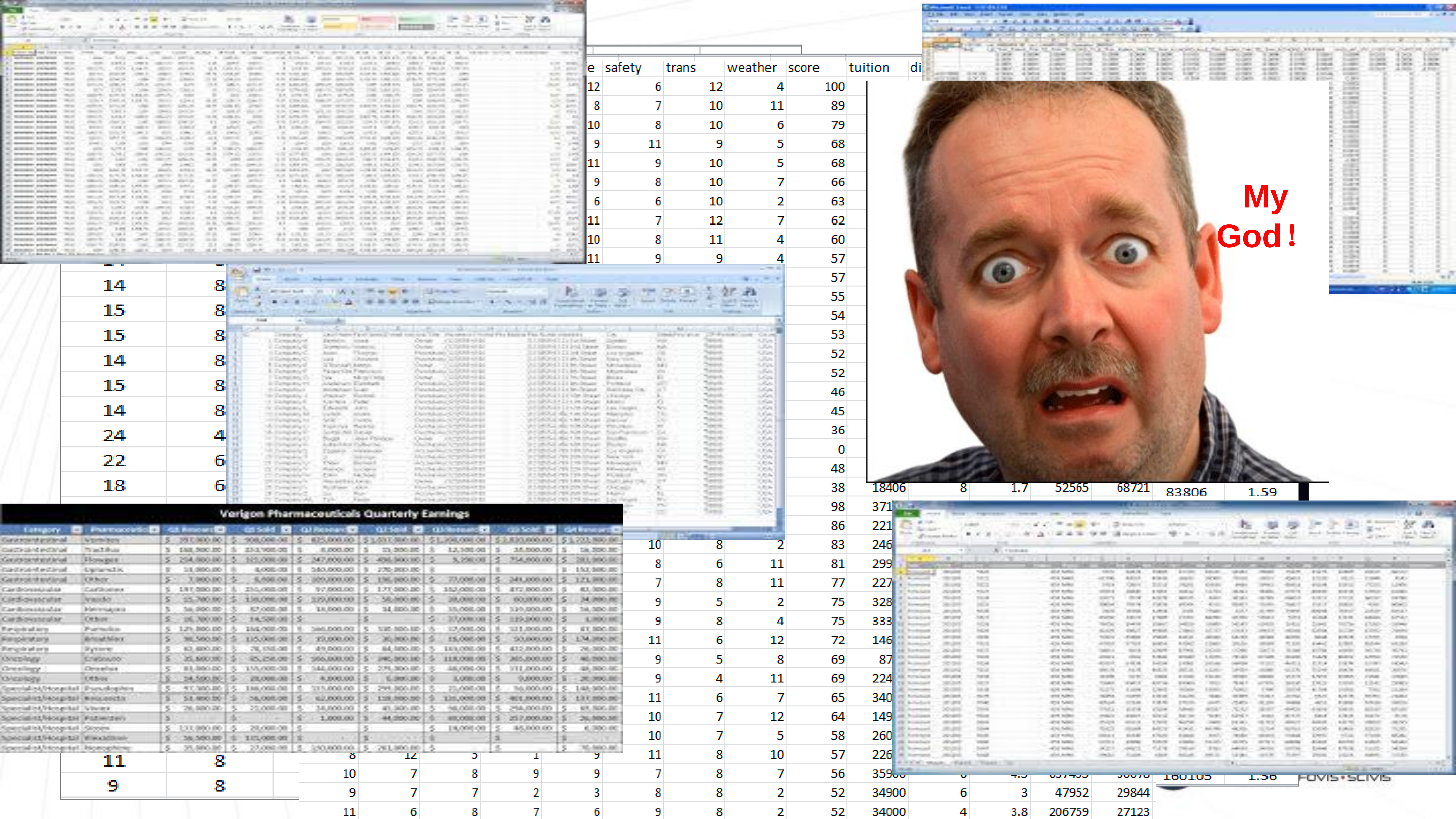
The Data Context Map: Fusing Data and Attributes into a Unified Display

Shenghui Cheng and Klaus Mueller

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

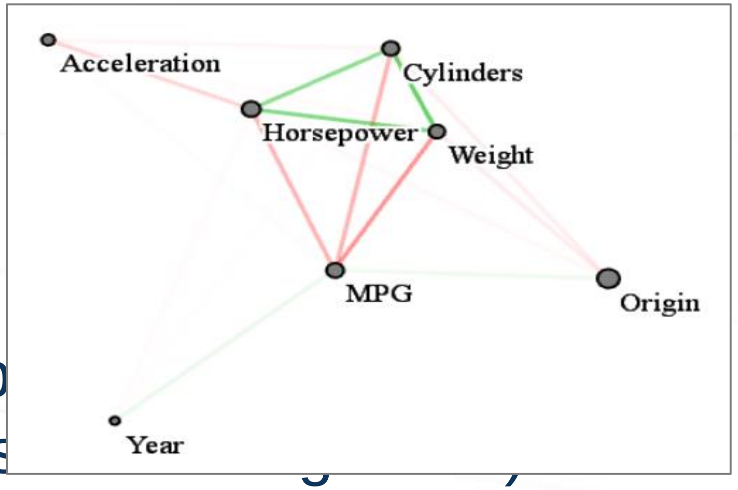
Data Explosion



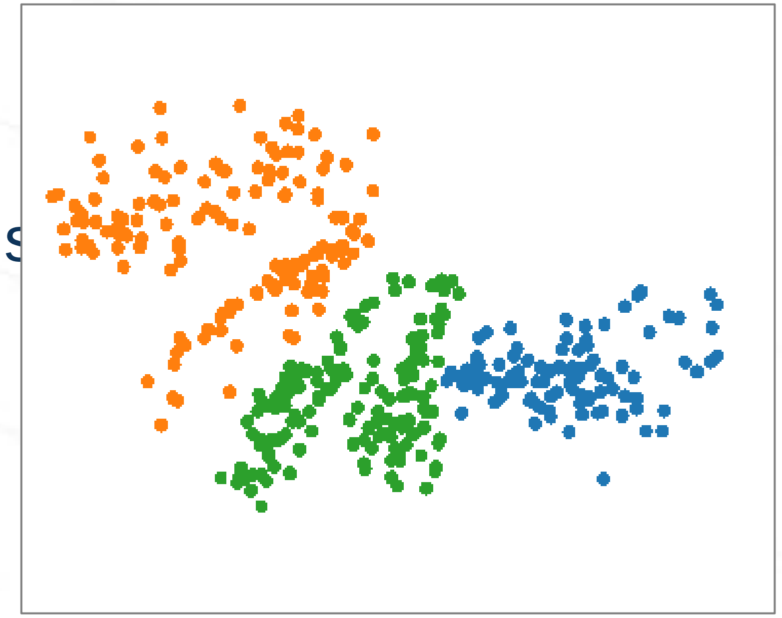

safety
trans
weather
score
tuition
Verigon Pharmaceuticals Quarterly Earnings
My
God !

Context Analysis

MPG	CYL	Hpower	Weight	Accel	Year	Origin
18	8	130	3504	12	70	1
15	8	165	3693	11.5	70	1
18	8	150	3436	11	70	1
16	8	150	3433	12	70	1
17	8	140	3449	10.5	70	1
15	8	198	4341	10	70	1
14	8	220	4354	9	70	1
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



- Attributes (Variables) to Attrib
Horsepower and weight? Pos)



- Data to Data Relation
(which cars are similar? Clus

- Context：Data to Attributes Relation ? (which cars are lastest?)

VIS2015
VAST • INFOVIS • SCIVIS

Fusion

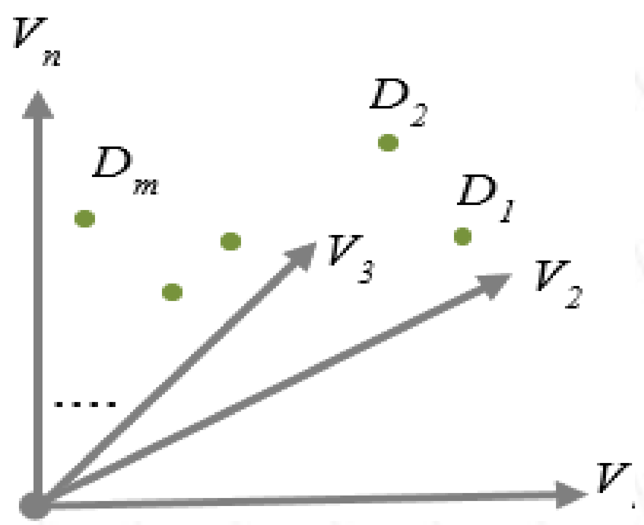
Data Matrix

$$DM = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

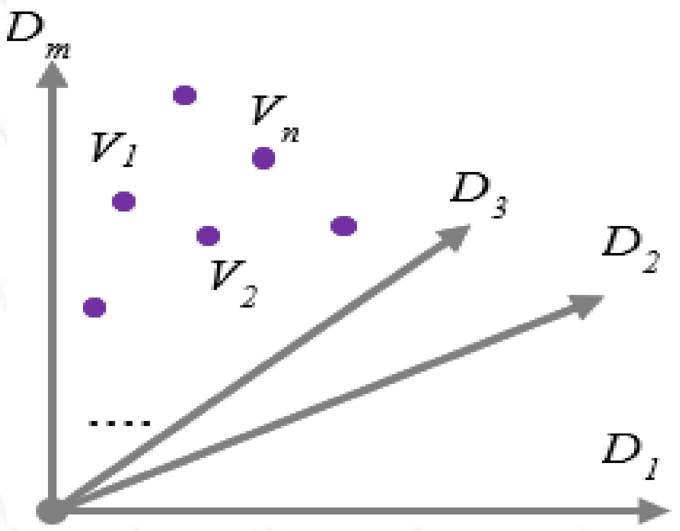
$$D_i = [x_{i1}, x_{i2}, \dots, x_{in}] \qquad (i = 1,2, \dots, m)$$

$$V_j = [x_{1j}, x_{2j}, \dots, x_{mj}]' \qquad (j = 1,2, \dots, n)$$

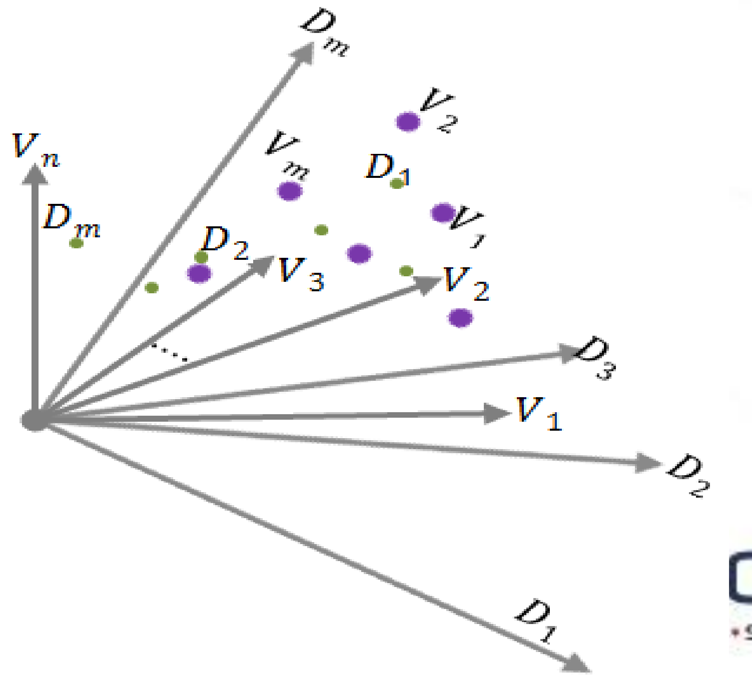
Data Space



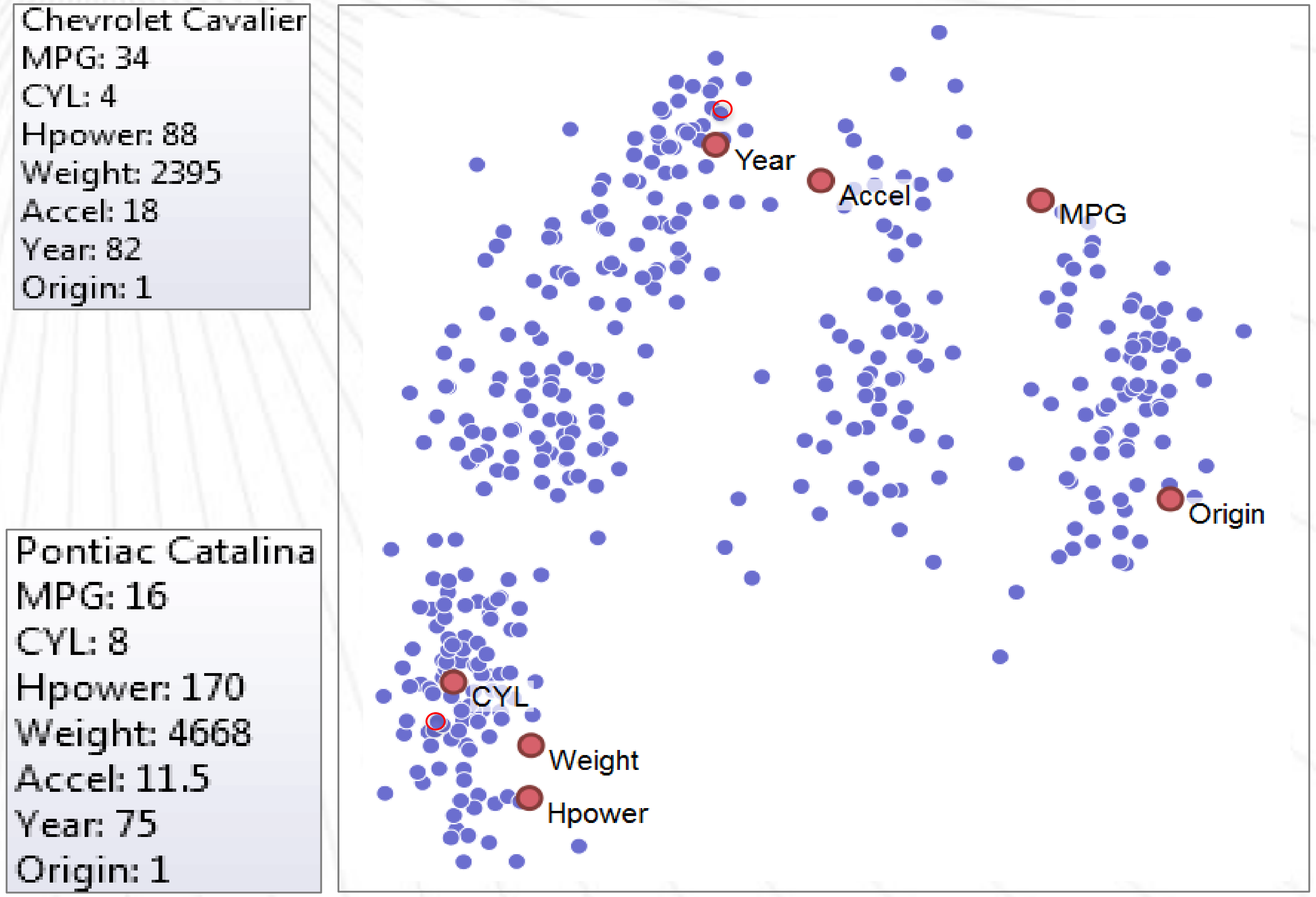
Variable Space

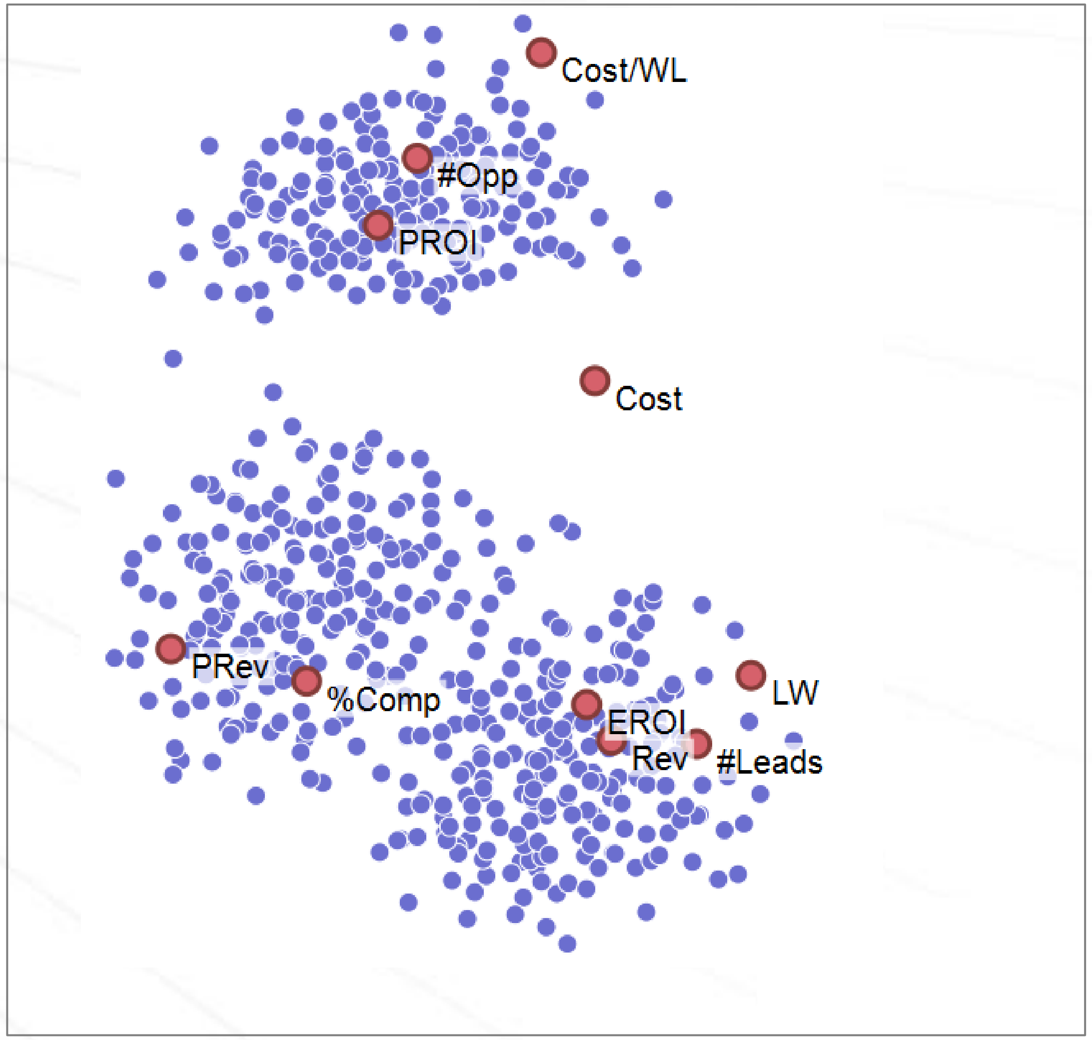


Fusion



Fusion

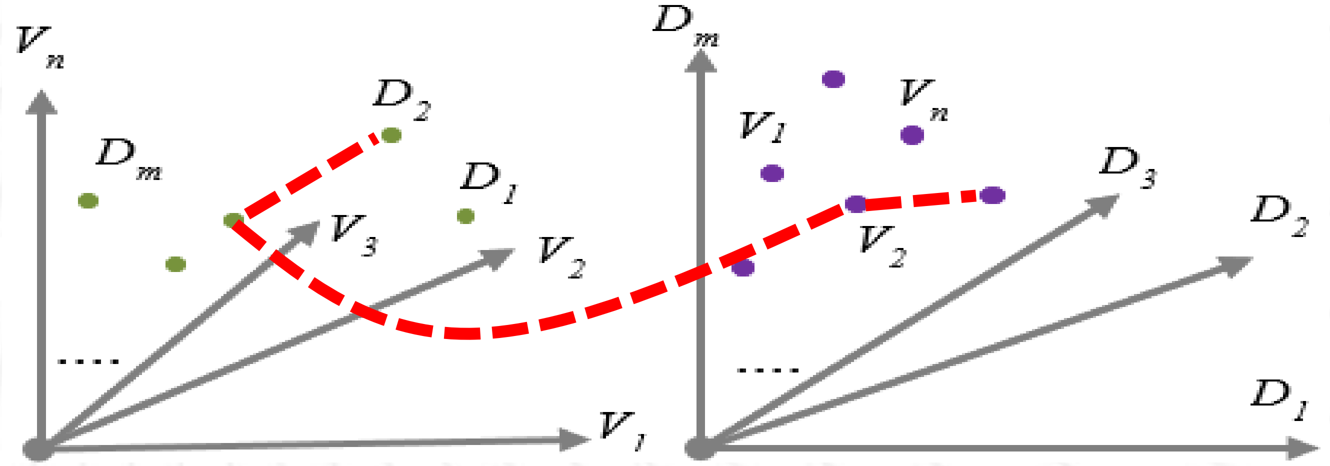

Chevrolet Cavalier
MPG: 34
CYL: 4
Hpower: 88
Weight: 2395
Accel: 18
Year: 82
Origin: 1
Pontiac Catalina
MPG: 16
CYL: 8
Hpower: 170
Weight: 4668
Accel: 11.5
Year: 75
Origin: 1
Year
Accel
MPG
Origin
CYL
Weight
Hpower


Cost/WL
#Opp
PROI
Cost
PRev
%Comp
EROI
Rev
LW
#Leads

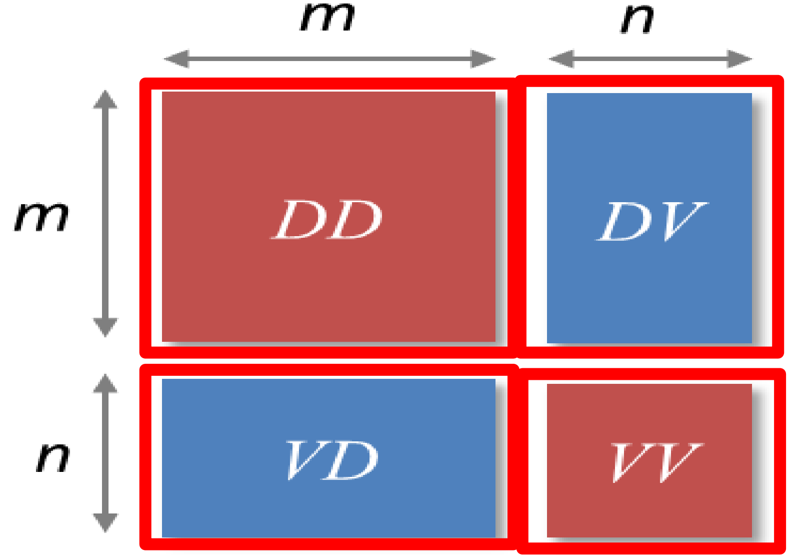
Fusion Pipeline

Data Space

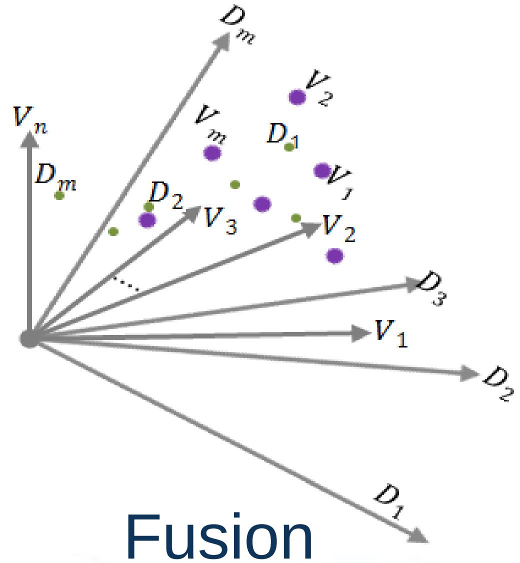
Variable Space



Distance Matrix

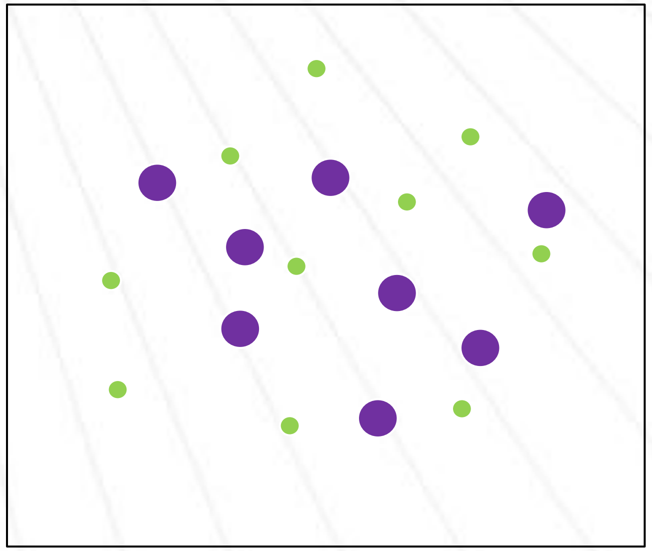


Transformation



Fusion

Glimmer MDS



Mapping

Distance Measurement

MPG	CYL	Hpower	Weight	Accel	Year	Origin
18	8	130	3504	12	70	1
15	8	165	3693	11.5	70	1
18	8	150	3436	11	70	1
16	8	150	3433	12	70	1
17	8	140	3449	10.5	70	1
15	8	198	4341	10	70	1
14	8	220	4354	9	70	1
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
27	4	88	2130	14.5	71	3
28	4	90	2264	15.5	71	1
25	4	95	2228	14	71	3

Variable to Variable Distance (*VV*)：

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Distance is opposite meaning !

1- Correlation`

Data to Data Distance (*DD*)： Euclidean Distance

Data to Variable (*DV*) or Variable to Data (*VD*) Distance???

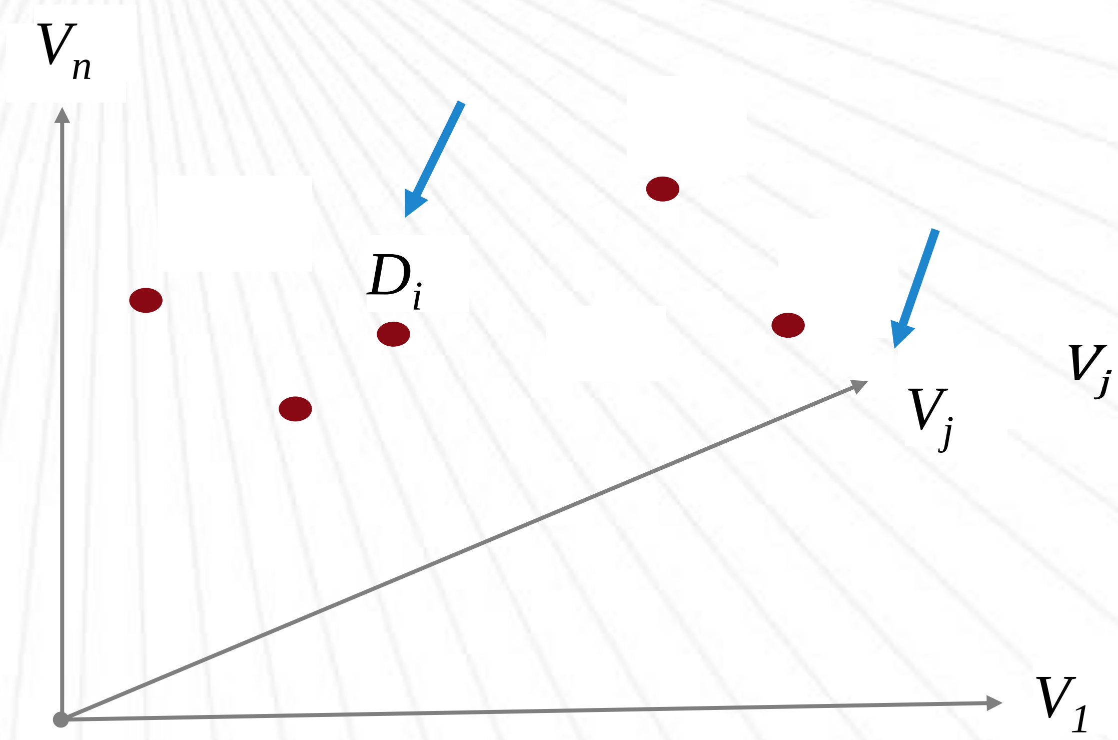
Not equal length!

V_1

Distance Measurement

n dimensional

$$D_i = [x_{i1}, x_{i2}, \dots, x_{in}]$$



Project to *j*th dimension: 1

Project to the other dimensions: 0

$$V_j = [x_{1j}, x_{2j}, \dots, x_{mj}]'$$

n dimensional !

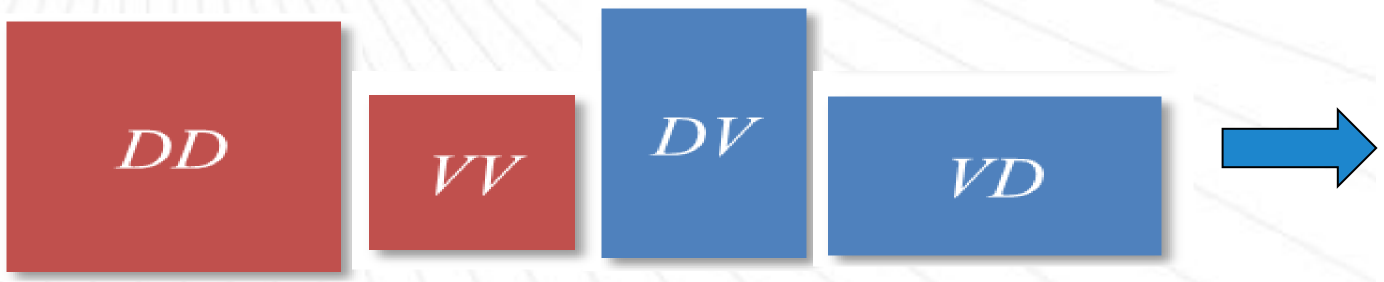
$$V_j = [\overbrace{0,0,\dots,1}^{\text{j numbers}}, 0, \dots, 0]^T$$

D_i and V_j are equal length now.

Transpose?

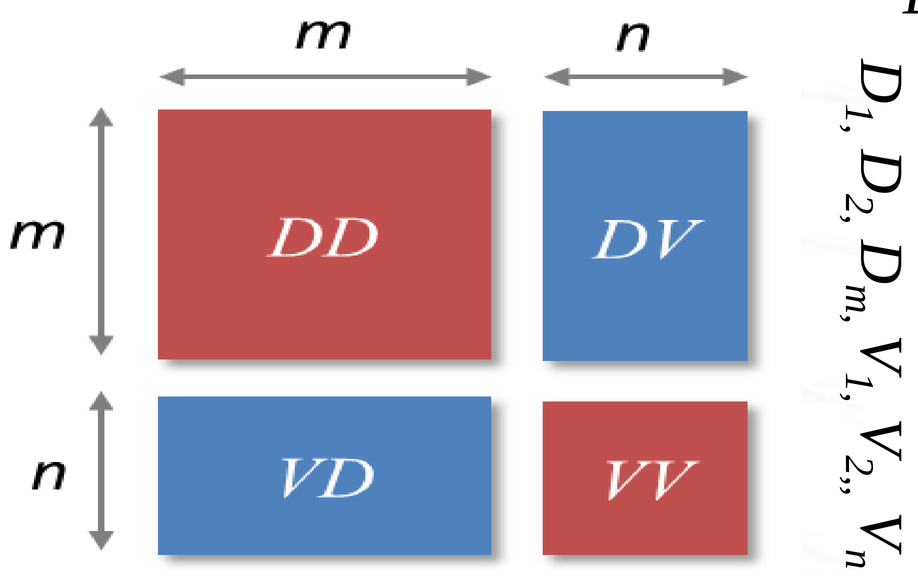
DV=VD !

Composite Distance Matrix

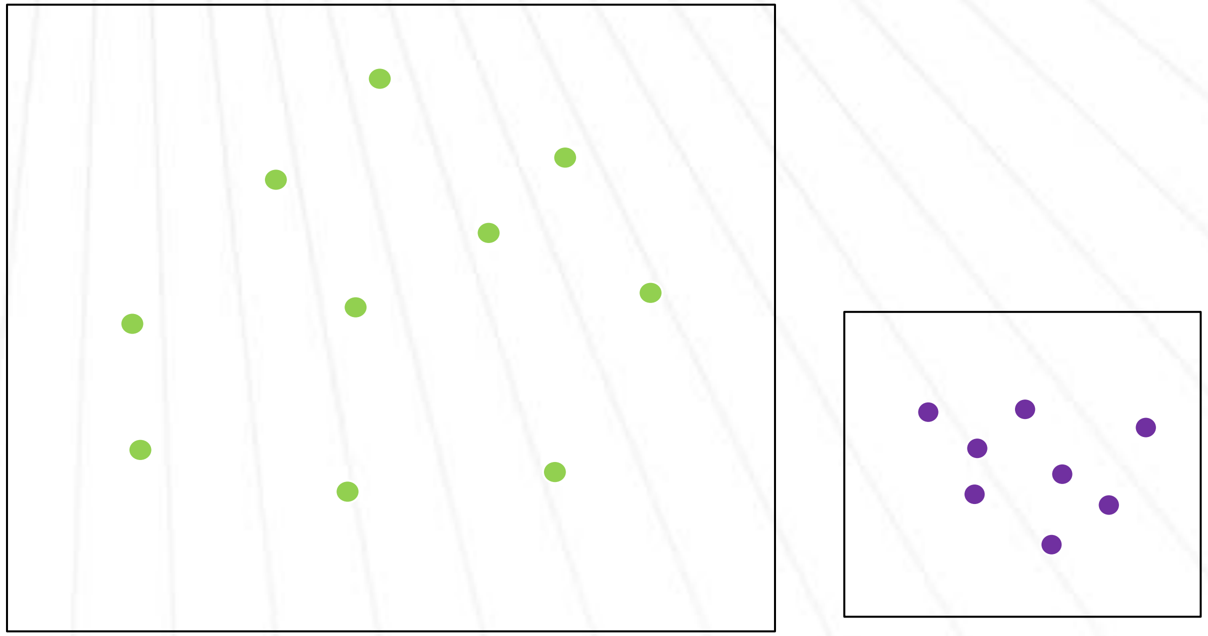


Distance Matrix

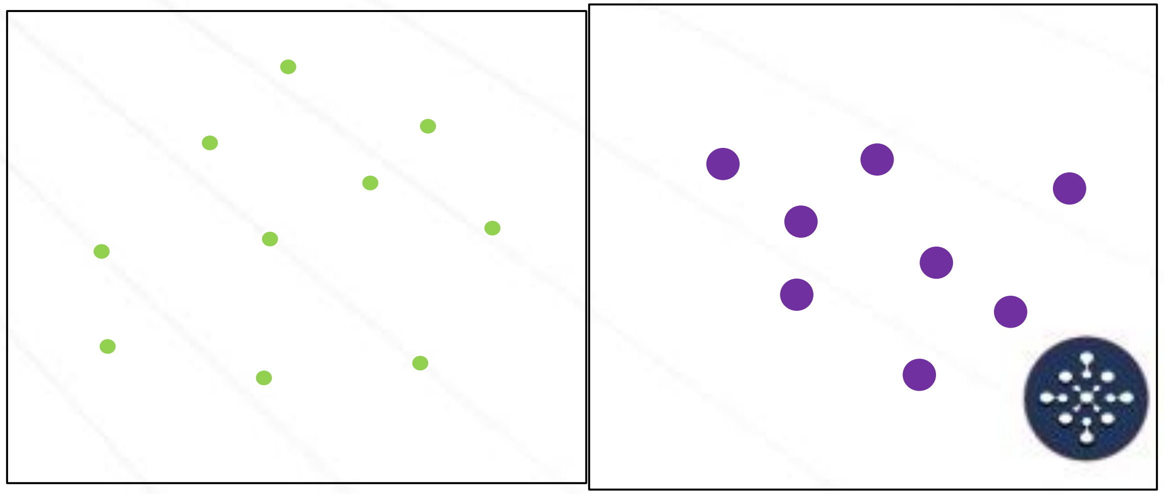
$D_1, D_{2,\ldots,} D_m, V_1, V_{2,\ldots,} V_n$



Scale?



$$\overline{DD} = \overline{DV} = \overline{VD} = \overline{VV}$$

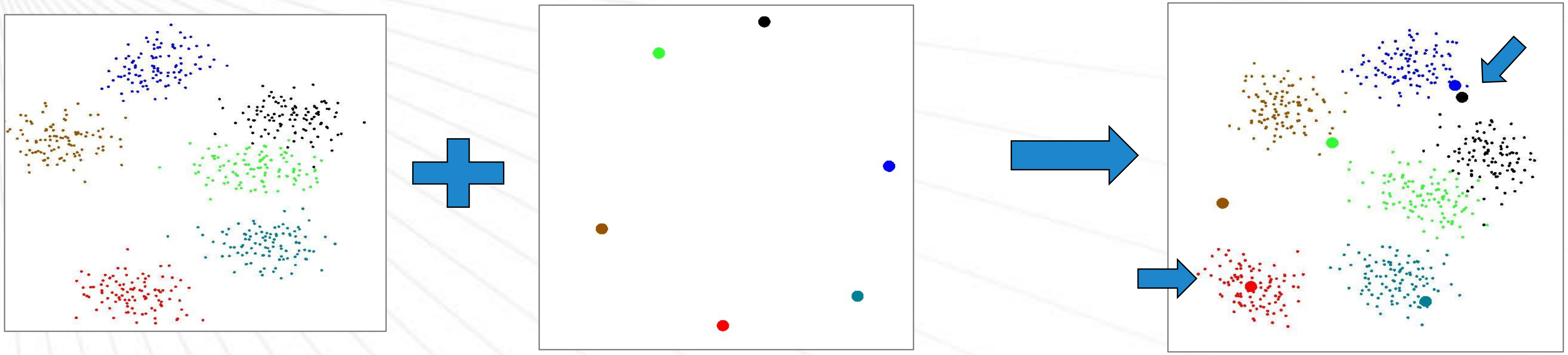


At a Glance

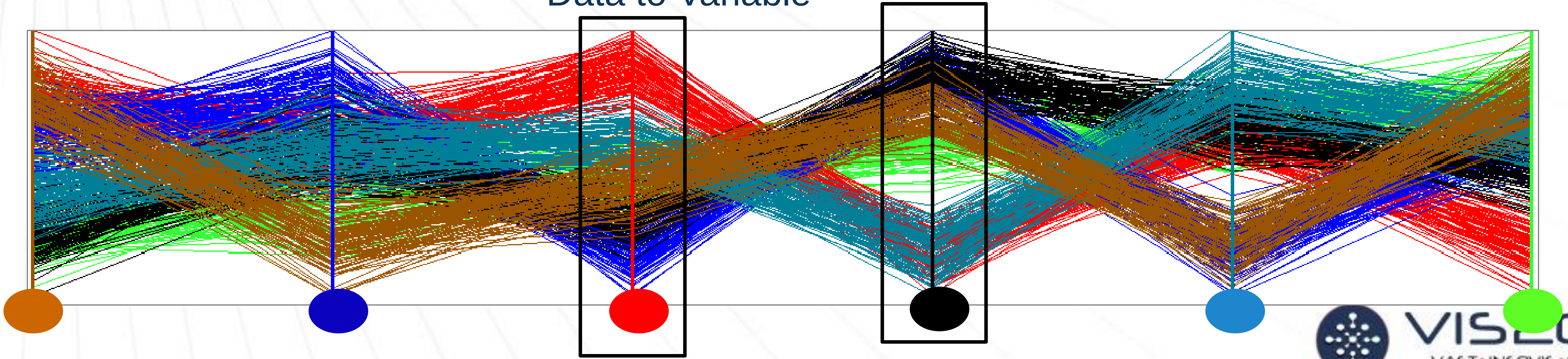
Data to Data

Variable to Variable

Data and Variable

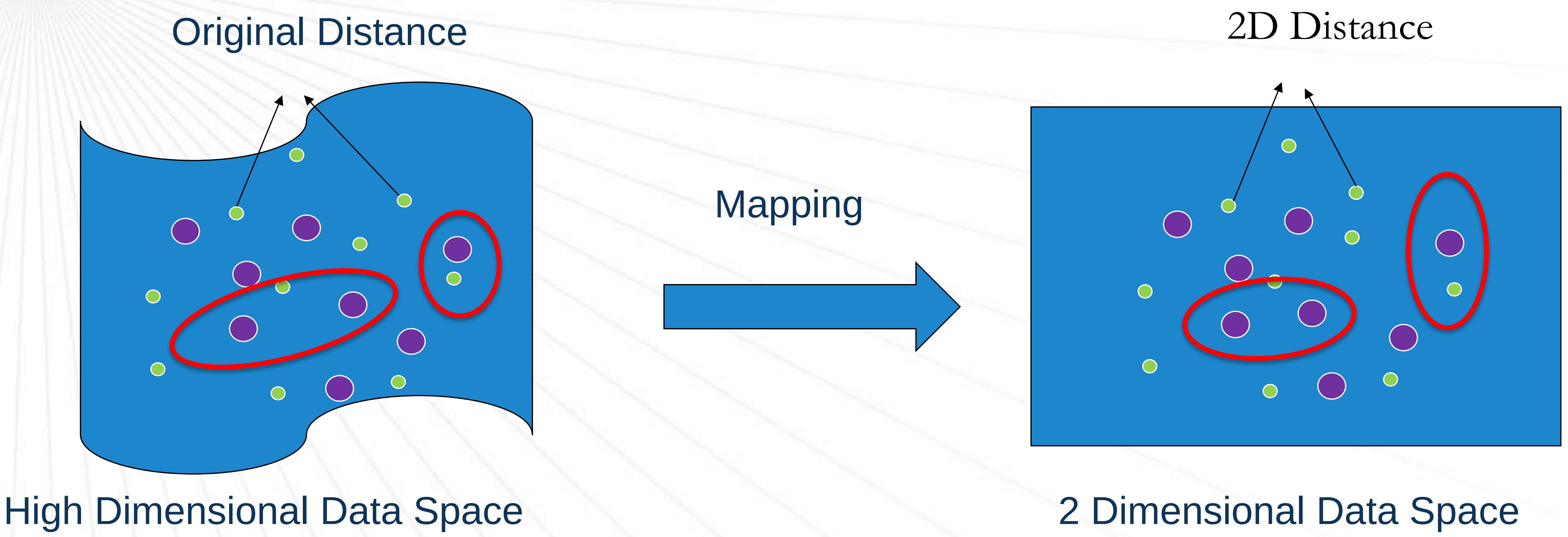


Data to Variable



VIS2015
VAST • INFOVIS • SCIVIS

Layout Schedules

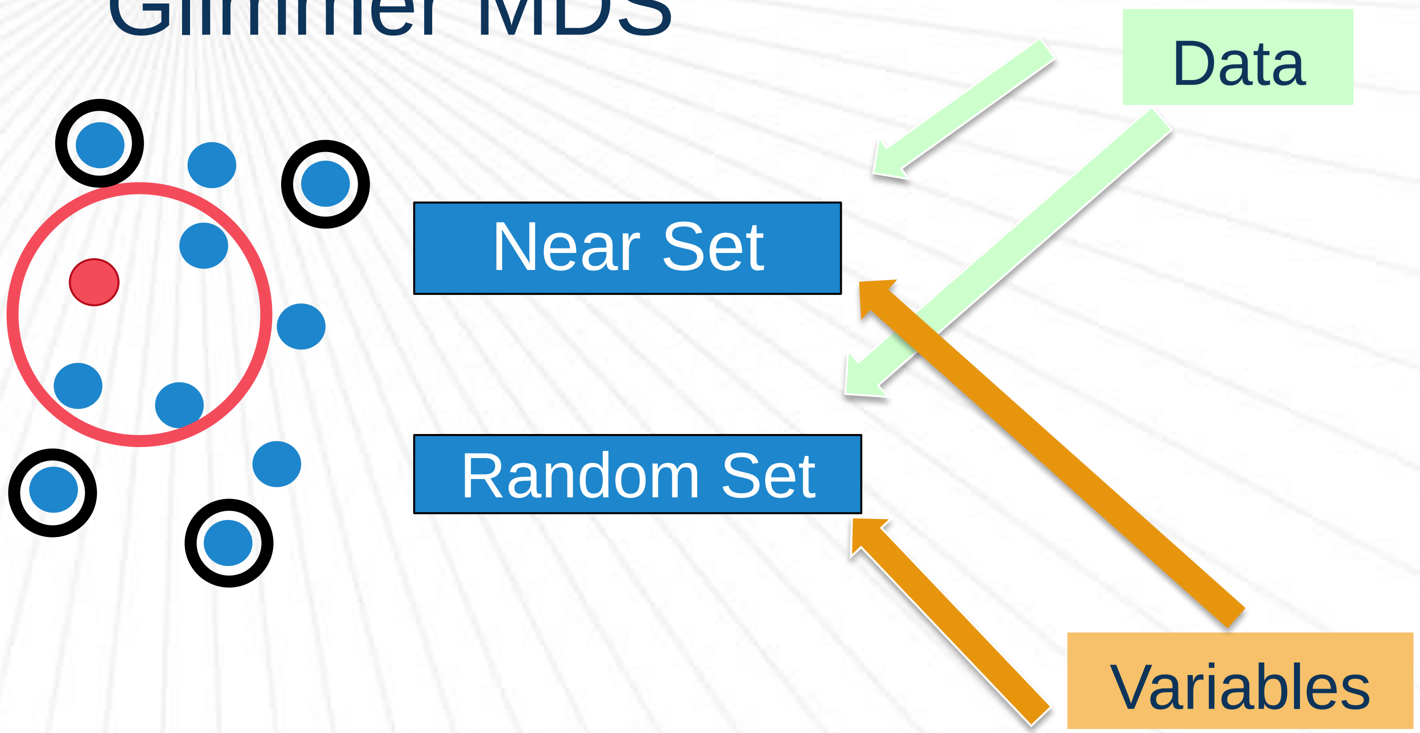


Data or Variable Error?

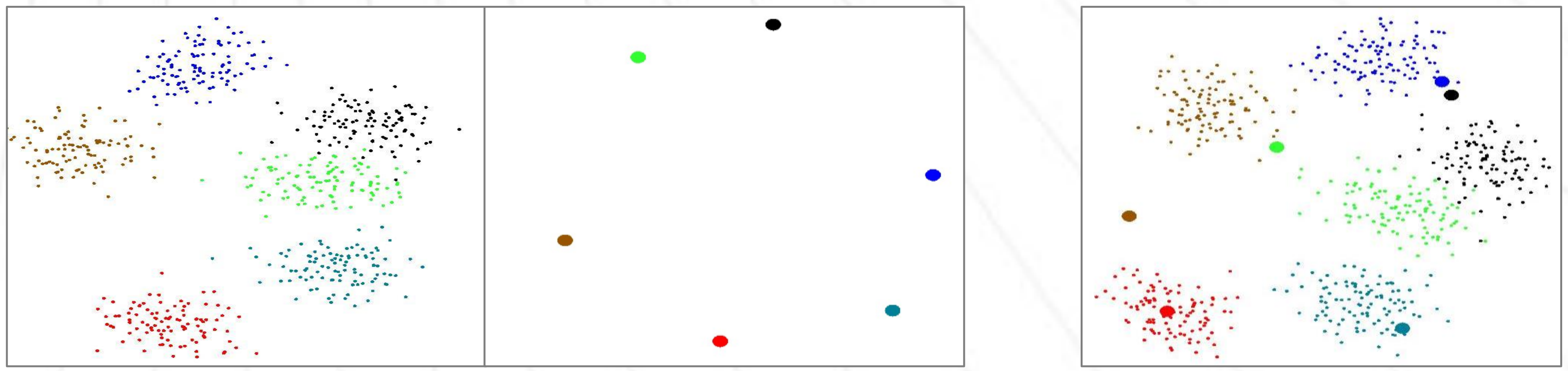
VIS2015
VAST • INFOVIS • SCIVIS

Layout Schedules

Glimmer MDS

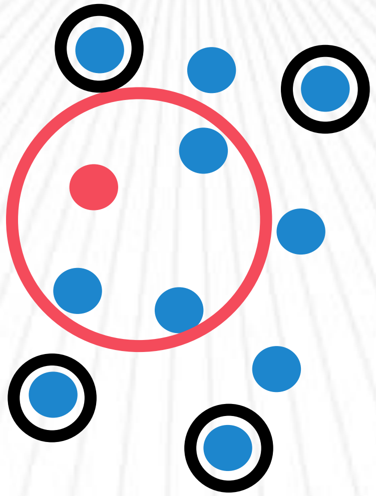


Data Variables



Layout Schedules

Glimmer MDS



Data

Near Set

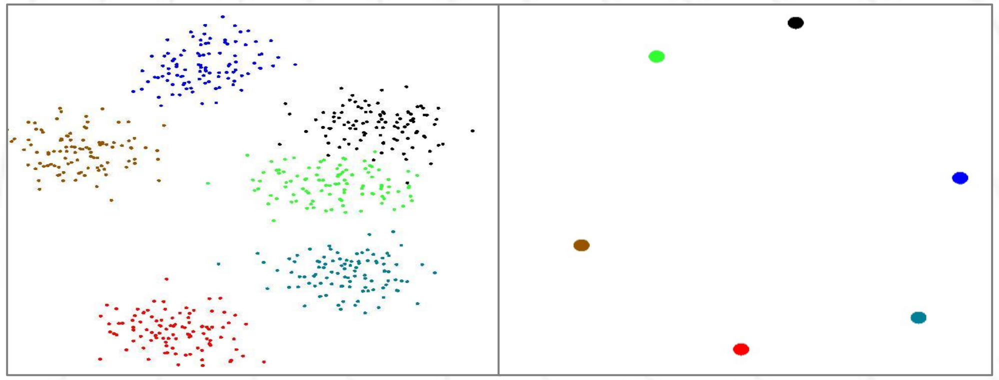
Random Set

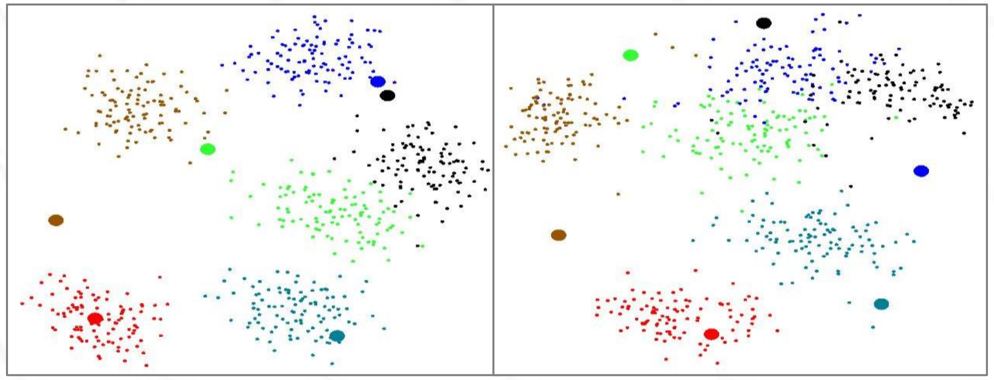
Variables

Data Variables

Variables

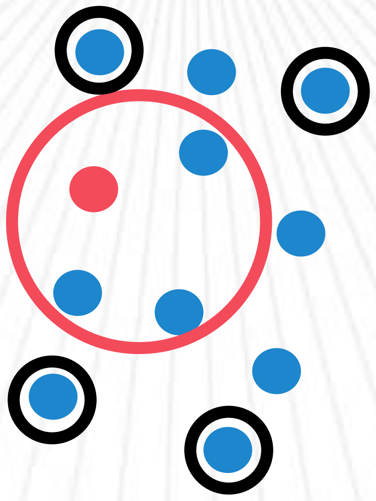
Data Variables





Layout Schedules

Glimmer MDS



Near Set

Random Set

Data

Variables

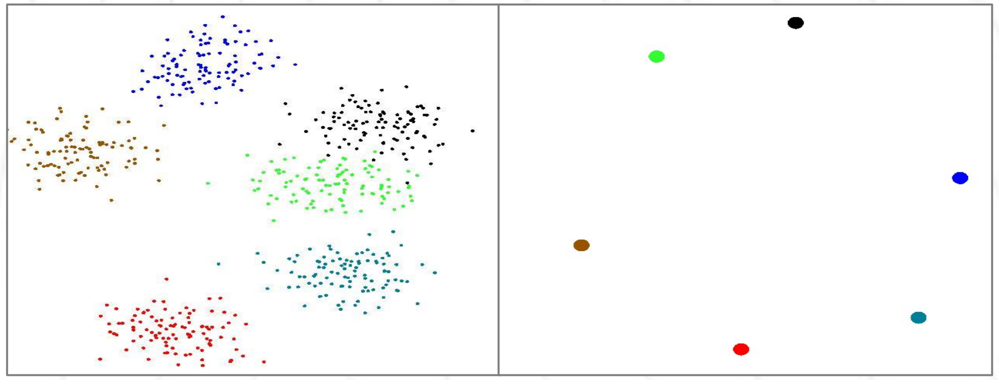
Data Variables

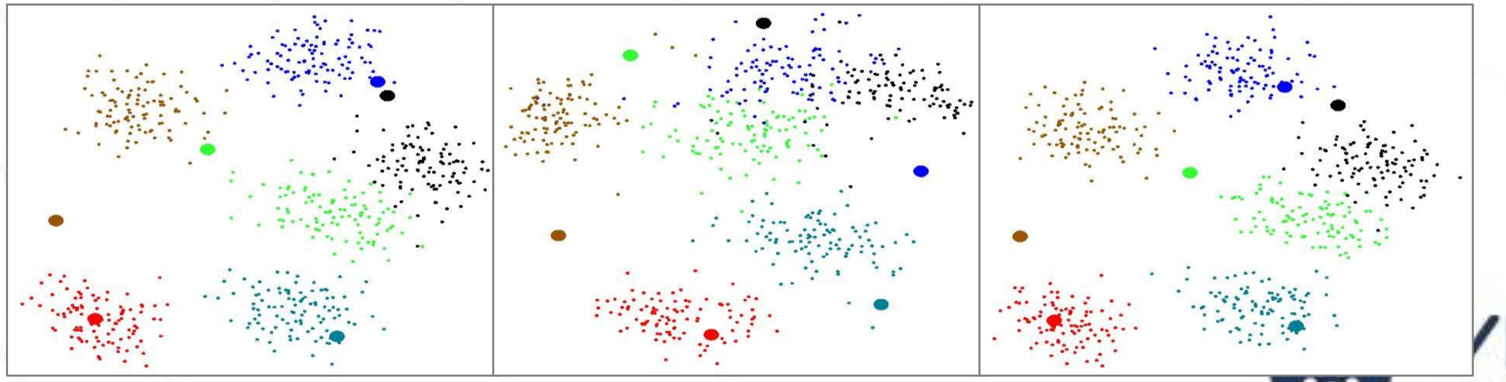
Variables

Data Variables

Data

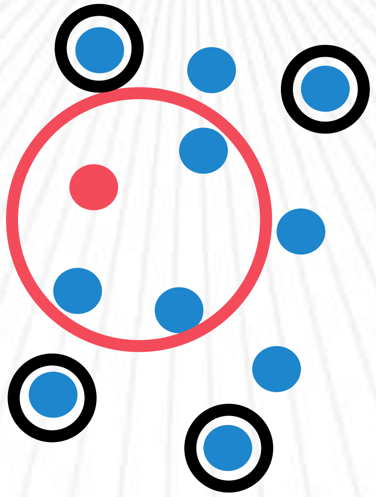
Data Variables





Layout Schedules

Glimmer MDS



Near Set

Random Set

Data

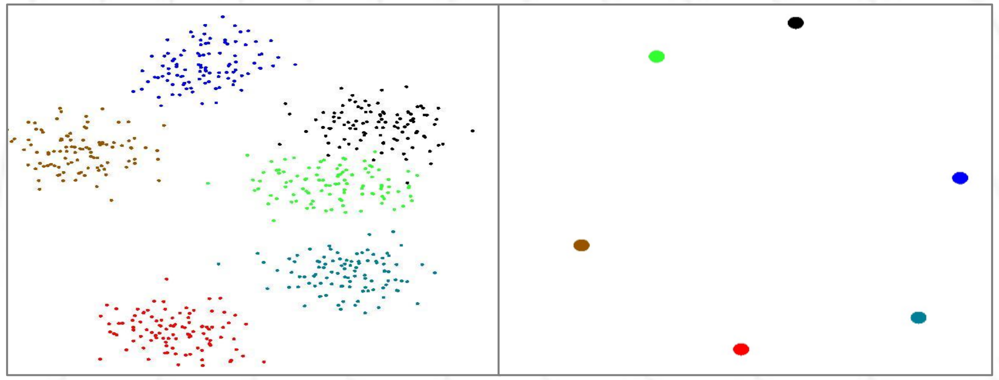
Variables

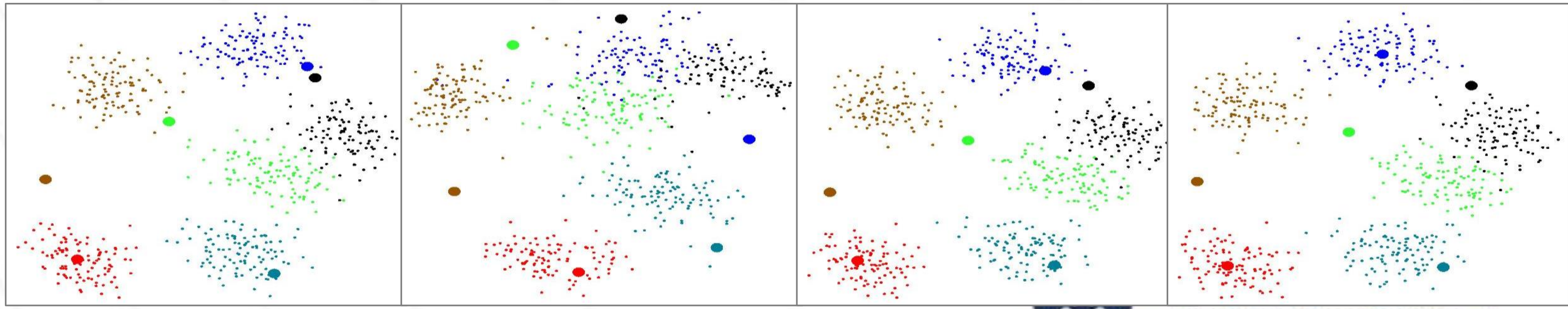
Data Variables

Variables Data Variables

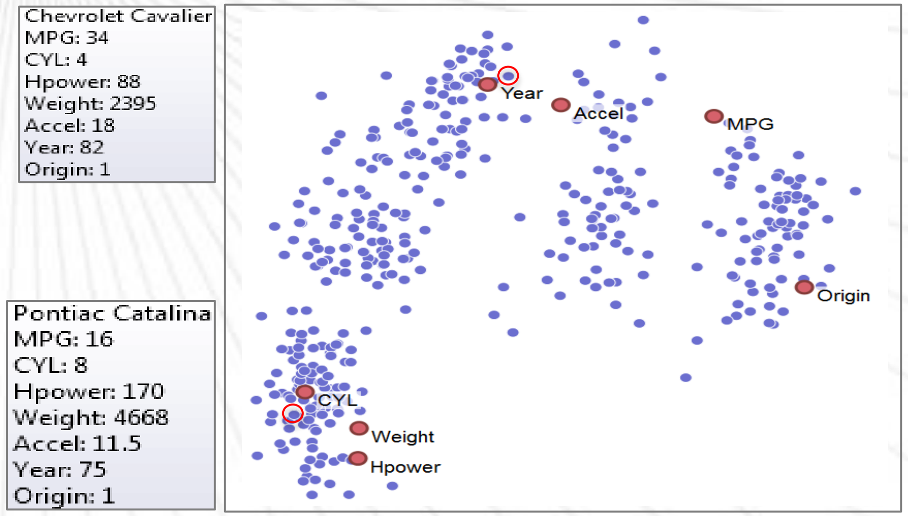
Data Data Variables

Variables Data Data Variables

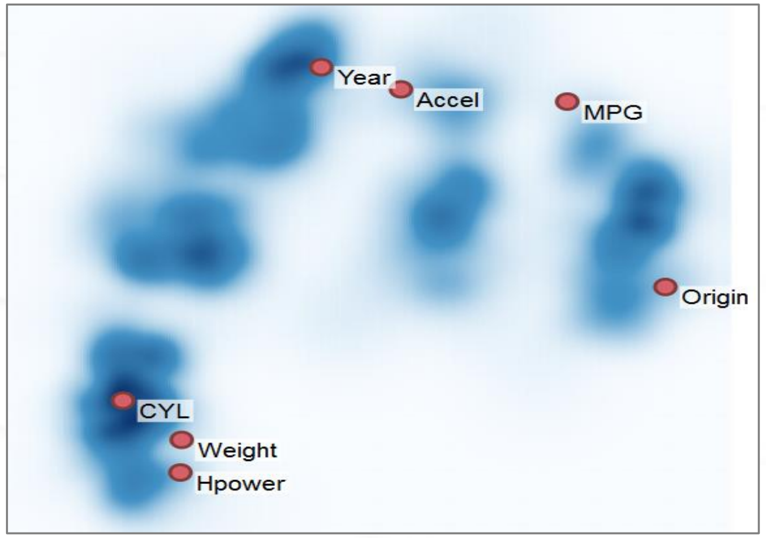




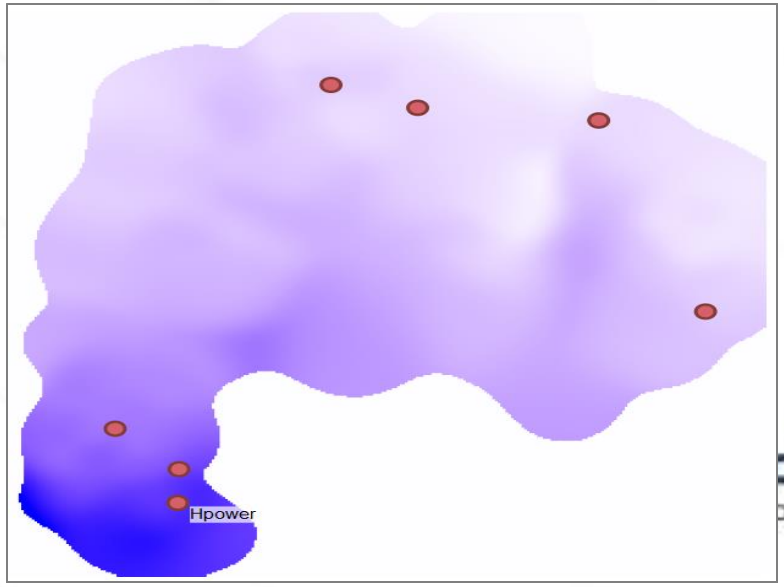
Data Context Map



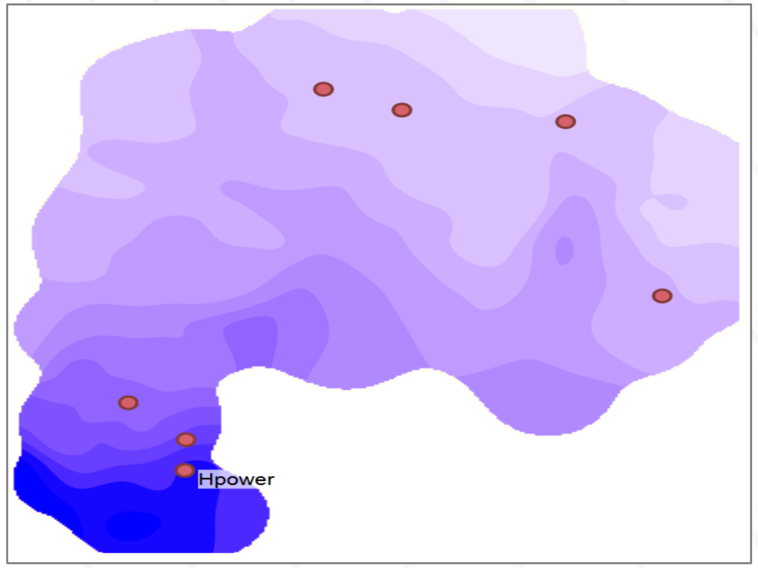
Adaptive Kernel Density Estimation



Nadaraya-Watson kernel regression



Contour boundary

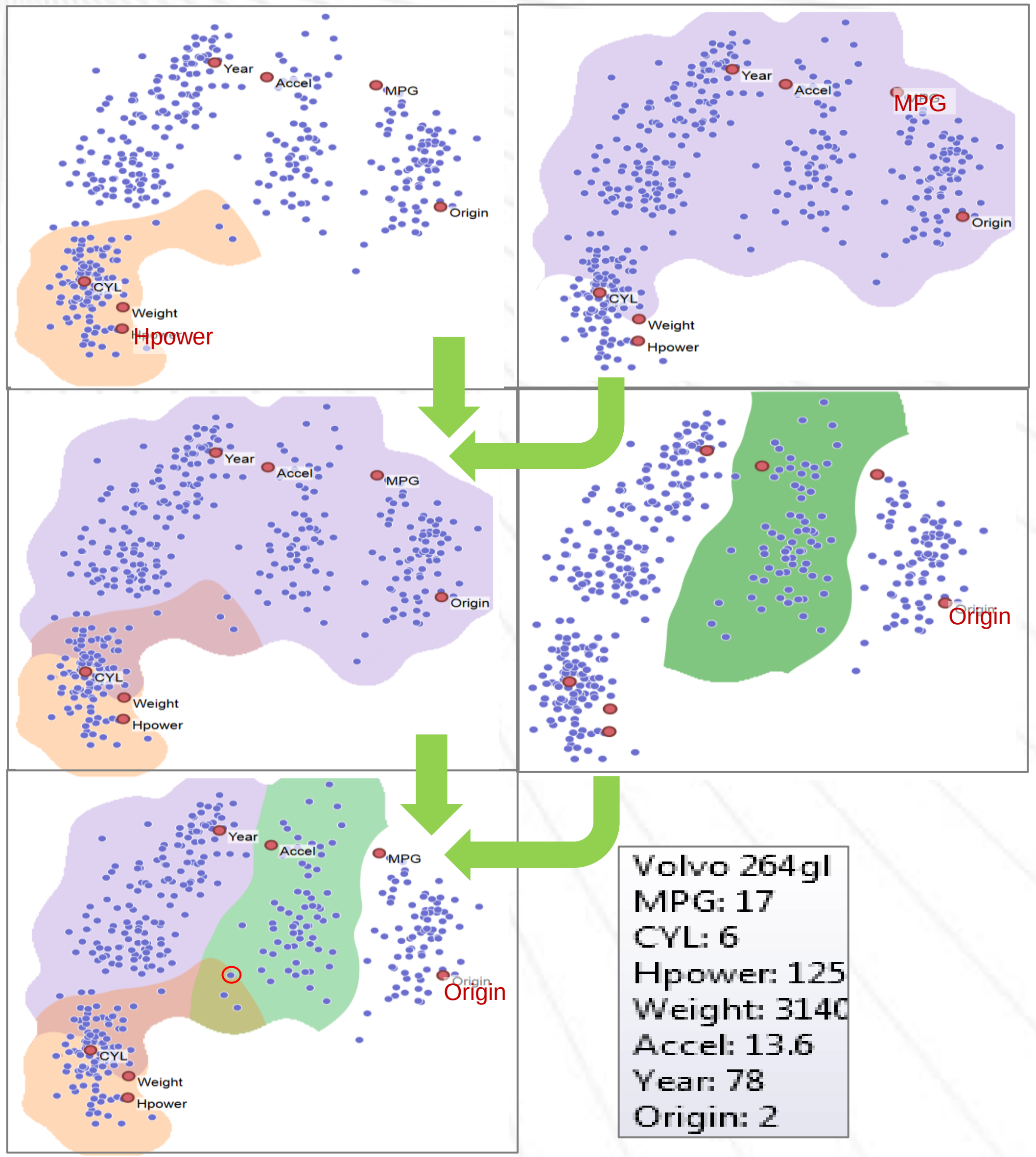


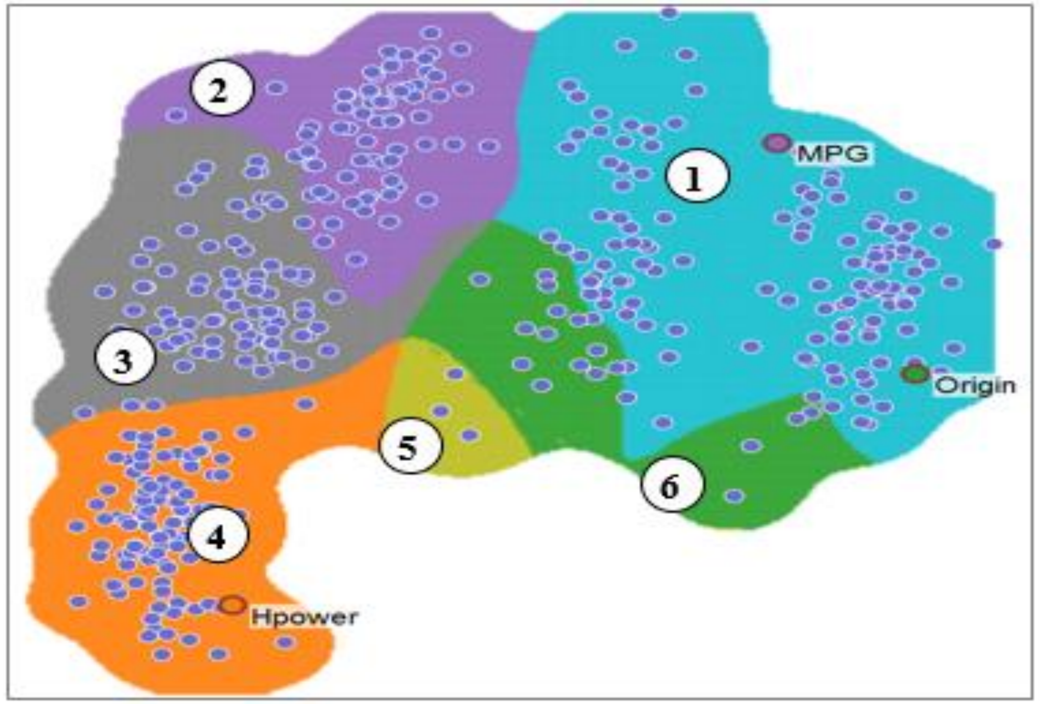
Data Context Map

Horsepower: 116~192

MPG: 16~30

European Car





- Euro-Japanese efficient compact cars
- US efficient compact cars
- US semi-efficient medium-power cars
- US big block gas guzzlers
- Euro-Japanese gas guzzlers
- Euro-Japanese semi-efficient medium-power cars

Data Context Map: Fusing Data and Attributes into a Unified Display

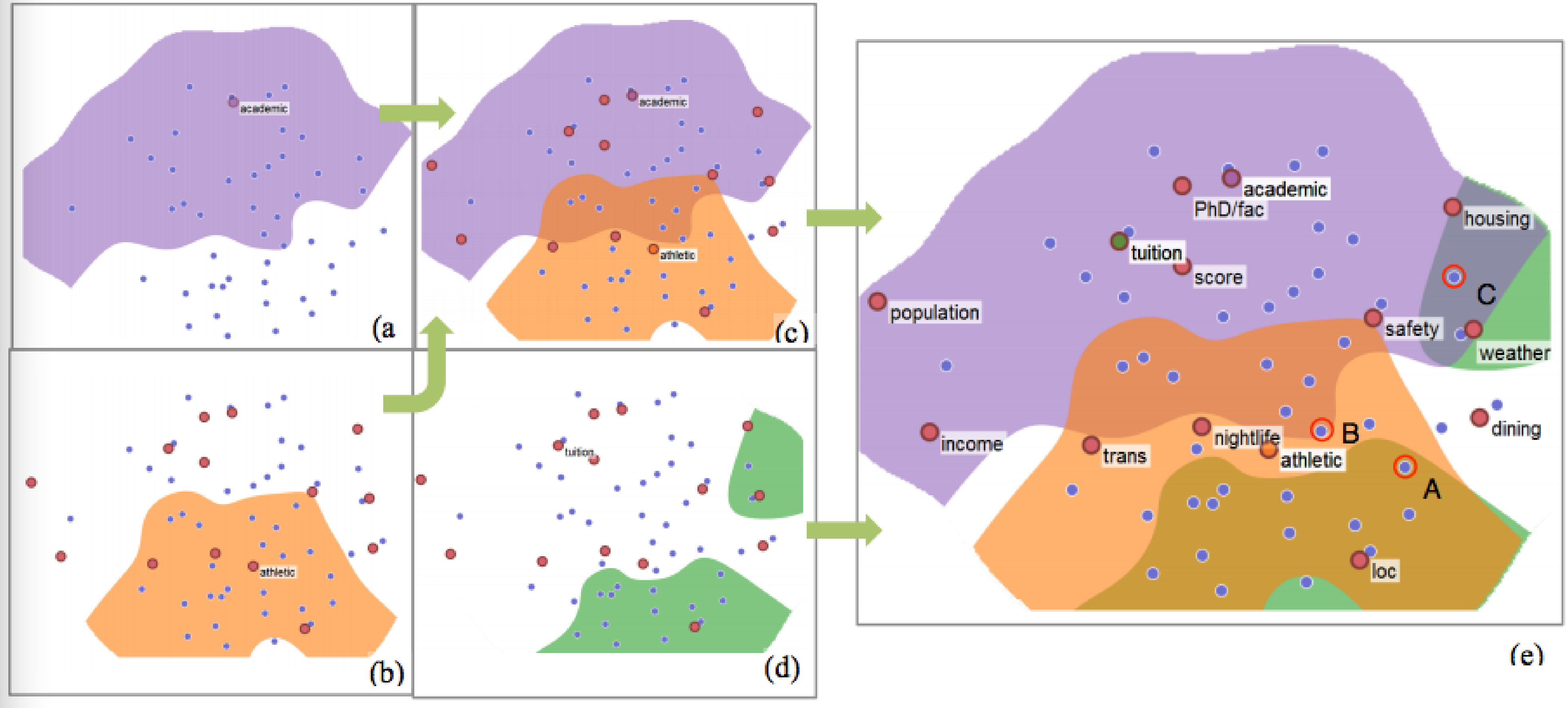
Shenghui Cheng and Klaus Mueller

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

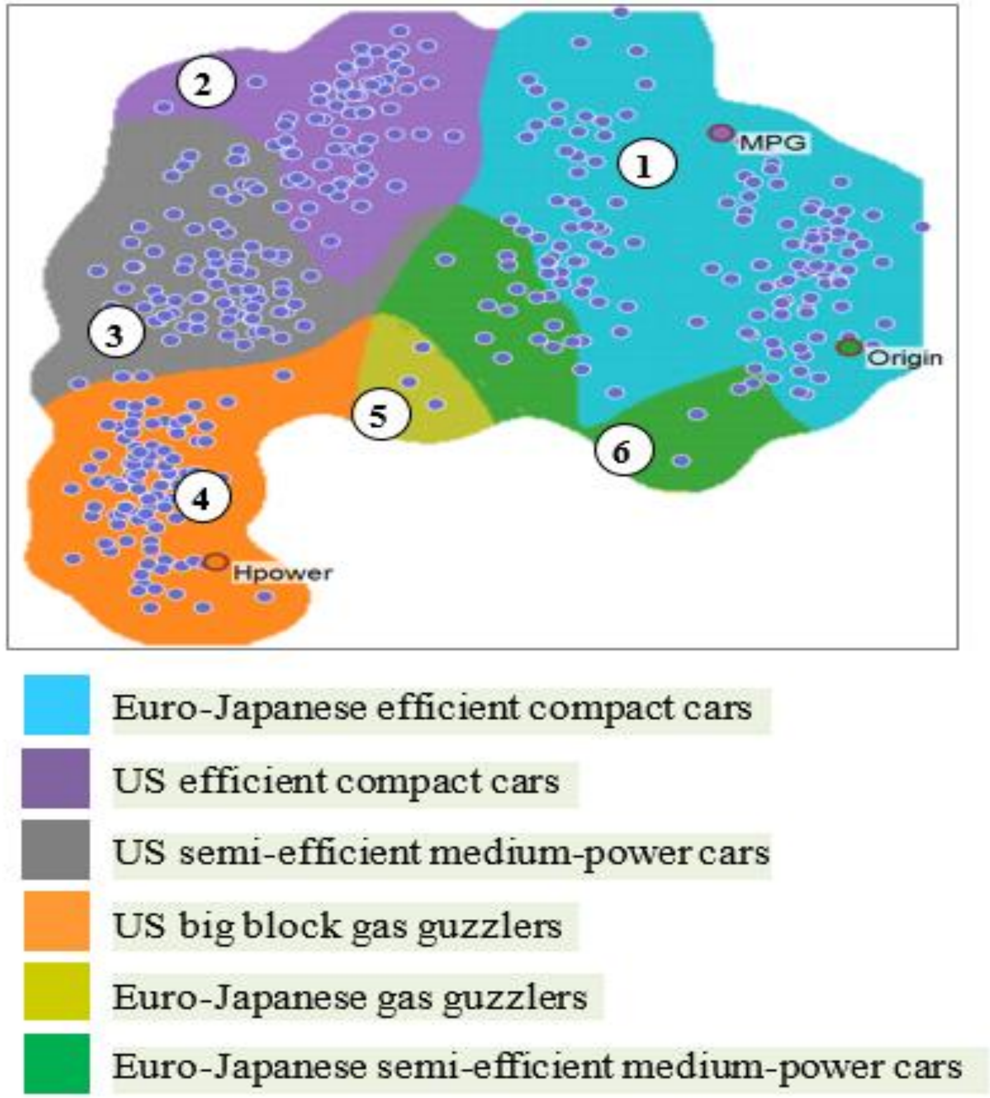
Reference

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.

VIS2015
VAST • INFOVIS • SCIVIS



(a)
(b)
(c)
(d)
(e)
academic
PhD/fac
housing
tuition
score
C
population
safety
weather
dining
income
trans
nightlife
B
athletic
A
loc



1
2
3
4
5
6
MPG
Origin
Hpower
Euro-Japanese efficient compact cars
US efficient compact cars
US semi-efficient medium-power cars
US big block gas guzzlers
Euro-Japanese gas guzzlers
Euro-Japanese semi-efficient medium-power cars

Thank you !