

ParallelSpaces: Simultaneous Exploration of Feature and Data for Hypothesis Generation

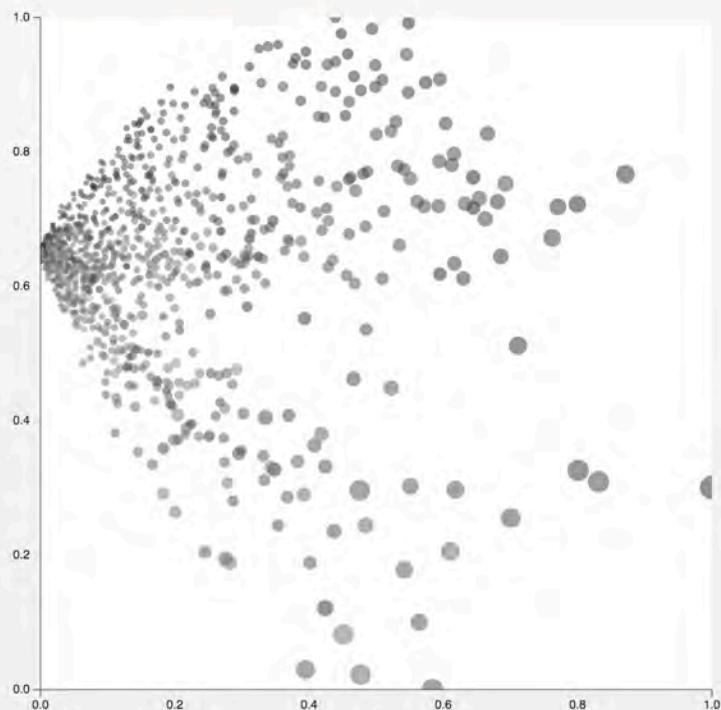
Deok Gun Park^{\$}, Jungu Choi^{*}, Niklas Elmqvist^{\$}

^{\$}University of Maryland

^{*}Purdue University

Movie Space

Search movies by name



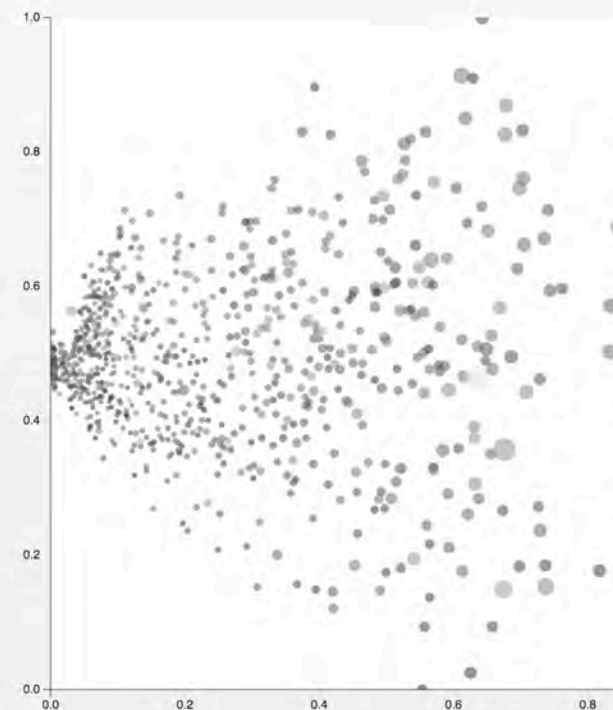
X Axis:

Similarity

Y Axis:

Similarity

User Space



X Axis:

Similarity

Y Axis:

Contour Plot Configuration

- + On/Off
- + Density Mode
- + Bandwidth

Selection Configuration

- + On/Off
- + Selection Criteria
- + Group Mode

Why Visual Analytics?

	User_Id	Japan	Type	Currency	Bitcoin	Money	Money_Rate	Money_JPY	Money_Fee	Money_Fee_Rate	Money_Fee_JPY	Bitcoin_Fee	Bitcoin_Fee_JPY
2013-06-04 6:48	238168	NJP	buy	USD	0.58932091	22.39419	93.542	2094.796	0	93.542	0	0.0031234	10.651
2013-06-04 6:48	109955	NJP	sell	USD	0.58932091	22.39419	93.542	2094.796	0.13437	93.542	12.569	0	0
2013-06-04 6:49	238168	NJP	buy	USD	0.39820074	15.13163	93.542	1415.442	0	93.542	0	0.00211046	7.197
2013-06-04 6:49	132909	NJP	sell	USD	0.39820074	15.13163	93.542	1415.442	0.08322	93.542	7.785	0	0
2013-06-04 6:49	698630	JP	buy	USD	1.70382	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	96376	NJP	sell	USD	1.70382	15.13163	93.542	1415.442	0.19462	93.542	18.205	0	0
2013-06-04 6:49	698630	JP	buy	USD	1	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	201601	NJP	sell	USD	1	15.13163	93.542	1415.442	0.17531	93.542	16.399	0	0
2013-06-04 6:49	698630	JP	buy	USD	21	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	95850	NJP	sell	USD	21	15.13163	93.542	1415.442	2.3209	93.542	217.101	0	0
2013-06-04 6:49	698630	JP	buy	USD	28.067686	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	95850	NJP	sell	USD	28.067686	15.13163	93.542	1415.442	3.10201	93.542	290.168	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.0326	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	67861	NJP	sell	USD	0.0326	15.13163	93.542	1415.442	0.00746	93.542	0.698	0	0
2013-06-04 6:49	698630	JP	buy	USD	41	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	74362	NJP	sell	USD	41	15.13163	93.542	1415.442	6.72056	93.542	628.654	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.211874	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	10078	NJP	sell	USD	0.211874	15.13163	93.542	1415.442	0.03473	93.542	3.249	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.1	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	99189	NJP	sell	USD	0.1	15.13163	93.542	1415.442	0.01907	93.542	1.784	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.01000001	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	90128	NJP	sell	USD	0.01000001	15.13163	93.542	1415.442	0.00229	93.542	0.214	0	0
2013-06-04 6:49	698630	JP	buy	USD	50	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	125	NJP	sell	USD	50	15.13163	93.542	1415.442	8.7722	93.542	820.569	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.14111	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	8456	NJP	sell	USD	0.14111	15.13163	93.542	1415.442	0.0323	93.542	3.021	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.5	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	178803	NJP	sell	USD	0.5	15.13163	93.542	1415.442	0.0687	93.542	6.426	0	0
2013-06-04 6:49	698630	JP	buy	USD	5	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	37547	NJP	sell	USD	5	15.13163	93.542	1415.442	1.1451	93.542	107.115	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.02	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	281813	NJP	sell	USD	0.02	15.13163	93.542	1415.442	0.00458	93.542	0.428	0	0
2013-06-04 6:49	698630	JP	buy	USD	1.49	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	87995	NJP	sell	USD	1.49	15.13163	93.542	1415.442	0.26162	93.542	24.472	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.1	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	99189	NJP	sell	USD	0.1	15.13163	93.542	1415.442	0.01909	93.542	1.786	0	0
2013-06-04 6:49	698630	JP	buy	USD	0.01000001	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	90128	NJP	sell	USD	0.01000001	15.13163	93.542	1415.442	0.00229	93.542	0.214	0	0
2013-06-04 6:49	698630	JP	buy	USD	5.28072	15.13163	93.542	1415.442	0	93.542	0	0	0
2013-06-04 6:49	90737	NJP	sell	USD	5.28072	15.13163	93.542	1415.442	0.92737	93.542	86.748	0	0

	User_Id	Japan	Type	Currency	Bitcoins	Money	Money_Rate	Money_JPY	Money_Fee	Money_Fee_Rate	Money_Fee_JPY	Bitcoin_Fee	Bitcoin_Fee_JPY
-2013 6:48	238168	NJP	buy	USD	0.58932091	22.39419	93.542	2094.796	0	93.542	0	0.0031234	10.651
-2013 6:48	109955	NJP	sell	USD	0.58932091	22.39419	93.542	2094.796	0.13437	93.542	12.569	0	0
-2013 6:49	238168	NJP	buy	USD	0.39820074	15.13163	93.542	1415.442	0	93.542	0	0.00211046	7.197
-2013 6:49	132909	NJP	sell	USD	0.39820074	15.13163	93.542	1415.442	0.08322	93.542	7.785	0	0
-2013 6:49	698630	JP	buy	USD	1.70382	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	96376	NJP	sell	USD	1.70382	15.13163	93.542	1415.442	0.19462	93.542	18.205	0	0
-2013 6:49	698630	JP	buy	USD	1	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	201601	NJP	sell	USD	1	15.13163	93.542	1415.442	0.17531	93.542	16.399	0	0
-2013 6:49	698630	JP	buy	USD	21	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	95850	NJP	sell	USD	21	15.13163	93.542	1415.442	2.3209	93.542	217.101	0	0
-2013 6:49	698630	JP	buy	USD	28.067686	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	95850	NJP	sell	USD	28.067686	15.13163	93.542	1415.442	3.10201	93.542	290.168	0	0
-2013 6:49	6												0
-2013 6:49													0
-2013 6:49	6												0
-2013 6:49													0
-2013 6:49	6												0
-2013 6:49													0
-2013 6:49	6												0
-2013 6:49													0
-2013 6:49	698630	JP	buy	USD	0.01000001	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	90128	NJP	sell	USD	0.01000001	15.13163	93.542	1415.442	0.00229	93.542	0.214	0	0
-2013 6:49	698630	JP	buy	USD	50	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	125	NJP	sell								20.569	0	0
-2013 6:49	698630	JP	buy								0	0	0
-2013 6:49	8456	NJP	sell								3.021	0	0
-2013 6:49	698630	JP	buy								0	0	0
-2013 6:49	178803	NJP	sell								6.426	0	0
-2013 6:49	698630	JP	buy								0	0	0
-2013 6:49	37547	NJP	sell								07.115	0	0
-2013 6:49	698630	JP	buy								0	0	0
-2013 6:49	281813	NJP	sell								0.428	0	0
-2013 6:49	698630	JP	buy	USD	1.49	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	87995	NJP	sell	USD	1.49	15.13163	93.542	1415.442	0.26162	93.542	24.472	0	0
-2013 6:49	698630	JP	buy	USD	0.1	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	99189	NJP	sell	USD	0.1	15.13163	93.542	1415.442	0.01909	93.542	1.786	0	0
-2013 6:49	698630	JP	buy	USD	0.01000001	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	90128	NJP	sell	USD	0.01000001	15.13163	93.542	1415.442	0.00229	93.542	0.214	0	0
-2013 6:49	698630	JP	buy	USD	5.28072	15.13163	93.542	1415.442	0	93.542	0	0	0
-2013 6:49	90737	NJP	sell	USD	5.28072	15.13163	93.542	1415.442	0.92737	93.542	86.748	0	0

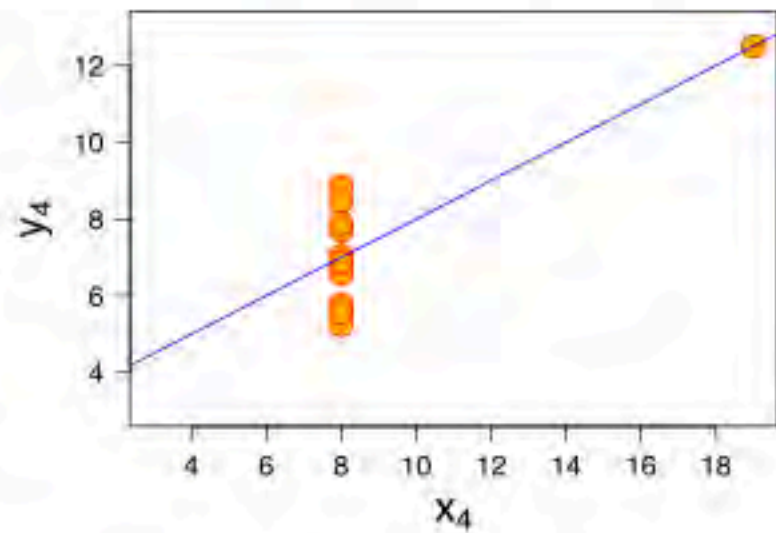
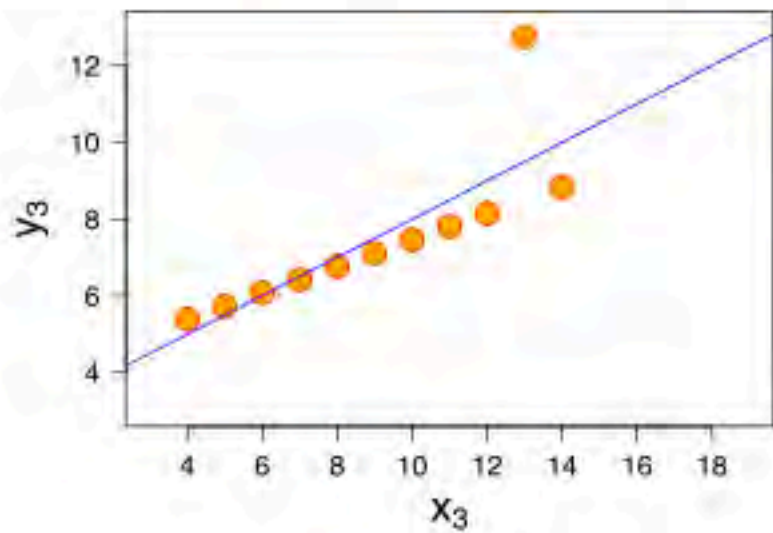
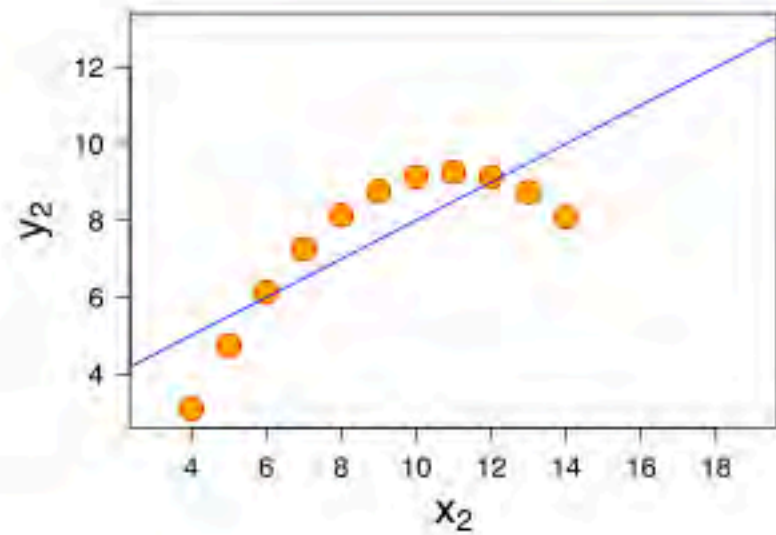
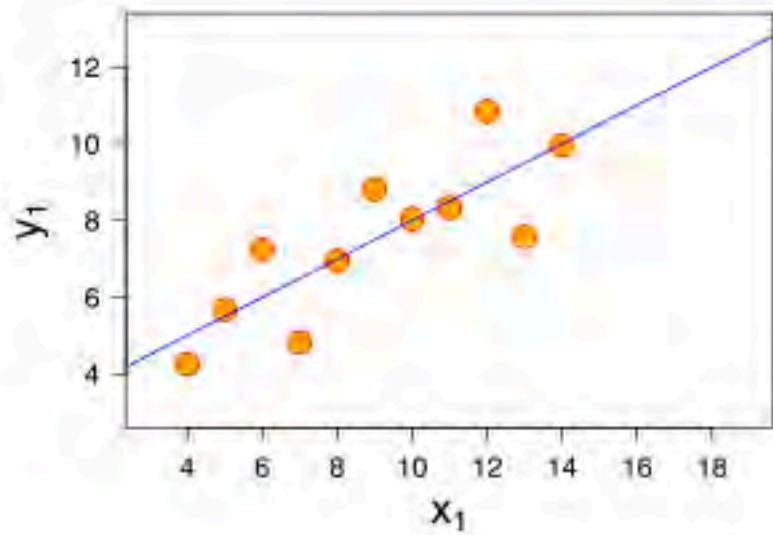
Statistical modeling to tackle
challenge of Big Data

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + e_i,$$

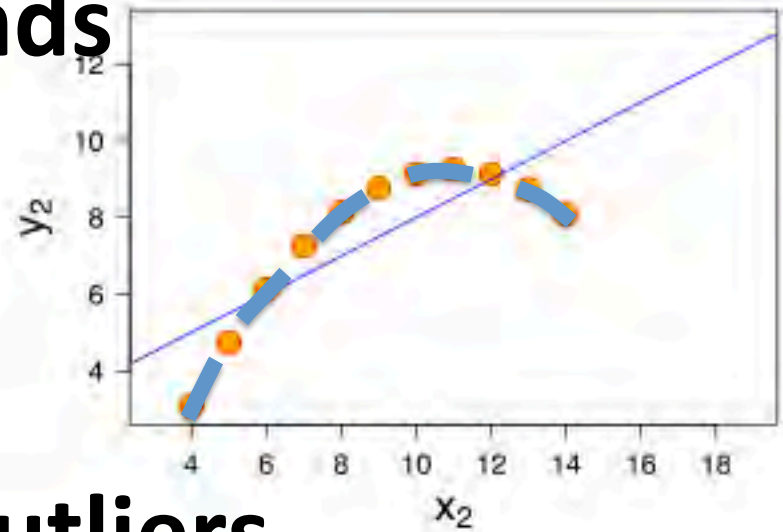
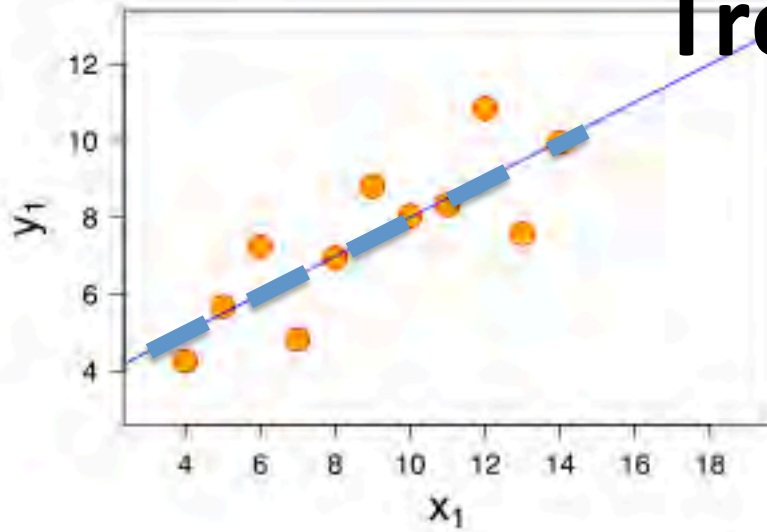
Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

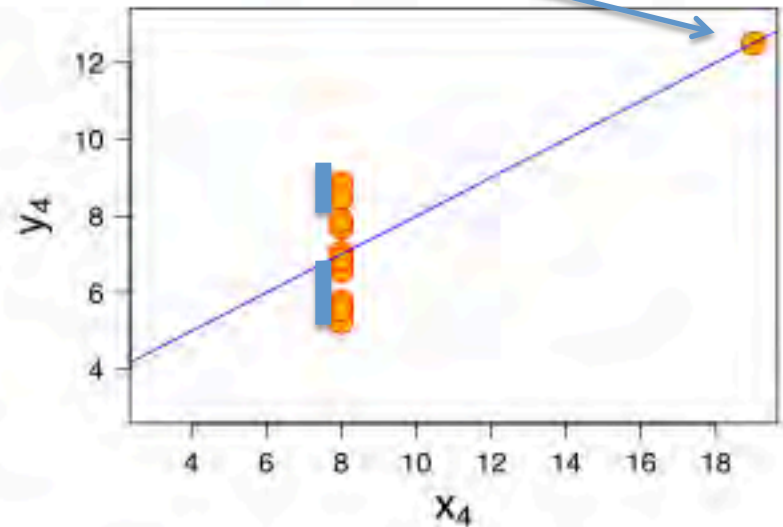
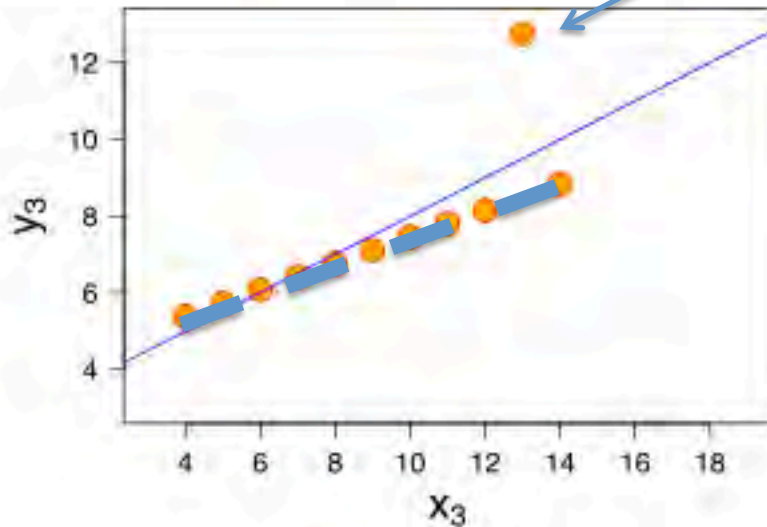
Property	Value
Mean of x in each case	9 (exact)
Sample variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Sample variance of y	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y	0.816 (to 3 decimal places)
Linear regression	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)



Trends



Outliers



Get answers to the questions you didn't ask yet.

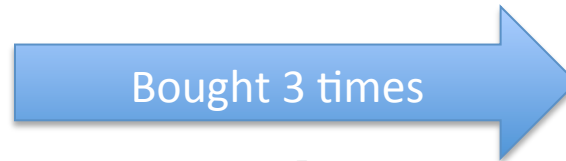


Statistical Modelling

Visual Analytics

E-COMMERCE DATABASE

E-Commerce Data



E-Commerce Data

1



**User
Table**

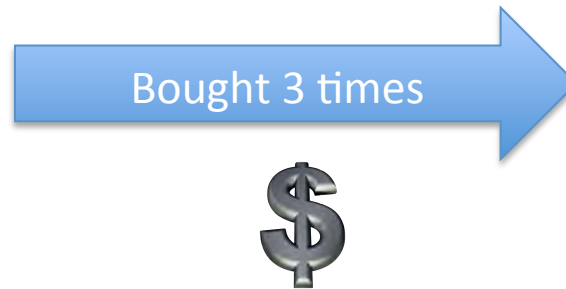
Age
Gender
Zip Code
Job ...

Bought 3 times

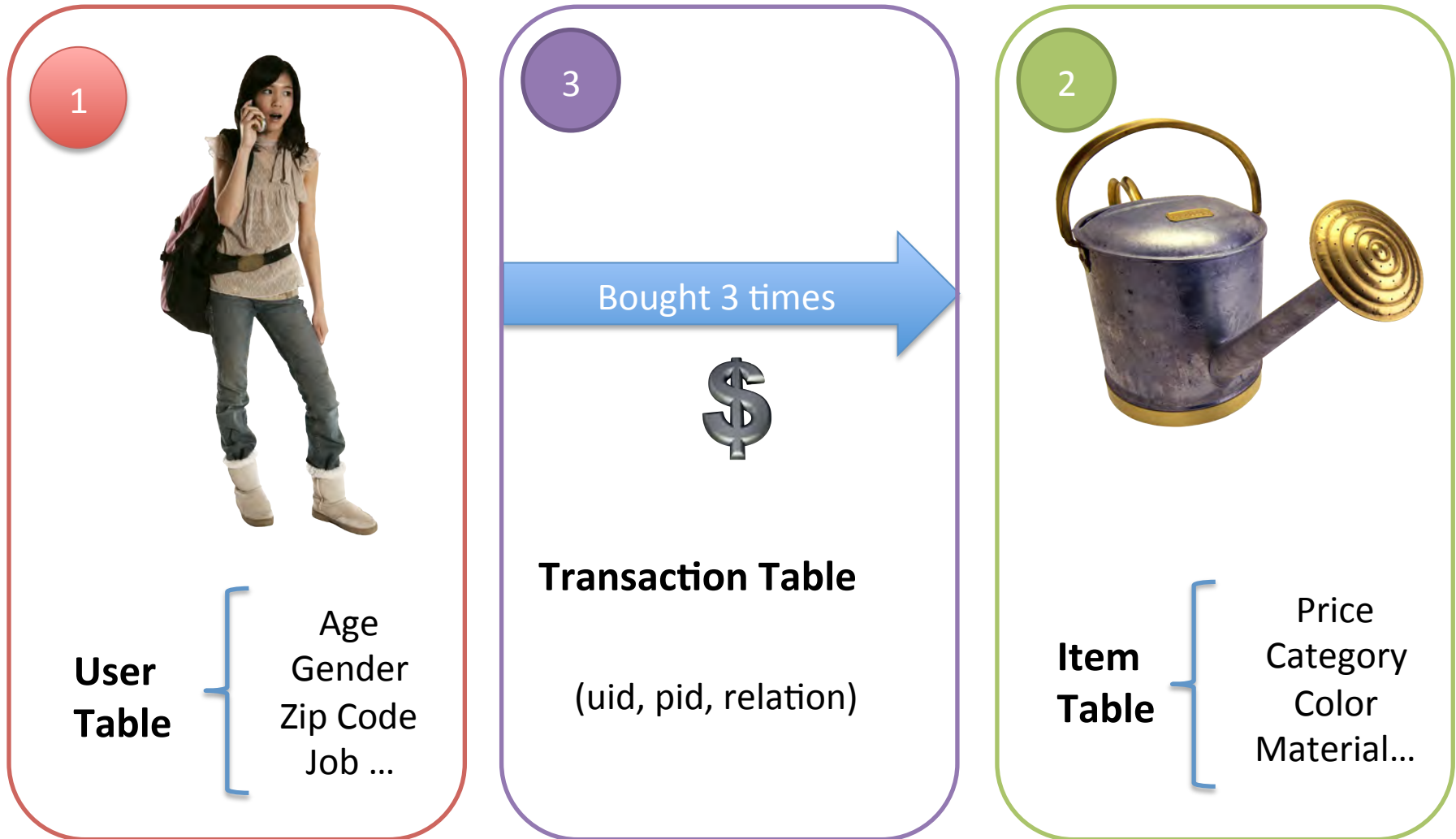


Price
Category
Color
Material...

E-Commerce Data



E-Commerce Data



Sparse Matrix for Dyadic Relation

	item 1	item 2	item 3	item 4	item 5
User A		★	★	★★	★★★★
User B	★★		★★★★	★★★	★★★
User C		★★★		★	
User D	★★				★
User E	★	★★★★	★★★★★	★	★★★★
User F	★★		?	★★★	★★★



In Stock.

Sold by **Eightarc** and **Fulfilled by Amazon** in [easy-to-open packaging](#). Gift-wrap available.

- Xbox 360, Playstation 3, PC compatible out of the box
- Sanwa JLF joystick & 8 Sanwa OBSF-30 buttons
- Xbox 360 headphone jack, headphones
- Retractable carrying handle, USB cord compartment
- Black felt bottom and 4 rubber feet for non-slip competitive play

4 new from **\$159.82**

This item's packaging will indicate what is inside. To cover it, select **Ship in Amazon box** on the checkout page.

Roll over image to zoom in

Item-based Filtering

User-based Filtering

Frequently Bought Together



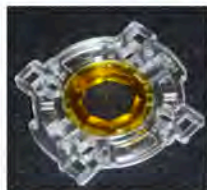
Total price: **\$184.98**

[Add both to Cart](#)

[Add both to List](#)

- ✓ **This item:** Qanba Q4 Q4RAF Black PS3 & Xbox 360 & PC Joystick (Fightstick) **\$169.99**
- ✓ Sanwa GT-Y Octagonal Restrictor Plate for JLF Joysticks **\$14.99**

Customers Who Bought This Item Also Bought



Sanwa GT-Y Octagonal Restrictor Plate for JLF Joysticks
★★★★★ 110
\$14.99 ✓Prime



Qanba Q4 Clear Plexi Overlay for custom artwork
★★★★☆ 5
\$19.99



Ultra Street Fighter IV - Xbox 360
Capcom
★★★★★ 61
Xbox 360
\$17.48 ✓Prime



Mayflash PS2 PS3 PC USB Universal Arcade Fighting Stick
Mayflash
★★★★☆ 277
PlayStation 3, PlayStation2
\$44.95 ✓Prime



Qanba Q1 Xbox 360 and PC Joystick (Fightstick)
★★★★★ 74
\$79.99 ✓Prime



Qanba Q4 Q4RAF White Red PS3 & Xbox 360 & PC Joystick (Fightstick)
★★★★★ 27
\$159.82



Qanba Q2 Glow LED XBOX 360 & PC Joystick (Fightstick)
★★★★★ 22
\$154.99 ✓Prime



Black Sony PS2 to Xbox 360 Controller Adapter Converter Cable
★★★★★ 23
\$19.88

Computers are good for
individualized recommendations.

But ...

What items should be on holiday promotion?

Who are our target user groups?

What are their favorite items?

Overview Questions



Which user group especially hates our product?

Which product shows interesting preference pattern?

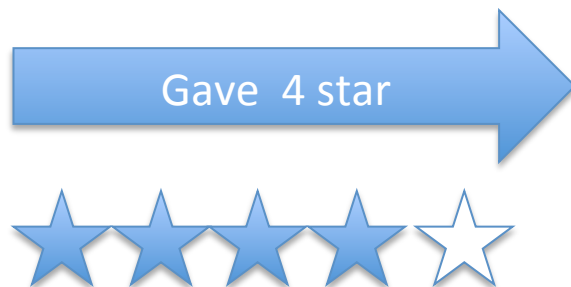
Outlier Questions



How e-commerce dataset can be
visualized?

Dataset

- Movielens dataset
 - GroupLens Research Project at the University of Minnesota.
 - 100,000 ratings (1-5) from 943 users on 1682 movies.
 - Each user has rated at least 20 movies.
 - Simple demographic info for the users (age, gender, occupation, zip)



	User E
Age	32
Job	Writer
Gender	Femaie
Zip Code	47906
Average Review	3.4
Number of Reviews	57

	User A	User B	User C	User D	User E
Movie 1		1	1		4
Movie 2	4	2	3	3	5
Movie 3	1			1	
Movie 4	2	4		1	1
Movie 5	3	3	2		4

	Movie 1
Release Date	1978
Average Review	4.3
Number of Reviews	345

Task Taxonomy

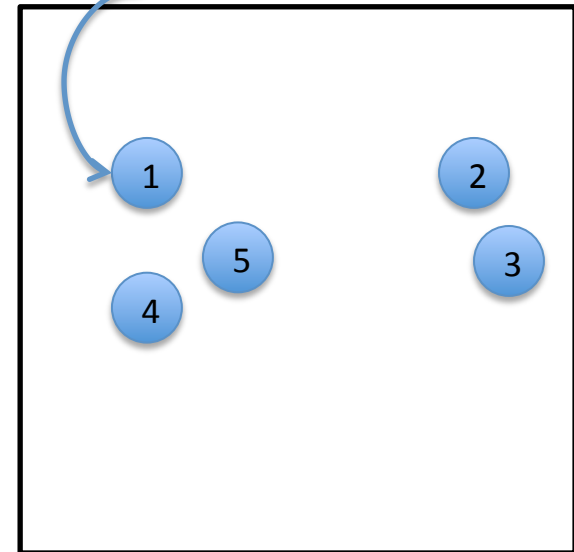
- Search and filter
- Data Distribution
- Finding similar entities
 - Based on properties
 - Based on the dyadic pattern
- Finding similar linked entities
- Estimate correlation

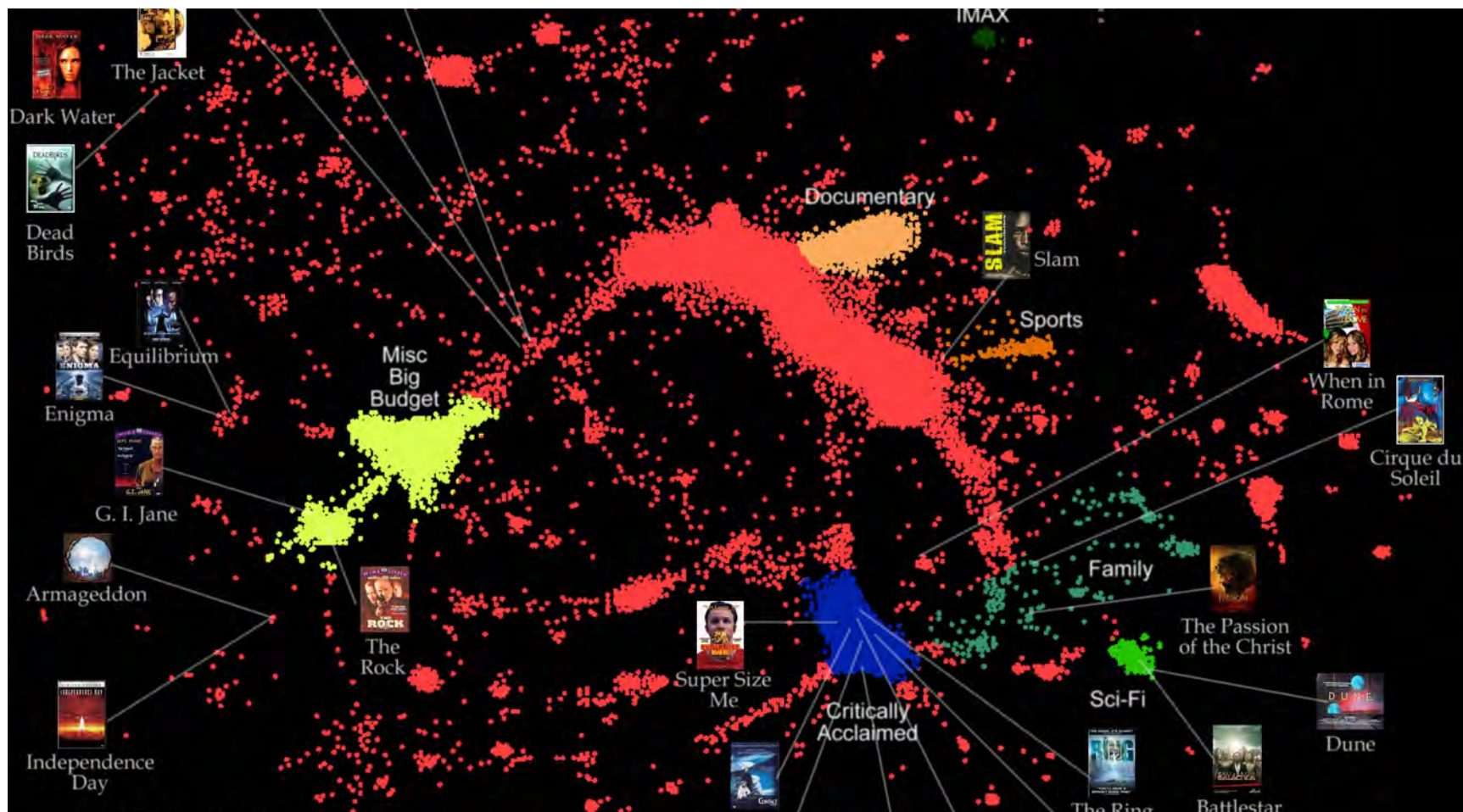
METHODS

Dimension Reduction

Multi dimensional scaling (MDS)
Independent component analysis (ICA).
Principal component analysis (PCA).
Singular value decomposition (SVD).
T-SNE

	Item 1	Item 2	Item 3	Item 4	Item 5
User A		1	1	4	4
User B	4		3	3	5
User C		5		1	
User D	2		5		1
User E	3	3	2	2	4

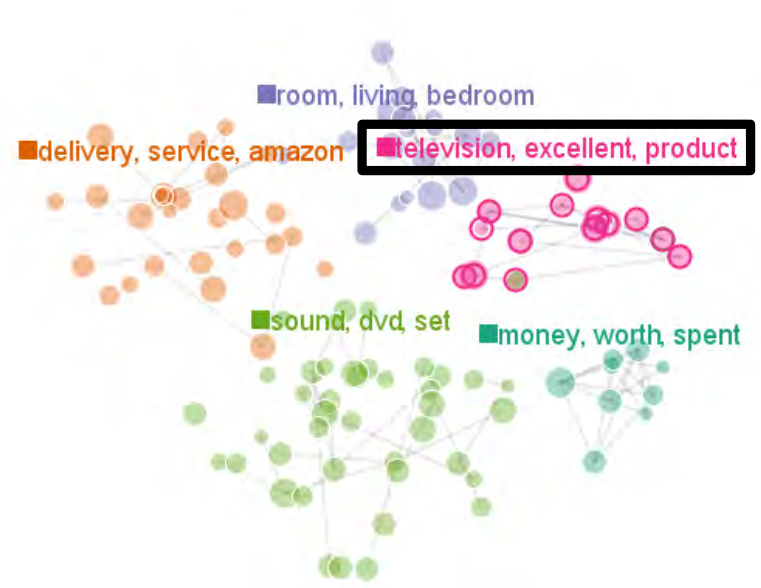




Top-K words



Galaxy view by INSPIRE

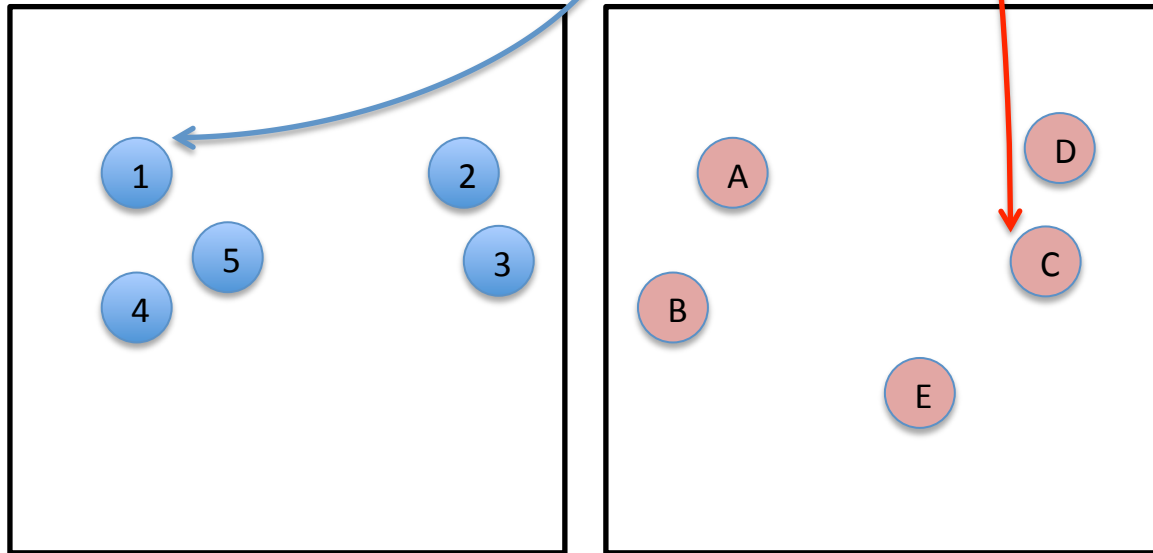


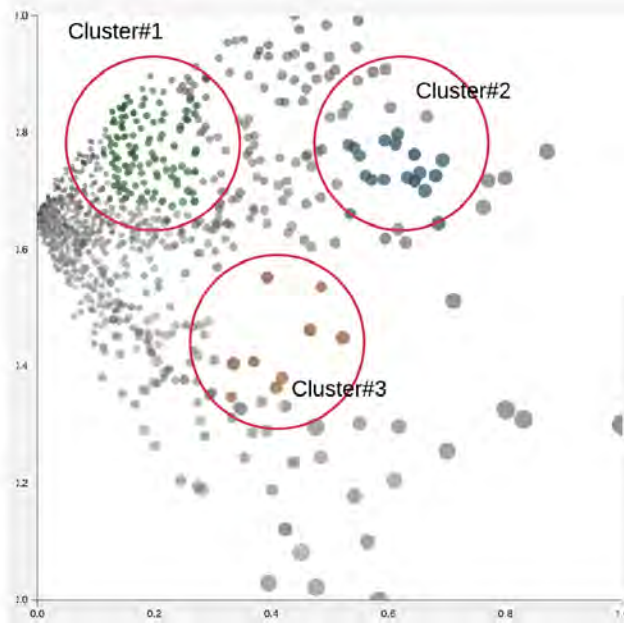
T-SNE by UTOPIAN

Two Ideas

- Use of Complementary space to identify
- Use of attribute features to identify the characteristics of cluster

	Item 1	Item 2	Item 3	Item 4	Item 5
User A		1	1		4
User B	4	2	3	3	5
User C	1			1	
User D	2	4		1	1
User E	3	3	2		4

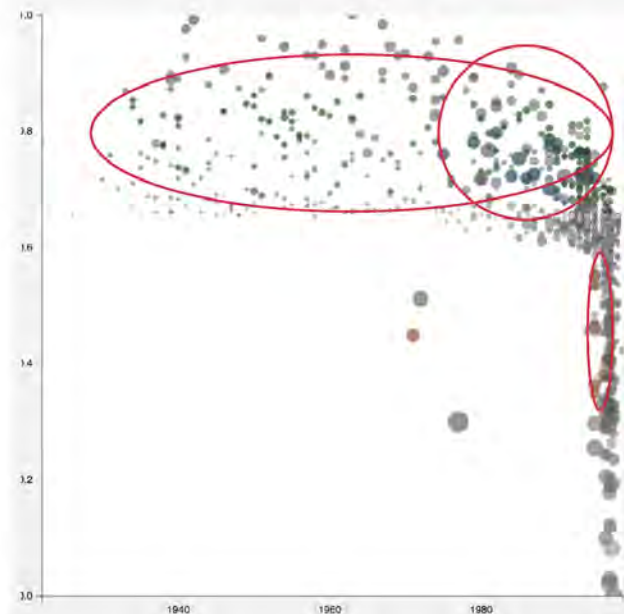




Let's say this is reduced dimension
display of multi-dimensional
dataset



When you apply single
concept vector on x-axis

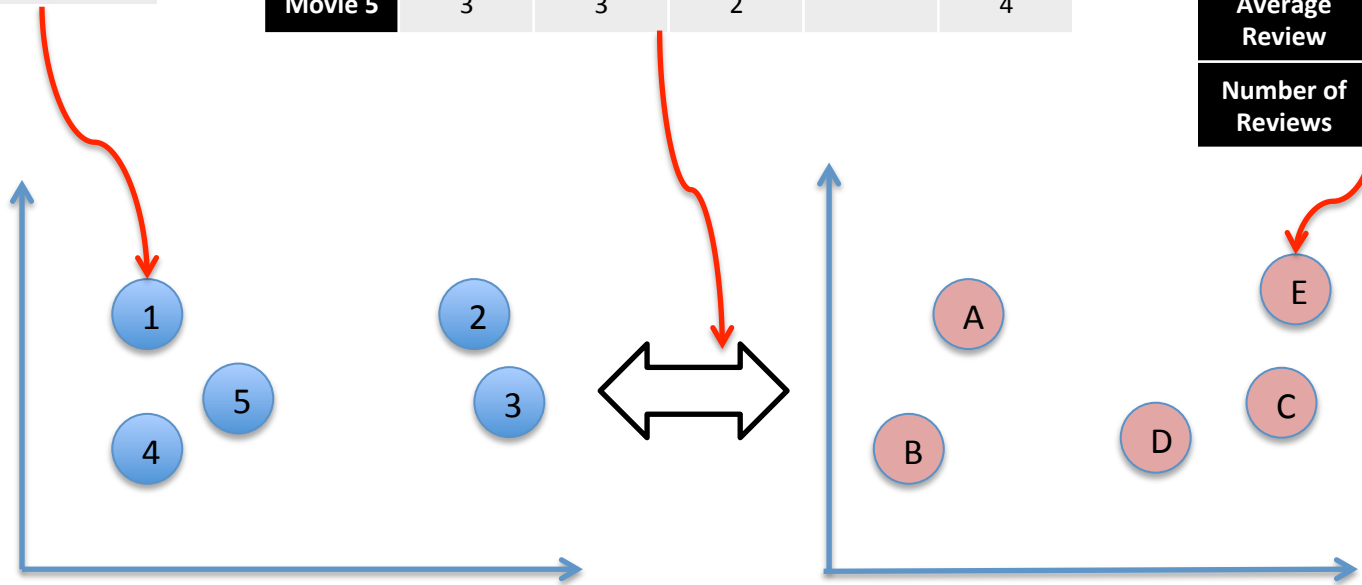


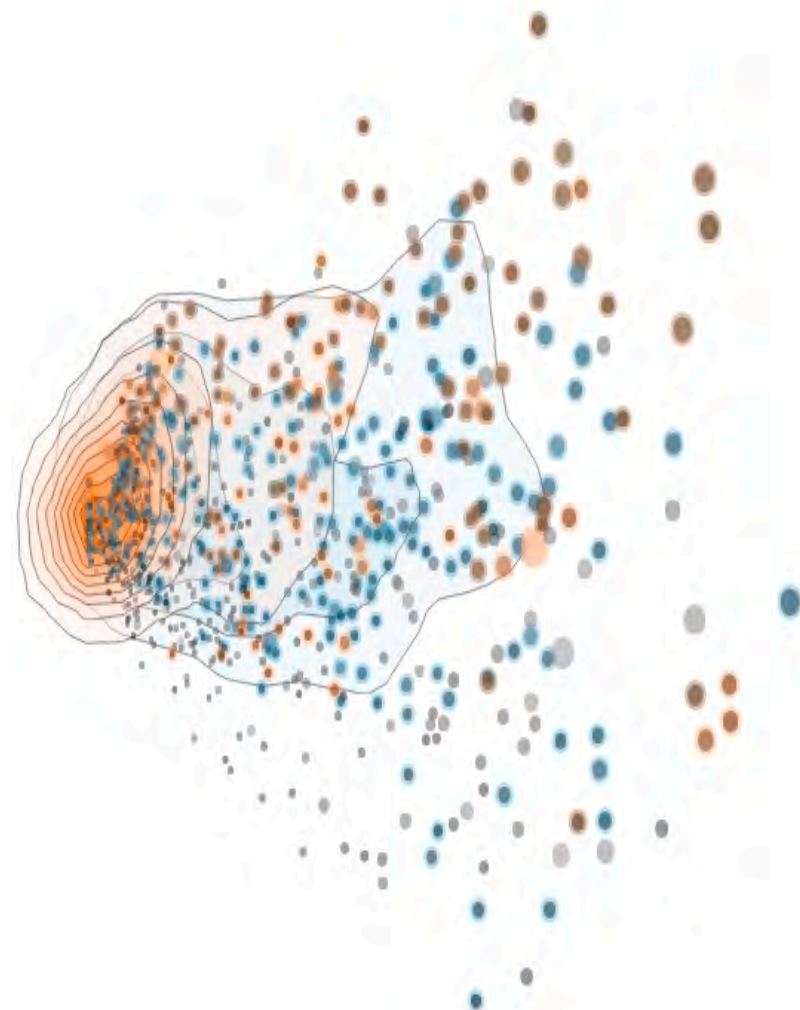
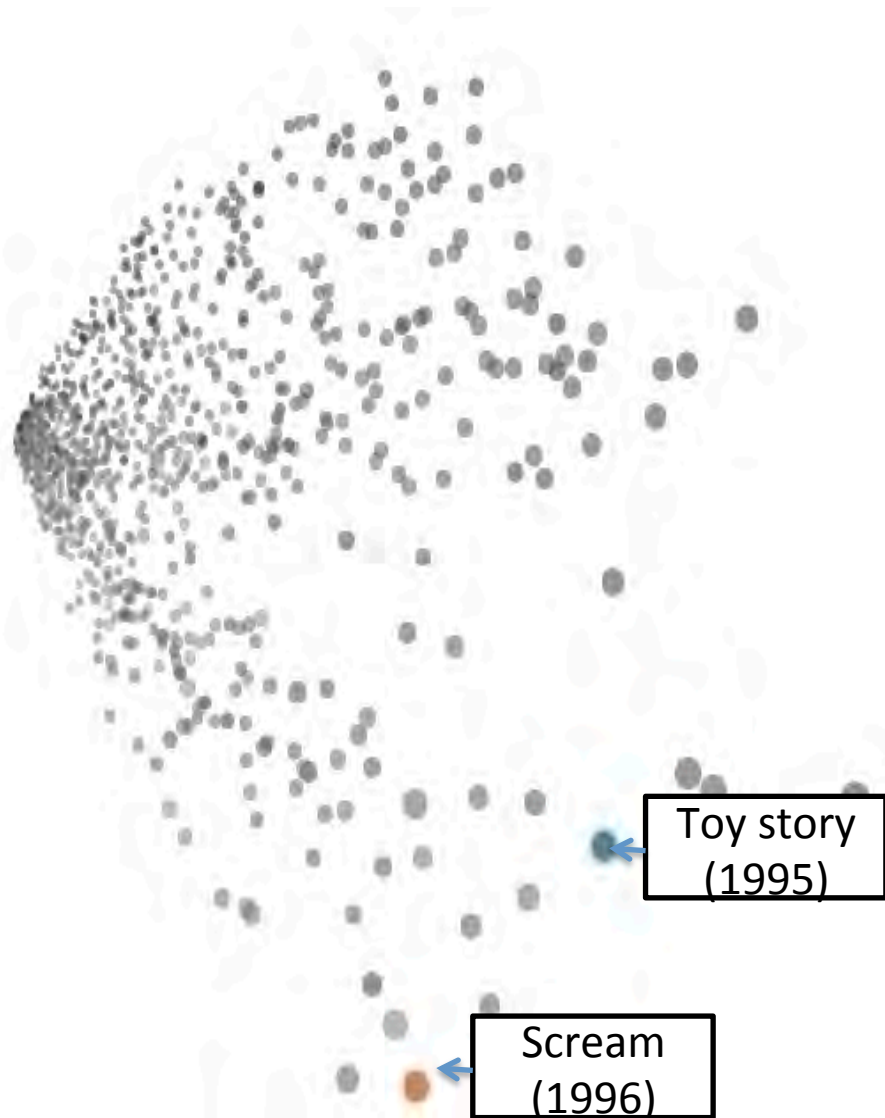
Cluster#1(Green dots) are dispersed, while
Cluster#3 is highly packed. It means that the
cluster#3 has a strong bias on the concept,
while cluster#1 is irrelevant.

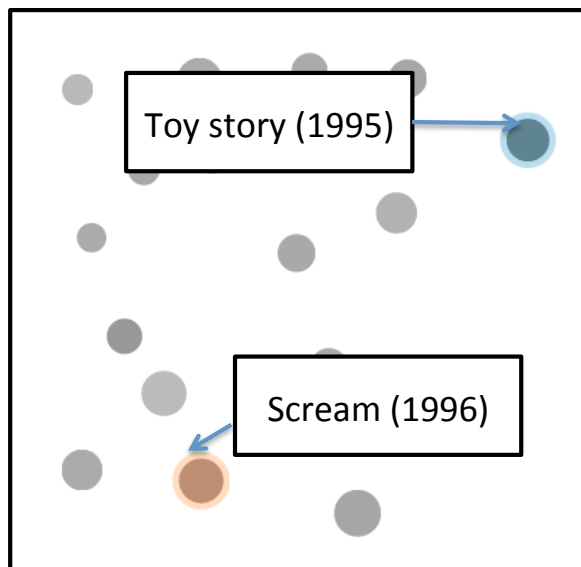
	Movie 1
Release Date	1978
Average Review	4.3
Number of Reviews	345

	User A	User B	User C	User D	User E
Movie 1		1	1		4
Movie 2	4	2	3	3	5
Movie 3	1			1	
Movie 4	2	4		1	1
Movie 5	3	3	2		4

	User E
Age	32
Job	Writer
Gender	Femaie
Zip Code	47906
Average Review	3.4
Number of Reviews	57

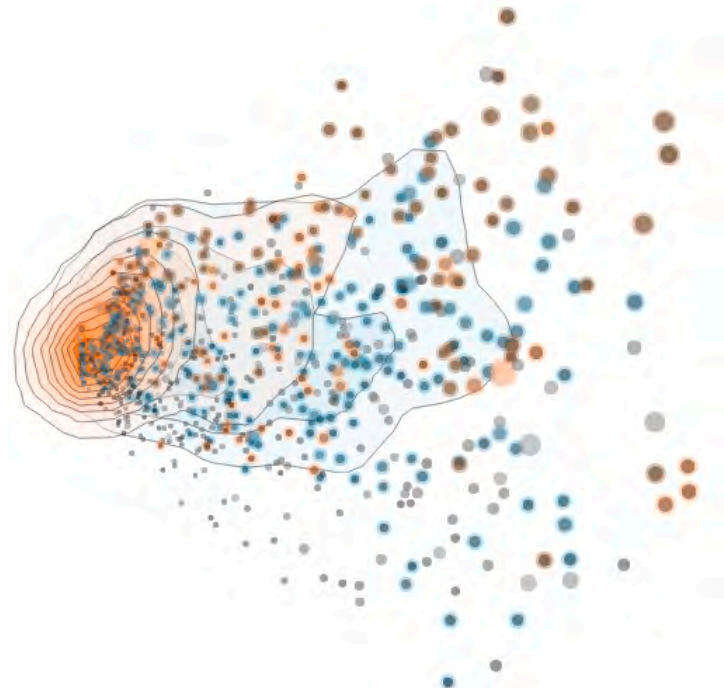






Similarity

A thick blue arrow pointing from the initial data box to the similarity plot.



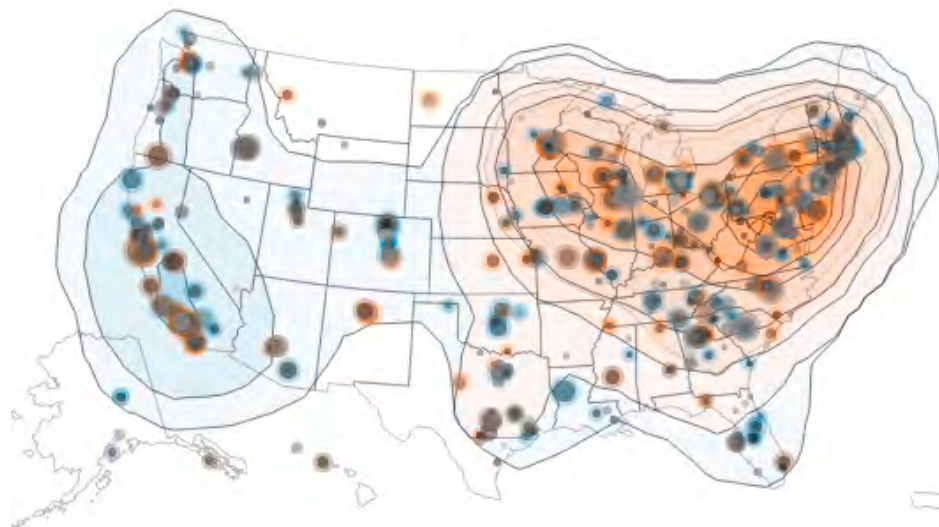
Age

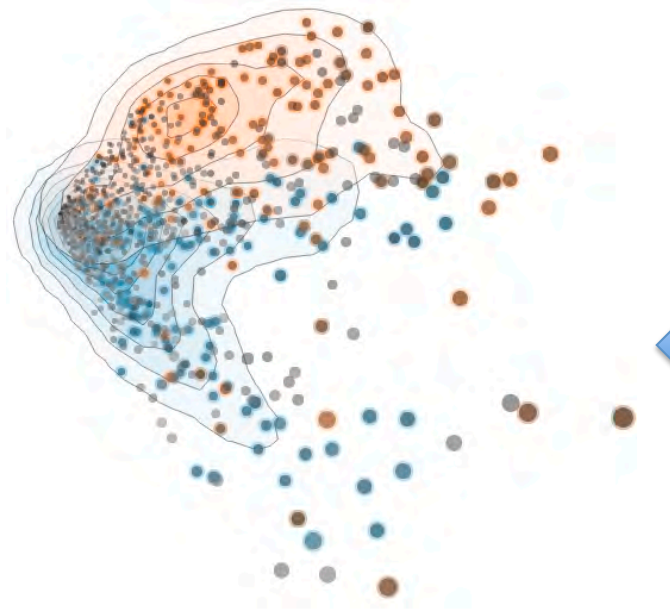
A thick blue arrow pointing from the initial data box to the age plot.



Location

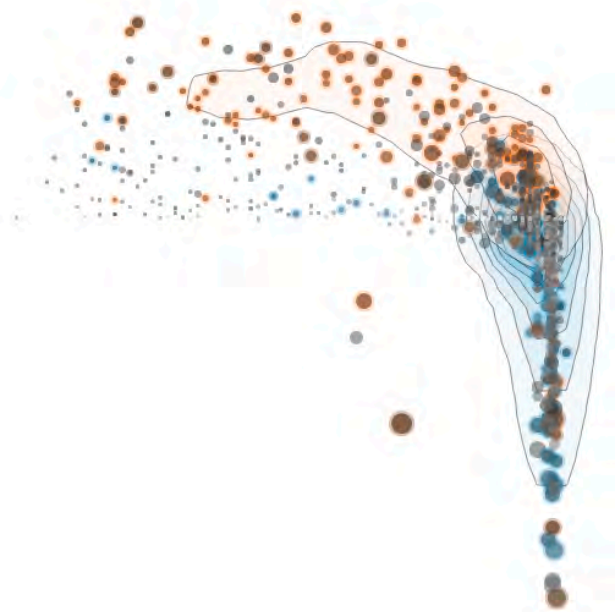
A thick blue arrow pointing from the initial data box to the location map.





Similarity

Release Date



19, M, Student

51, M, Educator

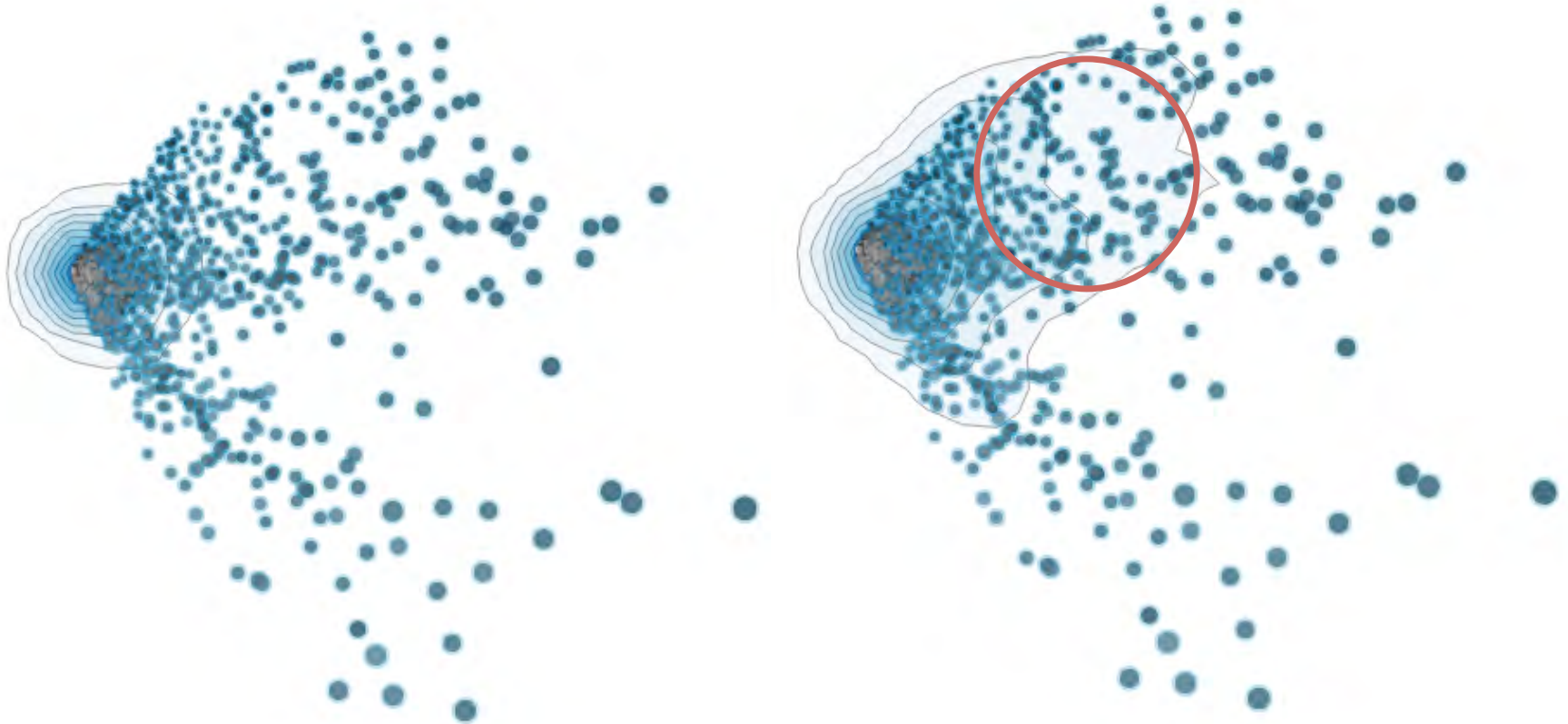
Average Review



Kernel Density Estimation



Density vs. Amplitude



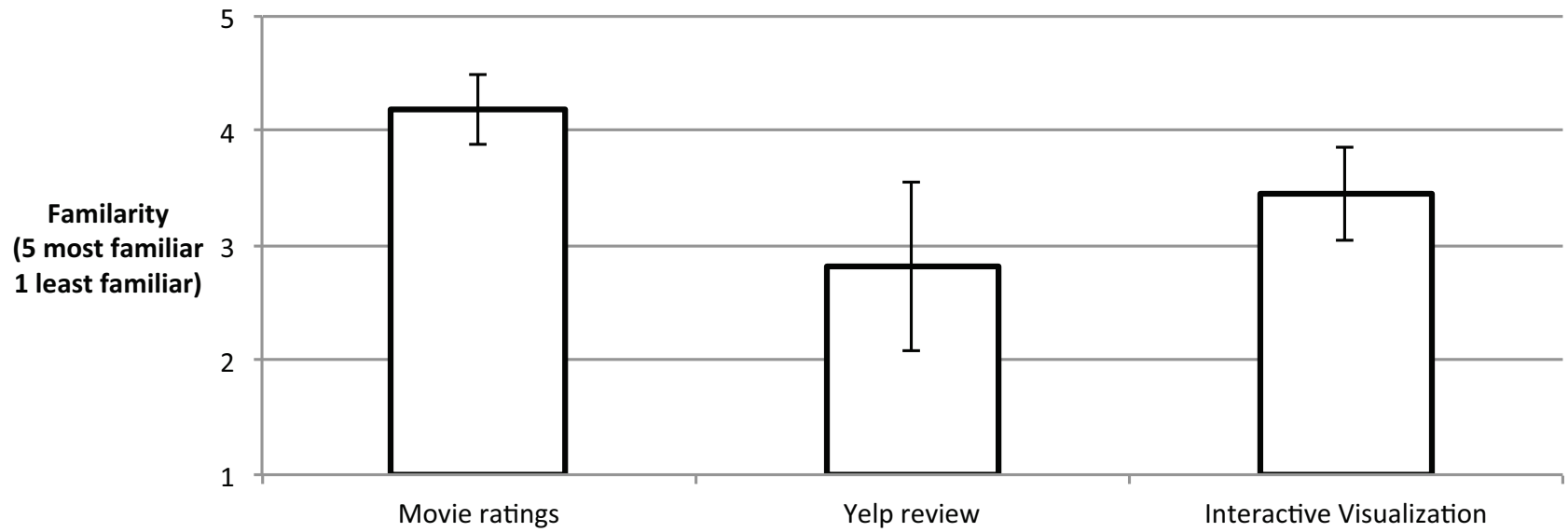
EVALUATION

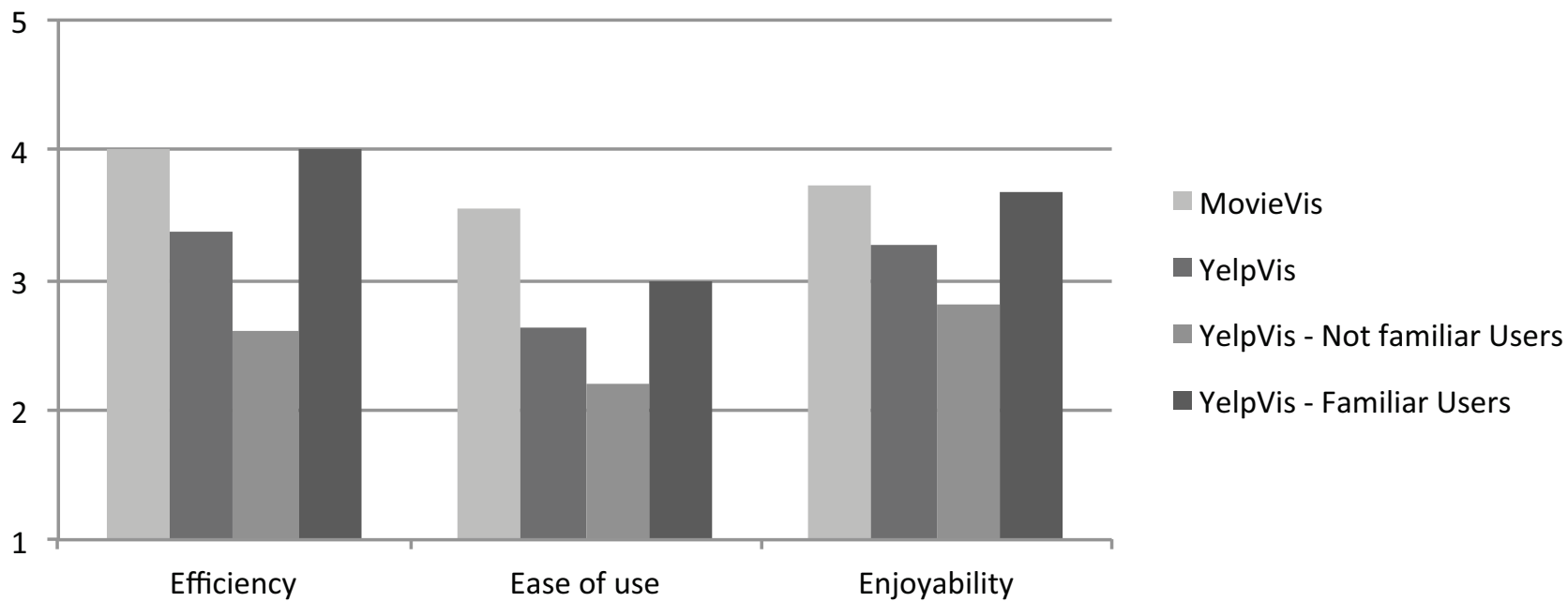
Two dataset

- Movie Lens
- Yelp Review text

Exploratory Analysis

- Get answers to questions to you didn't ask.
- 11 participants
- Explore movie dataset and find hypothesis
 - Average 6.5 (sd 4.8) comments per users





Testimonial

- *So Matilda and Contact are both good movies and both liked by a lot of people from all ages, but they have a 'far' similarity because Contact has way more reviews than Matilda and [is] closer to movies [...] like Star Wars...*
- *as a geek, I would like to play with this, but it is not for non-geeks.*

DISCUSSION

Limitation and Future works

- Occlusions
- Limited attributes
- Difficulty in group behavior visualization
 - Union vs Common
- Temporal Dimension

Conclusion

- Two novel methods to identify cluster characteristics
 - Use of feature and dimension reduction as an axis
 - Complementary entity for a brush and linking
- Web-based implementation with evaluation

Questions?

Deok Gun Park

intuinno@umd.edu

<http://vistalk.herokuapp.com/movievis/>

<http://vistalk.herokuapp.com/yelpvis/>