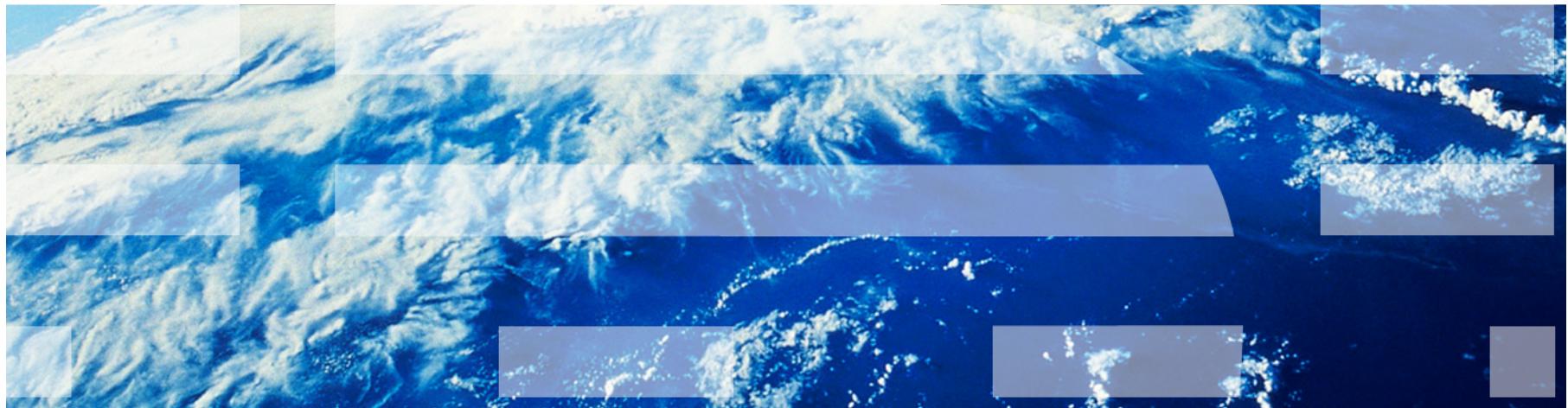


E6893 Big Data Analytics Lecture 7:

Big Data Visualization — I

Ching-Yung Lin, Ph.D.

Adjunct Professor, Dept. of Electrical Engineering and Computer Science



October 20th, 2023

Part I : Introduction

What is visualization ?

Why do we create visualisation ?

Existing Visualisation Techniques

Part II : Big Data Visualization

Challenges

Techniques

Part III : How can we visualize big data

Key techniques

Open source tools

Examples

Part IV : Visual Analysis of Big Data

Thanks to Dr. Nan Cao <http://nancao.org>

Part I : Introduction

What is visualization?

How can we acquire information?

Listen



Taste & Smell

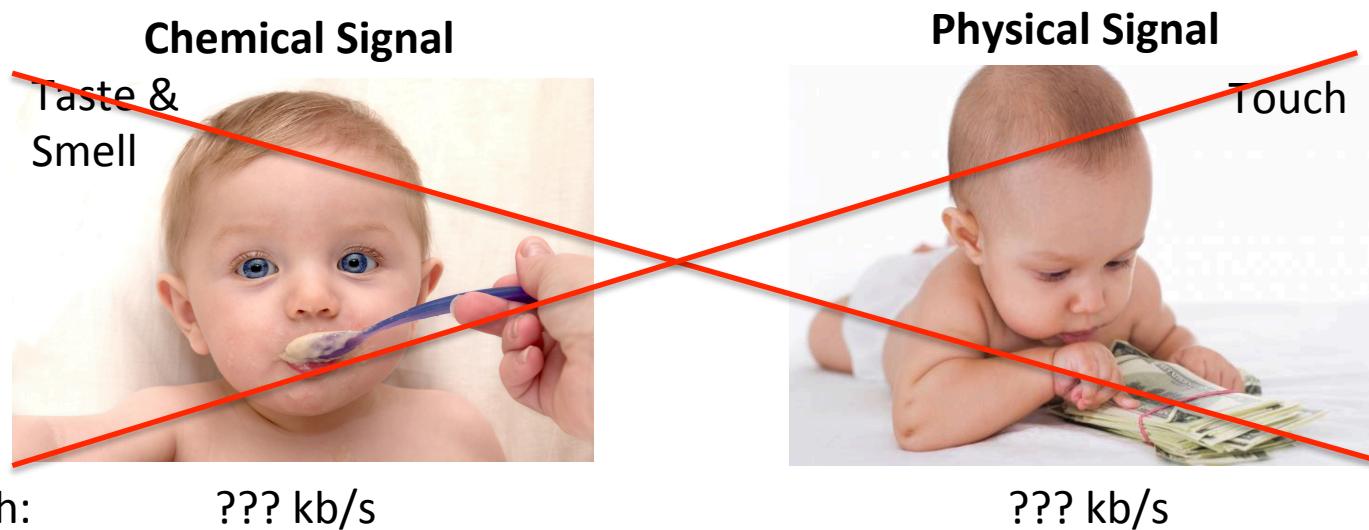


Touch



Look





**Can not be estimated by
information theory**

Sound Signal

Listen



Electronic / Light Signal

Look

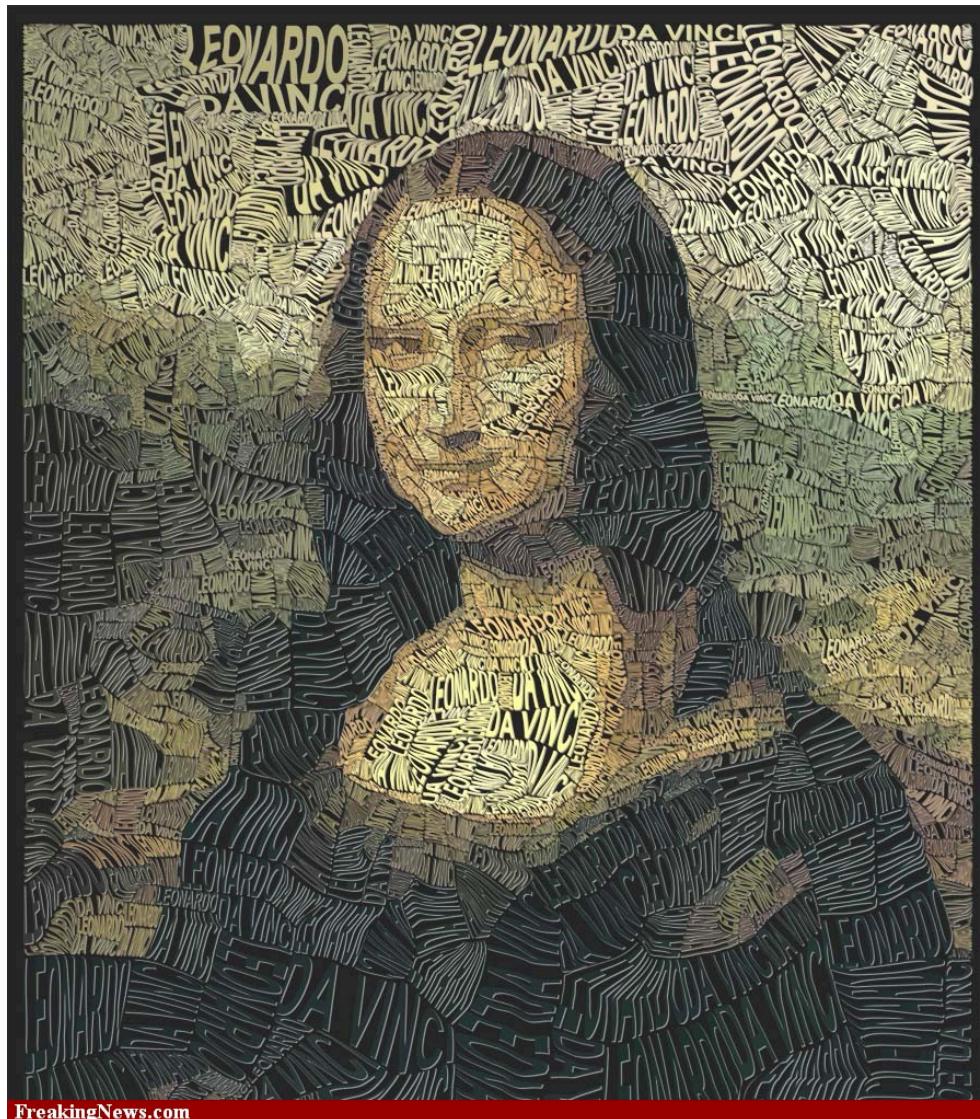


Bandwidth: **about 0.1 KB / s**

> 100 MB/s

“Information Visualization, Perception for Design” 3rd Edition, by Colin Ware

A picture worth 1000 Words

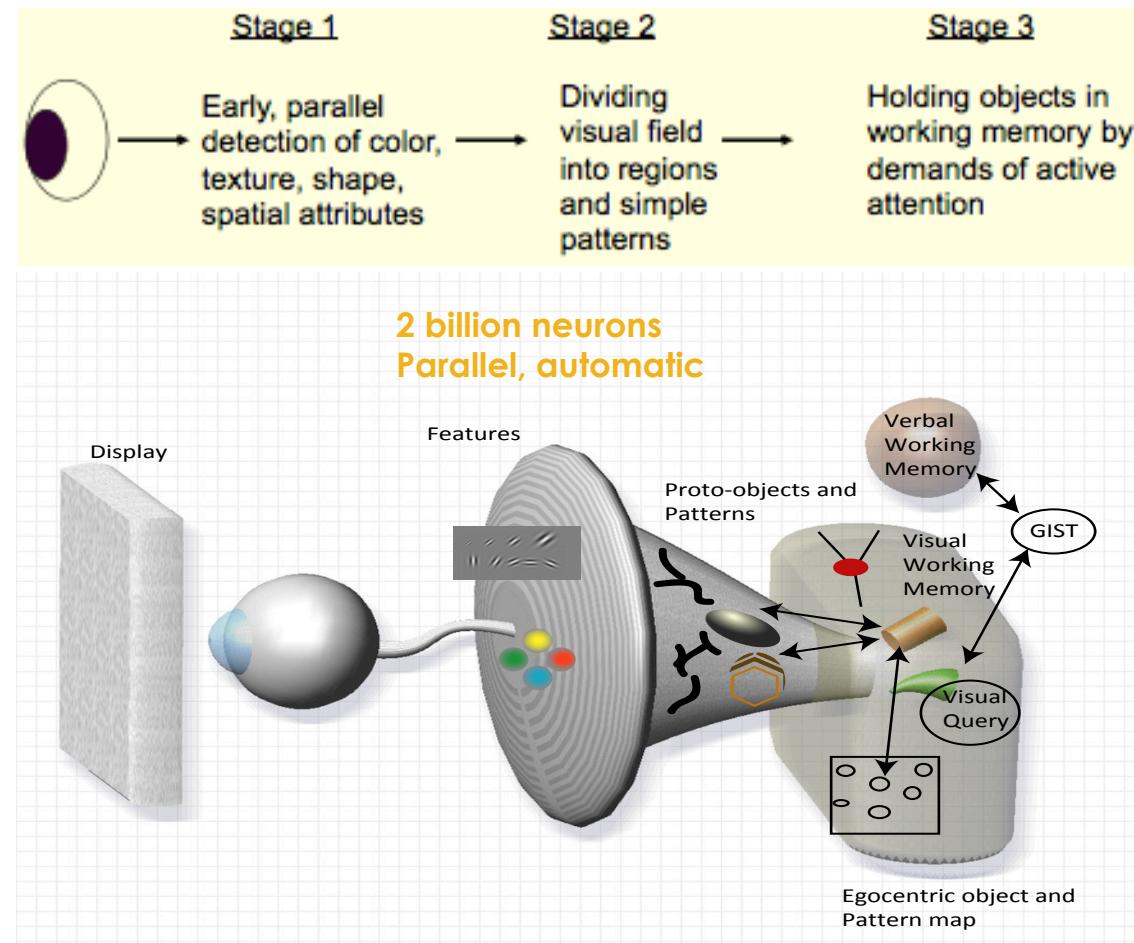


Why efficient ?

The Visual Thinking Pipeline

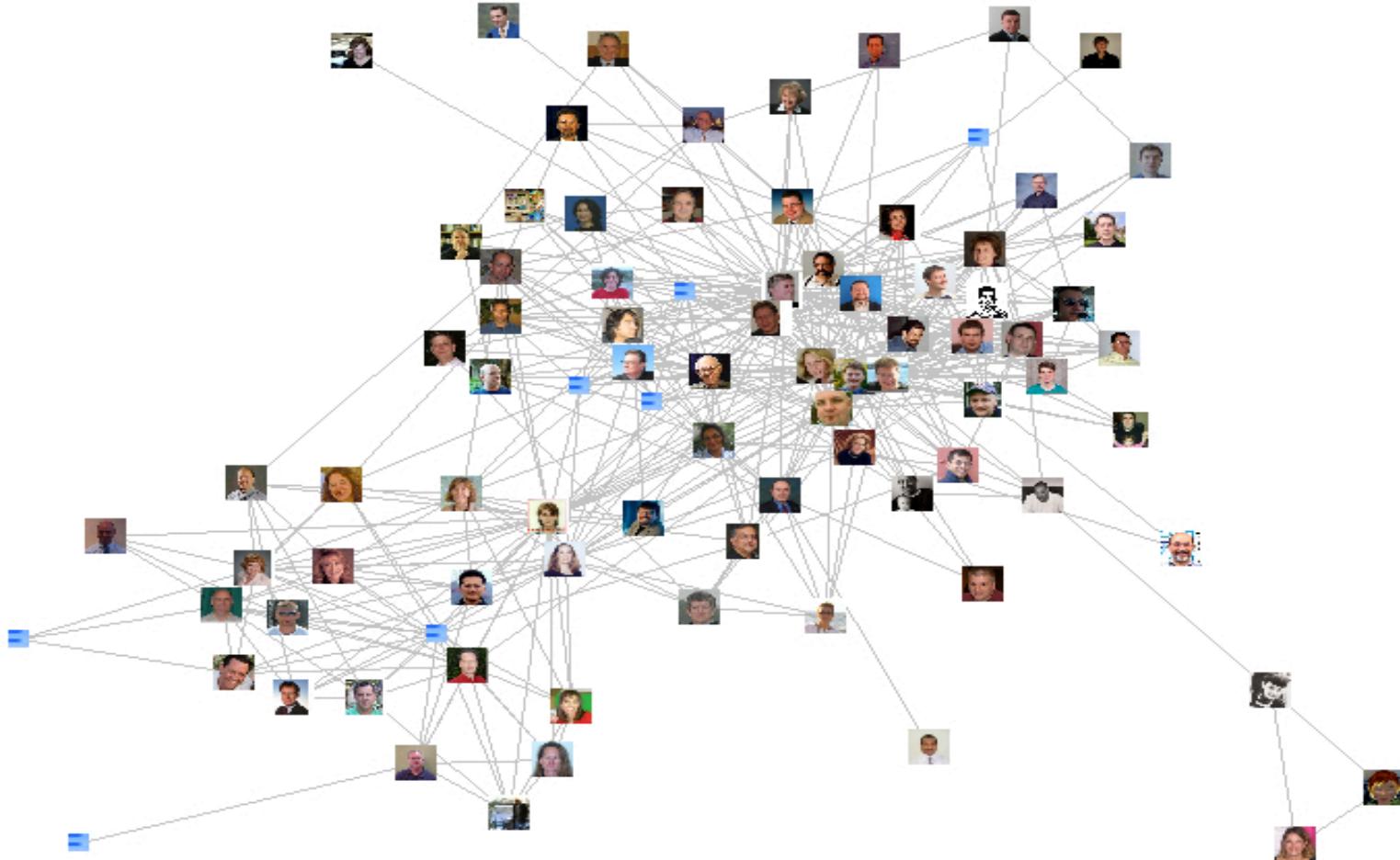
Parallel Processing to Extract Low-Level Visual Properties such as color, shape, etc

Sequential Goal-Oriented Processing



Example





Ching-Yung Lin, Nan Cao, Shixia Liu, Spiros Papadimitriou, Jimeng Sun, and Xifeng Yan. SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence. ICDE 2009

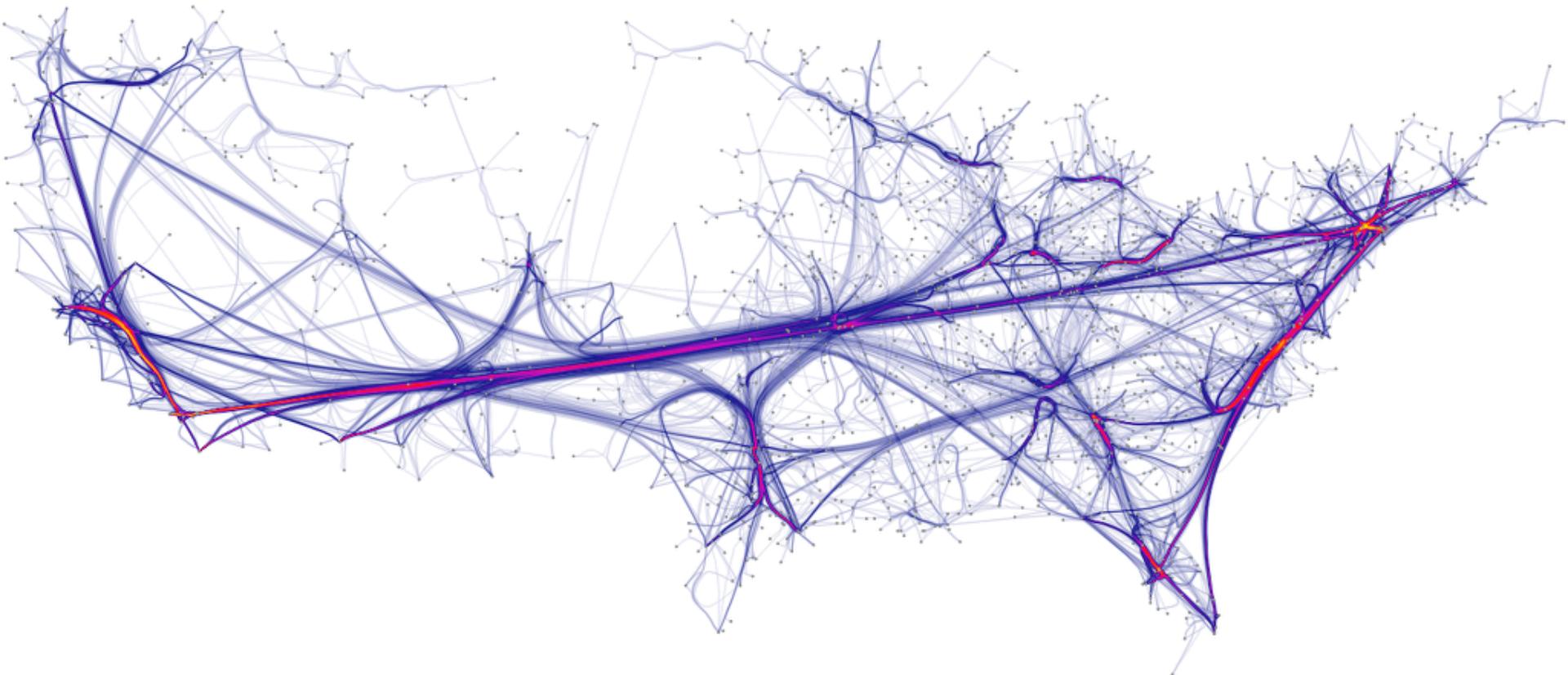
Record of human activities to help find data patterns

Example



Visualization is used for help reasoning and decision making

Summarization of Airlines in United States



Holten, Danny, and Jarke J. Van Wijk. "Force-Directed Edge Bundling for Graph Visualization." Computer Graphics Forum. Vol. 28. No. 3. Blackwell Publishing Ltd, 2009.

What is Information Visualization ?

“The action or fact of visualizing; the power or process of forming a mental picture or vision of something not actually present to the sight; a picture thus formed.”

-- Oxford English Dictionary

“... finding the artificial memory that best supports our natural means of perception.”

-- Bertin, 1983

The use of computer-supported, interactive, visual representations of abstract data to amplify cognition

-- Cart, Mackinlay, Shneiderman, 1999

Why do we create visualization?

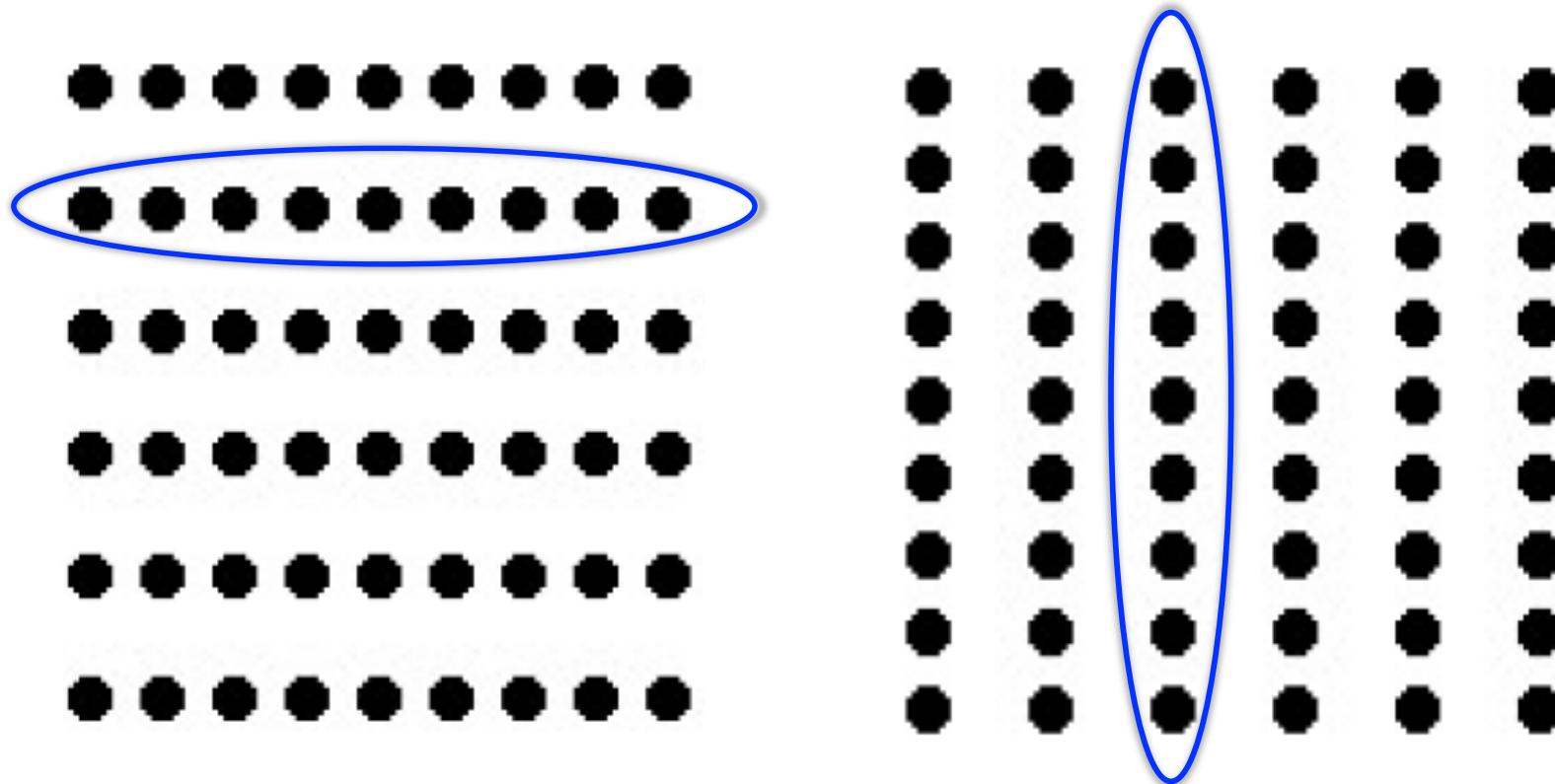
Counting the number of 3s in the
following Text:

1235693234870452973467
0378937043679709102539

Counting the number of 3s in the following Text:

1235693234870452973467
0378937043679709102539

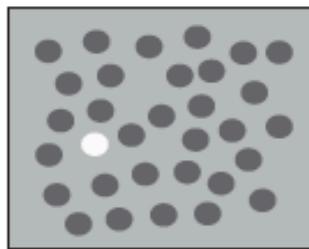
Can you identify the groups of dots in the following figures ?



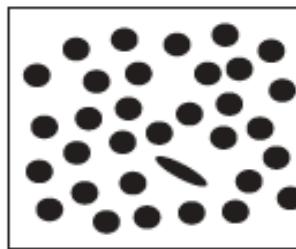
Law of Proximity

we tend to group elements that are closest to each other

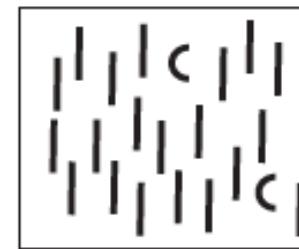
Find Patterns: Pre-Attentive Visual Channels



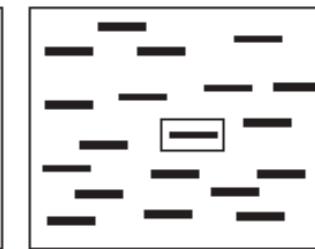
Grey value



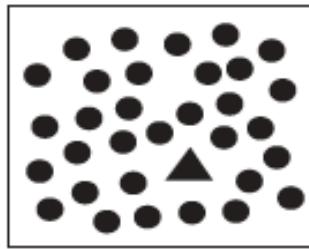
Elongation



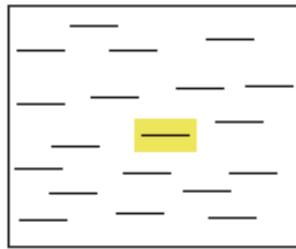
Curvature



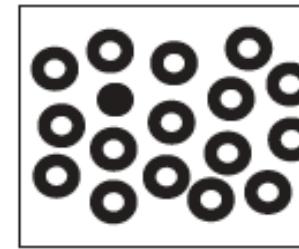
Added surround box



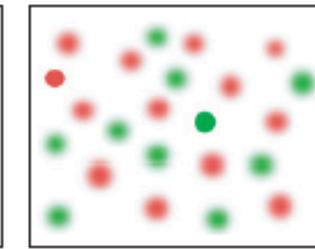
Shape



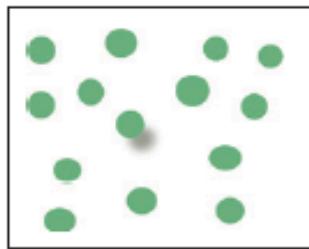
Added surround color



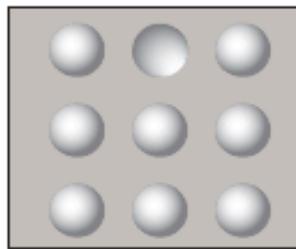
Filled



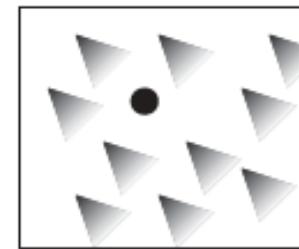
Sharpness



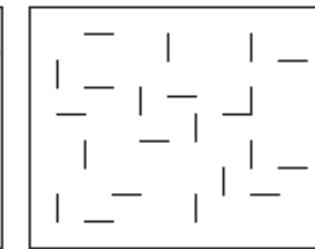
Cast shadow



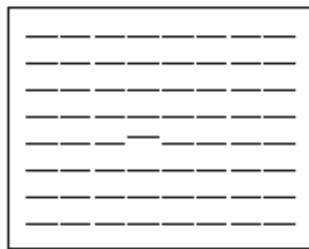
Convex and concave



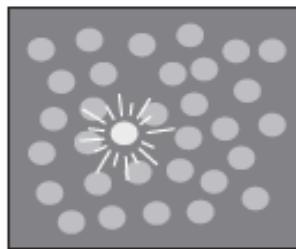
Sharp vertex



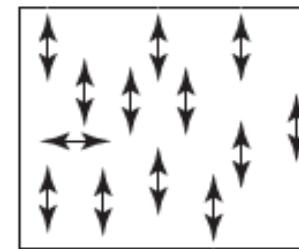
Joined lines



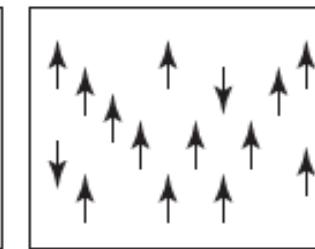
Misalignment



Blinking



Direction of motion

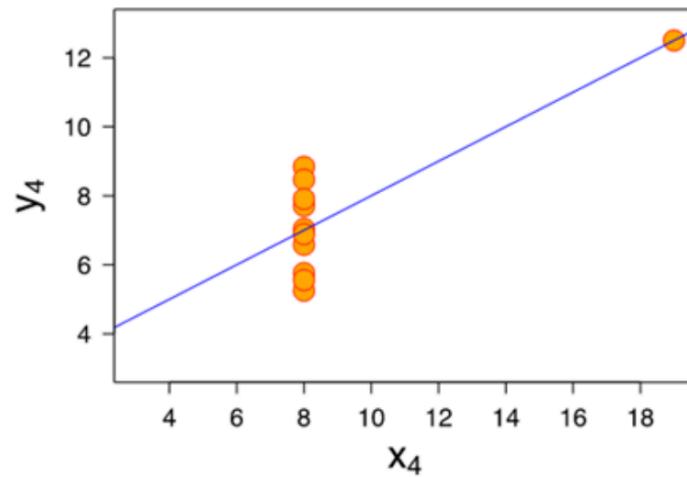
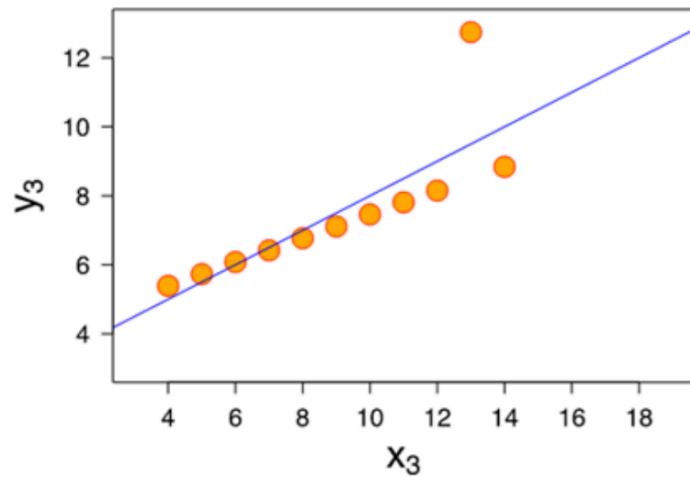
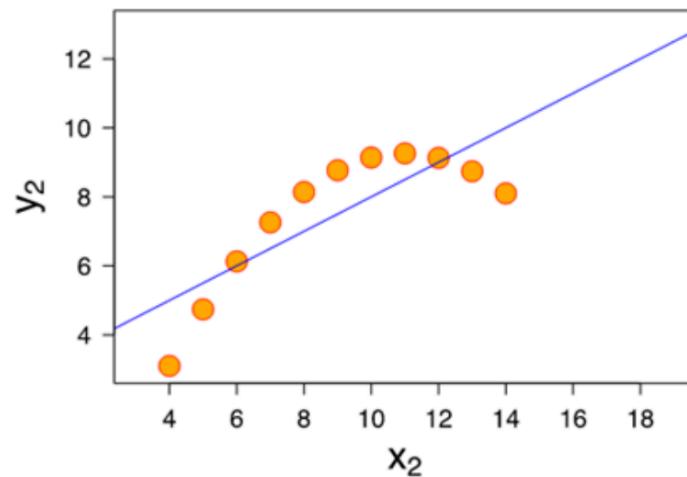
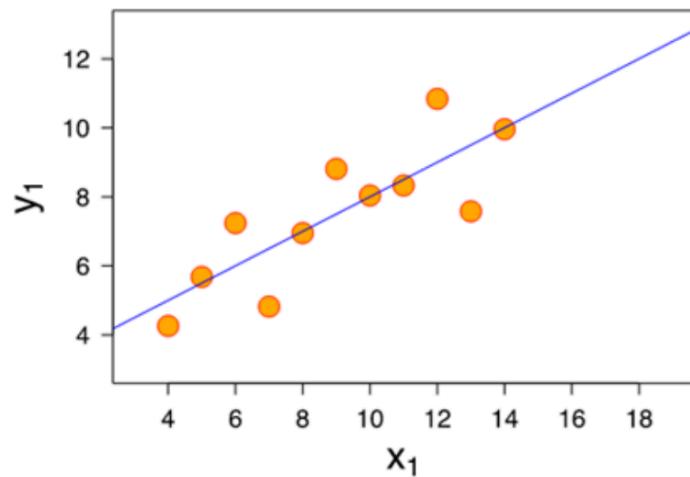


Phase of motion

Why do we create Visualization?

	Set A		Set B		Set C		Set D	
	X	Y	X	Y	X	Y	X	Y
0	10	8.04	10	9.14	10	7.46	8	6.58
1	8	6.95	8	8.14	8	6.77	8	5.76
2	13	7.58	13	8.74	13	12.74	8	7.71
3	9	8.81	9	8.77	9	7.11	8	8.84
4	11	8.33	11	9.26	11	7.81	8	8.47
5	14	9.96	14	8.10	14	8.84	8	7.04
6	6	7.24	6	6.13	6	6.08	8	5.25
7	4	4.26	4	3.10	4	5.39	19	12.50
8	12	10.84	12	9.13	12	8.15	8	5.56
9	7	4.82	7	7.26	7	6.42	8	7.91
10	5	5.68	5	4.74	5	5.73	8	6.89
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
std	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
corr	0.82		0.82		0.82		0.82	
lin. reg.	$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$	

Seeing data in context



Why do we create visualization ?

A picture is worth a thousand words

News illustrated



GANGNAM STYLE!!!
The 5 basic steps

The sudden explosion of a South Korean entertainer called Psy, has given the world Gangnam Style. It is setting the music and dance world on fire and has a set sequence. We simplify them for your perusal

Step 1
Riding the horse

Dress classy and dance cheesy

Cross your hands like taking the horse reins and pulse up and down

Do small jumps with your legs spread like you are riding a horse

Footsteps: R L R R L R L L

Step 2
Lassoing the sexy lady

Lassoing motion with your right arm

Continue with the horse-riding movement

Footsteps: R L R R L R L L

Step 3
Now everybody is looking at me

A: Hands in pockets or waist and small hip side movements combined with the foot steps

B: Finish this move dragging the right leg to the left leg.

Footsteps A: L R L R L R L R

Footsteps B: L R R L

Step 4
Combine a few 'sexy' moves

C: Now move your hips to the outside with quick movements twice

A: Cross your hands over your left leg

B: Spread and flex your legs. Move your body up and down three times

D: Rotate your right leg three times, maintaining the body position.

Footsteps: R L R R L R L L

Step 5
Finish with a cool pose

A: "L" shape with your fingers

B: Spread your arms and raise your right leg (position A). Now get down quickly on your right leg and flex the left one. Now rotate your right arm and with your hand touch your chin doing a "L" shape with your thumb and index fingers (position B)

★ When to use the steps during the chorus ★

Step 1 Oppa is Gangnam style, ahhhhh...	Step 2 Gangnam style... Oh, oh oh oh, Oppa is Gangnam style...	Step 3 or Step 1 (in the last chorus) Gangnam style... Oh, oh oh oh, Oppa is Gangnam style... Eeeeh- Sexy Lady...
Step 1 or Step 2 (in the last chorus) Oppa is Gangnam style... Eeeeh- Sexy Lady oh oh oh oh. Oppa is Gangnam style.	Step 4	Step 5 (only at the end)

Source: YouTube

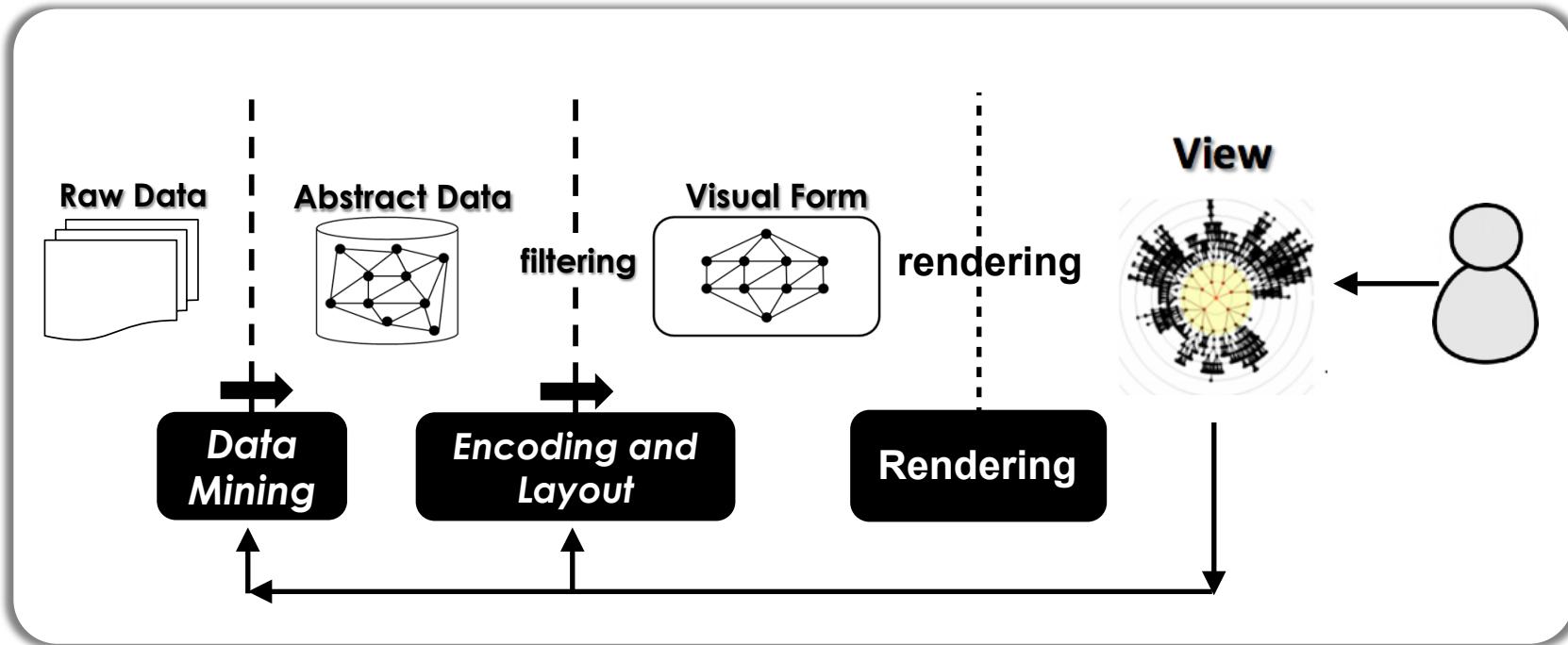
HUGO A. SANCHEZ©Gulf News

A better communication method

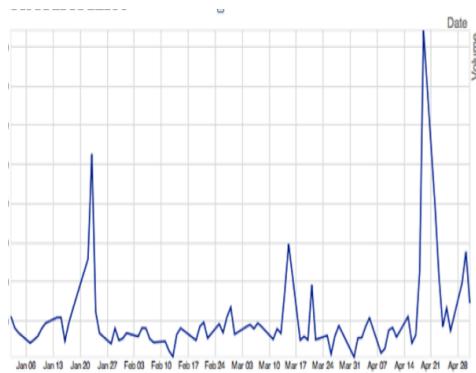
See data in context
Find patterns
Telling a story
Attract attentions
Communicate information with others
Summarization and interpretation
Graphical calculation
Expend memory
Inspire people

Existing Visualisation Techniques

Visualization & Visual Analysis Reference Model



Taxonomy by data types



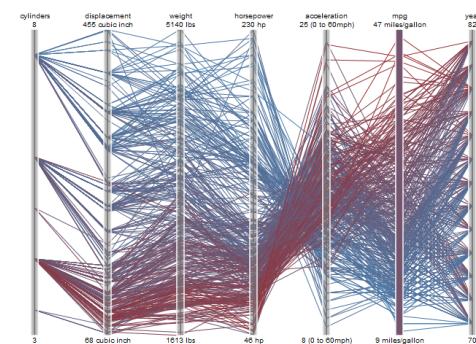
1D



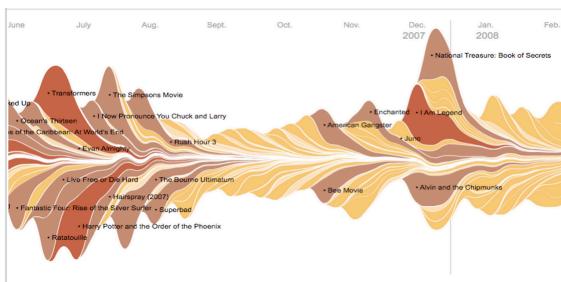
2D



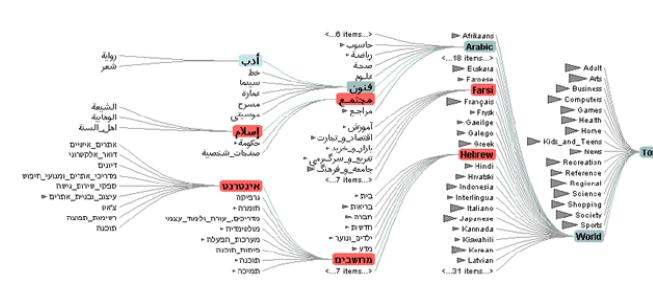
3D



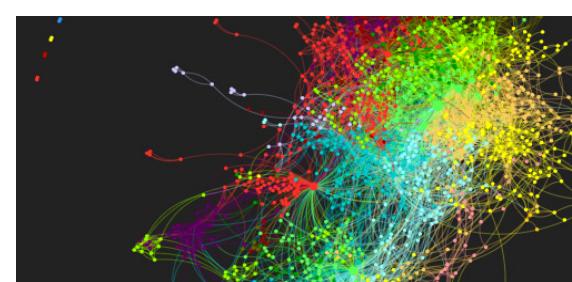
Multi-D



Temporal

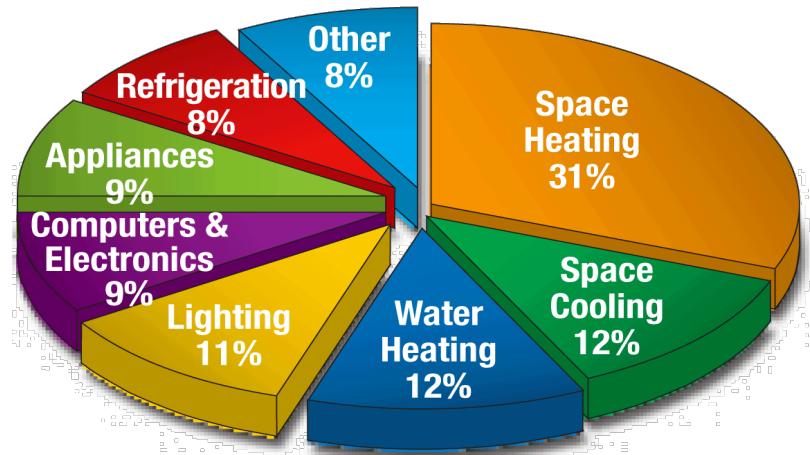


Tree

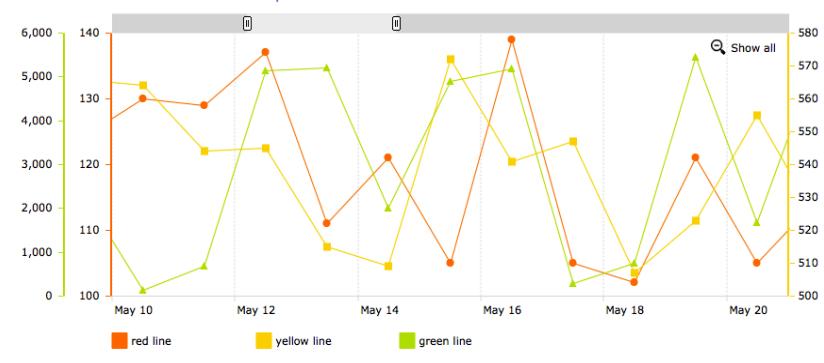


Graph

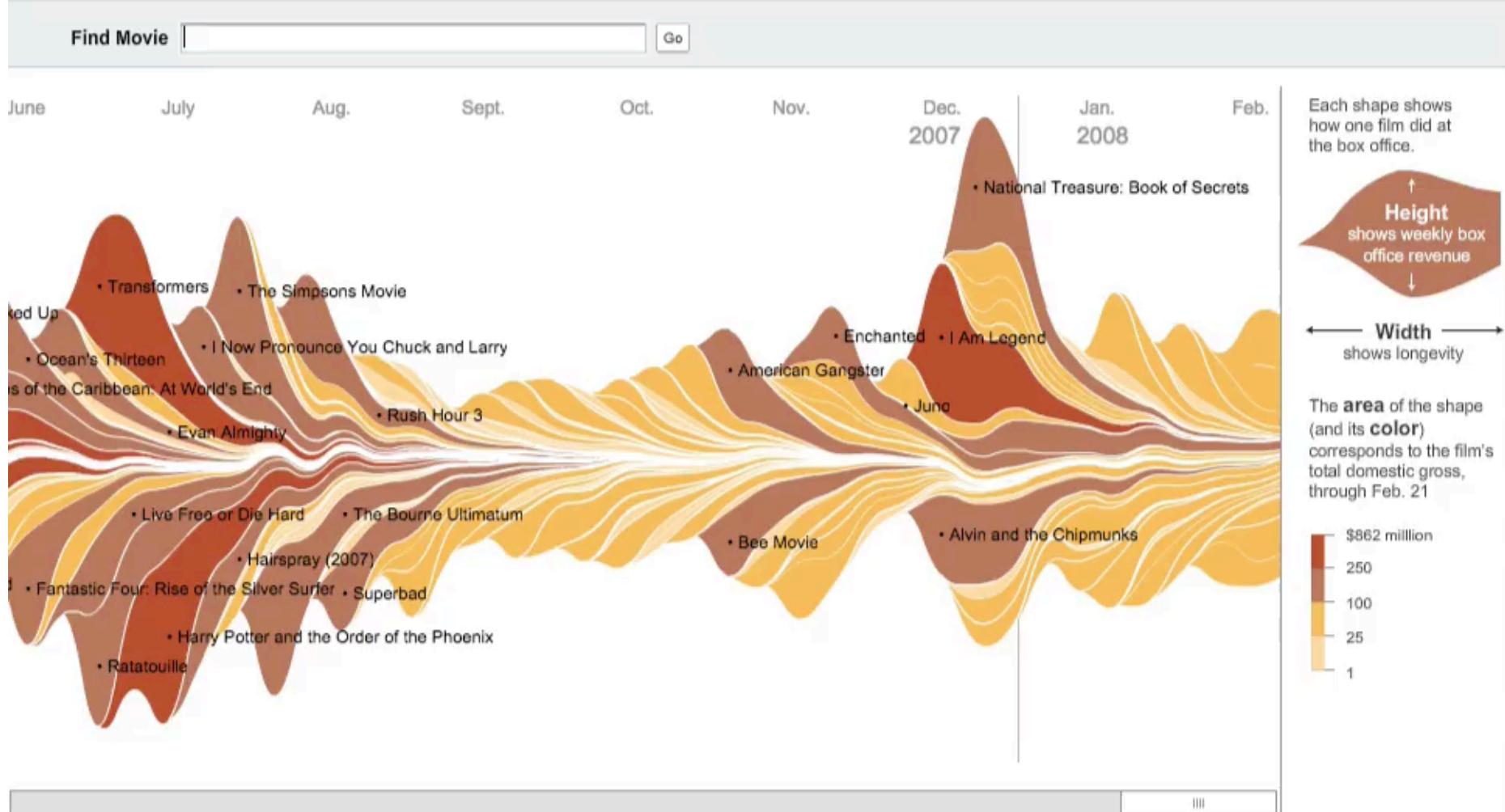
Examples: Visualizing 1D Numerical Data



▶ Line chart with multiple value axes



Example : Visualizing 1D Ordinal Data



Sources: Baseline StudioSystems; Box Office Mojo

Mathew Bloch, Lee Byron, Shan Carter and Amanda Cox

28

Examples: 2D Data

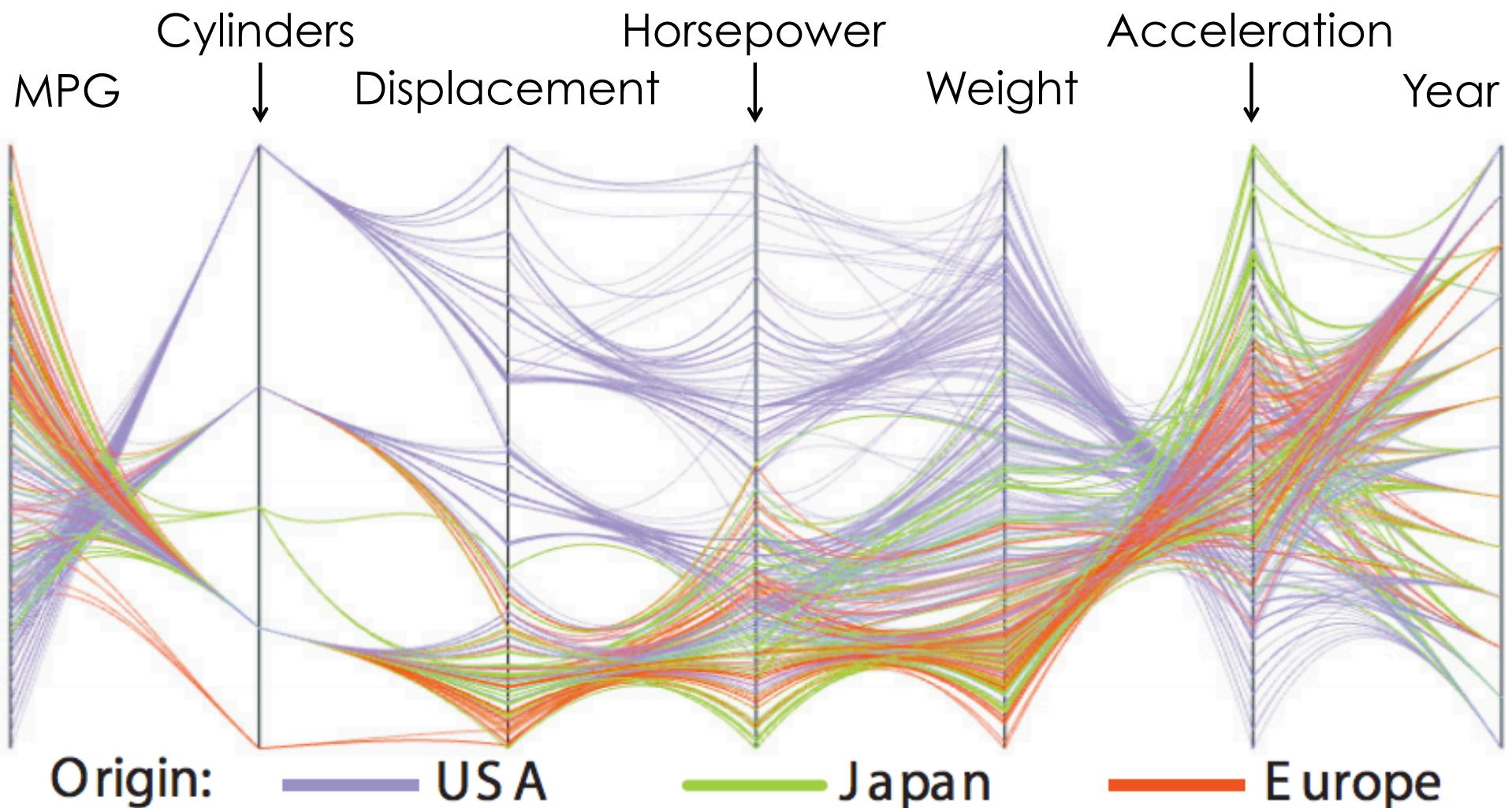
S&P 500 AUG 29 2008 04:00 PM

 finviz.com

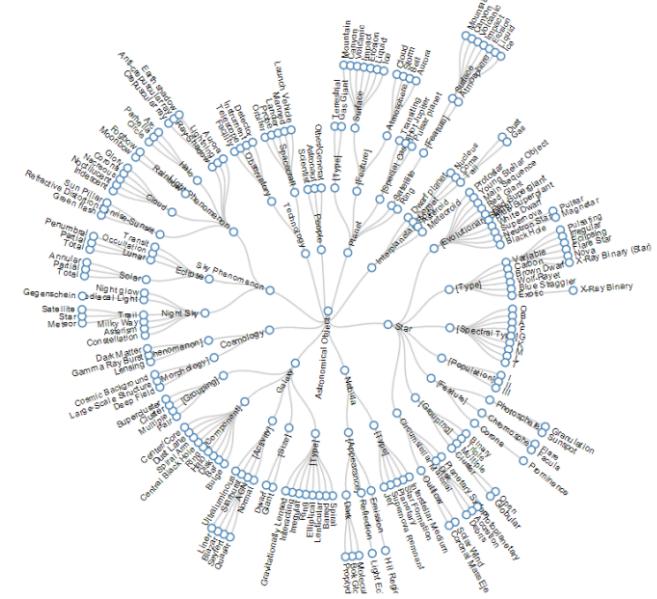
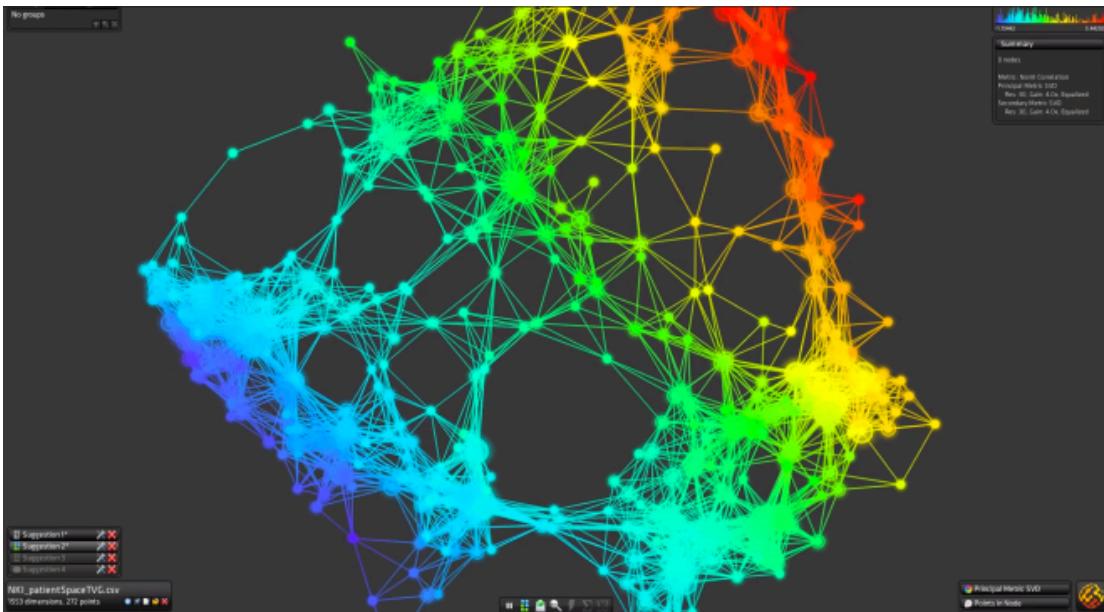


Size of each Cell: Stock Market Value
 Color: Stock Change

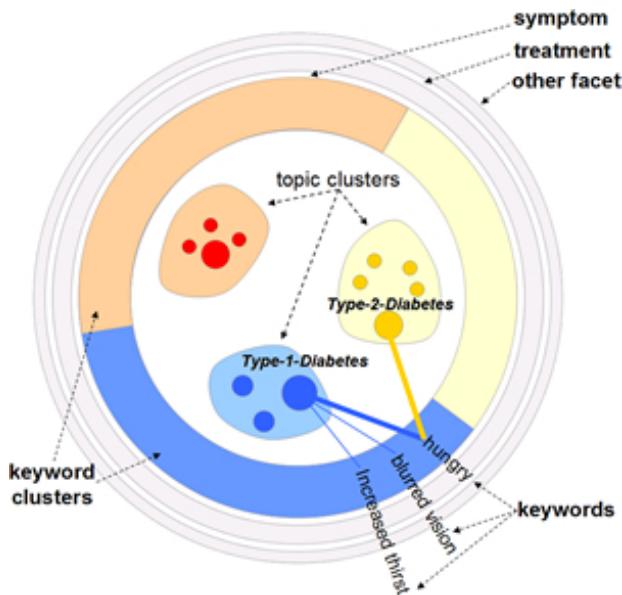
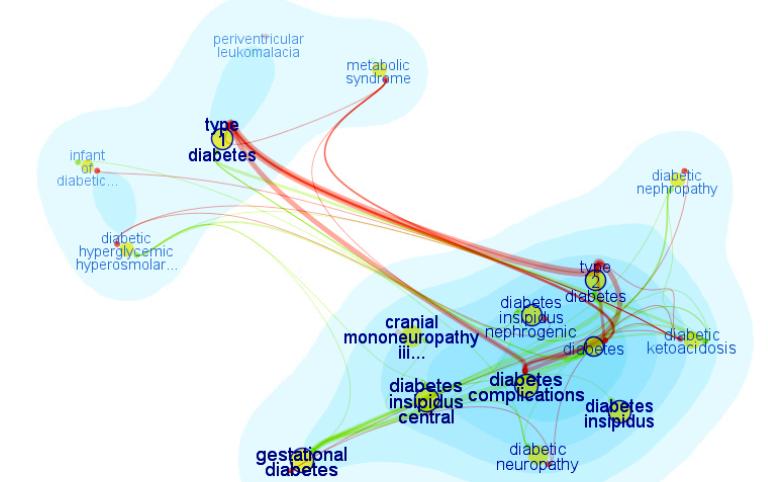
Example : Multi-Dimensional Data



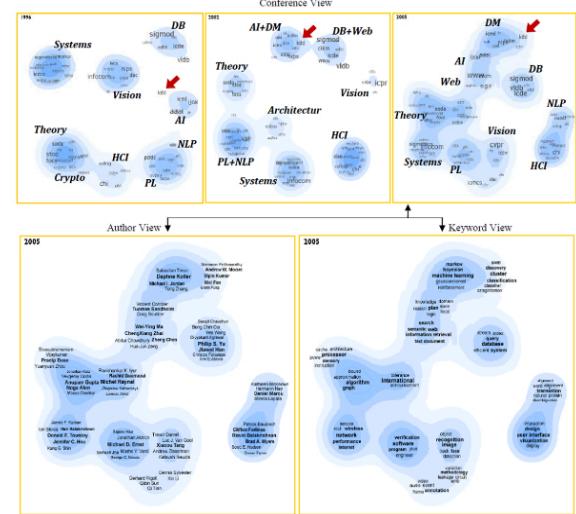
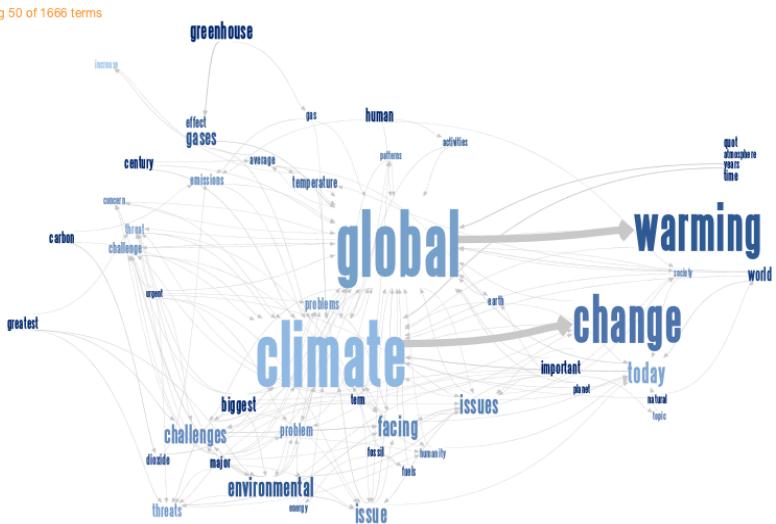
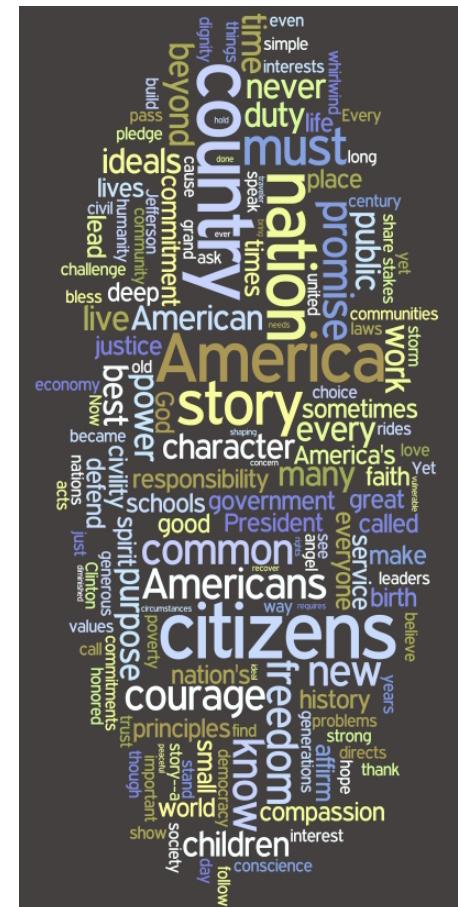
Examples: Visualizing Structured Data



Examples: Visualizing Unstructured Data



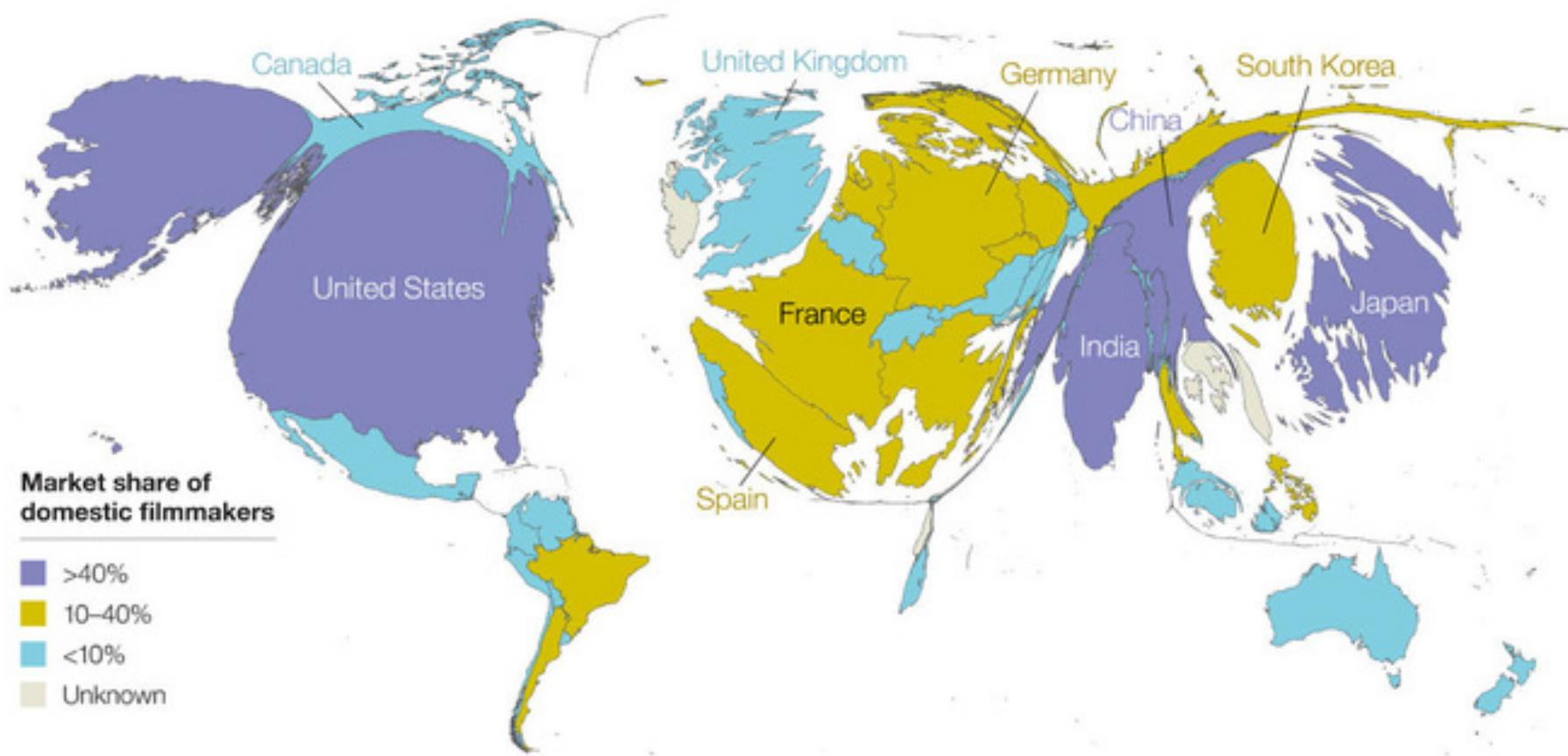
Visualization of Text Documents



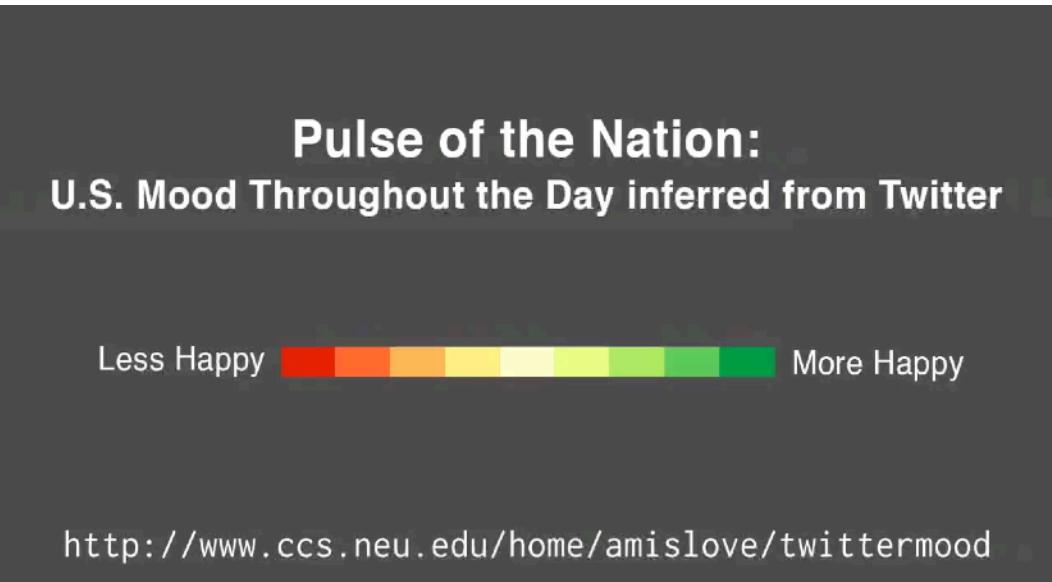
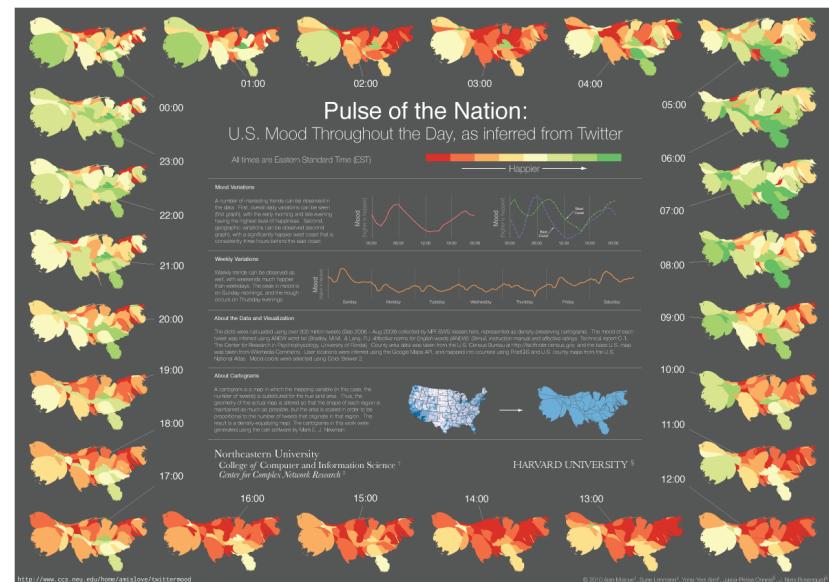
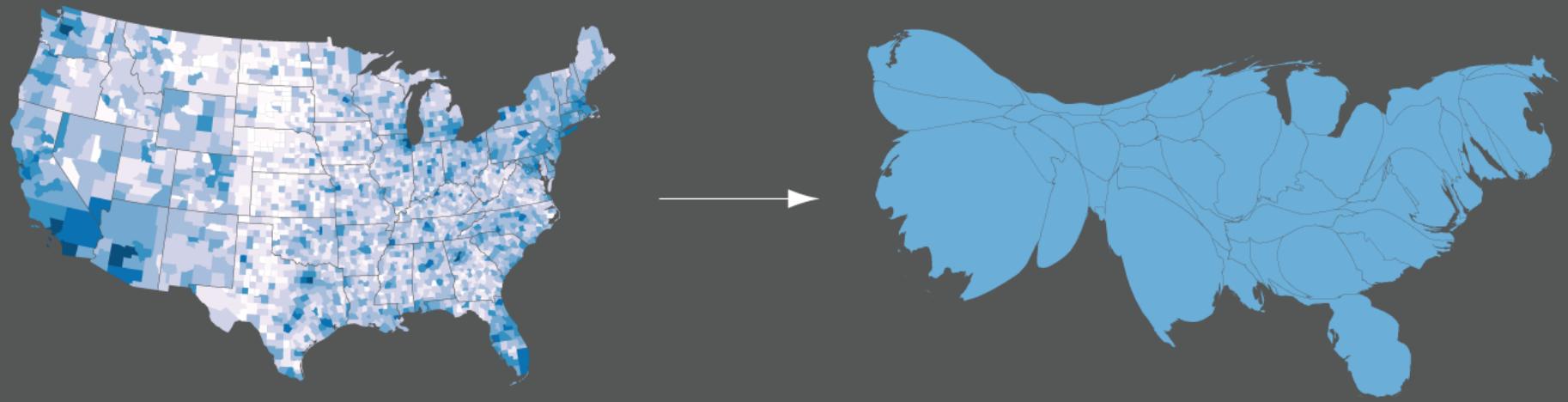
Examples: Geospatial

Larger cinema markets support stronger domestic film industries.

Countries sized by relative share of worldwide box office revenue, 2009



Examples: Visualizing Spatial Temporal Data



Examples: Visualizing Spatial Temporal Data

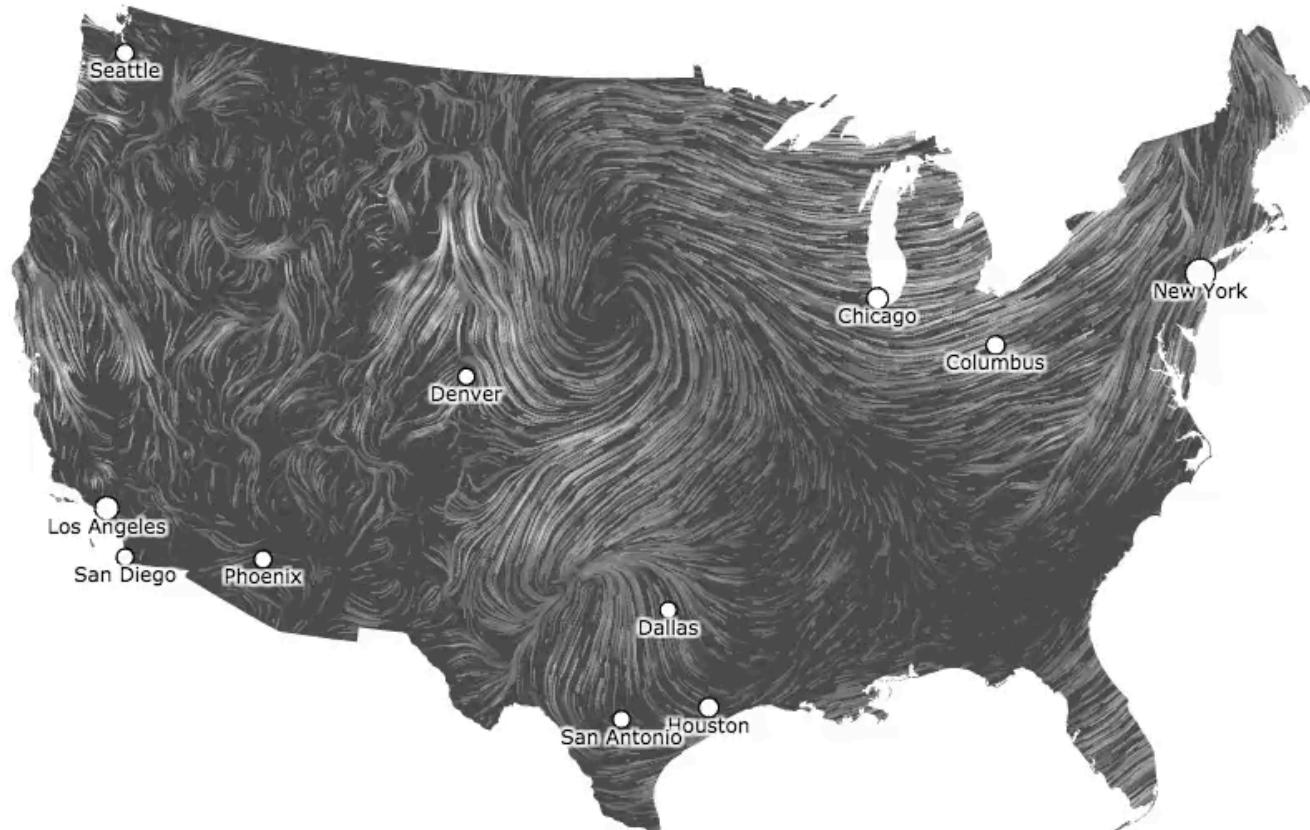
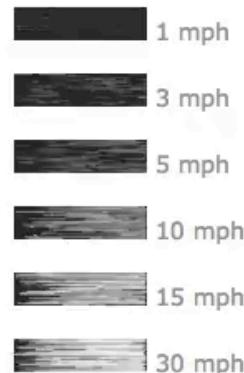
wind map

Dec. 3, 2014

11:35 am EST

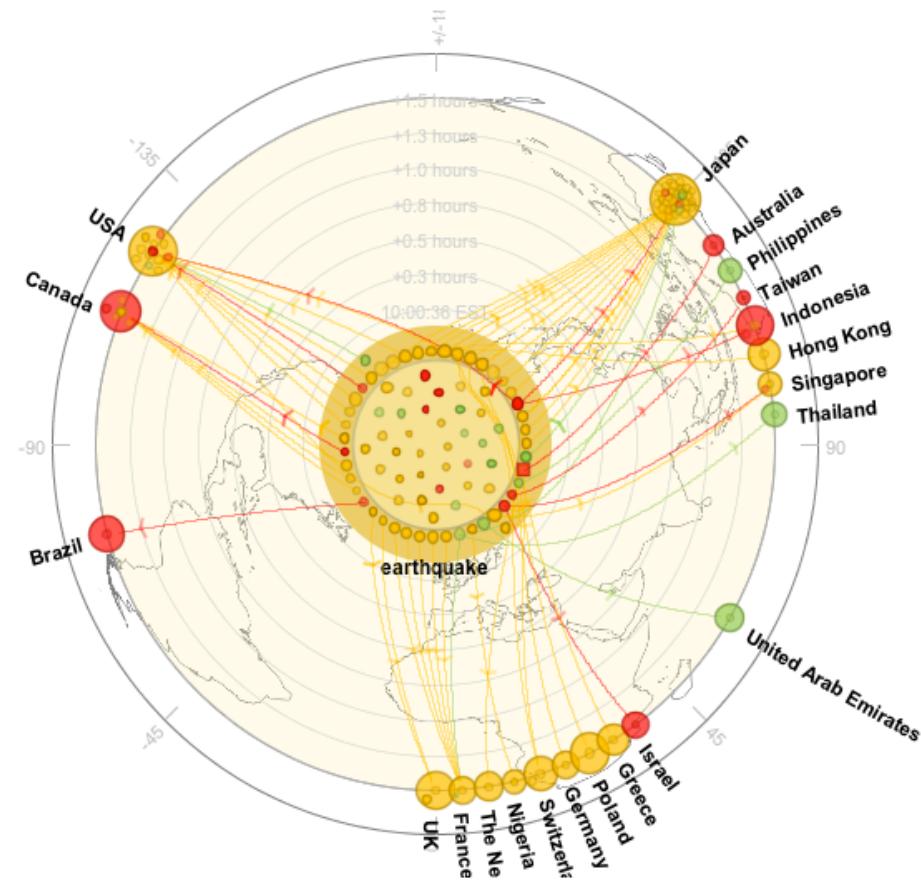
(time of forecast download)

top speed: 31.5 mph
average: 8.2 mph



<http://hint.fm/wind/>

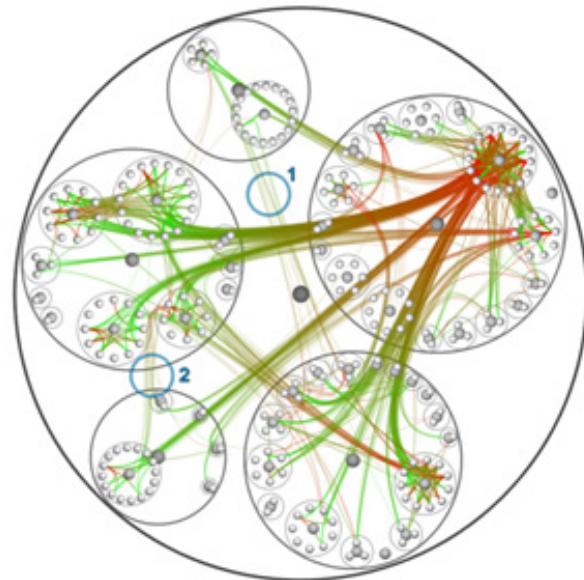
Visualization is not just a beautiful picture



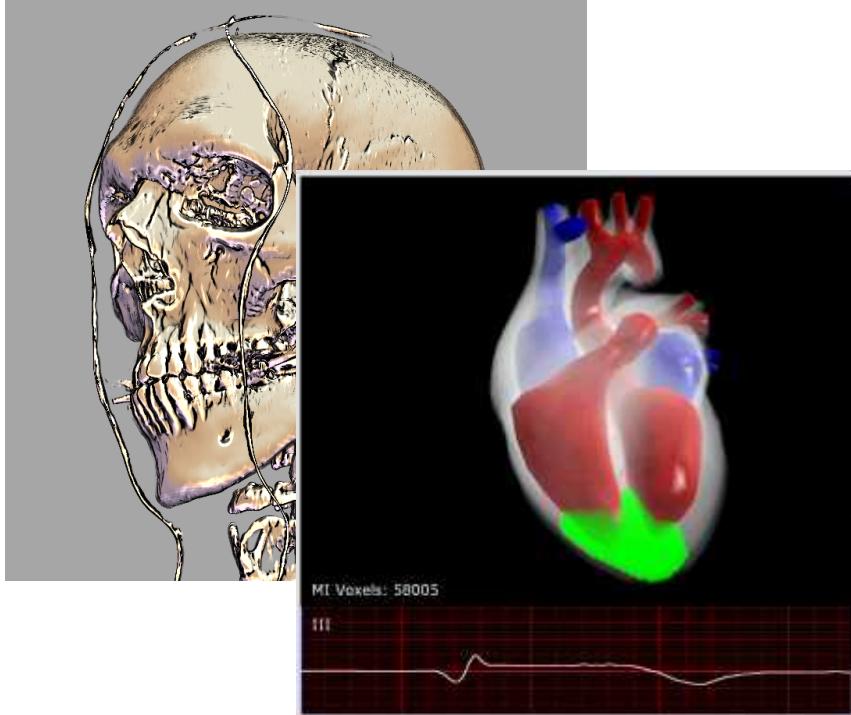
The purpose of visualization is to reveal the insight of the data



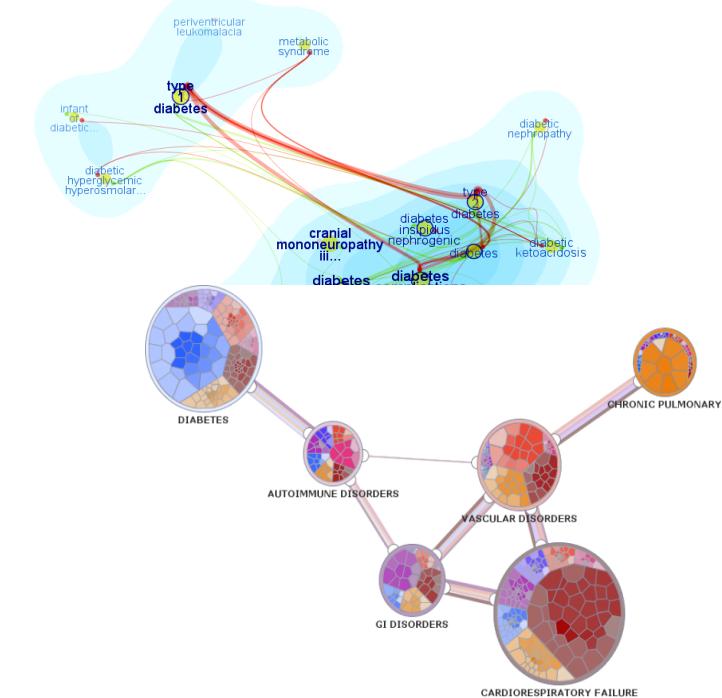
Realism



Information



Physical Data



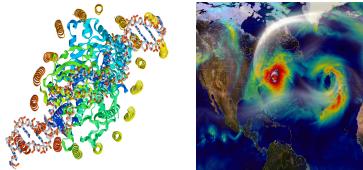
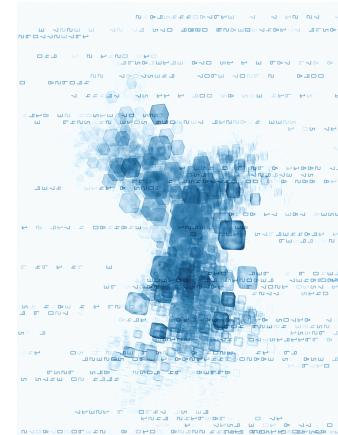
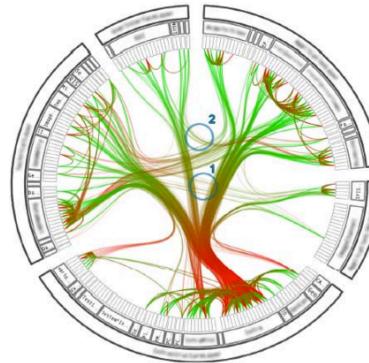
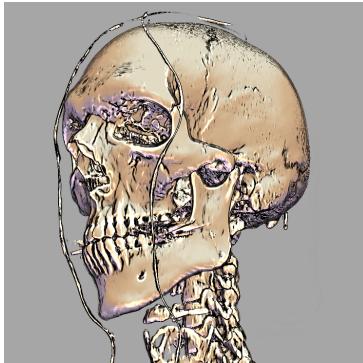
Artificial Data

Three Sub-areas

Scientific
Visualization
(SCIVIS)

Information
Visualization
(InfoVis)

Visual Analysis
(VA)



Physical Data

Abstract Data

InfoVis
Data Mining
Machine Learning

Traditional

Research Trend

Modern Technique

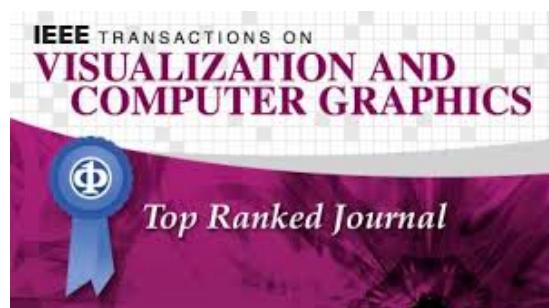
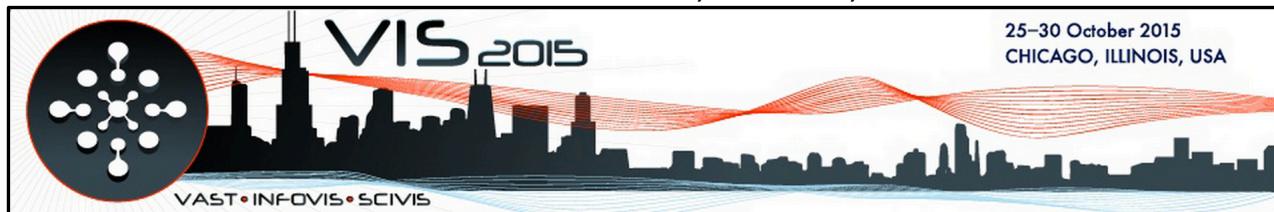
Major Conferences and Journals

Scientific
Visualization
(SCIVIS)

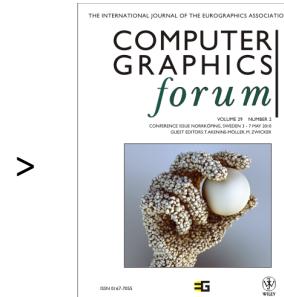
Information
Visualization
(InfoVis)

Visual Analysis
(VA)

VisWeek: IEEE SCIVIS, INFOVIS, VAST



IEEE Transaction on Visualization
and Computer Graphics

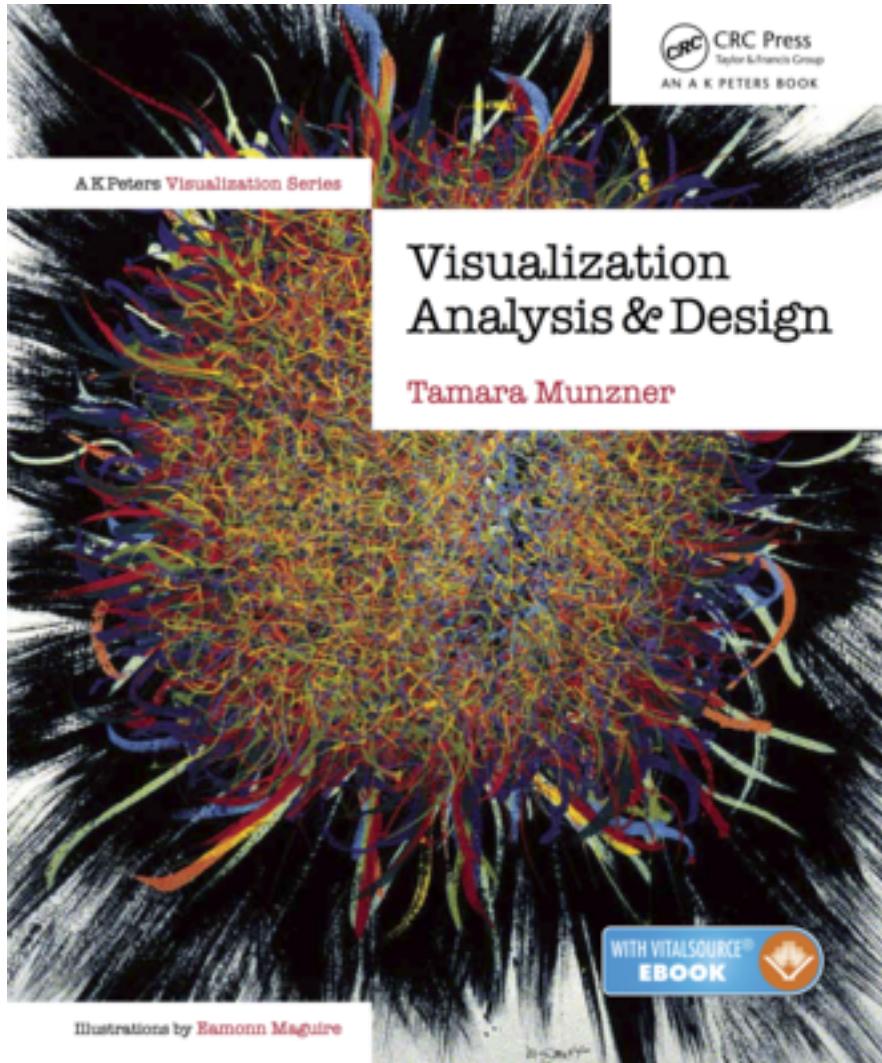
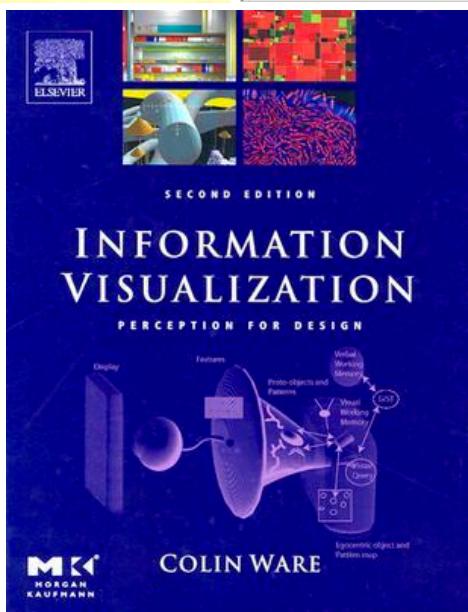
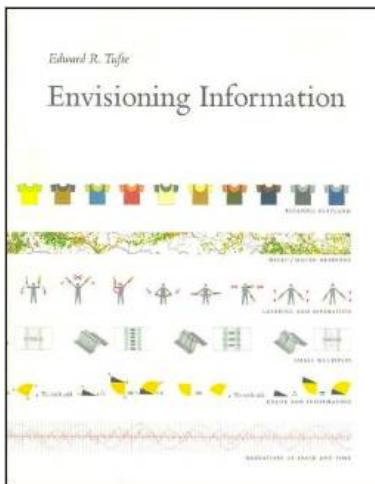
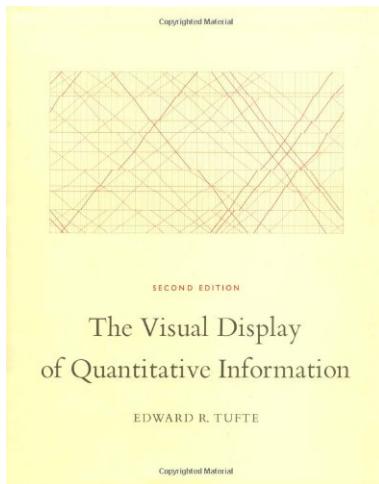


Computer Graphics
Forum



IEEE Computer Graphics
and Application

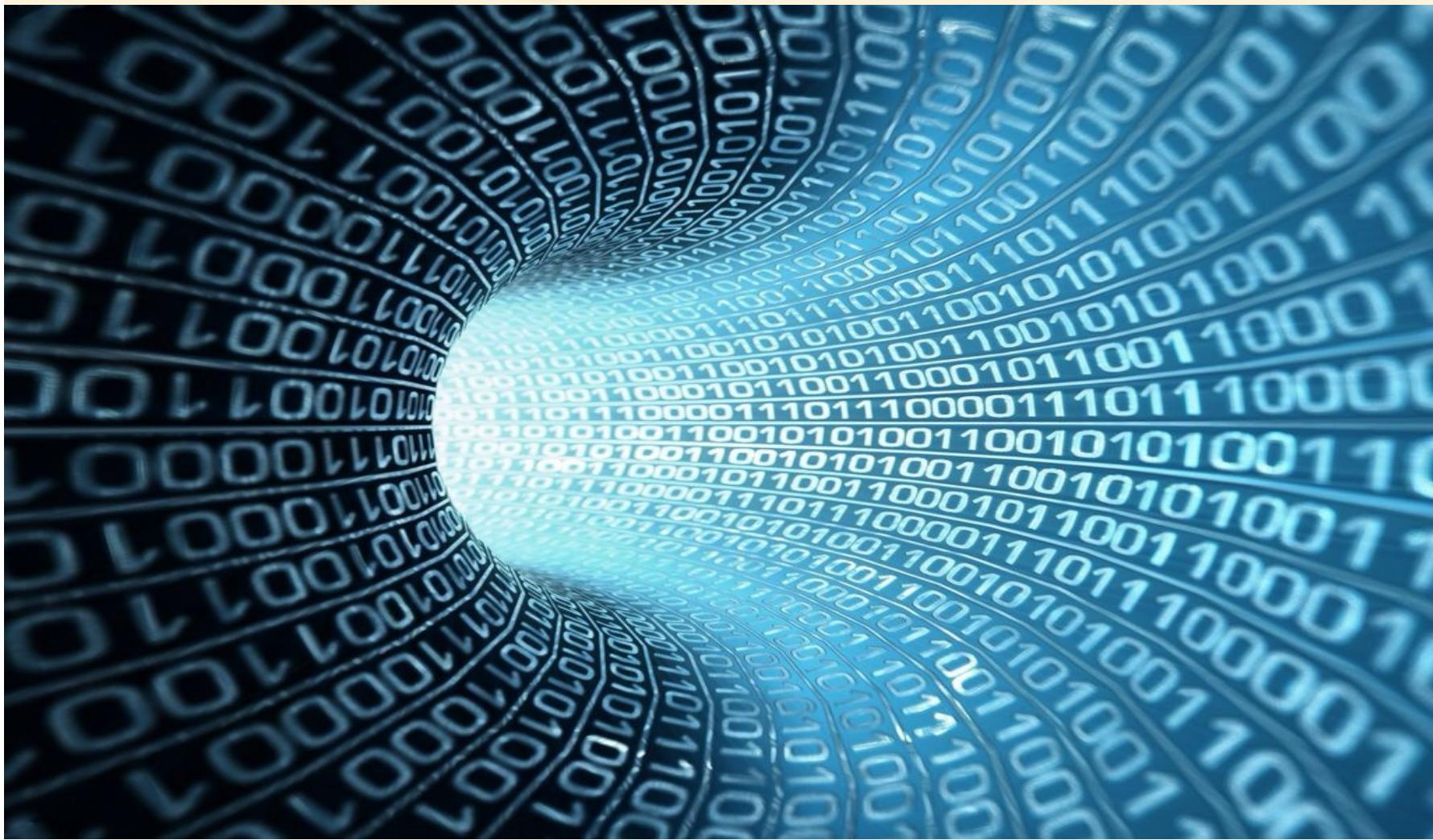
Recommended Books

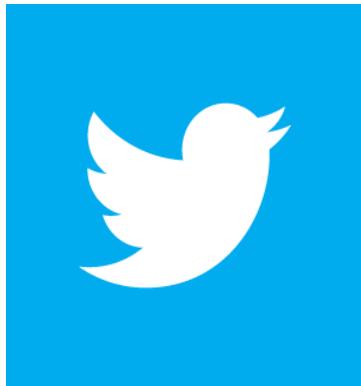


- SVG
- D3.js
- Bootstrap
- (Of course — HTML, Javascript, and CSS)

Part II: Visualising Big Data

Are you ready to Big Data?





***340 million tweets
a day!***



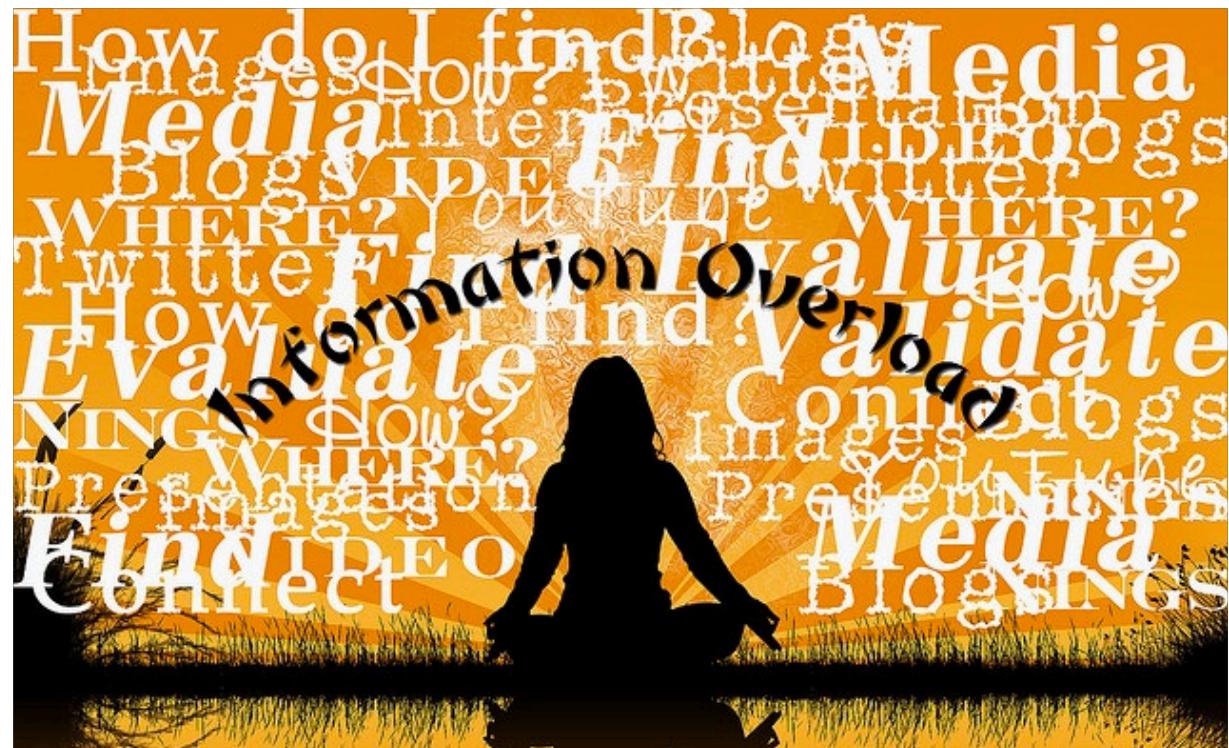
***4 billion messages
a day!***

Information Overload

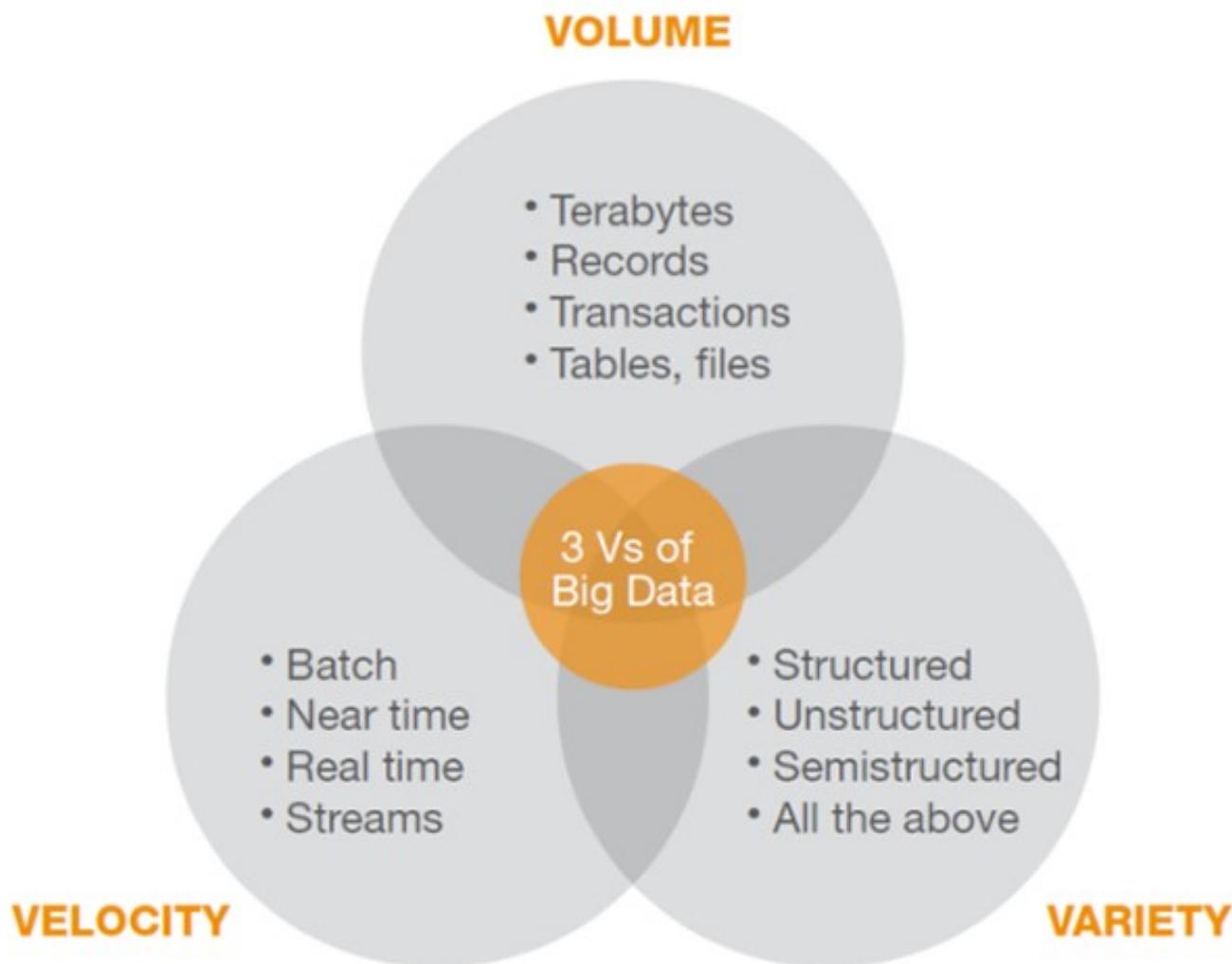
difficult caused by too much information



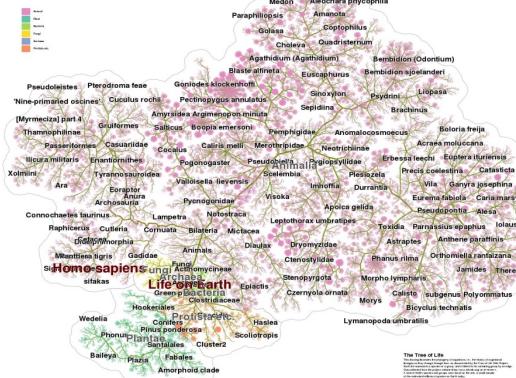
How can we acquire useful information
from the overwhelming data



3Vs of Big Data



76425 species



Tree of Life by Dr. Yifan Hu

14.8 million tweets



The information diffusion graph of the death of Osama bin Laden by Gilad Lotan

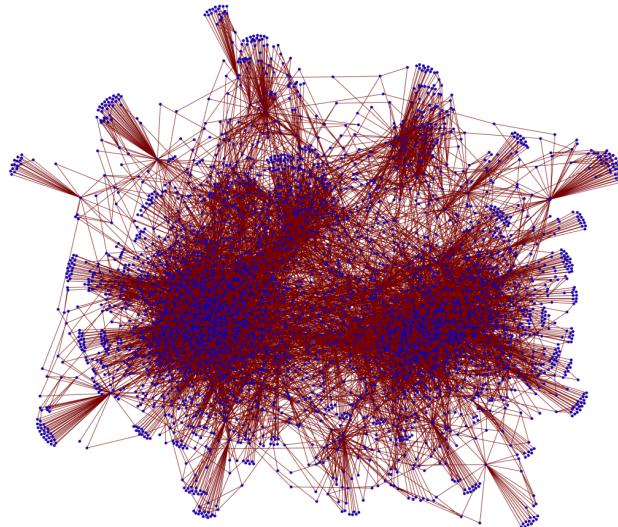
500 million users



Facebook friendship graph by Paul Butler

Challenging Task:

Squeezing millions and even billions of records into million pixels
($1600 \times 1200 = 2$ million pixels)



Visual clutter

How can we avoid visual clutter like overlaps and crossings?



Performance issues

How can we render the huge datasets in real time with rich interactions?



Limited cognition

How can users understand the visual representation when the information is overwhelming?

data item

attr1	
attr2	
attr3	
attr4	
attr5	
attr6	

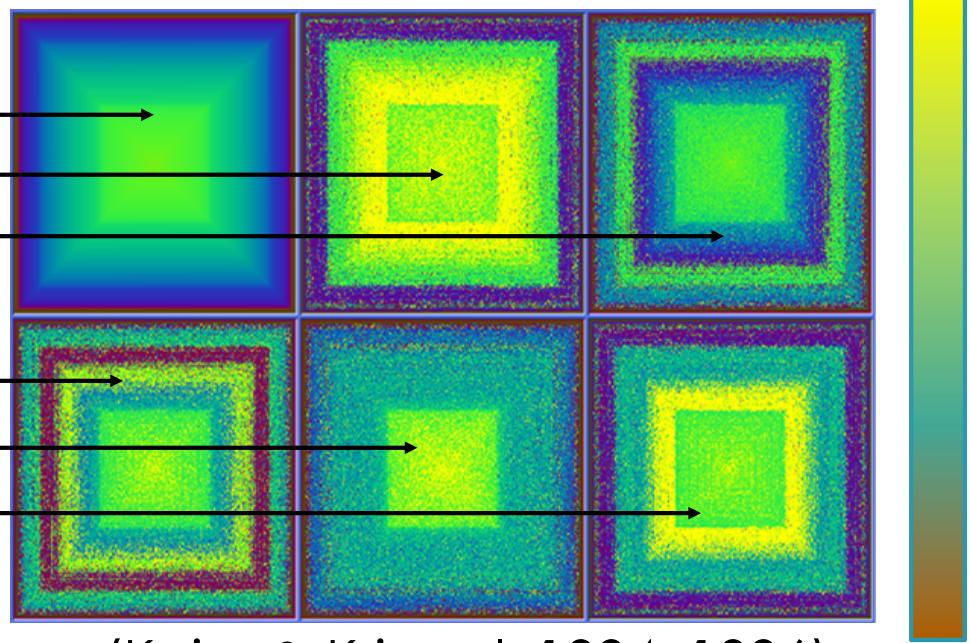
- ✿ A multidimensional data item contains 6 attributes

Technique(1) : Pixel Oriented Visualization

Database visualization (10,000 items, 6 dimensions)

Jan	Feb	Mar	Apr	May	Jun
-99.99	-99.99	315.7	317.45	317.5	317.26
315.62	316.38	316.71	317.72	318.29	318.16
316.43	316.97	317.58	319.02	320.03	319.59
316.93	317.7	318.54	319.48	320.58	319.77
317.94	318.56	319.68	320.63	321.01	320.55
318.74	319.08	319.86	321.39	322.24	321.47
319.57	-99.99	-99.99	-99.99	322.24	321.89
319.44	320.44	320.89	322.13	322.16	321.87
320.62	321.59	322.39	323.87	324.01	323.75
322.06	322.5	323.04	324.42	325	324.09
322.57	323.15	323.89	325.02	325.57	325.36
324	324.42	325.64	326.66	327.34	326.76
325.03	325.99	326.87	328.14	328.07	327.66
326.17	326.68	327.18	327.78	328.92	328.57
326.77	327.63	327.75	329.72	330.07	329.09
328.55	329.56	330.3	331.5	332.48	331.07
329.35	330.71	331.48	332.65	333.09	332.25
330.4	331.41	332.04	333.31	333.96	333.6
331.75	332.56	333.5	334.58	334.87	334.34
332.93	333.42	334.7	336.07	336.74	336.27
334.97	335.39	336.64	337.76	338.01	337.89
336.23	336.76	337.96	338.89	339.47	339.29
338.01	338.36	340.08	340.77	341.46	341.17
339.23	340.47	341.38	342.51	342.91	342.25
340.75	341.61	342.7	343.57	344.13	343.35
341.37	342.52	343.1	344.94	345.75	345.32
343.7	344.5	345.28	347.08	347.43	346.79
344.97	346	347.43	348.35	348.93	348.25
346.3	346.96	347.86	349.55	350.21	349.54
348.02	348.47	349.42	350.99	351.84	351.25
350.43	351.73	352.22	353.59	354.22	353.79
352.76	353.97	353.68	355.42	355.67	355.13
353.66	354.7	355.39	356.2	357.16	356.23
354.72	355.75	357.16	358.6	359.33	358.24
355.98	356.72	357.81	359.15	359.66	359.25
356.7	357.16	358.38	359.46	360.28	359.6
358.37	358.91	359.97	361.26	361.68	360.95
359.97	361	361.64	363.45	363.79	363.26
362.05	363.25	364.02	364.72	365.41	364.97
363.18	364	364.56	366.35	366.79	365.62
365.33	366.15	367.31	368.61	369.3	368.87
368.15	368.87	369.59	371.14	371	370.35
369.14	369.46	370.52	371.66	371.82	371.7
370.28	371.5	372.12	372.87	374.02	373.3
372.43	373.09	373.52	374.86	375.55	375.41
374.68	375.63	376.11	377.65	378.35	378.13
376.79	377.37	378.41	380.52	380.63	379.57
378.37	379.69	380.41	382.1	382.28	382.13
381.38	382.03	382.64	384.62	384.95	384.06
382.45	383.68	384.23	386.26	386.39	385.87
385.07	385.72	385.85	386.71	388.45	387.64

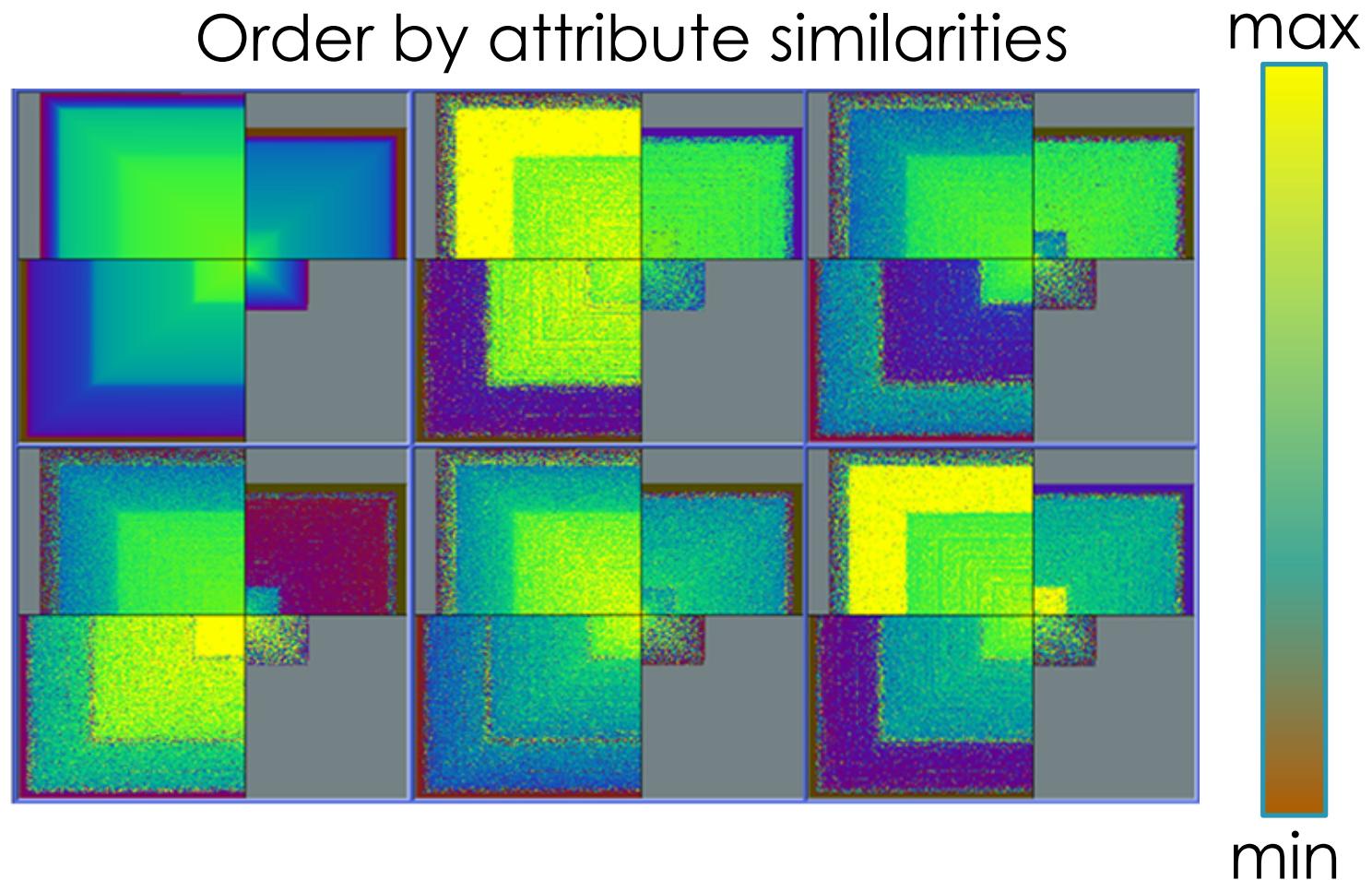
Order by degree of interests max



(Keim & Kriegel, 1994; 1996) min

Techniques (1): Pixel Oriented Visualization

Database Visualization (10,000 items, 6 dimensions)

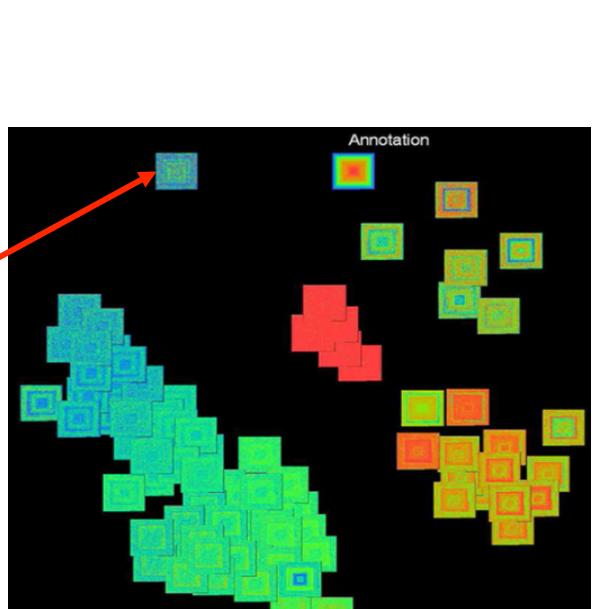


Techniques (1) : Pixel Oriented Visualization

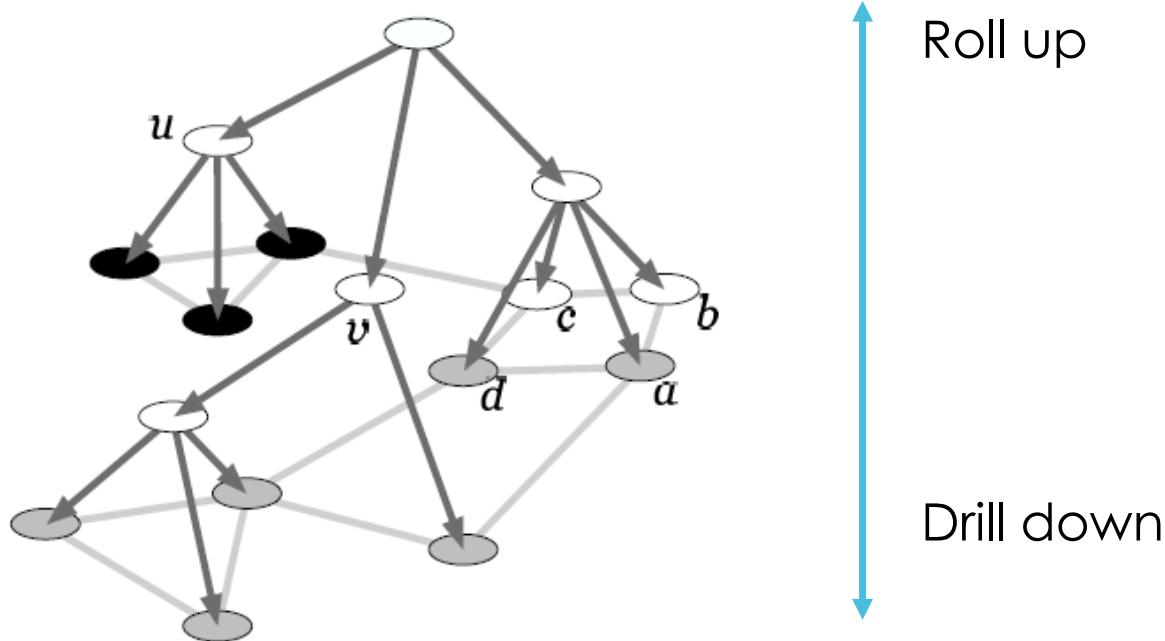
Different Ways for splitting the display region

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual Average
1958	-99.99	-99.99	315.7	317.45	317.5	317.26	315.86	314.93	313.2	312.44	313.33	314.6	-99.99
1959	315.62	316.38	316.71	317.72	318.29	318.16	316.54	314.8	313.84	313.26	314.8	315.58	315.98
1960	316.43	316.97	317.58	319.02	320.03	319.59	318.18	315.91	314.16	313.84	315	316.19	316.91
1961	316.93	317.7	318.54	319.48	320.58	319.77	318.57	316.79	314.8	315.38	316.1	317.01	317.64
1962	317.94	318.56	319.68	320.63	321.01	320.55	319.58	317.4	316.25	315.42	316.69	317.7	318.45
1963	318.74	319.08	319.86	321.39	322.24	321.47	319.74	317.77	316.21	315.99	317.12	318.31	318.99
1964	319.57	-99.99	-99.99	322.24	321.89	320.44	318.7	316.7	316.79	317.79	318.71	-99.99	
1965	319.44	320.44	320.89	322.13	322.16	321.87	321.39	318.8	317.81	317.3	318.87	319.42	320.04
1966	320.62	321.59	322.39	323.87	324.01	323.75	322.39	320.37	318.64	318.1	319.79	320.08	321.38
1967	322.06	322.5	323.04	324.42	325	324.09	322.55	320.92	319.31	319.31	320.72	321.96	322.16
1968	322.57	323.15	323.89	325.02	325.57	325.36	324.14	322.03	320.41	320.25	321.31	322.84	322.05
1969	324	324.42	325.64	326.66	327.34	326.76	325.88	323.67	322.38	321.78	322.85	324.12	324.63
1970	325.03	325.99	326.87	328.14	328.07	327.66	326.35	324.69	323.1	323.16	323.98	325.13	325.68
1971	326.17	326.68	327.18	327.78	328.92	328.57	327.34	325.46	323.36	323.57	324.8	326.01	326.32
1972	326.77	327.63	327.75	329.72	330.07	329.09	328.05	326.32	324.93	325.06	326.5	327.55	327.45
1973	328.55	329.56	330.3	331.5	332.48	332.07	330.87	329.31	327.51	327.18	328.16	328.64	329.68
1974	329.35	330.71	331.48	332.65	333.09	332.25	331.18	329.4	327.43	327.37	328.46	329.57	330.25
1975	330.4	331.41	332.04	333.31	333.96	333.6	331.91	330.06	328.56	328.34	329.49	330.76	331.15
1976	331.75	332.56	333.5	334.58	334.87	334.34	333.05	330.94	329.3	328.94	330.31	331.68	332.15
1977	332.93	333.42	334.7	336.07	336.74	336.27	334.93	332.75	331.59	331.16	332.4	333.85	333.9
1978	334.97	335.39	336.64	337.76	338.01	337.89	336.54	334.68	332.76	332.55	333.92	334.95	335.51
1979	336.23	336.76	337.96	338.89	339.47	339.29	337.73	336.09	333.91	333.86	335.29	336.73	336.85
1980	338.01	338.36	340.08	340.77	341.46	341.17	339.56	337.6	335.88	336.02	337.1	338.21	338.69
1981	339.23	340.47	341.38	342.51	342.91	342.25	340.49	338.43	336.69	336.86	338.36	339.61	339.9
1982	340.75	341.61	342.7	343.57	344.13	343.35	342.06	339.81	337.98	337.86	339.26	340.49	341.13
1983	341.37	342.52	343.1	344.94	345.75	345.32	343.99	342.39	339.86	339.99	341.15	342.99	342.78
1984	343.7	344.5	345.28	347.08	347.43	346.79	345.4	343.28	341.07	341.35	342.98	344.22	344.42
1985	344.97	346	347.43	348.35	348.93	348.25	346.56	344.68	343.09	342.8	344.24	345.55	345.9
1986	346.3	346.96	347.86	349.55	350.21	349.54	347.94	345.9	344.85	344.17	345.66	346.9	347.15
1987	348.02	348.47	349.42	350.99	351.84	351.25	349.52	348.1	346.45	346.36	347.81	348.96	349.93
1988	350.43	351.73	352.22	353.59	354.22	353.79	352.38	350.43	348.72	348.88	350.07	351.34	351.48
1989	352.76	353.07	353.68	355.42	355.67	355.13	353.9	351.67	349.8	349.99	351.29	352.52	352.91
1990	353.66	354.7	355.39	356.2	357.16	356.23	354.82	352.91	350.96	351.18	352.83	354.21	354.19
1991	354.72	355.75	357.16	358.56	359.33	358.24	356.17	354.02	352.15	352.21	353.75	354.99	355.59
1992	355.98	356.72	357.81	359.15	359.66	359.25	357.02	355	353.01	353.31	354.16	355.4	356.37
1993	356.7	357.16	358.38	359.46	360.28	359.6	357.57	355.52	353.69	353.99	355.34	356.8	357.04
1994	358.37	358.91	359.97	361.26	361.68	360.95	359.55	357.48	355.84	355.99	357.58	359.04	358.89
1995	359.97	361	361.64	363.45	363.79	363.26	361.9	359.46	358.05	357.76	359.56	360.7	360.88
1996	362.05	363.25	364.02	364.72	365.41	364.97	363.65	361.48	359.45	359.6	360.76	362.33	362.64
1997	363.18	364	364.56	366.35	366.79	365.62	364.47	362.51	360.19	360.77	362.43	364.28	363.76
1998	365.33	366.15	367.31	368.61	369.3	368.87	367.64	365.77	363.9	364.23	365.46	366.97	366.63
1999	368.15	369.87	369.59	371.14	371	370.35	369.27	366.93	364.63	366.13	366.67	368.01	368.31
2000	369.14	369.46	370.52	371.66	371.82	371.7	370.12	368.12	366.62	366.73	368.29	369.53	369.48
2001	370.28	371.5	372.12	372.87	374.02	373.3	371.62	369.55	367.96	368.09	369.68	371.24	371.02
2002	372.43	373.09	373.52	374.86	375.55	375.41	374.02	371.49	370.7	370.25	372.08	373.78	373.1
2003	374.68	375.63	376.11	377.65	378.35	378.13	376.62	374.5	372.99	373.01	374.35	375.7	375.64
2004	376.79	377.37	378.41	380.52	380.63	379.57	377.79	375.86	374.07	374.24	375.86	377.47	377.38
2005	378.37	379.69	380.41	382.1	382.28	382.13	380.66	378.71	376.42	376.88	378.32	380.04	379.67
2006	381.38	382.03	382.64	384.62	384.95	384.06	382.29	380.47	378.67	379.06	380.14	381.71	381.84
2007	382.45	383.68	384.23	386.26	386.39	385.87	384.39	381.78	380.73	380.81	382.33	383.65	383.55
2008	385.07	385.72	385.85	386.71	388.45	387.64	386.1	383.95	382.73	383.96	385.07		

Values above represent monthly concentrations adjusted to represent 2400 hours on the 15th day of each month. Units are parts per million by volume (ppmv) expressed in the 2003A SIO manometric mole fraction scale. The "annual average" is the arithmetic mean of the twelve monthly values where no monthly values are missing.

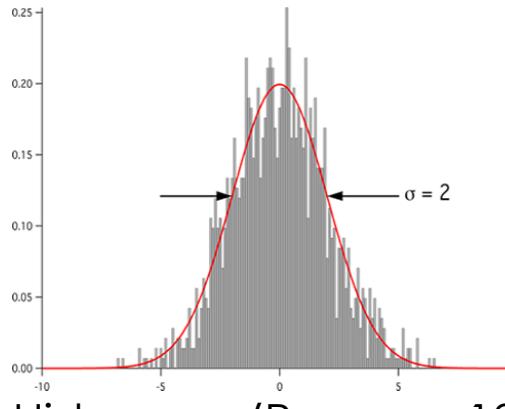


(Yang et al., 2006)

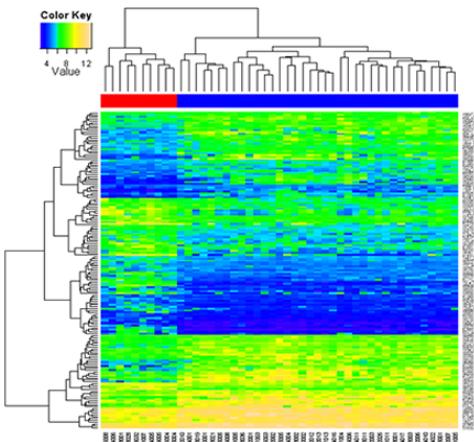


Building a tree for aggregating data items in either a bottom-up or top-down approach

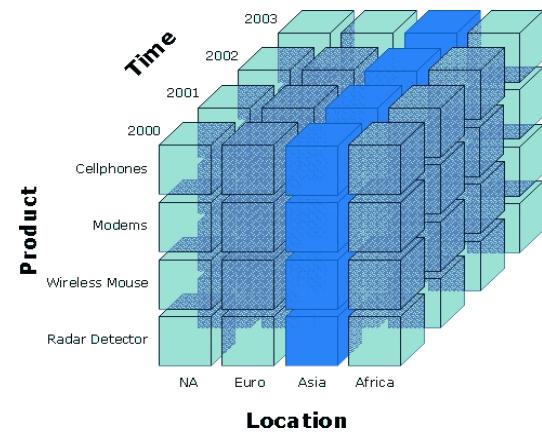
Technique (2) : Aggregation & LOD



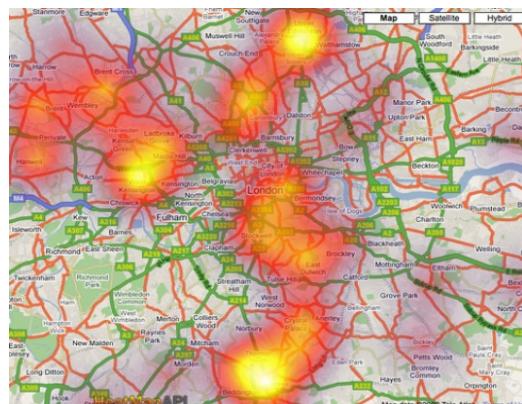
Histogram (Pearson, 1895)



Heatmap
(Wilkinson & Friendly, 2009)



InfoCube
(Stolte et al., 2003)

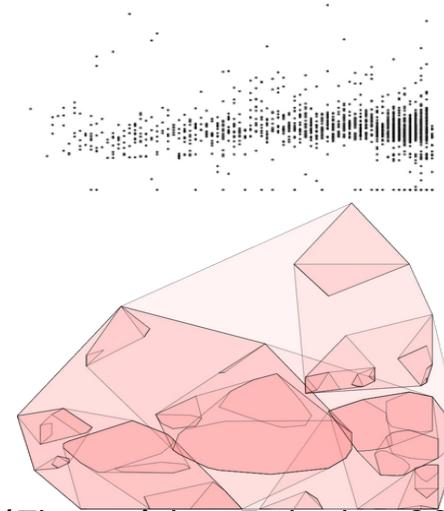


(Lin et al., 2010)



Techniques (2) : Aggregation

Scatter Plots

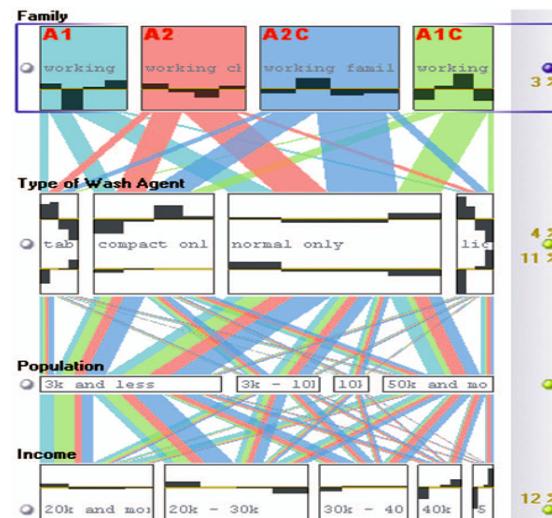


(Elmqvist & Fekete, 2010)

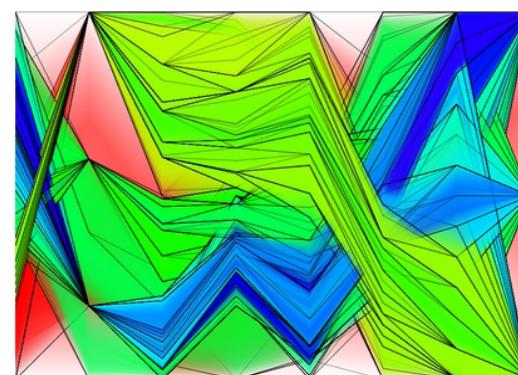


(Yang et al., 2003b)

Parallel Coordinates

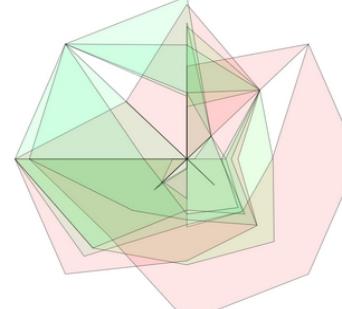
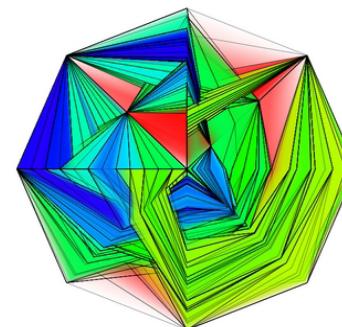
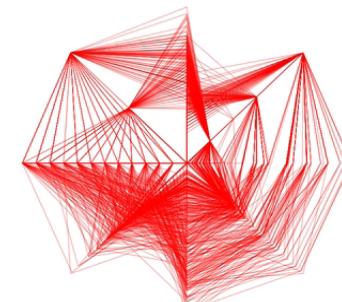


(Kosara et al., 2006)



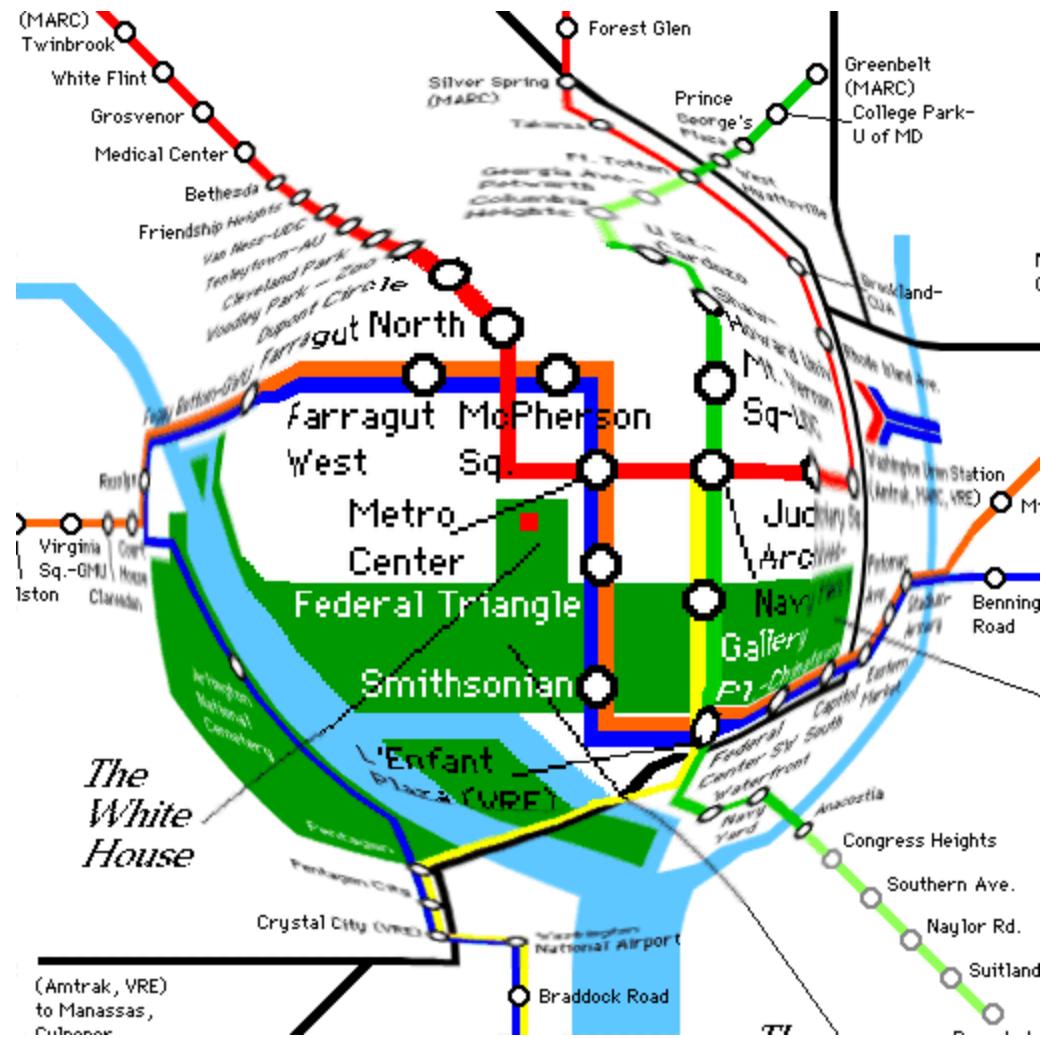
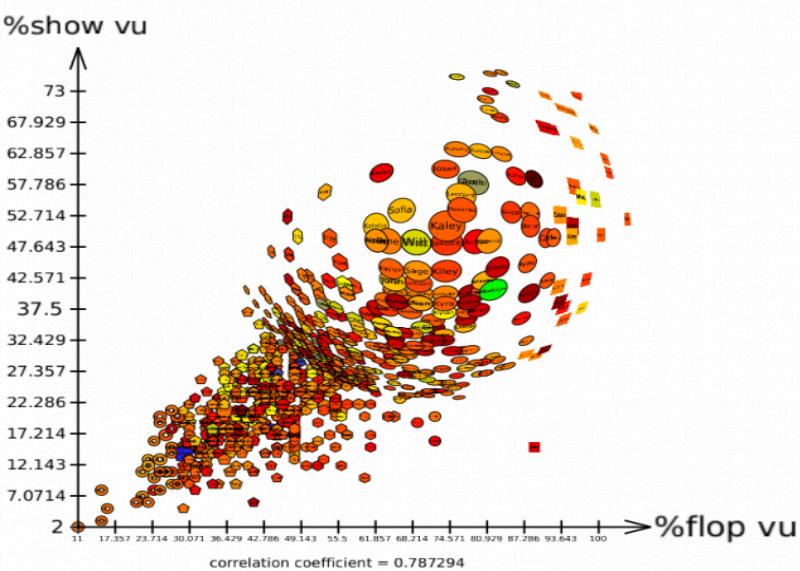
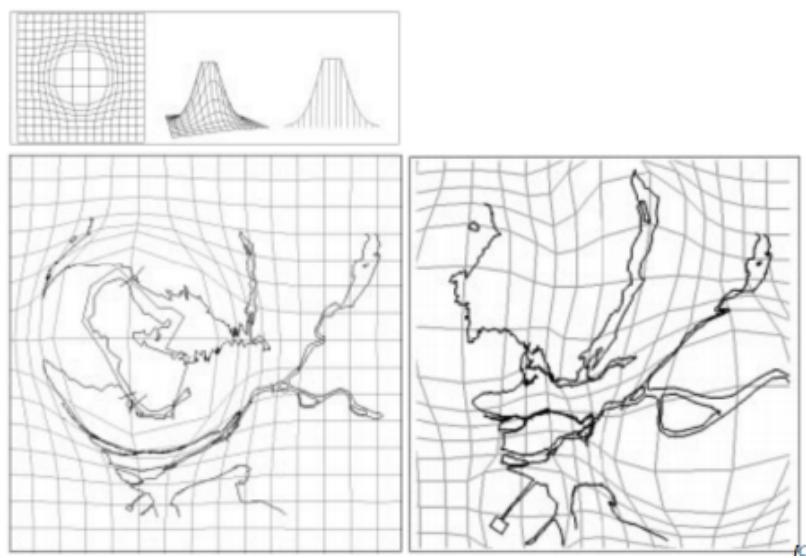
(Fua et al. 1999)

Star Plots

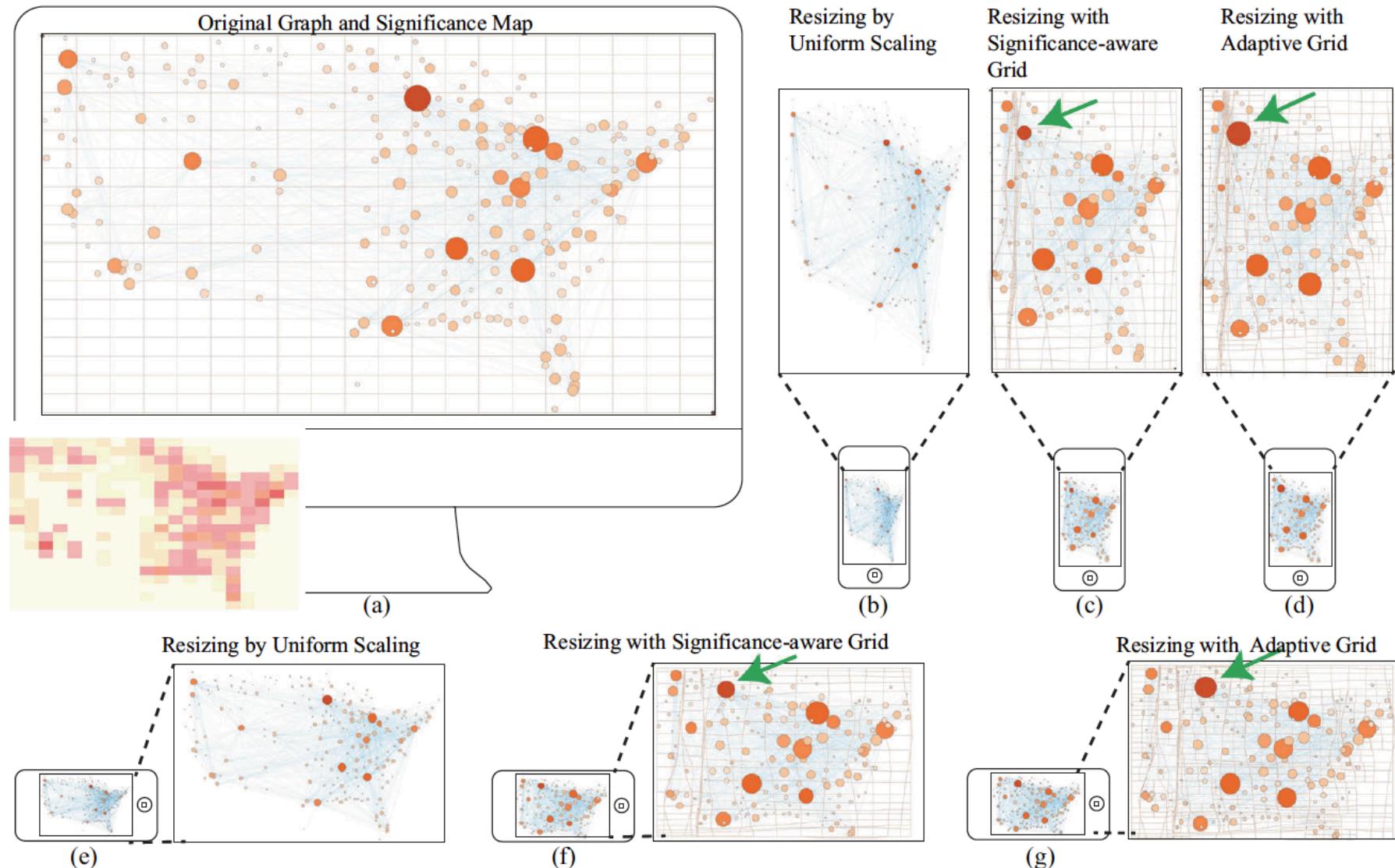


(Fua et al. 1999)

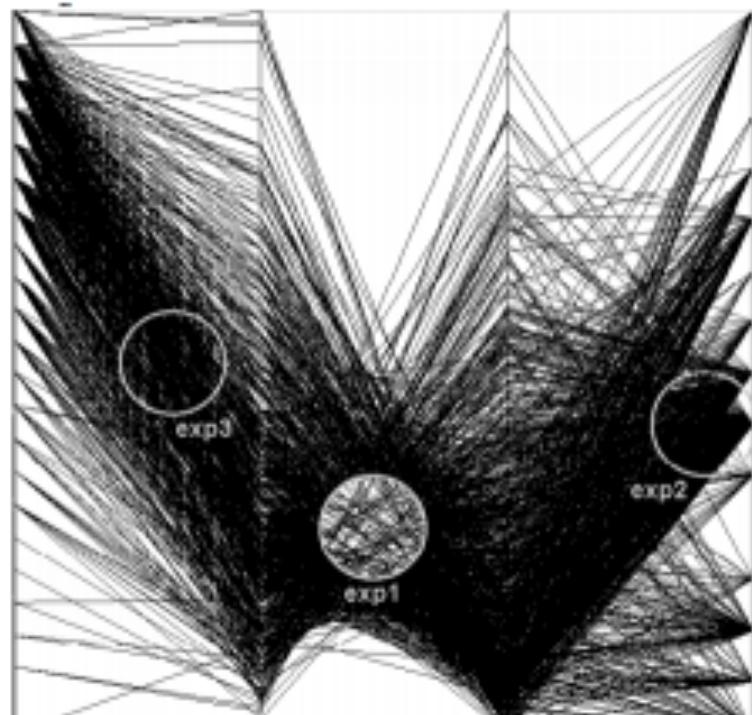
Technique (3) : Distortion



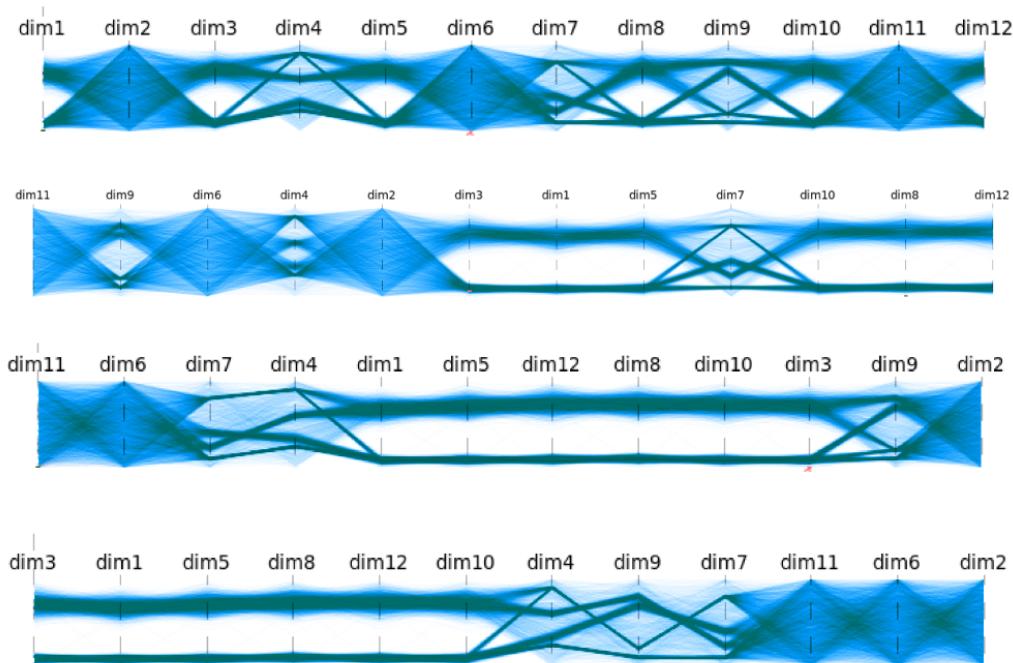
Techniques (3) : Distortion



Technique (4) : Clutter Reduction



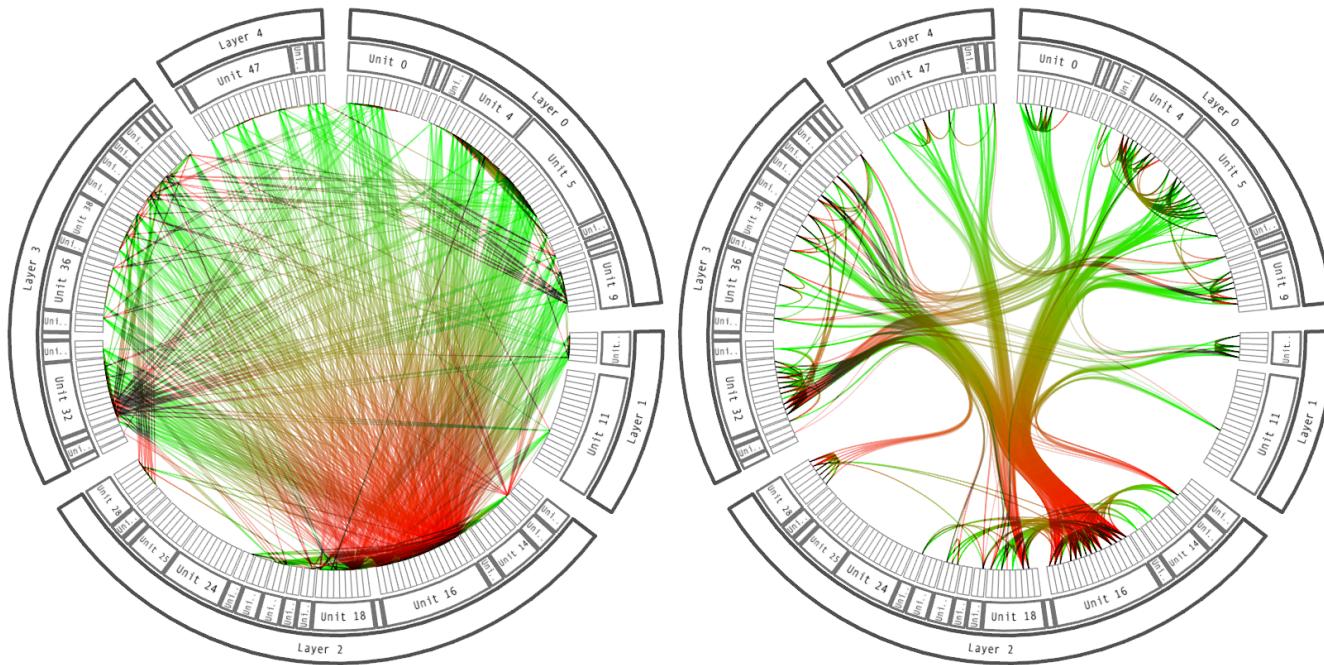
Sampling



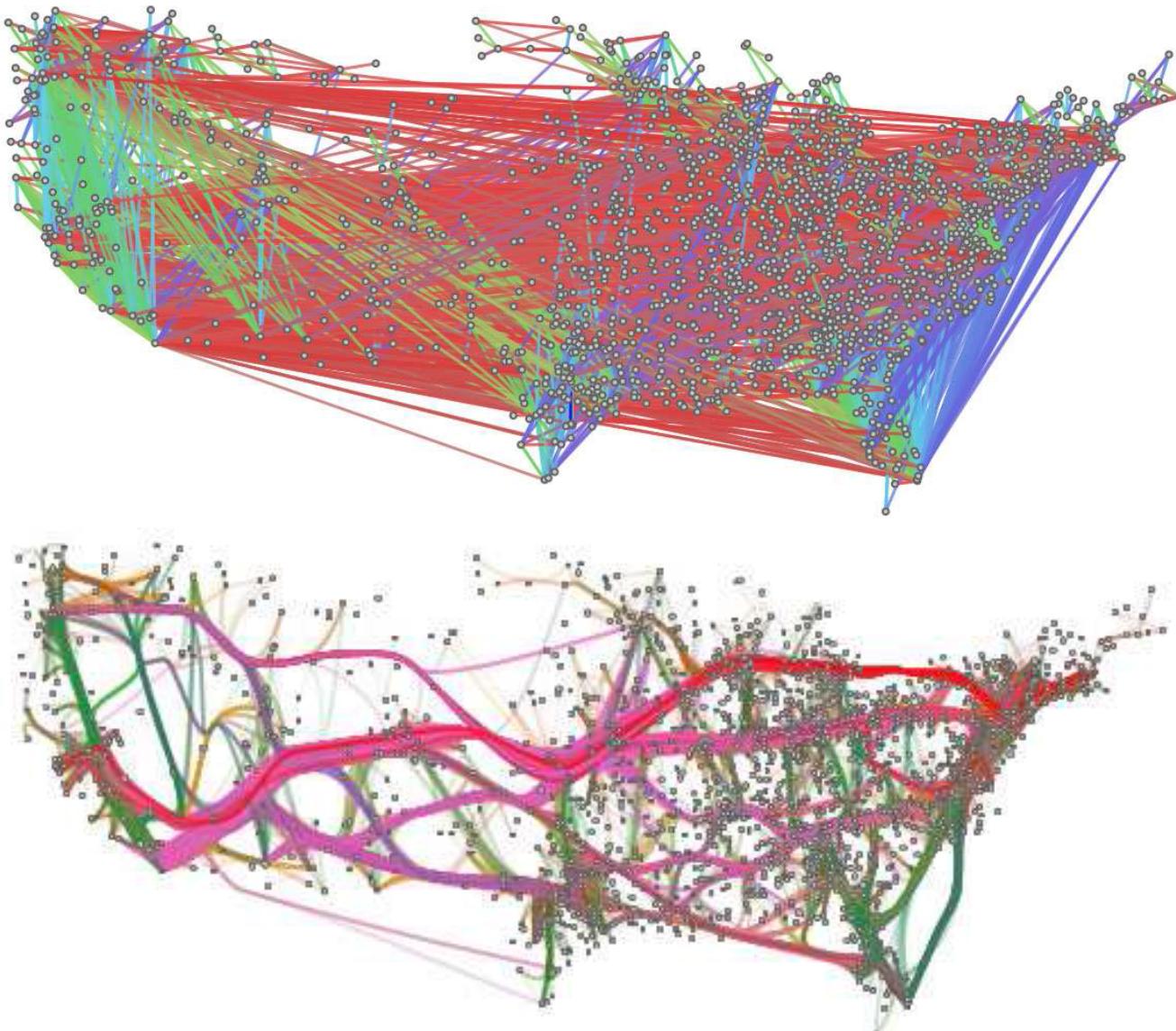
Reordering

Technique (4): Clutter Reduction

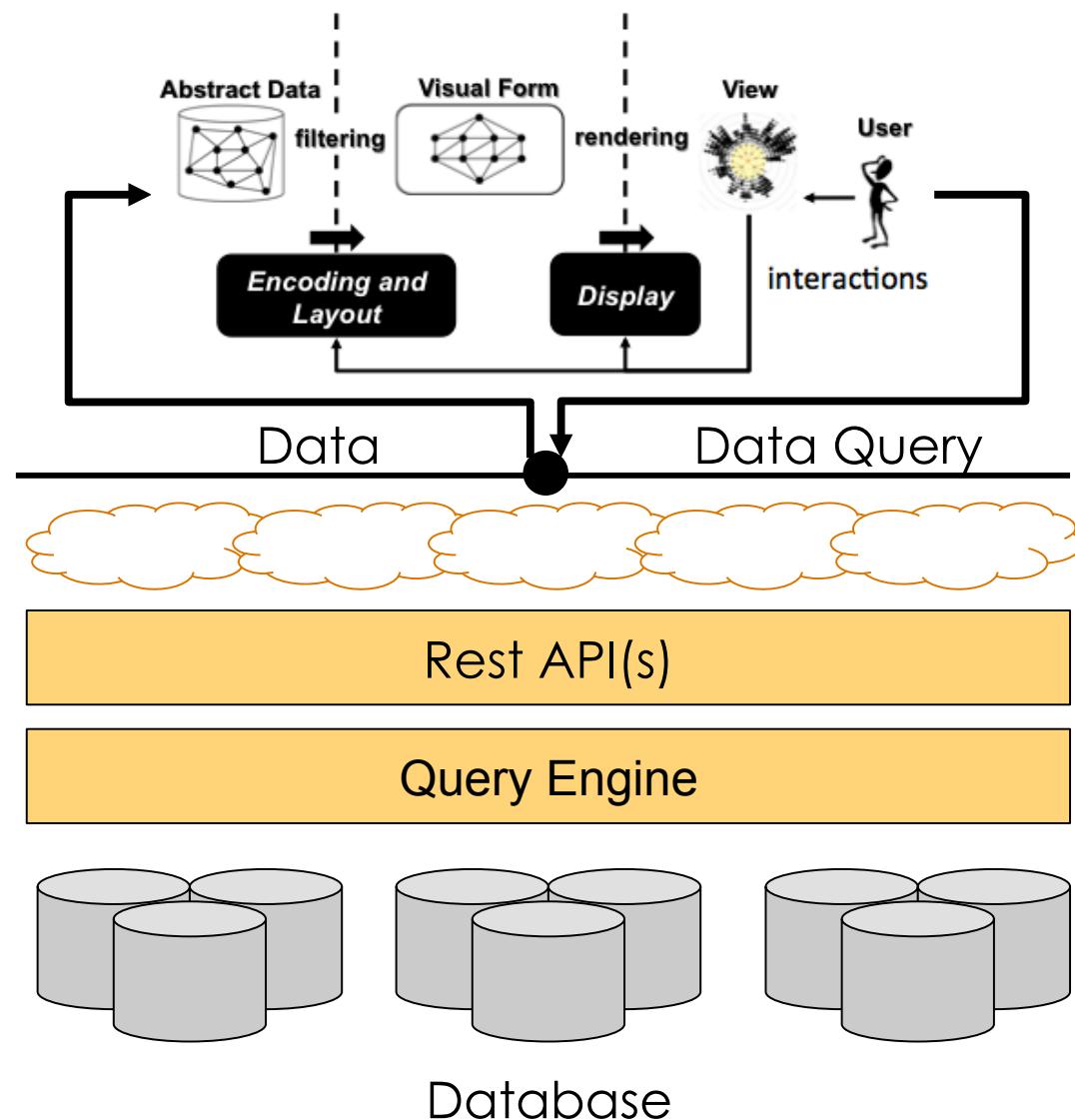
Edge Bundling



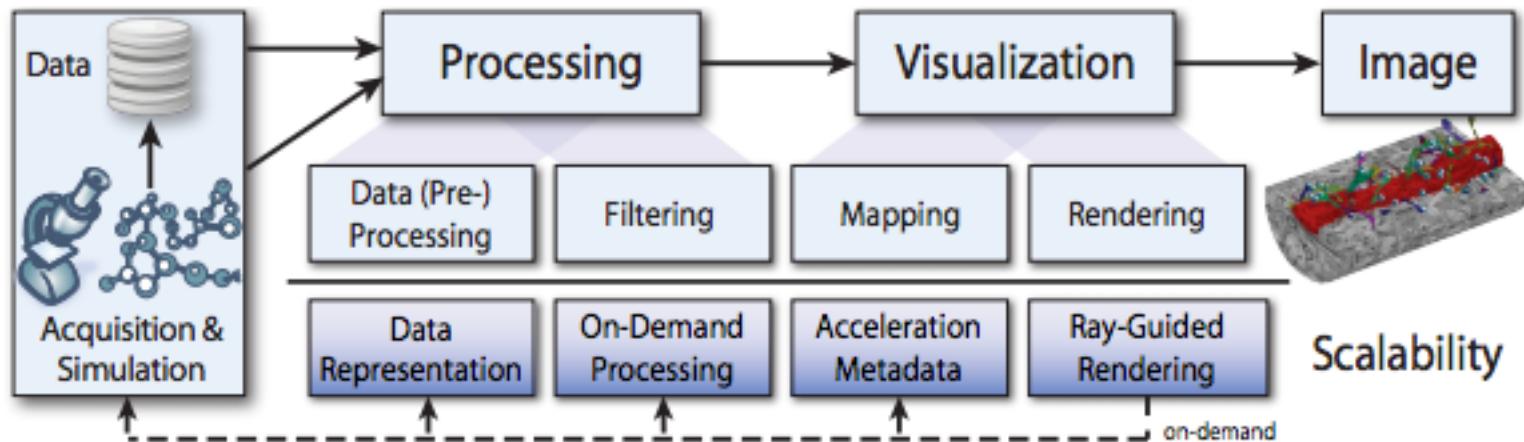
Technique (4): Clutter Reduction



Technique (4): Query based Visualization

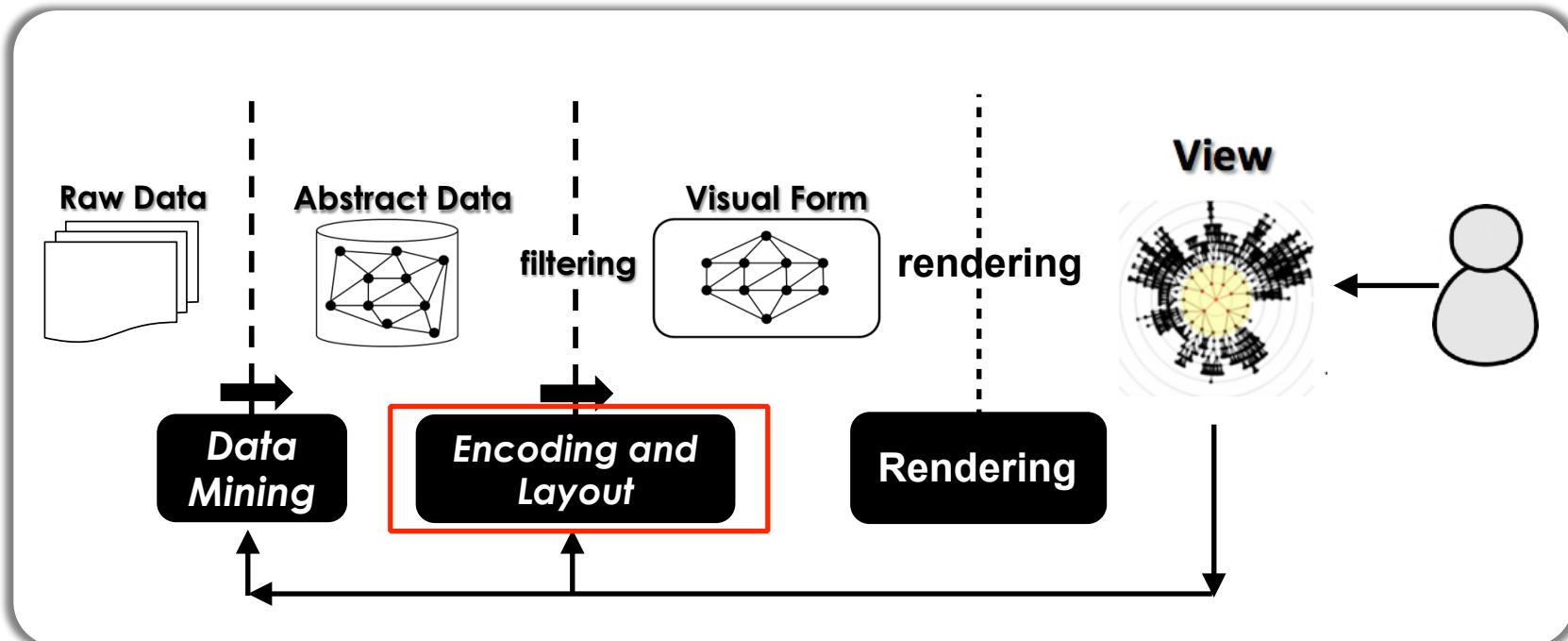


Technique (5): Parallel Computing via GPU or GUGPU



A Survey of GPU-Based Large-Scale Volume Visualization, EuroVis, 2014

Part III : **How can we visualize big data ?**



Encoding : Visual Design

Technique : Layout Algorithm

Using existing tools are easy



D3.js Data-Driven Documents

GitHub



The screenshot shows the Tableau software interface. On the left, there's a navigation sidebar with options like 'Admin', 'Content', 'Workbooks', 'Views', 'Data Sources', 'Filters', 'PROJECT Demo', 'PUBLISHER', and 'Any Publisher'. The main area displays a grid of data visualizations under categories such as 'Marketing', 'Market Statistics', 'Product Costs', 'Real Estate Dashboard', 'Sales Commission Model', and 'Sales Forecast'. Each visualization has a preview, a title, and a 'Download' button.

Tableau

The screenshot shows the ManyEyes website. It features a search bar at the top right. Below it, there's a section titled 'Try Our Featured Visualizations' with links to 'US Targets', 'US Cell Production', 'Academic papers over time', and 'Distributions of Presidential candidates as of 3/31/08'. There's also a 'many eyes' logo with a cartoon character. A central 'Explore' section lists options like 'visualizations', 'data sets', 'comments', 'topic hubs', 'participate', 'register', 'create visualization', 'upload data set', and 'create topic hub'. To the right, there are sections for 'Featured Topic Hubs' (Food Safety, Transportation, OIC Feedback 2007) and a 'new compare' feature.

ManyEyes

Python:

iGraph : <http://igraph.org/redirect.html>

Networkx : <https://networkx.github.io/>

JavaScript:

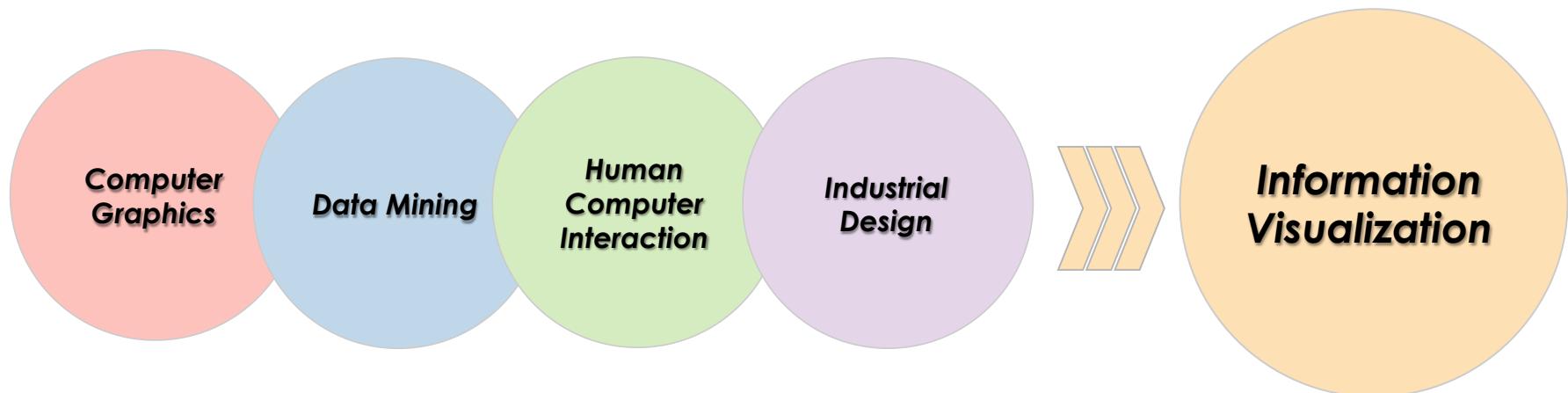
D3.js (2D, SVG): <http://d3js.org/>

Tree.js (3D, WebGL): <http://threejs.org/>

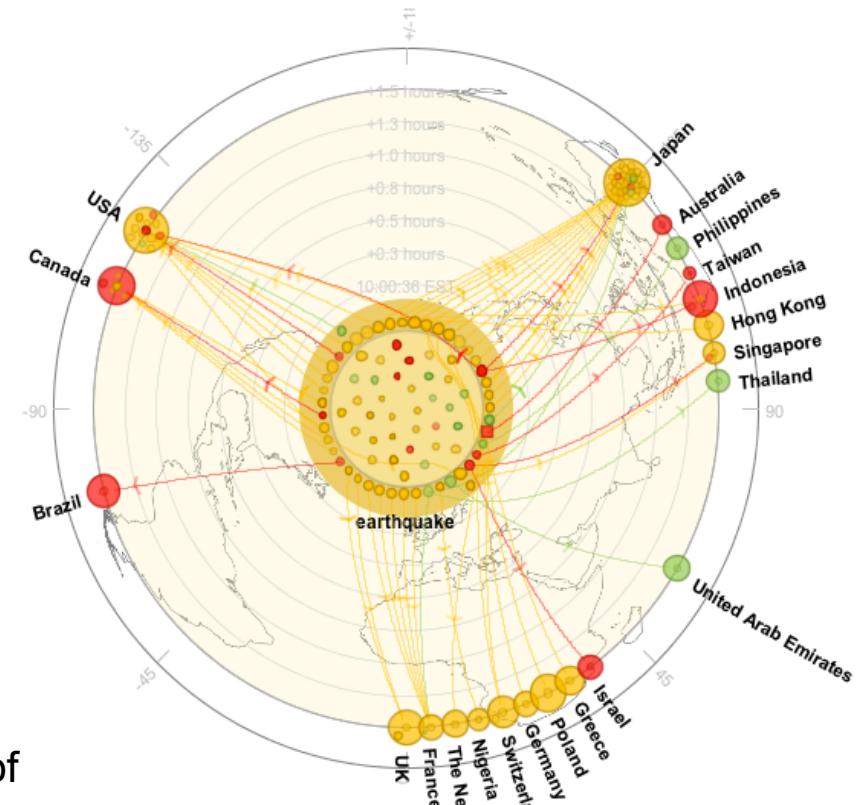
Java:

prefuse: <http://prefuse.org/>

InfoVis Toolkit: <http://ivtk.sourceforge.net/>



Example 1: Visualising Streaming Data



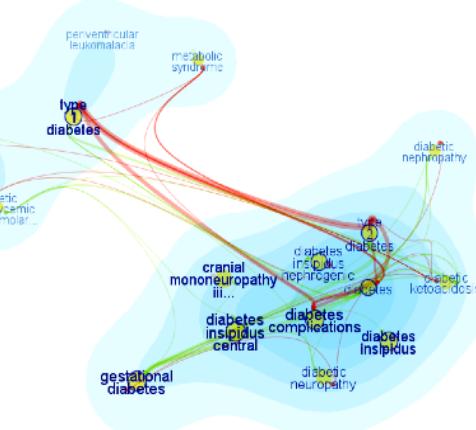
Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time
IEEE InfoVis 2012

Example 2: Visualizing Large Text Corpus



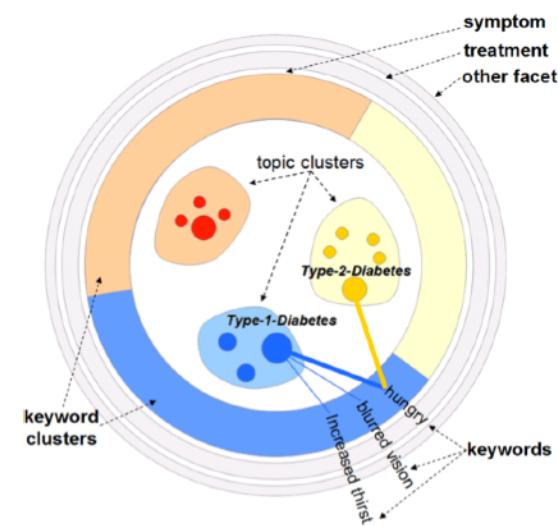
ContexTour SDM 2010

Visualizing Heterogeneous Clusters



FacetAtlas
TVCG (InfoVis 2010)

Visualizing Multi-relational Clusters



SolarMap
ICDM 2011

Cluster Interpretation

Part IV : Visual Analysis of Big Data

Computational Power



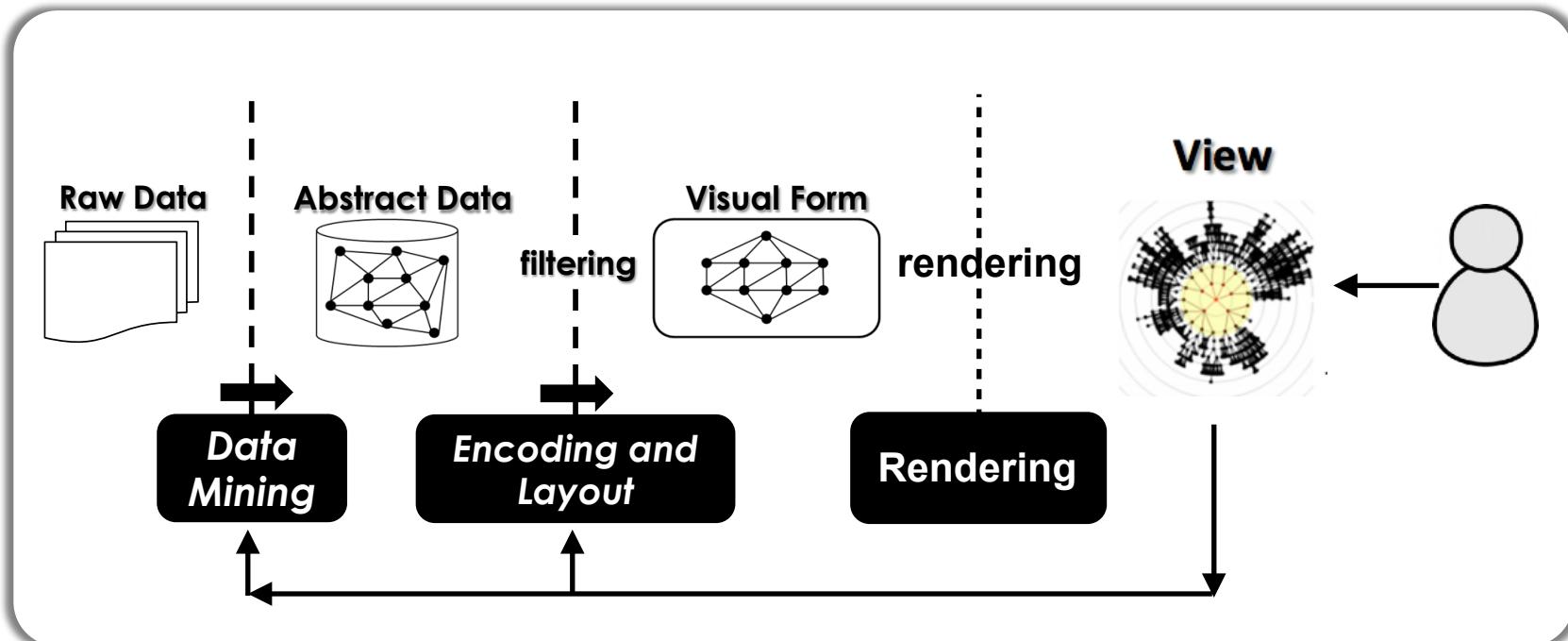
Data Mining



Human Intelligence

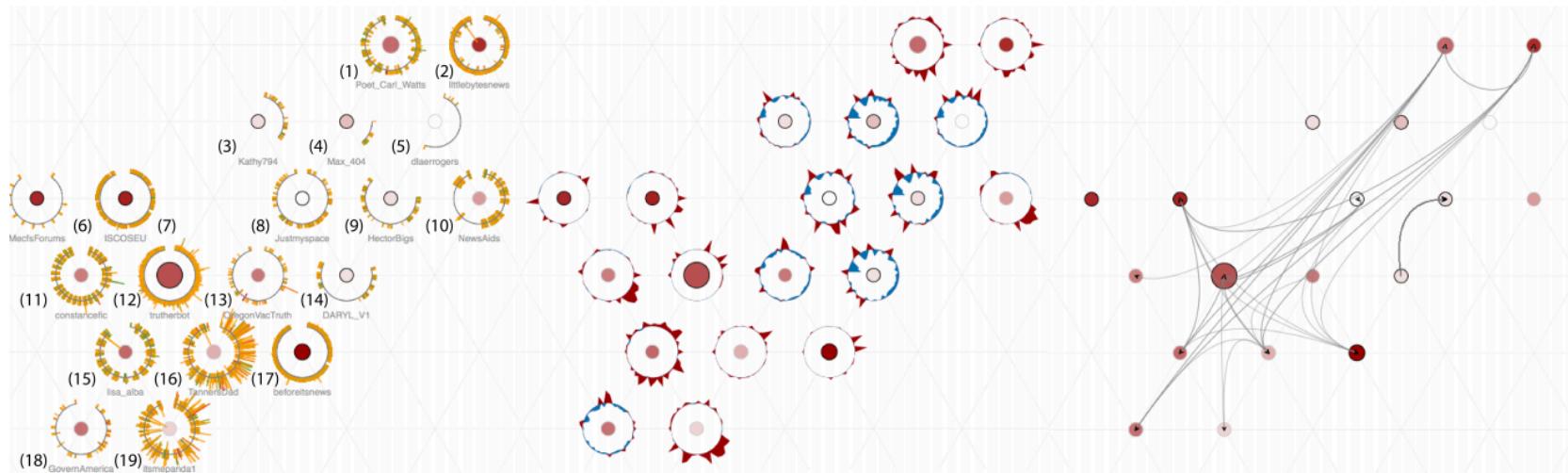


Visual Analysis



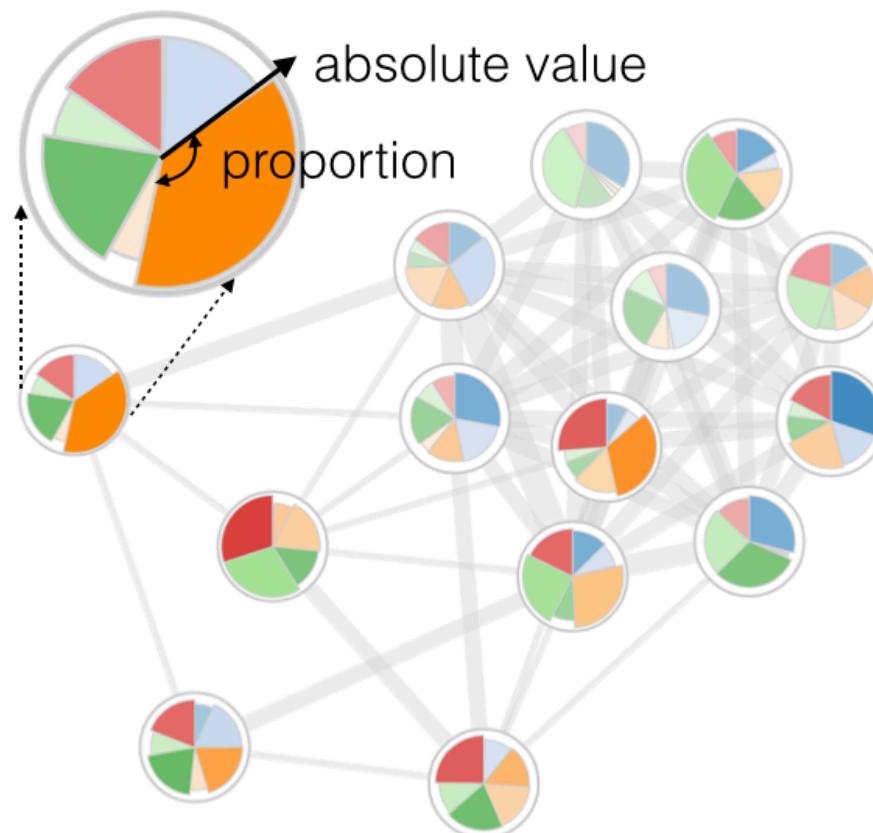
Analysis + Visualisation + Interaction

Example 3: Detect Anomalous Users in Twitter



TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems, IEEE Transactions on Visualisation and Computer Graphics (VAST'15)

Example 4: Visualizing Large Graphs



g-Miner: Interactive Visual Group Mining on Multivariate Graphs, ACM CHI 2015



Visualization Viewpoints

Editor:
Theresa-Marie Rhyne

The Top 10 Challenges in Extreme-Scale Visual Analytics

Pak Chung Wong
Pacific Northwest National Laboratory

Han-Wei Shen
Ohio State University

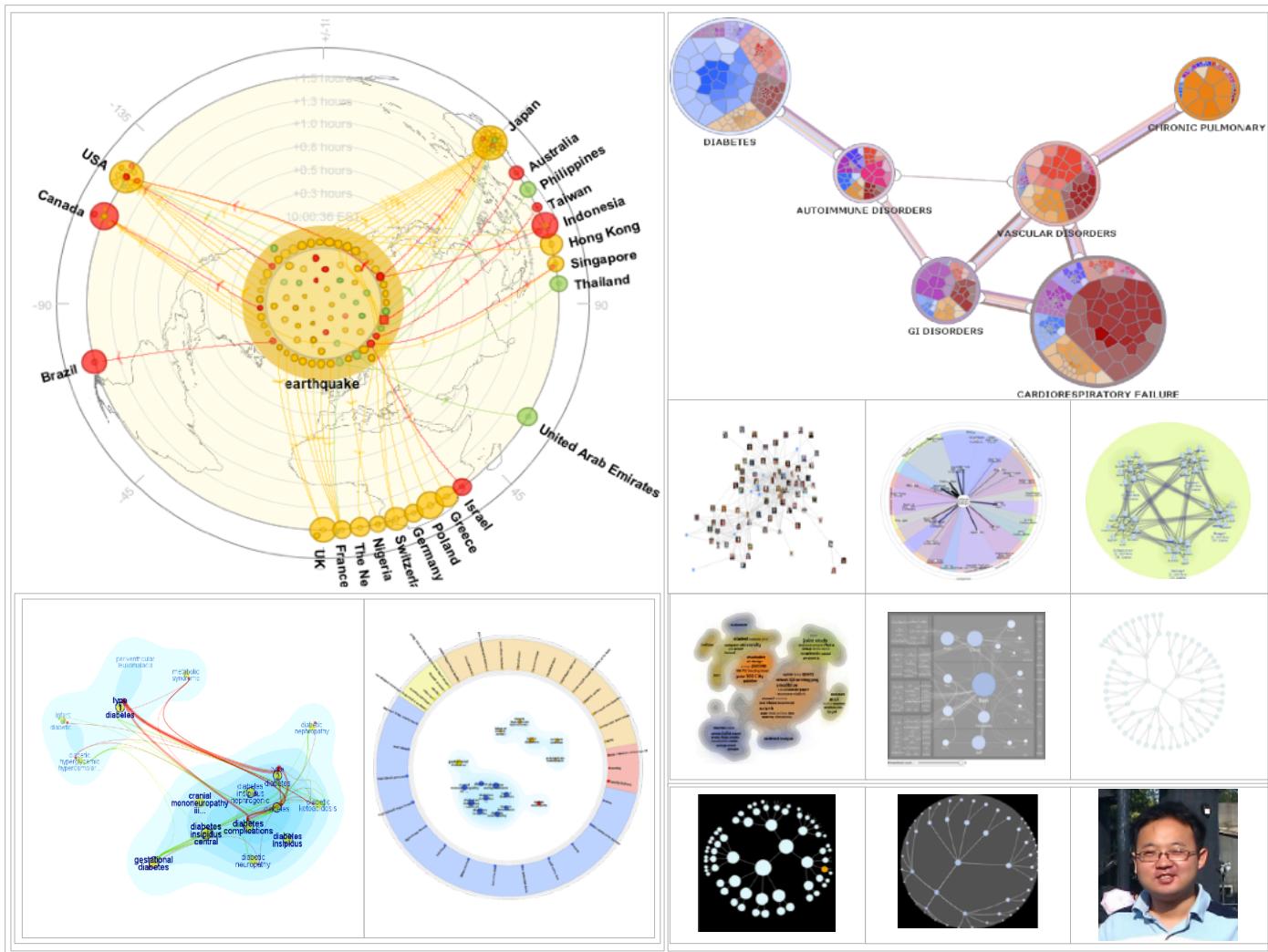
Christopher R. Johnson
University of Utah

Chaomei Chen
Drexel University

Robert B. Ross
Argonne National Laboratory

Wong, P. C., Shen, H. W., Johnson, C. R., Chen, C., & Ross, R. B. (2012). The top 10 challenges in extreme-scale visual analytics. *IEEE computer graphics and applications*, 32(4), 63.

Thank You!



Thanks to Dr. Nan Cao <http://nancao.org>

Homework #3 (Due 11/10/2023, 5pm)

Please see the assignment slides

Final Project Proposal Scoring

Proposal — preparing about 5 pages (each item 1/5 of the proposal score):

Goal — novel? challenging?

Data — 3Vs? New dataset? Existing dataset?

Methods — planning of methodologies and algorithms? Feasible?

System — an overview of system. What will be implemented?

Schedule — what to achieve by what time, and by whom?

Final Project Report Scoring

- **Title, Author(s)**
- **Abstract:** Briefly describe your problem, approach, and key results.
- **Introduction (5%)**: Describe and define the problem you are working on. Why is it important? Include an overview of your methods and results.
- **Related Work (5%)**: Discuss published works or approaches that are related to your project. What's the benefit or drawback of the previous works? What kind of problems have they solved? How is your approach similar or different from others?
- **Data (10%)**: Describe the data you are working with for your project. What type of data is it? Where did it come from? How much data are you working with? Did you have to do any preprocessing, filtering, feature engineering or other special treatment to use this data in your project?
- **Methods (25%)**: Discuss your approach for solving the problems that you set up in the introduction. Why is your approach the right thing to do? Did you consider alternative approaches? Have you tried some methods that didn't work out? It may be helpful to include figures, diagrams, or tables to describe your method or compare it with other methods.
- **Experiments (20%)**: Discuss the experiments that you performed to demonstrate your approach solves the problem. The experiments will vary depending on the project, but you might compare with previously methods, determine the impact of the components of your system, experiment with different hyper-parameters, architectures, or algorithms, use visualization techniques to gain insight of how your model works, etc. Graphs, tables, and figures are highly recommended to be included to illustrate your experimental results.
- **System Overview (25%)**: Describe the software architecture and tech stacks of your application. Discuss potential bottlenecks and improvements that could be made. Mention the software packages that you used. Mention how to use your application. You could provide screenshots to your application.
- **Conclusion (5%)**: Summarize your key results. What have you learned? What problems have you discovered and solved? Suggest ideas for future extensions or new applications.
- **Writing / Formatting (5%)** Is your paper clearly written and nicely formatted?