CS 644: Introduction to Big Data

Daqing Yun (daqing.yun@njit.edu)
New Jersey Institute of Technology

Jaguar SC@ORNL

https://flic.kr/p/7fkQ2c

Big Data Visualization

- Introduction to Computer Graphics
  - Images and Displays
  - Ray Tracing
- Big Data Visualization
  - Scientific (3D Volume) Visualization
    - Ray Casting
    - Marching Cubes
  - Information Visualization
  - Challenges and Techniques

What is Computer Graphics (CG)?
What can we do with CG?

Video games


0
Menu
50
100
75
25
0
150
150
200
39,560
Level 2-8

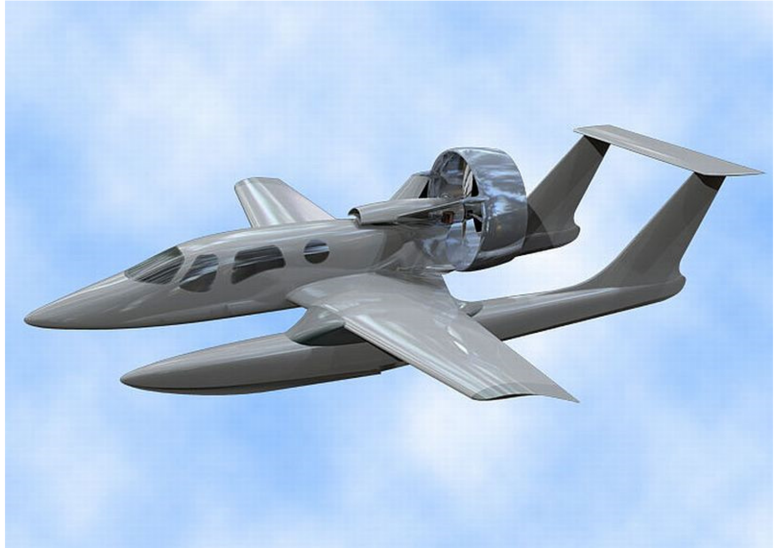

SCORE
430
HIGHSCORE
80100

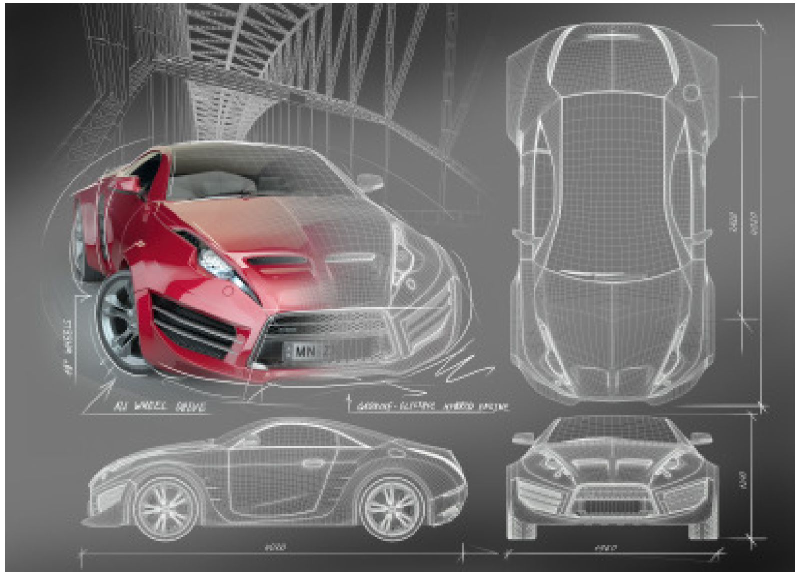

132

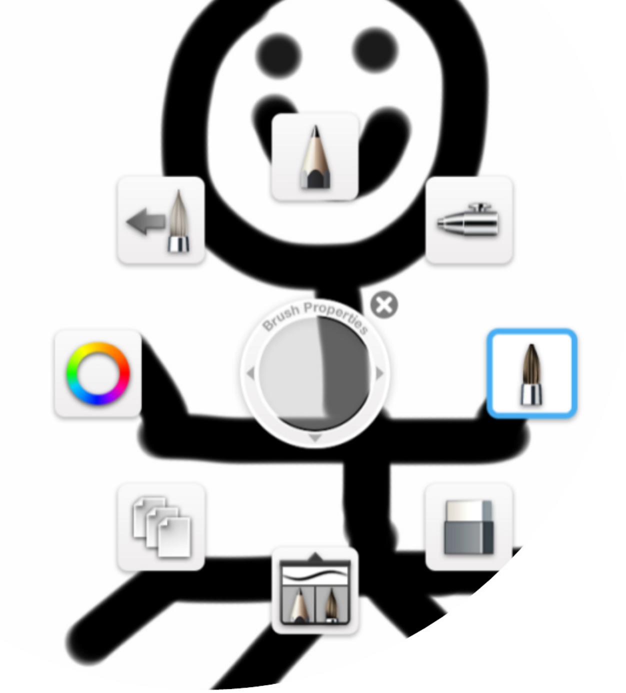
Cartoons and animated films

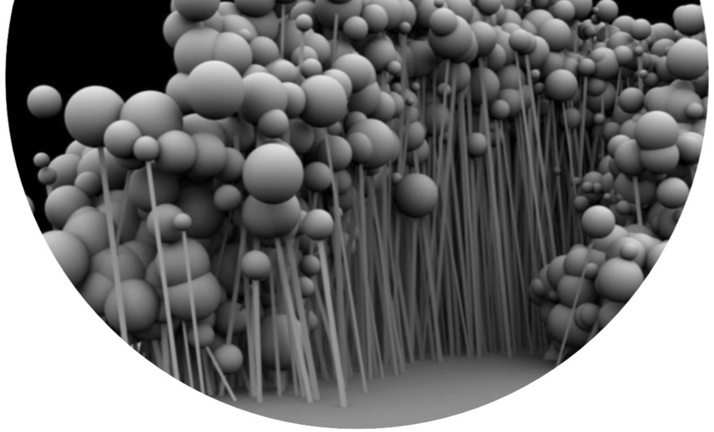



CAD/CAM



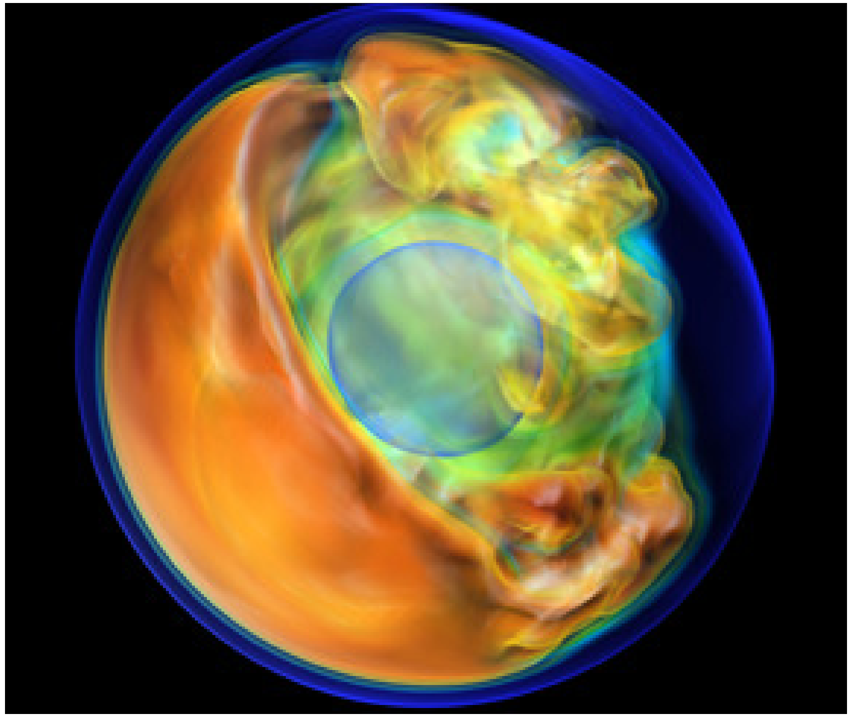






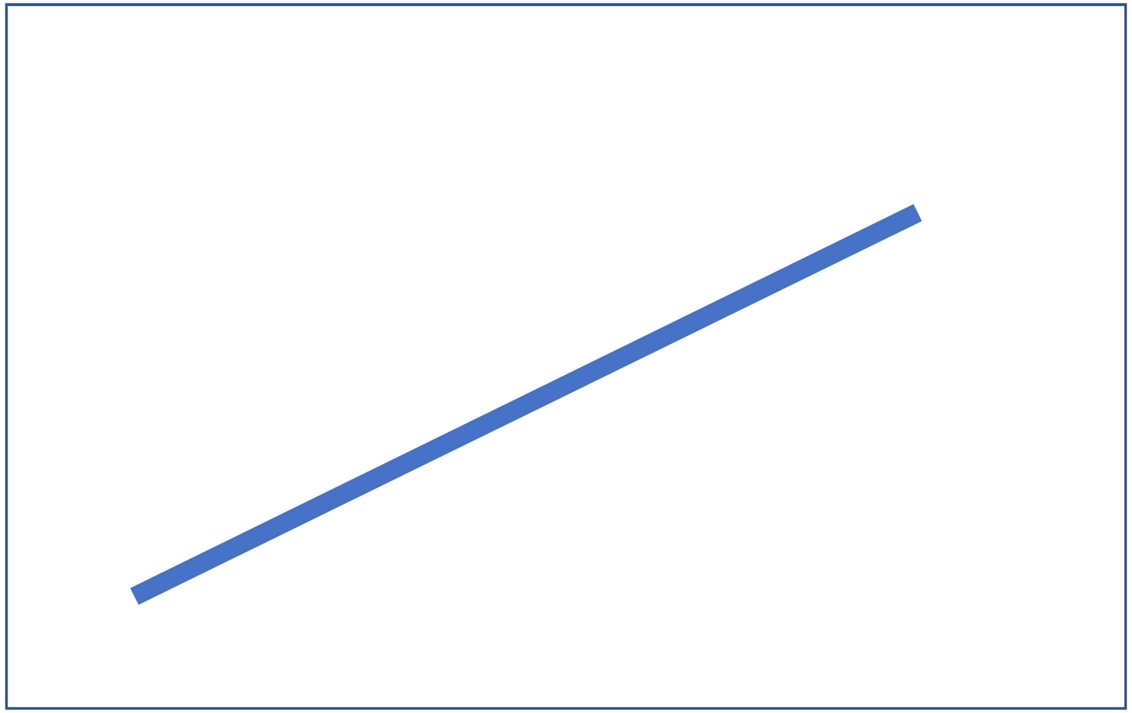
Artworks

Visualization



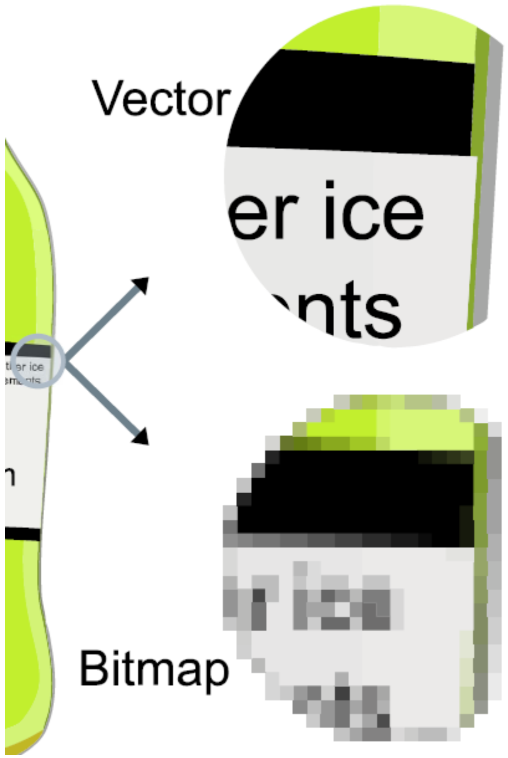
Size of each Cell: Stock Market Value
Color: Stock Change

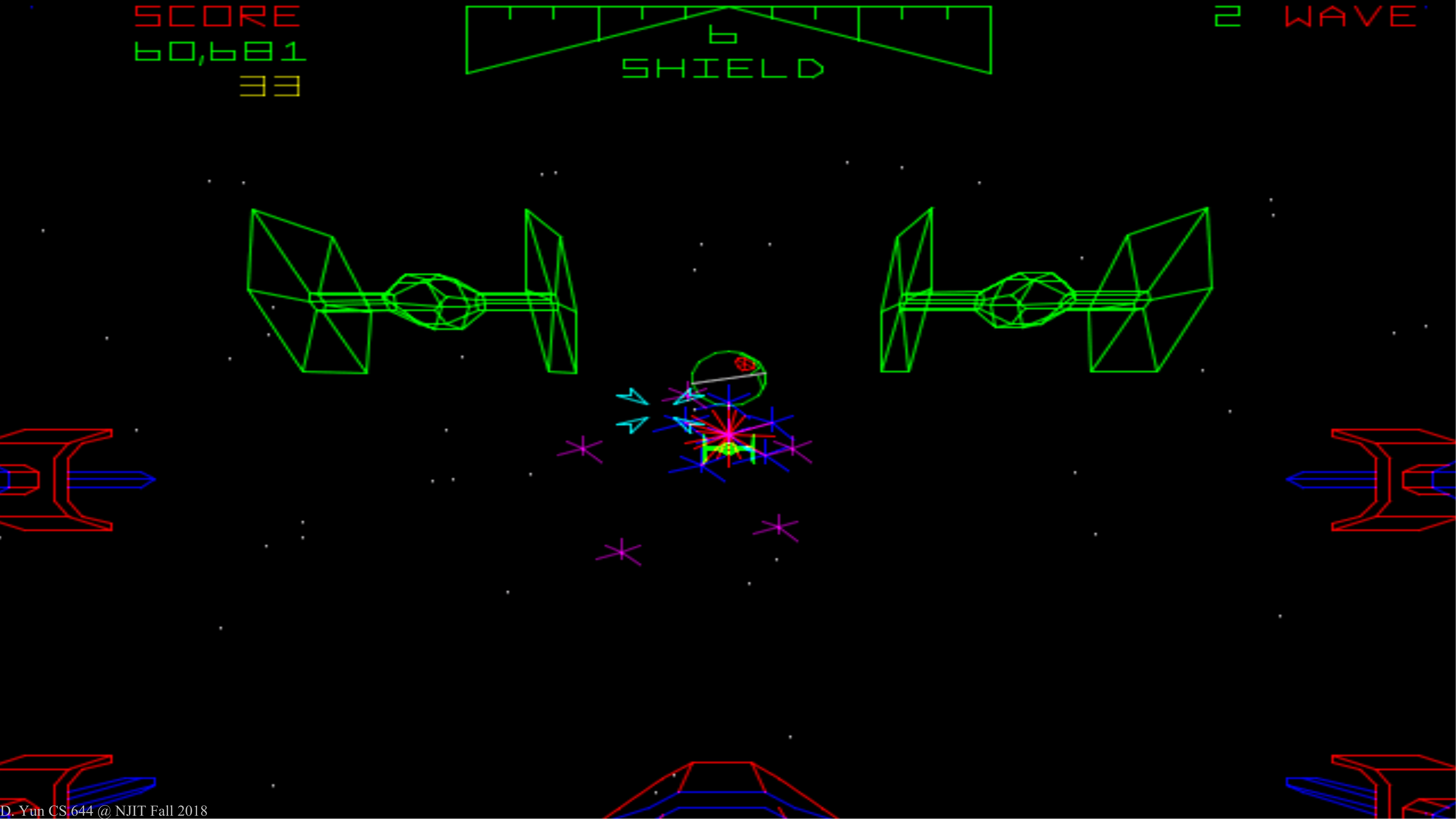
What is an image?



There are two ways to represent images

- Vector image
  - Use instructions to describe the shapes (lines or curves) with no reference to any particular pixel grid -- A simple line segment: Start (0,0), End (5,3)
- Advantage
  - Resolution independent, so can be displayed well on very high resolution devices
  - Require very little memory
  - No aliasing of lines/curves
- Disadvantage
  - Can only draw line segments -- More lines, more time needed
  - Must be rasterized before they can be displayed
    - Rasterization: converting a vector image (shapes) to a raster image (dots)
    - Virtually all displays used today are raster displays
    - Dots are the only things modern displays can understand
- Used for
  - Text, diagrams, mechanical drawings
  - Other applications
    - Crispness and precision are important
    - Photographic images and complex shading are not needed



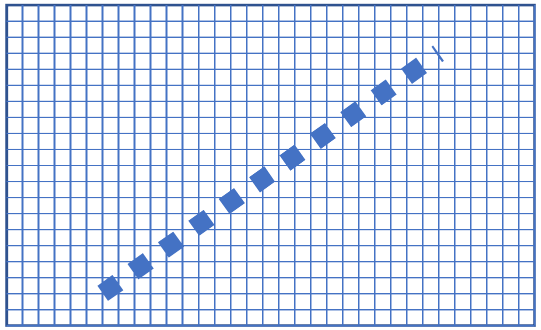

SCORE
60,681
33
6
SHIELD
2 WAVE

There are two ways to represent images

- Raster image
  - A 2D distribution of intensity or color (pixel?)
  - A function defined in 2D plane

$$I : \mathbb{R}^2 \rightarrow \ldots$$

  - A natural representation



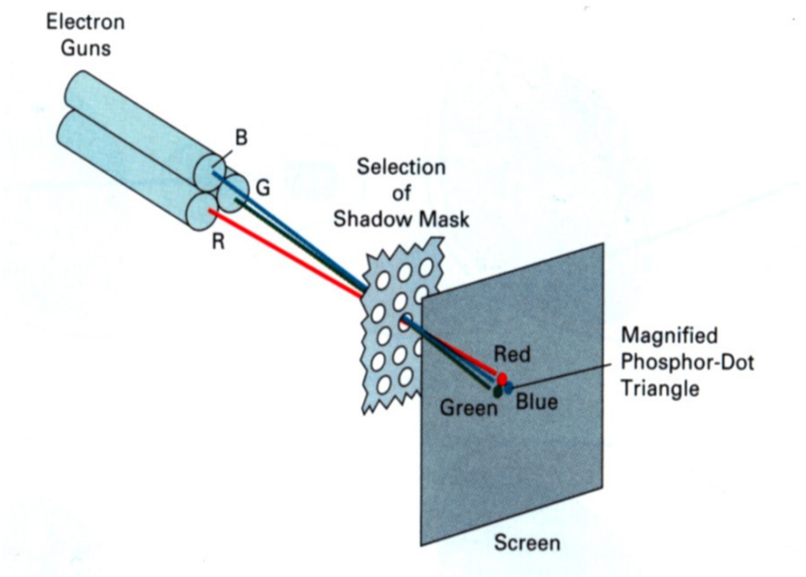
- To do graphics, we must
  - Represent images: encode them numerically
    - Vector or raster
  - Display images: realize them as actual intensity distributions
    - Various display devices

Representative Display Technologies

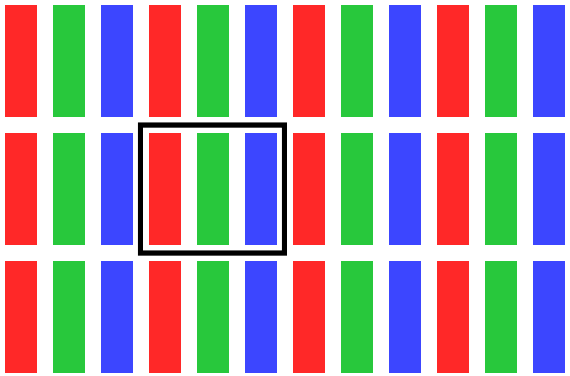
- Computer displays
  - Raster CRT display
  - LCD display
- Printers
  - Laser printer
  - Inkjet printer

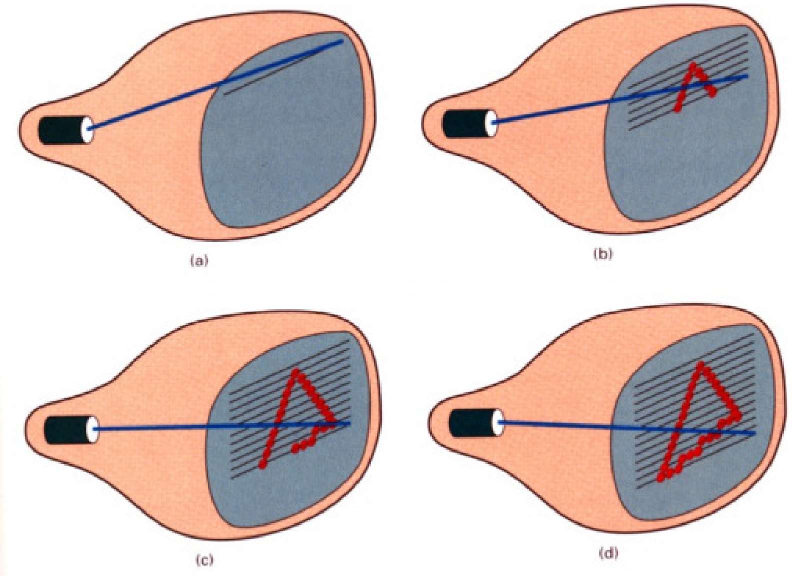
Color Displays

- CRT
  - Phosphor dot to produce finely interleaved color images



- LCD
  - Interleaved R,G,B pixels



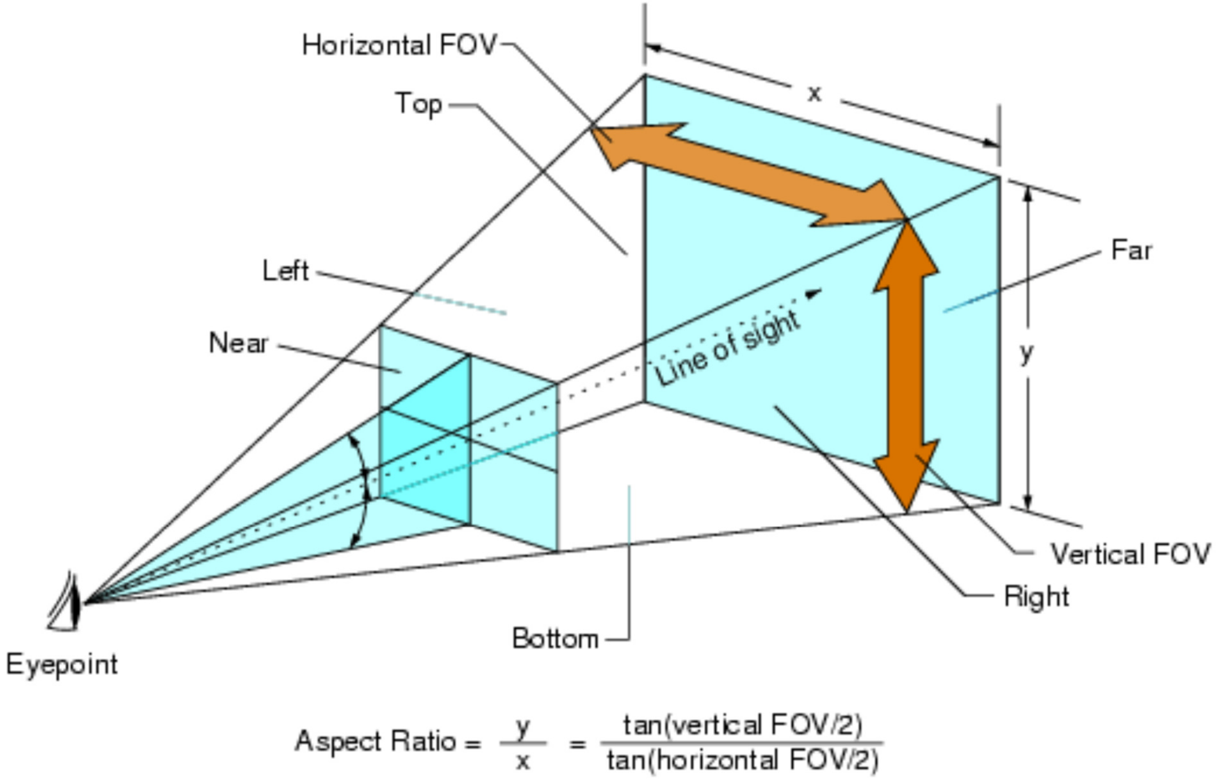


Raster Image Representation

- All these devices suggest 2D arrays of numbers
- Big advantage: represent arbitrary images
  - Approximate arbitrary functions by increasing resolution
    - Just need more memory for more pixels
  - Works because memory is cheap (brute force approach!)

Raster Image Representation

- Disadvantage
  - Memory demand
    - Draw the whole screen “at once”
    - Need a framebuffer to hold the information for the whole image
  - Aliasing
    - This is what causes “jaggies”
    - The incoming signal (the desired image) can only be sampled at pixel centers on the display
      - Image is a sampled representation (image reconstruction)
    - Pixel means “this is the intensity around here”
      - LCD: intensity is constant over square regions
      - CRT: intensity varies smoothly across pixel grid



- The physical world (real-life objects) is 3D
- The display is, virtually always, only 2D
- Projection: transform 3D model into 2D model

How do we draw an image?

Rendering Process

- Start
  - Original 3D geometric model
- Shading
  - Compute color of original geometry
    - Based on lighting and surface color
- Projection
  - Project original 3D geometry to 2D model
- Clipping
  - Clip original geometry outside FOV
- Rasterization
  - Generate fragments from projected 2D model
- Fragment processing
  - Compute pixel colors from fragments
- End
  - Display pixels

Data Types for Raster Images

- Bitmaps: **`boolean`** per pixel (1 bpp): $I : \mathbb{R}^2 \rightarrow \{0, 1\}$
  - black and white; e.g., fax

- Grayscale: integer per pixel: $I : \mathbb{R}^2 \rightarrow [0, 1]$
  - shades of gray; e.g., black-and-white print
  - precision: usually **`byte`** (8 bpp); sometimes 10, 12, or 16 bpp

- Color: 3 integers (RGB) per pixel: $I : \mathbb{R}^2 \rightarrow [0, 1]^3$
  - full range of displayable color; e.g., color print
  - precision: usually **`byte[3]`** (24 bpp)
  - sometimes 16 (5+6+5), 30, 36, 48 bpp

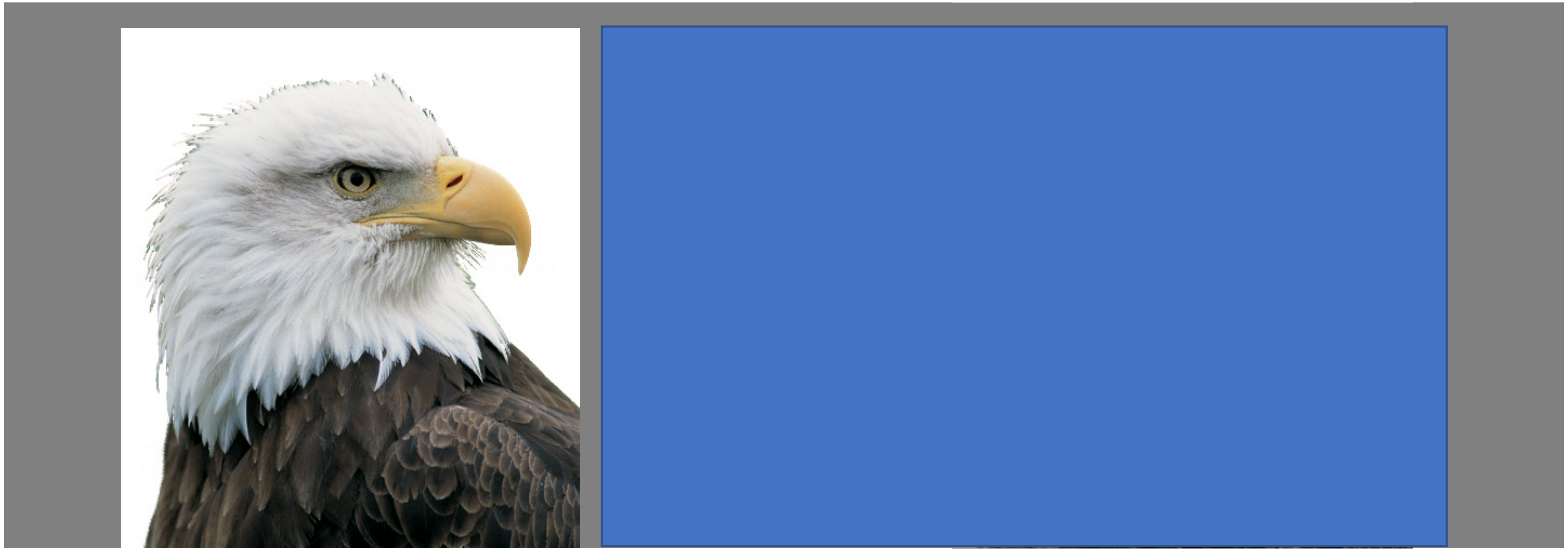
Data Types for Raster Images

- Floating point: $I : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ or $I : \mathbb{R}^2 \rightarrow \mathbb{R}_+^3$
  - more abstract, because no output device has infinite range
  - provides *high dynamic range* (HDR)
  - represent real scenes independent of display
  - becoming the standard intermediate format in graphics processors

- Clipping
  - first compute floating point (FP), then convert to integer
  - full range of values may not "fit" in display's output range
  - simplest solution: choose a maximum value, scale so that value becomes full intensity (2^n–1 in an n-bit integer image)

Data Types for Raster Images

- For color or grayscale, sometimes add *alpha* channel
  - describes transparency of images



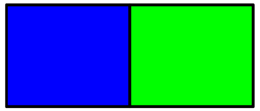
Storage Requirements for Images

- 1024 x 1024 image (1 megapixel)
  - bitmap:
  - grayscale 8bpp:
  - grayscale 16bpp:
  - color 24bpp:
  - floating-point HDR color:
- What is the resolution of your camera?

Converting Pixel Formats

- Color to gray
  - could take one channel (blue, say)
    - leads to odd choices of gray value
  - combination of channels is better
    - but different colors contribute differently to lightness
    - which is lighter, full blue or full green?

Same pixel values.

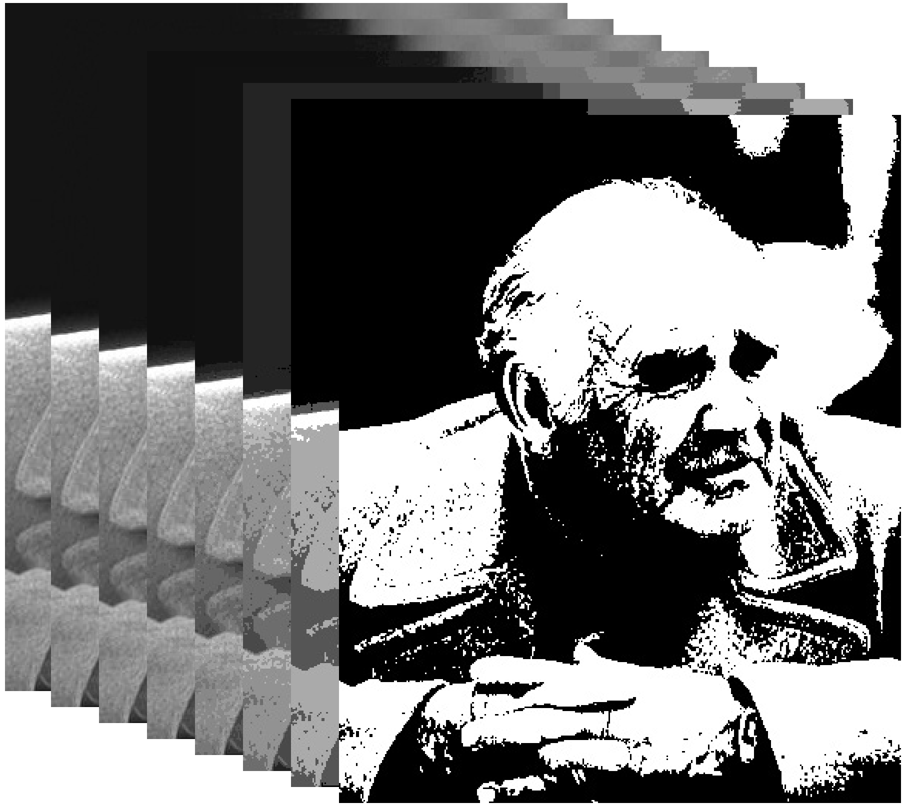


Same luminance?

    - good choice: gray = 0.2 R + 0.7 G + 0.1 B

Converting Pixel Precision

- Up is easy; down loses information — be careful



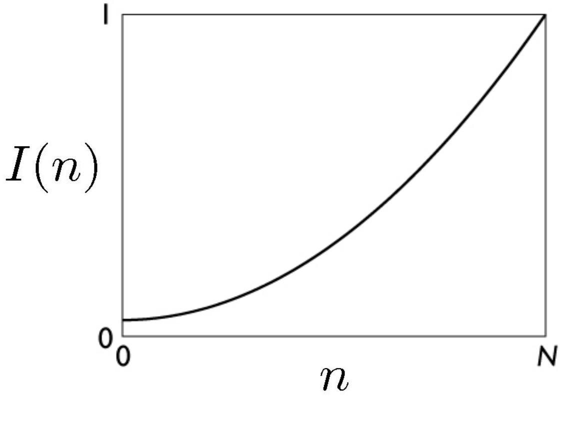
1 bpp (2 grays)

Intensity Encoding in Images

- What do the numbers in images (pixel values) mean?
  - they determine how bright that pixel is
  - bigger numbers are (usually) brighter
- *Transfer function*: function that maps input pixel value to luminance (intensity) of displayed image

$$I = f(n) \quad f : [0, N] \to [I_{\min}, I_{\max}]$$

- What determines this function?
  - physical constraints of device or medium
  - desired visual characteristics



Constraints on Transfer Function

- Maximum displayable intensity, I_{max}
  - how much power can be channeled into a pixel?
    - LCD: backlight intensity, transmission efficiency (<10%)
    - projector: lamp power, efficiency of imager and optics
- Minimum displayable intensity, I_{min}
  - light emitted by the display in its "off" state
    - LCD: polarizer quality
    - CRT: stray electron flux
- Viewing flare k: light reflected by the display
  - very important factor determining image contrast in practice
    - 5% of I_{max} is typical in a normal office environment
    - requires much effort to make very black CRT and LCD screens

Dynamic Range

- Dynamic range:

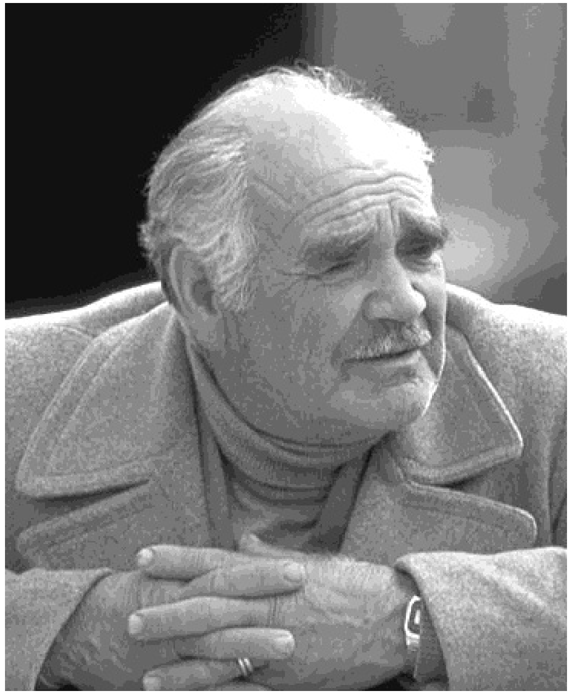
$$R_d = I_{max} / I_{min} \quad \text{or} \quad (I_{max} + k) / (I_{min} + k)$$

  - determines the degree of image contrast that can be achieved
  - a major factor in image quality!

- Ballpark values of common display devices
  - Desktop display in typical conditions: 20:1
  - Photographic print: 30:1
  - Desktop display in good conditions: 100:1
  - Photographic transparency (directly viewed): 1000:1
  - High Dynamic Range (HDR) display: 10,000:1

Banding : noticeable intensity change between neighboring pixels

- Desirable property: the change from one pixel value to the next highest pixel value should not produce a visible contrast
  - otherwise smooth areas of images will show visible bands
- What contrasts are visible?
  - rule of thumb: under good conditions we can notice a **2%** change in intensity
  - we generally need smaller quantization steps in the darker tones than in the lighter tones (why?)
    - Darker tones have a lower intensity value
    - A smaller denominator leads to a higher percentage change
  - most efficient quantization is logarithmic



an image with severe *banding*

How many levels (pixel value range) are needed?

- Depends on dynamic range
  - 2% steps are most efficient:

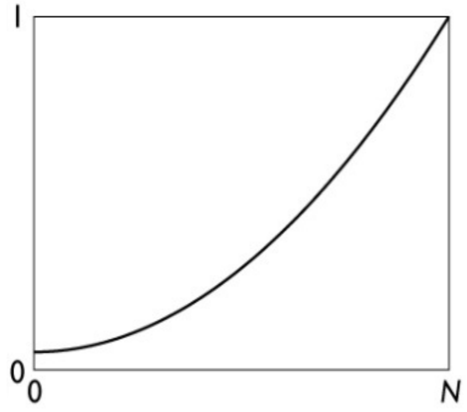
$$0 \mapsto I_{\min};\, 1 \mapsto 1.02 I_{\min};\, 2 \mapsto (1.02)^2 I_{\min};\, \ldots \quad N \to (1.02)^N I_{\min}$$

  - How many steps (levels) needed per decade (10:1) of dynamic range?

$$R_d = \frac{I_{\max}}{I_{\min}} = \frac{(1.02)^N I_{\min}}{I_{\min}} = (1.02)^N = \frac{10}{1} = 10$$

$$\log(1.02)^N = N \log(1.02) = \log 10 = 1$$

$$N = \frac{1}{\log(1.02)} = \frac{1}{1/120} = 120$$

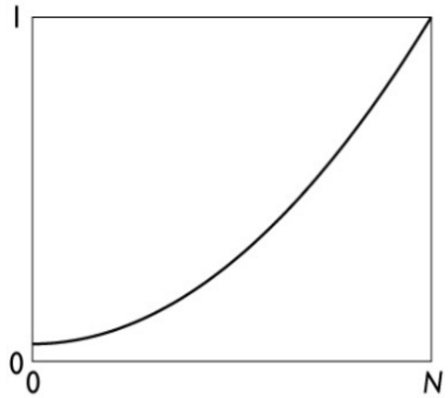


    - 240 for desktop display with R_d 100:1
    - 360 to print to film with R_d 1000:1
    - 480 to drive HDR display with R_d 10,000:1

How many levels (pixel value range) are needed?

- If we want to use linear quantization (equal steps), how many levels are needed for R_d=10?
  - one step must be < 2% (1/50) of I_{min}
  - need to get from ~0 to $I_{min} \cdot R_d$, so need about 50 R_d levels

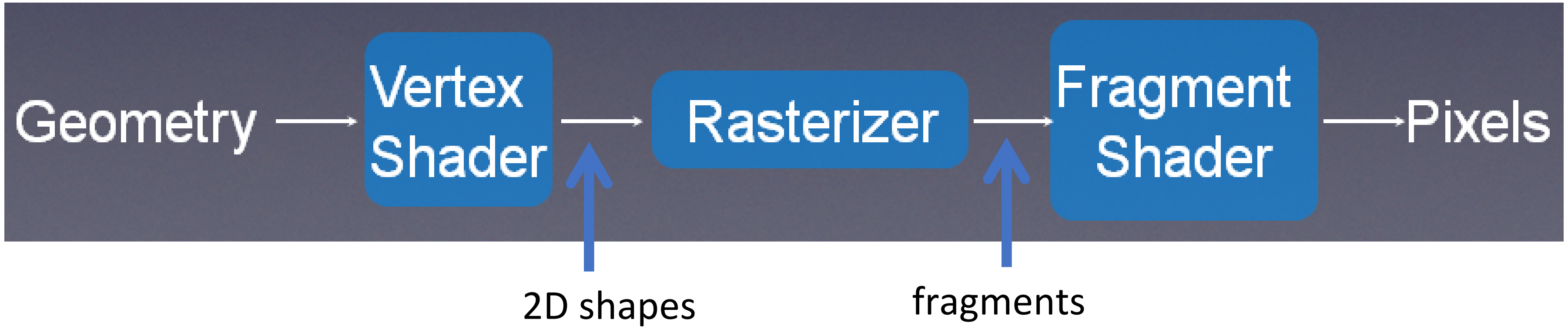
$$\frac{I_{\max} - I_{\min}}{step\ size} = \frac{I_{\min} \cdot R_d - 0}{2\% \cdot I_{\min}} = 50 R_d$$



  - 1500 for a print with R_d 30:1
  - 5000 for desktop display with R_d 100:1
  - 500,000 for HDR display with R_d 10,000:1
- Moral: 8 bits (256 levels) is just barely enough for low-end applications
  - but only if we are careful about quantization

How do we view the world?

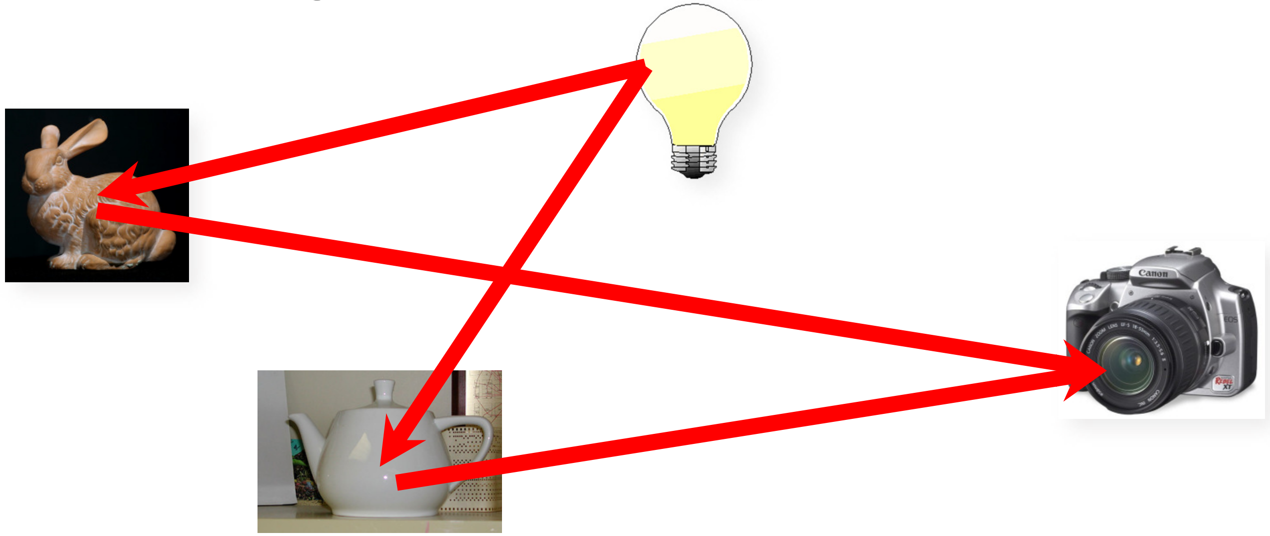
- From the perspective of the graphics pipeline



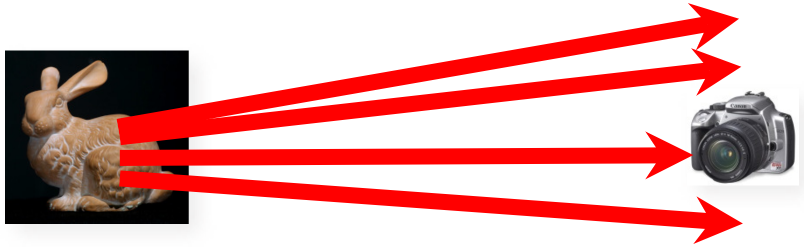
- Light, surface, and camera
  - Light
    - determines the color of the surface
  - Surface
    - represents the 3D geometry in the scene
  - Camera
    - projects the 3D geometry onto the 2D view plane

A Model of the Universe

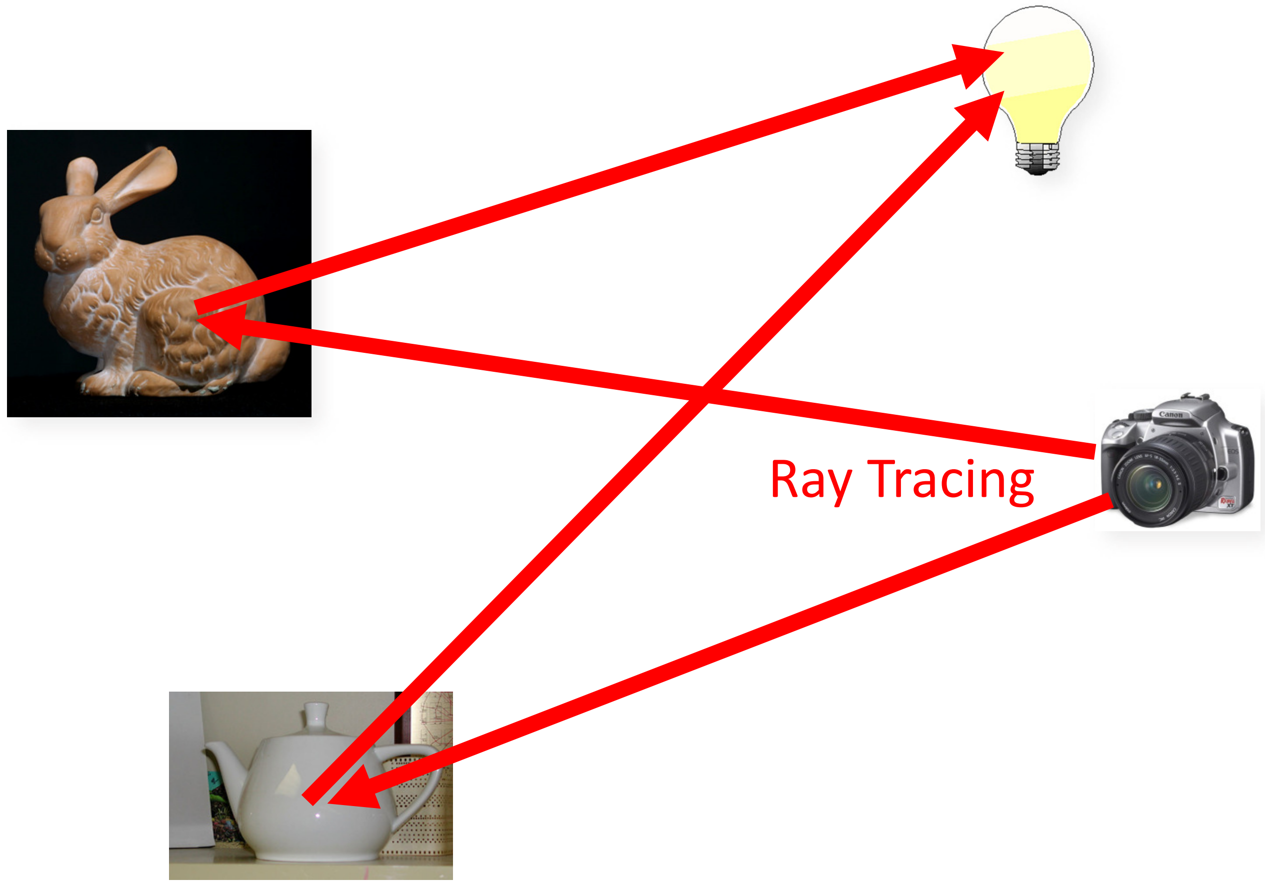
- Implement a straightforward algorithm based on this model



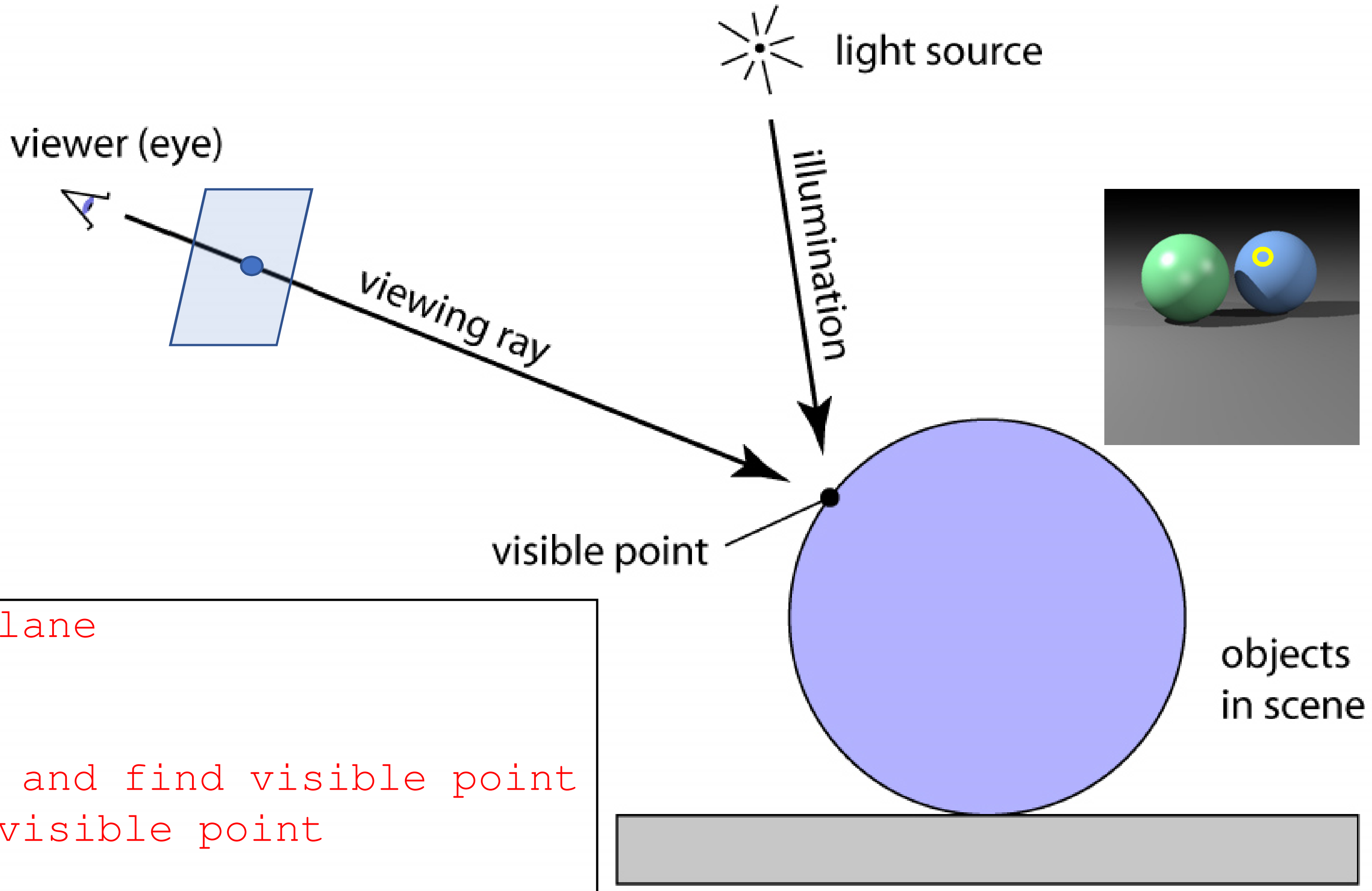
- What's the biggest issue with this model?
  - Inefficient
    - Many (probably most) light rays in a scene would never hit the image plane



A Solution: Ray Tracing!



Ray Tracing Algorithm

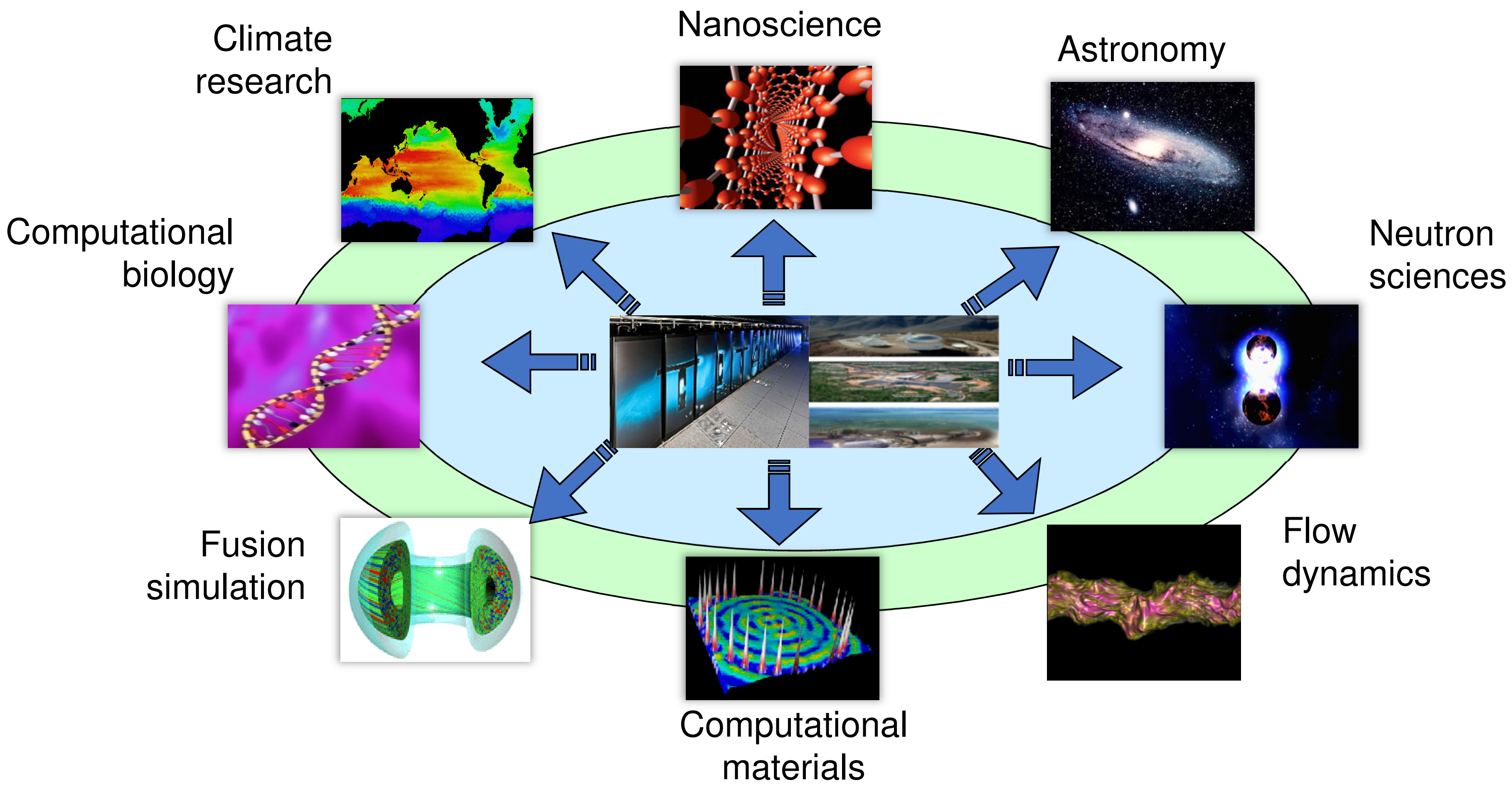



```
for each pixel in the 2D view plane
{
    1. compute a viewing ray
    2. intersect ray with scene and find visible point
    3. compute illumination at visible point
    4. put result into image
}
```

Big Data Visualization

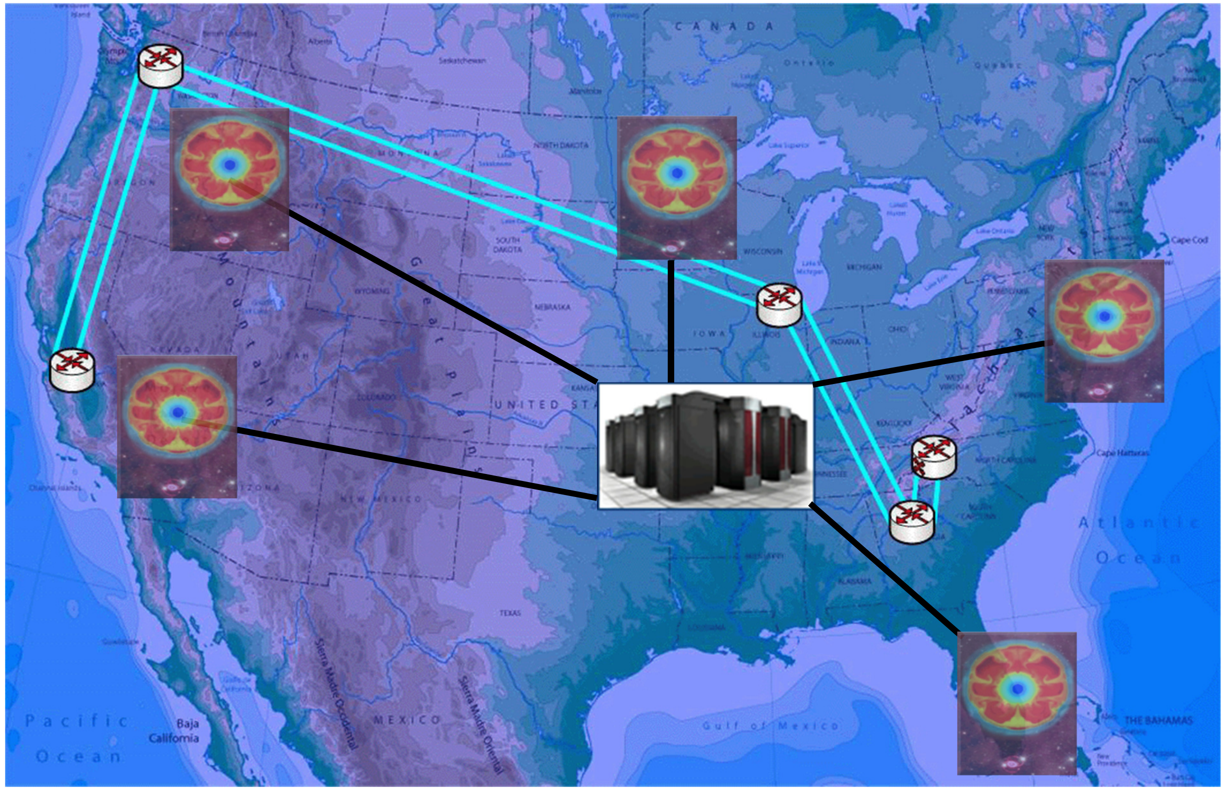
- Scientific (Volume) Visualization
- Information Visualization

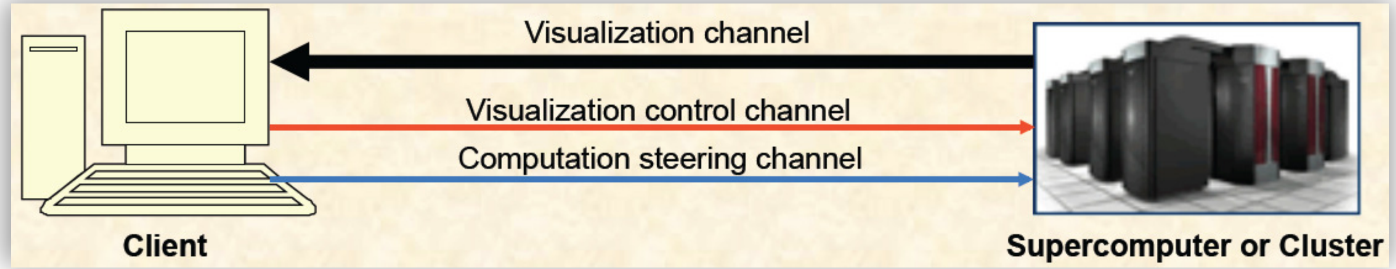
Scientific Visualization



Terascale Supernova Initiative (TSI)

- Collaborative project
  - Supernova explosion
- TSI simulation
  - 1 terabyte a day with a small portion of parameters
  - From TSI to PSI
- Transfer to remote sites
  - Interactive distributed visualization
  - Collaborative data analysis
  - Computation monitoring
  - Computation steering

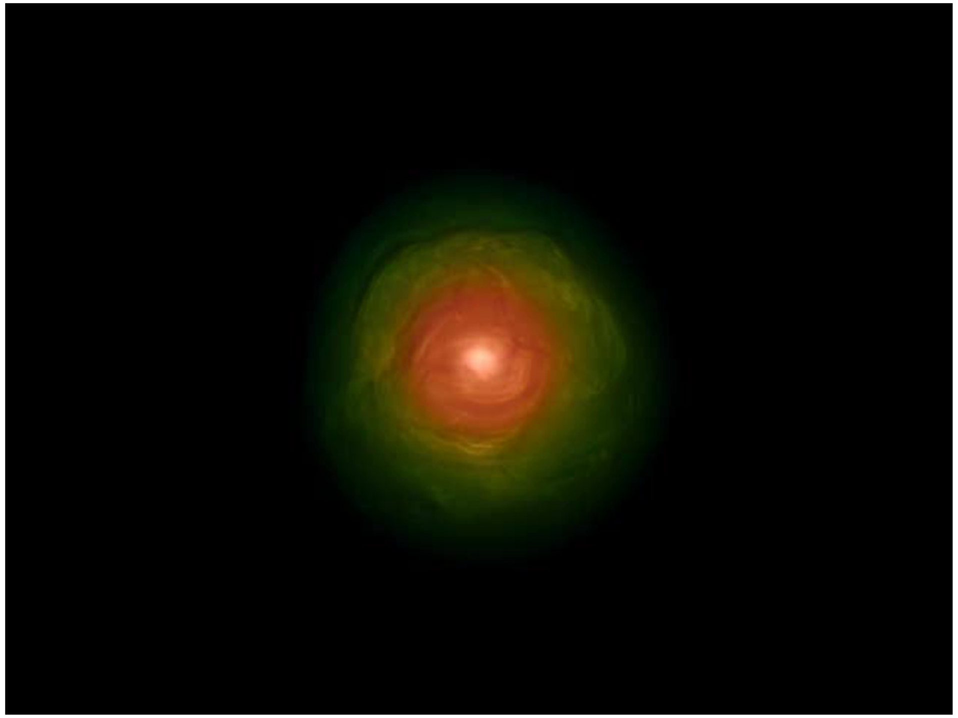




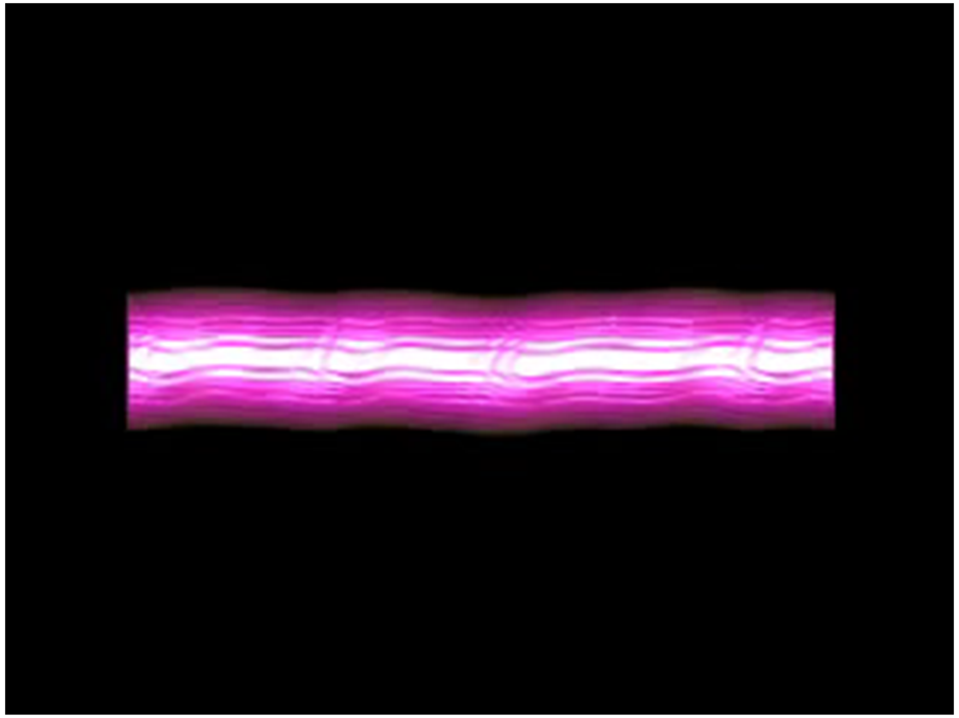
A Prototype System: Distributed Remote Intelligent Visualization Environment (DRIVE)

- Two Examples in the Visualization of Large-scale Scientific Applications

TSI explosion (density, raycasting)

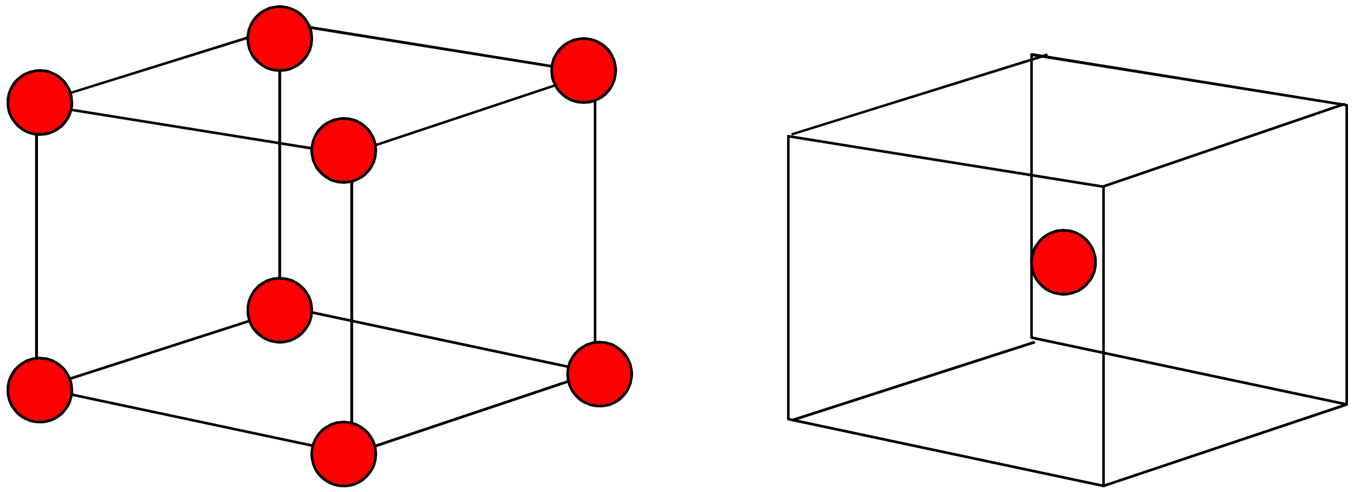


Jet air flow dynamics (pressure, raycasting)



Wu et al. Self-adaptive configuration of visualization pipeline over wide-area networks. IEEE Transactions on Computers, 2008

Volume Rendering
Samples in a volume dataset



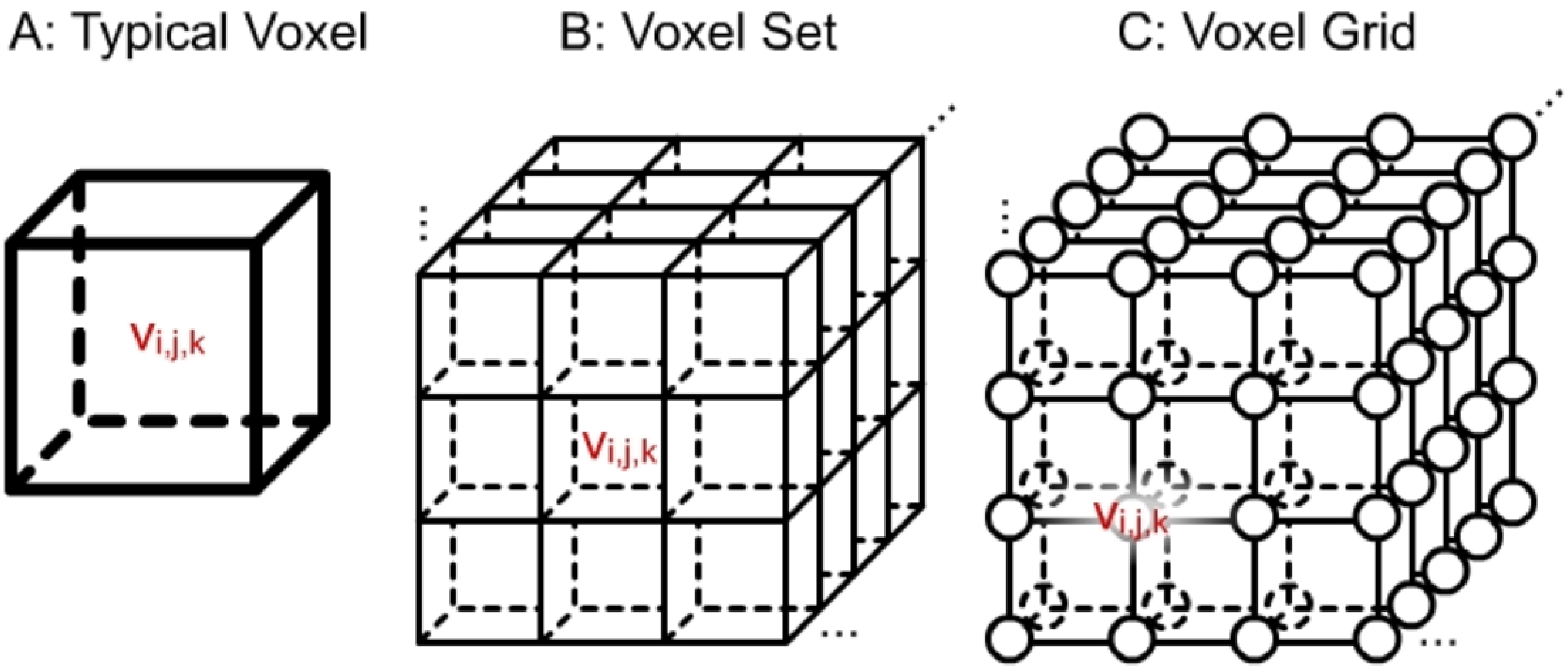
Voxel with samples
at vertices.

Voxel with sample
at center.

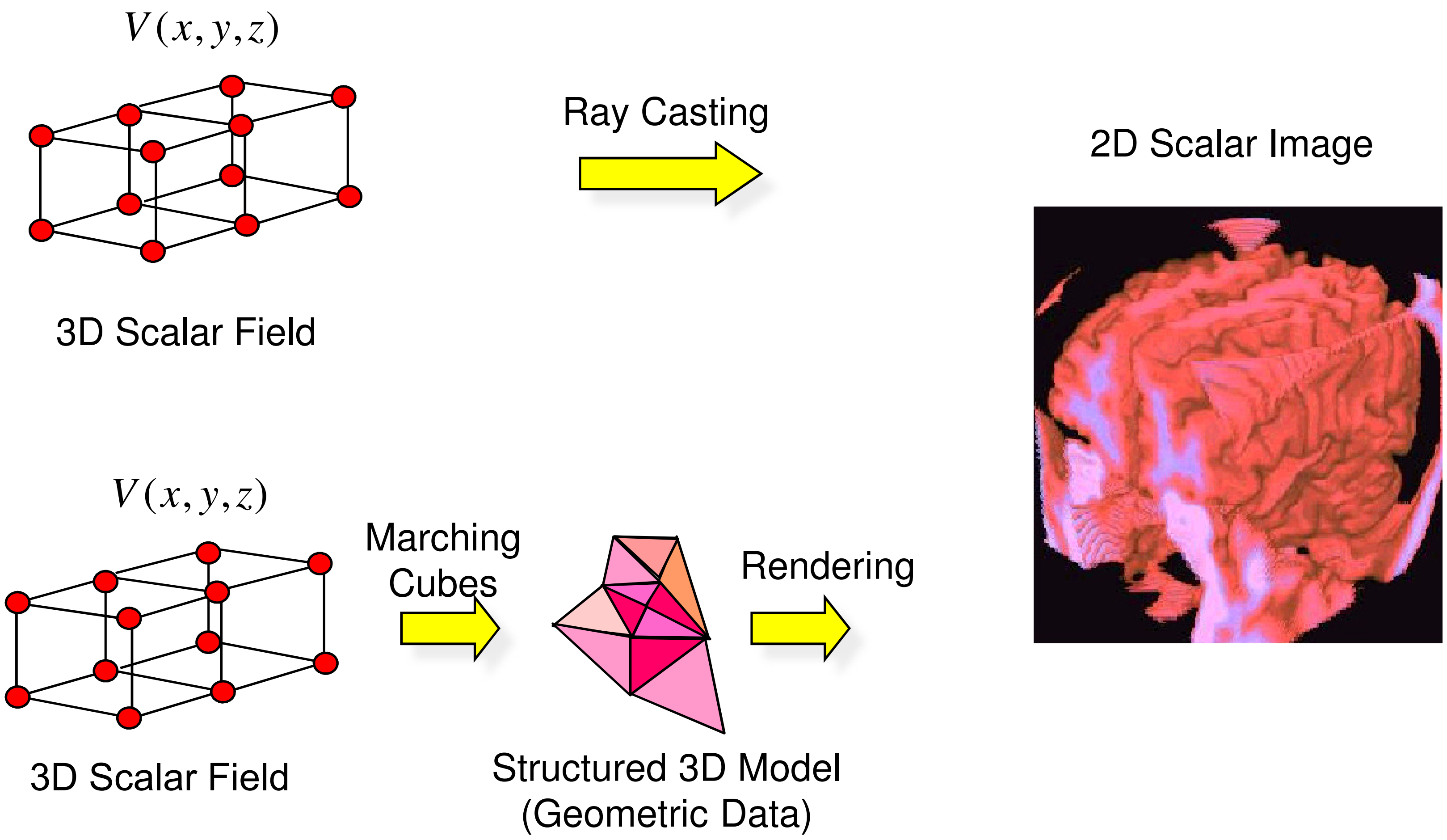
Voxels are for any data representation: <u>temperature, density, pressure</u>, etc.

- Pixels are just for colors

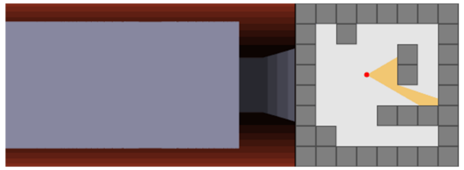
Volume Dataset



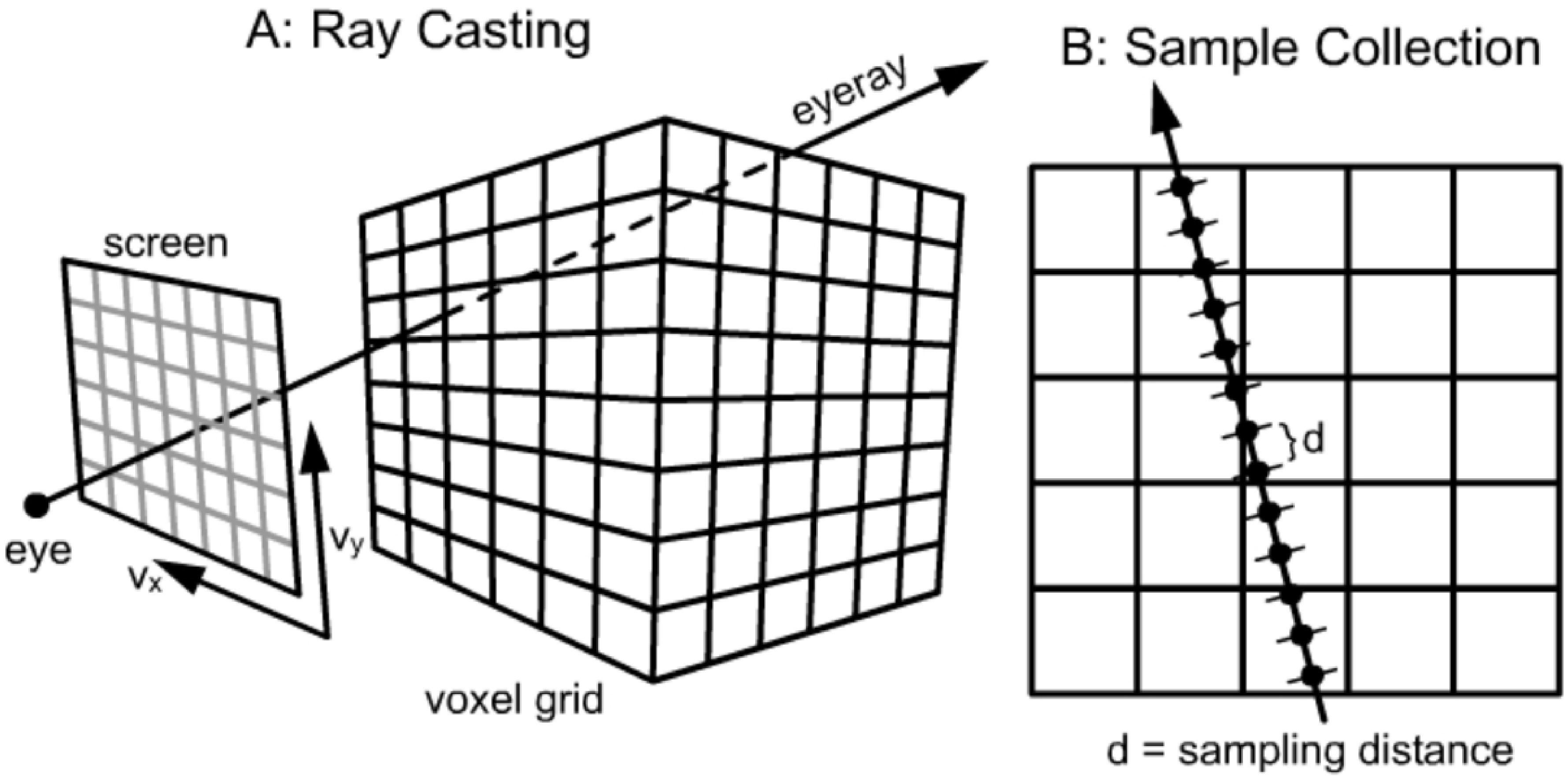
Volume Rendering Process



Ray Casting



Ray Casting and Sample Collection

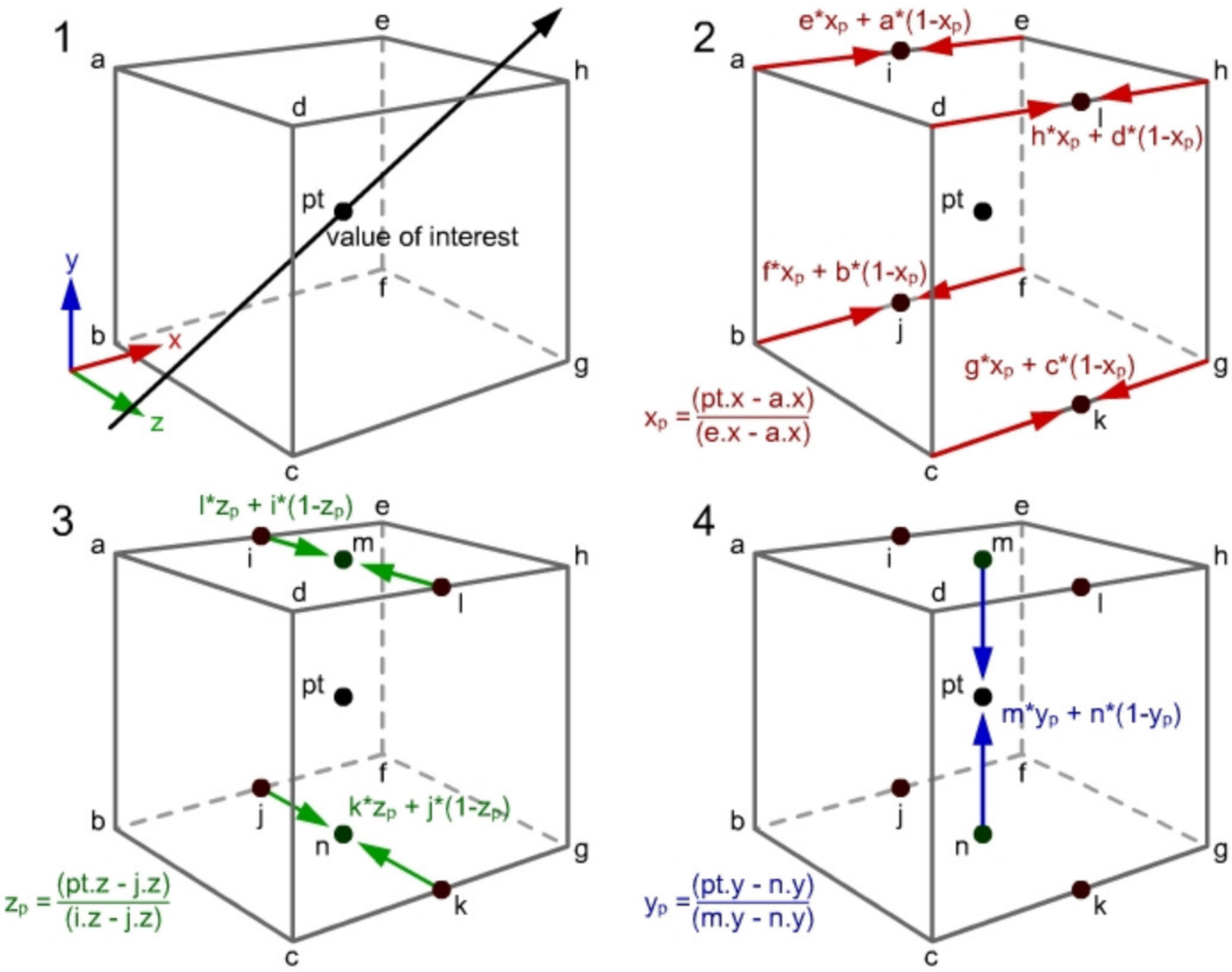


Sampling distance is a user-defined value

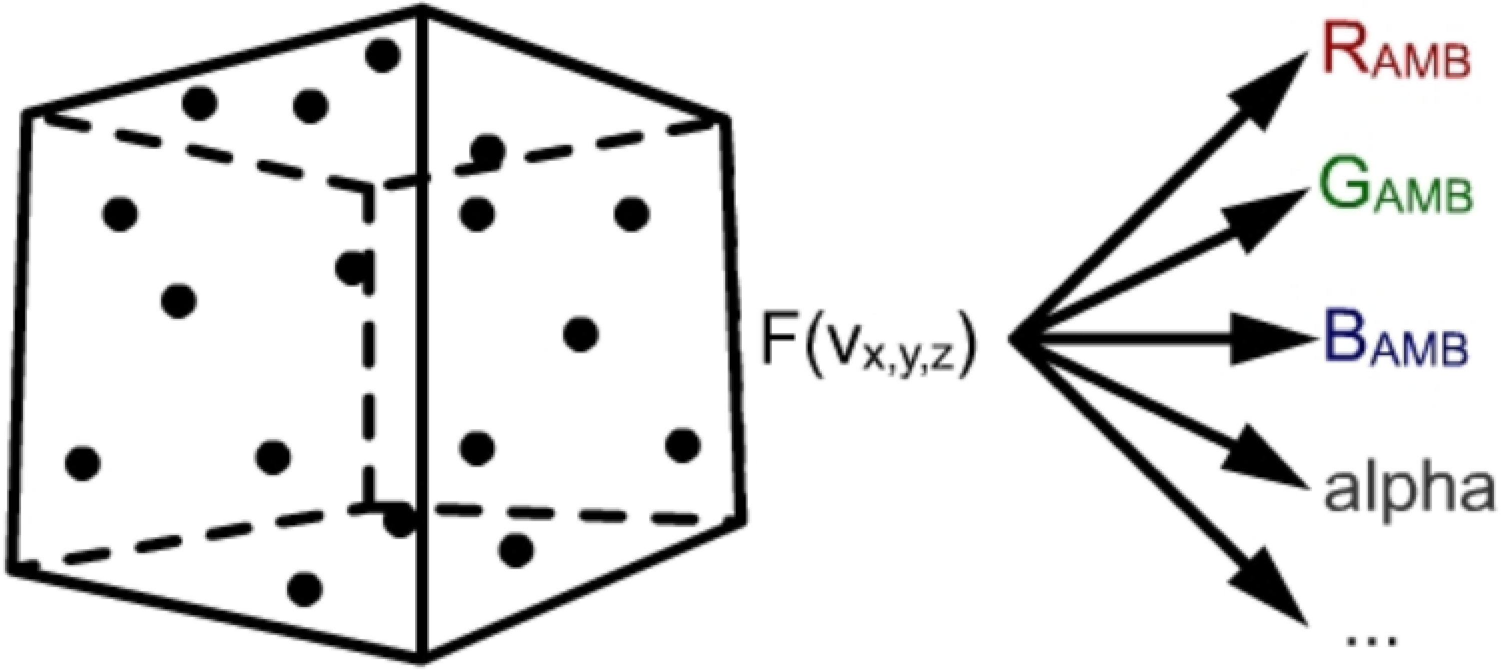
- More sampling results in a more clear, well approximated surface

In general, the sampling distance should be less than the size of a voxel

3D Linear Interpolation



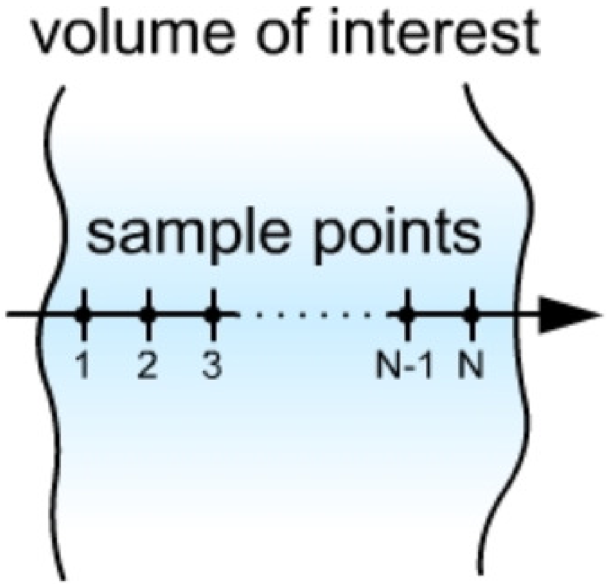
Transfer Function



Transfer functions are used to convert sampled values to color and alpha values to describe the surface

- Transfer functions are entirely user defined and are manipulated to make the surface coherent

Alpha Combination of Sample Color

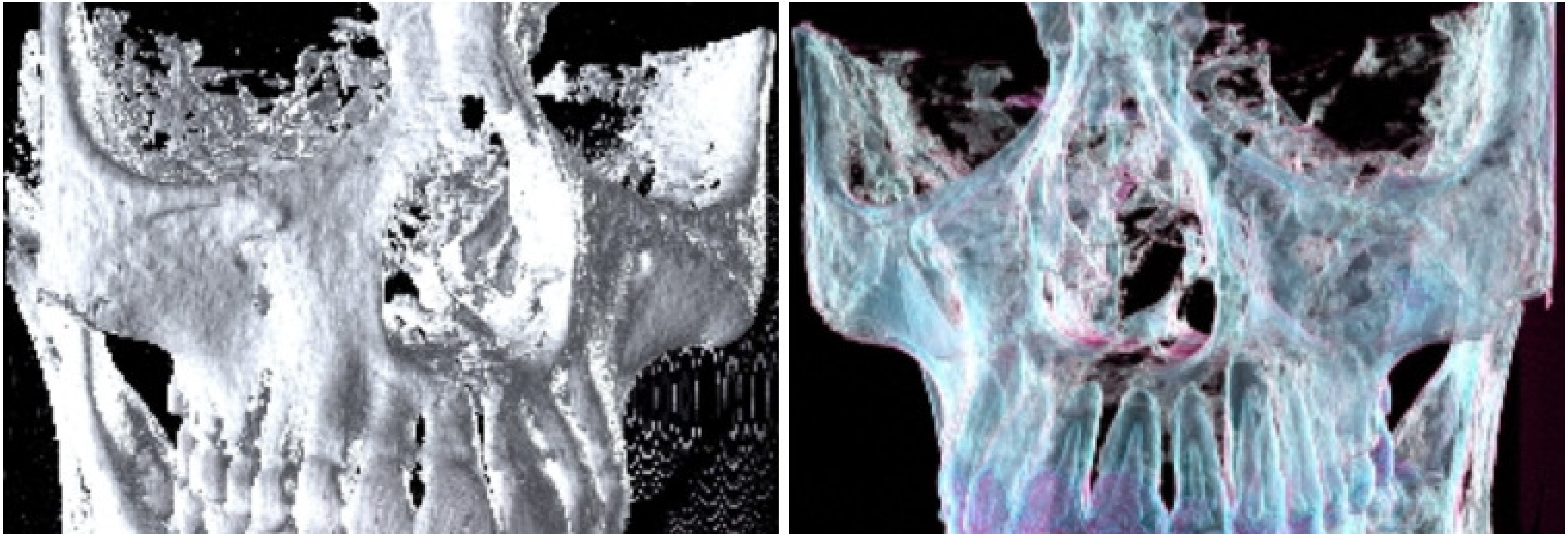


$$c(i,j) = c_1\alpha_1 + c_2\alpha_2(1-\alpha_1) + \ldots + c_{N+1}\alpha_{N+1}(1-\alpha_1)(1-\alpha_2)\ldots(1-\alpha_N)$$

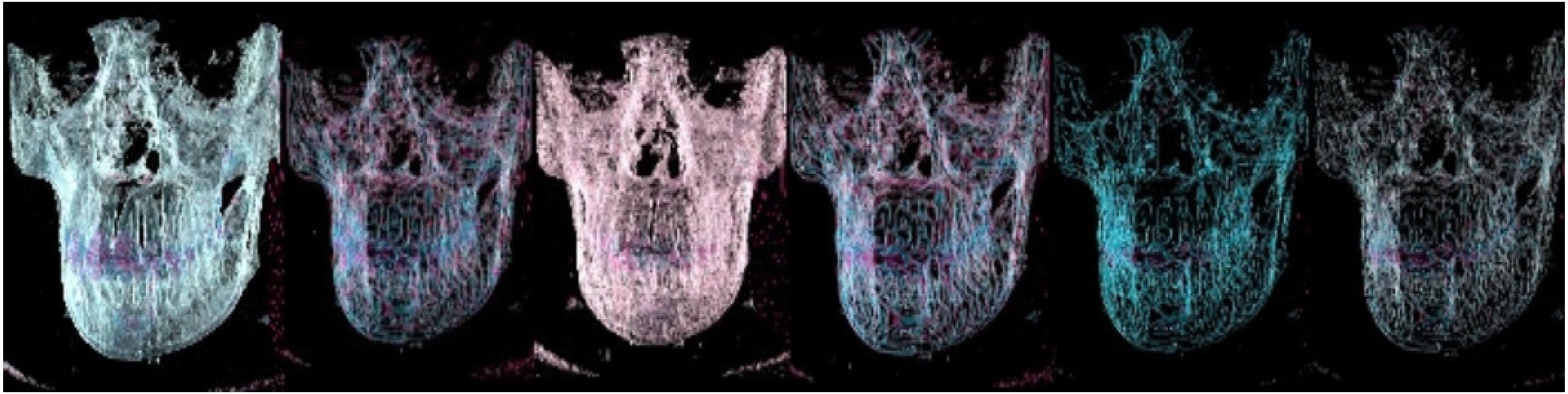
$$color_{pixel}(i,j) = c(i,j) = \sum_{n=1}^{N+1} c_n\alpha_n\left[\prod_{m=0}^{n-1}(1-\alpha_m)\right]$$

$$\alpha_0 = 0 \;\&\; c_{N+1} = c_{BG}$$

The combining of the samples is performed in such a way that the samples nearer to the observer (eye) obscure those behind it according to the surface alpha values.



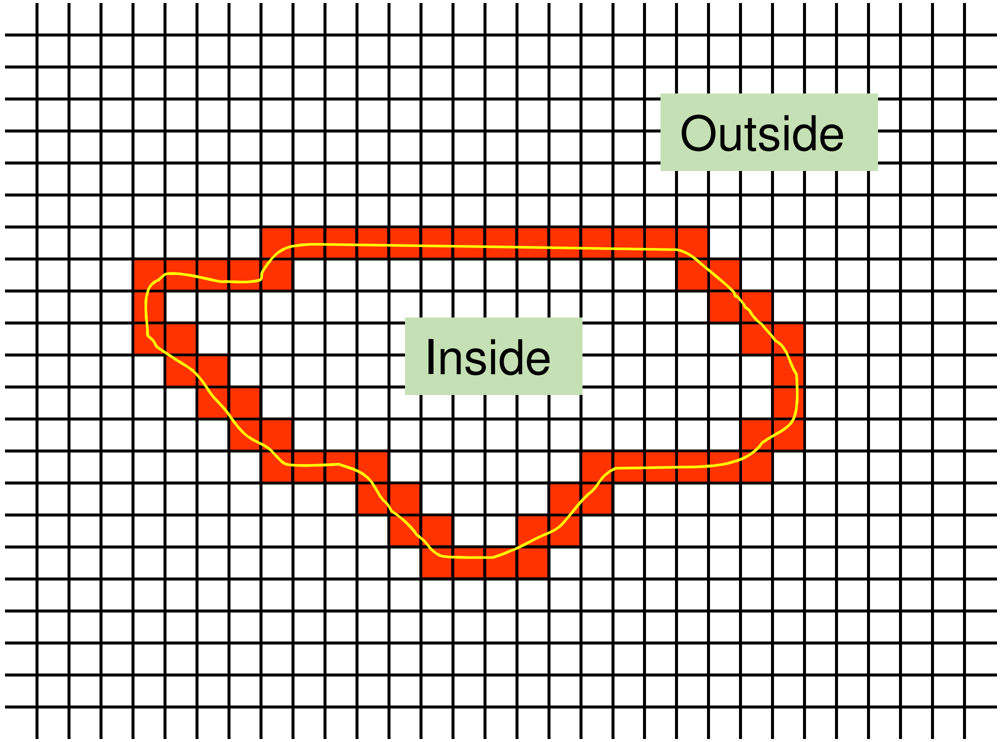
Examples
(without and with transparency)



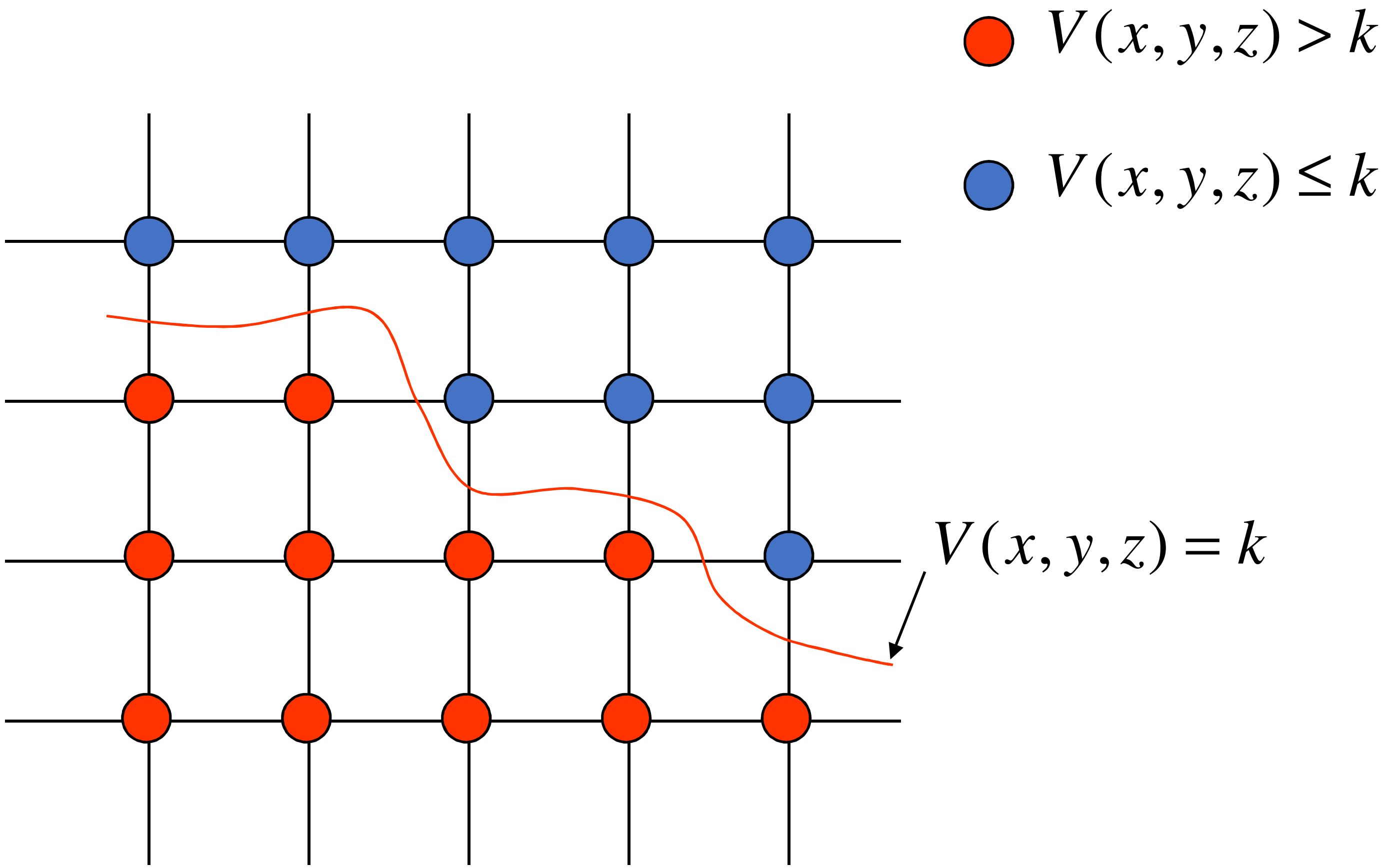
Using different transfer functions

Marching Cubes

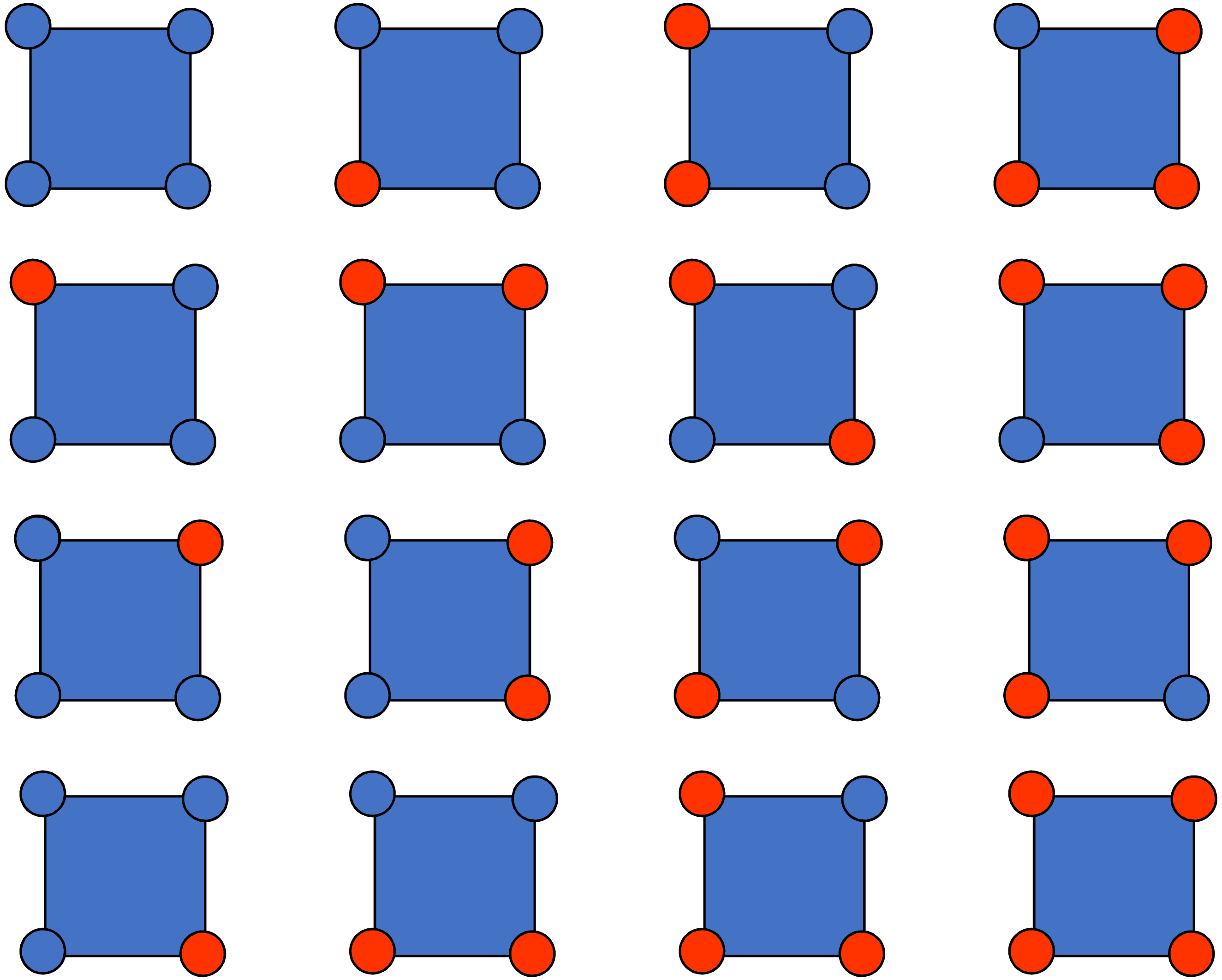
The "Marching Cubes" Algorithm Lorenson & Cline 1987



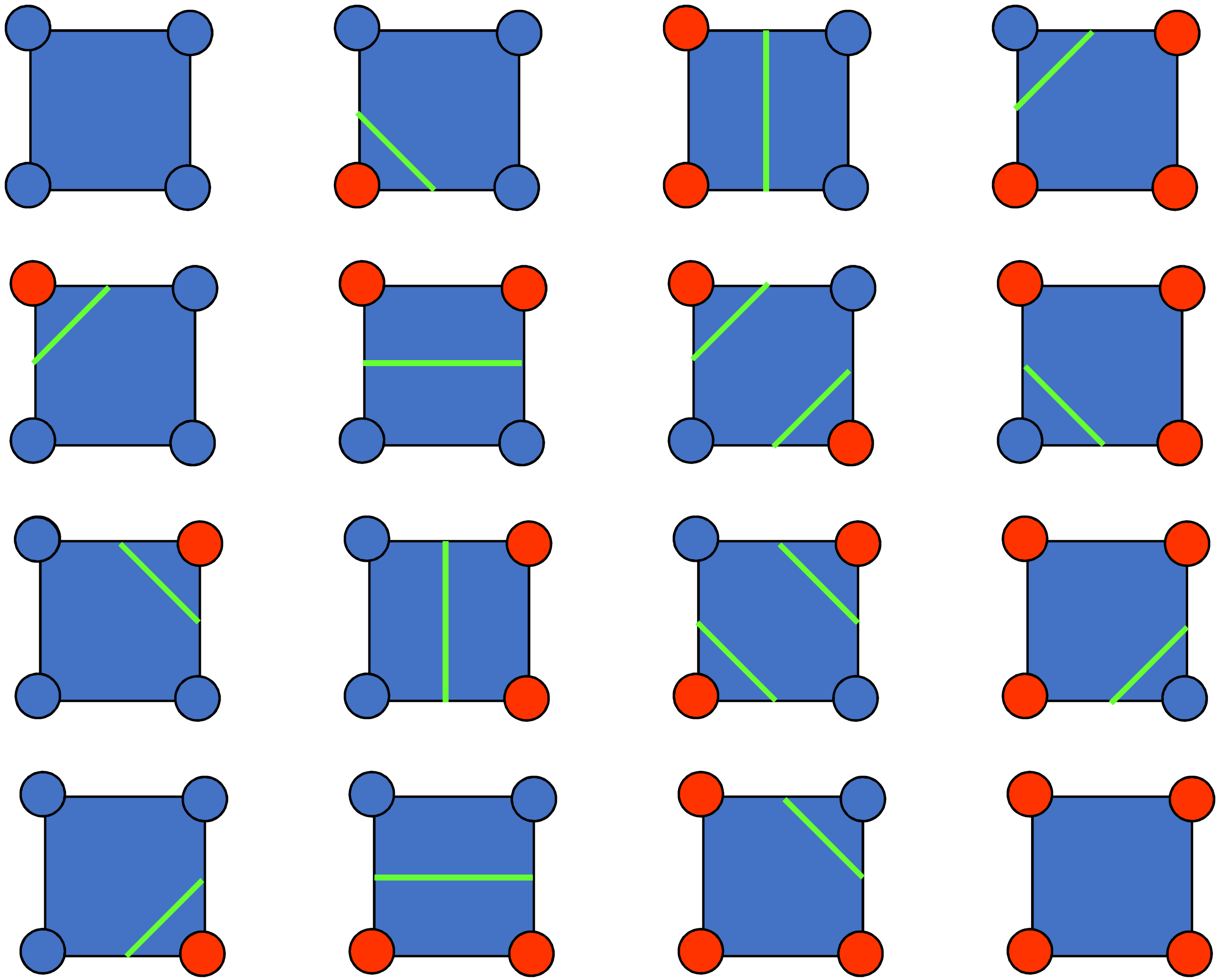
The "Marching Cubes" Algorithm



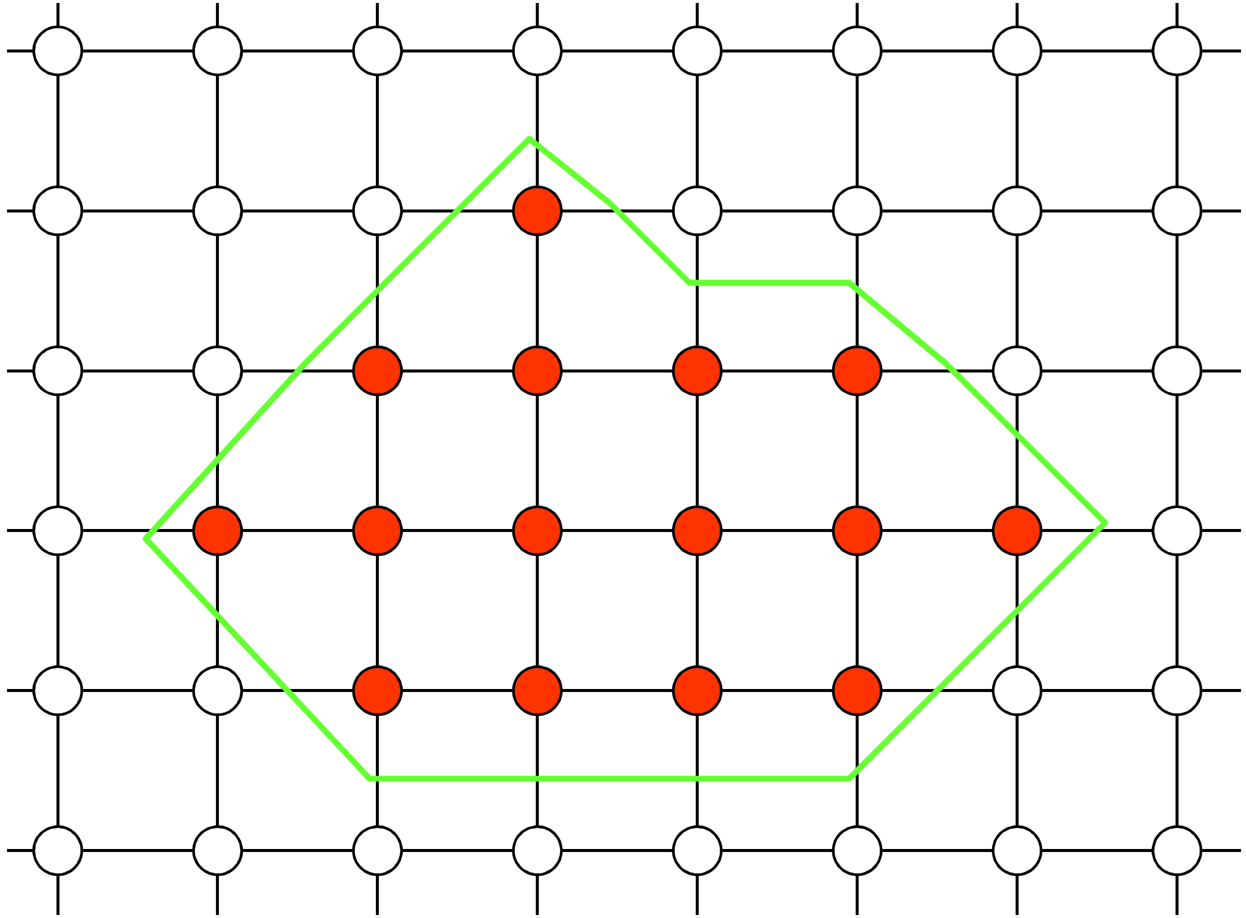
The "Marching Cubes" (Marching Square) Algorithm Possible Vertex States



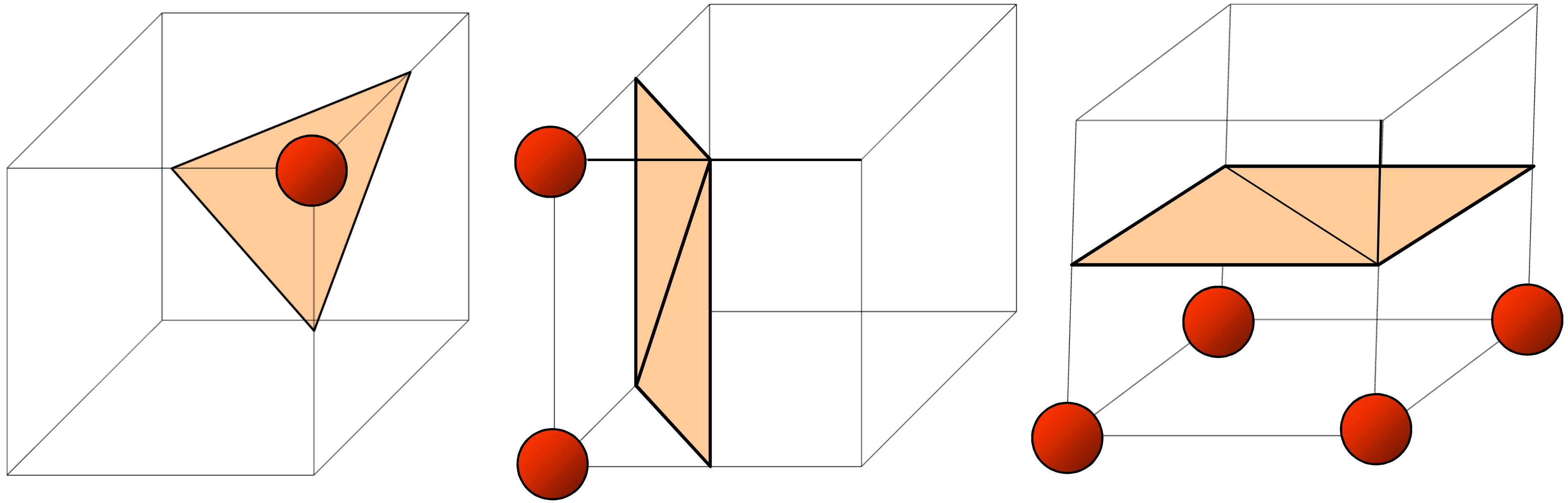
The "Marching Cubes" Algorithm
Generated contour



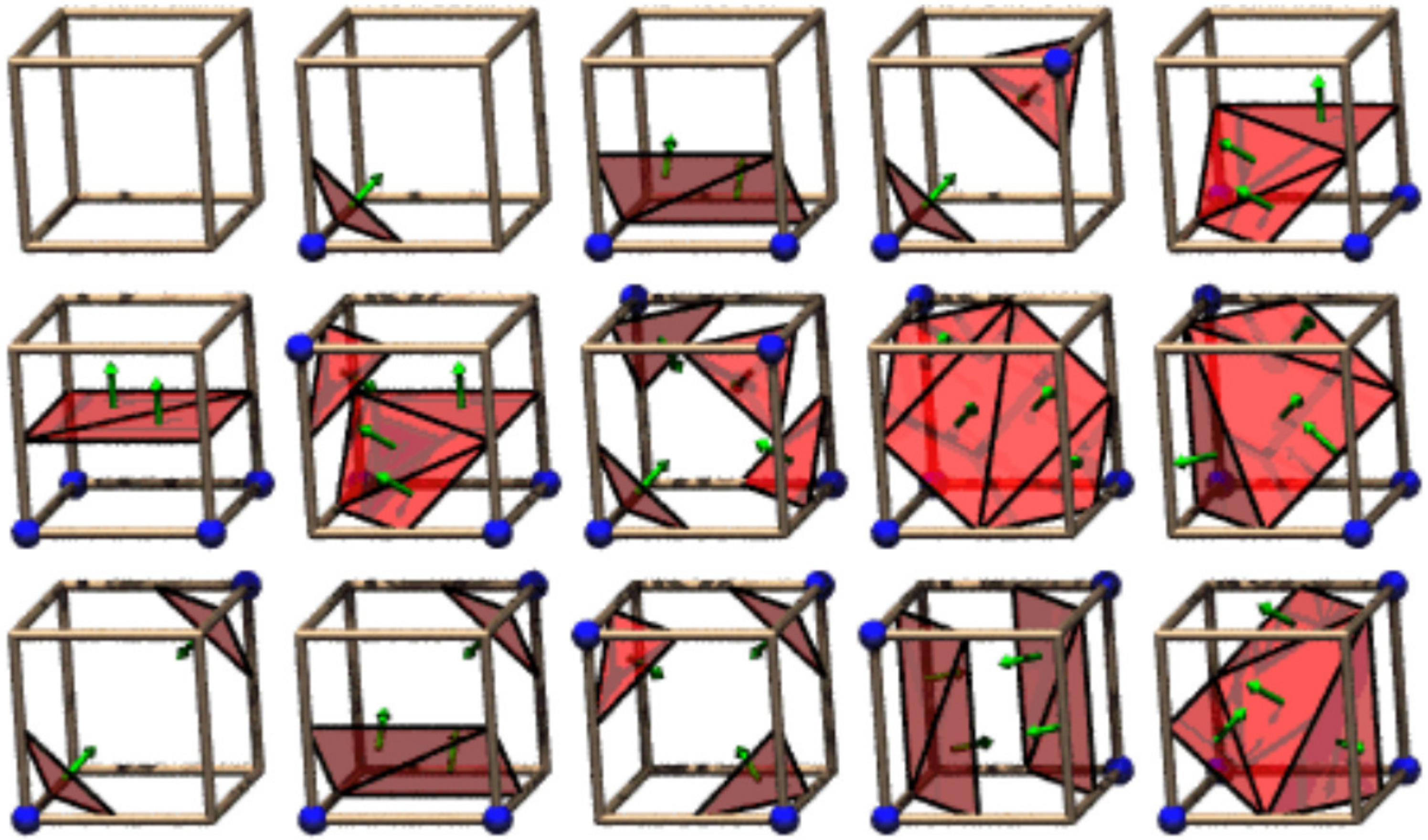
Generation of contour from subcontours



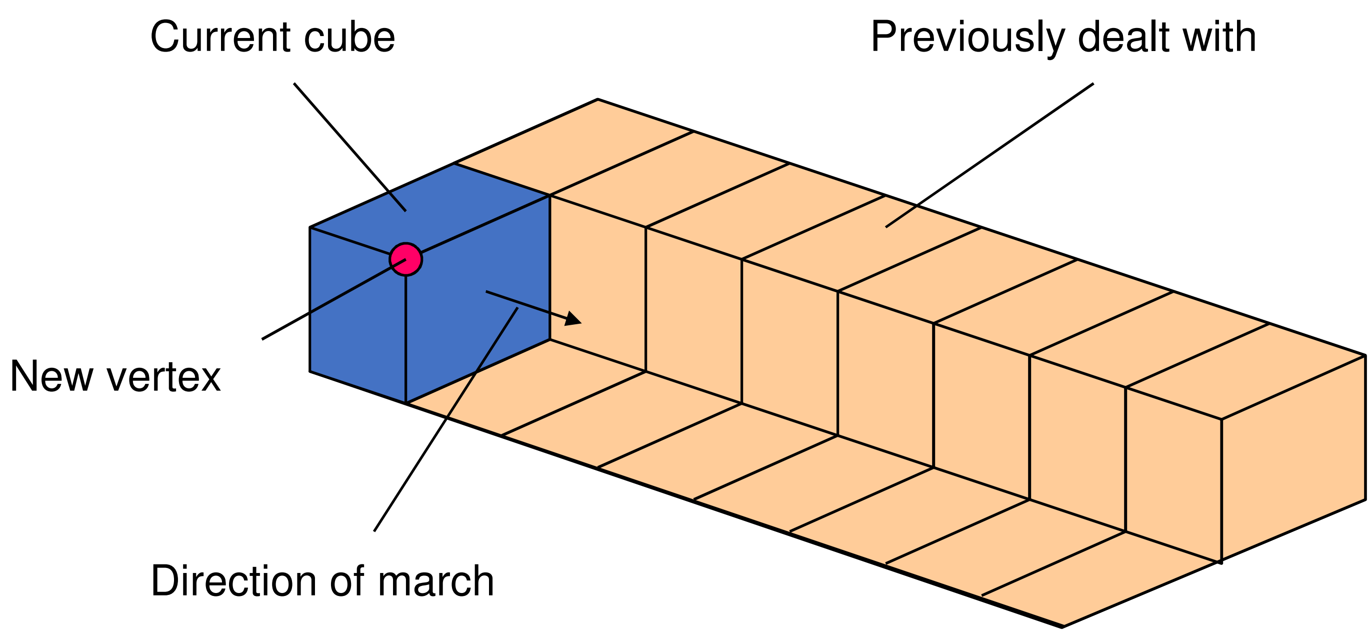
Marching Cubes Algorithm in 3D Isosurface generation



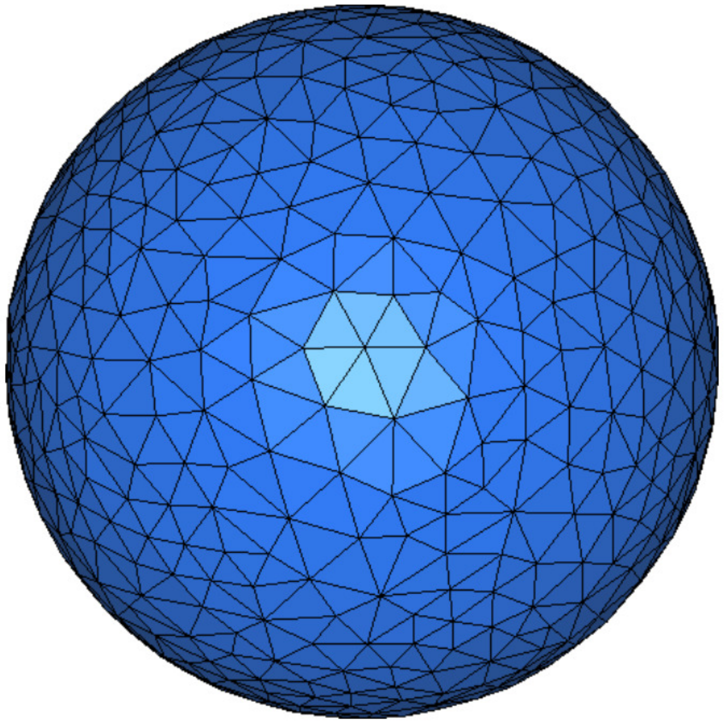
15 Cases

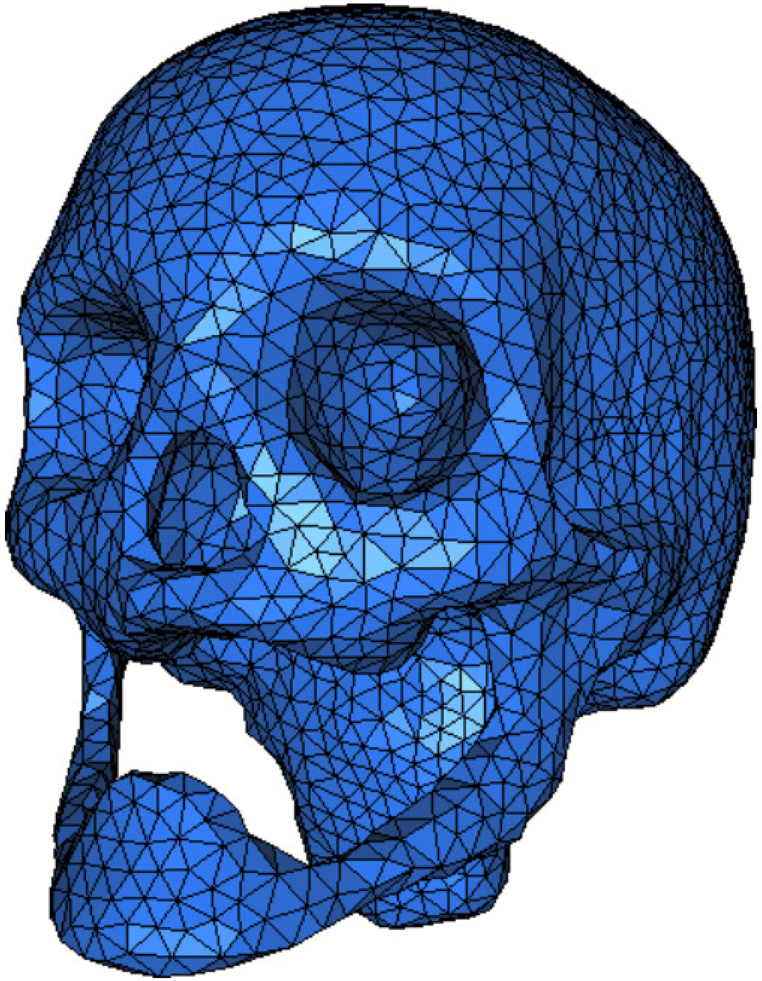


Marching Cubes



Triangulation Examples





Information Visualization

- “... 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.” (Card, Mackinlay, Shneiderman, 1999)

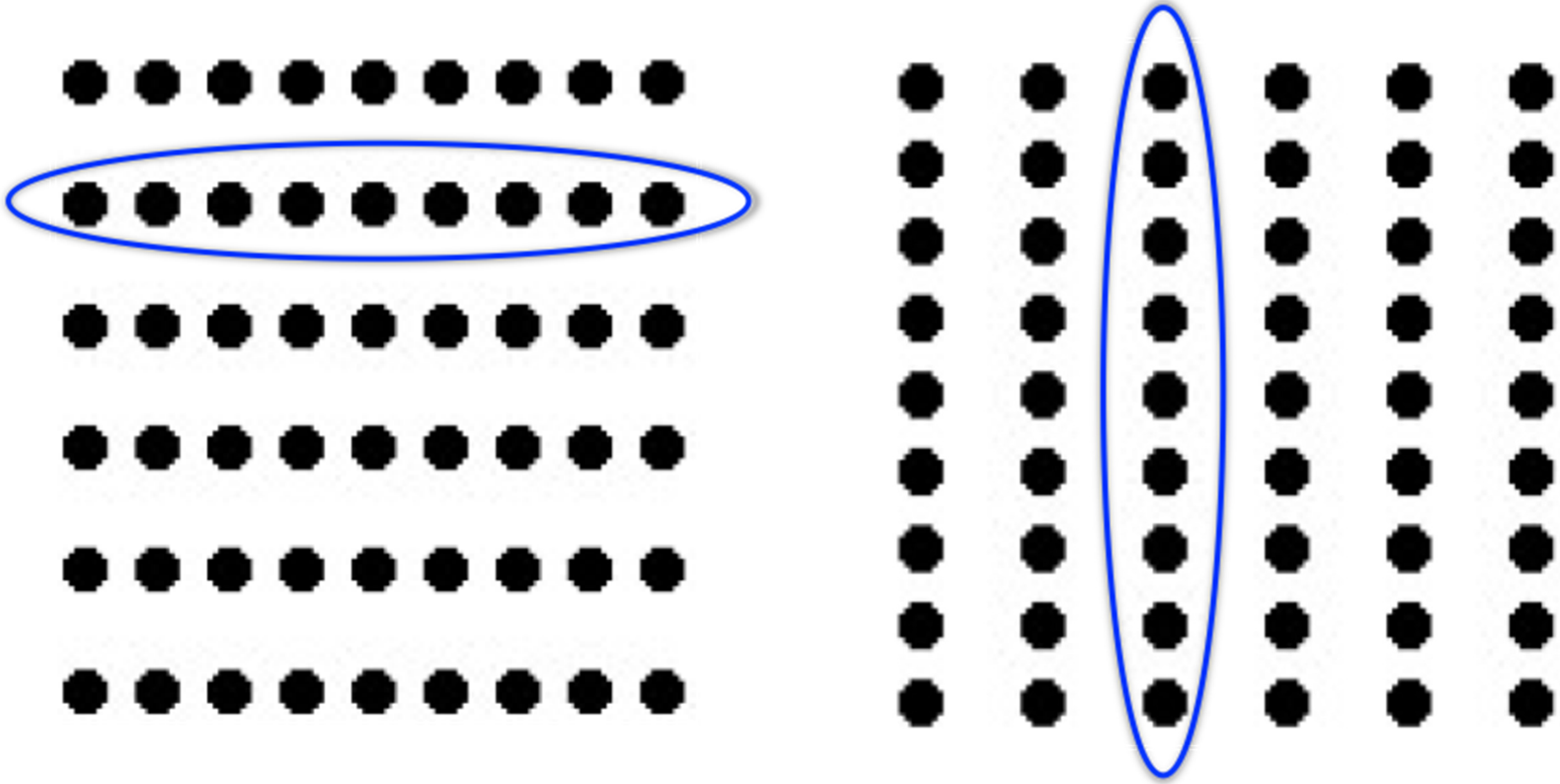
Visual Thinking: Example 1

- Counting the number of 3s in the following Text:

1235693234870452973467
0378937043679709102539

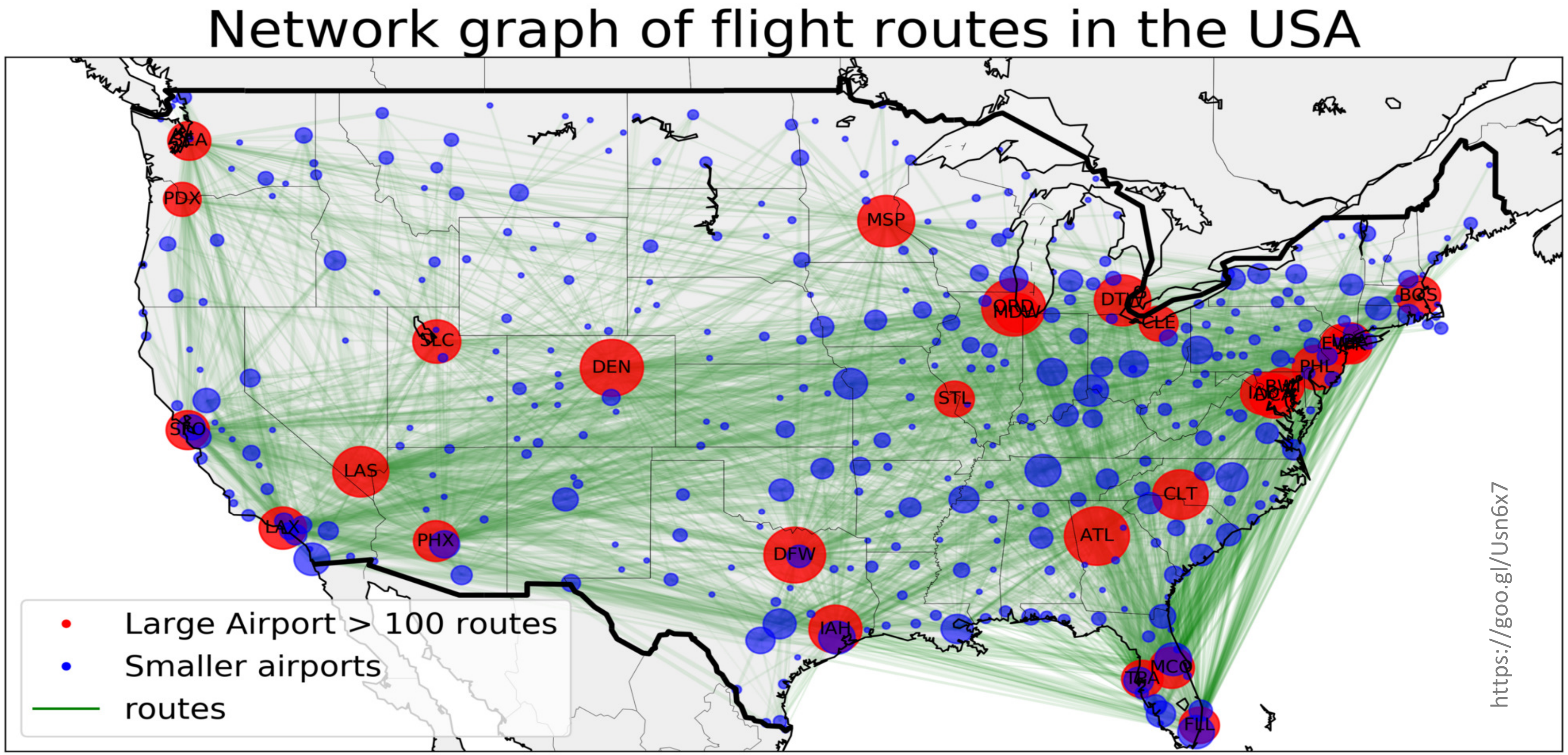
Visual Thinking: Example 2

- Identify the groups of dots in the following figures

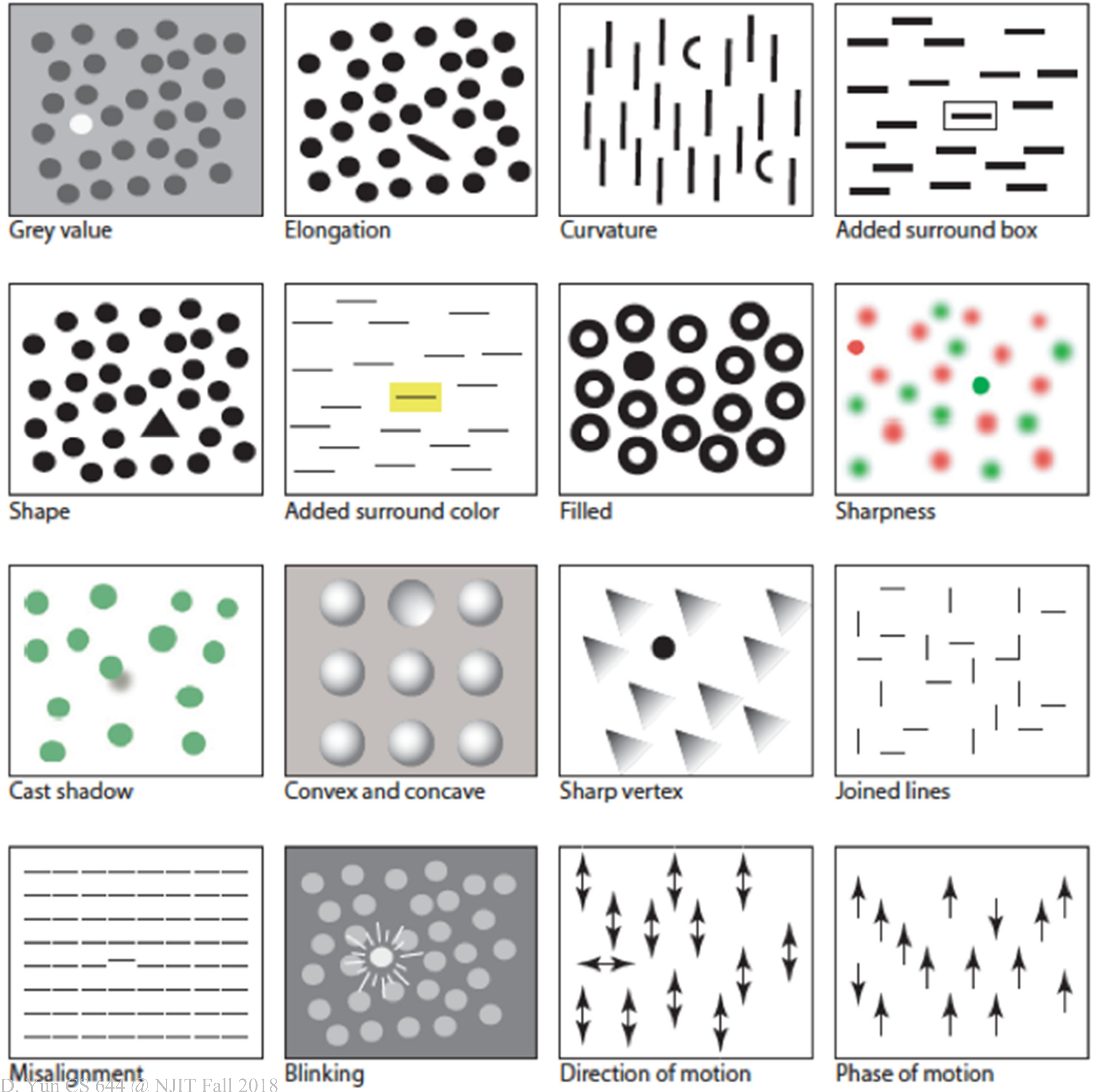


Law of Proximity
we tend to group elements that are closest to each other

Visual Thinking: Example 3



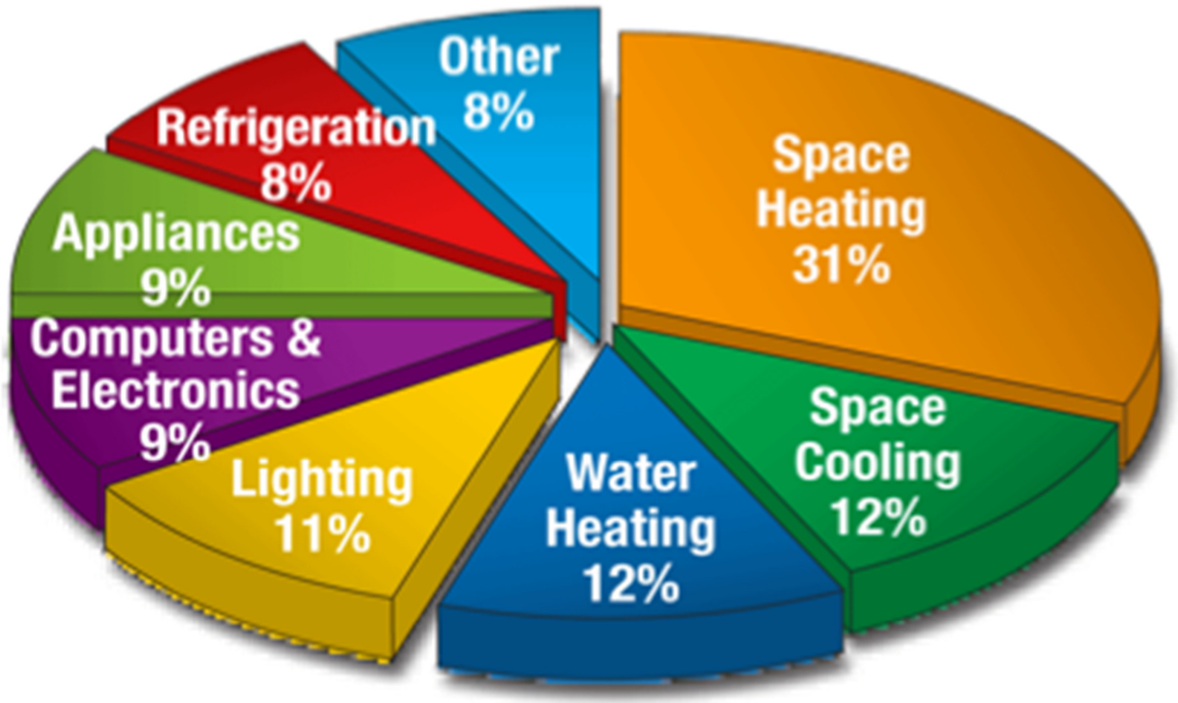
Pre-Attentive Visual Attributes



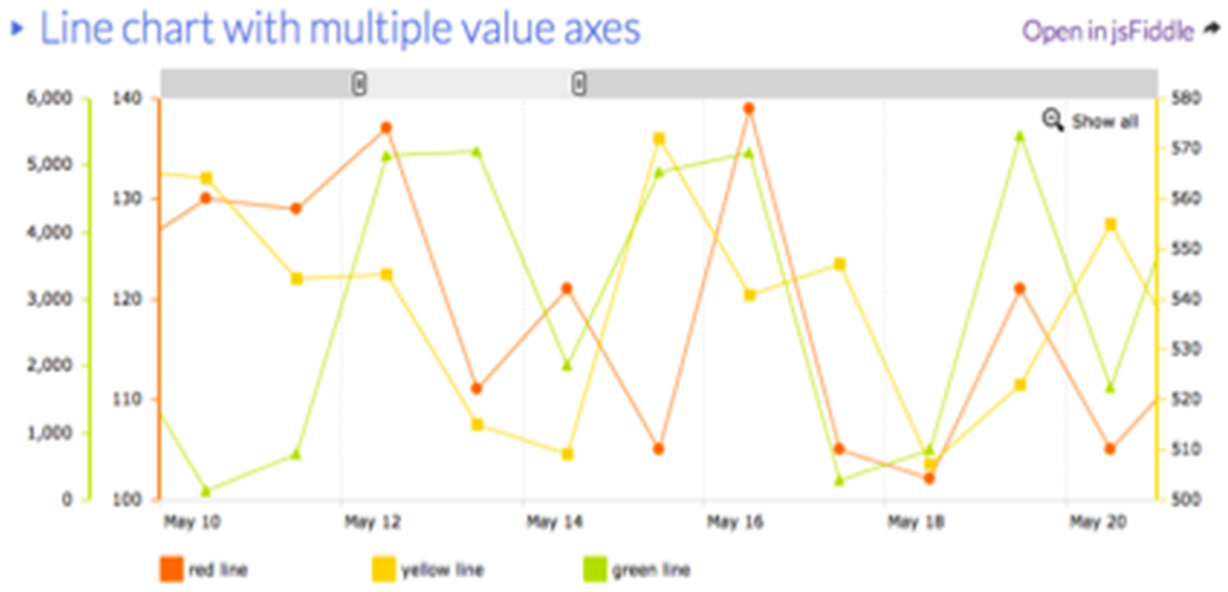
Big Data Era: Data, Data, and Data
How do we make sense of the data?

340 million tweets
a day!

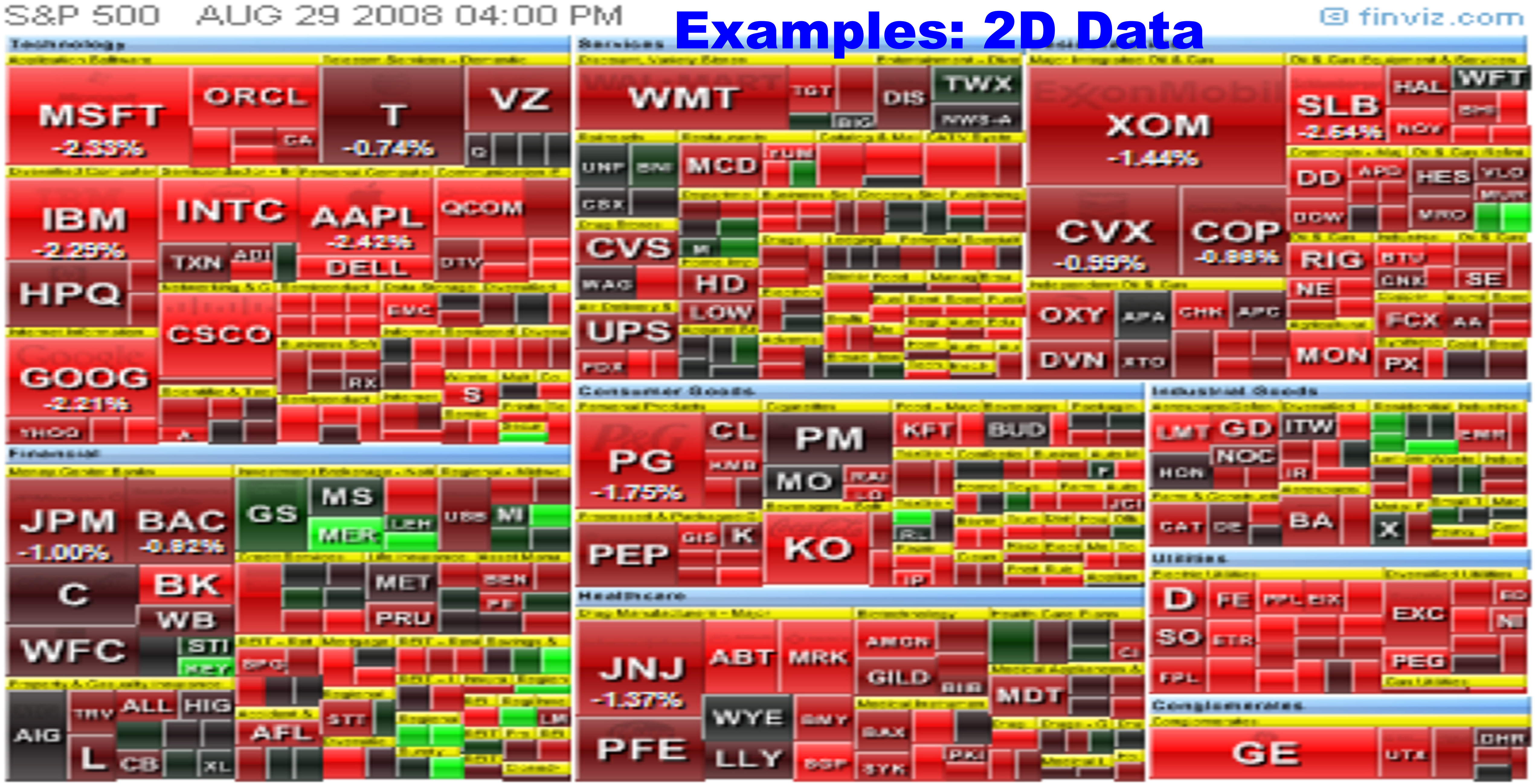
4 billion messages
a day!



Examples: Visualizing Numerical Data

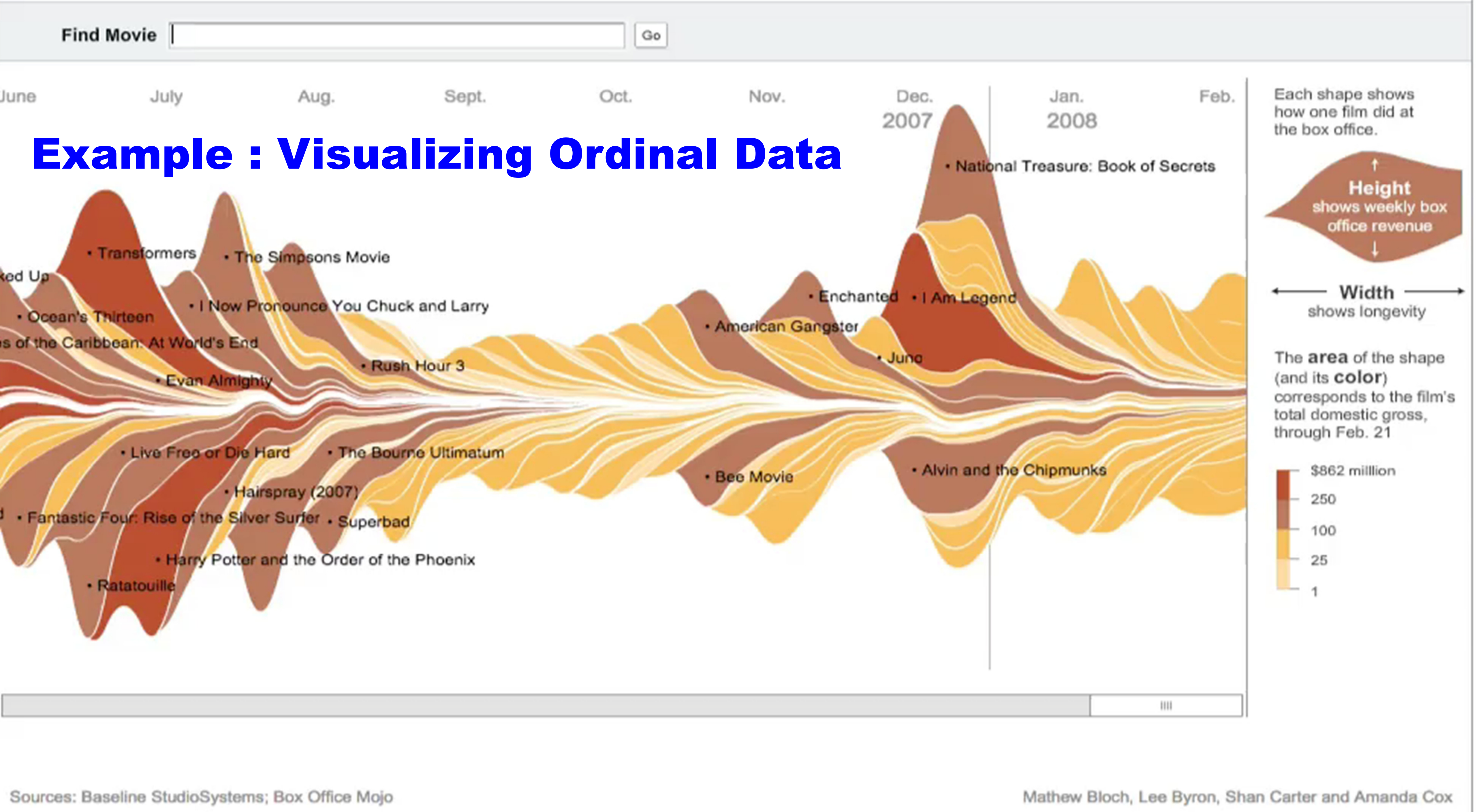


Examples: 2D Data

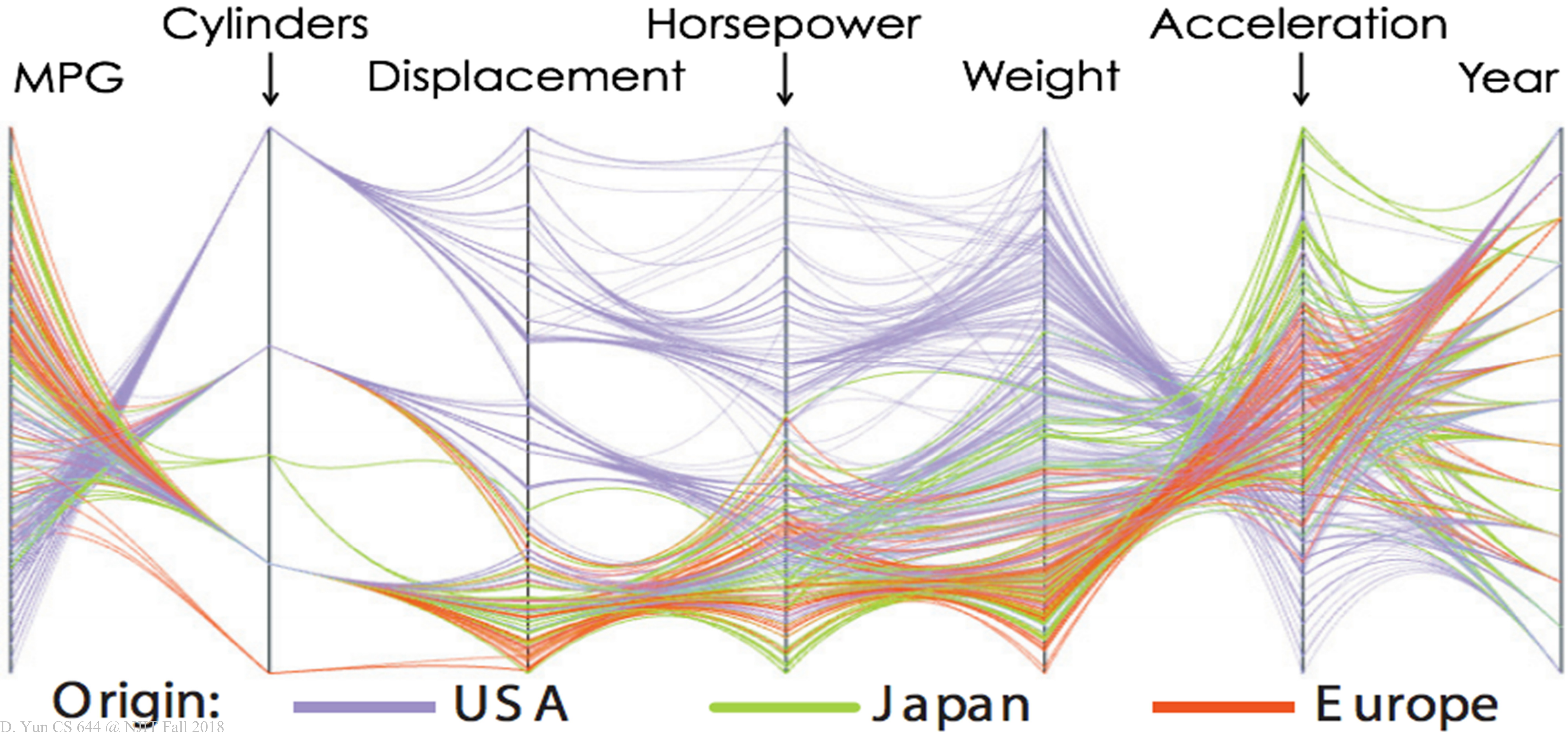


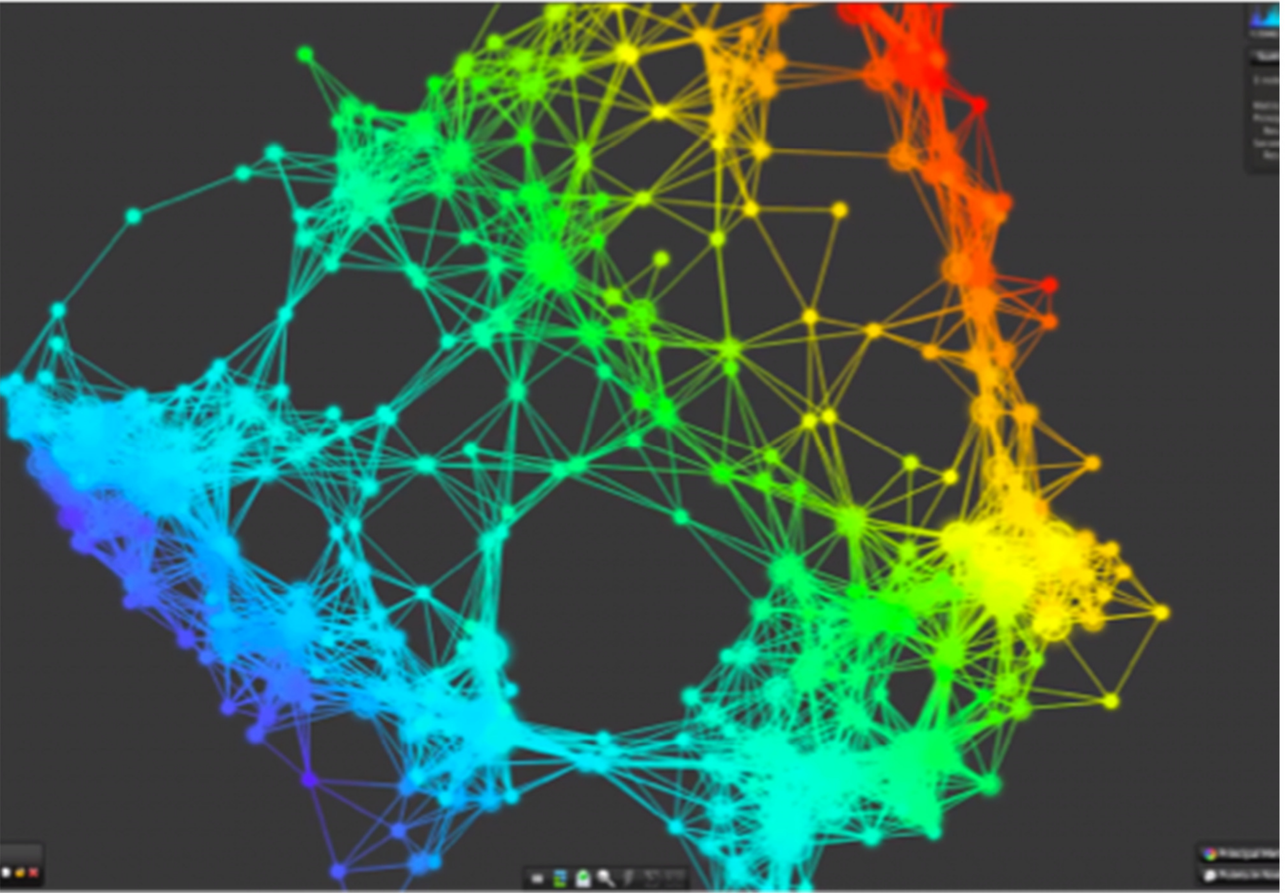
Size of each cell: Stock Market Value
Color: Stock Change

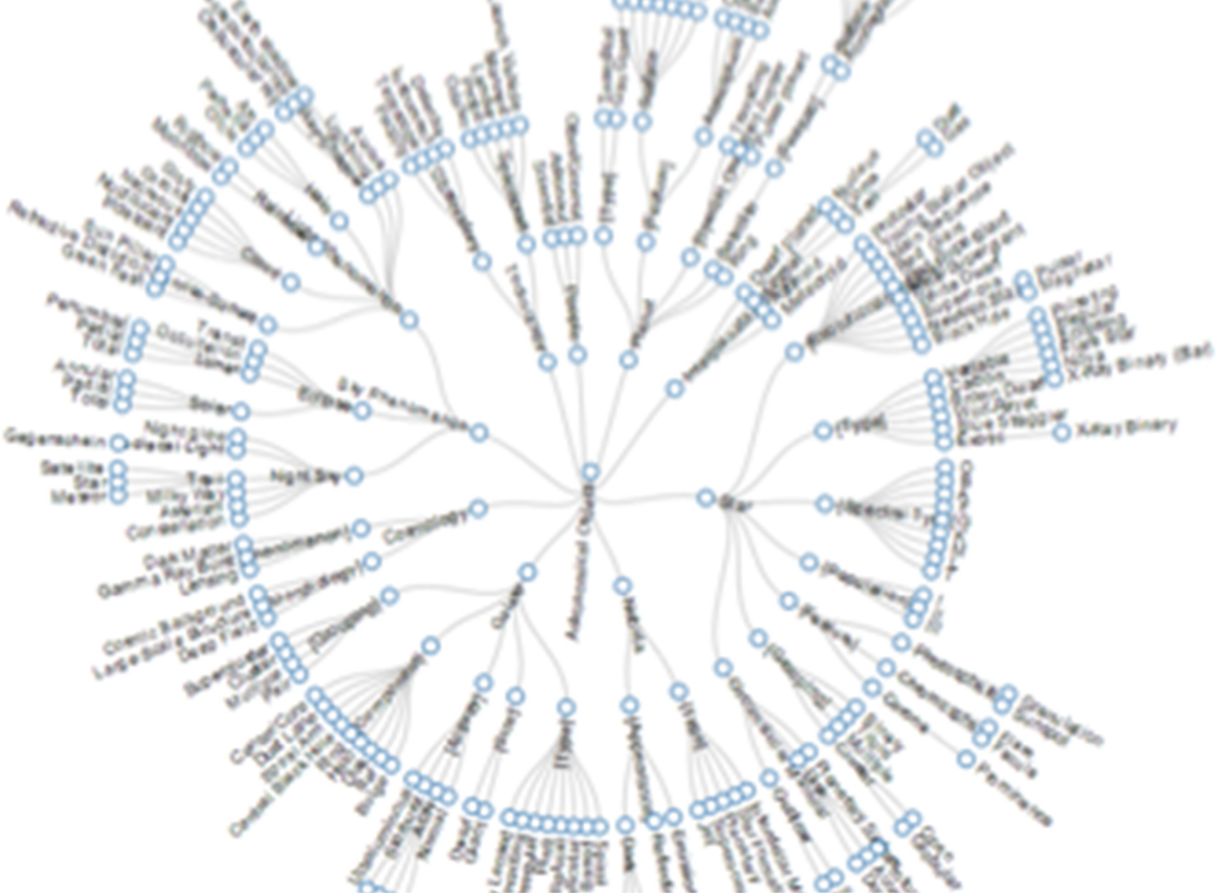
Example : Visualizing Ordinal Data



Example: Multi-Dimensional Data

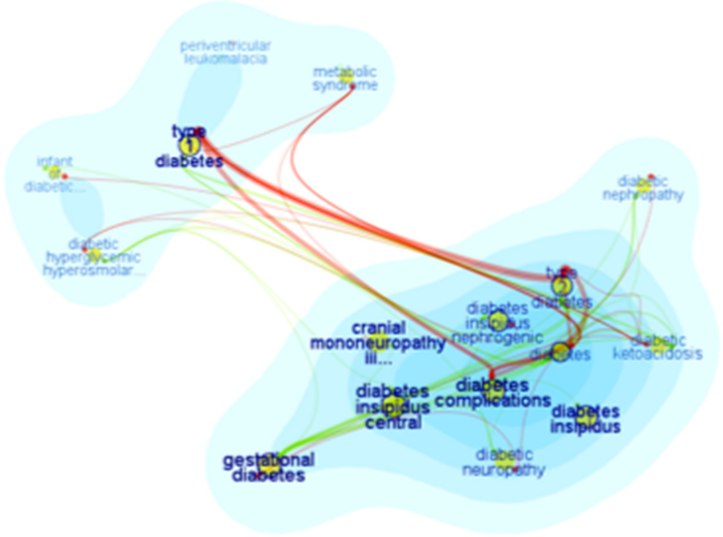


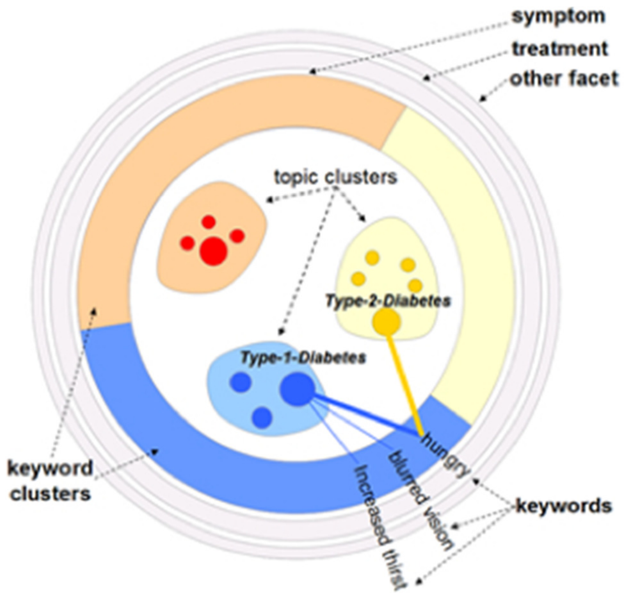




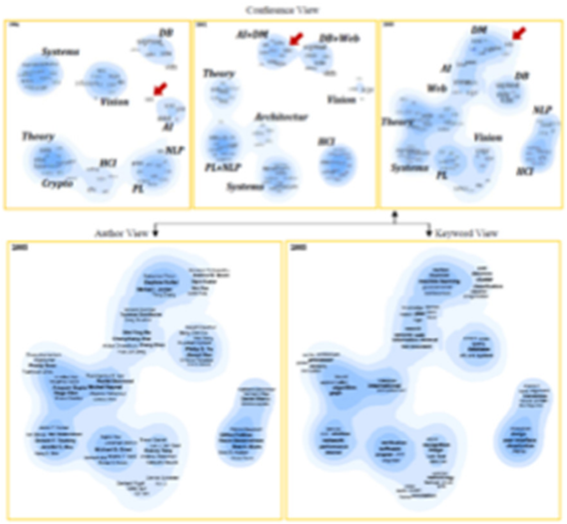
Examples: Visualizing Structured Data

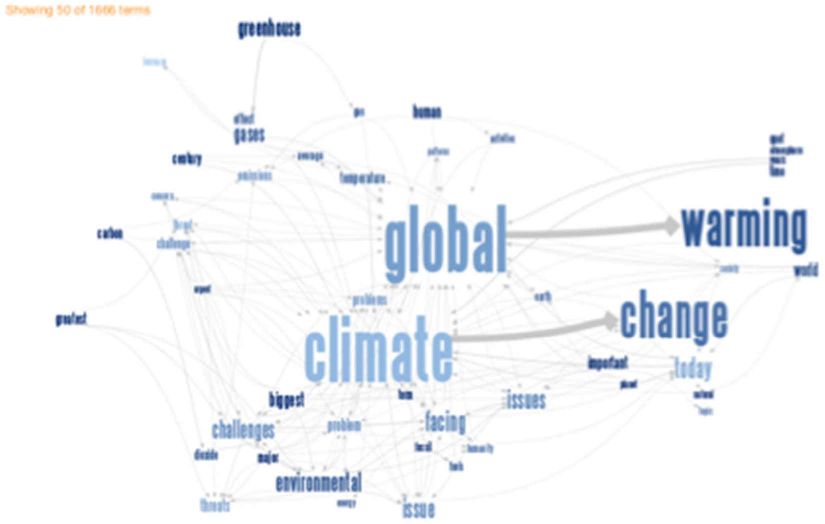
Examples: Visualizing Unstructured Data

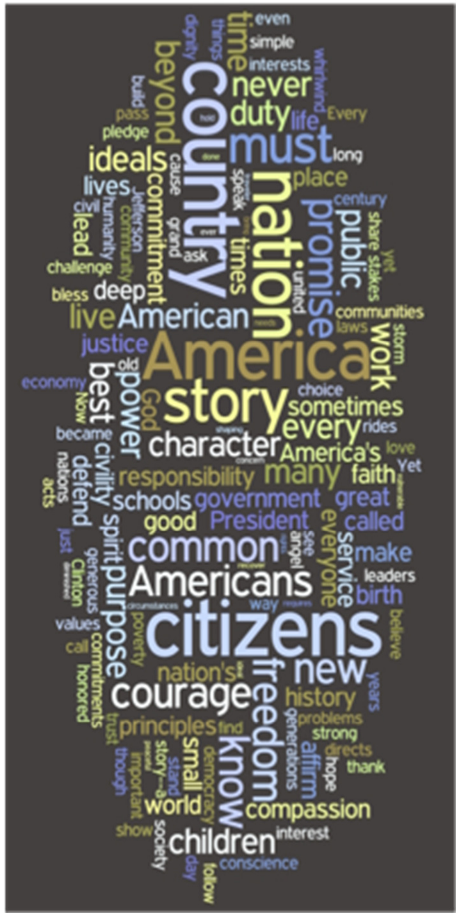




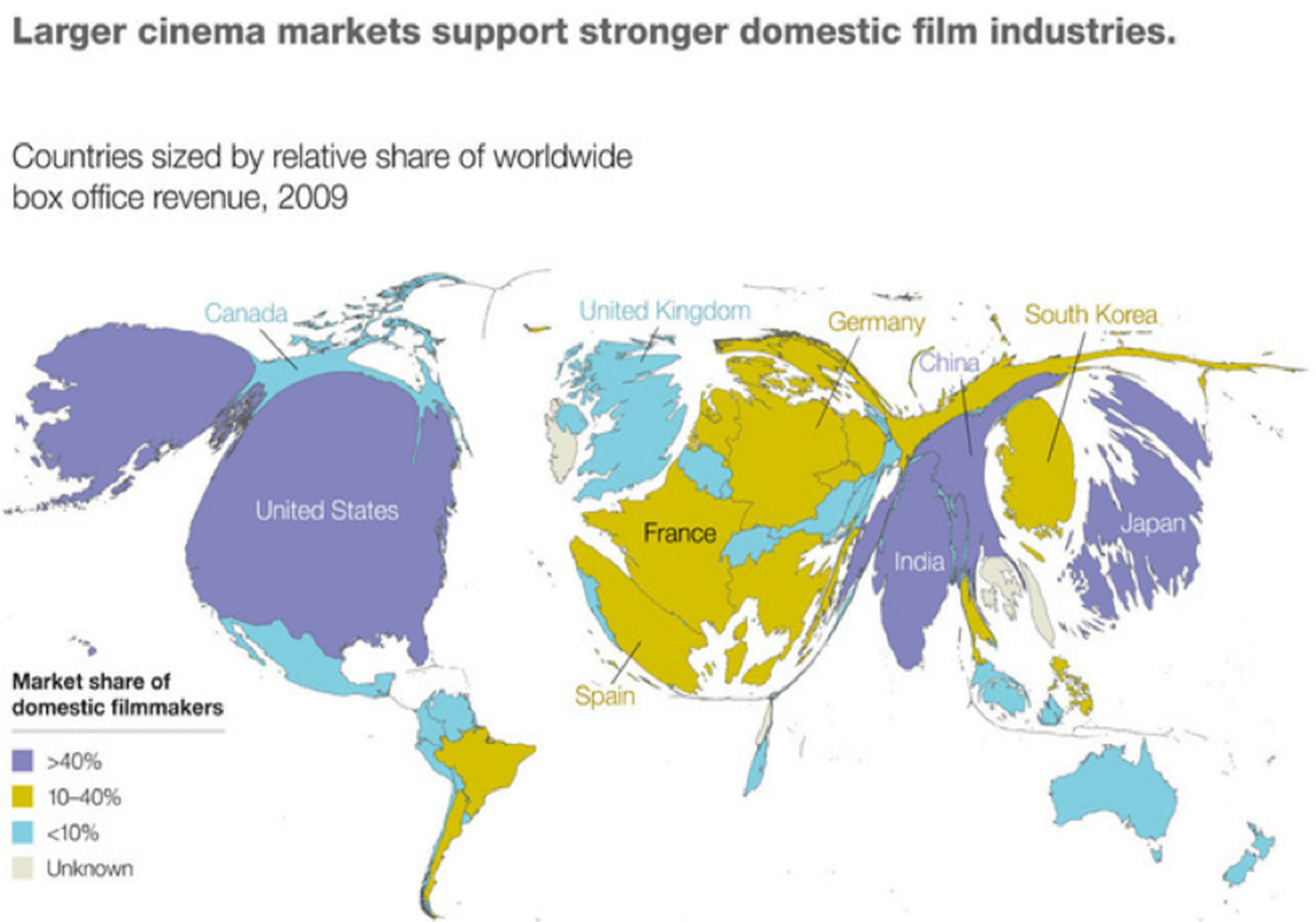
Visualization of
Text Documents



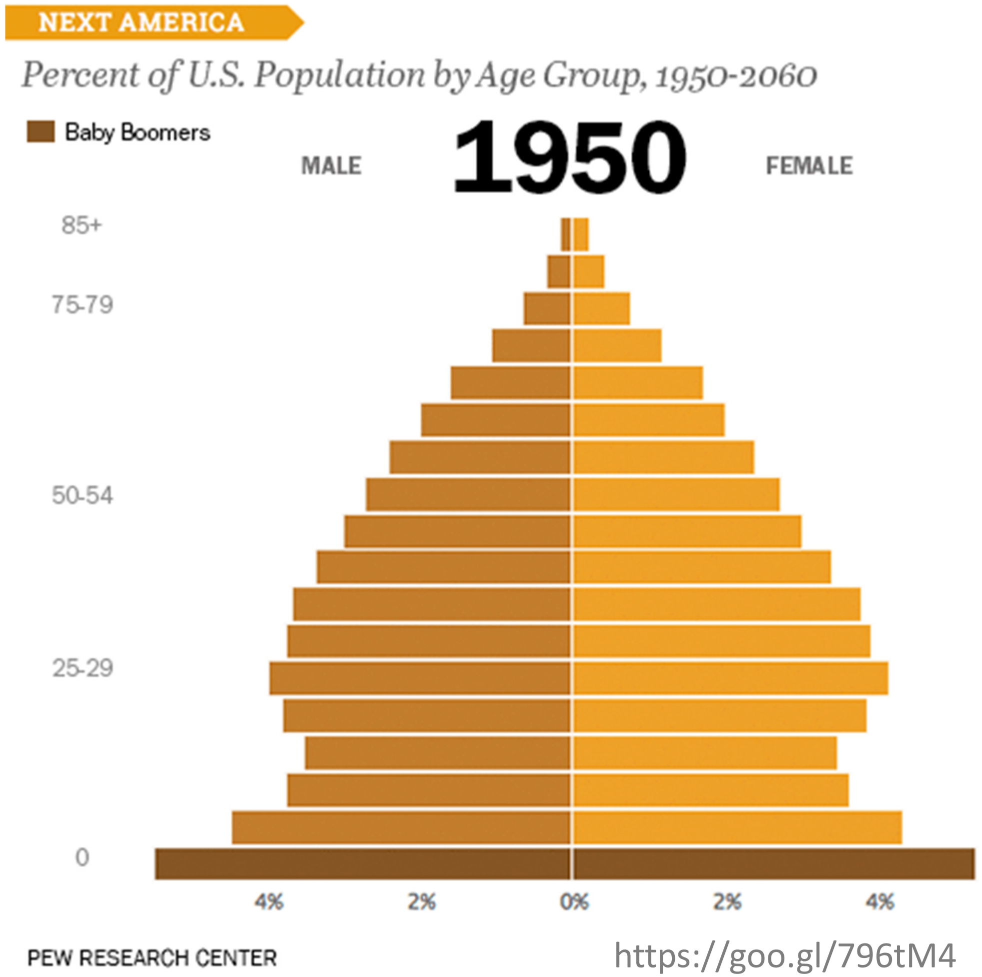




Examples: Geospatial

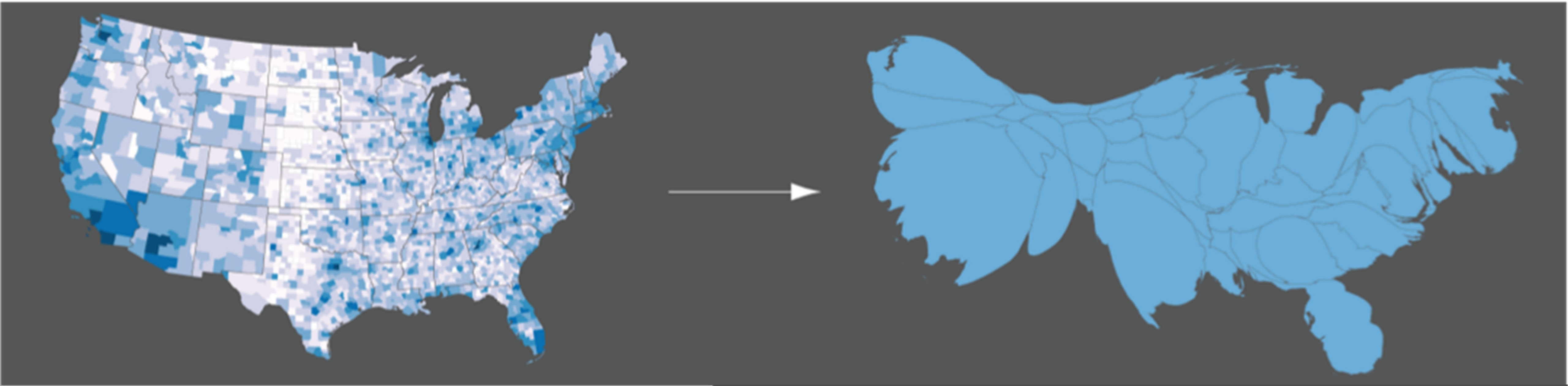


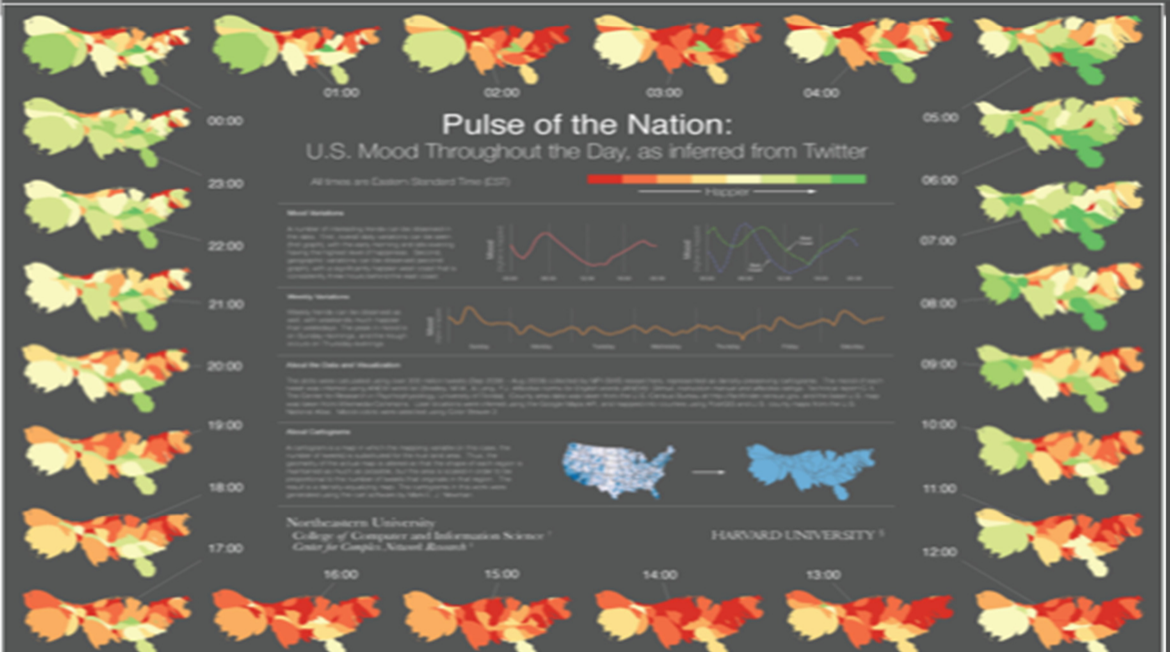
Example : Visualizing Temporal Data



https://goo.gl/796tM4

Example : Visualizing Spatial Temporal Data





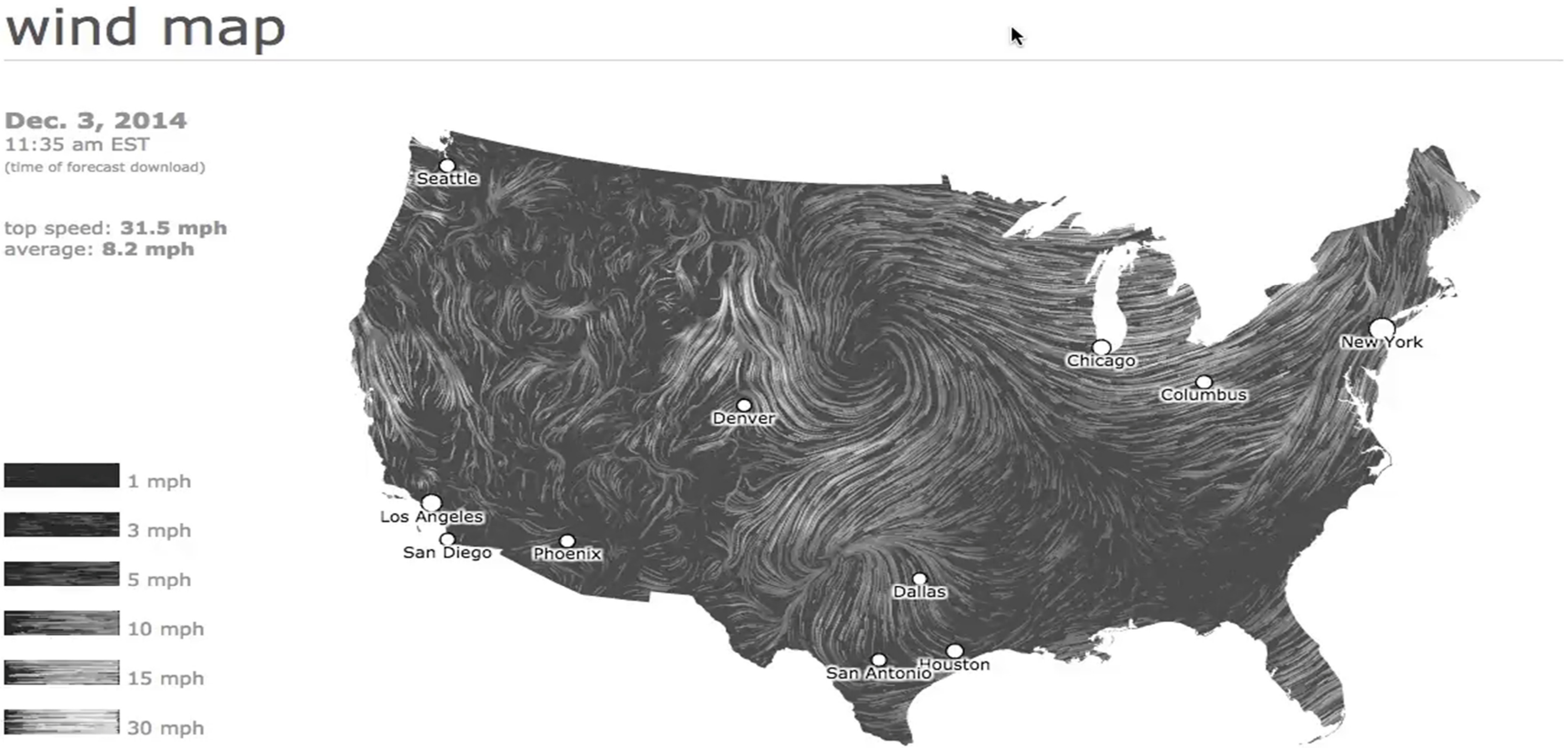
Pulse of the Nation:

U.S. Mood Throughout the Day inferred from Twitter

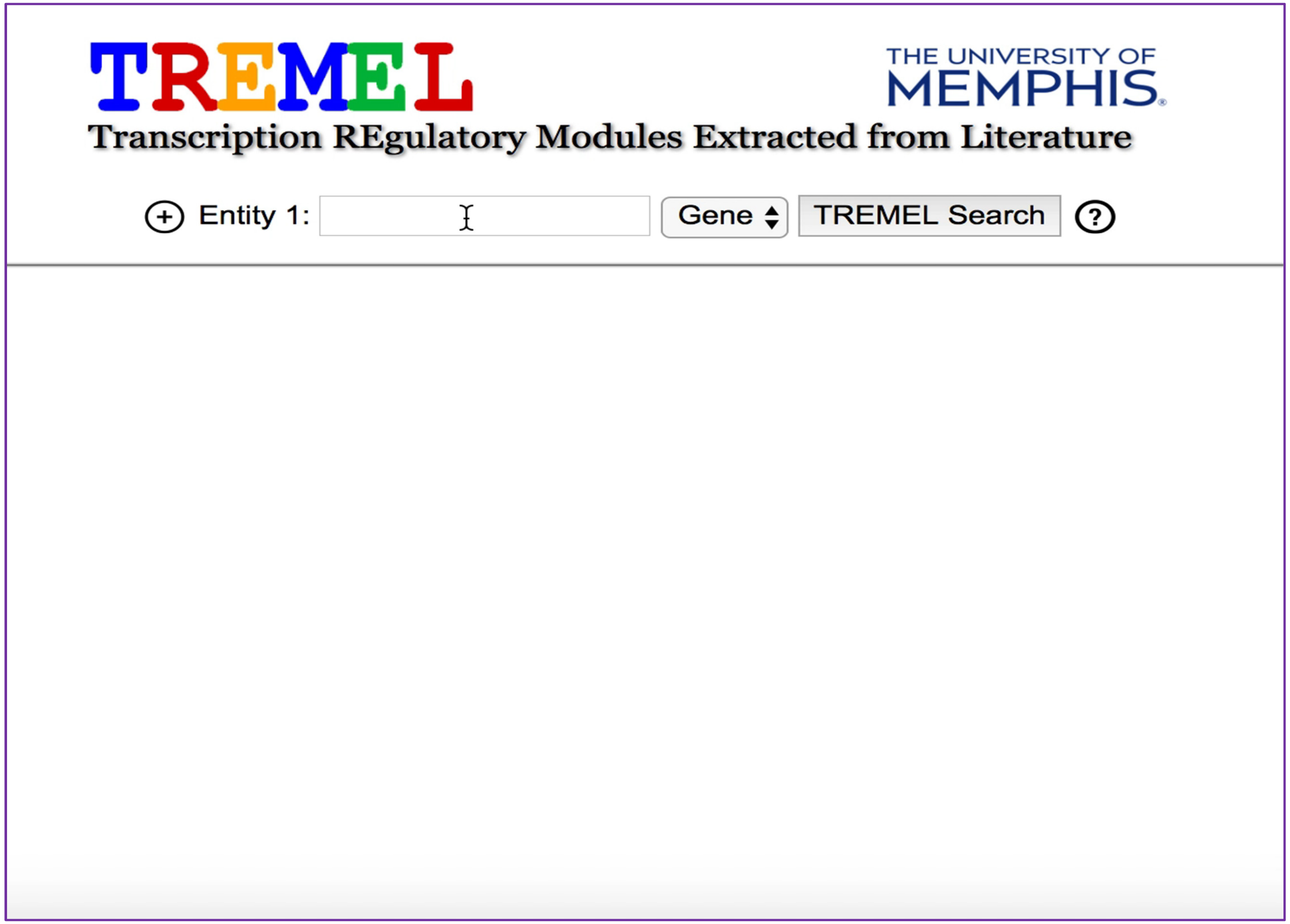
Less Happy More Happy

http://www.ccs.neu.edu/home/amislove/twittermood

Examples: Visualizing Spatial Temporal Data

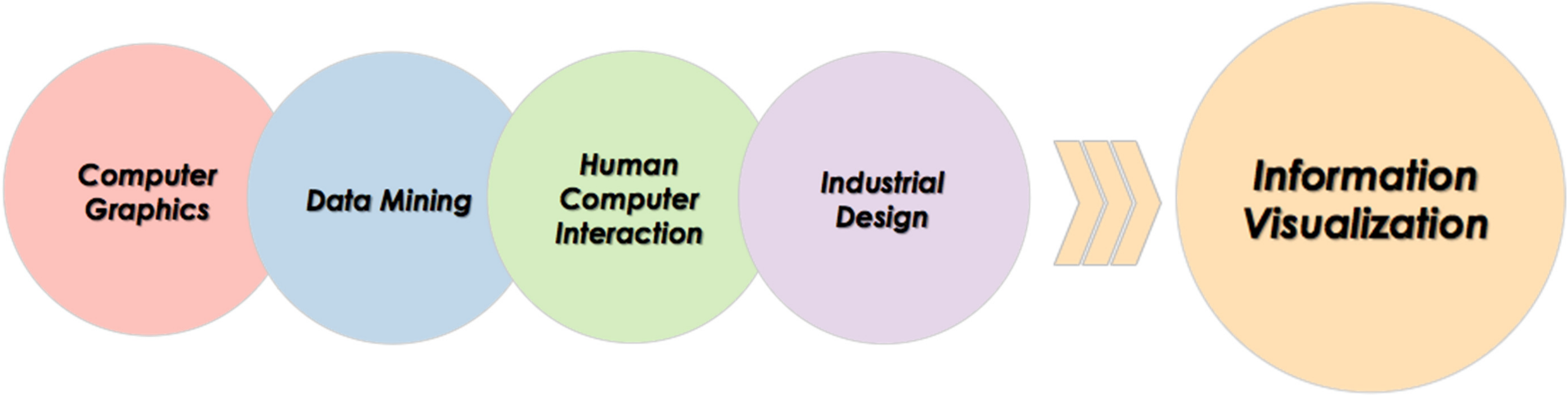


Query based (huge) graph visualization and interaction

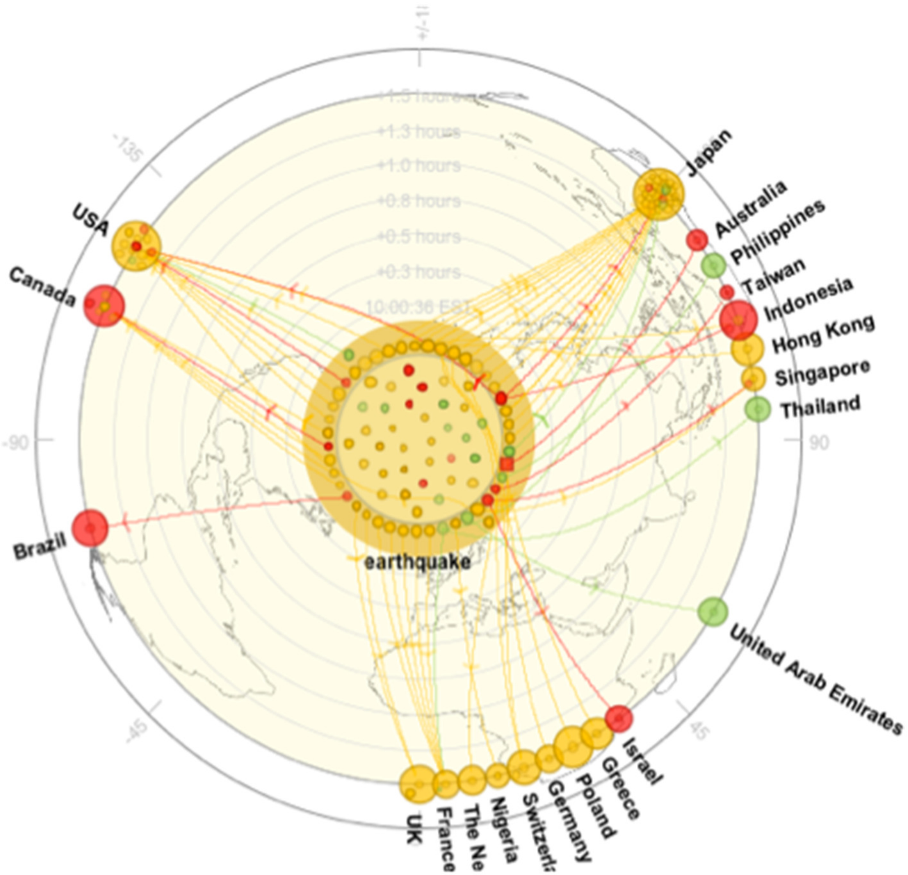


http://binf1.memphis.edu/tremel/

An Interdisciplinary Field

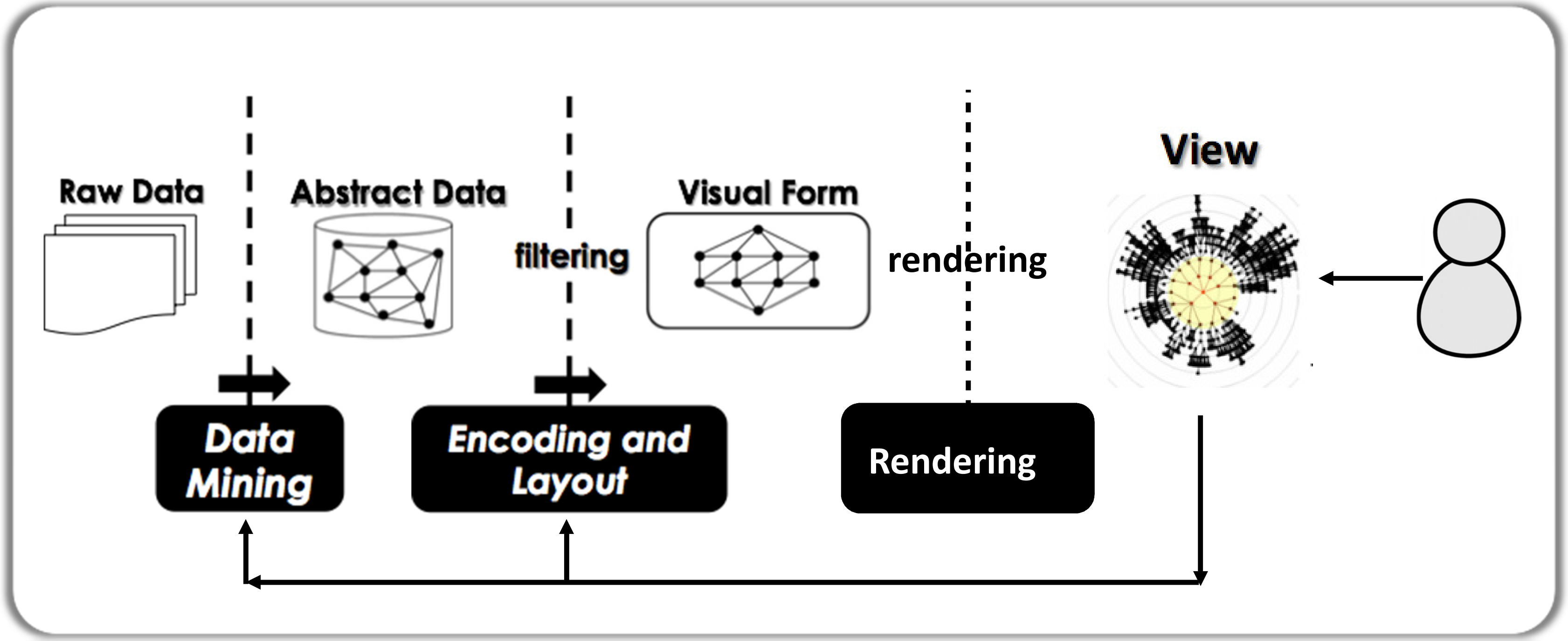


Visualization is not just about producing a beautiful picture



<u>The purpose of visualization is to reveal the insight of the data!</u>

Visualization & Visual Analysis Reference Model



Functions of Visualization

- Record information (store & summarize):
  - Photographs, blueprints, …
- Explore information (analyze):
  - Process and calculate
  - Reason about data
  - Feedback and interaction
- Explain information (present):
  - Convey information to others
  - Share and persuade
  - Collaborative and revise
  - Emphasize important aspect of data

Big Data Visualization Challenges and Techniques

Big Data Visualization

76425 species

Tree of Life by Dr. Yifan Hu

14.8 million tweets

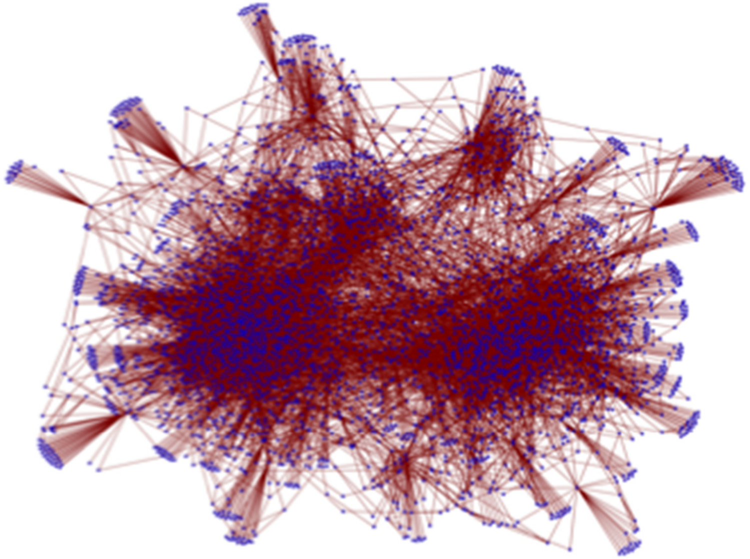
The information diffusion graph of the death of Osama bin Laden by Gilad Lotan

500 million users

Facebook friendship graph by Paul Butler

Challenging Task:
Squeezing millions and even billions of records into million pixels (1600 X 1200 = 2 million pixels)

Challenges



Visual clutter

How can we avoid visual clutters 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?

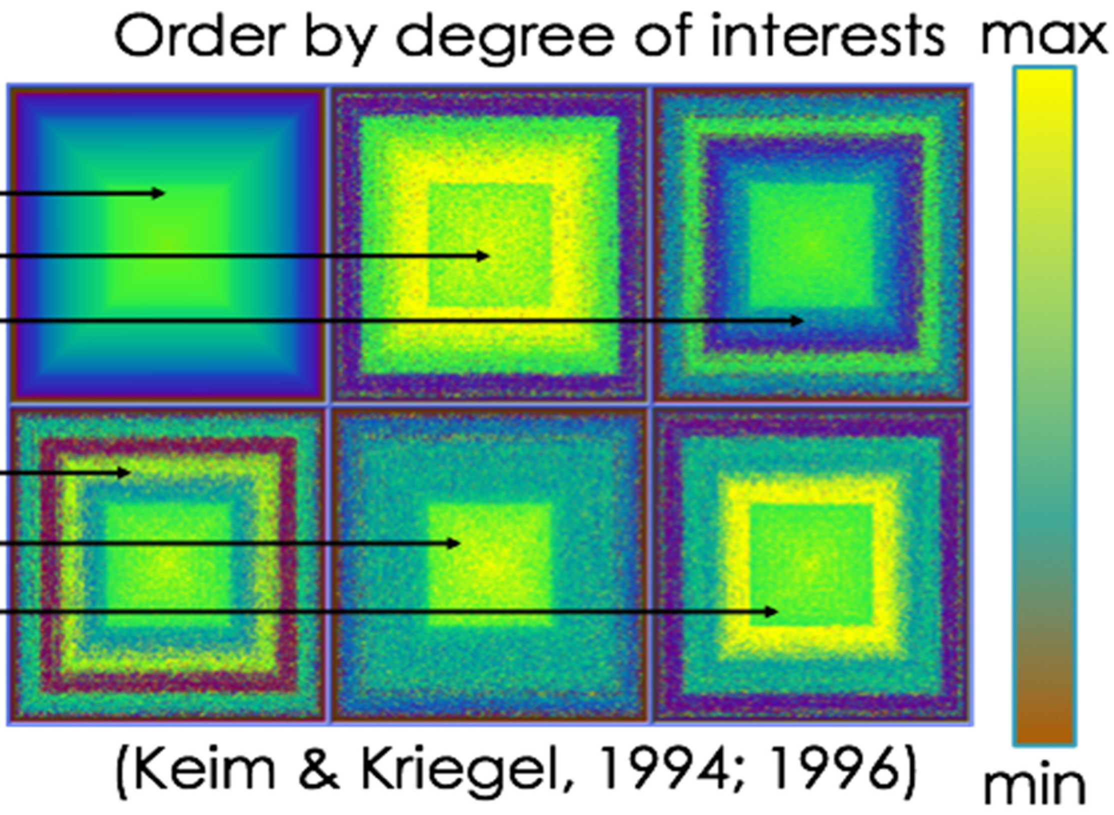
Techniques (1) : Pixel Oriented Visualization

- A multidimensional data item contains 6 attributes

Techniques (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	332.07
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	356.67	356.13
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

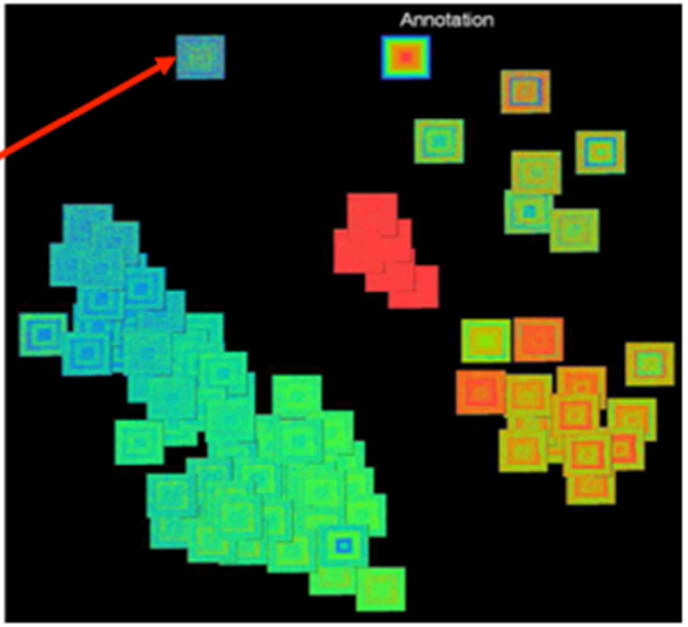


Techniques (1) : Pixel Oriented Visualization

• Different Ways for splitting the display region

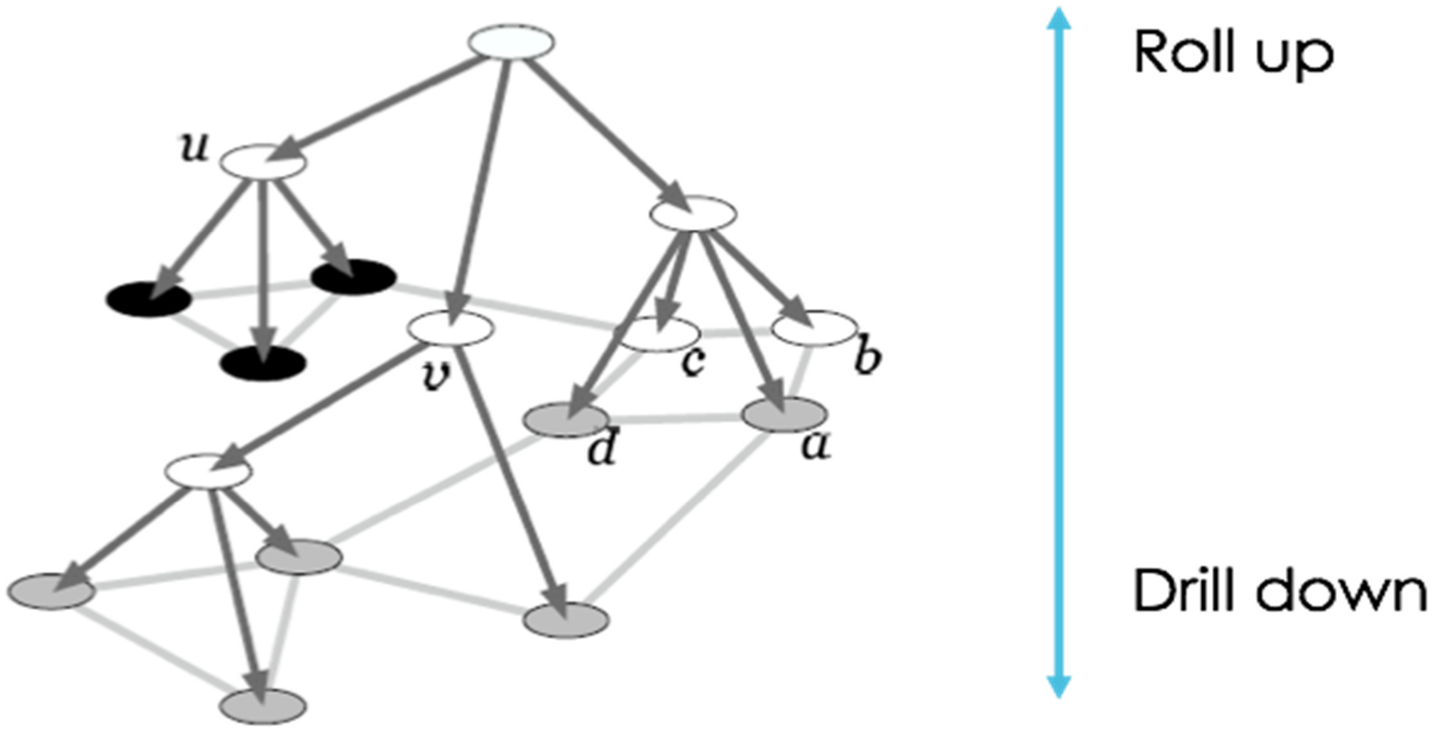
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual Average
1958	-99.99	-99.99	315.7	317.45	317.5	317.26	315.86	314.93	313.2	312.44	313.33	314.6	-99.99
1959	315.62	316.38	316.71	317.72	318.29	318.16	316.54	314.8	313.84	313.26	314.8	315.58	315.98
1960	316.43	316.97	317.58	319.02	320.03	319.59	318.18	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	-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	321.08	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	323.05
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	[illegible]
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	[illegible]
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	348.93
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	358.6	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	368.87	369.59	371.14	371	370.35	369.27	366.93	364.63	365.13	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	381.74	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	383.69	383.55
2008	385.07	385.72	385.85	386.71	388.45	387.64	386.1	383.95	382.91	382.73	383.96	385.02	385.34

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.



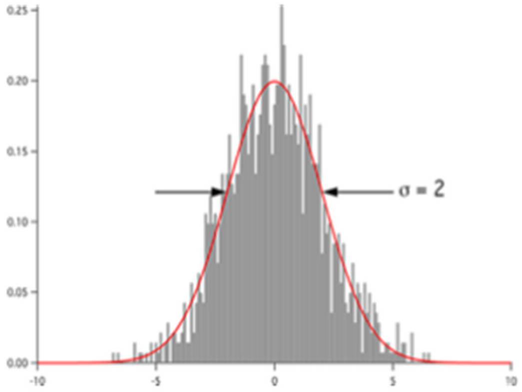
(Yang et al., 2006)

Techniques (2): Aggregation & Level of Details (LOD)

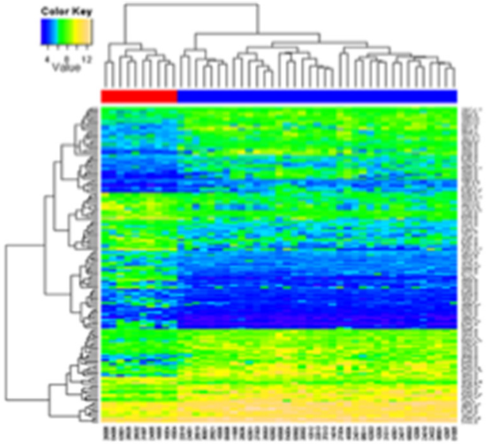


Building a tree for aggregating data items in either a bottom-up or top-down approach

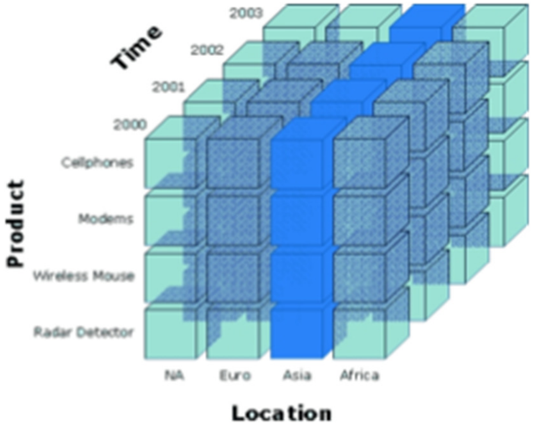
Technique (2) : Aggregation & LOD



Histogram (Pearson, 1895)



Heatmap
(Wilkinson & Friendly, 2009)

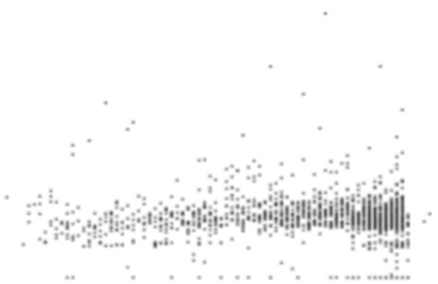


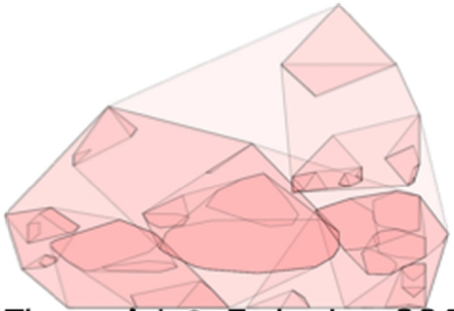
InfoCube
(Stolte et al., 2003)


(Lin et al., 2010)

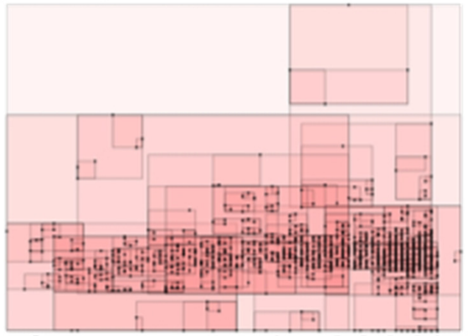
Techniques (2) : Aggregation & LOD

Scatter Plots



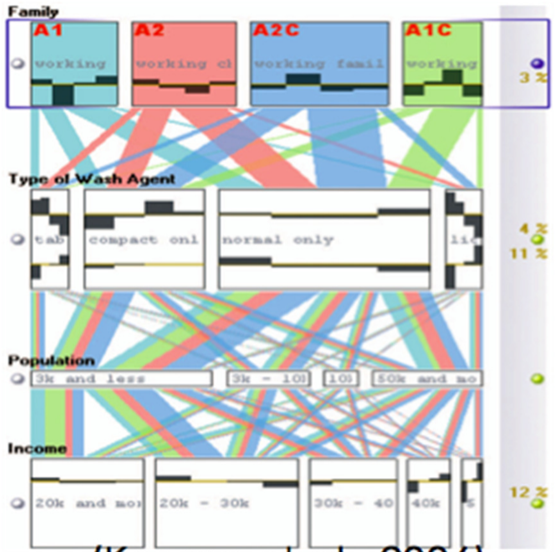


(Elmqvist & Fekete, 2010)

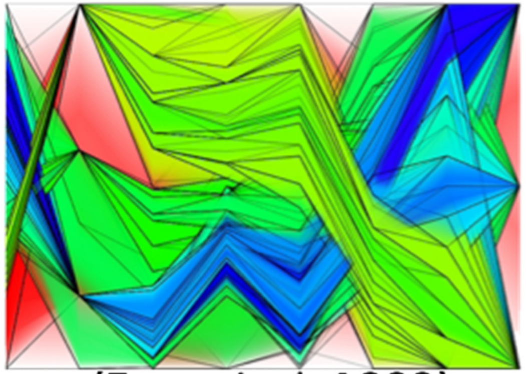


(Yang et al., 2003b)

Parallel Coordinates

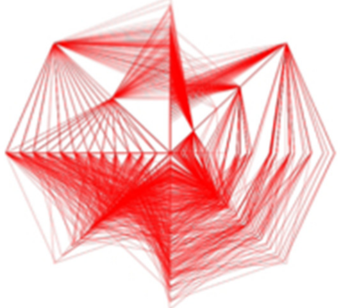


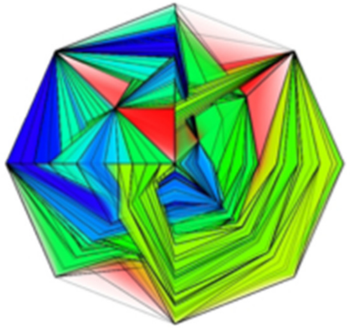
(Kosara et al., 2006)



(Fua et al. 1999)

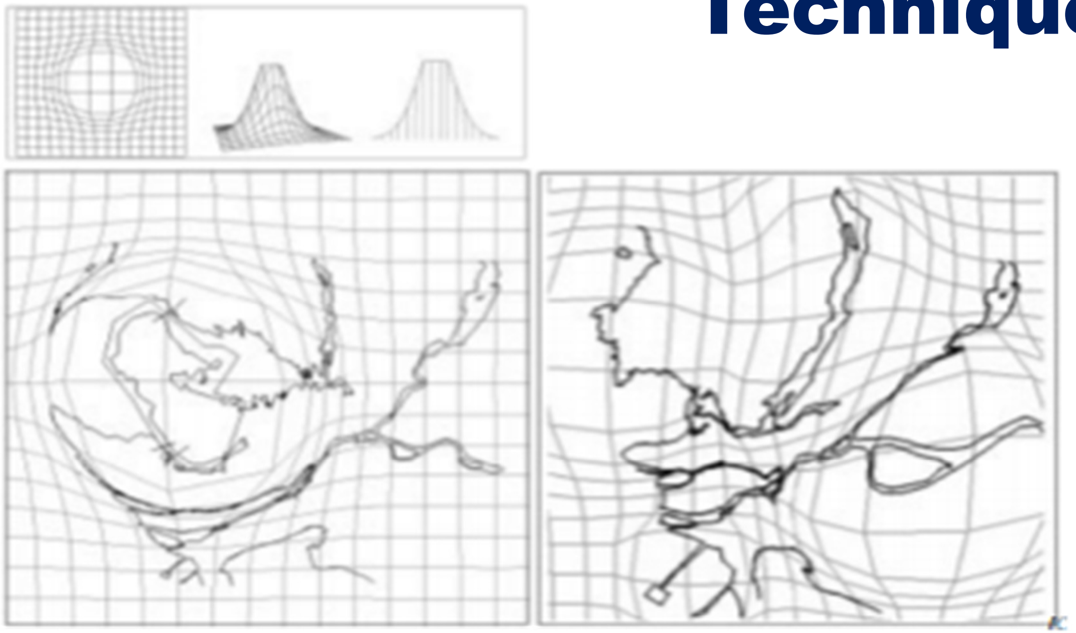
Star Plots

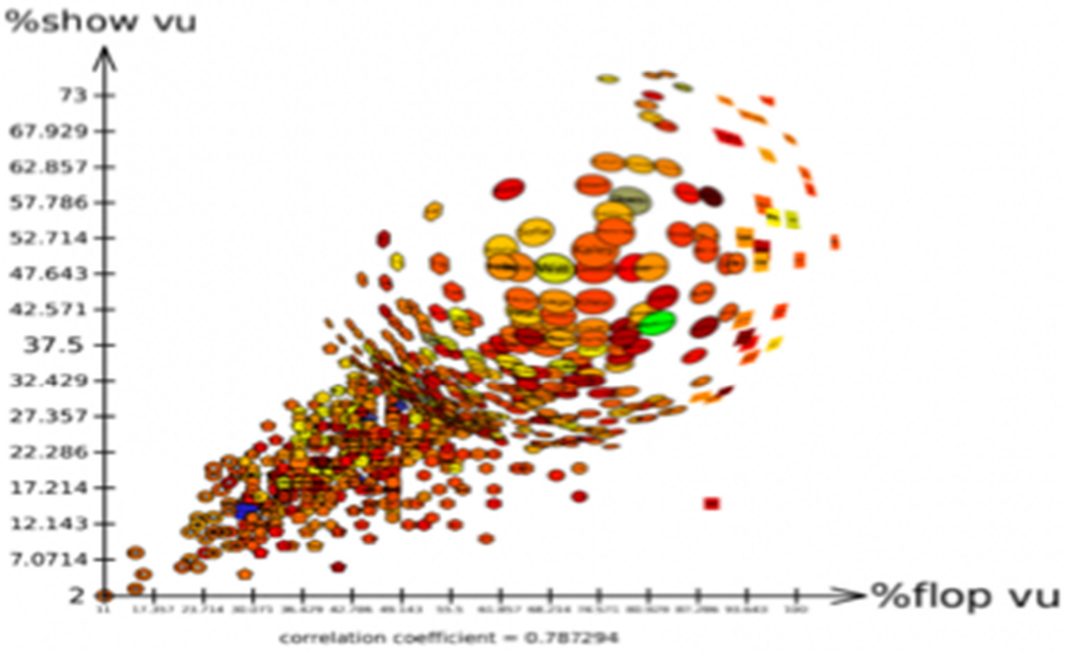


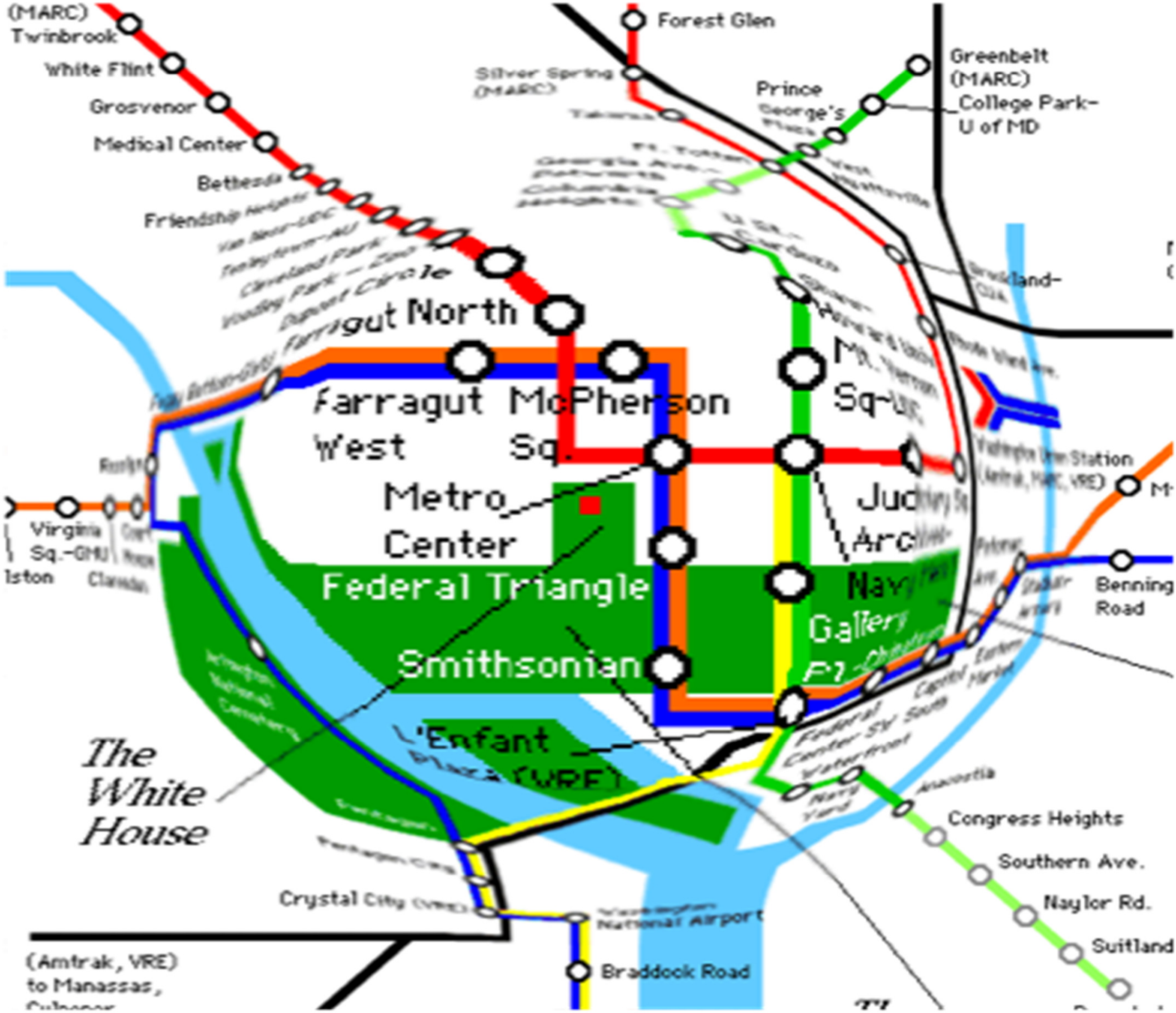


(Fua et al. 1999)

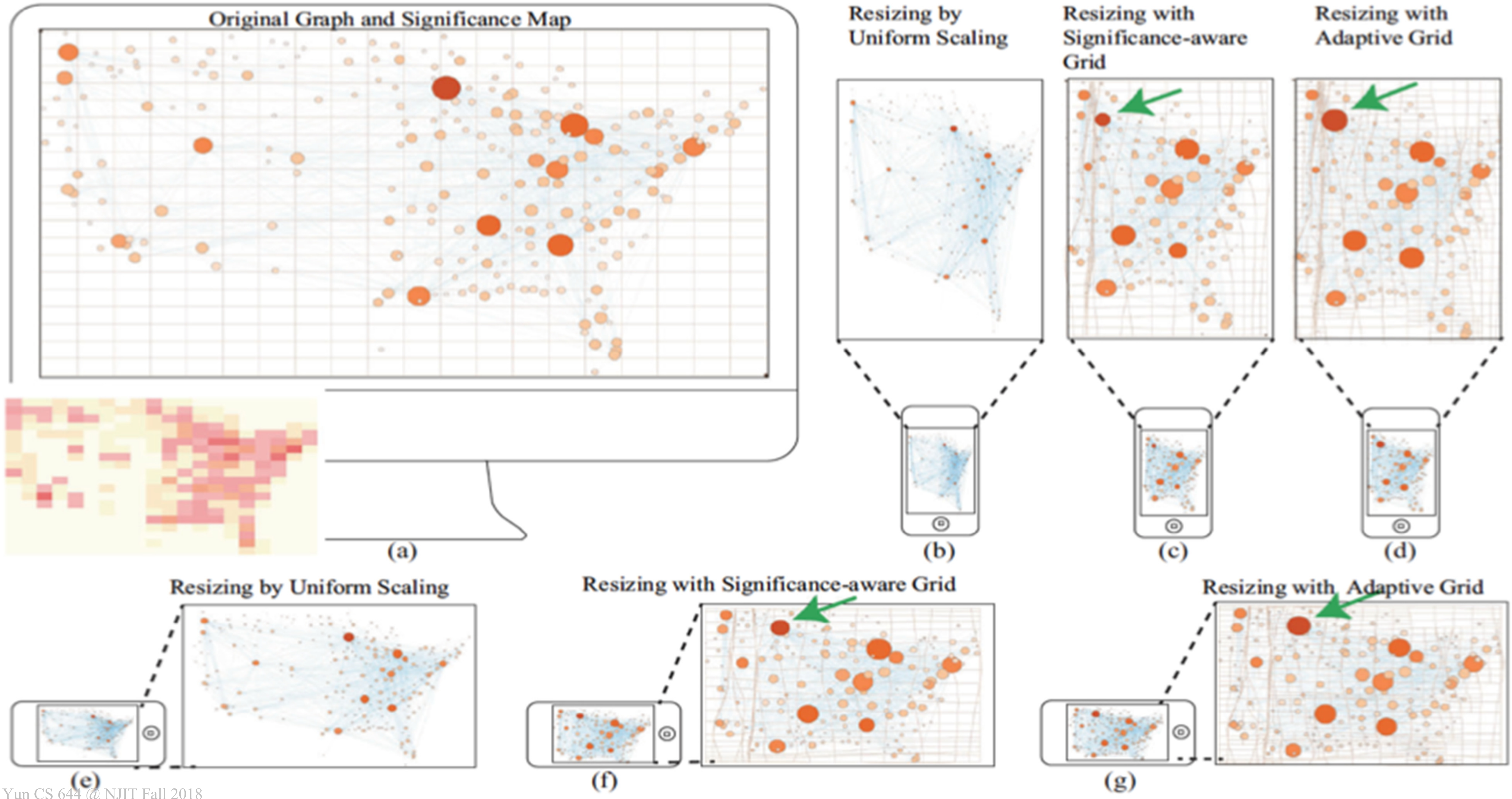
Technique (3) : Distortion



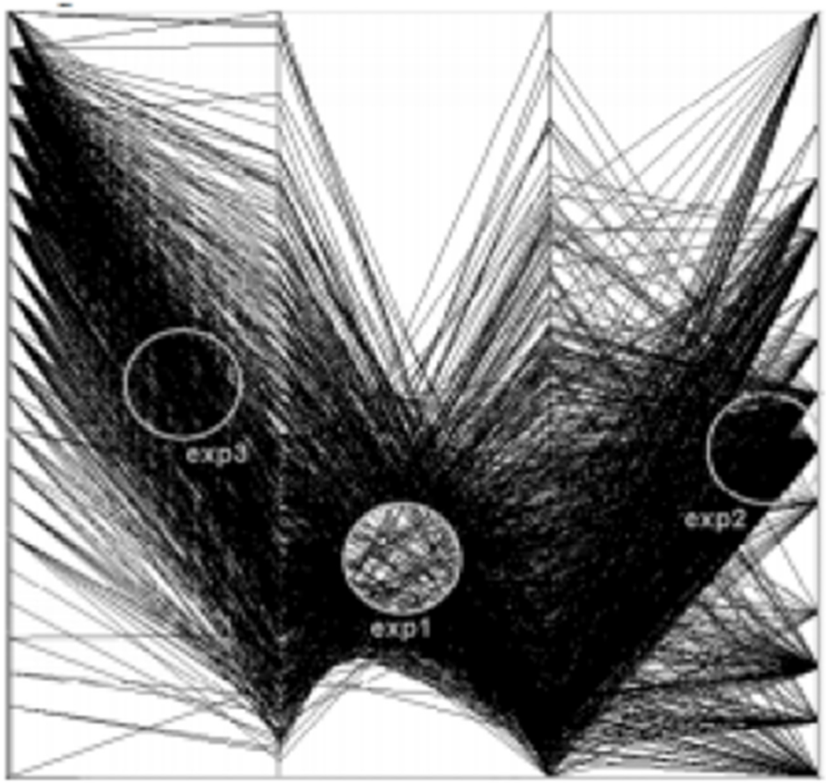




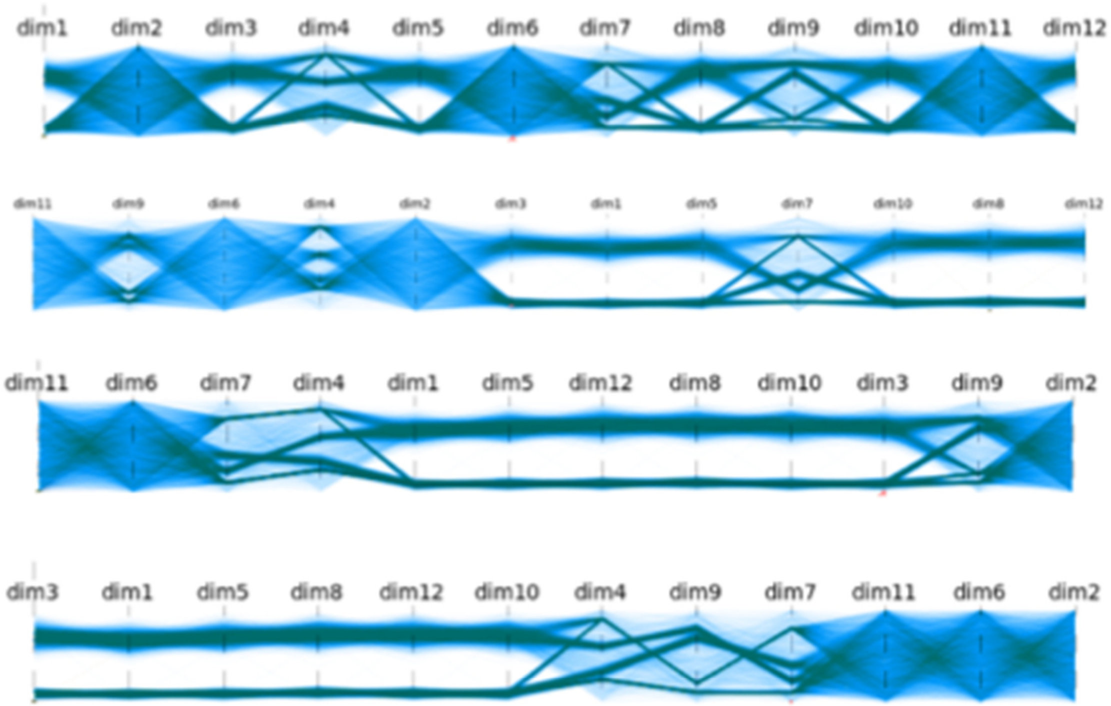
Technique (3) : Distortion



Technique (4) : Clutter Reduction

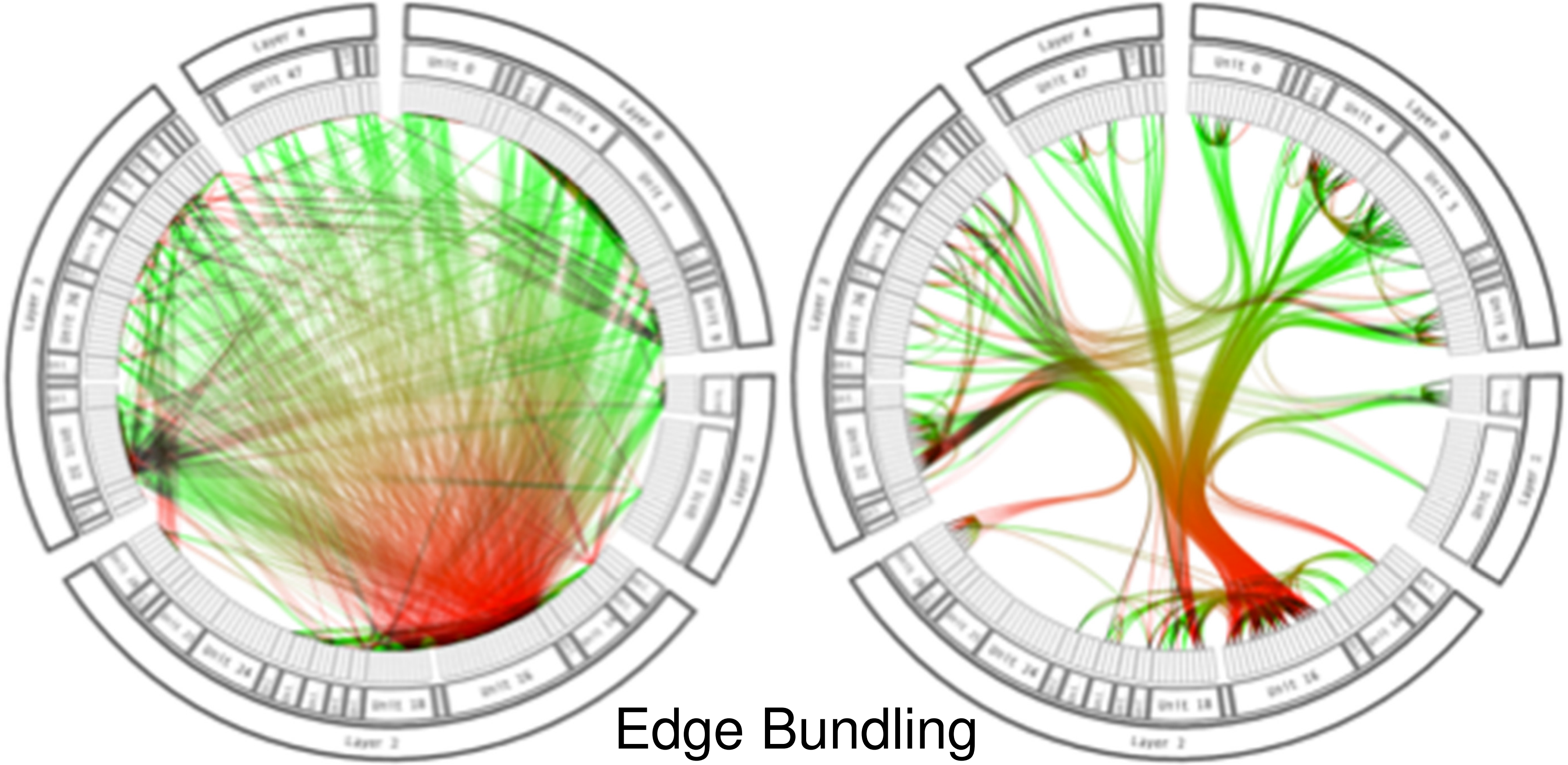


Sampling



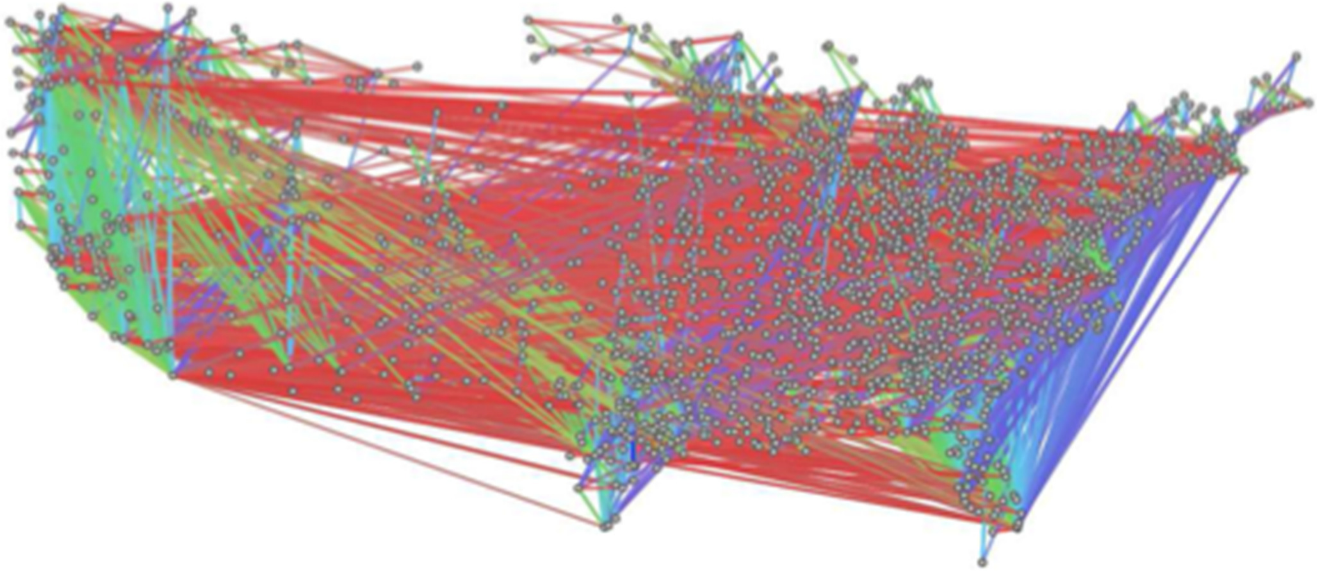
Reordering

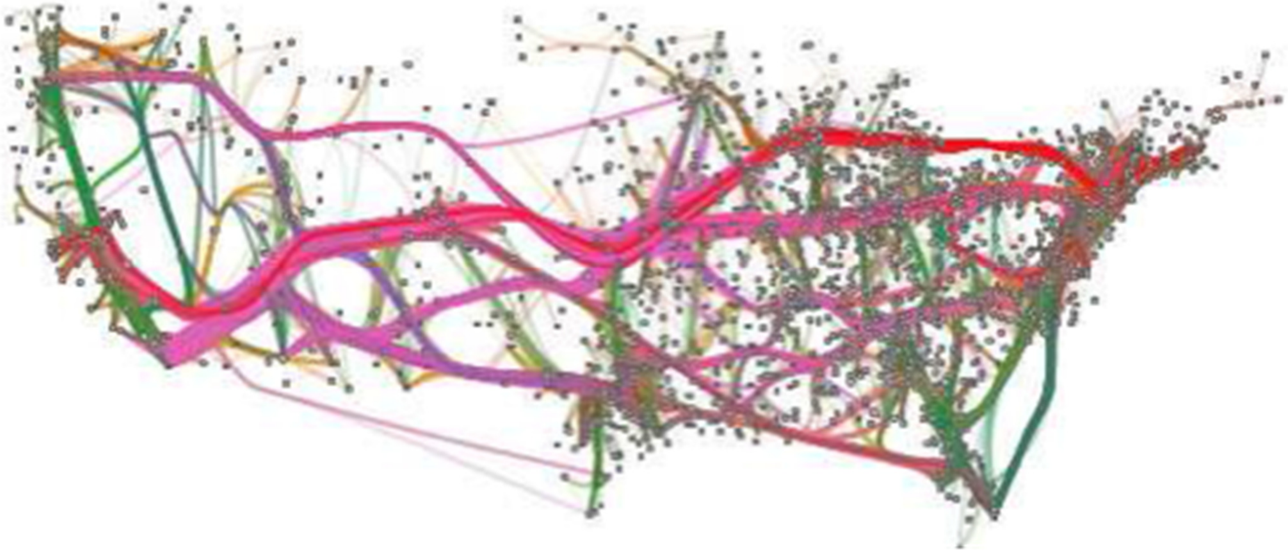
Technique (4) : Clutter Reduction



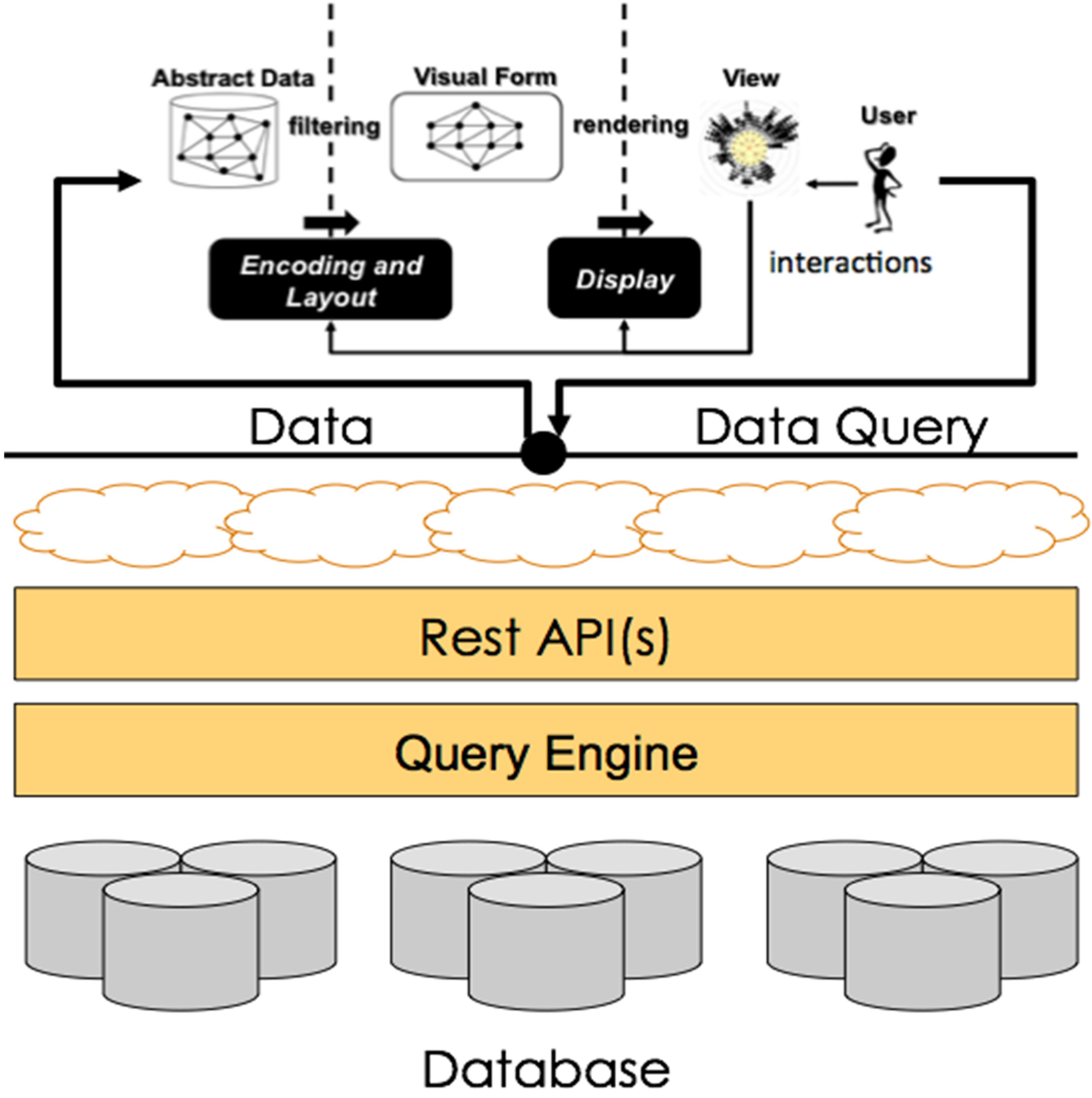
Edge Bundling

Technique (4) : Clutter Reduction





Technique (5): Query-based Visualization

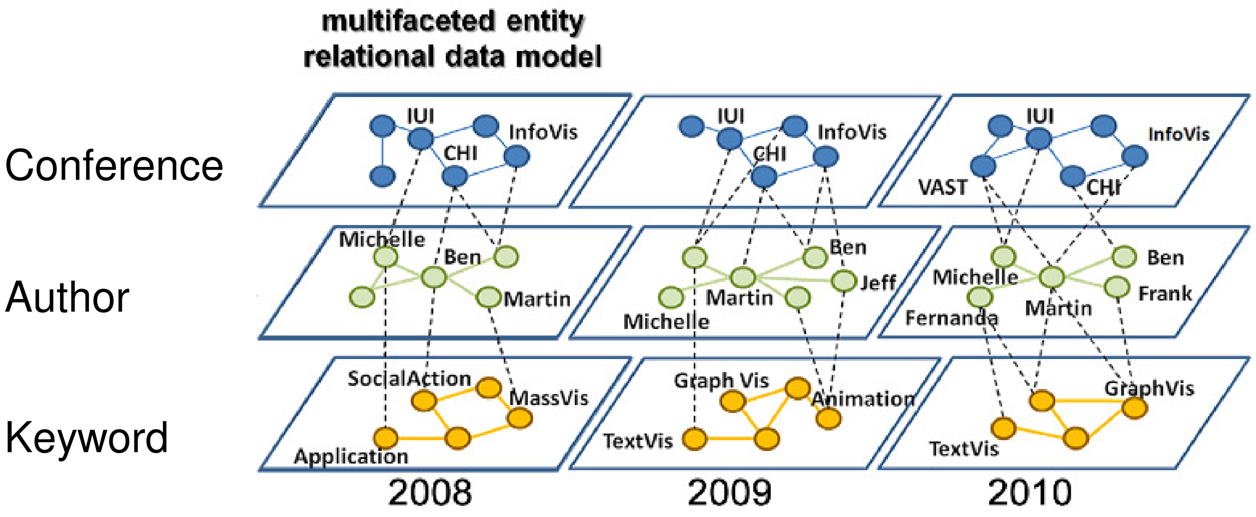


Case Study

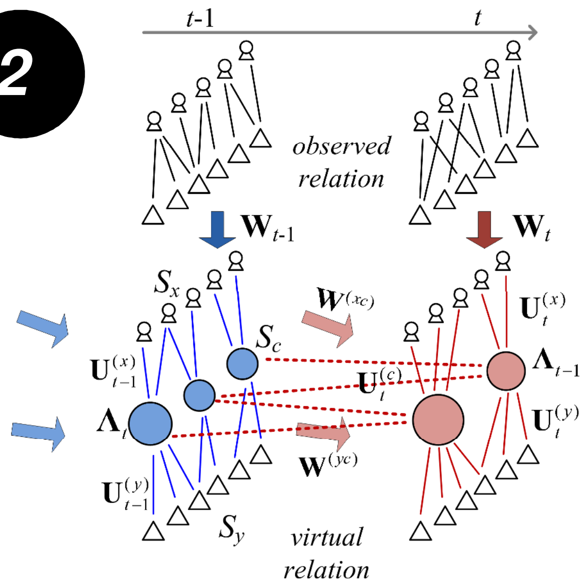
ContexTour:
Multifaceted Visuailzation of Research Communities

Context Tour Data Transformation & Analysis

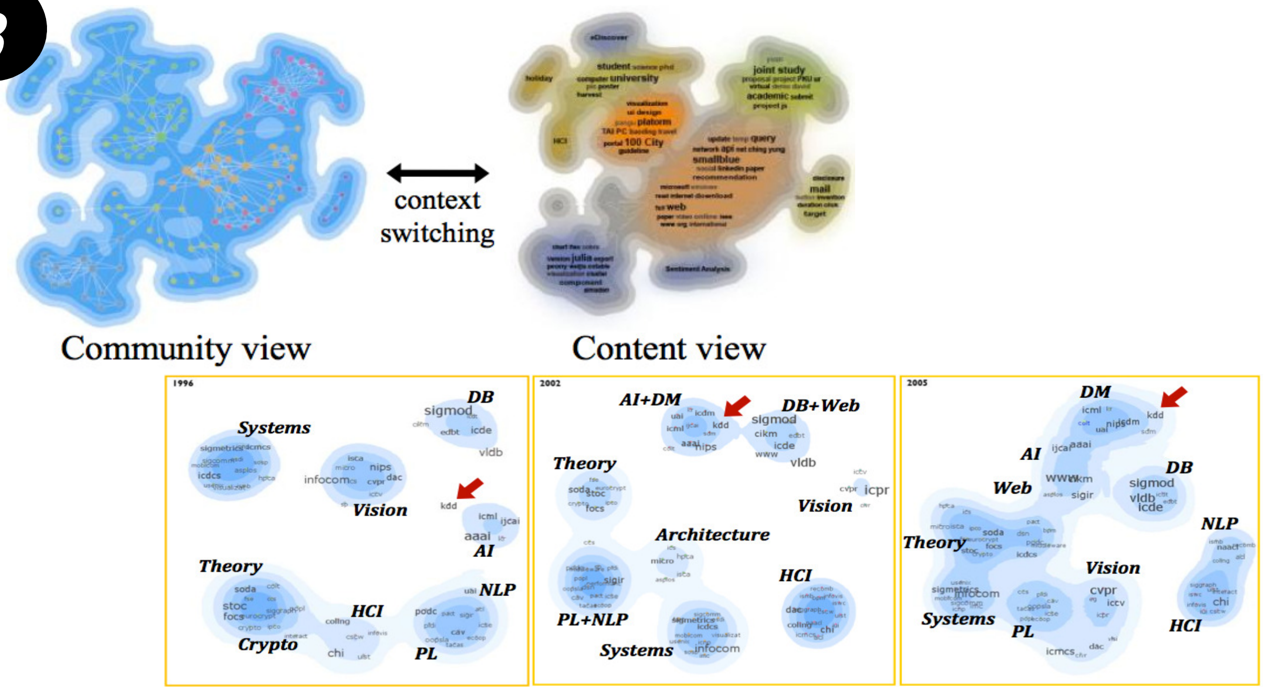
1

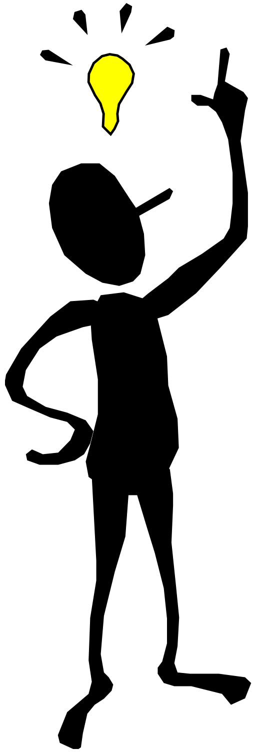


2



3





Thanks ! ☺

Questions ?