Scalable Visualization for Information Discovery

Hongfeng Yu
Associate Professor
Computer Science & Engineering, University of Nebraska-Lincoln

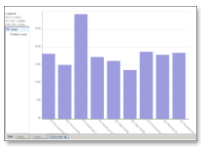
UNL CSCE 100
November 5, 2019

1

What is Visualization?

- **Communication** of **information** using **graphical** representations

- Everyday and everywhere
  - Bar chart in a newspaper
  - Subway map
  - Weather chart
  - Stock market analysis
  - ….



2

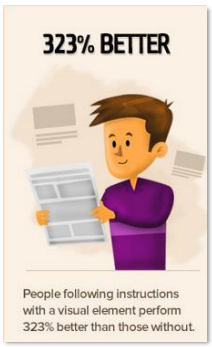
Why Visualization?

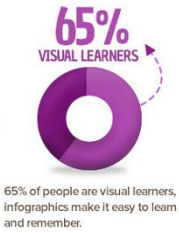
- Pictures have been used for communication since before the formalization of written language

- Our brain is well built for processing pictures
  - **Speed**
    - Image interpretation is performed in parallel
    - Text analysis is limited by sequential process of reading
    - One biological study estimates the transmission speed of the optic nerve at around 9Mb/sec

  - **Pattern matching**
    - Our visual system can quickly identify important patterns from massive data
      - E.g. face recognition from a large number of people

Why Visualization?





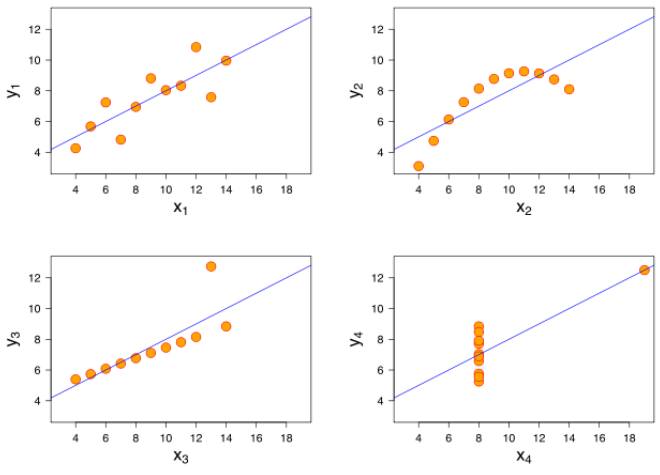
http://www.quicksprout.com/2013/08/08/this-is-your-brain-on-visualization/?display=wide

A Classic Example

- Four data sets
  - Each containing a set of data points in (x, y)

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

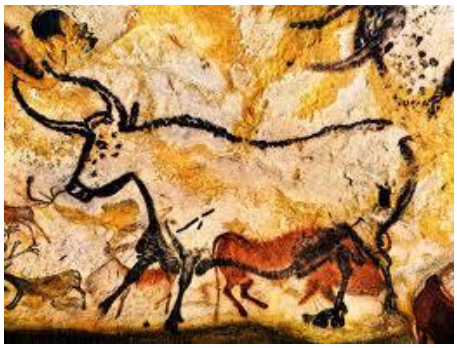
Magic!



Why is Visualization Challenging?

Early Visualization

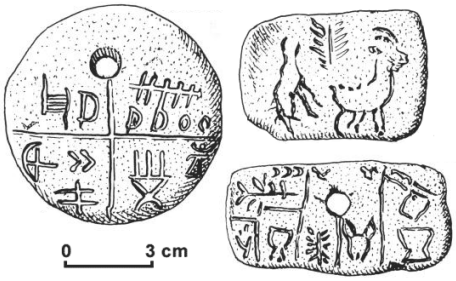
- Perhaps the first technique for **graphically** recording and presenting **information**



Cave paintings by early man approximately **30,000** years ago

Early Visualization

- Early graphical writing



Tartaria Tablets, **5500BC**

Kish Tablet, **3500 BC**

Early Visualization

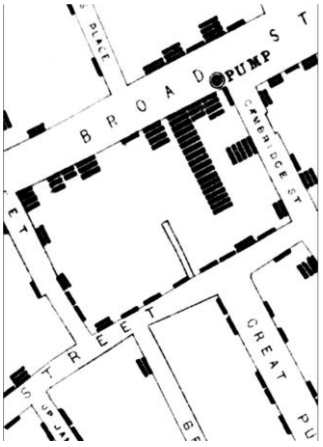
- Some necessary need for survival

Peutinger map: the road network in the Roman Empire.
Created in **15th century** based on a **4th-century map**

Early Visualization

The Lord of the Rings, 20th century

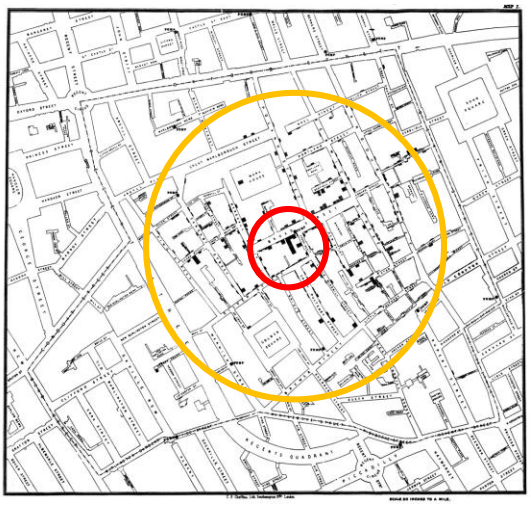
Early Visualization



A section of John Snow's map of the deaths from cholera in London in 1663.

Each bar within the houses represents one deceased individual.

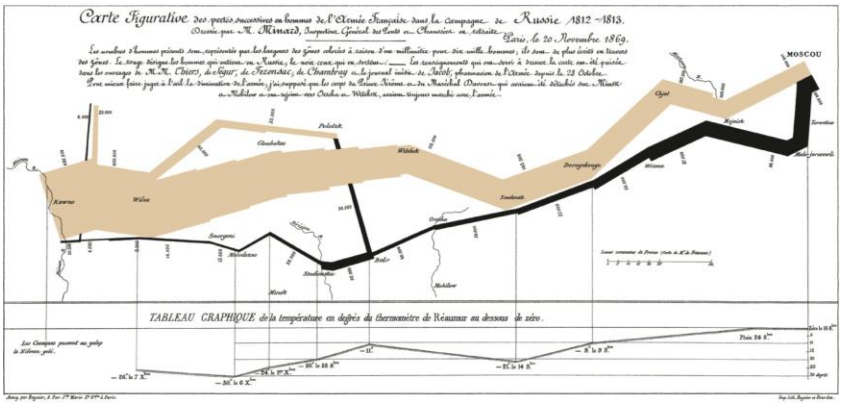
Early Visualization



Overview map of the deaths from cholera in London in 1663.

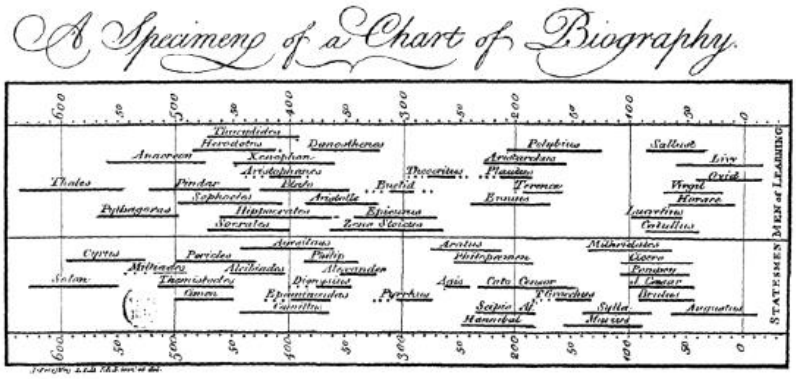
Note the concentration around the Broad Street Water Pump. Note as well the outliers.

Early Visualization



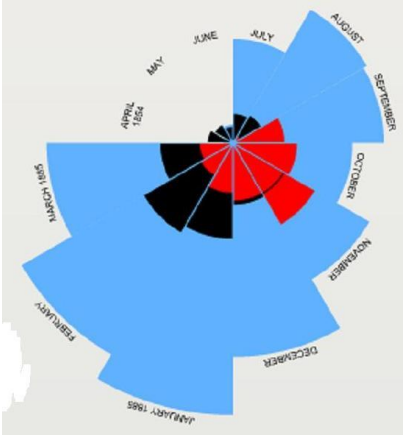
Minard's map, showing Napoleon's march on Moscow. The width of the line conveys the size of the army at that location. Color indicates the direction of movement. The temperature is plotted at different points along the retreat at the bottom. (1869)

Early Visualization



Joseph Priestley's display of the longevity of famous people (1765).

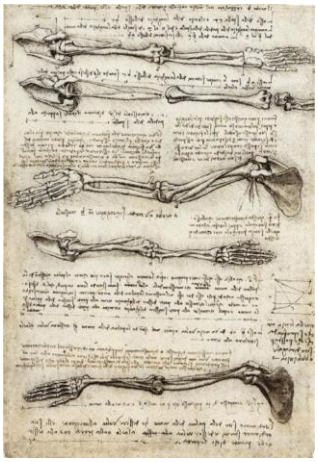
Early Visualization



Florence Nightingale's coxcomb chart showing monthly deaths from battle and other causes (1858).

Blue represents the deaths from disease, red represents deaths from wounds, and black represents all other deaths.

Early Visualization



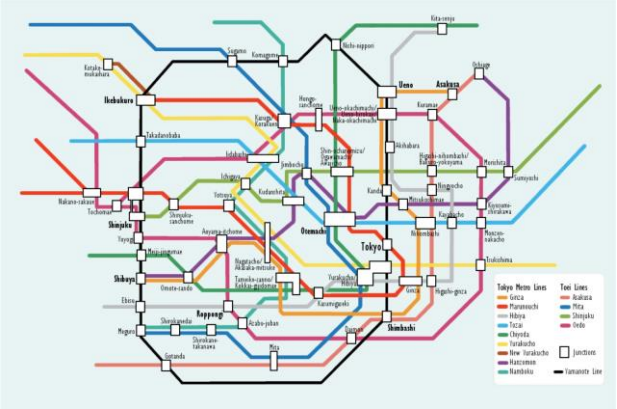
Leonardo Da Vinci's study of the motion of the human arm (1510).

Early Visualization

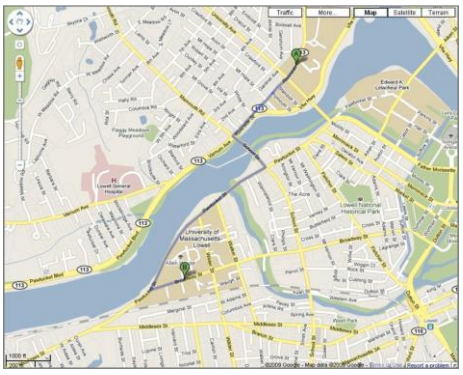
Leonardo Da Vinci's picture of water flow

Visualization Today

The Tokyo Underground map. A logical representation of the metro highlighting **qualitative** relationships between the stops.



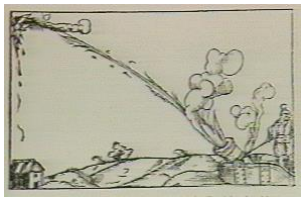
Visualization Today

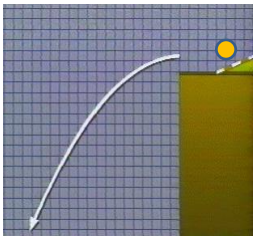


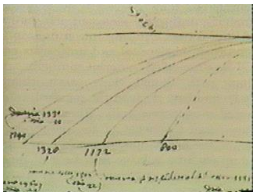
The google.com map directions from 198 Riverside St., Lowell, MA (UMass Lowell, North Campus) to 883 Broadway St., Lowell, MA (UMass Lowell, South Campus). Google.com maps provide graphical cues drawn on top of road maps to indicate driving directions from point A to point B.

Why is visualization a necessity?

- Galileo's Analysis of Projectile Motion







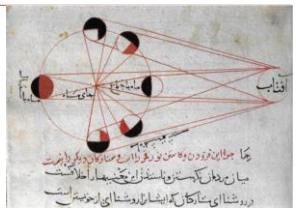
Parabola

$$y = ax^2 + bx + c$$

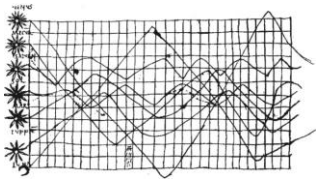
http://www.mcm.edu/academic/galileo/ars/arshtml/mathofmotion2.html

Why is visualization a necessity?

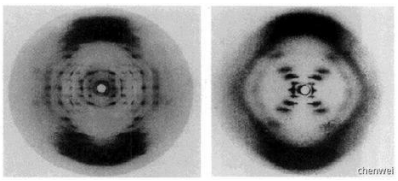
- Many many examples
  - Kepler's laws
  - Newton's laws
  - DNA structure
  - … …
  - Numerous scientific publications



Produced by Biruni circa 1030. Shows the phases of the moon in orbit.



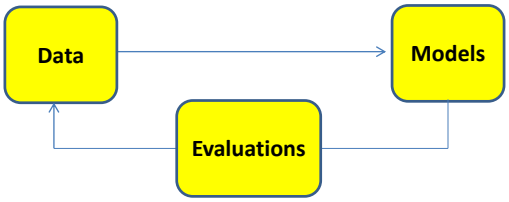
Planetary motion.



The x-ray diffraction image

Why is visualization a necessity?

- Human's Knowledge Discovery Pipeline



- Visualization
  - Can be used at every step of the KD pipeline
  - Often a part of this larger process
  - Tightly coupled with analysis

Visualization vs. Computer Graphics

- Computer Graphics
  - Graphical objects and organization of graphic primitives

- Visualization
  - More than simply computer graphics
  - Based on the underlying data (spatial positions, populations, or physical measures)
  - Include aspects from numerous other disciplines
    - Human-computer interaction
    - Psychology for human perception
    - Databases
    - Statistics
    - Data mining
    - …

Visualization vs. Computer Graphics

- Computer Graphics
  - Creation of images and animations for visual realism
  - Video games, cartoons, advertisements, and movie special effects
- Visualization
  - Not emphasize visual realism, but effective communication of information
  - Many applications do not deal with physical objects

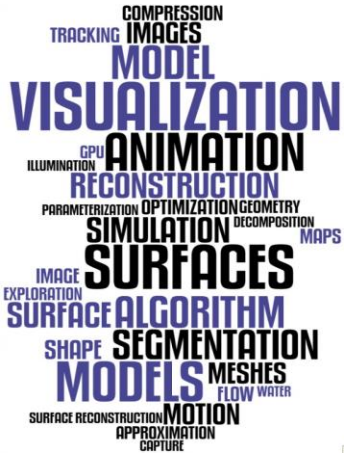
Visualization vs. Computer Graphics

- Computer Graphics provides tools for visualization
  - Graphics-programming language
    - OpenGL, DirectX, Processing, Java3D
  - Underlying graphics Hardware
    - Intel, Nvidia or AMD graphics cards
  - Rendering process
    - Different shading
  - Output format
    - JPEG, TIFF, AVI, MPEG

Visualization vs. Computer Graphics

- Paper keywords from Computer Graphics Forum

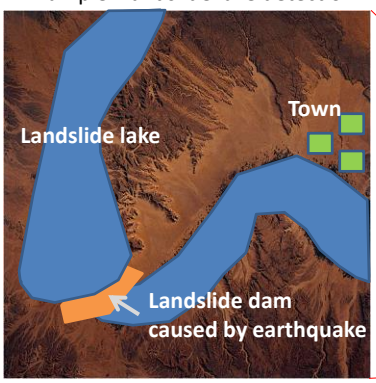
This image itself is a visualization. →

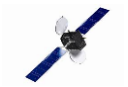


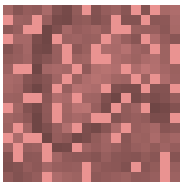
Why is Visualization Challenging in Big Data Era?

- Big Data – **Dilemma**

Example: Landslide lake detection







Coarser resolution: missing details

Finer resolution: Finding the needle in a haystack

Why is Visualization Challenging in Big Data Era?

Our Research Theme

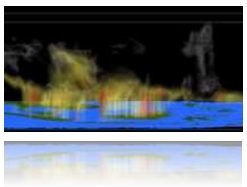
Our Research Theme

Holistically address every stage
of a visualization pipeline

Big Data → Processing & Analysis → Visualization → Information Discovery

Scalable Data Store

High-Performance Computing



Visual Scalability

Our Research Theme

Holistically address every stage
of a visualization pipeline

Big Data → Processing & Analysis → Visualization → Information Discovery

Scalable Data Store

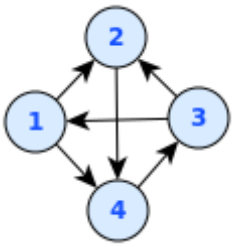
High-Performance Computing

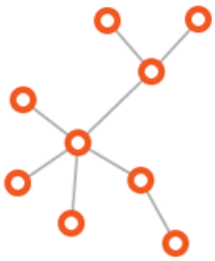
Visual Scalability

Visual Scalability

Massive Data vs. Limited Screen

Example: Traditional node-link diagram for network data

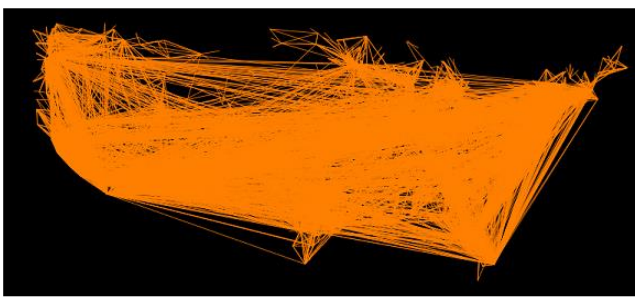




Visual Scalability

Massive Data vs. Limited Screen

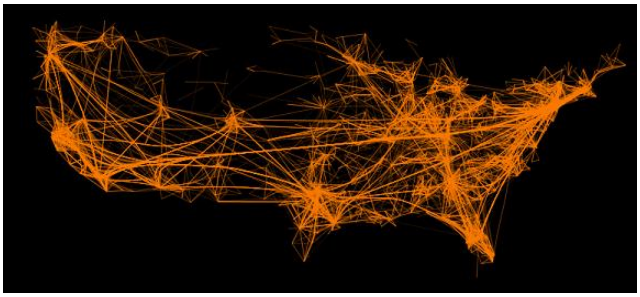
Example: Traditional node-link diagram for network data



Visual Scalability

Massive Data vs. Limited Screen

Our research: Edge bundling for network data



[Yves 2019]

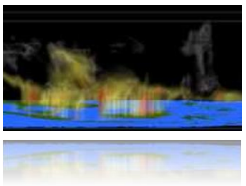
Our Research Theme

Holistically address every stage of a visualization pipeline

Big Data → Processing & Analysis → Visualization → Information Discovery

Scalable Data Store

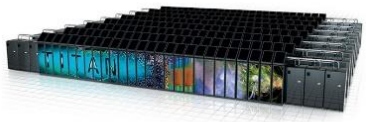
High-Performance Computing



Visual Scalability

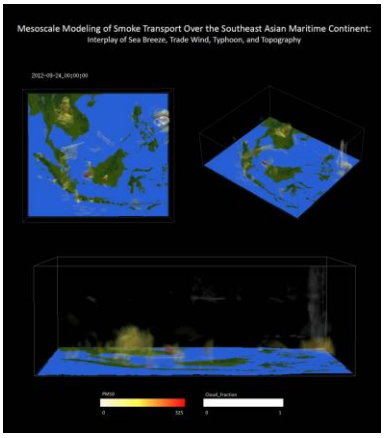
High-Performance Computing

Our research: Exploit high-performance compute technology
- Accelerate data processing
- Ensure **interactivity**



High-Performance Computing

Our research: Exploit high-performance compute technology
- Accelerate data processing
- Ensure **interactivity**



Our Research Theme

Holistically address every stage
of a visualization pipeline

Big Data → Processing & Analysis → Visualization → Information Discovery

Scalable Data Store | High-Performance Computing | Visual Scalability

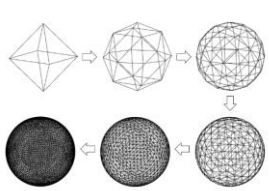
Scalable Data Store

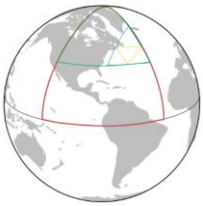
Our research: Design new distributed data indexing and placement solutions

- **Petabytes and beyond**
- **Diverse data types and structures**
  (vector, raster, relation, binary, and text)
- **Complex data presentations**
  (shapefiles, images, tables and point clouds)

Scalable Data Store

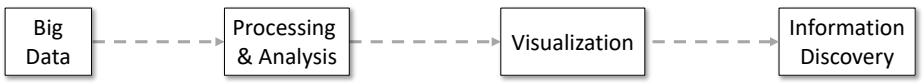
Our example work: SpatioTemporal Adaptive-Resolution Encoding (STARE)



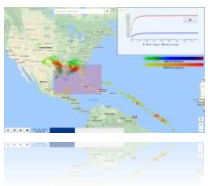


Our Research Theme

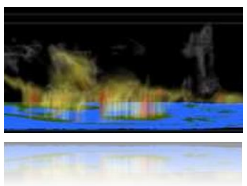
Lead to new scientific applications and discoveries



Scalable Data Store



High-Performance Computing

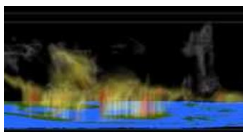


Visual Scalability

Our Research Theme

Scalable Data Store

High-Performance Computing



Visual Scalability

Thank You!