



Lecture 11: Information Search and Visualization

Naureen Nizam

CSCC10H3: Human-Computer Interaction

Department of Computer and Mathematical Science

July 27, 2022



UNIVERSITY OF TORONTO SCARBOROUGH
1265 Military Trail, Toronto, Ontario M1C 1A4

Administrivia

- Project:
 - P4 –
 - Presentations during tutorials (Aug 4, 5, 9, 10) via Zoom
 - Presentation Slide Deck Due: Aug 3
 - Report Due – Aug 10
- Tutorial (P4 Report Writing)
- Discussion Question – this is the last one
- Final Exam
 - August 22nd 9 – 11 am (MW 170)
- Any Questions?

The plan for today...

- Information Search
- Data Visualization
 - Netlytics



Information Search

Topics:

1. Introduction
2. Five-phase search framework
3. Dynamic queries and faceted search
4. Command languages and “natural” language queries
5. Multimedia Document Search & specialized search
6. The Social aspects of search

Information Search

- Information search should be a joyous experience, but it still can take a skilled searcher with robust tools to perform an effective search
- Information retrieval and database management have evolved into:
 - Information seeking, filtering, collaborative filtering, sensemaking, and visual analytics.
- All the above is complicated by the increased volume of material to search
 - Data mining
 - Deep learning

5

Information Search (cont'd)

<https://www.loc.gov/>



Search Terminology

- **Task objects** (such as movies for rent) are stored in structured relational databases, textual document libraries, or multimedia document libraries
- A **structured relational database** consists of *relations* and a *schema* to describe the relations
- **Relations** have *items* (usually called *tuples* or *records*), and each item has multiple *attributes* (often called *fields*), which each have *attribute values*
- A **library** consists of a set of *collections* plus some *descriptive attributes* or *metadata* about the library (for example, name, location, owner)

7

Search Terminology (cont'd)

- **Digital libraries** are generally sets of carefully selected and cataloged collections
 - *Digital archives* tend to be more loosely organized
- **Directories** hold metadata about the items in a library and point users to the appropriate locations
 - for example, the NASA Global Change Master Directory helps scientists locate datasets in the many NASA's archives
- Items in **unstructured collections** like the web have no (or very few) attributes

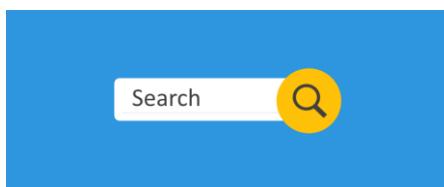
8

Search Terminology (cont'd)

- Task actions are decomposed into *browsing* or *searching*
- Here are some examples of task actions:
 - **Specific fact finding (known-item search)**
 - Find the e-mail address of the President of the United States.
 - **Extended fact-finding**
 - What other books are by the author of "Jurassic Park"?
 - **Exploration of availability**
 - Is there new work on voice recognition in the ACM digital library?
 - **Open-ended browsing and problem analysis**
 - Is there promising new research on fibromyalgia that might help my patient?

9

5-Phase Framework for Search User Interfaces



1. *Formulation*: expressing the search
2. *Initiation of action*: launching the search
3. *Review of results*: reading messages and outcomes
4. *Refinement*: formulating the next step
5. *Use*: compiling or disseminating insight

10

1. Formulation

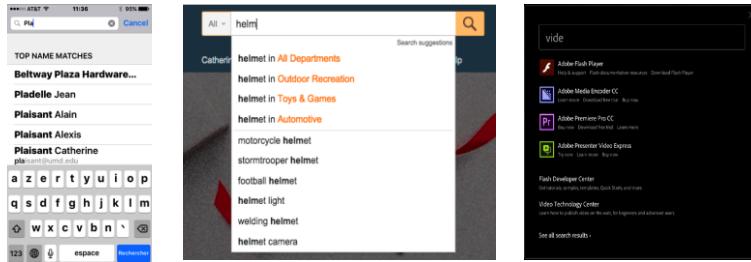
- Use simple and advanced search
- Limit the search using structured fields such as year, media, or location
- Recognize phrases to allow entry of names, such as George Washington
- Permit variants to allow relaxation of search constraints (e.g. phonetic variations)
- Control the size of the initial result set.
- Use scoping of source carefully
- Provide suggestions, hints, common sources

An Example – Search Formulation

The screenshot shows the Yahoo! Help search results for the query "boolean queries". The page has a dark blue header with the Yahoo! logo and navigation links for Home, Mail, Search, News, Sports, Finance, Weather, Games, Answers, Screen, Flickr, Mobile, and More. Below the header is a search bar containing the query. To the right of the search bar are buttons for "Search Help" (purple) and "Search Web" (blue). Further right are links for "Sign In" and "Mail". The main content area has a purple banner at the top with "All Results" and a "Search" button. Below the banner, there is a section titled "Yahoo Help Community Forums: Join the discussion" with a link to "Search Results". There is also a section about "Ads in Yahoo Search results" and a link to "Report abuse or spam on Yahoo". At the bottom, there is a list of bullet points:

- Yahoo help search box has 2 buttons of different colors to search two different sources of information
 - Purple for searching the help information and Blue for searching the web
 - Pressing the purple button "scopes" the results to the help information only and shows results below a purple banner
 - Searching the web jumps to a different page (the normal search) that reuses the blue button color, helping users keep track of which

More Search Formulation Examples



a) Apple iPhone contact manager, b) Amazon.com c) Adobe webpage

- Autocomplete suggestions can speed data entry and guide users toward successful queries
 - a) In a mobile phone address book typing one character filters the list to all names that contain that character, and the lists is updated continuously as users type
 - b) Typing "helm" in Amazon's search box shows suggestions for "helmet light" or "welding helmet" but also suggestions to narrow the scope of the search to relevant departments
 - c) In the Adobe website suggestions include products, e.g. typing the beginning of the word "video" suggests several video editing tools

13

2. Initiation of action

- Explicit actions are initiated by buttons with consistent labels (such as "Search")
- Implicit actions are initiated by changes to a parameter and update results immediately
- Guide users to successful or past queries with auto-complete

A screenshot of a search bar from the Library of Congress website. The search bar has a dropdown menu set to 'Everything'. To the right of the search bar is a large black search button with a white magnifying glass icon.

3. Review of results

- Keep search terms and constraints visible
- Provide an overview of the results (e.g. total #)
- Categorize results using metadata (by attribute value, topics, and so on).
- Provide descriptive previews of each result item
- Highlight search terms in results
- Allow examination of selected items
- Provide visualizations when appropriate (e.g. maps or timelines)
- Allow adjustment of the size of the result set and which fields are displayed.
- Allow change of sequencing (alphabetical, chronological, relevance ranked, etc.)

Review of Results Page

Google HCIL twinlist

Web Maps News Shopping Images More Search tools

About 560 results (0.84 seconds)

Twinlist and Medication Reconciliation Interfaces
www.cs.umd.edu/hcil/sharp/twinlist/ ▾ University of Maryland, College Park ▾
Novel User Interface Design for ... June 2015: Story on our Twinlist prototype for
medication reconciliation appears in User Experience, the Magazine of the User
Experience Professionals Association. ... Those interfaces were built on the substratum
of a novel medication reconciliation ...
You've visited this page many times. Last visit: 10/19/15

Ben Shneiderman, Catherine Plaisant and HCIL Twinlist ...
https://www.cs.umd.edu/~ben-shn... ▾ University of Maryland, College Park ▾
Nov 21, 2013 - Ben Shneiderman, Catherine Plaisant and HCIL Twinlist Team receive
Distinguished Paper Award at the American Medical Informatics ...

ManyLists
www.cs.umd.edu/hcil/manylists/ ▾ University of Maryland, College Park ▾
We applied the design concept from our recent comparison tools: Twinlist. It aims to
meet the following three goals: 1. Support the comparison of at least four ...
You've visited this page 4 times. Last visit: 9/8/15

Novel user interface design for medication reconciliation: an ...
www.ncbi.nlm.nih.gov/... ▾ National Center for Biotechnology Information ▾
by C Plaisant - 2015 - Related articles
Feb 8, 2015 - (2)Human-Computer Interaction Lab and Department of Computer ...
www.cs.umd.edu/hcil/sharp/twinlist compared to a Control interface ...
You've visited this page 3 times. Last visit: 9/29/15

TwinList - AMIA 2011 demo - YouTube
https://www.youtube.com/watch?v=YoSxIK0pCo
Oct 28, 2011 - Uploaded by Catherine Plaisant

Another Results Page Example

17

Another Results Page Example

18

4. Refinement

- Guide users in progressive refinement with meaningful messages
- Make changing of search parameters convenient
 - Provide related searches
 - Provide suggestions for error correction (without forcing correction)

5. Use

- Embed actions in results when possible
- Allow queries, settings and results to be saved and annotated, sent to other applications
- Explore collecting explicit feedback (ratings, reviews, like, etc.)

Use of the Results Page

The figure shows two side-by-side screenshots of search results pages. The left screenshot is from Google Search, showing a search bar with 'how to call the fire department?' and a results page with a snippet about calling 911 for emergencies. The right screenshot is from Peapod.com, showing a search bar with 'grape' and a results page displaying a grid of organic grape products.

View as:	Grid	List
Qty		Qty
Description		Description
Size		Size
Unit Price		Unit Price
Price		Price

90 Results for: "grape"
Sorted by Best Match

View as: Grid List

Qty Description Size Unit Price Price

Grapes Red Seedless Organic APX 2 LB (\$4.99 / LB) \$9.98

Grapes Green Seedless Organic APX 2 LB (\$4.99 / LB) \$9.98

left: Google Search, and right: Peapod.com

Dynamic queries and faceted search for structured collections (example)

The screenshot shows the Kayak search interface for Chicago, IL, from Oct 22 to Nov 4. The main feature is a heatmap map of Chicago, color-coded by attraction type and popularity. A legend on the right identifies categories: Sightseeing (orange), Eating (yellow), Shopping (green), and Nightlife (blue). On the left, there are several filter panels:

- Stars:** 4 stars (checked), 5 stars, 3 stars, 2 stars.
- Review Score:** Excellent (checked), Good, Okay, Mediocre.
- Price:** Price range slider from \$44 to \$894.
- Check-in:** Options for Free Breakfast, Free Cancellation, and Free Internet.

At the bottom, there are links for "Map" and "Satellite" views, and a "Google" logo.

21

Faceted search example

The screenshot shows a faceted search interface for "REI tents". The left sidebar contains various filters:

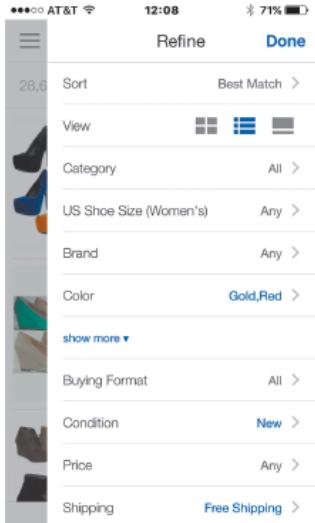
- Categories:** Backpacking Tents (checked).
- Sleeping Capacity:** Remove filters for 1-person (0), 2-person (7), 3-person (1), 4-person (1).
- Brand:** REI (checked).
- Seasons:** Remove filters for 3-4-season (1), 3-season (0).
- Best Use:** Remove filters for Backpacking (checked), Camping (0).
- Features:** Remove filter for Ultralight (0).
- Price:** Remove filters for \$100.00 to \$199.99 (0), \$200.00 to \$499.99 (1).
- Weight (lbs):** Remove filters for 1.00 to 2.99 (0), 3 to 4.99 (0), 5 to 7.99 (1).
- Number Of Doors:** Remove filter for 1 (0).

The right side displays the results for "REI tents" (9 matches), sorted by Relevance. Each result card includes a thumbnail image, a star rating, the product name, price, and a "Compare" button.

Product	Rating	Price
REI Passage 2 Tent	★★★★★ (14)	\$159.00
REI Half Dome 2 Plus Tent	★★★★★ (0)	\$219.00
REI Quarter Dome 2 Tent	★★★★★ (0)	\$299.00
REI Half Dome 2 Tent	★★★★★ (0)	\$169.00
REI Half Dome 4 Tent	★★★★★ (0)	\$199.00 - \$299.00
REI Arctiq ASL 2 Tent	★★★★★ (0)	\$369.00

22

Another Faceted Search Example



23

Command Language & “Natural Language” Example

```
SELECT DOCUMENTNO  
FROM JOURNAL-DB  
WHERE (DATE >= 2014 AND DATE <= 2018)  
AND (LANGUAGE = ENGLISH OR FRENCH)  
AND (PUBLISHER = ASIST OR HFES OR ACM)
```

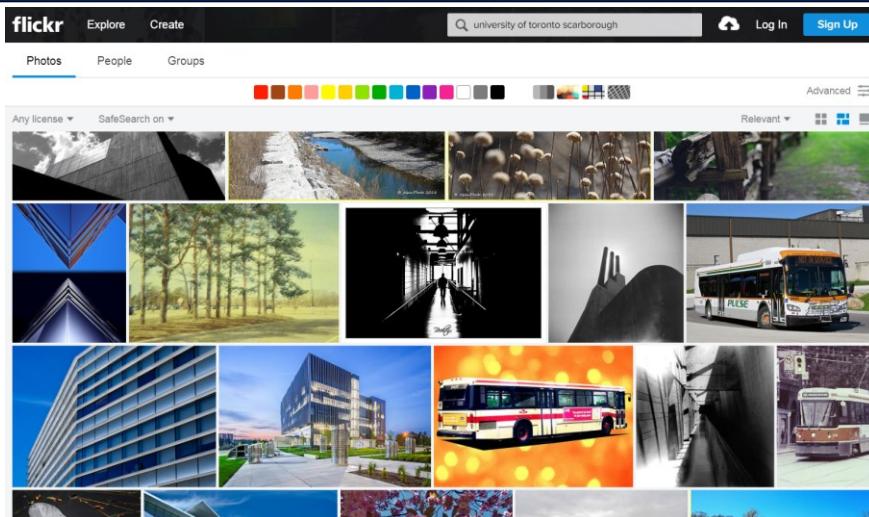
24

Multimedia Document Search and other specialized searches

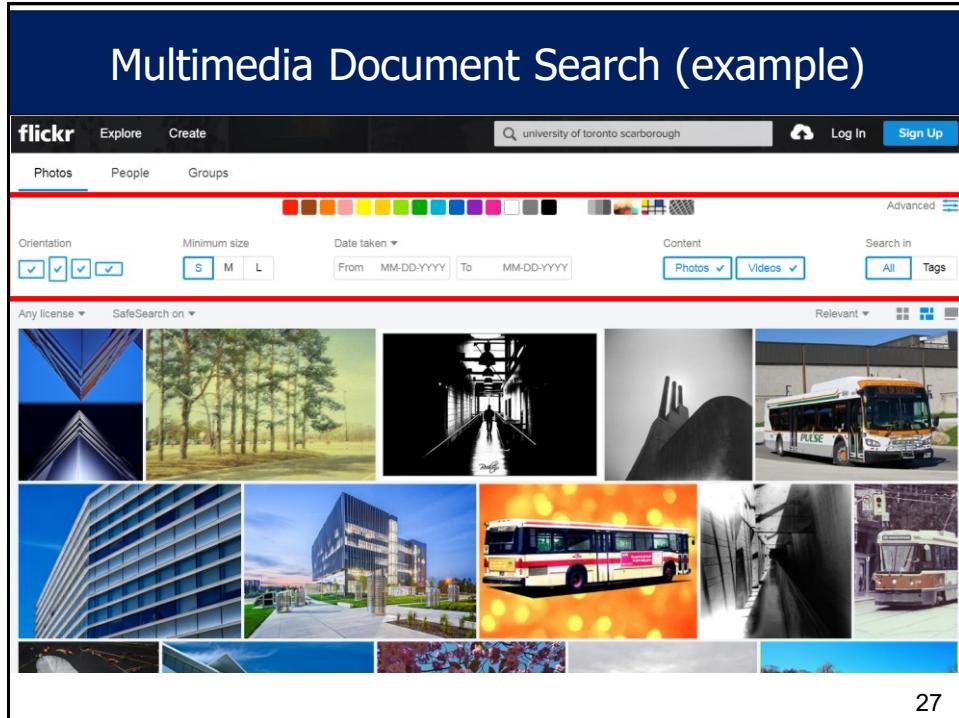
- Image search
- Video search
- Audio search
- Geographic information search
- Multilingual search
- Other specialized searches??

25

Multimedia Document Search (example)



26



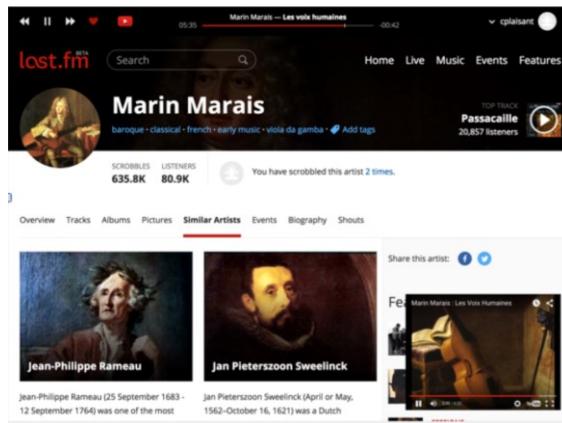
27

The social aspects of search

- *Social search* as “an umbrella term” describing search acts that make use of social interactions with others
 - May be explicit or implicit, co-located or remote, synchronous or asynchronous
 - Social bookmarking and ranking, e.g. Reddit
 - Personalized search built on user profiles, e.g. past site visits
 - Collaborative filtering and recommender systems, e.g. Netflix
 - Music recommendation, e.g. Pandora

28

The social aspects of search (example)



- Last.fm is an example of online radio using playlists created automatically
- The process starts by users selecting a start point (e.g. a song or artist they like) then users provide feedback on the suggestions by clicking on the heart or skipping the track

29

Data Visualization

Topics

1. Introduction
2. Tasks in Data Visualization
3. Visualization by Data Type
4. Challenges for Data Visualization

30

Data Visualization - Intro

- The best medium for many tasks and types of data is a visual representation—after all, a picture is supposedly worth a thousand words!
 - Successful designers should adapt the data presentation based on what the user needs to do
- This idea of data-driven pictures is called *visualization*
- Visual information seeking mantra:
 - Overview first, zoom and filter, then details on demand

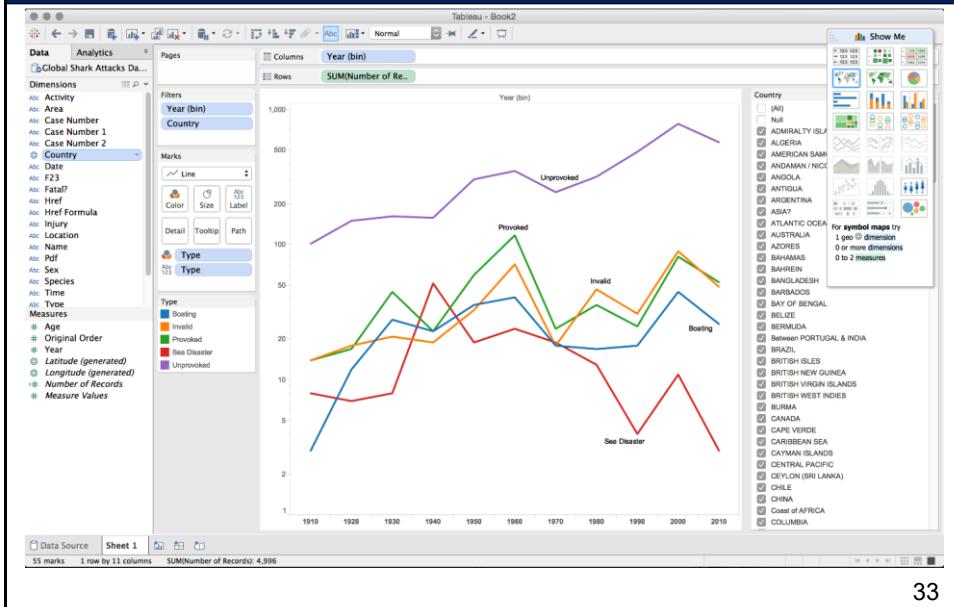
31

Tasks in Data Visualization

Task Categories	Task Types
<i>Data and view specification</i>	Visualize data by choosing visual encodings Filter out data to focus on relevant items Sort items to expose patterns Derive values of models from source data
<i>View manipulation</i>	Select items to highlight, filter, or manipulate Navigate to examine high-level patterns and low-level detail Coordinate views for linked exploration Organize multiple windows and workspaces
<i>Process and provenance</i>	Record analysis histories for revisit, review, and sharing Annotate patterns to document findings Share views and annotations to enable collaboration Guide users through analysis tasks or stories

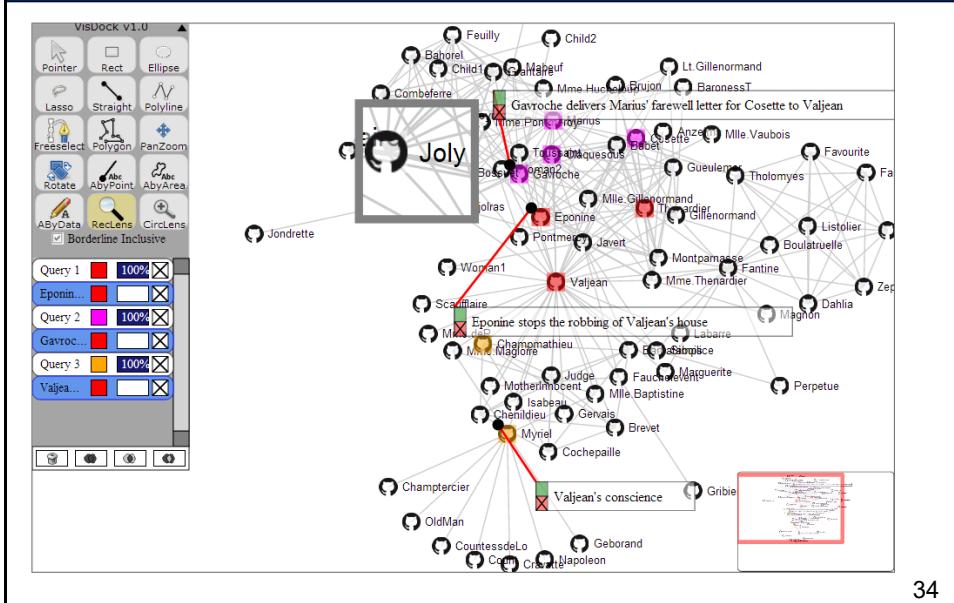
32

Visualize data by choosing visual encodings



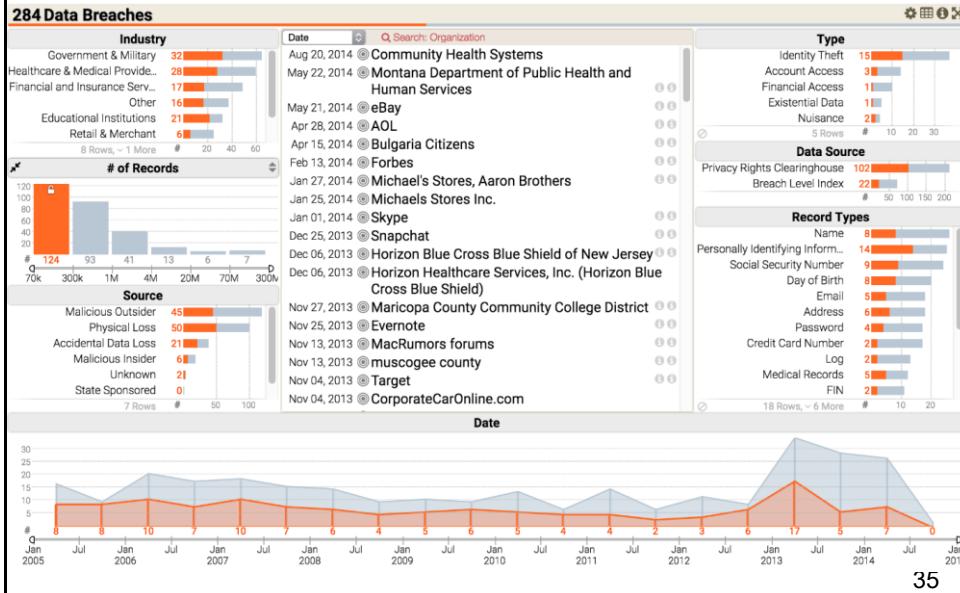
33

Select items to highlight, filter, or manipulate

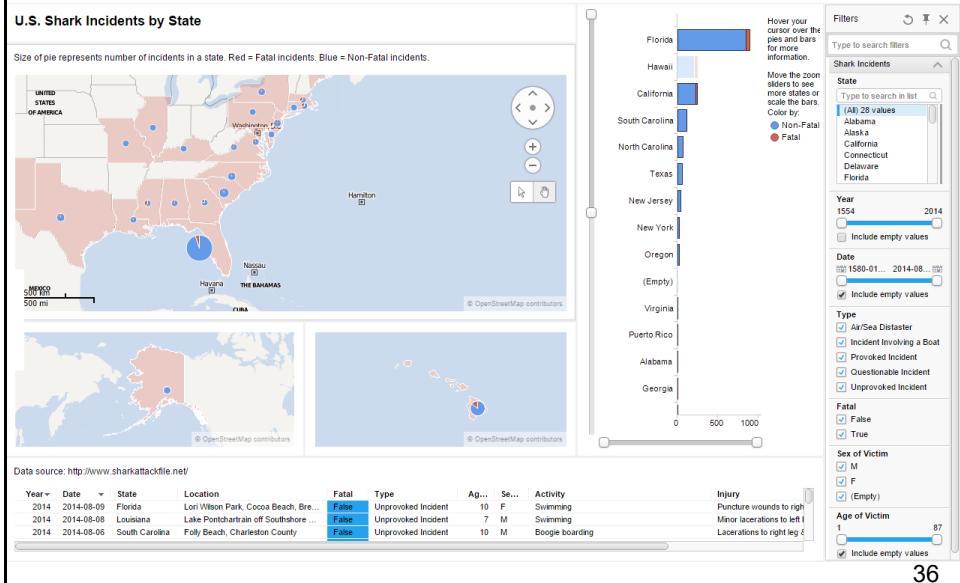


34

Coordinate views for linked exploration



Share views and annotations to enable collaboration



Guide users through analysis tasks or stories

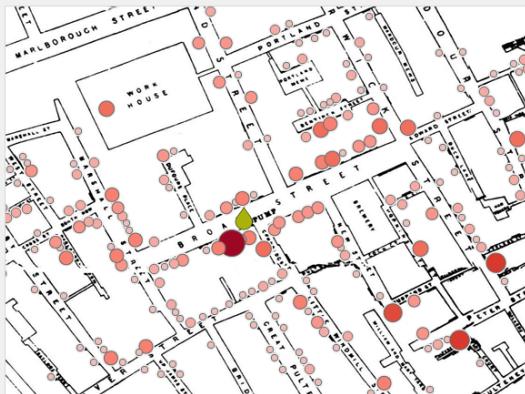
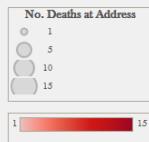
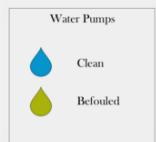
London's 1854 Cholera Outbreak:
Data Mapping Halts an Epidemic

An Outbreak Begins < Collecting the Data Mapping the Results Snow's Analysis: Focus on Broad St. Ending an Epidemic >

He noted that the bulk of deaths were concentrated in an area that generally used the same municipal water pump: Broad Street.

Snow realized that the Broad Street water pump—and poor water quality—was likely the source of the outbreak.

Armed with this information, he went to city authorities.



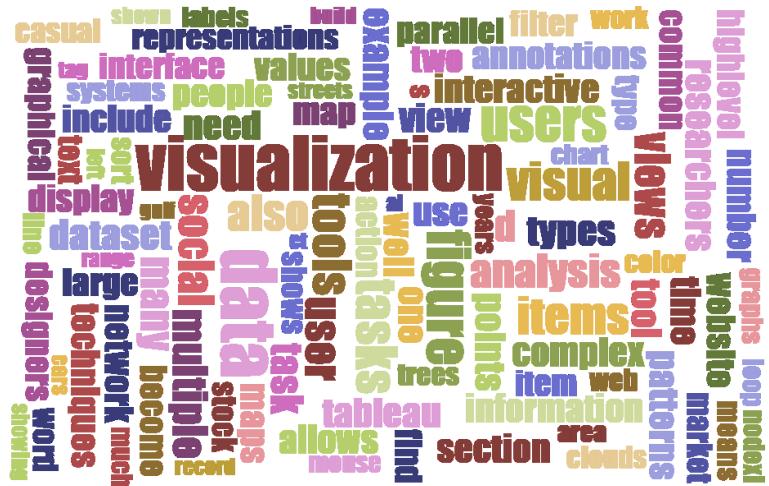
37

Visualization by Data Types

Data Type	Visualization Techniques and Systems
1-D linear	Tag clouds, Wordle, PhraseNets, parallel tag clouds
2D space	Geographic information systems (GIS), self-organizing maps
3D volume	Volume rendering, medical visualization, molecule visualization
Multi-dimensional	Tableau, parallel coordinates, scatterplot matrices
Temporal	Google Finance, EventFlow, LifeLines, TimeSearcher
Tree	Treemaps, degree of interest trees, space trees
Network	Node-link diagrams, adjacency matrices, NodeXL, Cytoscape

38

Linear data types are one-dimensional



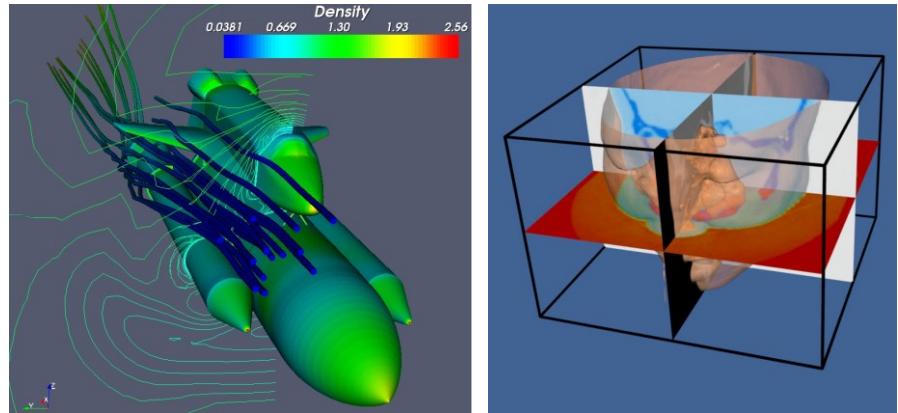
39

Another 2D space data example



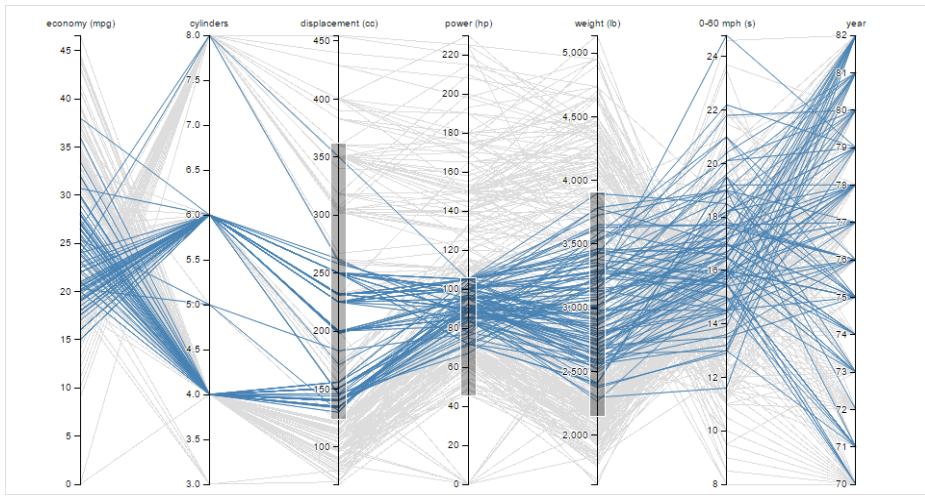
40

3D volume data

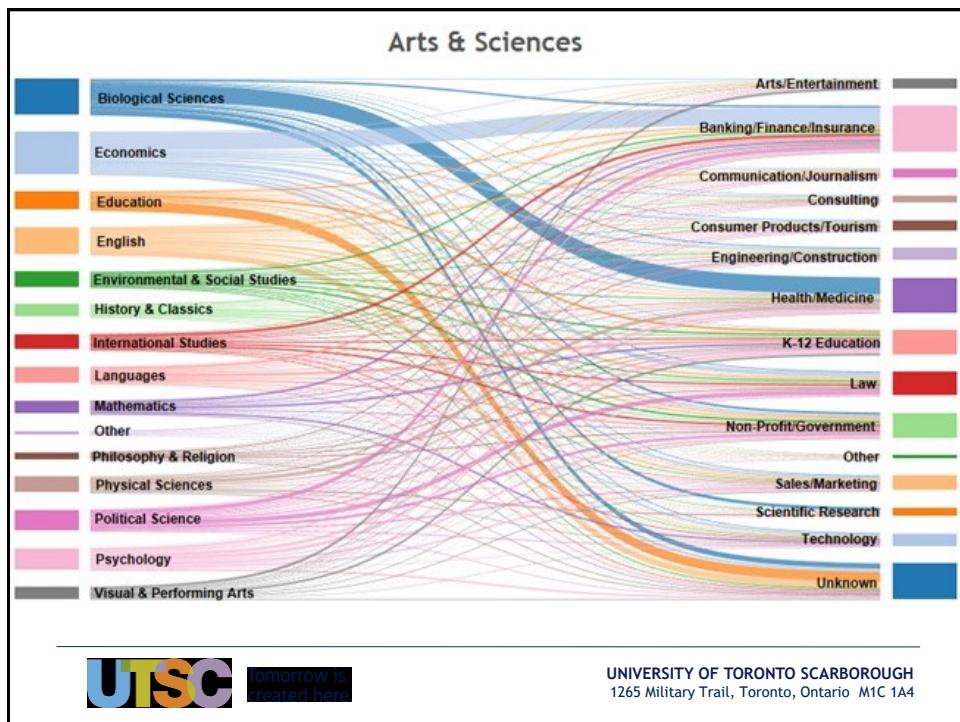


41

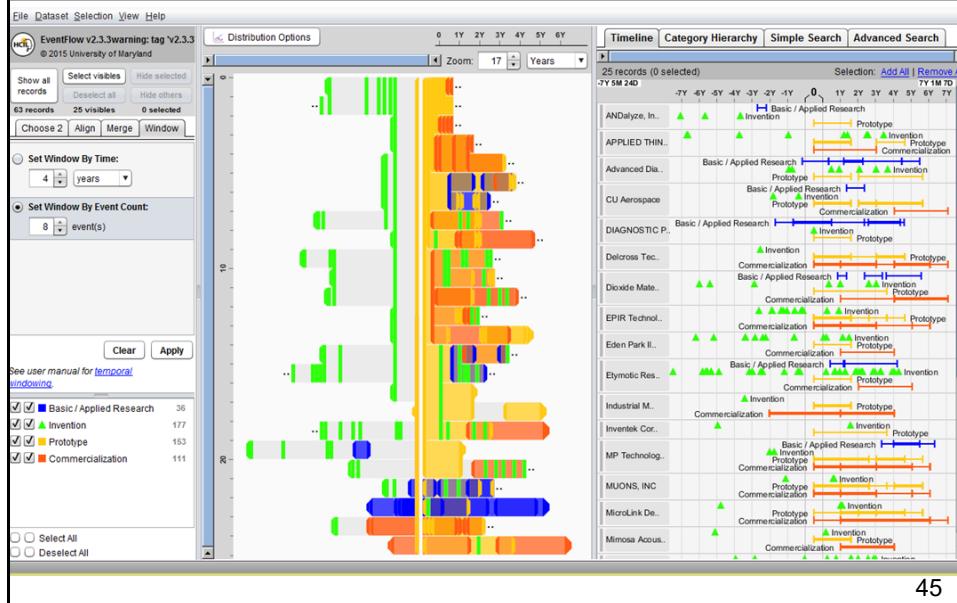
Multi-dimensional data



42

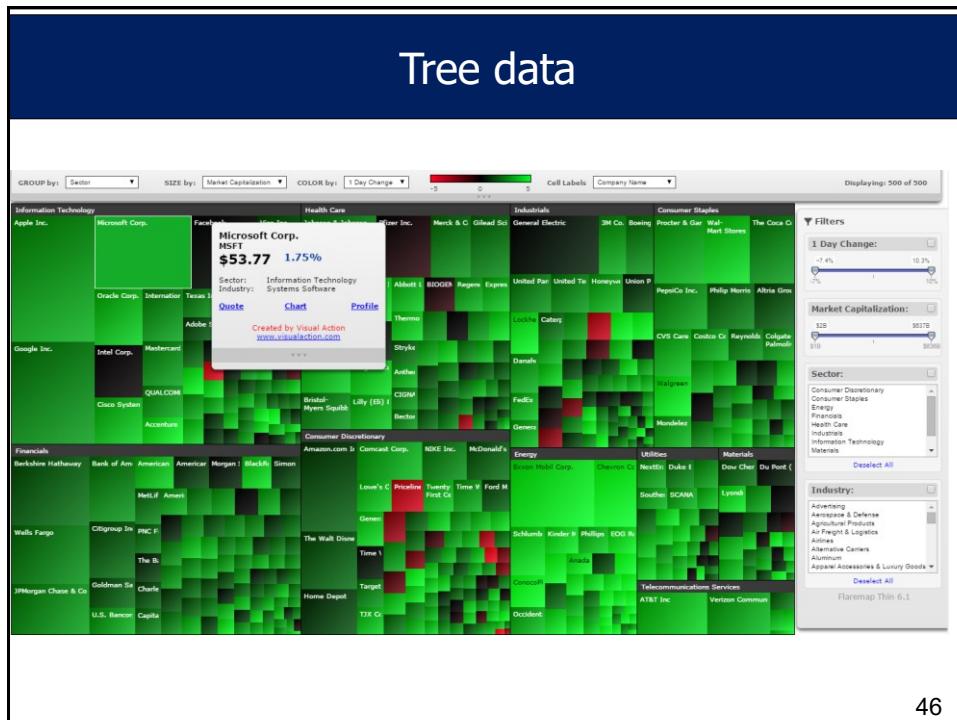


Another Temporal data example



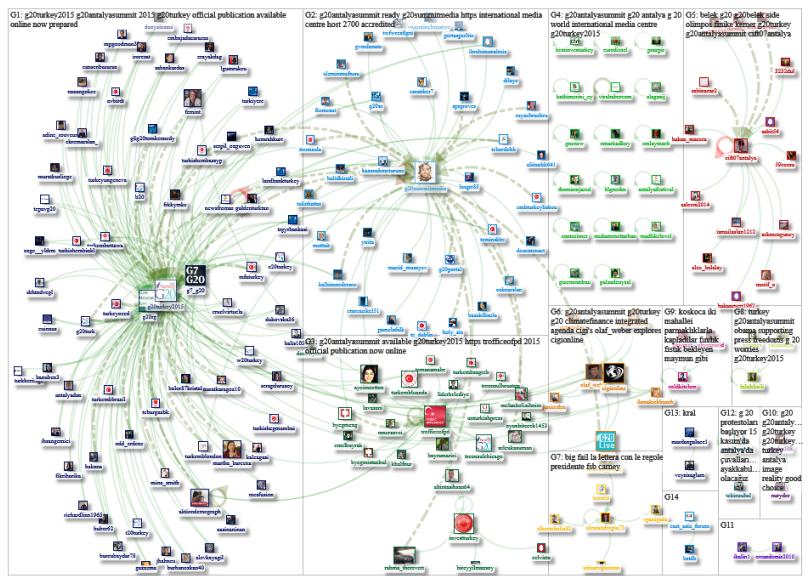
45

Tree data



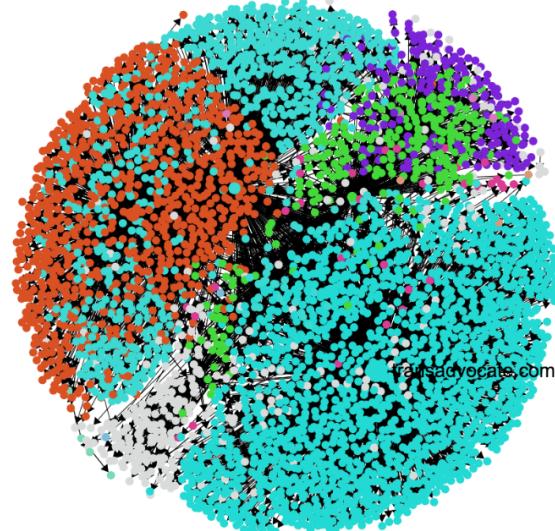
46

Network data



47

Network Visualizations



University of Victoria's Trans* Web Archive Link Structure as a Network Graph

48

Netlytics – Network Visualizations

Netlytic Demo

<https://netlytic.org/>

Making sense of online conversations

49

Challenges for Data Visualization

- Importing and cleaning data
- Integrating data mining
- Viewing big data
- Achieving universal usability
- Supporting casual users
- Dissemination and storytelling
- Adapting to any device
- Evaluation

50