CHAPTER 16:
Data Visualization

Designing the User Interface:

Strategies for Effective Human-Computer Interaction

Sixth Edition

Ben Shneiderman, Catherine Plaisant,
Maxine S. Cohen, Steven M. Jacobs, and Niklas Elmqvist

in collaboration with

Nicholas Diakopoulos

Addison Wesley
is an imprint of

PEARSON

Data Visualization

Topics

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

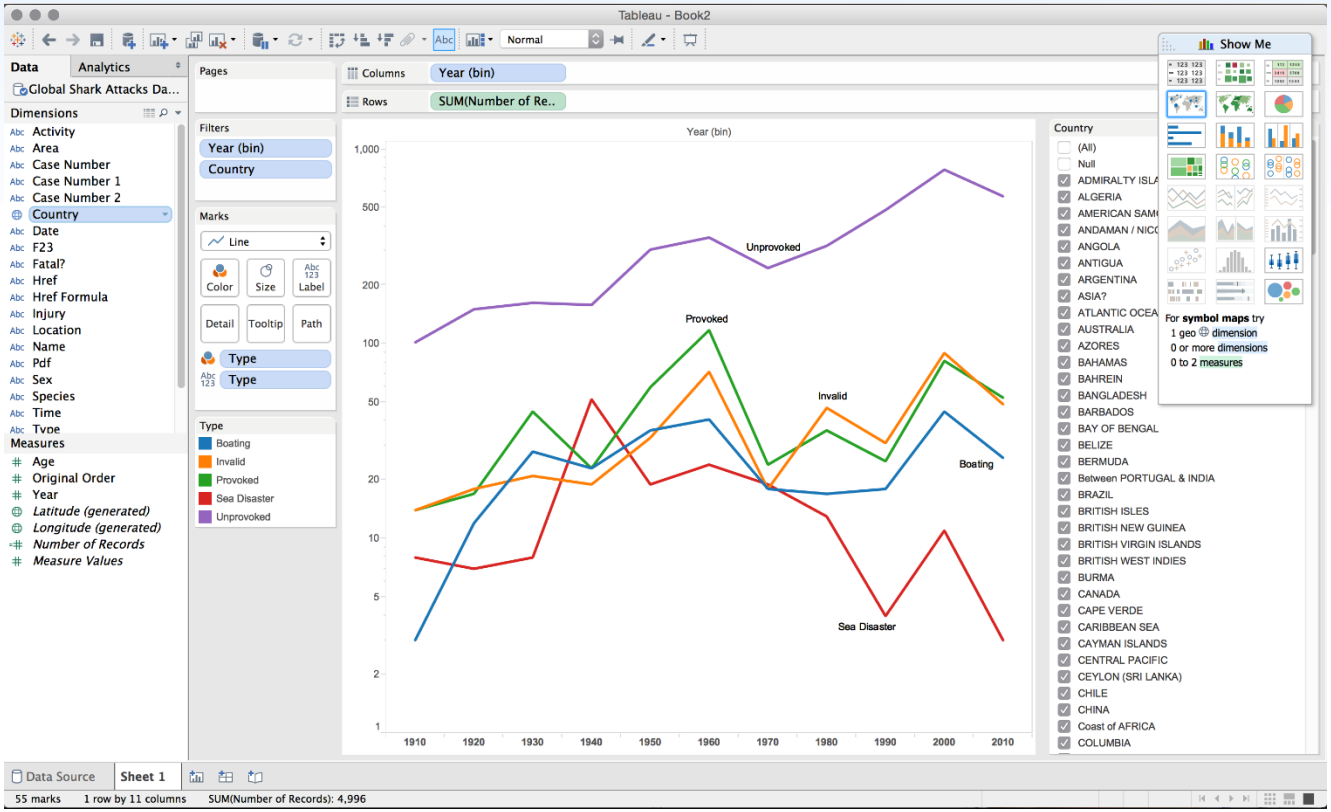
Introduction

- 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*
  - Visualization provides compact graphical presentations and user interfaces for interactively manipulating large numbers of items, often extracted from large datasets
  - Sometimes called *visual data mining*, it uses the enormous *visual bandwidth* and the remarkable *human perceptual system* to enable users to make discoveries, take decisions, or propose explanations about patterns, groups of items, or individual items.
- Visual information seeking mantra:
  - Overview first, zoom and filter, then details on demand

Tasks in Data Visualization

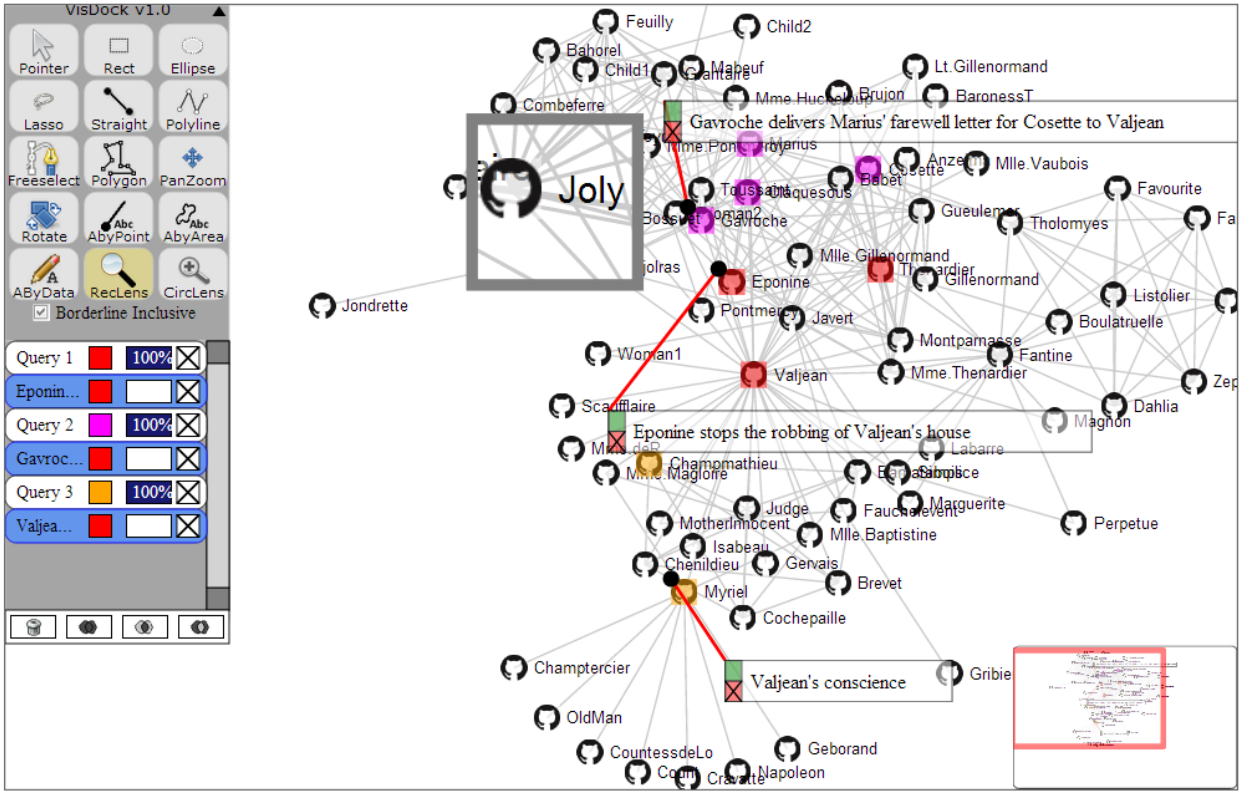
Task Categories	*Task Types*
Data and view specification	**Visualize** data by choosing visual encodings
	Filter out data to focus on relevant items
	Sort items to expose patterns
	Derive values of models from source data
View manipulation	**Select** items to highlight, filter, or manipulate
	Navigate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

Visualize data by choosing visual encodings



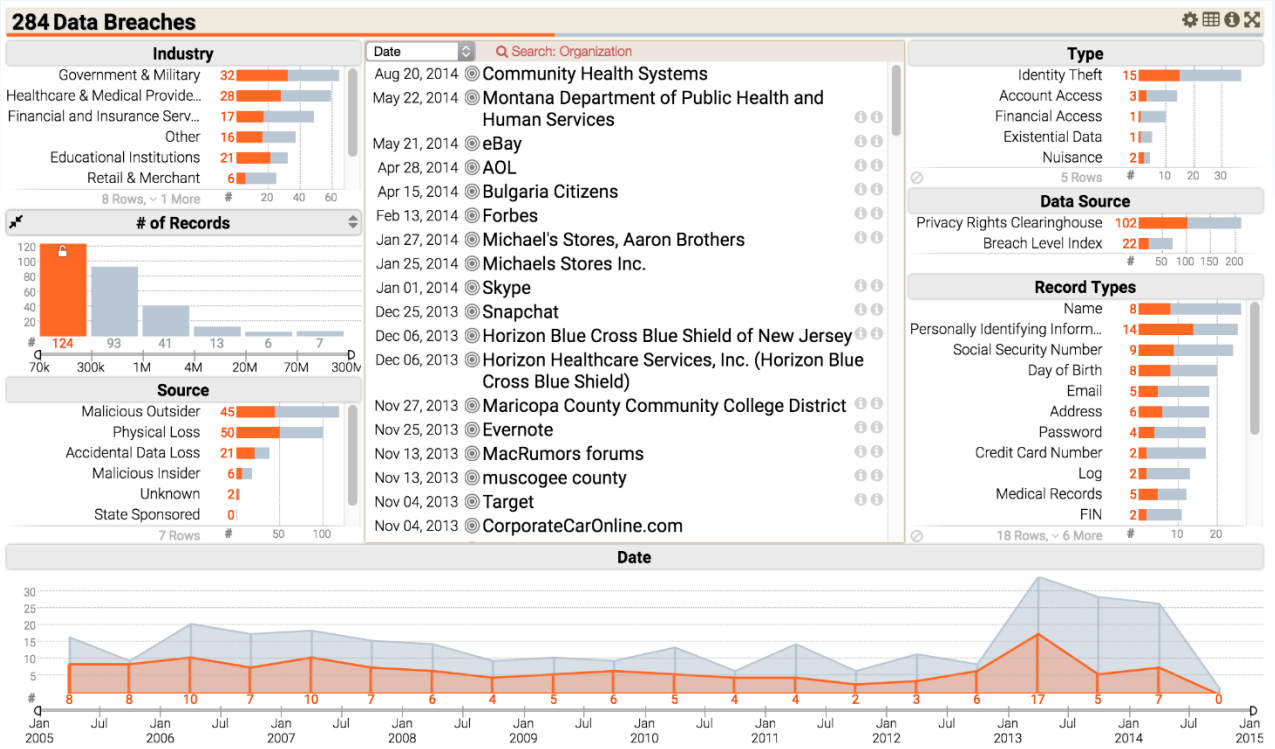
- Visualization palette (upper right) in the Tableau Desktop application for a dataset of shark attacks
- The “show me” feature in the tool (Mackinlay, 2007) will automatically highlight the suitable charts that can be used for the selected data

Select items to highlight, filter, or manipulate



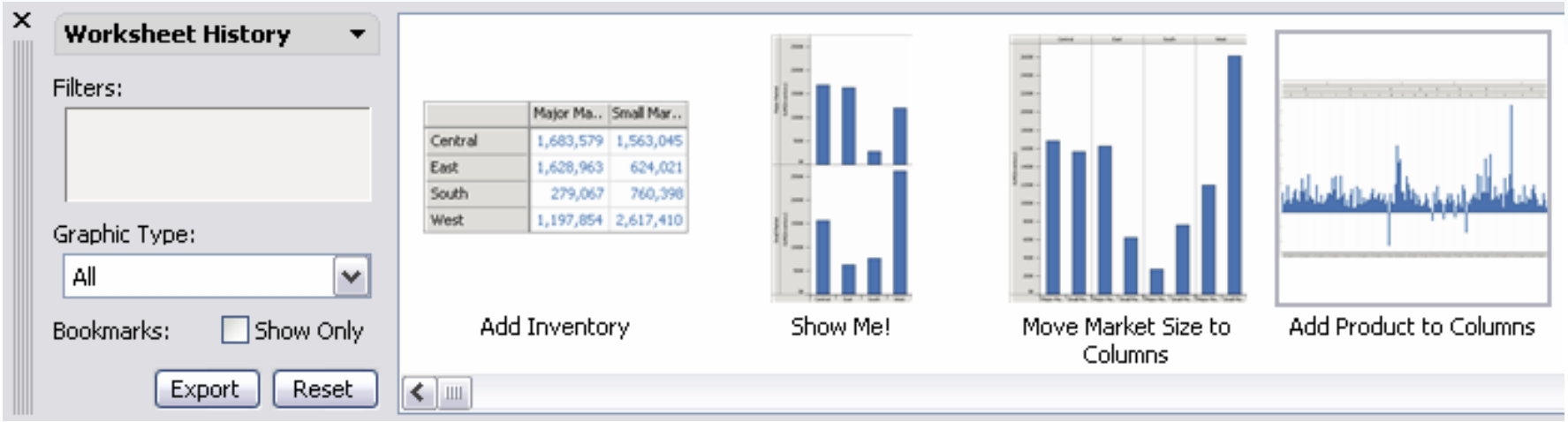
- Selection tools and data-aware annotations in an interactive node-link diagram representation of the social network for all of the characters in Victor Hugo's Les Misérables
  - Characters are linked together if they appear in the same chapter in the book
  - The textual annotations are connected to nodes using red lines and stay connected as the graph layout changes
  - The toolbar on the upper left is part of the VisDock toolkit, and provides tools for annotation, navigation, and selection (Choi, 2015)

Coordinate views for linked exploration



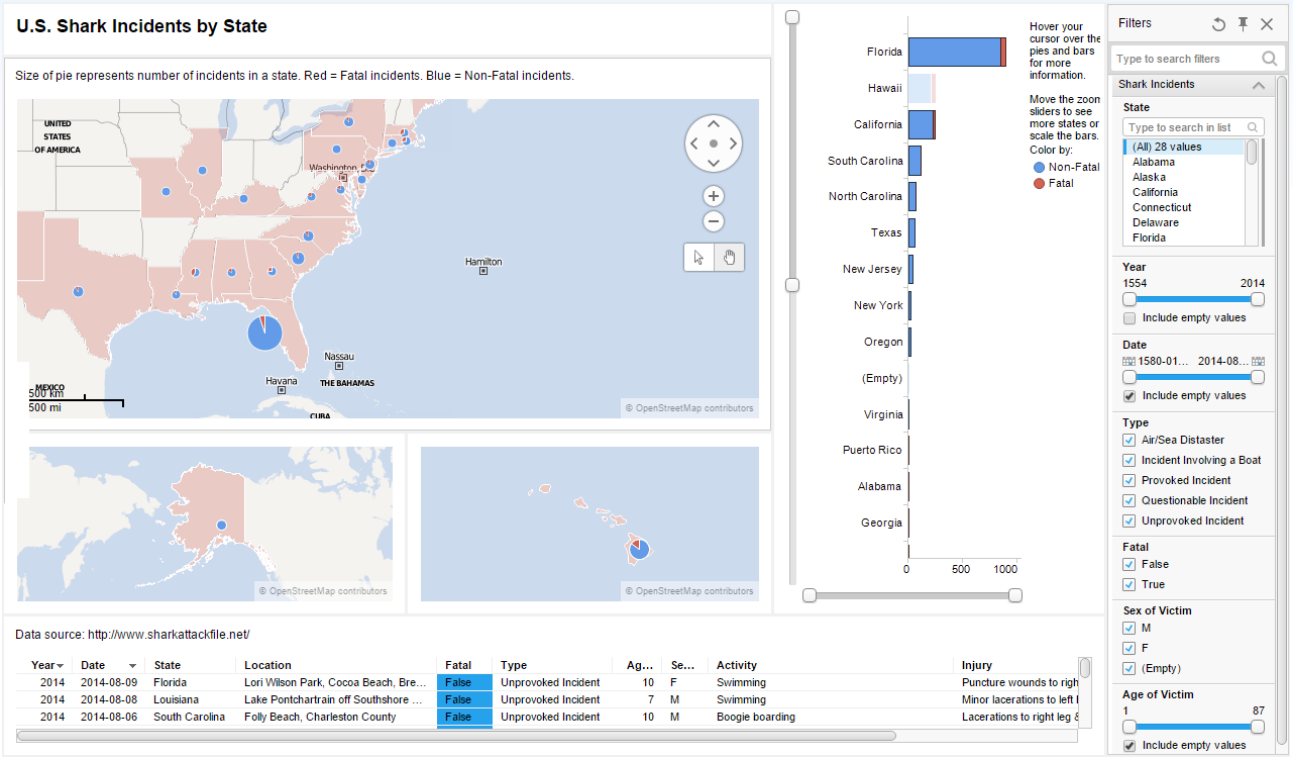
- Exploring 284 data breaches in the United States using the Keshif tool (http://keshif.me/), a multi-view visualization tool that shows different aspects of the data in separate views (Yalcin, 2016)
- Selecting items in one view highlights them in others; for example, the user is currently hovering over the bar for “70k-300k” in the view titled “# of Records”, which causes those breaches (i.e. 124 breaches that exposed in the range of 70,000 to 300,000 records) to also be highlighted in orange in other views, including in the timeline at the bottom

Record analysis histories for re-visitation, review, and sharing



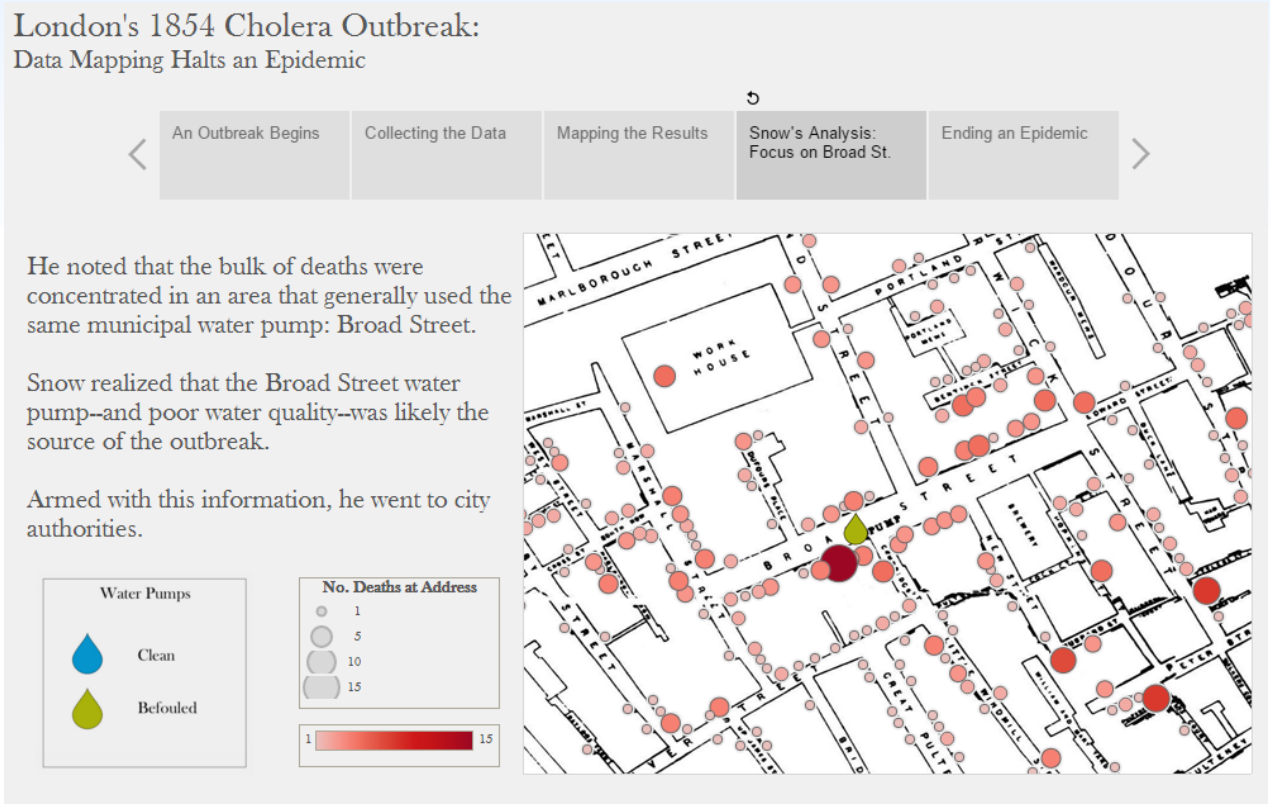
- Graphical history interface using thumbnails of previous visualization states organized in a comic-strip layout (Heer, 2008)
  - The labels describe the actions performed

Share views and annotations to enable collaboration



- Spotfire visualization dashboard of shark attacks published on the web
  - The dashboard can be interacted, causing views to update dynamically
  - The tool also allows for application bookmarking (storing the state for specific insights) as well as sharing the analysis on social media platforms such as Facebook, Twitter, and LinkedIn

Guide users through analysis tasks or stories



- Web-based visualization of London's 1854 cholera outbreak showing physician Jon Snow's use of visualization to find its source
  - This visualization was created in Tableau using its Story Points feature that allows users to build a narrative from data
  - The horizontal list of five boxes at the top of the display are the main points in the story, and viewers can be automatically guided through the story by moving to each point from left to right

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

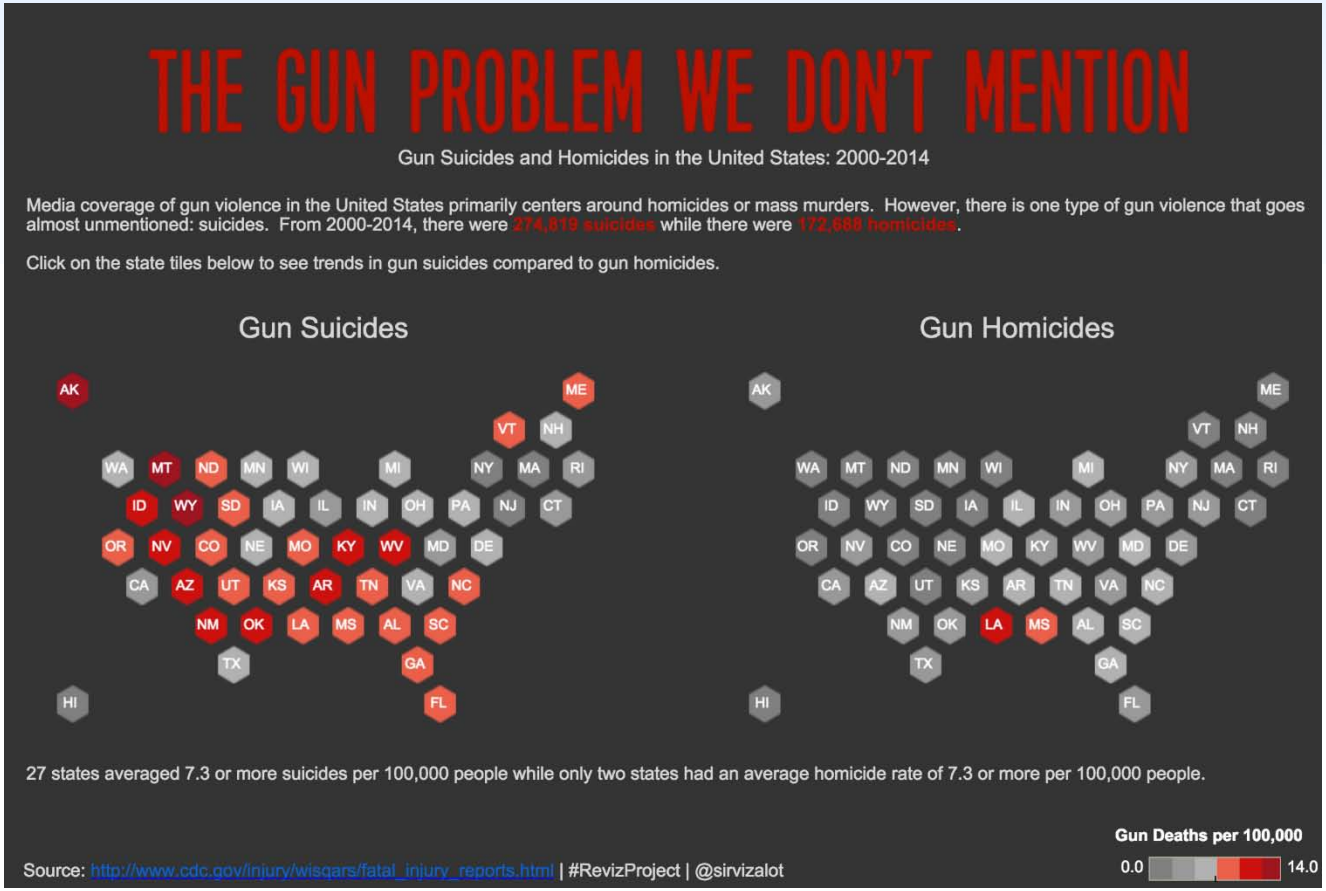
Linear data types are one-dimensional

- While tag clouds summarize popular tags used in collaborative tagging applications, word clouds display statistics about word usage in a text collection
  - Here, a word cloud generated by the online generator at https://www.jasondavies.com/wordcloud/ shows the most frequent words in Chapter 16 of this book

2D space data

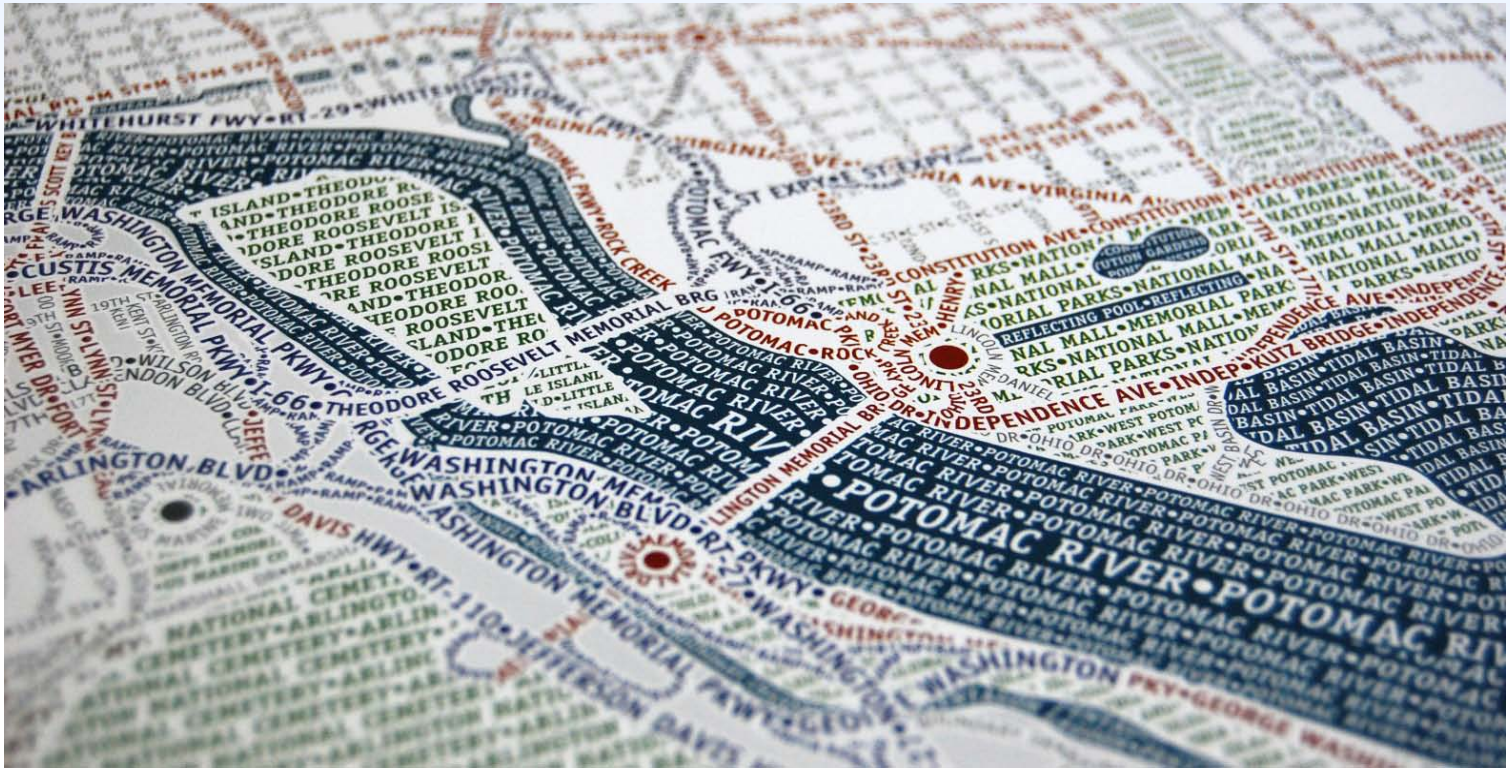
- New York Times Electoral Map for the 2012 U.S. presidential election showing the final results for each of the states (http://elections.nytimes.com/2012/electoral-map)
  - The main map uses square shapes for each state, sized according to the number of electoral votes
  - This gives a quick indication of the importance of each state on the overall race
  - The smaller map on the right shows the actual geographic view

Another 2D space data example



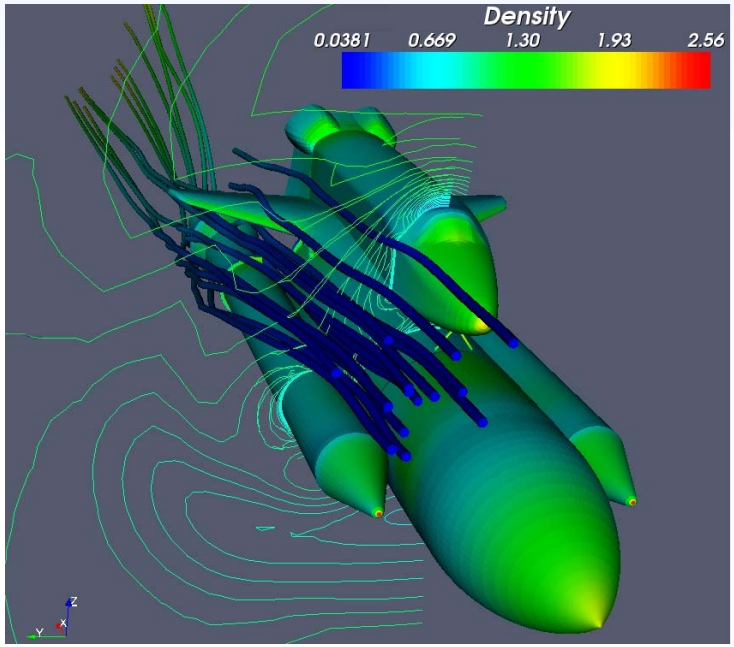
- Geographic visualization of gun deaths (suicides versus homicides) in the United States from 2000 to 2014 using data from the Center for Disease Control and Prevention (CDC).
  - Instead of using the actual geographic boundaries of the individual states, this map replaces states with uniform hexagons that have been color-coded using the color scale on the bottom right. The benefit of this representation is to prevent large states from dominating the visual appearance of the overall map. The hexagons have been placed so that they largely preserve the topology of the original map.

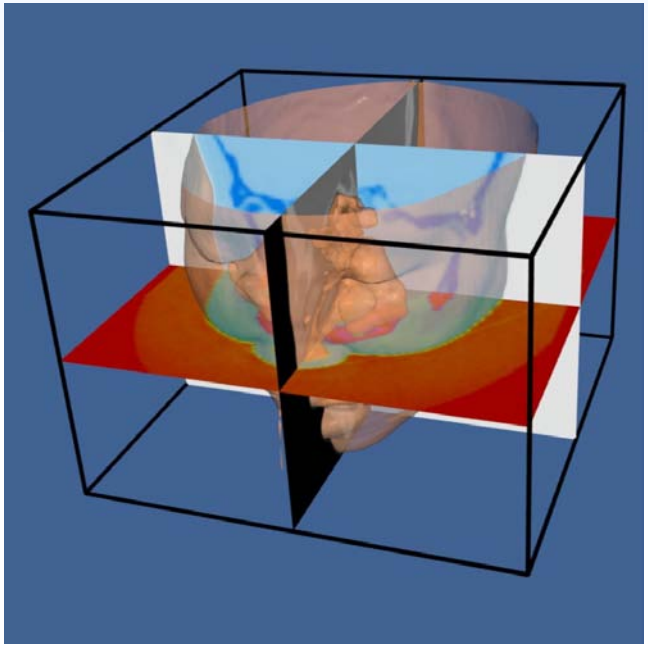
Another 2D space data example



- Typographic map of Washington D.C. created by Axis Maps
- A typographic map consists entirely of text organized into shapes using colored labels of streets, parks, highways, shorelines, and neighborhoods
- While this map took a skilled cartographer hundreds of painstaking hours to create, Afzal et al. (2012) later proposed an automatic approach taking mere minutes

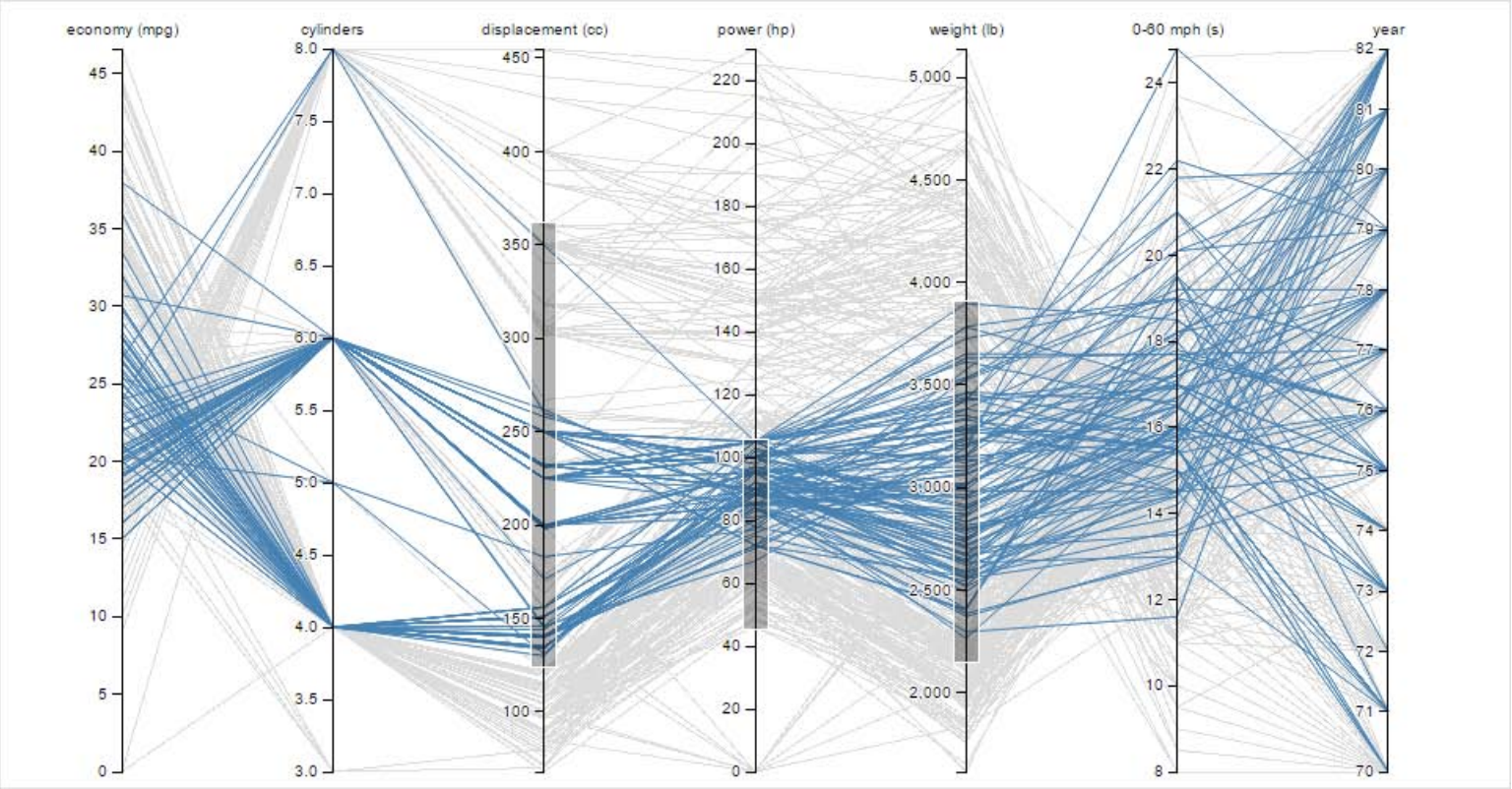
3D volume data





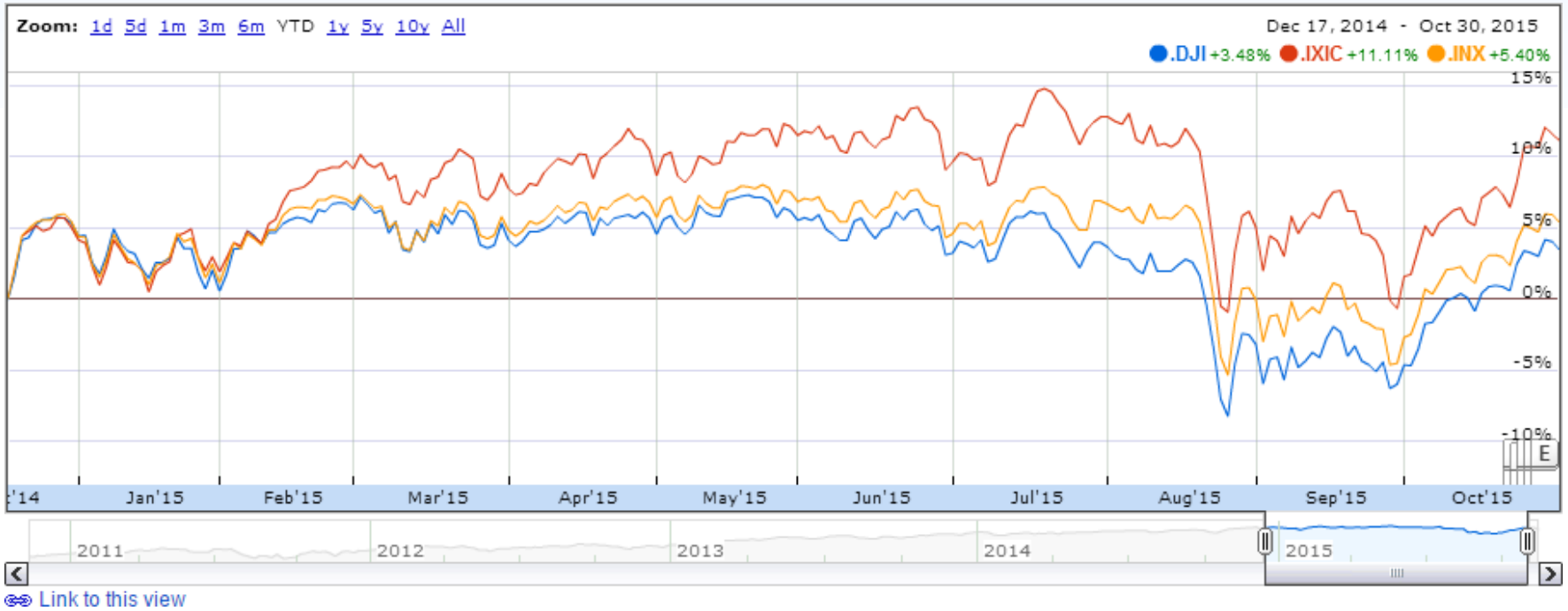
- Two 3D visualizations created using the Visualization Toolkit (VTK), a commercial software development library by Kitware, Inc (http://www.kitware.com/)
  - The left image shows flow density around the space shuttle using a rainbow color scale
  - The right image shows a CT scan of a human head with cross-sectional planes through the data

Multi-dimensional data



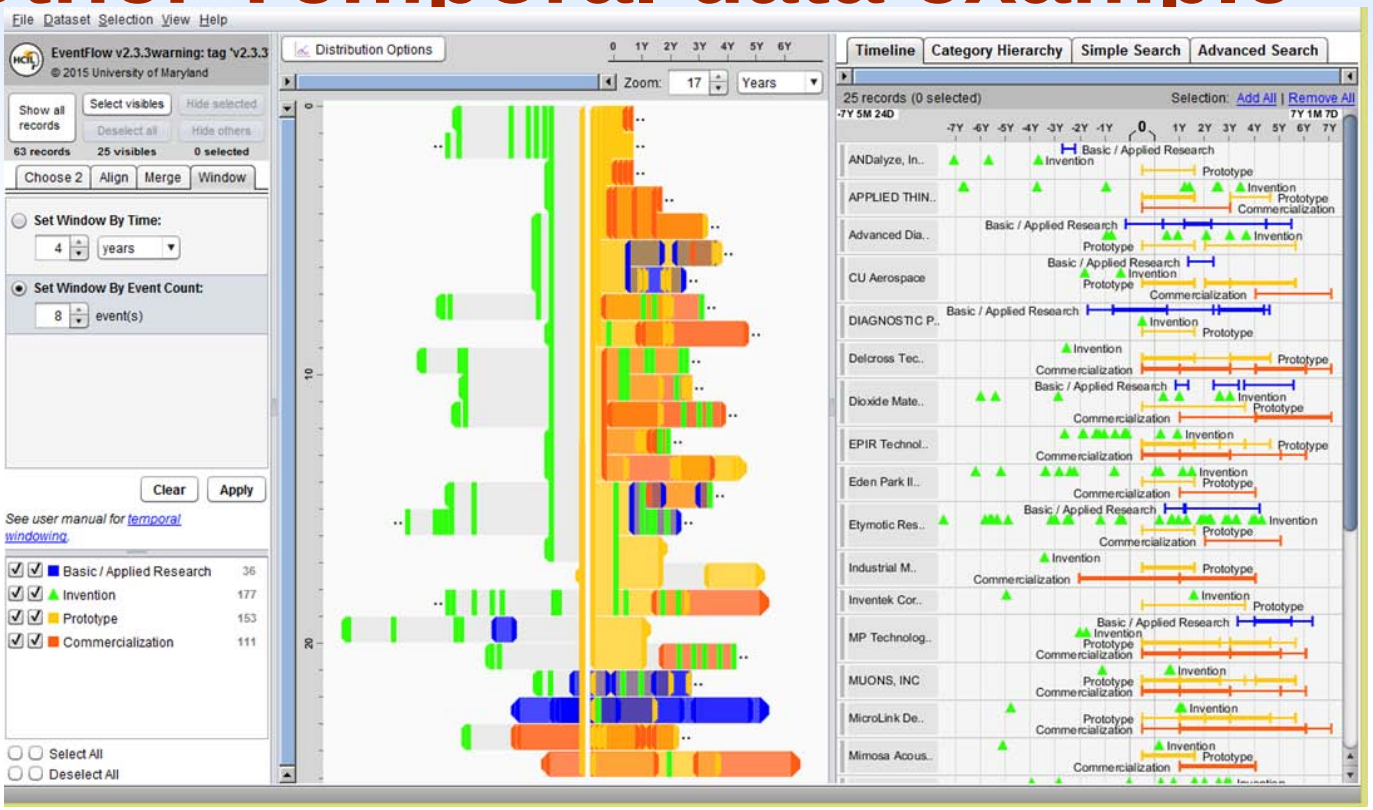
- Parallel coordinate visualization of cars from the 1970s and 1980s created using the D3 library
- This visualization supports axis filtering where selecting data ranges on the dimension axes filters out the cars that do not meet all of the criteria (gray lines)

Temporal data



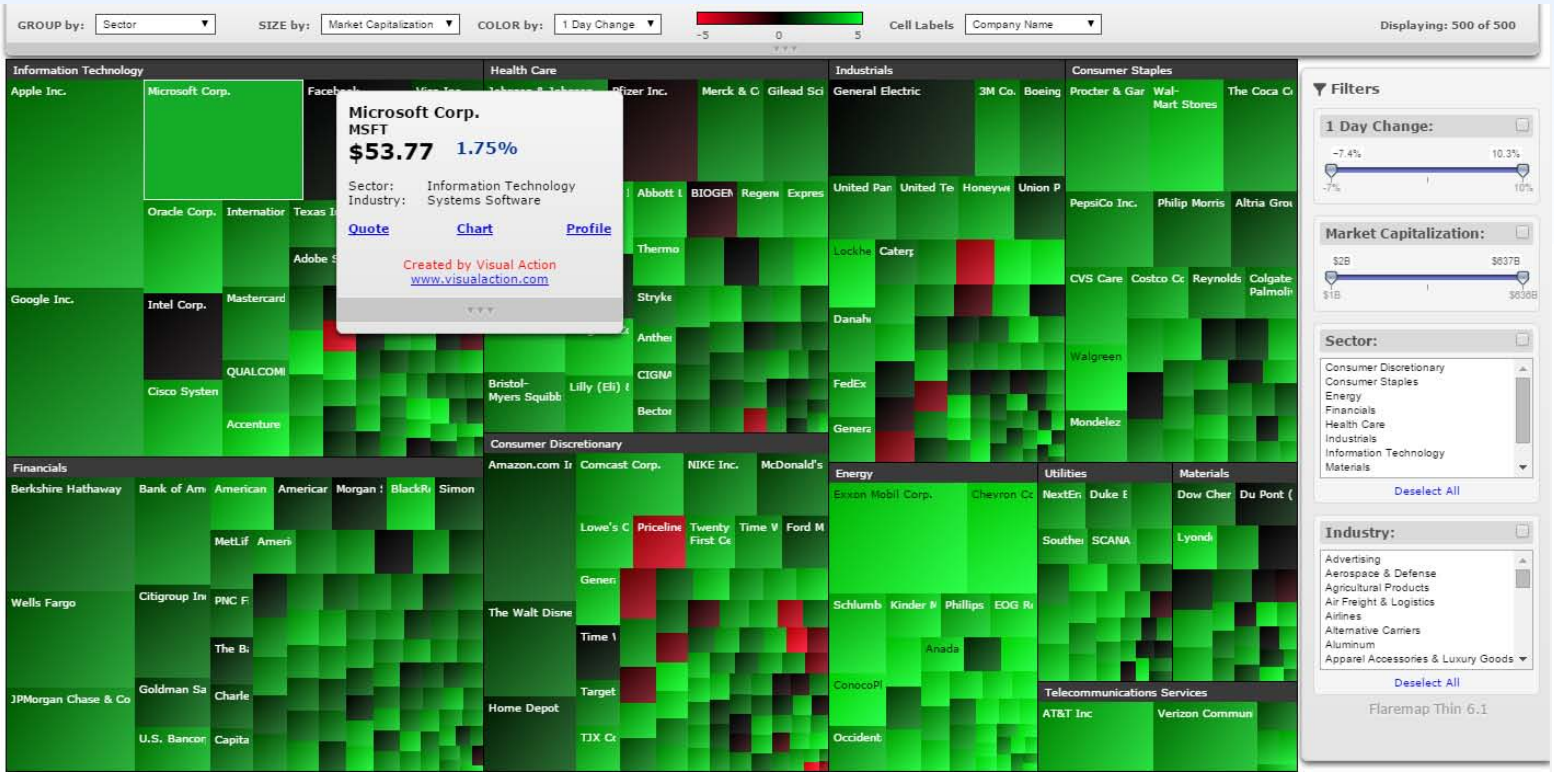
- Google Finance line graph showing the year-to-date performance of three stock market indices: the Dow Jones Industrial Average (.DJI, blue), the NASDAQ Composite (.IXIC, red), and the S&P 500 (.INX, yellow)
- The overview window at the bottom shows several years of data from 2011 to present day; grabbing the window allows for panning and resizing the detail view (top)

Another Temporal data example



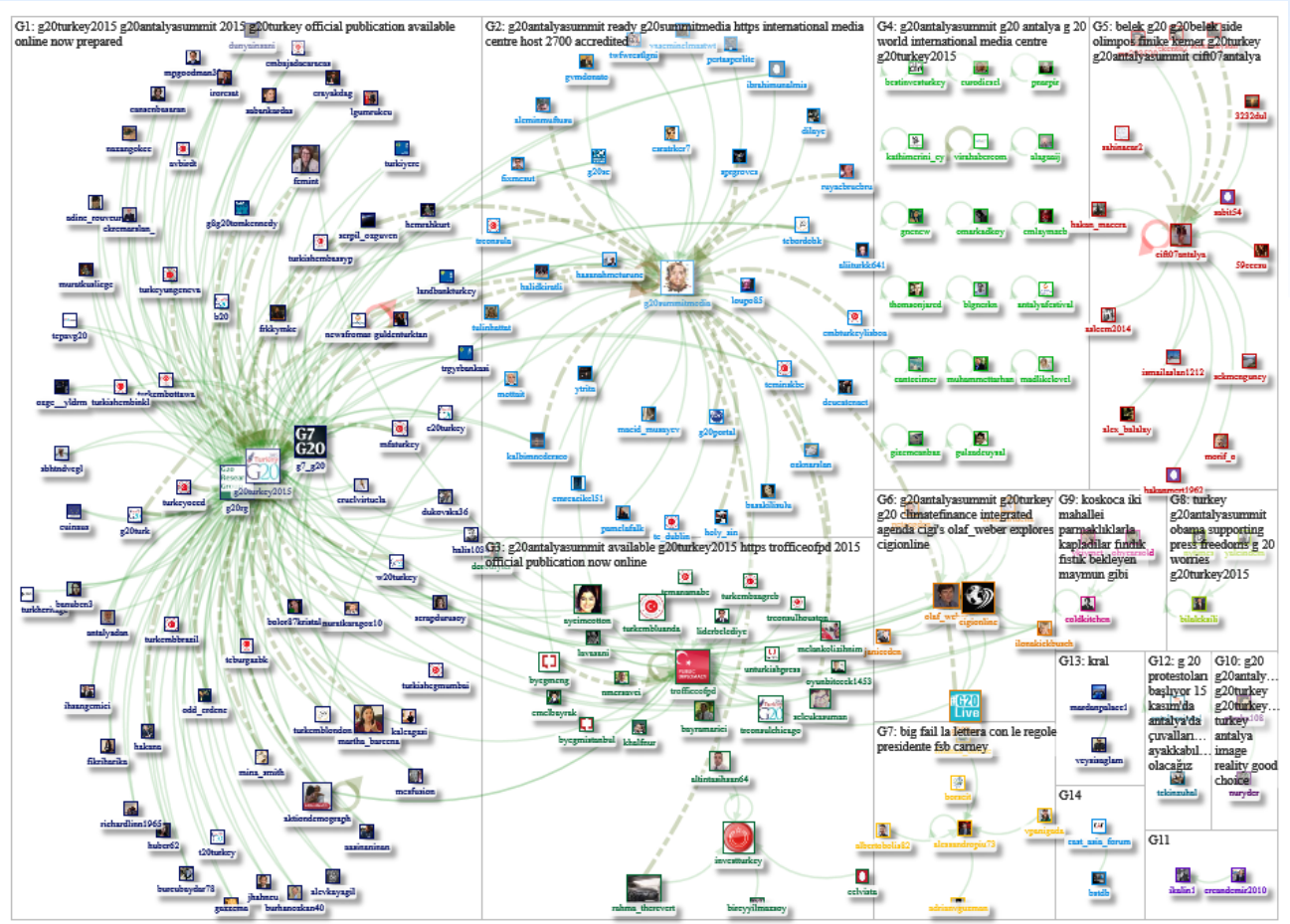
- The EventFlow (http://www.cs.umd.edu/hcil/eventflow) temporal event visualization system used to visualize sequences of innovation activities by Illinois companies
  - Activity types include research, invention, prototyping and commercialization
  - The timeline (right panel) shows the sequence of activities for each company
  - The overview panel (center) summarizes all the records aligned by the first prototyping activity of the company
  - In most of the sequences shown here the company's first prototype is preceded by two or more patents with a lag of about one year between the last patent application and the first prototype

Tree data



- The S&P 500 Market Monitor by Visual Action (http://www.visualaction.com/), a web-based treemap visualization showing stock performance of the 500 large companies making up the S&P 500 index on the NYSE and NASDAQ stock markets
- Each rectangle represents a company, sized according to its market capitalization, colored based on its 1-day change, and organized into their sectors

Network data



- Social network visualization built using NodeXL (Hansen, 2010) of 191 Twitter users tweeting with the hashtag "#G20AntalyaSummit" on November 9, 2015
- The hashtag refers to the 2015 G-20 summit held in Antalya, Turkey on November 15-16, 2015
- The users have been grouped and laid out in boxes based on the contents of the tweets
- NodeXL (https://nodexl.codeplex.com/) allows social scientists to collect, analyze, and visualize network graphs using a familiar interface by plugging into Microsoft Excel

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