8. Time-Based Data

Visualizing Change over time

Dr. Thorsten Büring, 13. Dezember 2007, Vorlesung Wintersemester 2007/08

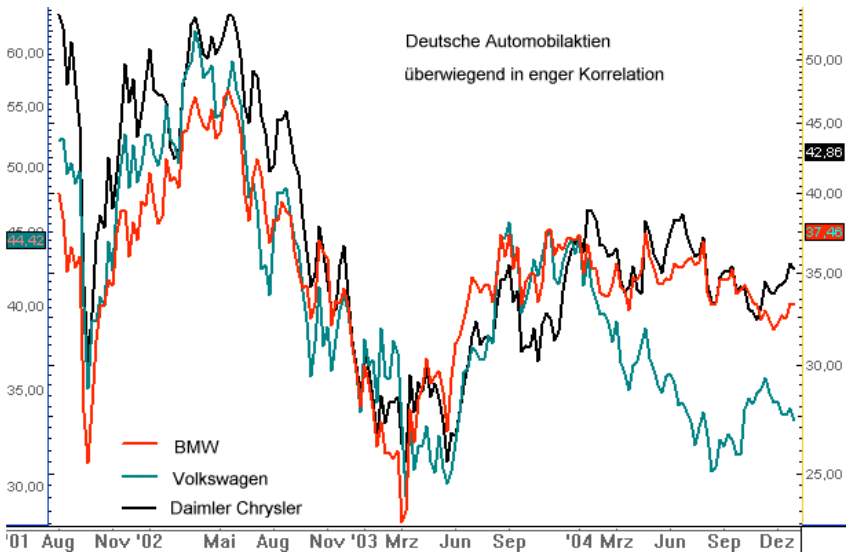
Ludwig
Maximilians-
Universität
München

Outline

- Term clarification, user tasks & taxonomies
- Historic time-based graphics
- Timeline representations in Information Visualization
- Use case: time-based computer desktops
- Searching time-based data

Time-Based Data

- Data set to visualize has a fundamental chronological component
- Usually one data variable (e.g. price) is measured at several points in time / value changes over time
- Tufte 2001: From 4,000 graphics drawn from 15 newspapers and magazines published from 1974 to 1980, more than 75% were time-series data
- Stasko: sometimes it is hard to distinguish between data entity and data cases
  - Example: stock prices for multiple stocks
  - Is each stock a data case, or is a price on a particular day a case, with the stock as one of the other variables?
- Standard representation: 2D line graph



Potential User Tasks

- MacEachern 1995
  - Does data element exist at time t?
  - When does a data element exist?
  - How long does a data element exist?
  - How often does a data element occur?
  - How fast are data elements changing?
  - In what order do data elements appear?
  - Do data elements exist together?
- Stasko
  - When was something greatest/least?
  - Is there a pattern?
  - Are two series similar?

Taxonomies

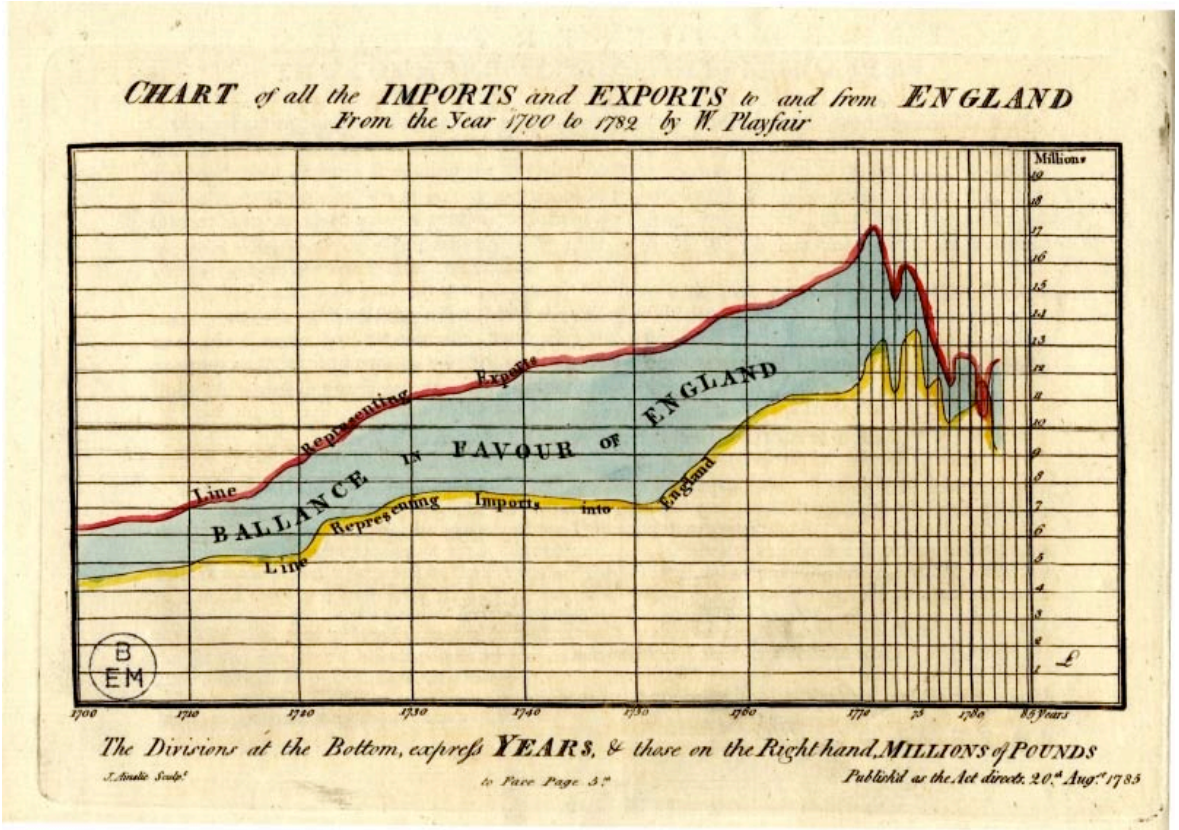
- Time axis - Frank 1998
  - Discrete points vs. interval points
  - Linear time vs. cyclic time
  - Ordinal time vs. continuous time
  - Ordered time (events happen one after the other) vs. branching time (e.g. sequence diagram of decision making) vs. time with multiple perspectives (e.g. parallel events in time)
- Time-dependency of the visualization – Müller & Schumann 2003
  - Static representation
    - Allow comparing of and focusing on data features without temporal limitations
    - Modifications of the visualization result from user interaction only
  - Dynamic representation
    - Changes dynamically and automatically over time (animation)
    - Get a feel for the process & change of the data over time

Outline

- Term clarification, user tasks & taxonomies
- Historic time-based graphics
- Timeline representations in Information Visualization
- Use case: time-based computer desktops
- Searching time-based data

Imports and Exports

- Playfair 1768



Train Schedule Paris - Lyon

- Marey 1885 (from Tufte 2001)

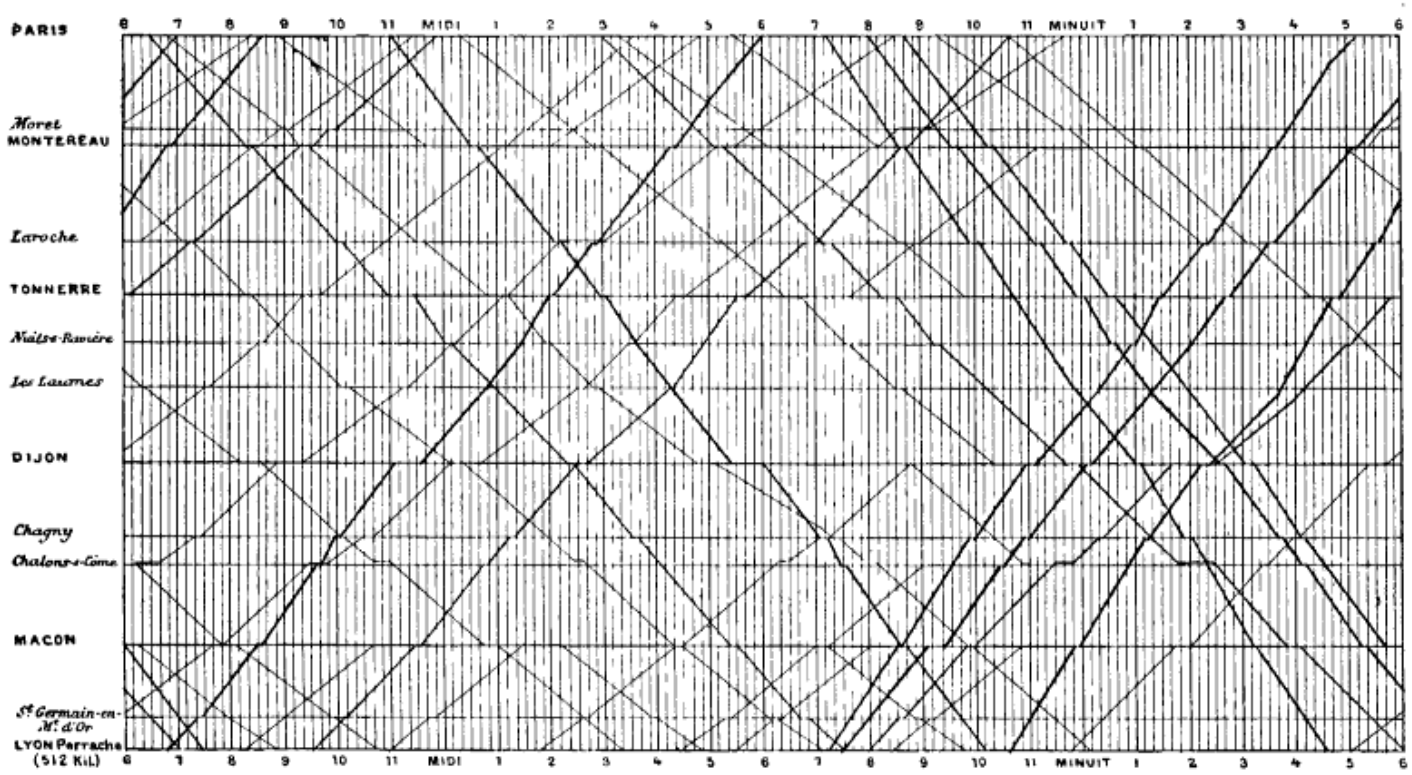


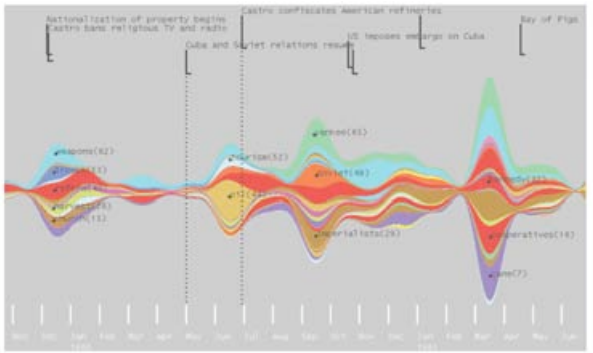
Fig. 5. — Graphique de la marche des trains sur un chemin de fer, d'après la méthode de Ibry.

Outline

- Term clarification, user tasks & taxonomies
- Historic time-based graphics
- Timeline representations in Information Visualization
- Use case: time-based computer desktops
- Searching time-based data

ThemeRiver

- Havre et al. 2002
- River metaphor for visualizing thematic variations over time within a document collection
  - Example: collection of Fidel Castro's speeches, interviews and articles (1959-1961)
  - Flow from left to right represents movement through time
  - Selected thematic content shown as color currents
  - Vertical width of current indicates increase or decrease in thematic strength at any point in time
  - Smoothing of curves
- Enable users to find patterns and trends, e.g. is there a causal relationship between external events and changes of thematic strength?
- Problem: scalability, i.e. number of themes
- **Movie**



ThemeRiver

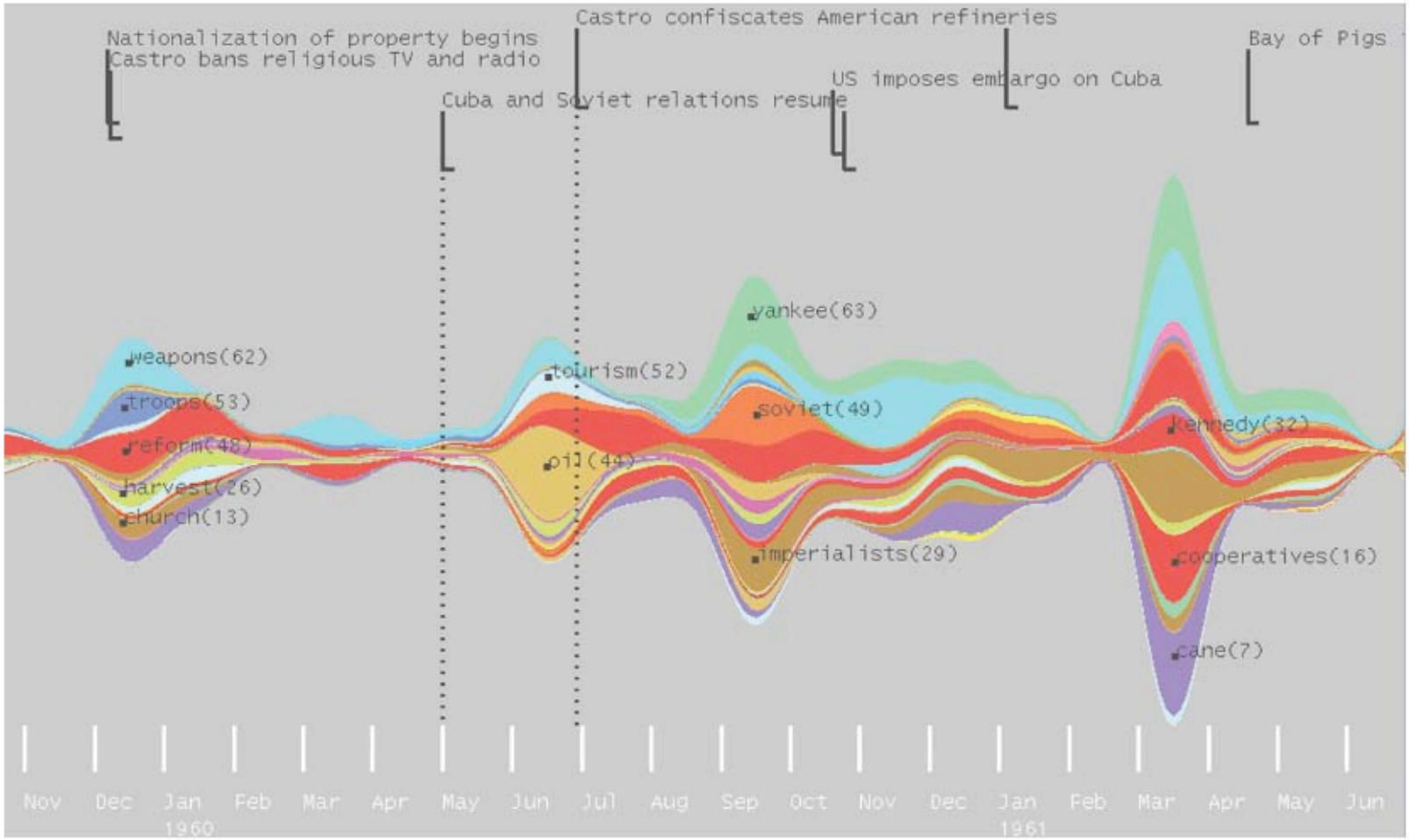
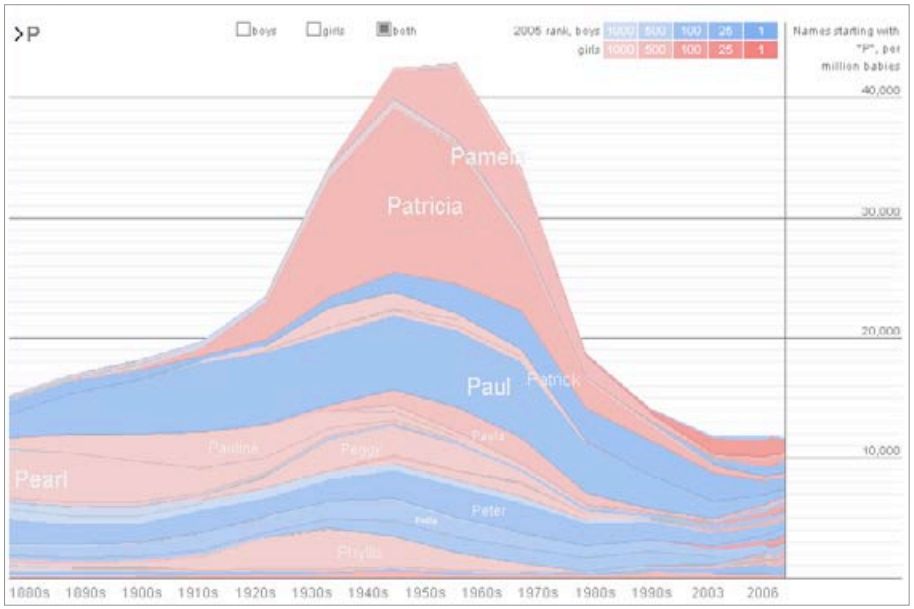


Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.

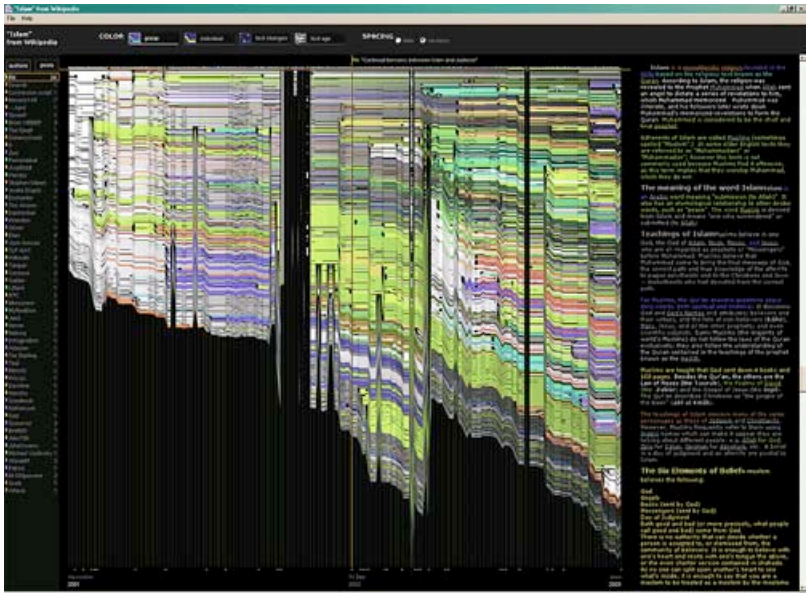
Baby Name Wizard

- The NameVoyager charts the 1000 most frequent girl and boy names in the USA over time
- Filled stripes that are stacked on top of each other
- Color mapping
  - Hue: sex
  - Saturation: current popularity of the name
- Direct manipulation by text-entry
- On-demand
  - Detail view of individual graph (click)
  - Value for points in time
- Animated transitions
- **Demo**



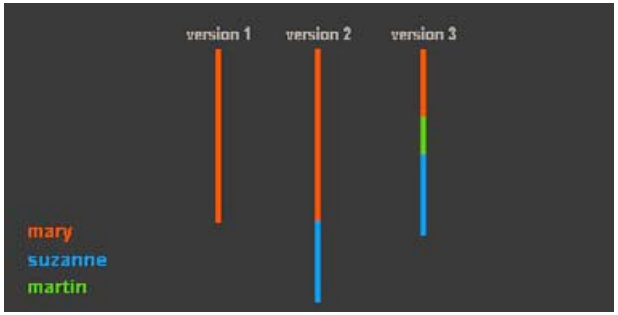
History Flow

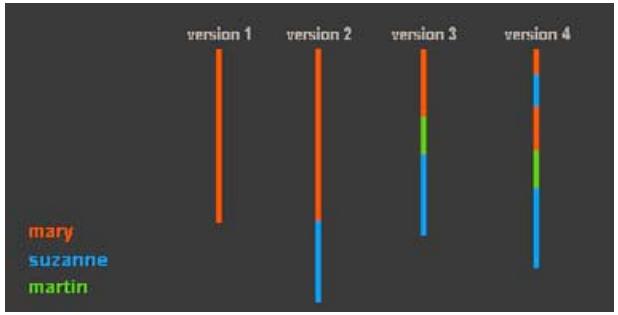
- Viégas et al. 2004
- http://www.research.ibm.com/visual/projects/history_flow/
- Visualizes
  - Dynamic, evolving documents
  - Interactions of multiple collaborating authors
- Example: evolutionary history of wiki pages on Wikipedia
- Highlight deletions and insertions of text by by different authors over time

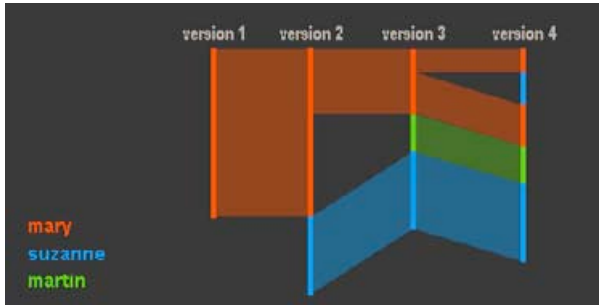


History Flow

- Visualization construction



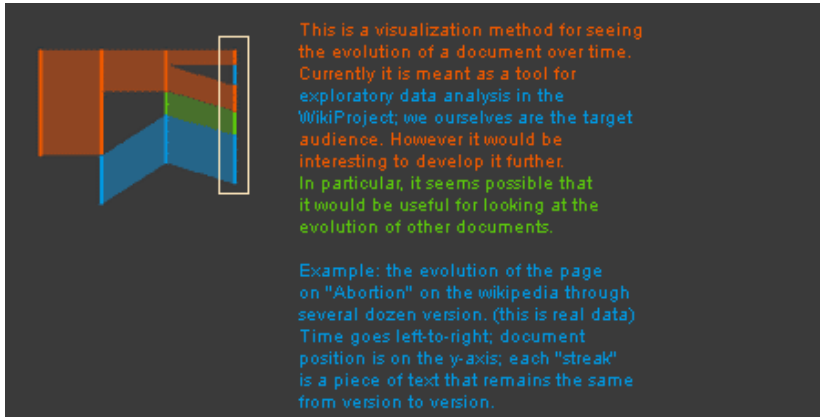


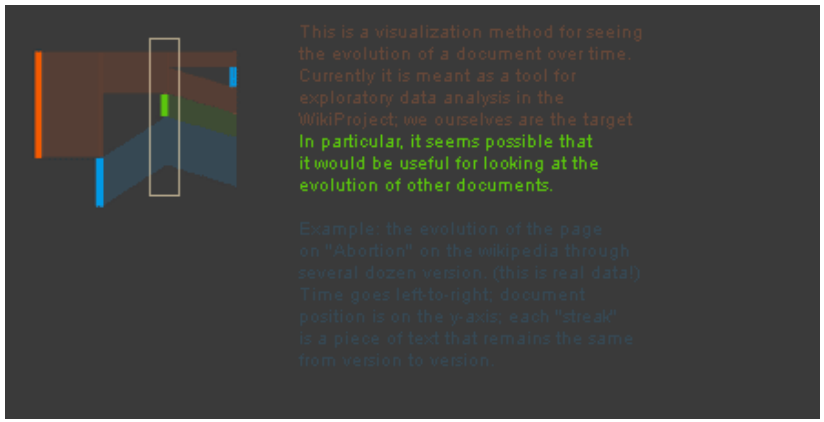


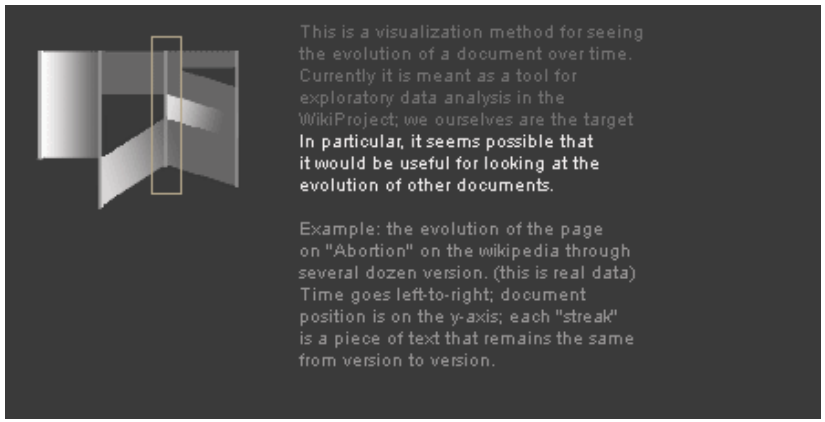
Ludwig Maximilians Universität München

History Flow

Visualization modes

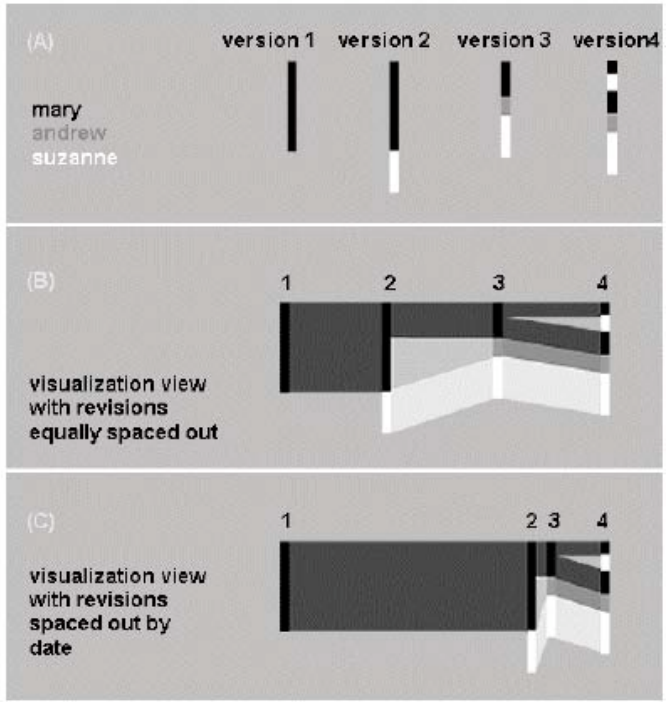


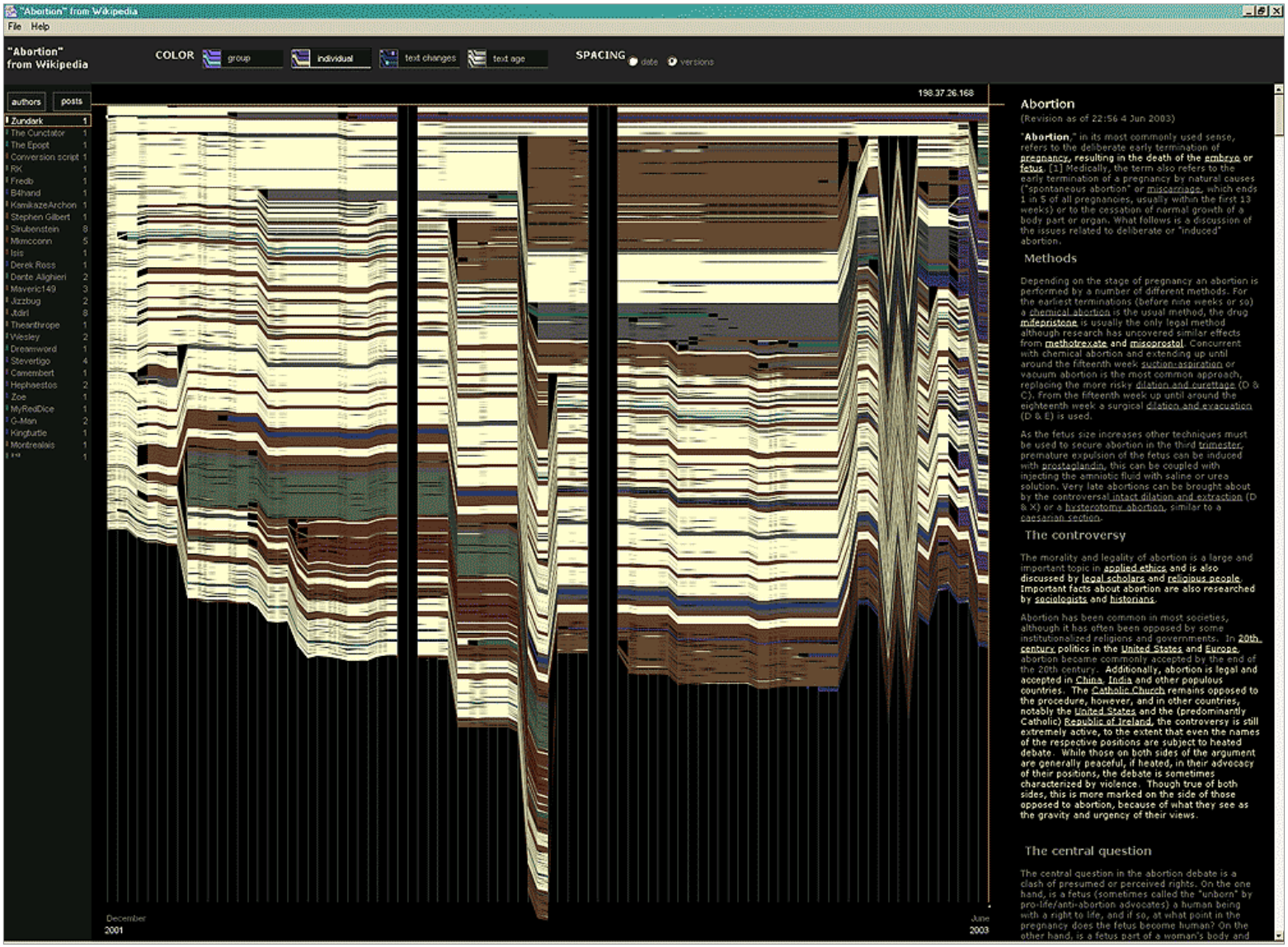


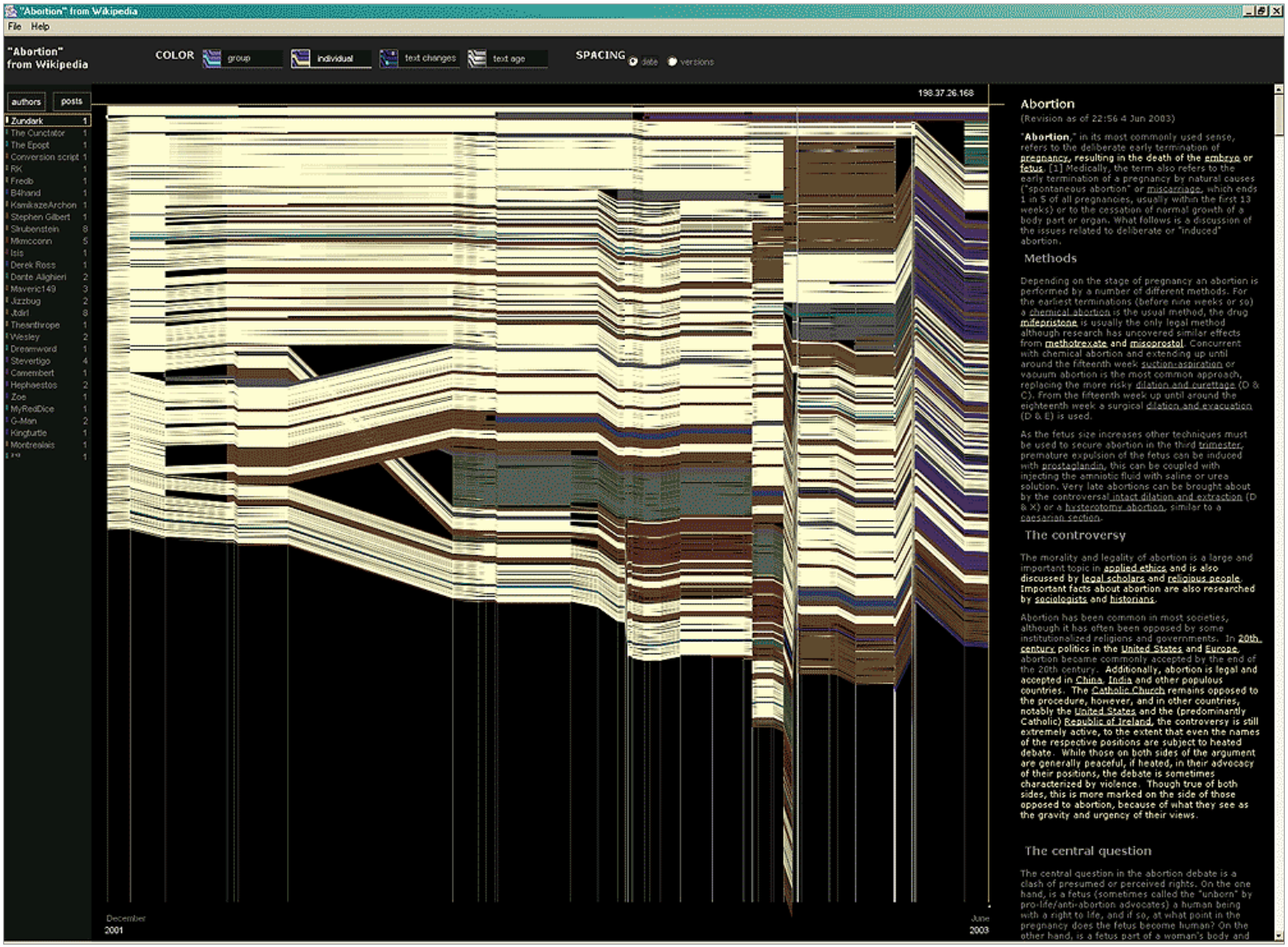


History Flow

- Space by versions: regular spacing
- Space by date
  - Space between successive revision lines is set proportional to the time between revision dates
  - Deemphasizes revisions that come in rapid succession
  - Indicates rhythms of collaboration among authors

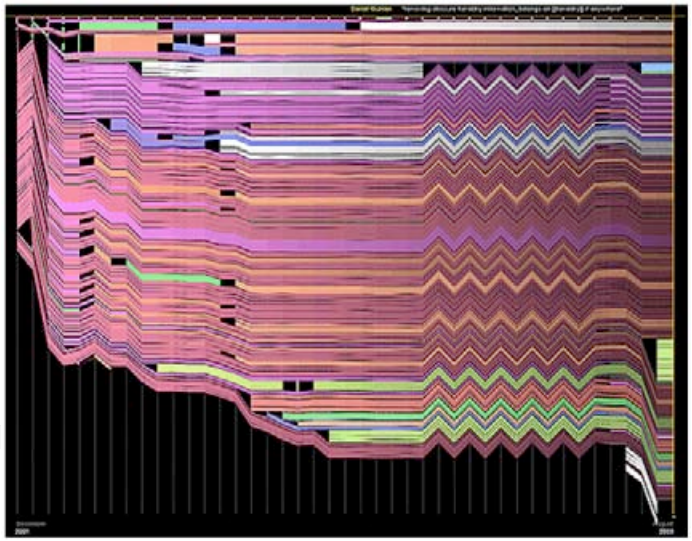



"Abortion" from Wikipedia
File Help
"Abortion" from Wikipedia
COLOR group individual text changes text age
SPACING date versions
authors posts
198.37.26.168
December 2001
June 2003
Abortion
(Revision as of 22:56 4 Jun 2003)
Methods
The controversy
The central question


"Abortion" from Wikipedia
File Help
"Abortion" from Wikipedia
COLOR group individual text changes text age
SPACING date versions
authors posts
198.37.26.168
December 2001
June 2003
Abortion
(Revision as of 22:56 4 Jun 2003)
Methods
The controversy
The central question

History Flow

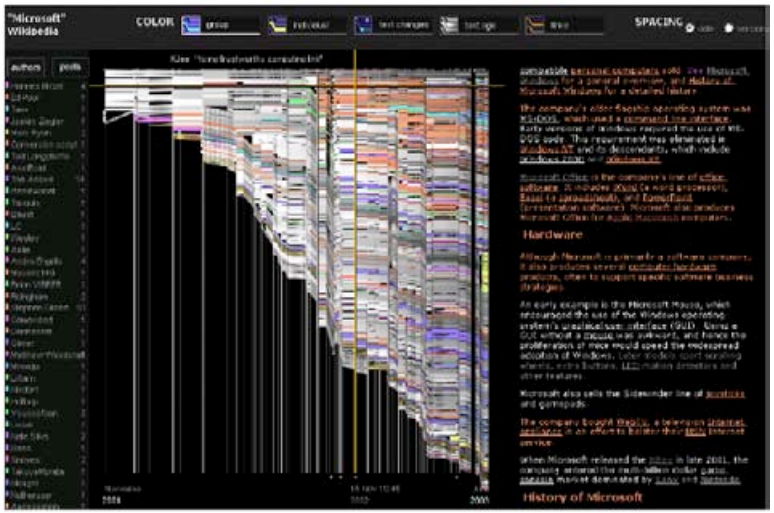
- Patterns of cooperation and conflict in wiki editing
- Vandalism and repair
  - Mass deletion: deletion of all contents on a page
  - Offensive copy: insertion of vulgarities or slurs
  - Phony copy: insertion of text unrelated to the page topic
  - Phony redirection: redirecting to an unrelated / malicious
  - Idiosyncratic copy: clearly one-sided, inflammatory text

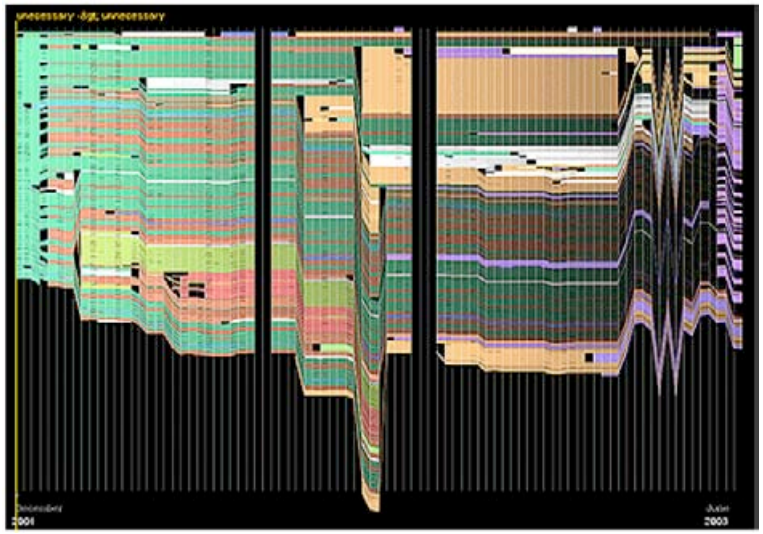


- Negotiation
  - Zigzag pattern that is dying out after a few versions
  - Also called edit wars
  - Example: two users fought over whether a chocolate sculpture called “coulage” really existed
  - 12 consecutive versions of reverting back and forth between two versions of the entry for chocolate

History Flow

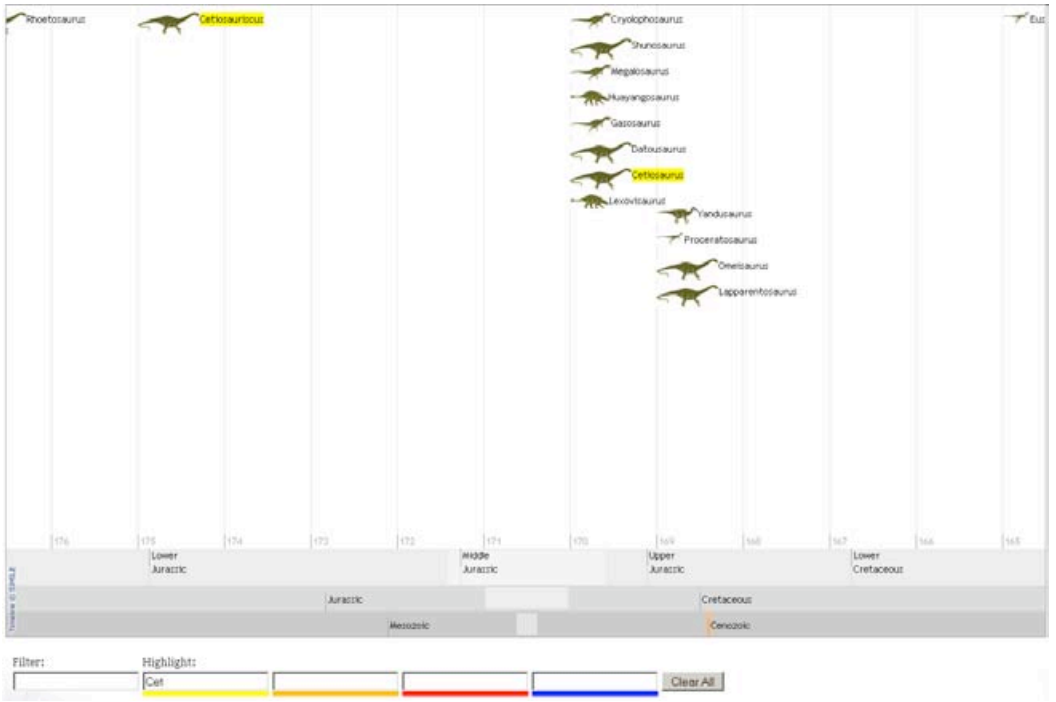
- Temporal patterns observed
- Length of page does not stabilize but change in size over time
  - Microsoft article shows constant growth
  - Article about abortion shows growth and shrinkage (some text was shifted to a separate entry)
- People tend to delete and insert text more frequently than moving text in an article
  - More gaps than crossings in visualizations
  - Proposed reason: editing window of wiki only shows 25 lines at once – without a decent overview, users may find it difficult to move text around





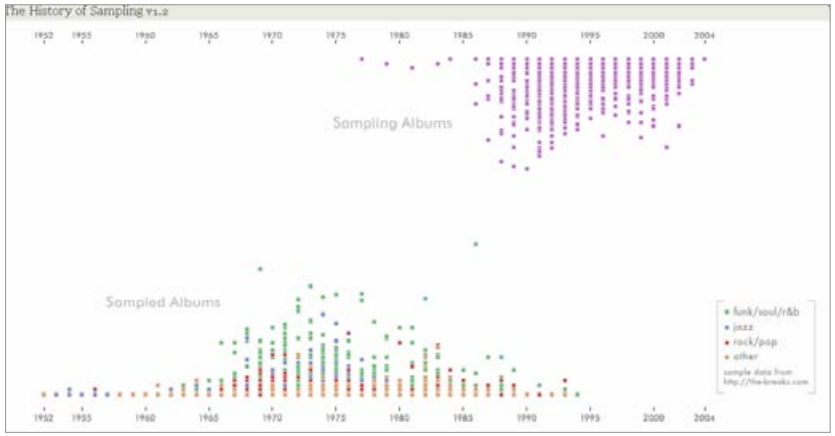
SMILE Timeline

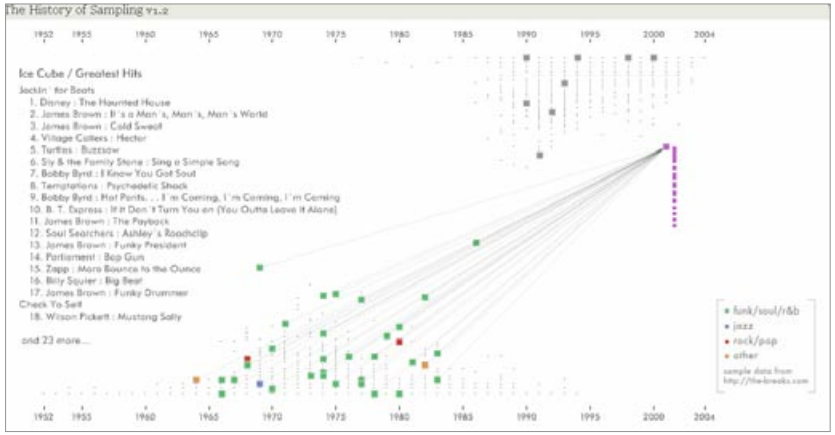
- http://simile.mit.edu/timeline/
- DHTML-based AJAX widget for visualizing time-based events
- Open source software (BSD license)
- Navigation by dragging time bends of different granularity
- Text-based highlight functionality
- **Demo**



History of Sampling

- http://jessekriss.com/projects/samplinghistory
- Each square represents an album
- Color-coding for genres
- Upper half: sampling artist
- Lower half: sampled artists
- X-axis: album release date
- Y-axis: number of samples on / from album – middle represents area of most sampling
- On-demand information (mouse or text-entry with automatic completion)
  - Song titles
  - Rectangles representing songs – taller rectangles indicate higher sample count
- **Demo**



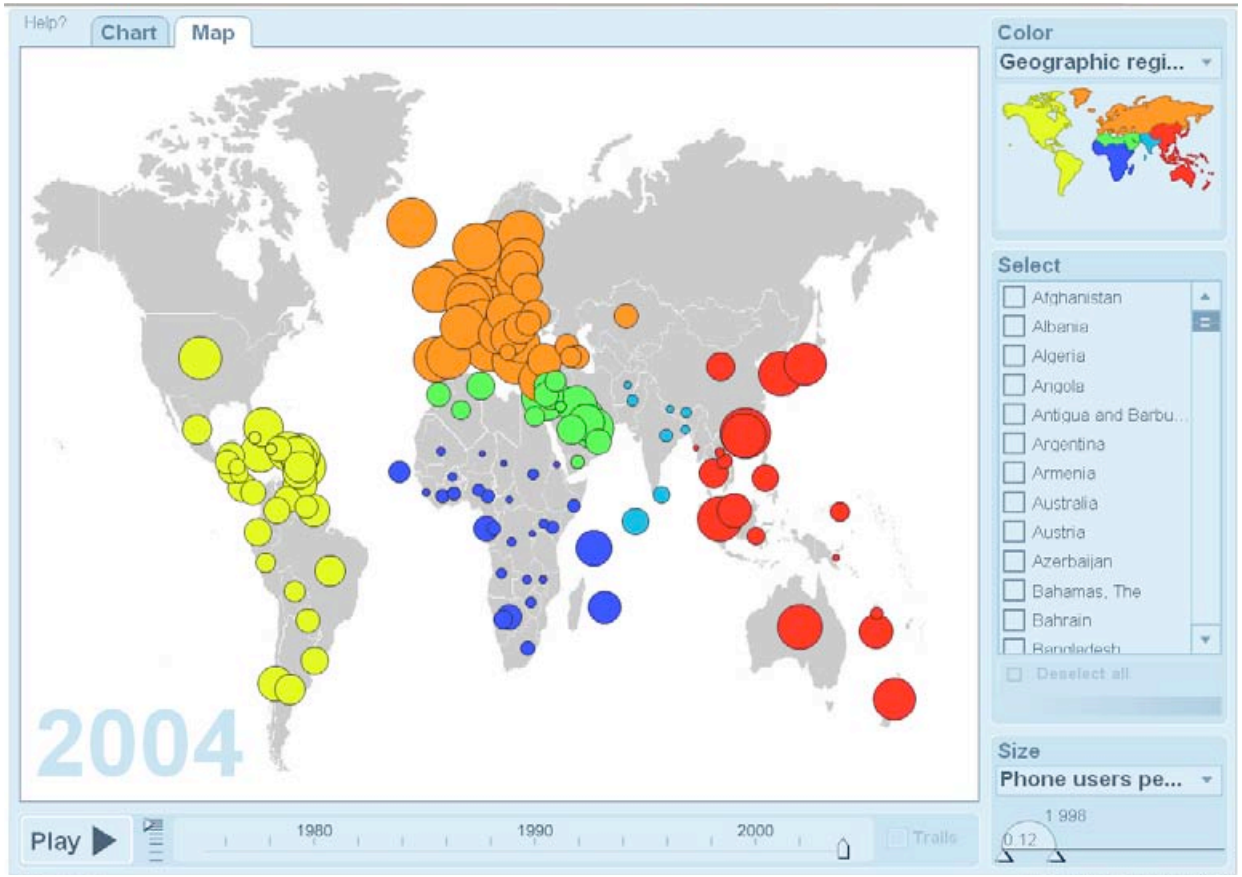


LifeLines

- Plaisant et al. 1998
- Visualizing patient records as timelines and dots on a single screen
  - Problems
  - Diagnoses
  - Test results
  - Medications, etc.
- Details-on-demand
- Zooming and panning
- Keyword search
- **Movie & Demo**

Gap Minder

- Use animation to visualize change over time - **demo**

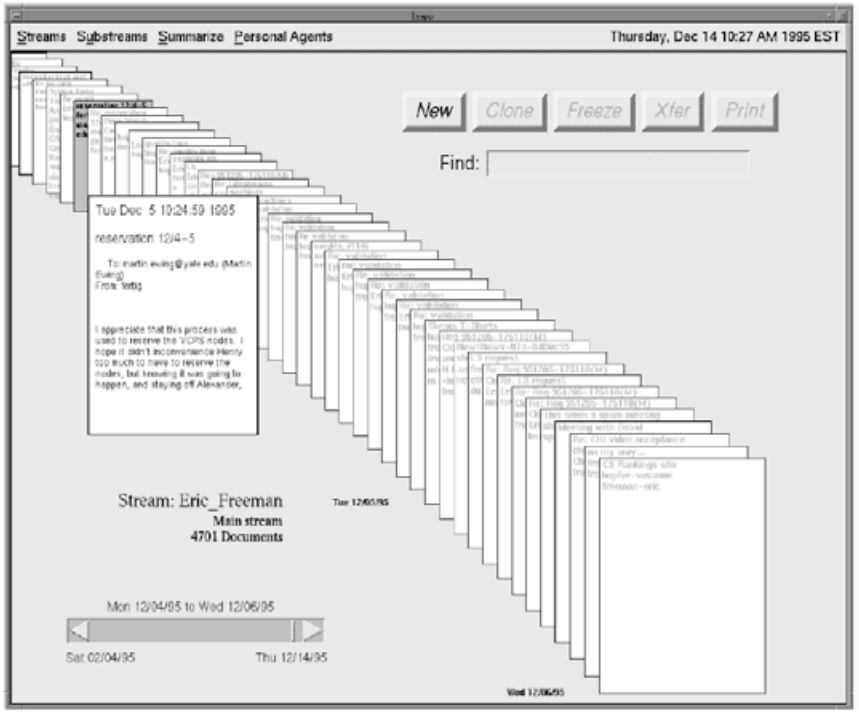


Outline

- Term clarification, user tasks & taxonomies
- Historic time-based graphics
- Timeline representations in Information Visualization
- Use case: time-based computer desktops
- Searching time-based data

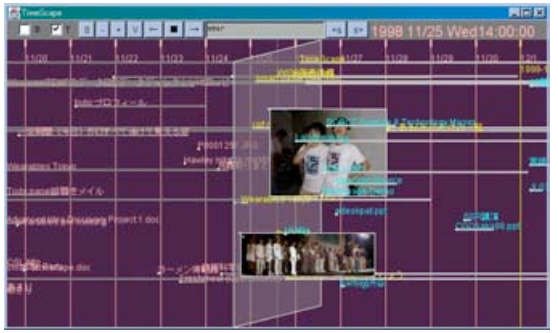
LifeStreams

- Freeman & Gelernter 1996
- Use time-ordered stream (and substreams) of documents as organizational metaphor for a computer desktop
- Design objectives
  - No need to name files or to choose a storage location for them
  - Automatic archiving
  - Inherent calendar functionality
  - Personal data should be accessible via a network connection and via any machine
- **Movie**



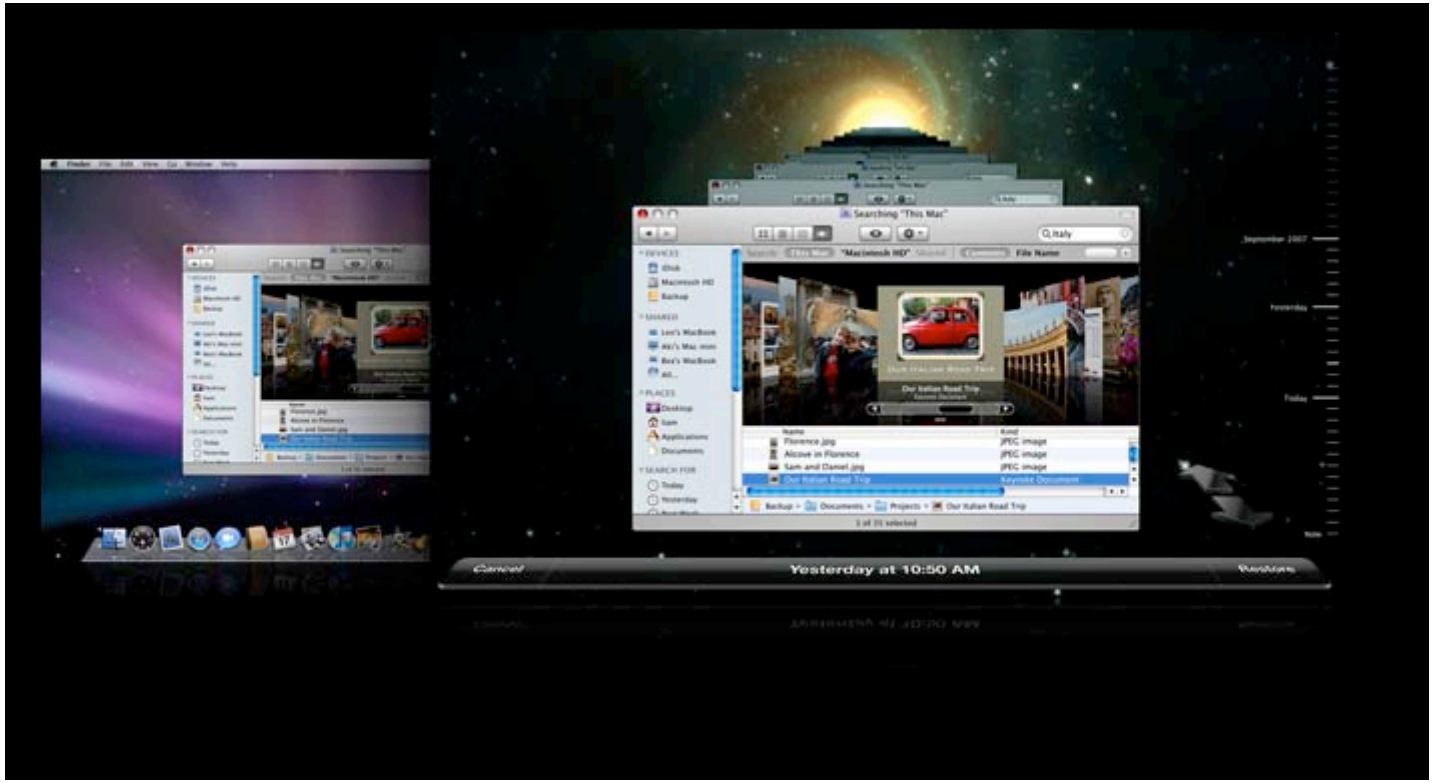
TimeScape

- Rekimoto 1999
- Spatial metaphor plus chronological navigation mechanism
  - People tend to pile files on the computer desktop for casual organization (no folders)
  - Time-order system intuitive for archiving
- Items fade out after some tine
- Time-travel dial to restore the desktop according to a designated point in time
- Thus users can remove items from the desktop, which currently are not of interest, without actually deleting them
- Users can also travel to the future to make a schedule
- Scheduled objects automatically appear on the desktop at the given time
- Scalable timeline view to support traveling over time
- Keyword search for past and future objects



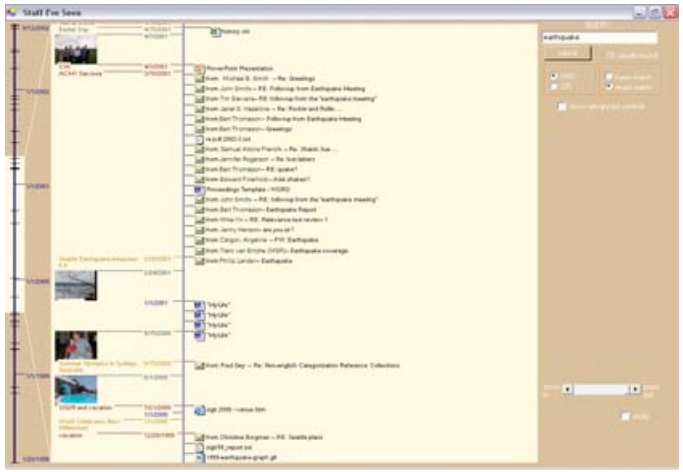
Time Machine

- Hmm, sounds familiar...
- Feature in Mac OS X Leopard (http://www.apple.com/de/macosx/features/timemachine.html)



Landmarks for Timeline Visualization

- Ringel et al. 2003
- Display results of queries on personal content, e.g. web pages, emails, documents
- Overview+detail timeline
  - Summary view showing the distribution of search hits over time
  - Detail view: inspection of individual search results
- Enhance visualization with public landmarks (e.g. holidays, important news events) and personal landmarks (e.g. photos and important calendar events)
- User study indicated significantly faster search times for landmark-augmented timeline compared to a timeline marked only by dates



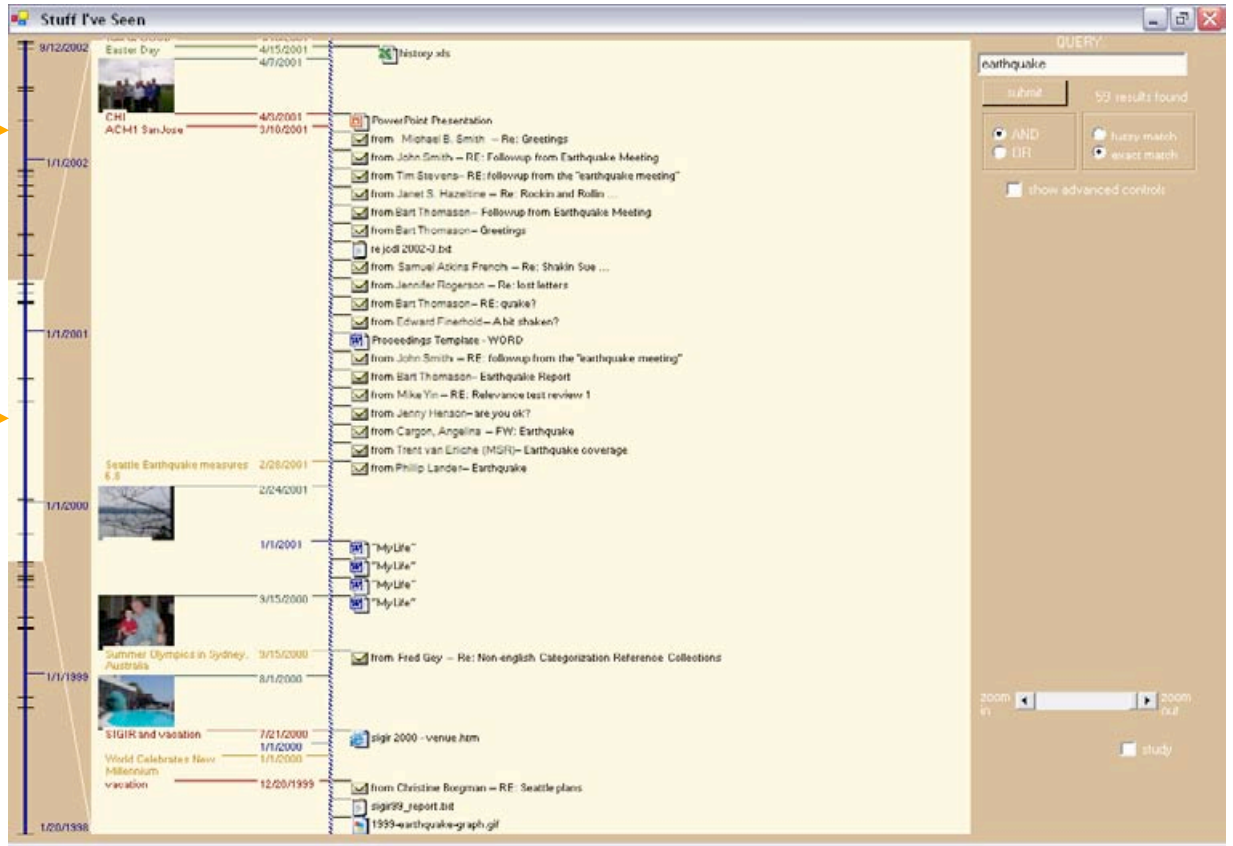
Landmarks for Timeline Visualization

Landmarks providing context

Detail View

Overview with hash marks showing the distribution of search results over time

Highlighted region indicates the period displayed in the detail view



Outline

- Term clarification, user tasks & taxonomies
- Historic time-based graphics
- Timeline representations in Information Visualization
- Use case: time-based computer desktops
- Searching time-based data

Detect Patterns in Time-Series

- Wijk & Selow 1999
- Univariate time-series data
  - Energy consumption over time
  - Number of employees present at work
- How to represent such simple data: draw a graph
- Problems
  - Number of measurements can become very large (e.g. measuring a value every 10 min during a year yields 52,560 values)
  - Repetitive patterns have different time scales (e.g. seasons, weeks, days)

Detect Patterns in Time-Series

- Initial attempt
  - Display data as fingerprints
  - Time data is treated as 2D – days and hours are mapped on different axes
  - Data values are visualized by color and the third dimension

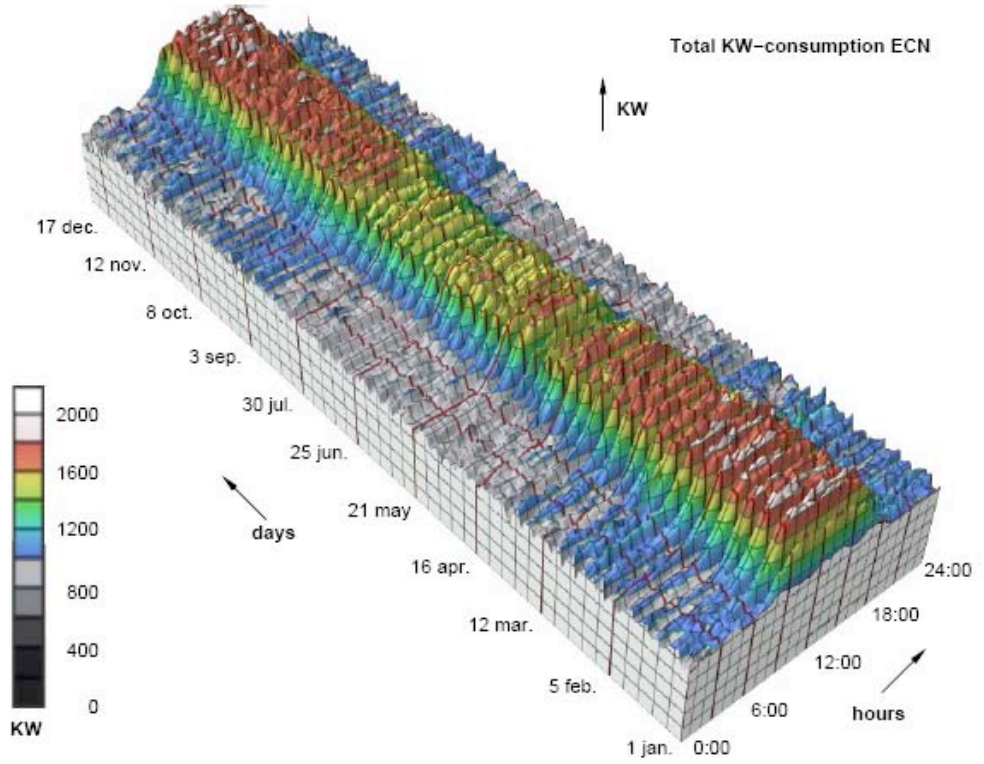


Figure 1. Power demand by ECN, displayed as a function of hours and days

Wijk & Selow 1999

Detect Patterns in Time-Series

- Advantages
  - All data is shown simultaneously
  - Seasonal trends can be observed
  - Day pattern can be observed
- Problems
  - Variation over the week is harder to discern
  - Day patterns of Saturdays and Sundays are obscured
  - Smoothing has been applied to make trends more easy to be perceived, but this eliminates fine details

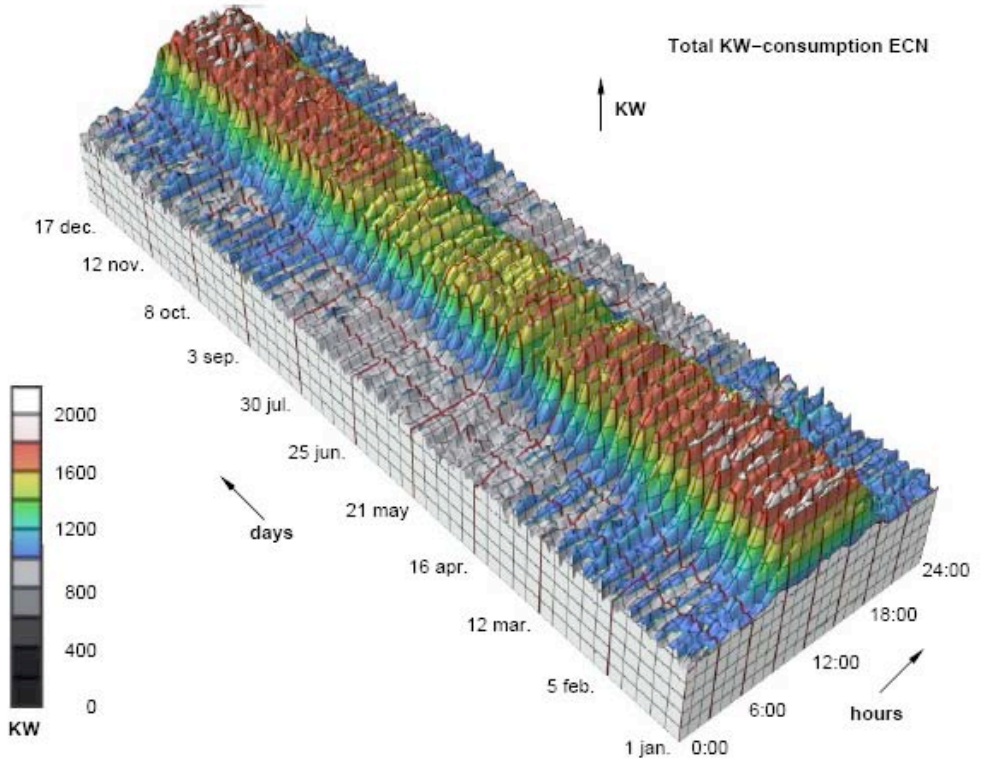


Figure 1. Power demand by ECN, displayed as a function of hours and days

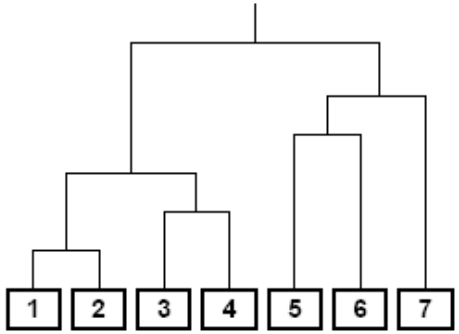
Wijk & Selow 1999

Detect Patterns in Time-Series

- Reduce amount of data to be displayed by using cluster analysis
- Objective: let the analysis tool decide which daily patterns are similar and show their distribution over the year
- Cluster analysis:
- Split time series into sequence of M day patterns
- Y_j, $j = 1,\ldots,M$
- M consists of sequence of pairs $(y_i,t_i), i = 1,\ldots,N$
- y_i denotes measured value
- t_i denotes the time elapsed since midnight

Detect Patterns in Time-Series

- Start with M clusters (each cluster contains one day pattern)
- Compute mutual differences between clusters
- Merge the two clusters being most similar into a new cluster
- Repeat clustering until a single large cluster results (contains average of all day patterns)
- Some proposed distance measures (y_i and z_i are day patterns)



Eliminate slow trends by subtracting the average difference

$$d_{sh} = \sqrt{\sum (y_i - z_i - \Delta)^2 / N},$$

$$\Delta = \sum (y_i - z_i) / N.$$

By average geometric distance

$$d_{rms} = \sqrt{\sum (y_i - z_i)^2 / N}.$$

By similarity of shape (normalize by maximum value in the sequence)

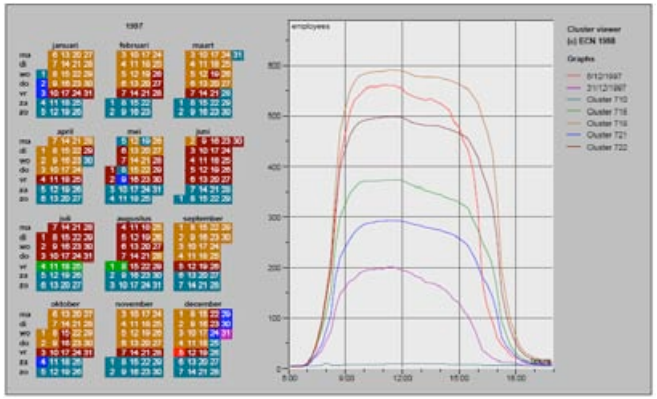
$$d_{nm} = \sqrt{\sum (y_i / y_{\max} - z_i / z_{\max})^2 / N}.$$

By the difference of peak values

$$d_{ma} = | \, y_{\max} - z_{\max} \, | \, .$$

Detect Patterns in Time-Series

- Based on cluster analysis a selection of the most significant clusters can be displayed
- Average daily pattern per cluster is shown as graph
- Problem: how to visualize whether similarities of day patterns in a cluster is due to season, week, day etc.?
- Clusters are shown on a calendar, i.e. each day in the calendar is colored according to the cluster to which it belongs



Wijk & Selow 1999

Detect Patterns in Time-Series

Wijk & Selow 1999

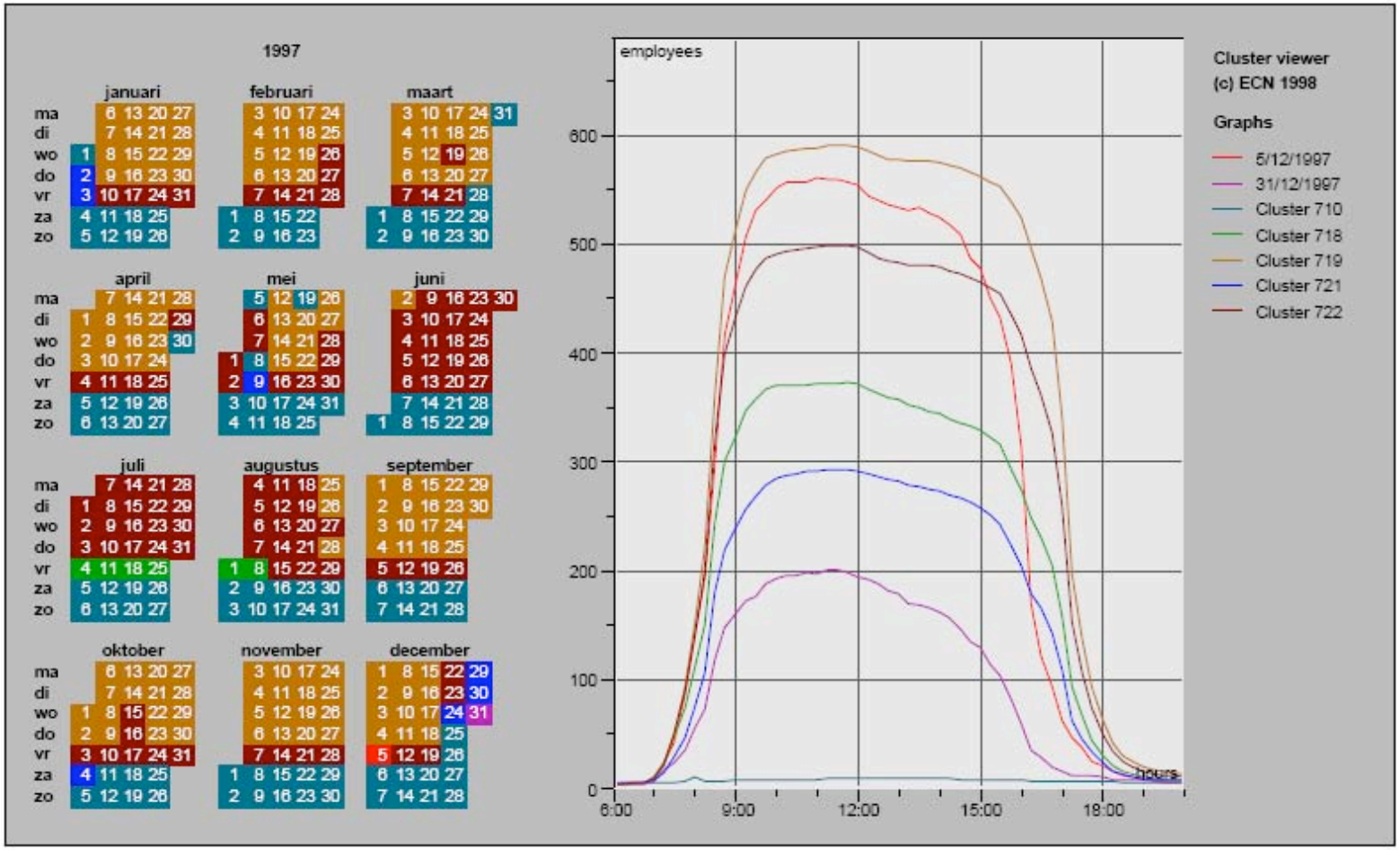


Figure 4. Calendar view of the number of employees

Detect Patterns in Time-Series

Wijk & Selow 1999

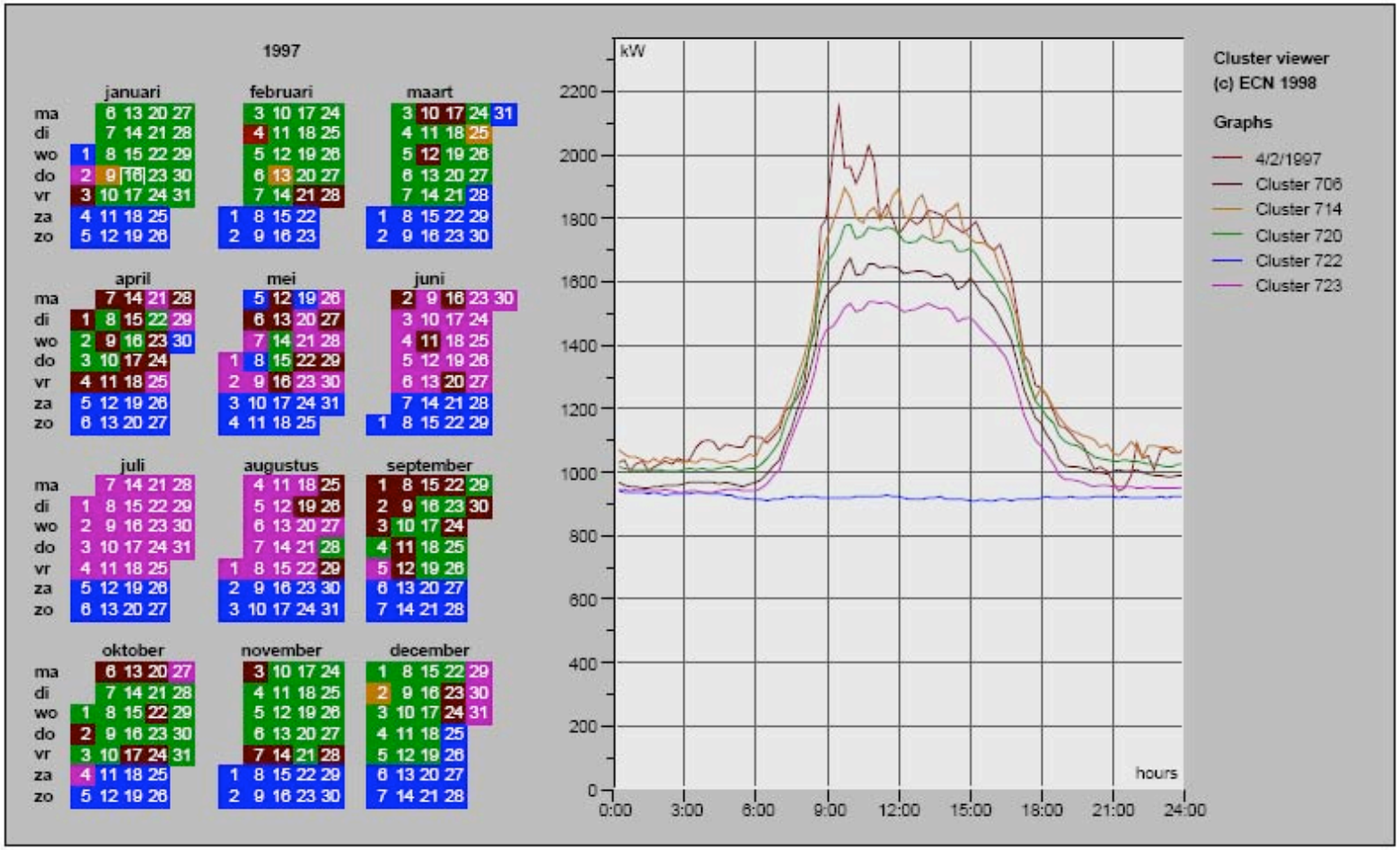
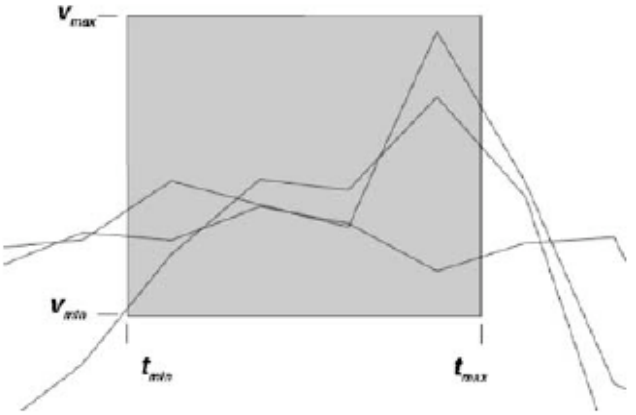
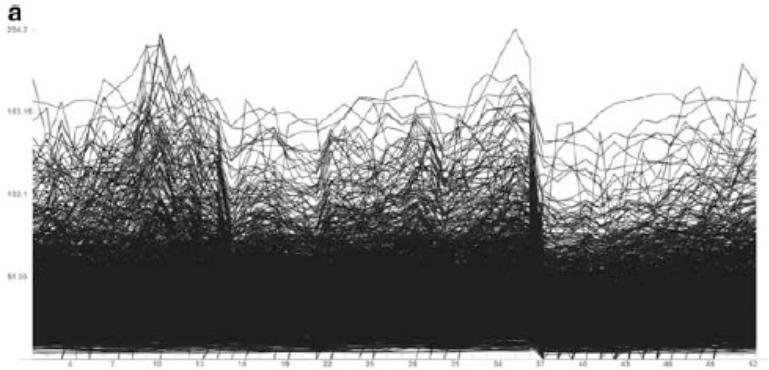


Figure 5. Cluster analysis of power demand by ECN

TimeSearcher

- Hochheiser & Shneiderman 2004
- Dynamic query tools for searching time series for trends and patterns
- Timebox: rectangular query regions (bounding box) drawn on a 2D representation of time series data
  - X-axis extent: time period of interest
  - Y-axis extend: constraint on the range of values in the time period
- Graphs that do not have values in the given ranges are removed
- Example: 52 weekly stock prices for 1430 stocks

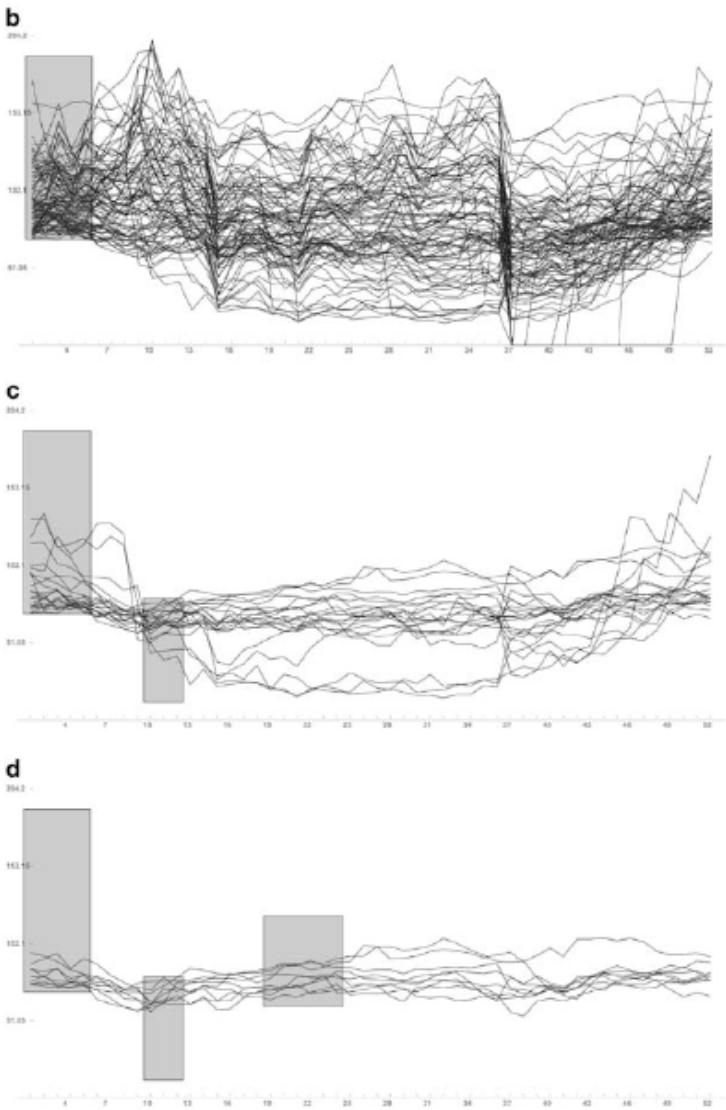




Ludwig Maximilians Universität München

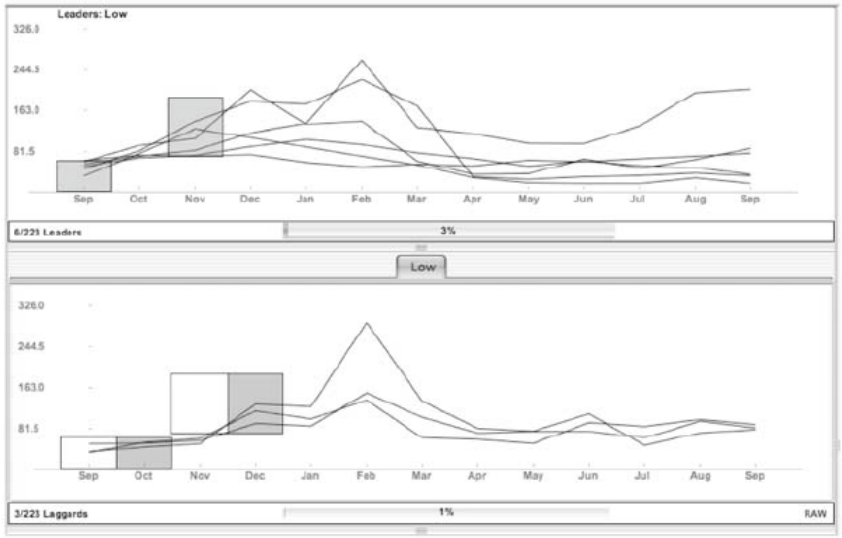
TimeSearcher

- Multiple timeboxes to define conjunctive queries
- Filter hierarchy by order of box creation
- Modifying queries: boxes can be scaled, translated, removed, singly or together
- Box manipulation also via range sliders
- Query by example
  - Pick one one graph as query
  - Timeboxes are generated for each data point
  - Boxes are centered around the values of the query graph
  - Manipulate all timeboxes for varying required similarity to graph



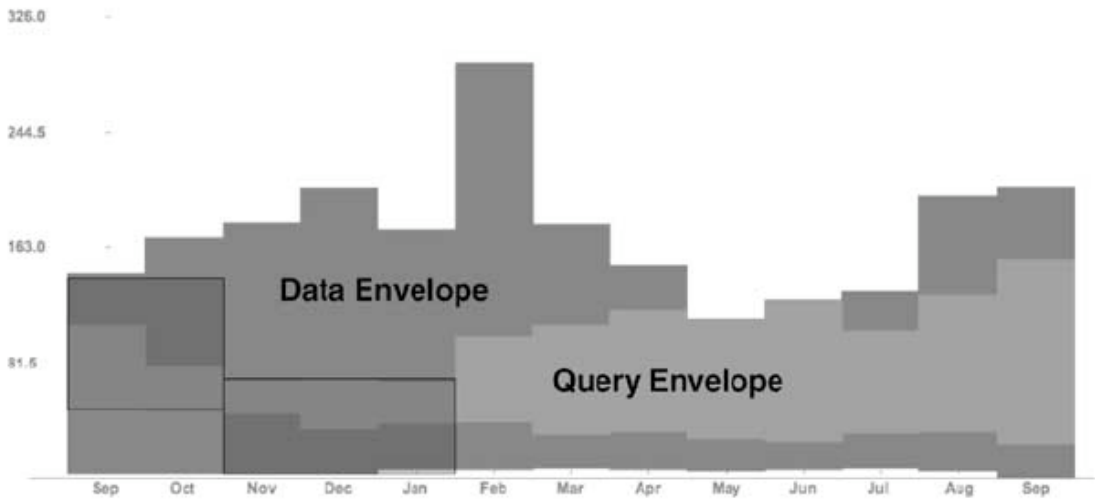
TimeSearcher

- Identify leaders & laggards
- Items with behavior trends that anticipate changes of other items in the data set
- Generate new queries by shifting old queries one time period to the right
- Graphs will be displayed that undergo the same transitions similar to leader graphs, i.e. that satisfy the query defined by new timeboxes
- Identify similar trends in other value ranges by translating the boxes



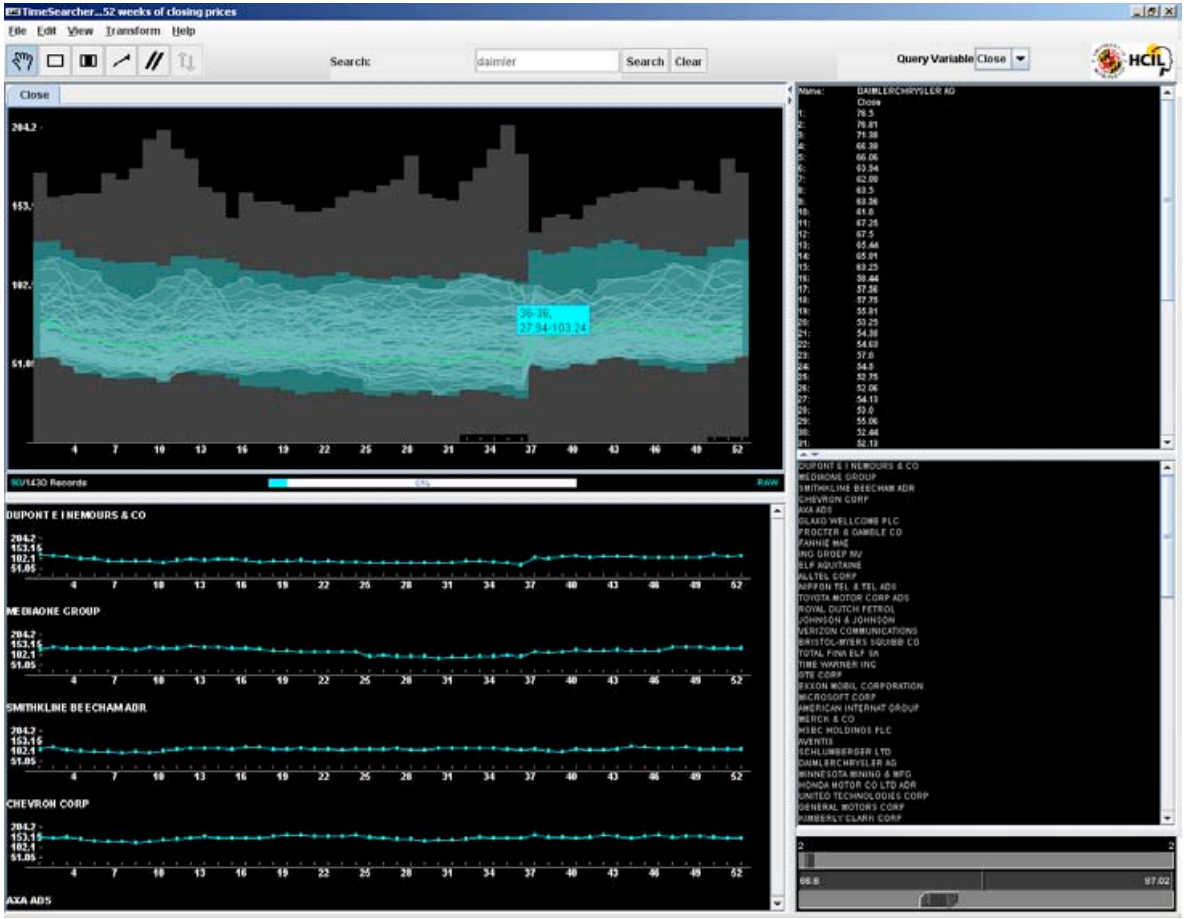
TimeSearcher

- Problem
  - Visual clutter when many graphs are displayed
  - Computational overhead of drawing all graphs and mouse-over handling
- Envelope: contour of extreme values as low-resolution overview
  - Data envelope: all graphs of the data set
  - Query envelope: graphs of the result set
- Threshold to define below which number graphs are displayed



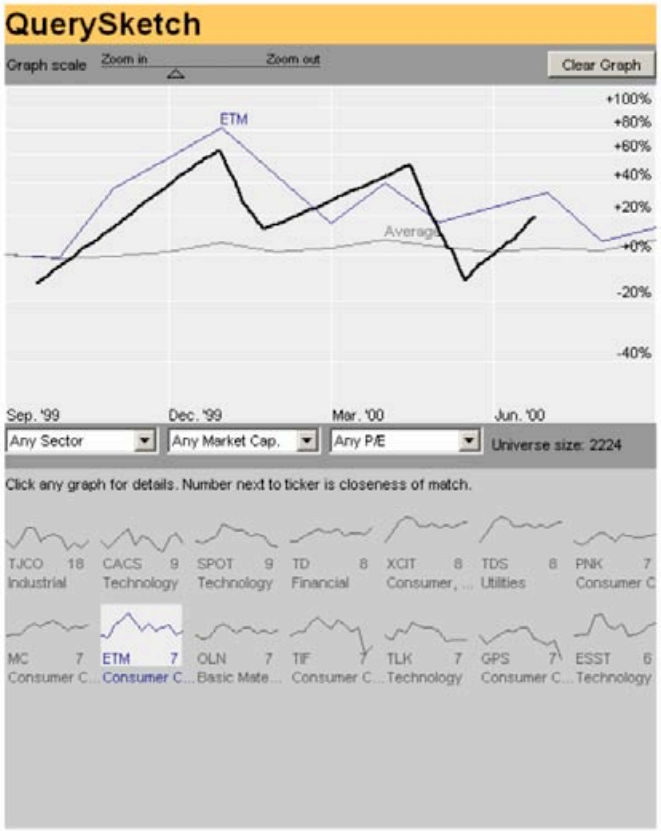
TimeSearcher

- Demo



QuerySketch

- Wattenberg 2001
- Query historical stock price data by sketching a graph
- Overview of result set by showing miniature thumbnails + numerical indicator of similarity
- Pattern matching: metric based on the Euclidean distance between sequences of monthly percentage price changes
- Also multiple partial queries possible
- Scale of the graph is logarithmic: stocks with steady growth rate can be queried via a straight line instead of drawing an accurate exponential curve
- Zooming functionality
- Proposed: querying for inequality, e.g. stocks that move above a certain rate
- **Demo**



Sources and Literature

- Obligatory literature to summarize
  - Jarke J. van Wijk & Edward R. van Selow: "Cluster and Calendar based Visualization of Time Series Data“, 2002.
  - http://www.win.tue.nl/~vanwijk/clv.pdf
- Additonal source
  - John Stasko, CS7450, lecture material