8. Time-Based Data
Visualizing Change over time

Lecture „Informationsvisualisierung”
Prof. Dr. Andreas Butz, WS 2012/13
Concept and slides: Thorsten Büring,
3rd, revised edition

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

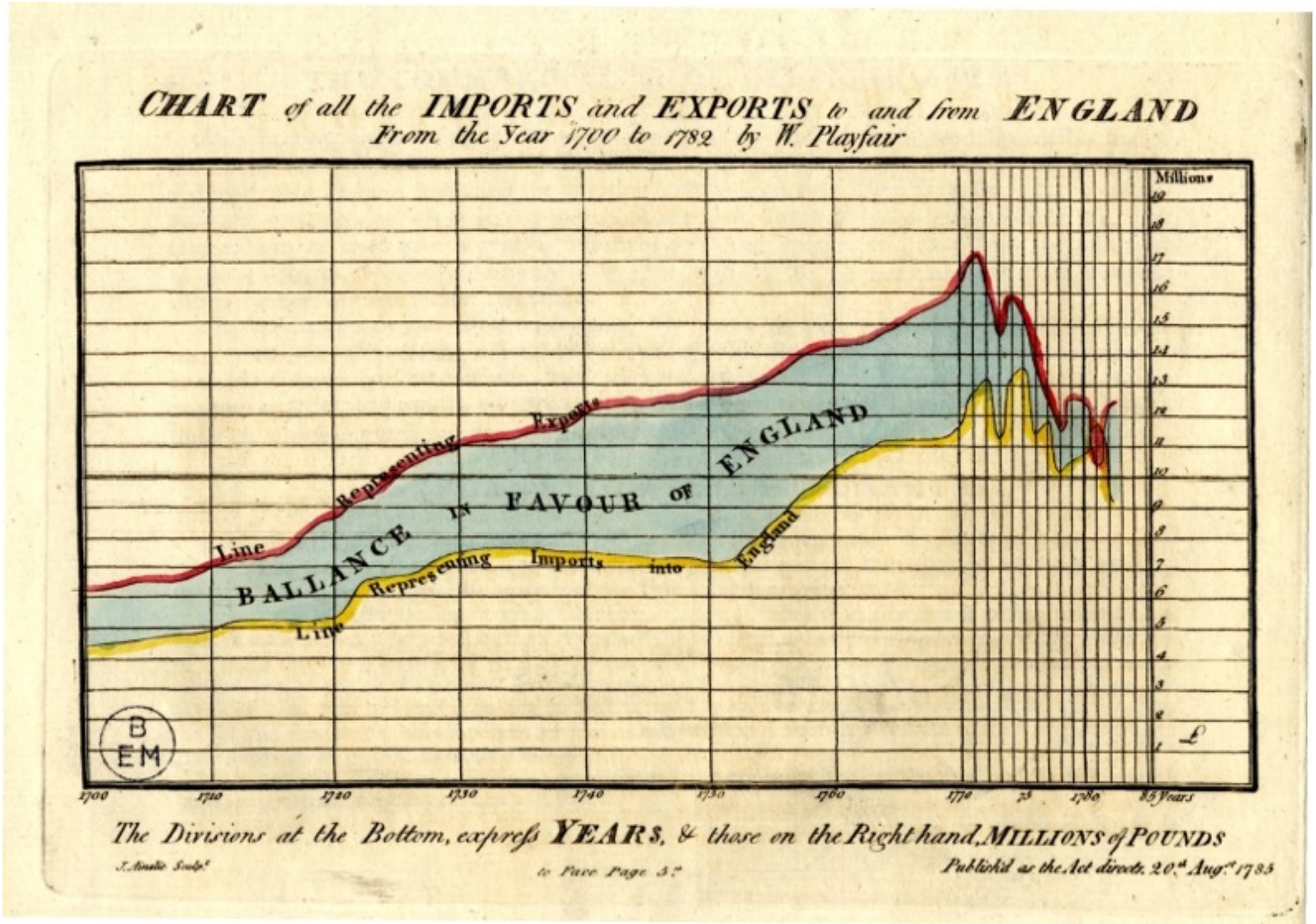
- MacEachern 1995
  - Does a data element exist at time t?
  - When does a data element exist?
  - How long does a data element exist?
  - How often does a data element occur?
  - How fast are data elements changing?
  - In what order do data elements appear?
  - Do data elements exist together?
- Stasko
  - 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
    - Allows 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

Imports and Exports

- Playfair 1785



Train Schedule Paris - Lyon

- Marey 1885 (from Tufte 2001)

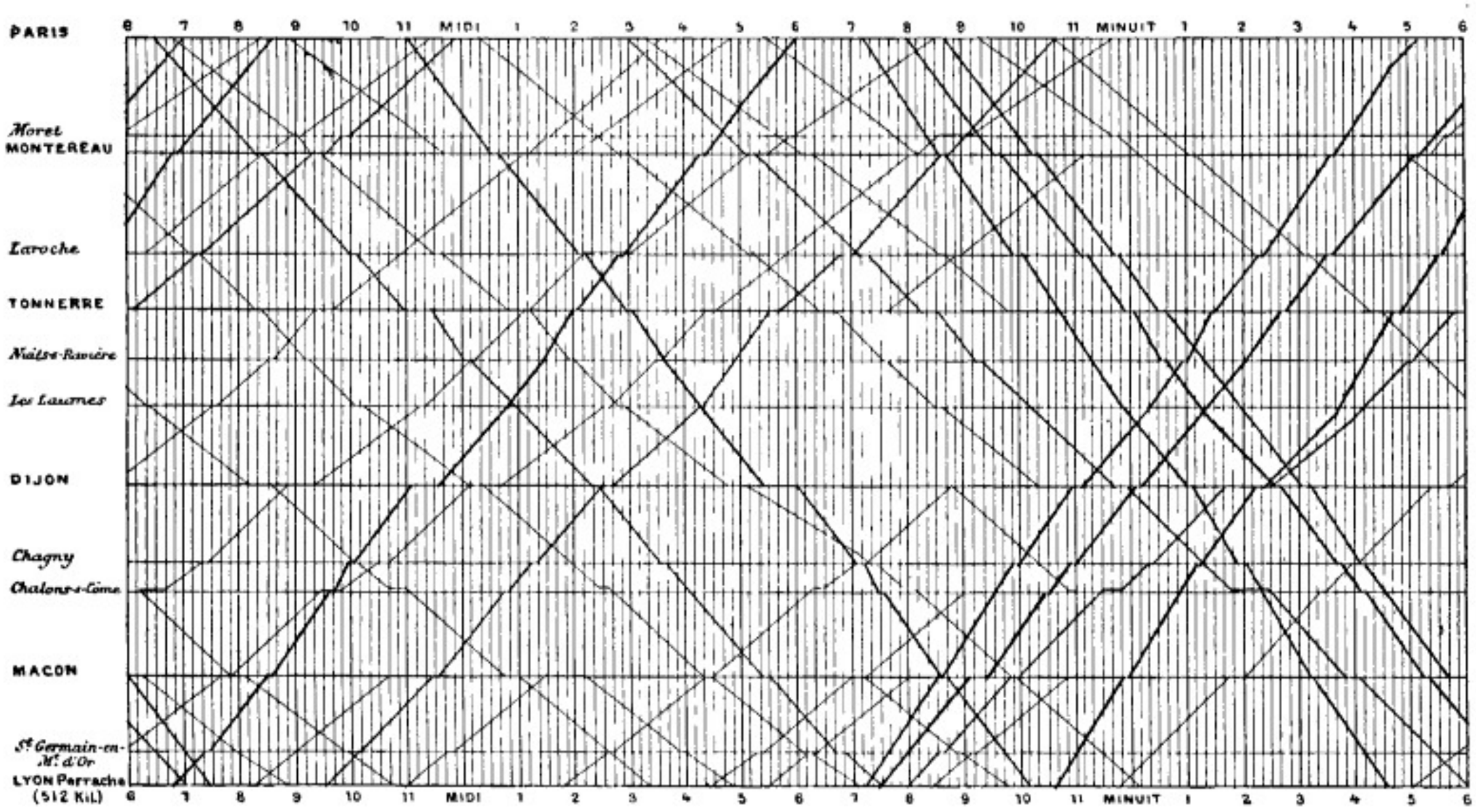
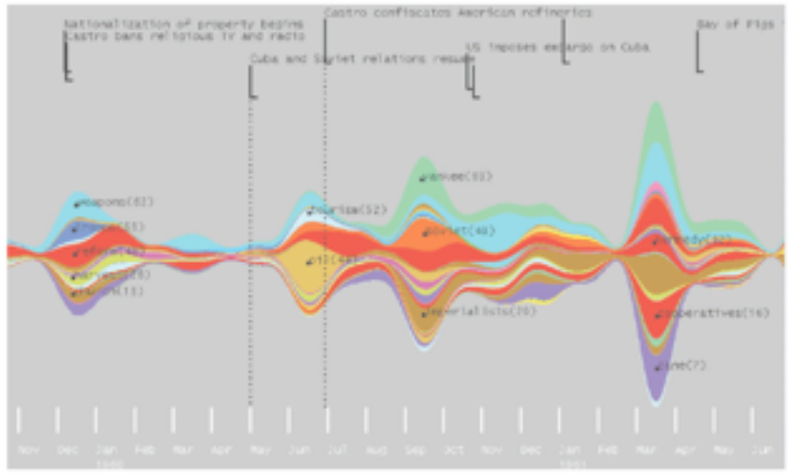


Fig. 5. — Graphique de la marche des trains sur un chemin de fer, d'après la méthode de Ibry.

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 (1959-1998)
  - Flow from left to right represents movement through time
  - Selected thematic content shown as color currents
  - Change of 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

ThemeRiver

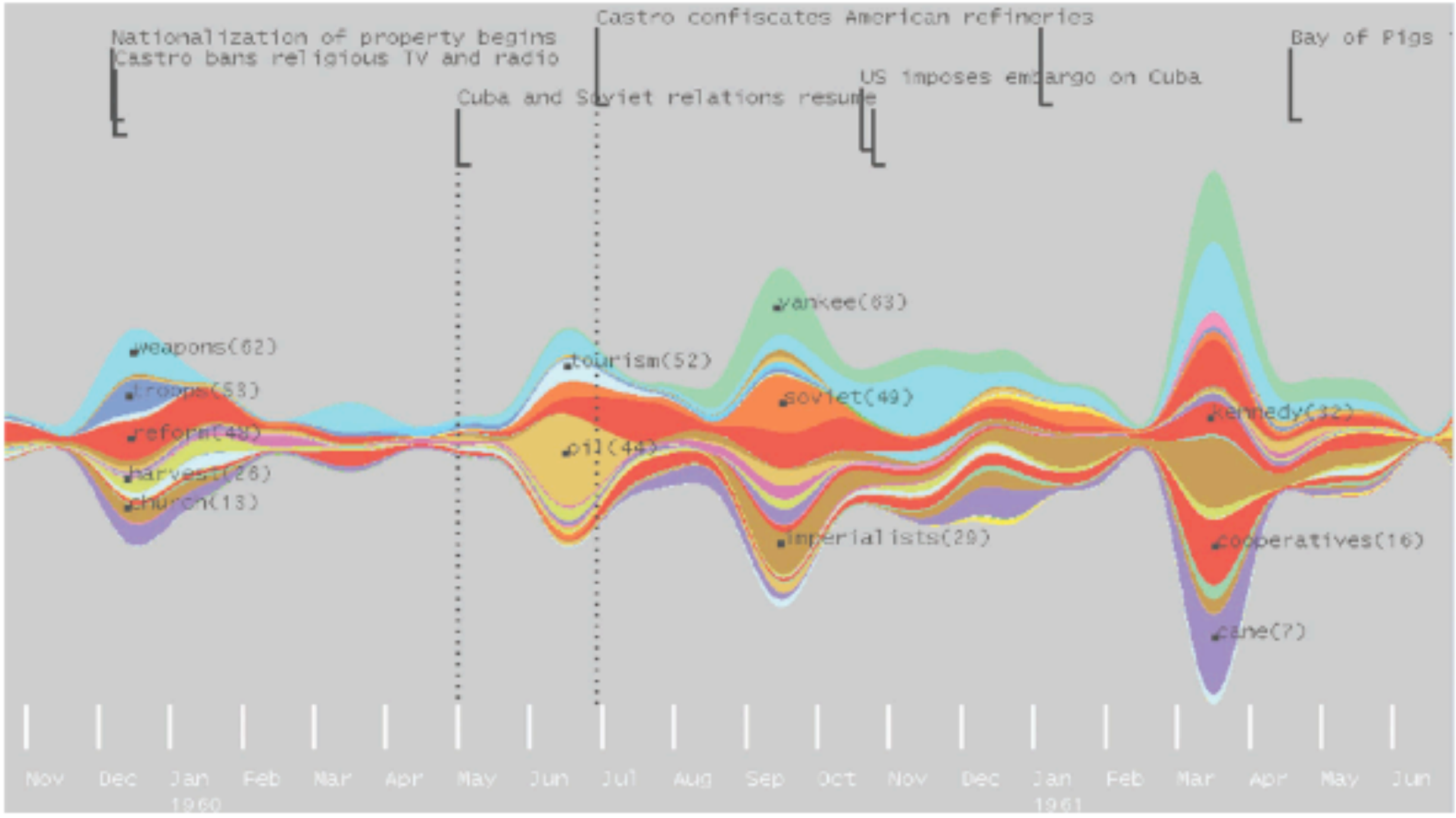
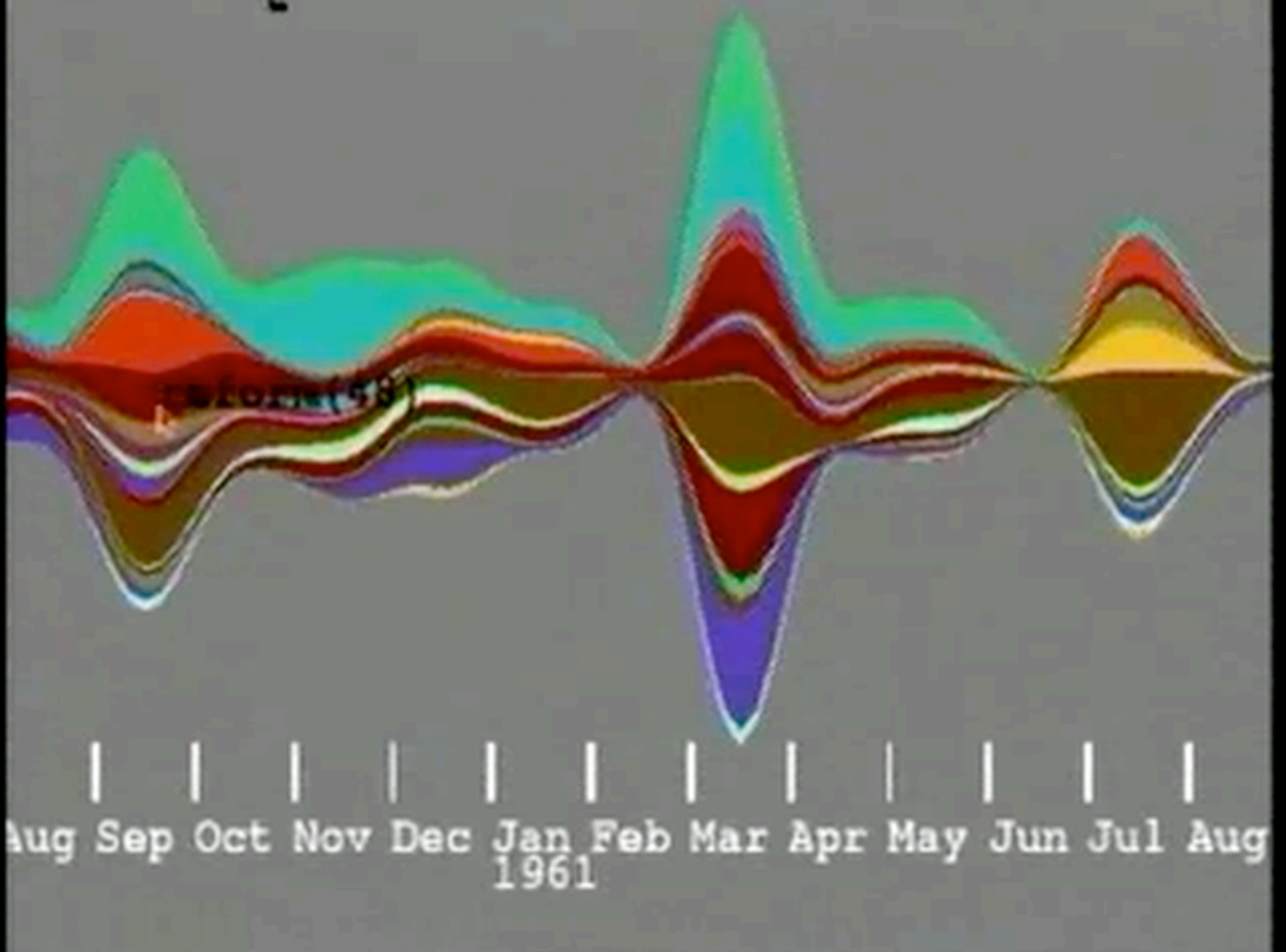
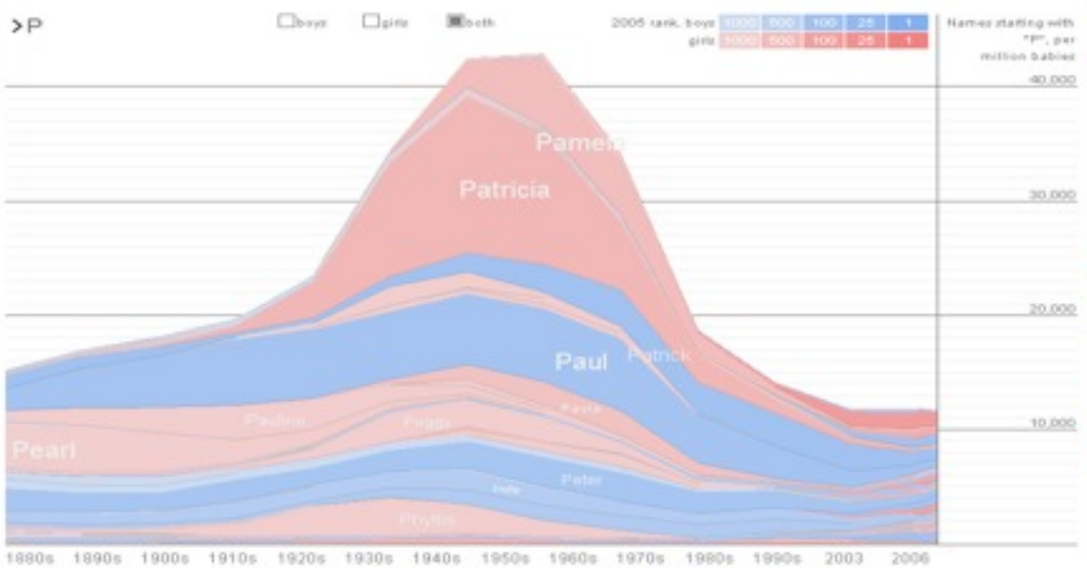


Fig. 2. ThemeRiver uses a river metaphor to represent themes in a collection of Fidel Castro's speeches, interviews, and articles from the end of 1959 to mid-1961.


Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug
1961

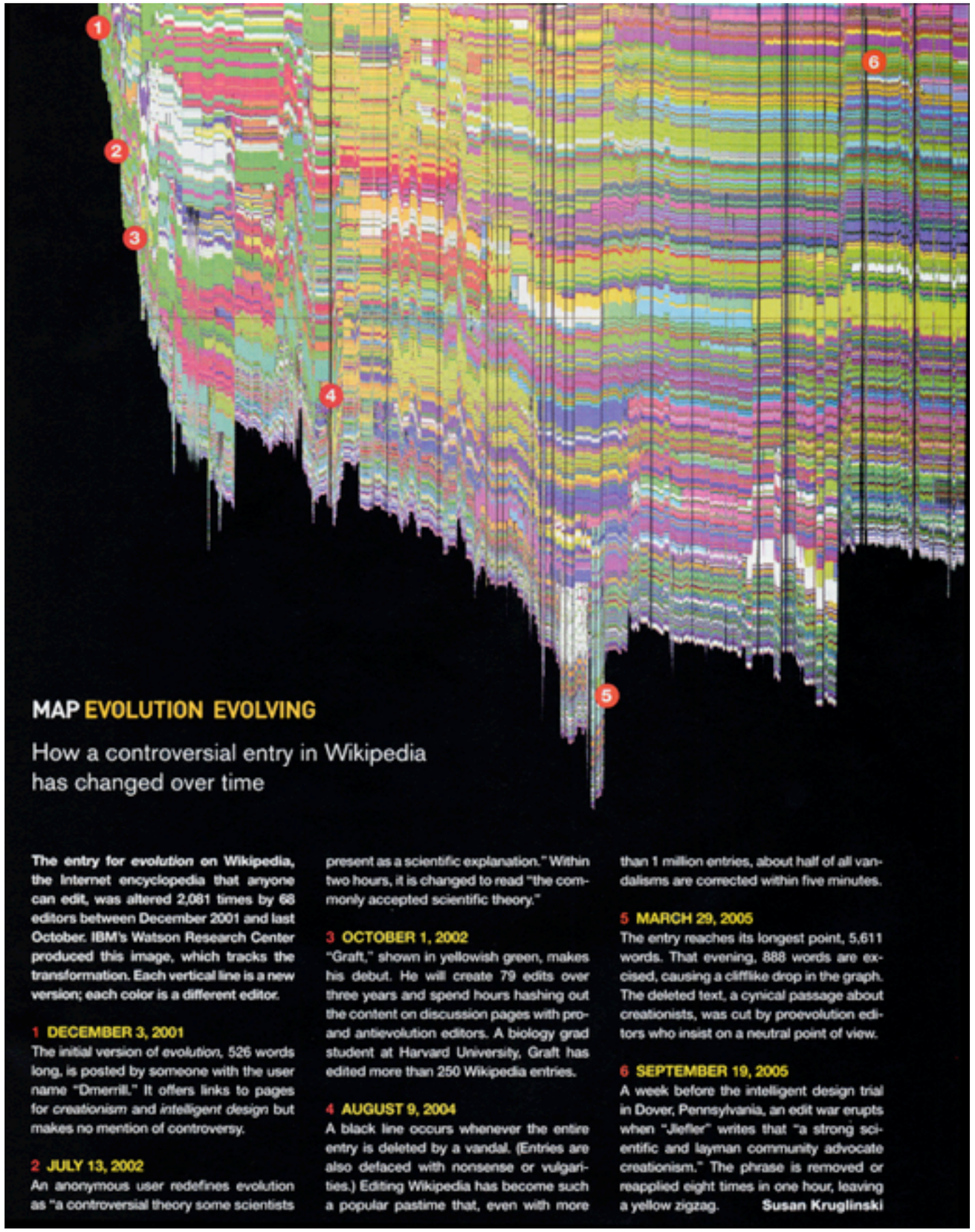
Baby Name Wizard

- 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
  - Value for points in time
- Animated transitions
- http://www.babynamewizard.com/voyager



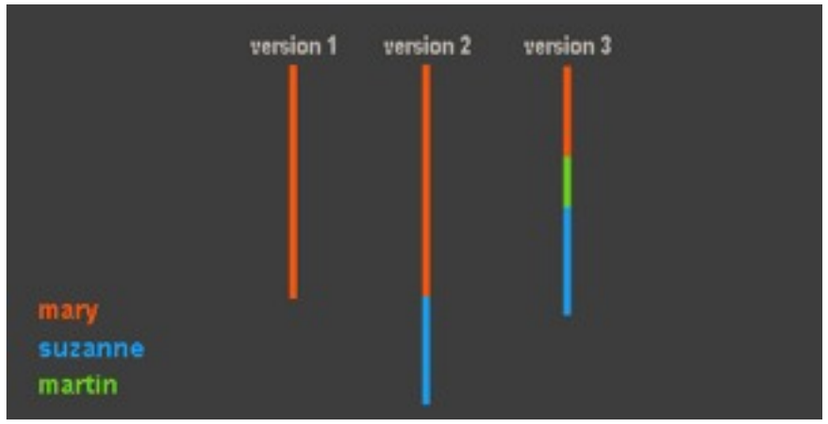
History Flow

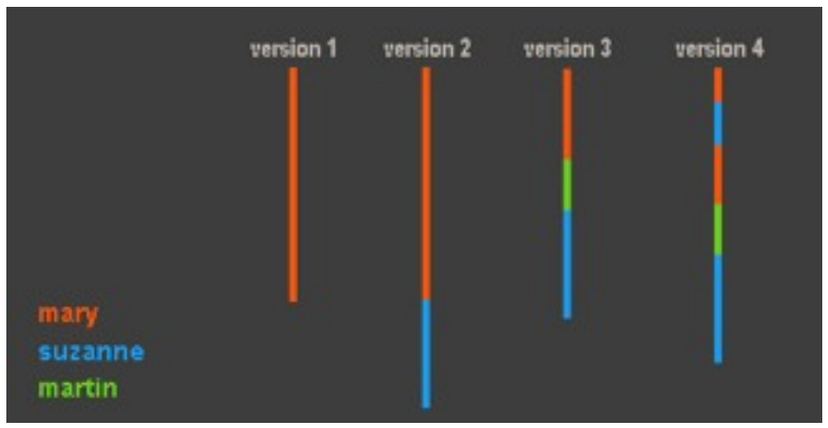
- Viégas et al. 2004
- http://hint.fm/projects/historyflow/
- http://fernandaviegas.com/wikipedia.html
- Visualizes
  - Dynamic, evolving documents
  - Interactions of multiple collaborating authors
- Example: evolutionary history of wiki pages on Wikipedia
- Highlight deletions and insertions of text by different authors over time

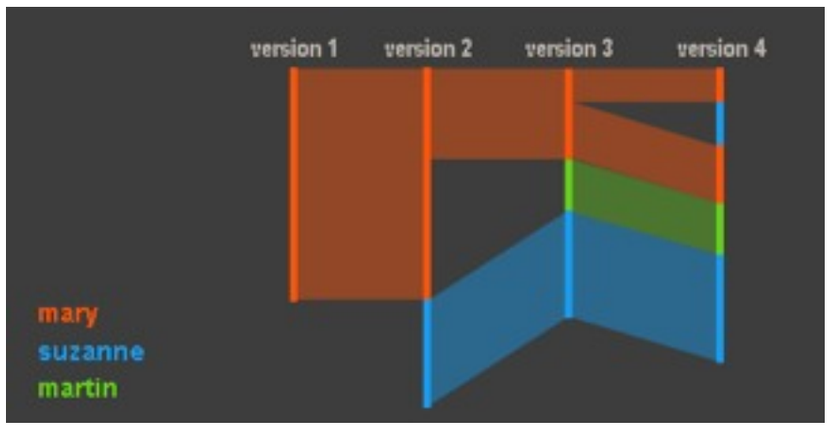


History Flow

- Visualization construction

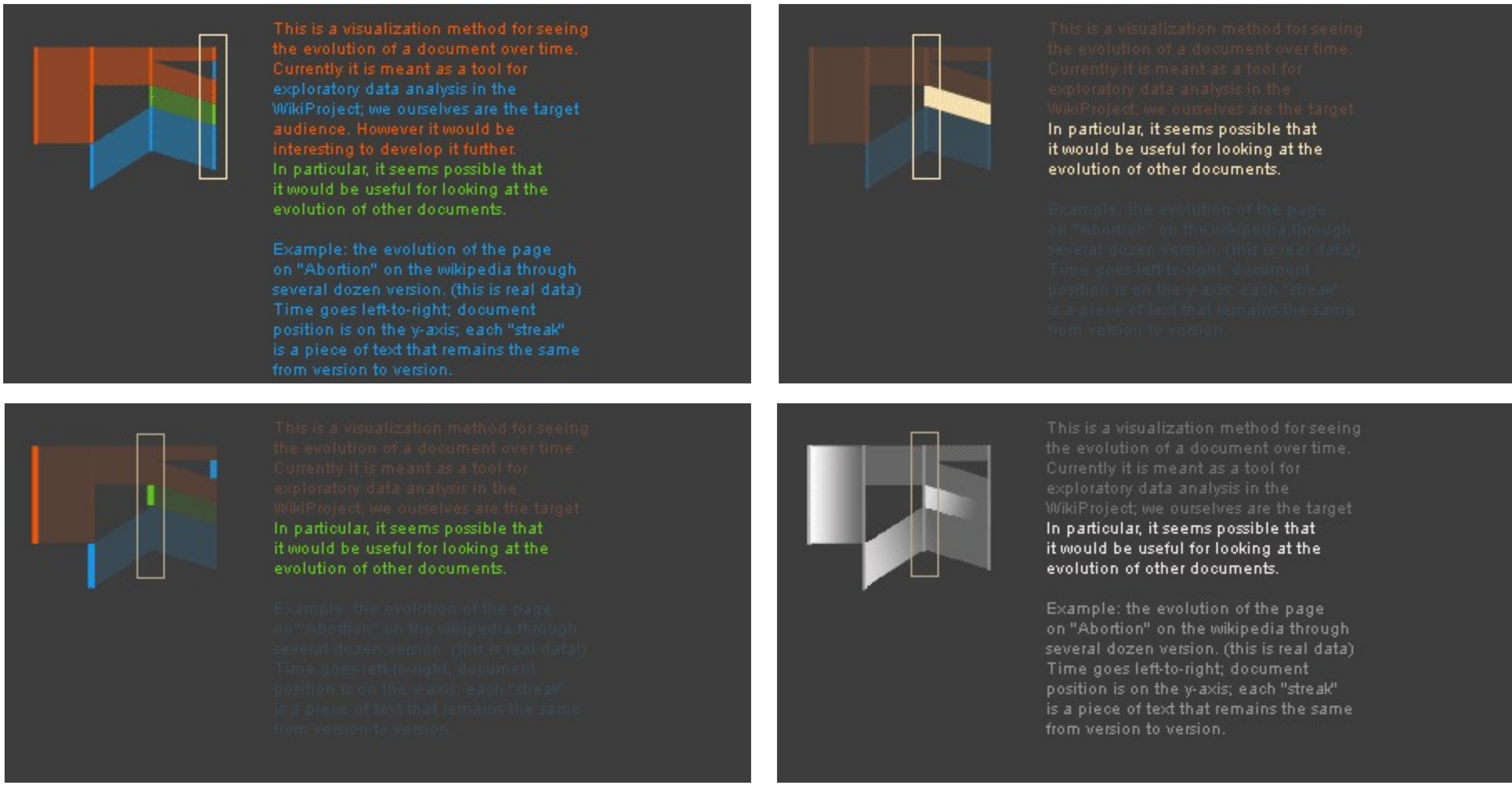






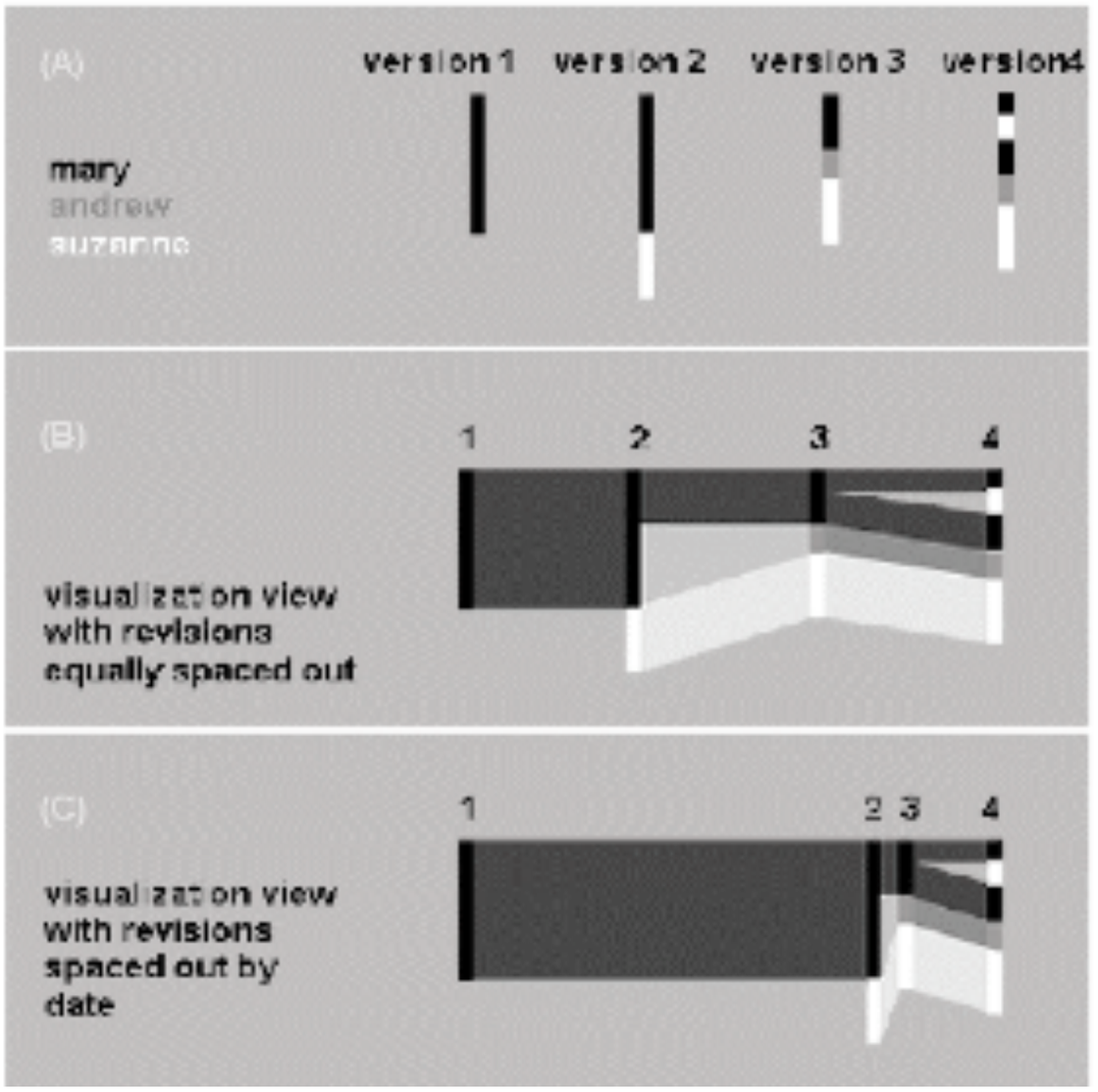
History Flow

- 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

- Zundark 1
- The Cunctator 1
- The Epopt 1
- Conversion script 1
- RK 1
- Fredb 1
- B4hand 1
- KamikazeArchon 1
- Stephen Gilbert 1
- Sirubenstein 8
- Mxmcconn 5
- Isis 1
- Derek Ross 1
- Dante Alighieri 2
- Maveric149 3
- Jizzbug 2
- Jdirl 8
- Theanthrope 1
- Wesley 2
- Dreamword 1
- Stevertigo 4
- Camembert 1
- Hephaestos 2
- Zoe 1
- MyRedDice 1
- G-Man 2
- Kingturtle 1
- Montrealais 1

198.37.26.168

December 2001

June 2003

Abortion

(Revision as of 22:56 4 Jun 2003)

"**Abortion**," in its most commonly used sense, refers to the deliberate early termination of **pregnancy, resulting in the death of the embryo or fetus**. [1] Medically, the term also refers to the early termination of a pregnancy by natural causes ("spontaneous abortion" or miscarriage, which ends 1 in 5 of all pregnancies, usually within the first 13 weeks) or to the cessation of normal growth of a body part or organ. What follows is a discussion of the issues related to deliberate or "induced" abortion.

Methods

Depending on the stage of pregnancy an abortion is performed by a number of different methods. For the earliest terminations (before nine weeks or so) a chemical abortion is the usual method, the drug **mifepristone** is usually the only legal method although research has uncovered similar effects from **methotrexate** and **misoprostol**. Concurrent with chemical abortion and extending up until around the fifteenth week suction-aspiration or vacuum abortion is the most common approach, replacing the more risky dilation and curettage (D & C). From the fifteenth week up until around the eighteenth week a surgical dilation and evacuation (D & E) is used.

As the fetus size increases other techniques must be used to secure abortion in the third trimester, premature expulsion of the fetus can be induced with prostaglandin, this can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought about by the controversal intact dilation and extraction (D & X) or a hysterotomy abortion, similar to a caesarian section.

The controversy

The morality and legality of abortion is a large and important topic in **applied ethics** and is also discussed by **legal scholars** and **religious people**. Important facts about abortion are also researched by **sociologists** and **historians**.

Abortion has been common in most societies, although it has often been opposed by some institutionalized religions and governments. In **20th century** politics in the **United States** and **Europe**, abortion became commonly accepted by the end of the 20th century. **Additionally, abortion is legal and accepted in China, India and other populous countries. The Catholic Church remains opposed to the procedure, however, and in other countries, notably the United States and the (predominantly Catholic) Republic of Ireland, the controversy is still extremely active, to the extent that even the names of the respective positions are subject to heated debate. While those on both sides of the argument are generally peaceful, if heated, in their advocacy of their positions, the debate is sometimes characterized by violence. Though true of both sides, this is more marked on the side of those opposed to abortion, because of what they see as the gravity and urgency of their views.**

The central question

The central question in the abortion debate is a clash of presumed or perceived rights. On the one hand, is a fetus (sometimes called the "unborn" by pro-life/anti-abortion advocates) a human being with a right to life, and if so, at what point in the pregnancy does the fetus become human? On the other hand, is a fetus part of a woman's body and

"Abortion" from Wikipedia

File Help

"Abortion" from Wikipedia

COLOR group | individual | text changes | text age

SPACING date | versions

authors | posts

Zundark 1
The Cunctator 1
The Epopt 1
Conversion script 1
RK 1
Fredb 1
B4hand 1
KamikazeArchon 1
Stephen Gilbert 1
Shrubenstein 8
Mkmccann 5
Isis 1
Derek Ross 1
Dante Alighieri 2
Maveric149 3
Jzzbug 2
Jdirl 8
Theanthrope 1
Wesley 2
Dreemword 1
Stevertigo 4
Camembert 1
Hephaestos 2
Zoe 1
MyRedDice 1
G-Man 2
Kingturtle 1
Montrealais 1

198.37.26.168

December 2001

June 2003

Abortion

(Revision as of 22:56 4 Jun 2003)

"**Abortion**," in its most commonly used sense, refers to the deliberate early termination of **pregnancy, resulting in the death of the embryo or fetus**. [1] Medically, the term also refers to the early termination of a pregnancy by natural causes ("spontaneous abortion" or miscarriage, which ends 1 in 5 of all pregnancies, usually within the first 13 weeks) or to the cessation of normal growth of a body part or organ. What follows is a discussion of the issues related to deliberate or "induced" abortion.

Methods

Depending on the stage of pregnancy an abortion is performed by a number of different methods. For the earliest terminations (before nine weeks or so) a chemical abortion is the usual method, the drug mifepristone is usually the only legal method although research has uncovered similar effects from methotrexate and misoprostol. Concurrent with chemical abortion and extending up until around the fifteenth week suction-aspiration or vacuum abortion is the most common approach, replacing the more risky dilation and curettage (D & C). From the fifteenth week up until around the eighteenth week a surgical dilation and evacuation (D & E) is used.

As the fetus size increases other techniques must be used to secure abortion in the third trimester, premature expulsion of the fetus can be induced with prostaglandin, this can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought about by the controversal intact dilation and extraction (D & X) or a hysterotomy abortion, similar to a caesarian section.

The controversy

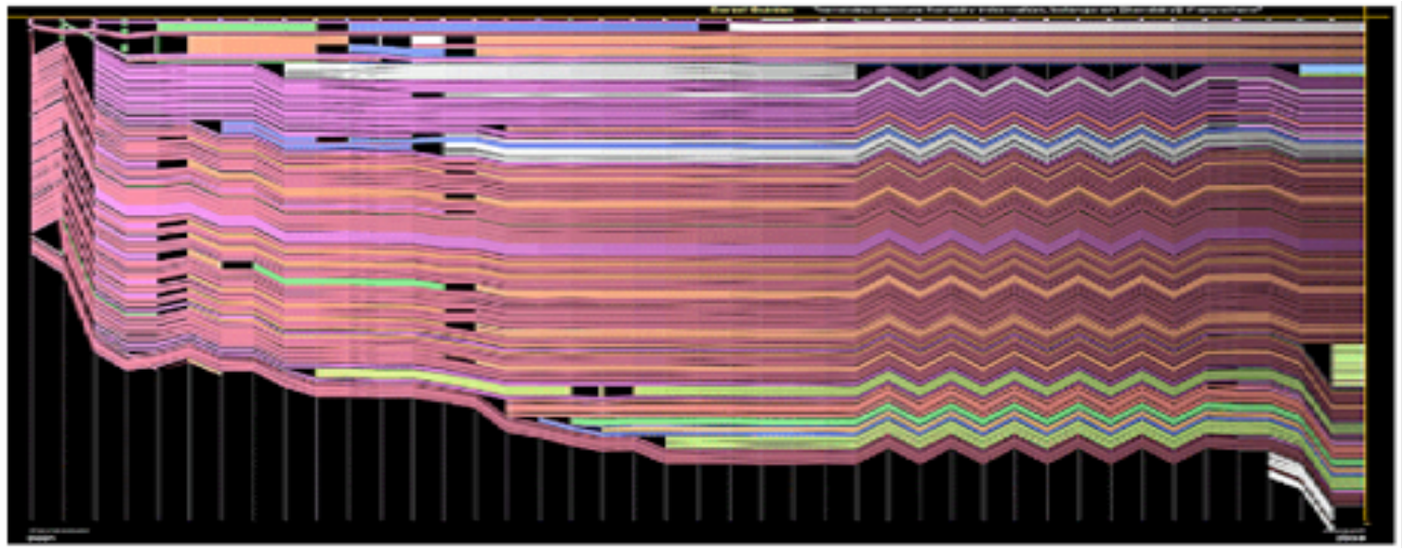
The morality and legality of abortion is a large and important topic in applied ethics and is also discussed by legal scholars and religious people. Important facts about abortion are also researched by sociologists and historians.

Abortion has been common in most societies, although it has often been opposed by some institutionalized religions and governments. In 20th century politics in the United States and Europe, abortion became commonly accepted by the end of the 20th century. Additionally, abortion is legal and accepted in China, India and other populous countries. The Catholic Church remains opposed to the procedure, however, and in other countries, notably the United States and the (predominantly Catholic) Republic of Ireland, the controversy is still extremely active, to the extent that even the names of the respective positions are subject to heated debate. While those on both sides of the argument are generally peaceful, if heated, in their advocacy of their positions, the debate is sometimes characterized by violence. Though true of both sides, this is more marked on the side of those opposed to abortion, because of what they see as the gravity and urgency of their views.

The central question

The central question in the abortion debate is a clash of presumed or perceived rights. On the one hand, is a fetus (sometimes called the "unborn" by pro-life/anti-abortion advocates) a human being with a right to life, and if so, at what point in the pregnancy does the fetus become human? On the other hand, is a fetus part of a woman's body and

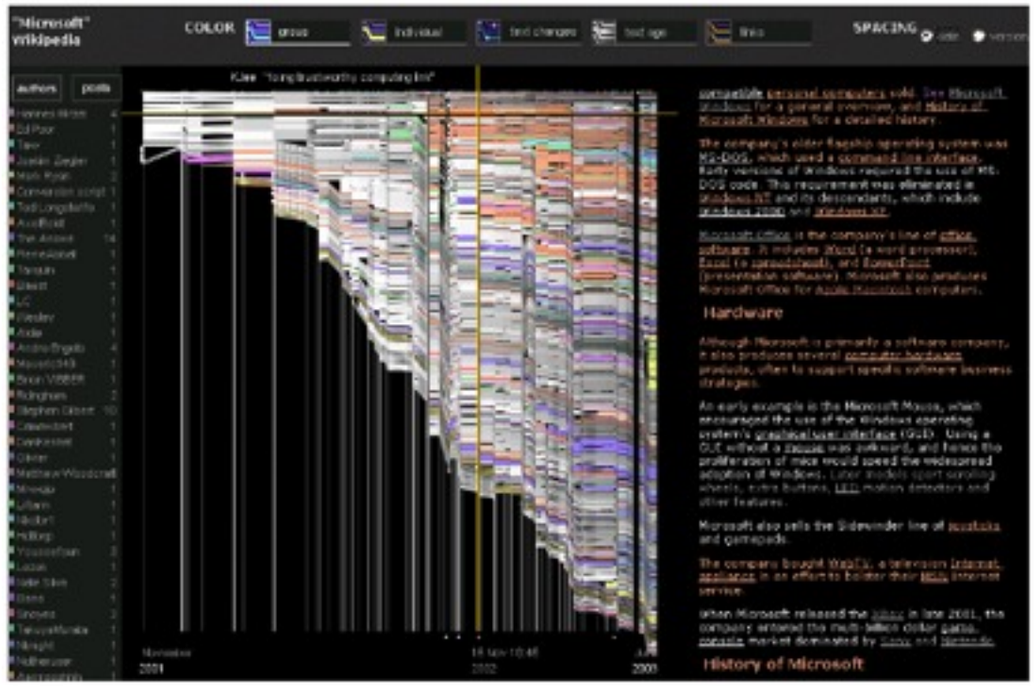
History Flow

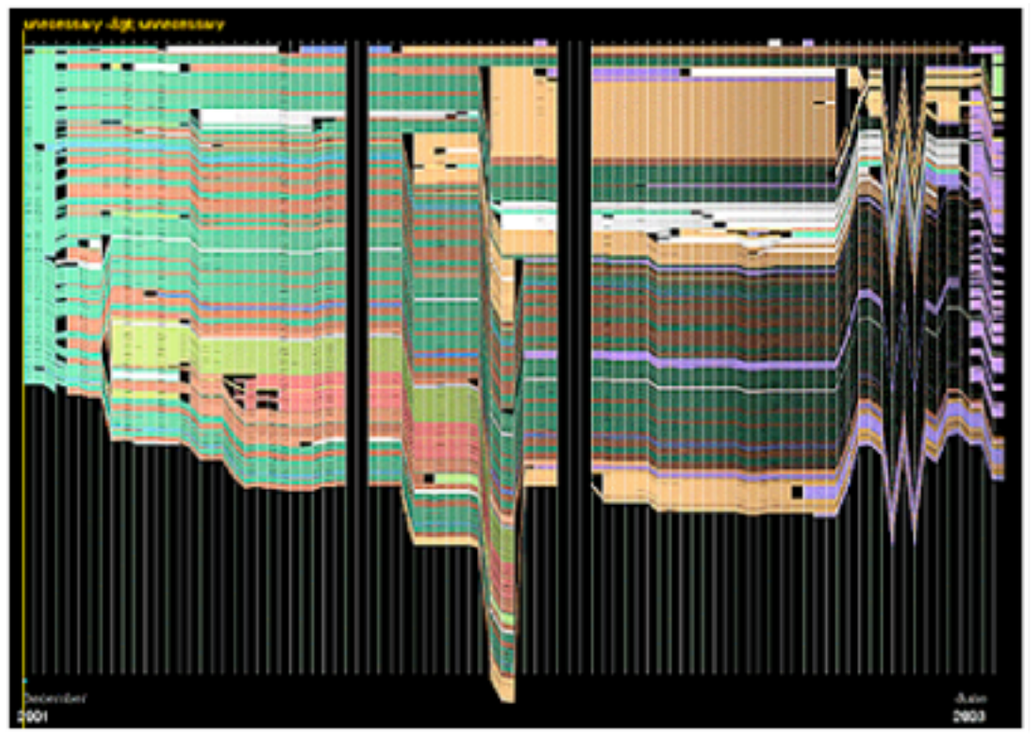


- Patterns in wiki editing
- Vandalism and repair
  - Mass deletion: deletion of all contents on a page
  - Offensive copy: insertion of vulgarities or slurs
  - Phony copy: insertion of text unrelated to the page topic
  - Phony redirection: redirecting to an unrelated / malicious page
  - Idiosyncratic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

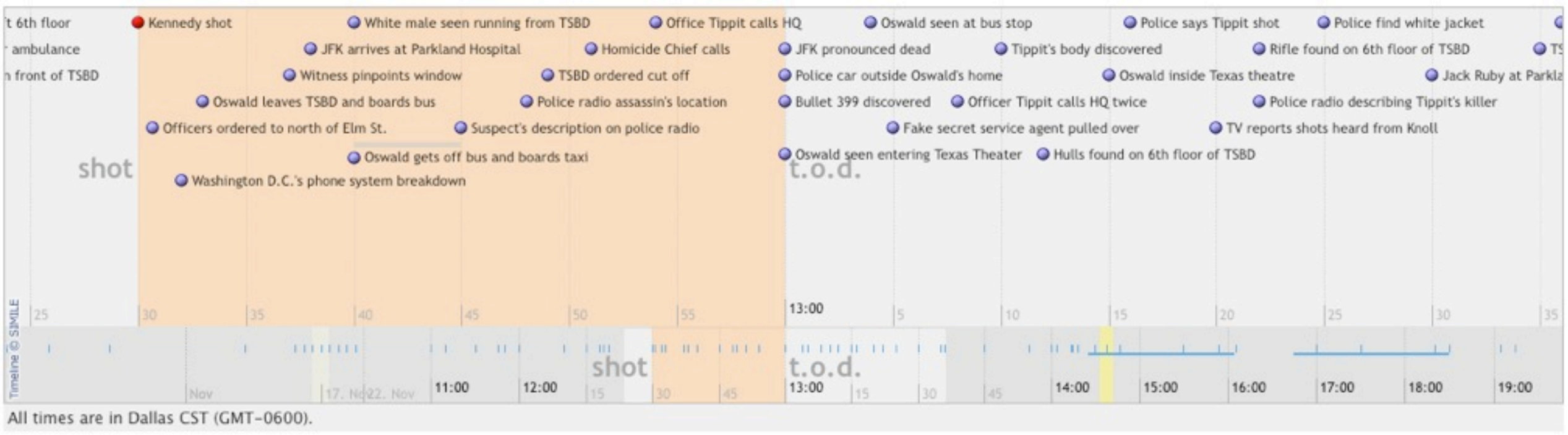
- Temporal patterns observed
- Length of page does not stabilize but change in size over time
  - Microsoft article shows constant growth
  - Article about abortion shows growth and shrinkage (some text was shifted to a separate entry)
- People tend to delete and insert text more frequently than moving 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





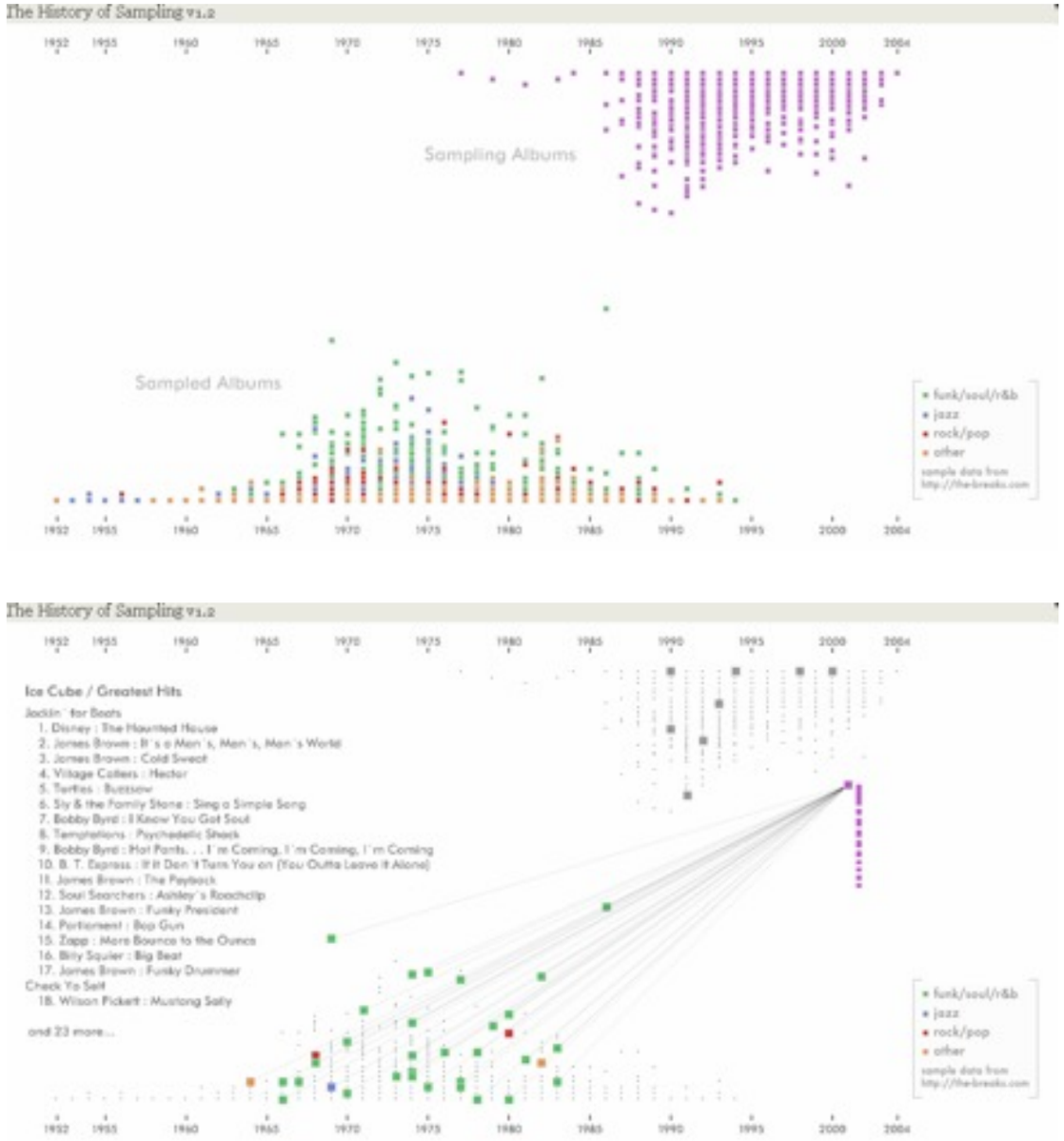
SMILE Timeline

- 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
- http://www.simile-widgets.org/timeline/



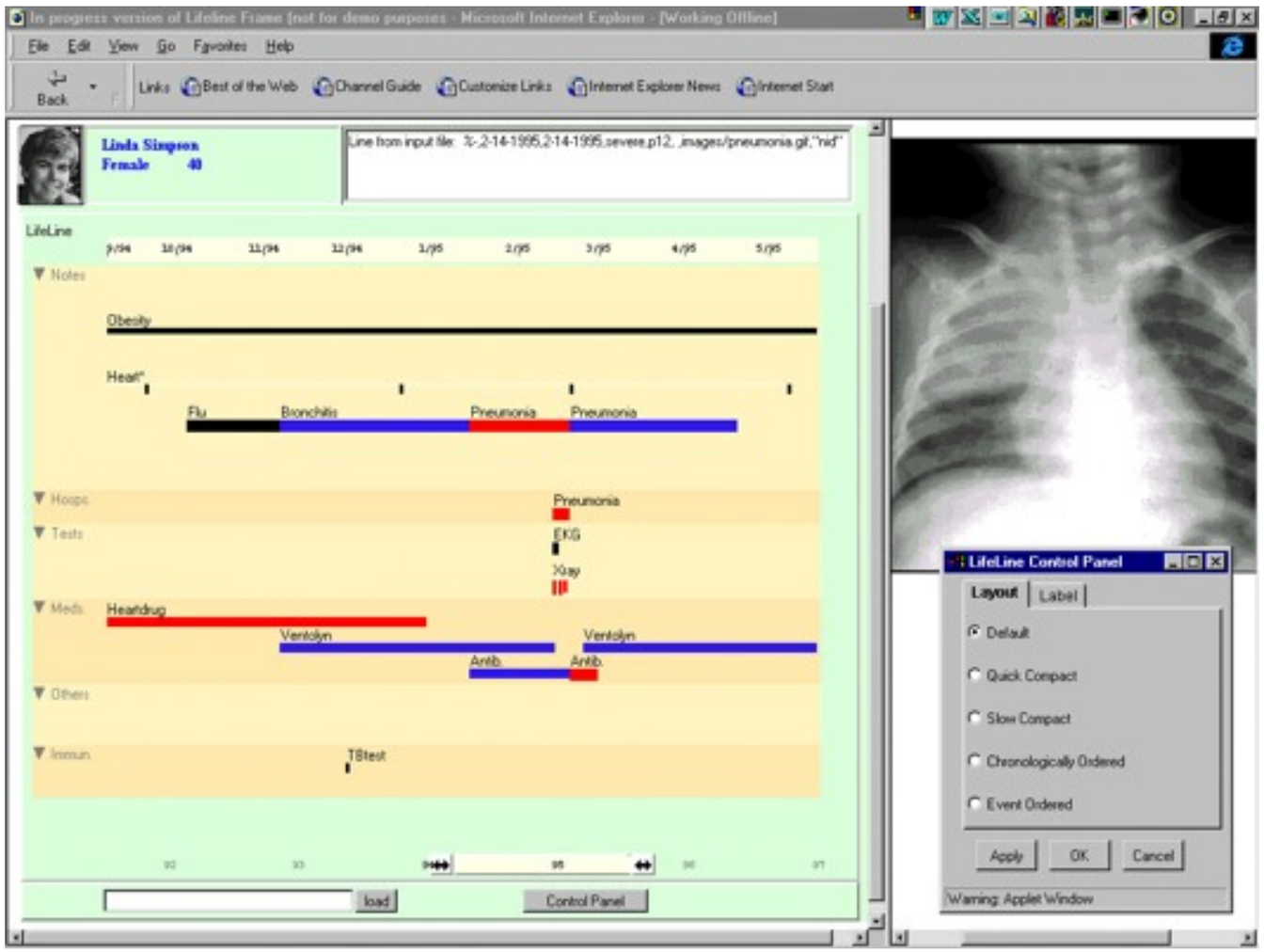
History of Sampling

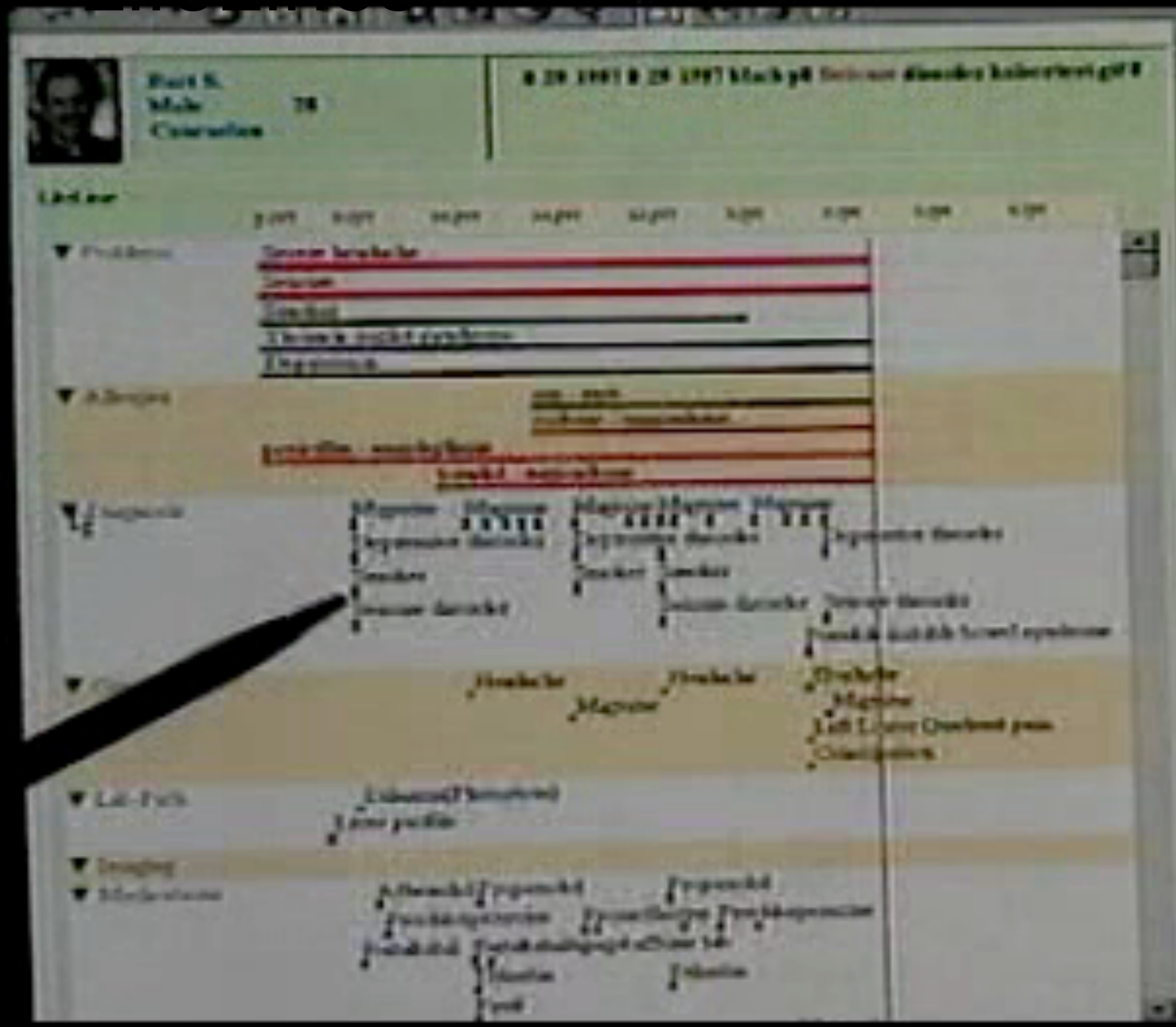
- Each square represents an album
- Color-coding for genres
- Upper half: sampling artist
- Lower half: sampled artists
- X-axis: album release date
- Y-axis: number of samples on / from album – middle represents area of most sampling
- On-demand information (mouse or text-entry with automatic completion)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

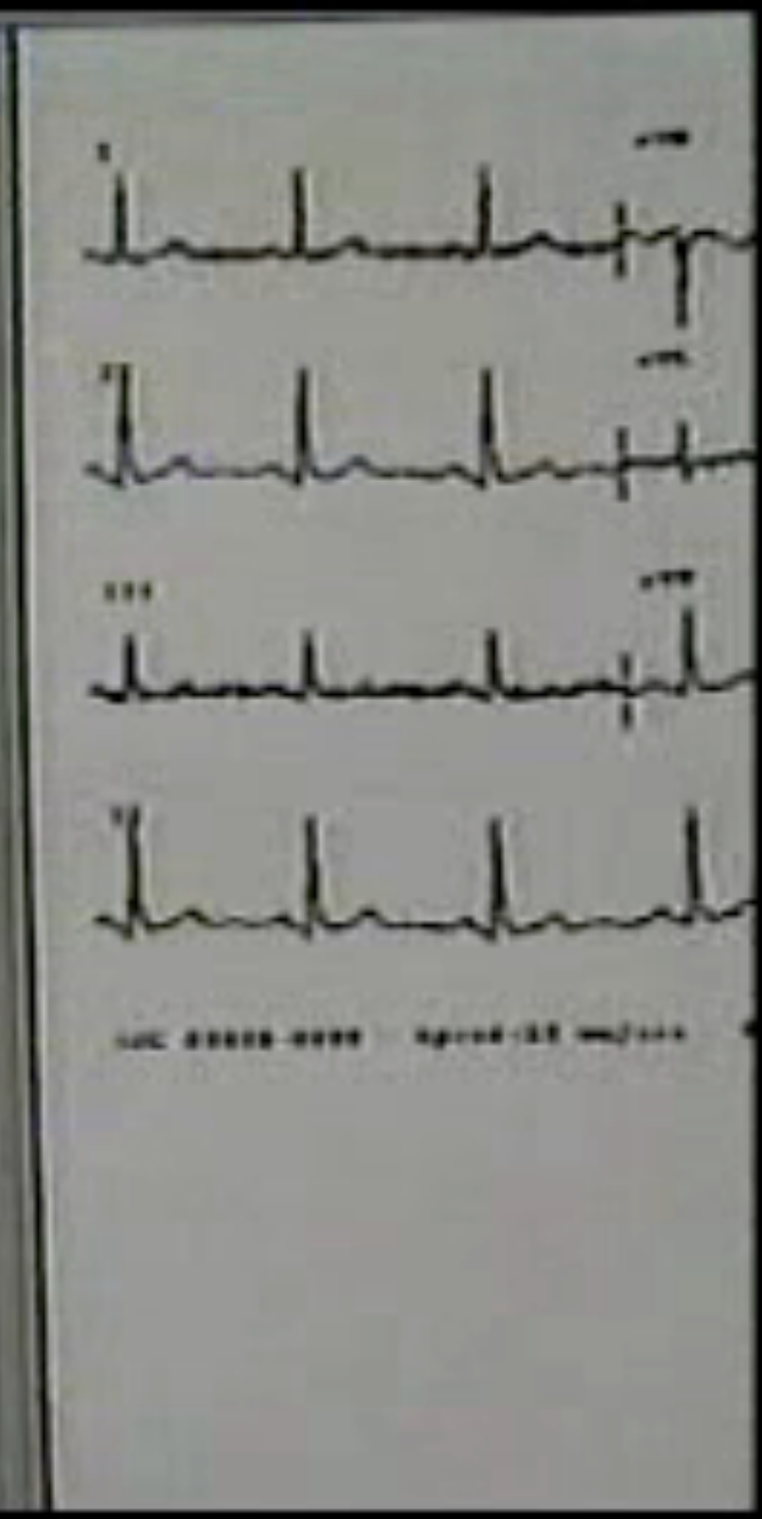


LifeLines

- 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

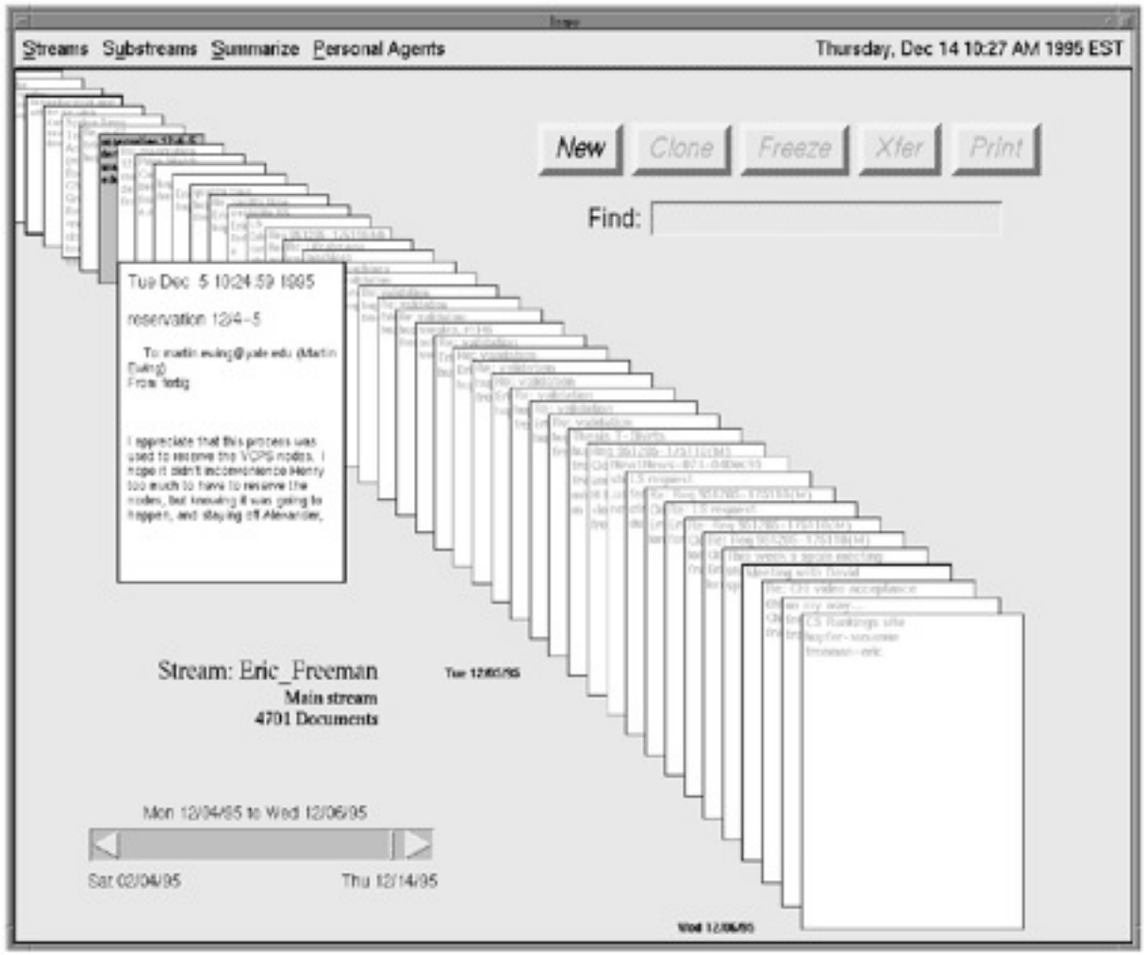


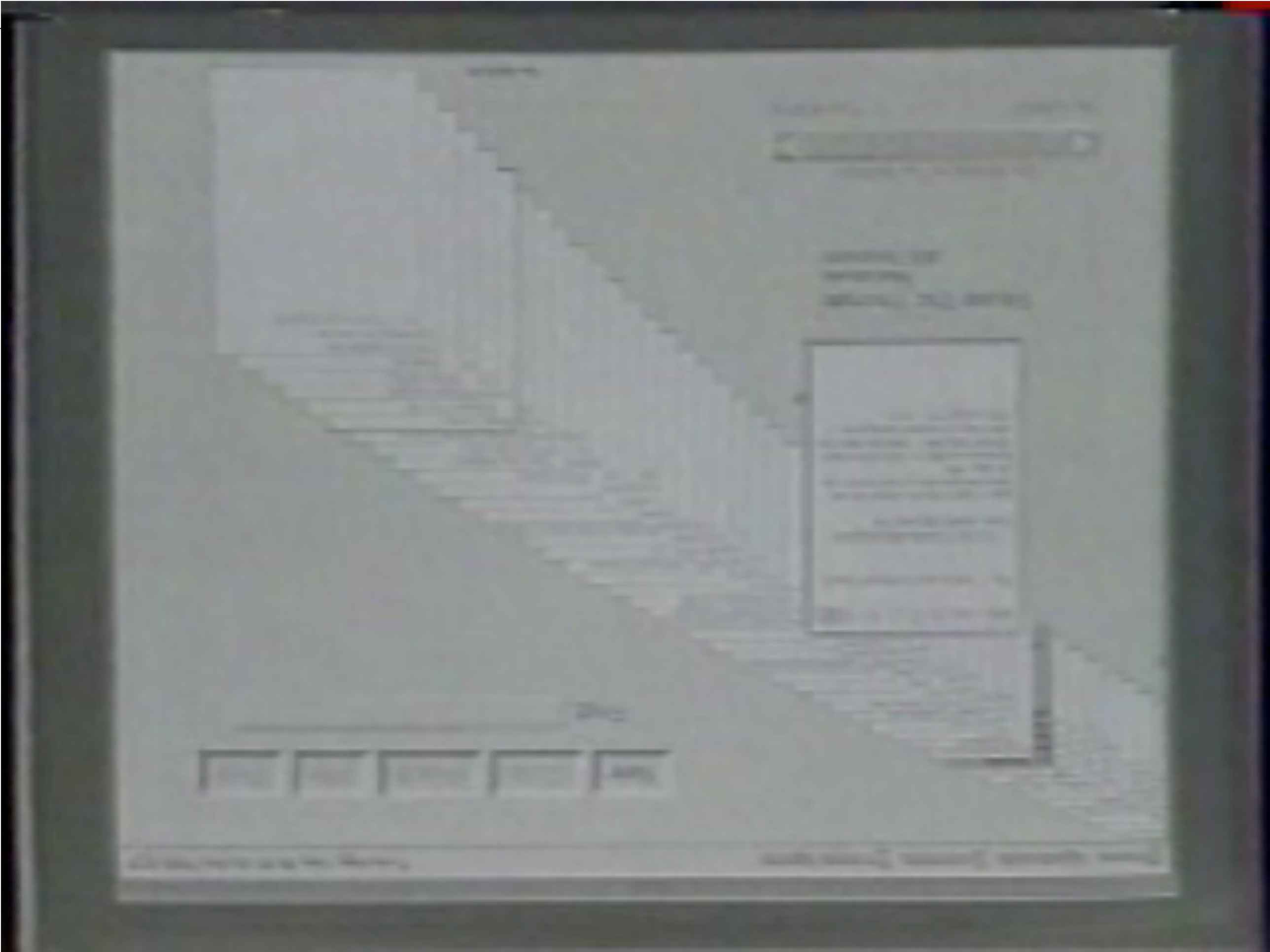




LifeStreams

- 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



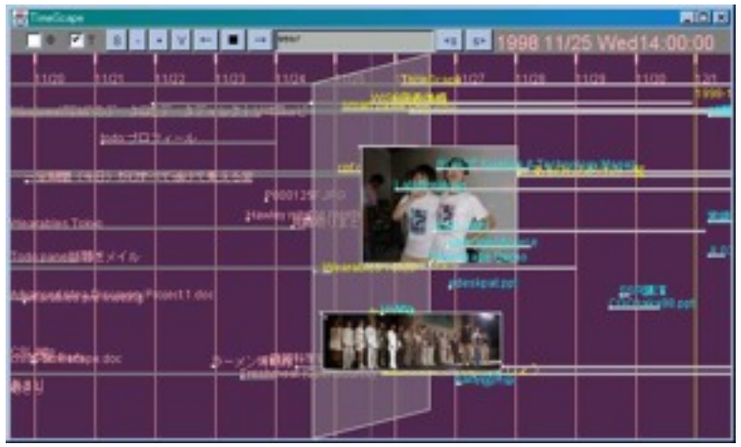


Time Machine

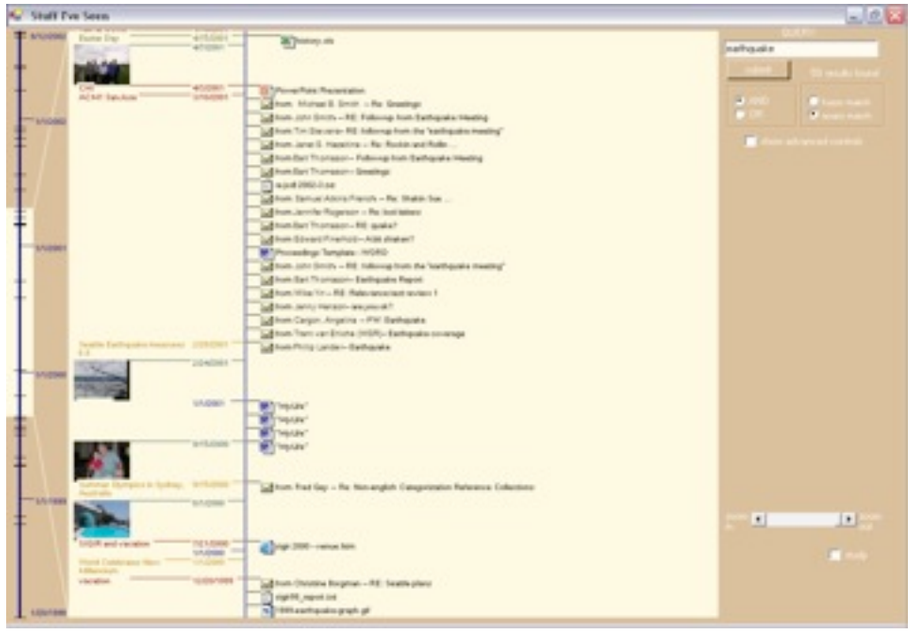
- Hmm, sounds familiar...
- Feature in Mac OS X Leopard, Introduced 2008(?)

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



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

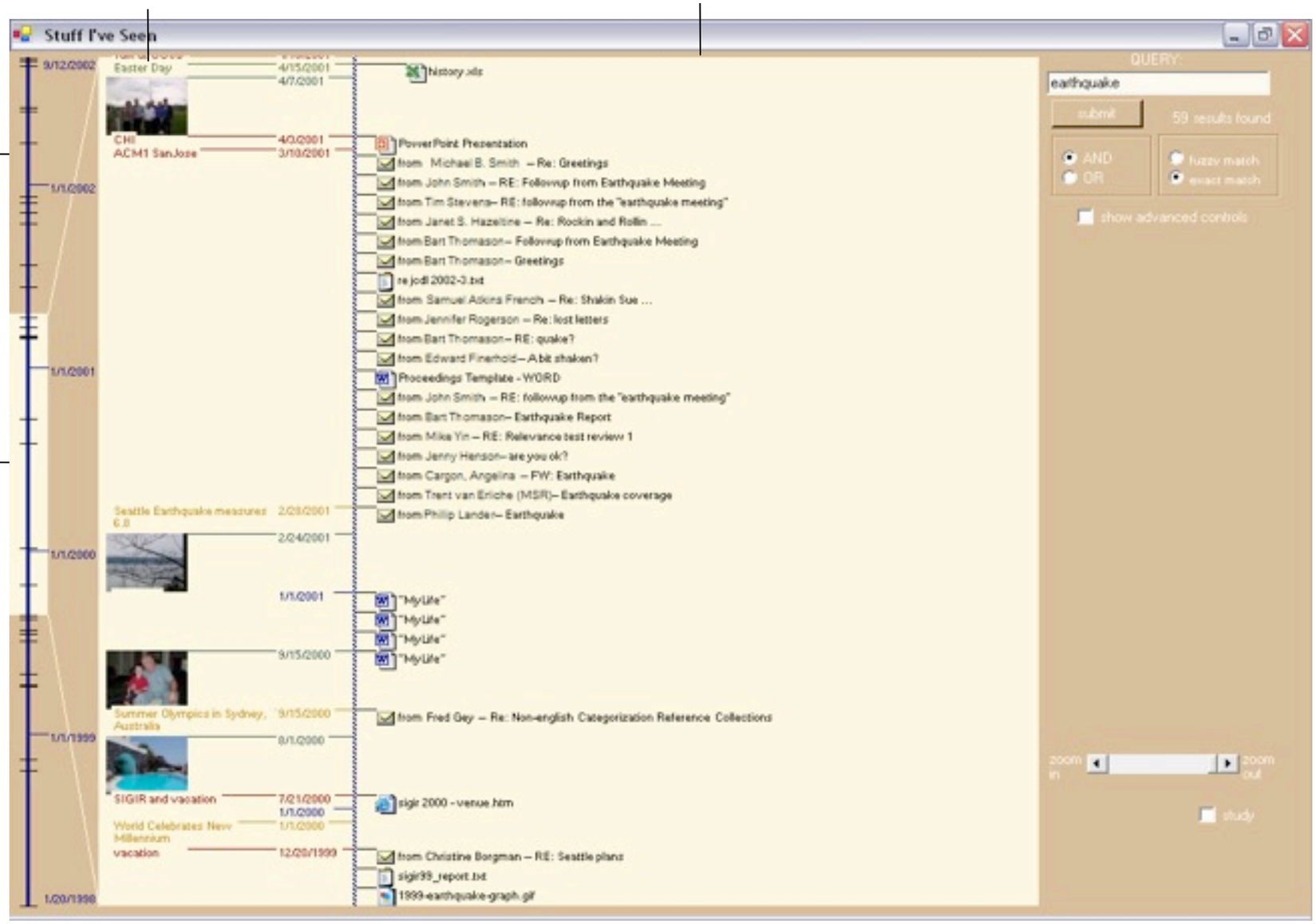
Landmarks for Timeline Visualization

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



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

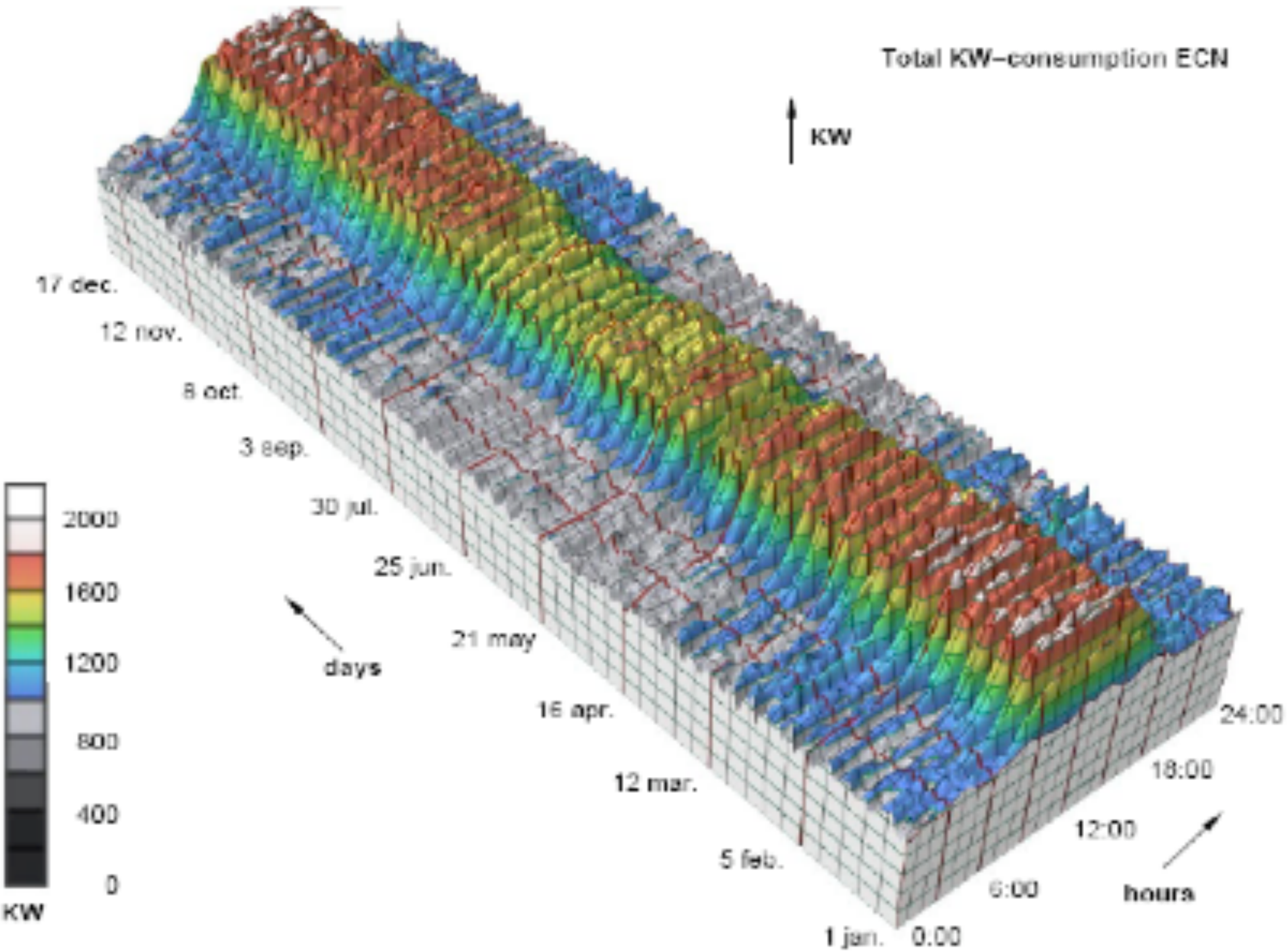


Figure 1. Power demand by ECN, displayed as a function of hours and days

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

Wijk & Selow 1999

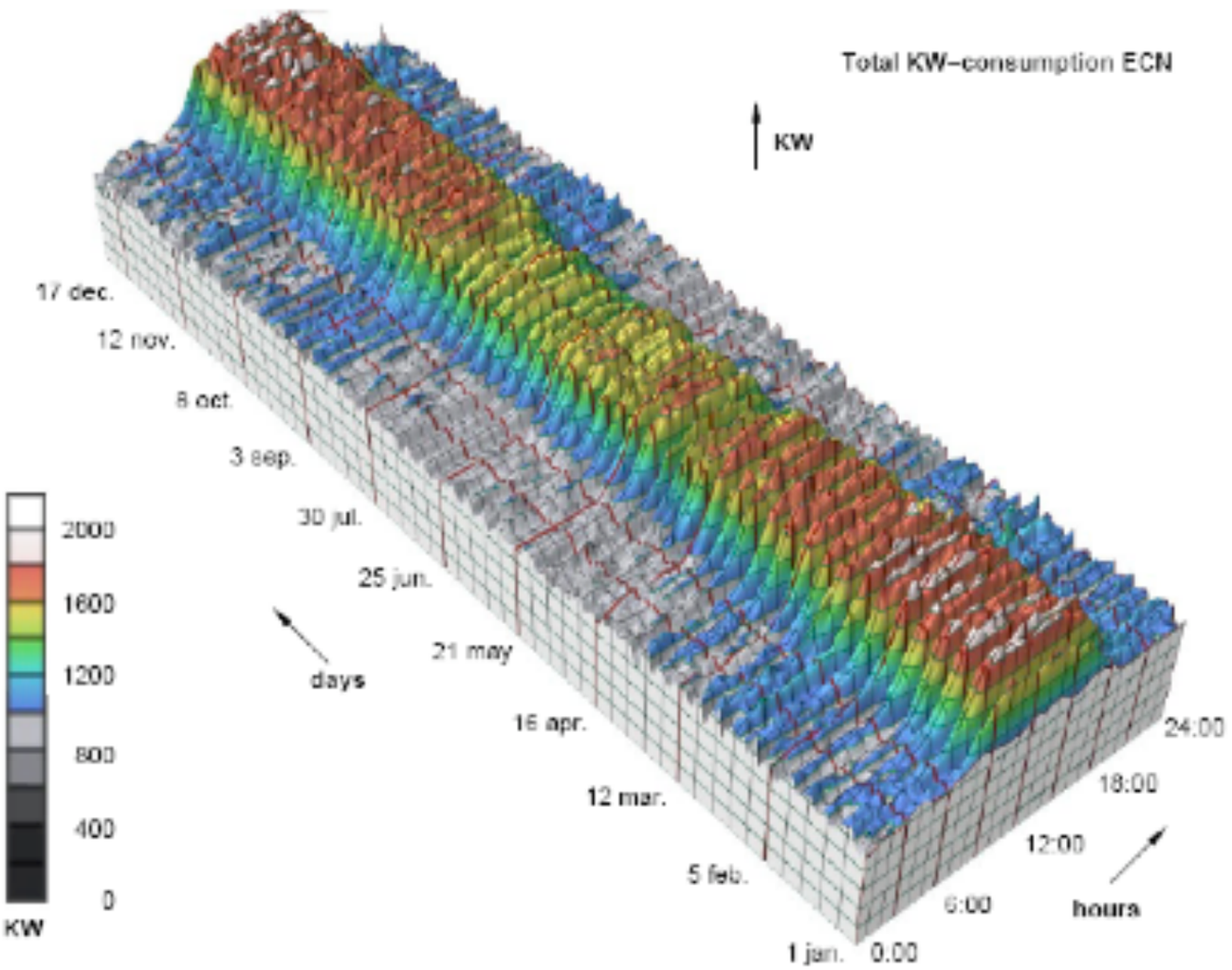


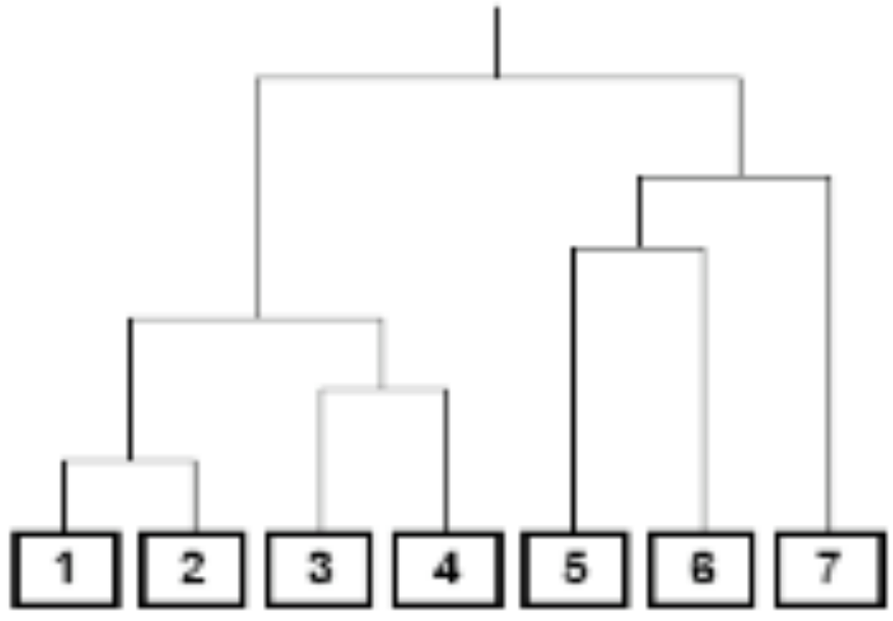
Figure 1. Power demand by ECN, displayed as a function of hours and days

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,…,M
- Each Y_j consists of seq. of pairs (y_i,t_i),i = 1,…,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 most similar clusters to 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 values in the 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)

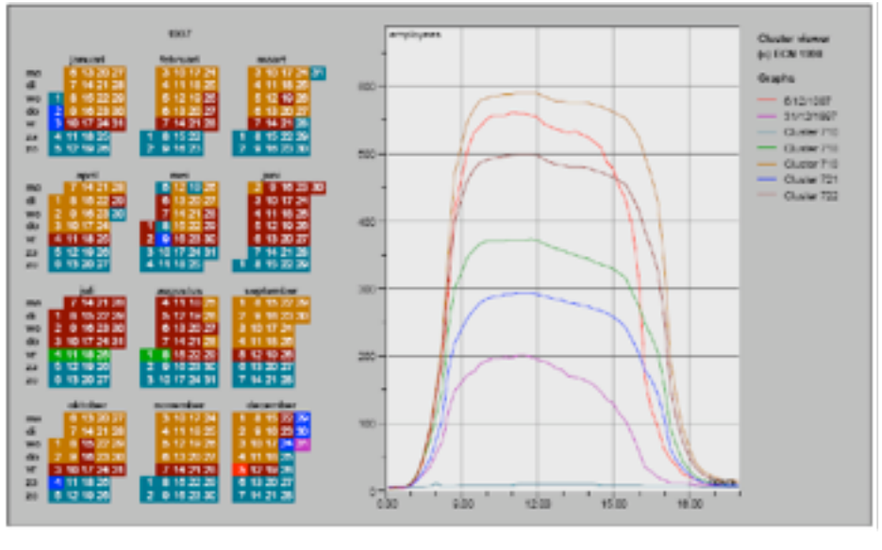
$$d_{nm} = \sqrt{\sum (y_i / y_{\max} - z_i / z_{\max})^2 / N}.$$

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



Wijk & Selow 1999

- 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

Detect Patterns in Time-Series

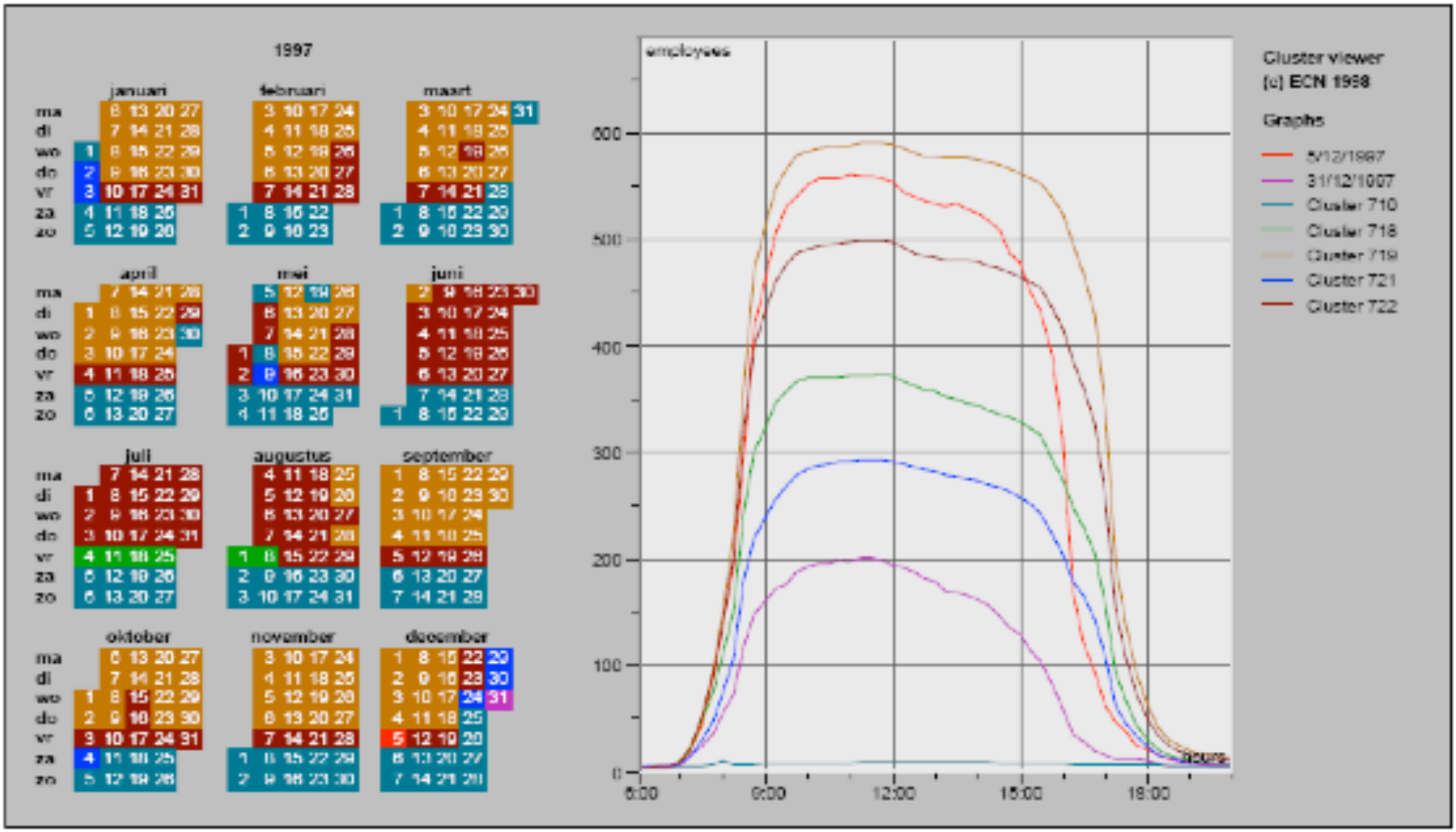


Figure 4. Calendar view of the number of employees

Wijk & Selow 1999

Detect Patterns in Time-Series

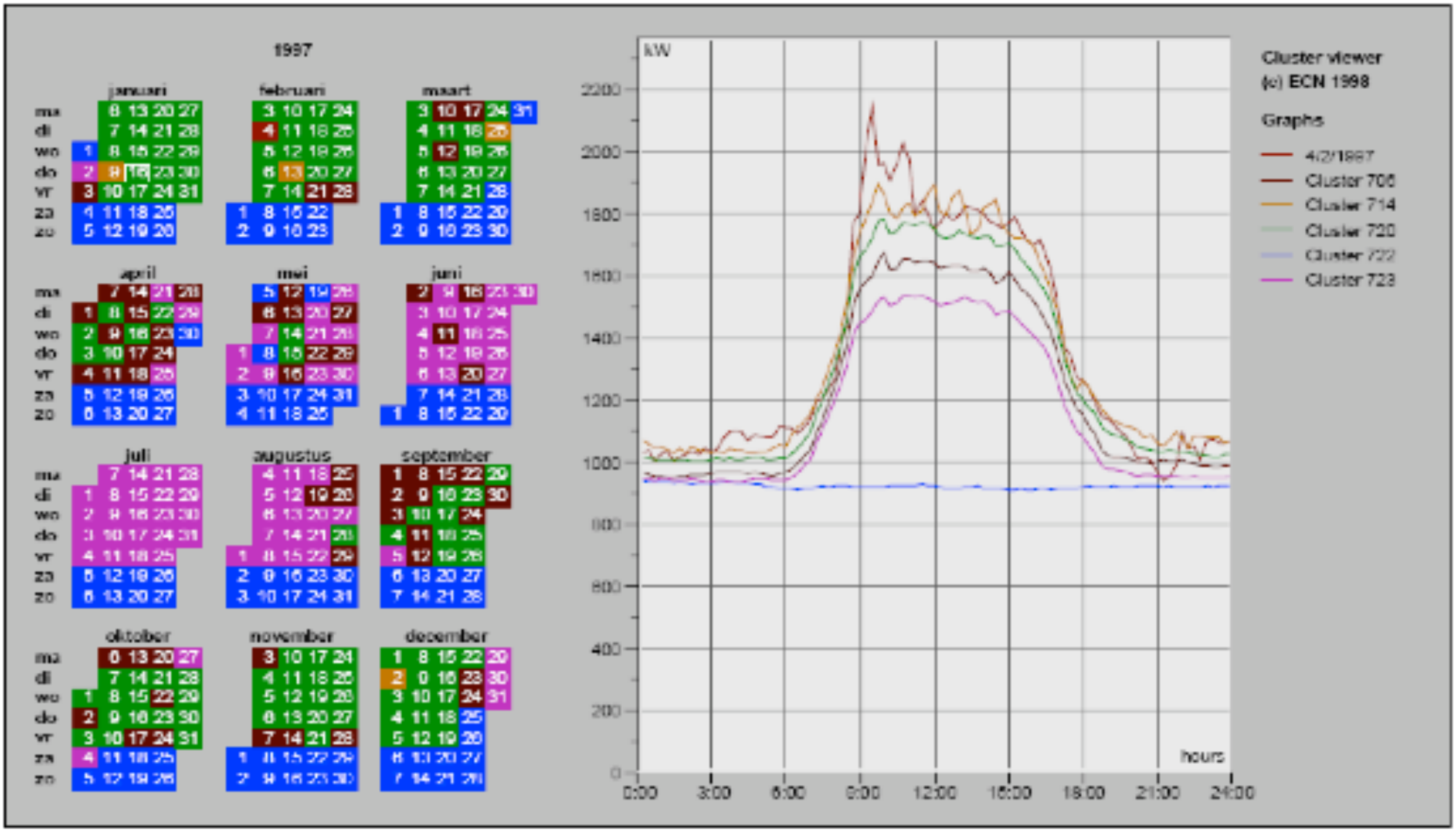
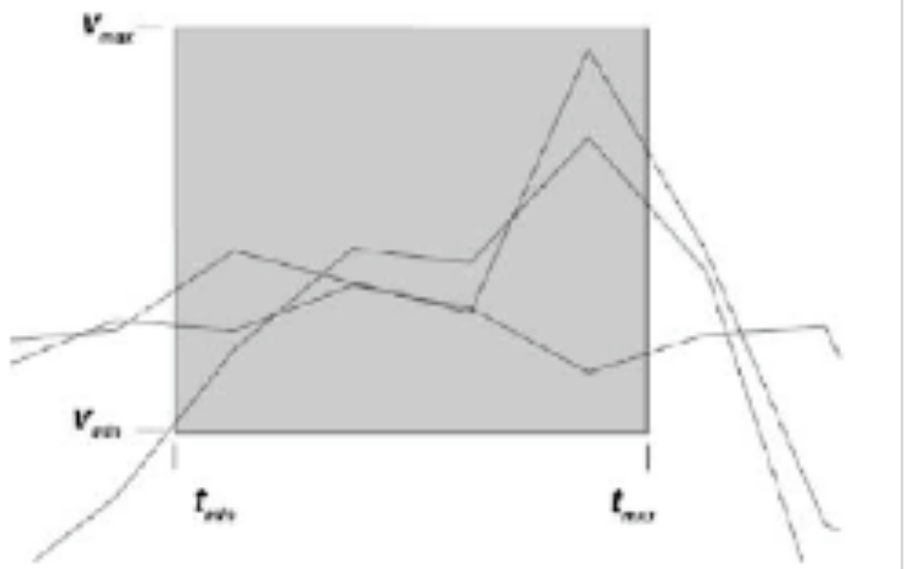


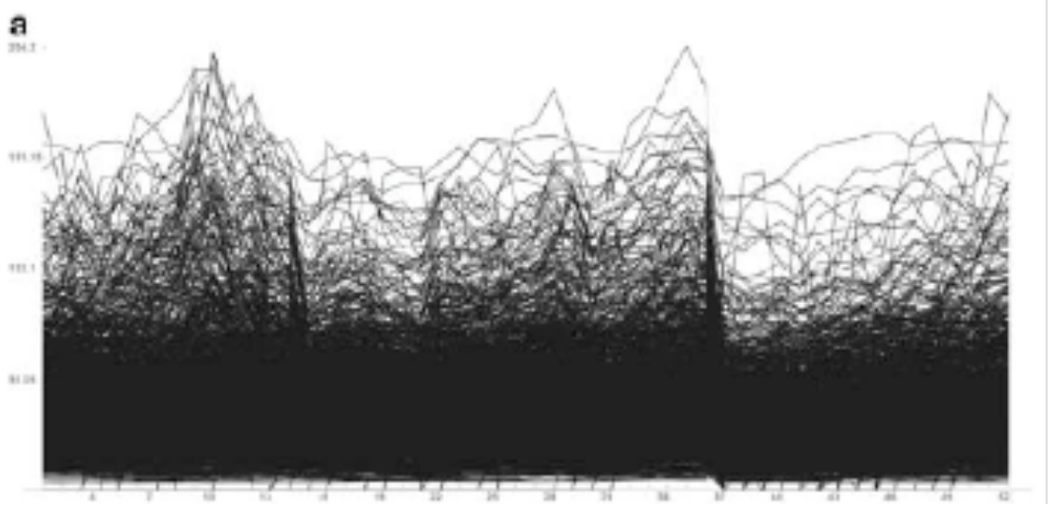
Figure 5. Cluster analysis of power demand by ECN

Wijk & Selow 1999

TimeSearcher

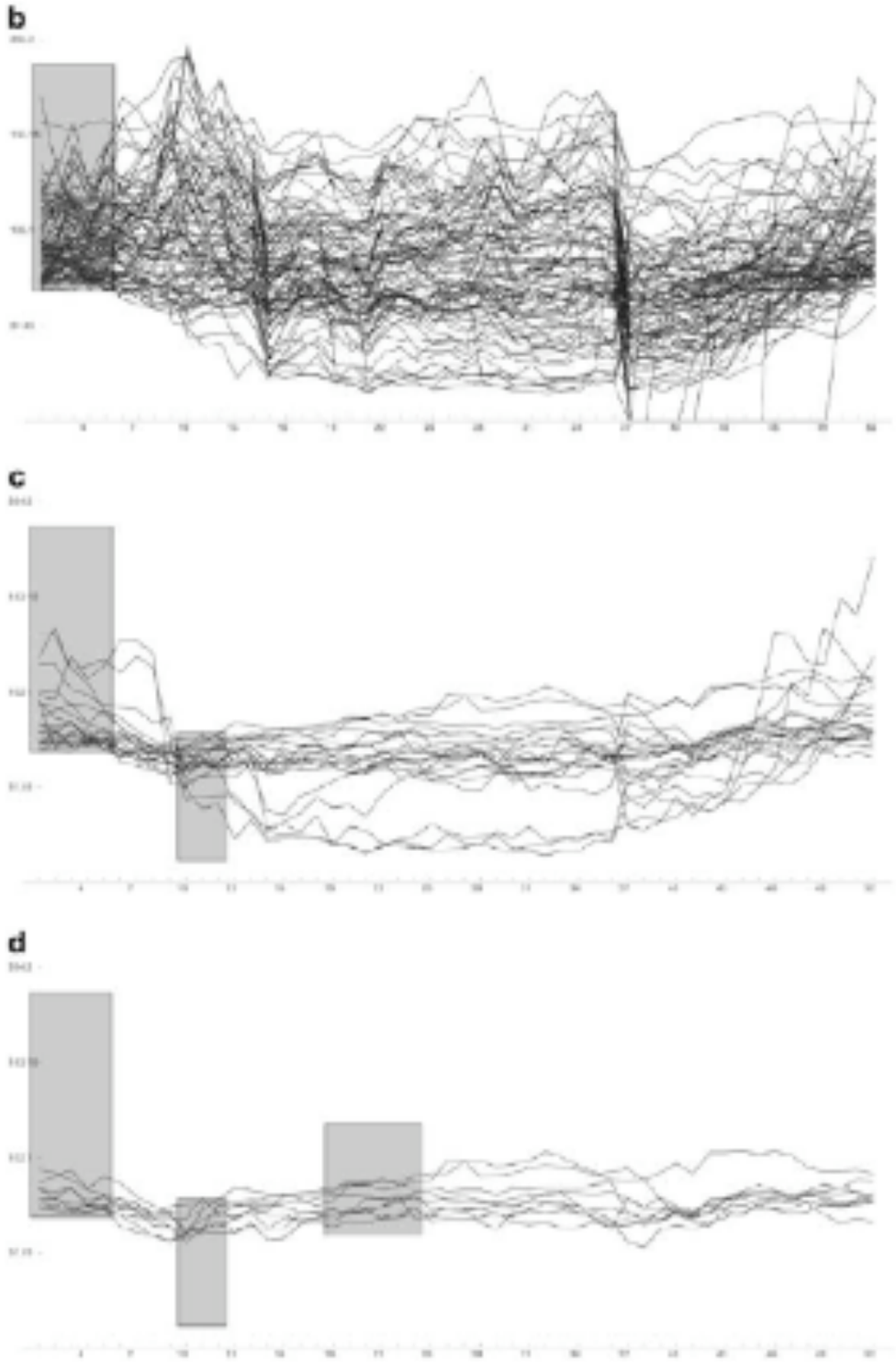
- Hochheiser & Shneiderman 2004
- Dynamic query tools for searching time series for trends and patterns
- Timebox: rectangular query regions (bounding box) drawn on a 2D representation of time series data
  - X-axis extent: time period of interest
  - Y-axis extent: constraint on the range of values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





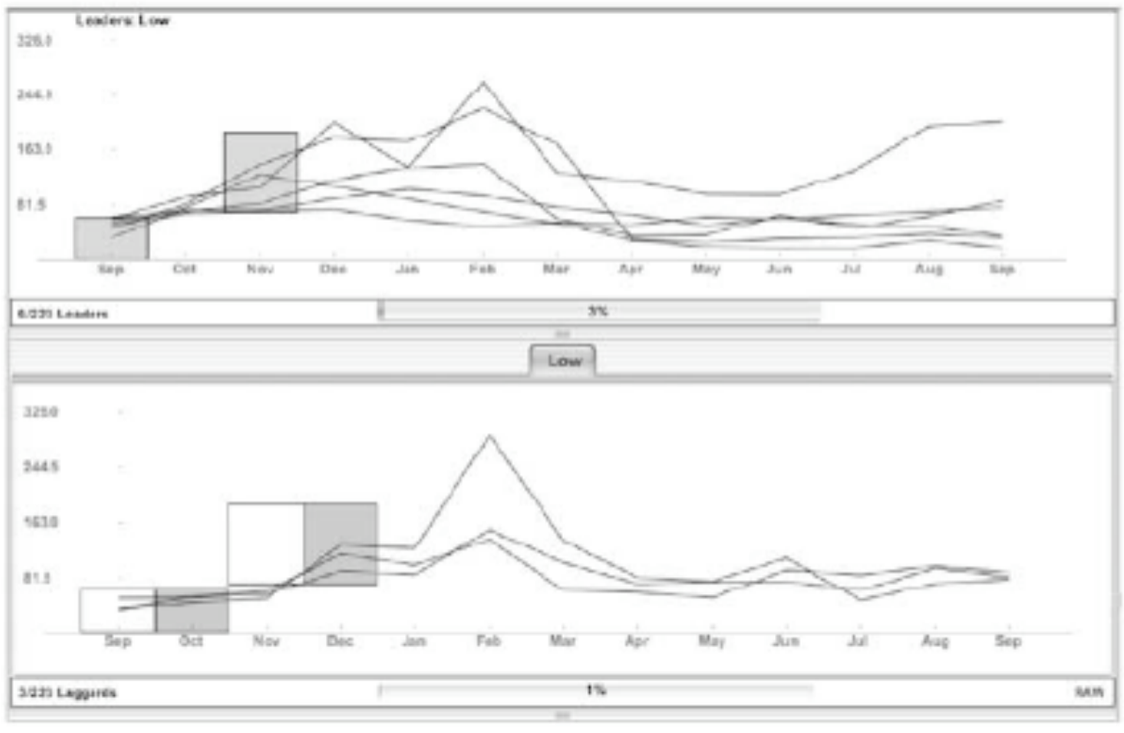
TimeSearcher

- 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



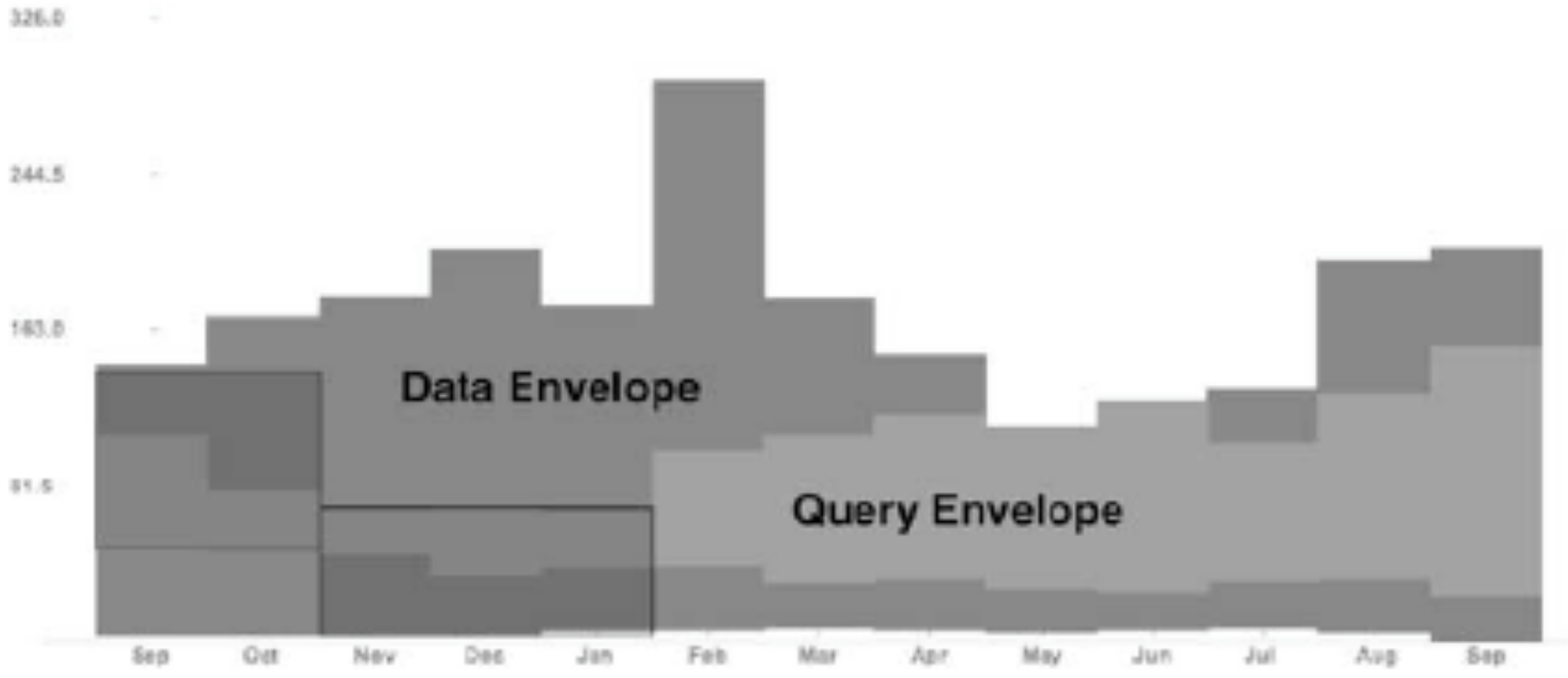
TimeSearcher

- 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



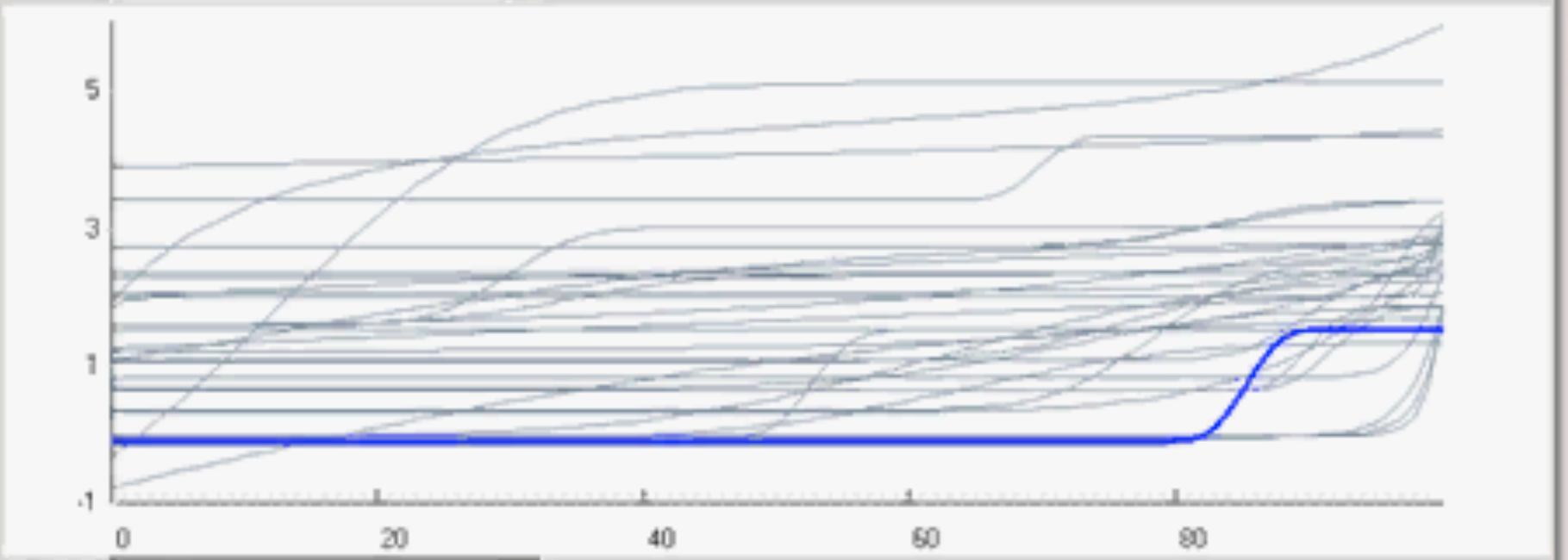
TimeSearcher 2.2 magazine50b.tqd

File Edit View Actions Tools Help

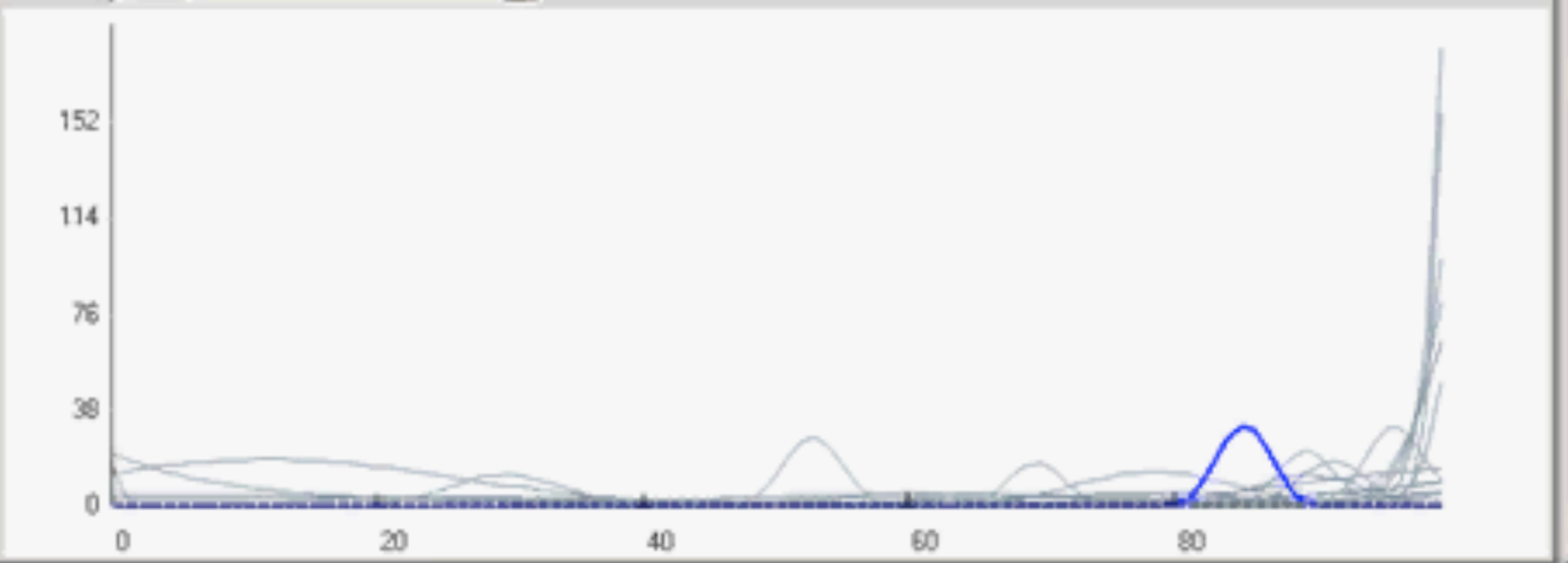
Number Of Variables 2 of 3

Variables | Single Auctions | River Plot

Variable: Price



Variable: Velocity



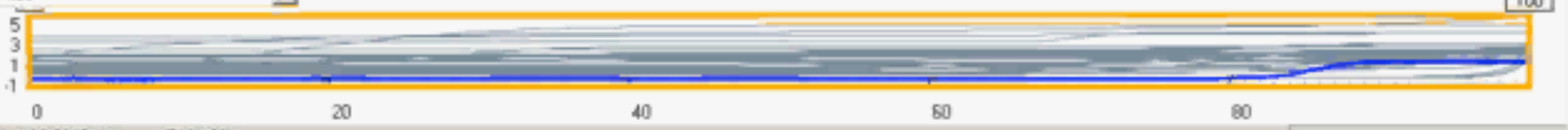
Time (1... | Price | Velocity | Acceler...

Items | Attribute Statistics

AuctionID (34/34)	End Day	Duration	No of bids
2494249392	Tuesday	7	11
2495274258	Wednesday	3	10
2495280562	Wednesday	3	5
2494662462	Thursday	7	29
2494771149	Thursday	7	8
2494916056	Friday	7	14
2495266391	Friday	5	5
2495277300	Friday	5	8
2495155447	Saturday	7	5
2495158450	Saturday	7	5
2495159407	Saturday	7	6
2495159443	Saturday	7	4
2495163132	Saturday	7	5
2495171689	Saturday	7	7
2495177180	Saturday	7	13
2495189479	Saturday	7	5
2495192458	Saturday	7	5
2495192773	Saturday	7	6
2494865057	Sunday	7	4
2494868219	Sunday	7	4
2494870487	Sunday	7	4
2495228215	Sunday	7	4
2495231854	Sunday	7	4
2495253247	Sunday	7	4
2495271486	Sunday	7	5
2495276028	Sunday	7	5
2495276761	Sunday	7	5
2495276905	Sunday	7	12
2495278146	Sunday	7	6
2495278940	Sunday	7	4
2495279983	Sunday	7	6
2495280344	Sunday	7	4
2495283294	Sunday	7	4
2495284428	Sunday	7	5

Price

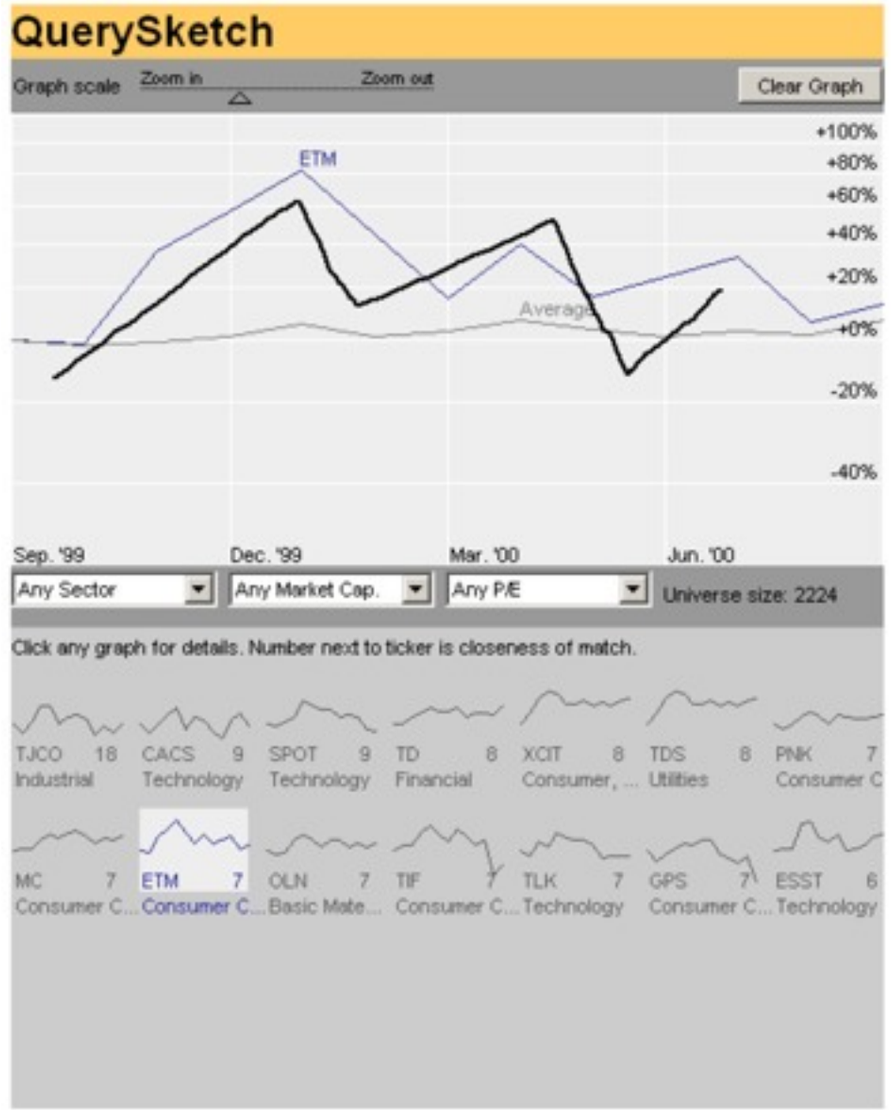
100



X: Y:69.3243 - Mouse: (710,166)

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



Sources and Literature

- Recommended literature
  - 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