Visualization

Visualizing Time

Overview

- Introduction
- Time-series Data
- Applications

Overview

- Introduction
- Time-series Data
- Applications

Time-oriented Data

- Nowadays, time-oriented data are ubiquitous in many application domains as, for example, in business, medicine, history, planning, or project. management.

- Fundamental chronological component to the data set

75 % of 4000 samples of graphics from newspapers and magazines (‘74-’80) were time-series data!

Tufte Vol. 1

Data Sets

- Each data case is likely an event of some kind
- One of the variables can be the date and time of the event
- Examples:
  - Office meetings
  - Cricket matches
  - Medicines taken
  - Cities visited
  - Stock prices
- Therefore there are many ways of visualizing time

Data Mining

- Data mining domain has techniques for algorithmically examining time series data, looking for patterns, etc.

- Good when objective is known a priori
- But what if not?
  - Which questions should I be asking?
  - InfoVis better for that

Tasks

- What kinds of questions do people ask about temporal data?
  - Is there an order that things occur in?
  - How long do events last?
  - Is there a cycle?
  - How does an event change over time?
  - How often does an event happen? When was something greatest/least?
  - Is there a pattern?
  - Are two series similar?
  - Do any of the series match a pattern?
  - Provide simpler, faster access to the series
  - 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?

Taxonomy – Temporal Data

- Discrete points vs. interval points
- Linear time vs. cyclic time
- Ordinal time vs. continuous time
- Ordered time vs. branching time vs. time with multiple perspectives

Muller & Schumann '03

Taxonomy – Temporal Data

- Ordered time domains consider things that happen one after the other.
- For branching time, multiple strands of time branch out
  - Facilitates description and comparison of alternative scenarios
  - For example, for project planning
  - This type of time supports decision-making processes, where only one alternative will actually happen.
- Time with multiple perspectives allows more than one point of view at observed facts
  - For example, eye-witness reports

Taxonomy – Temporal Data

- Continuous
  - Non-periodic
  - Periodic
- Discrete
  - Non-periodic (non-recurring)
  - Periodic

Continuous Time - Example

- Stock market prices
- Heart rate
- Daily temperature
- Height of ocean

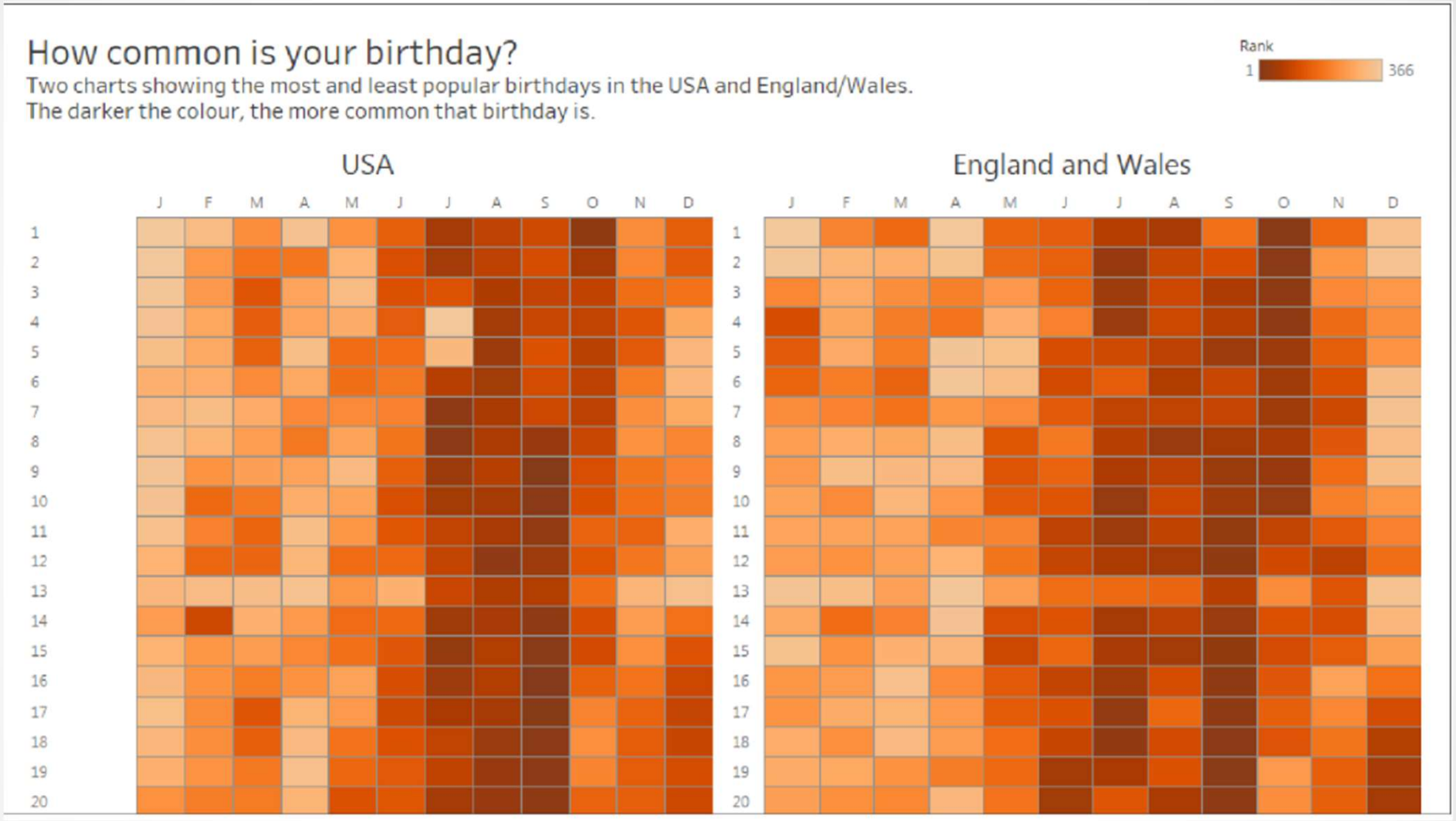
Discrete Time - Example

- Birth/marriage/death
- Historical events
- Train departure

- Some discrete events have duration:
  - Meeting or appointment
  - Hospital stay
  - Project phase

- Some discrete events are periodic
  - US Presidential elections – every four years

Time in 2D


How common is your birthday?
Two charts showing the most and least popular birthdays in the USA and England/Wales.
The darker the colour, the more common that birthday is.
Rank
1
366
USA
England and Wales
J F M A M J J A S O N D
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

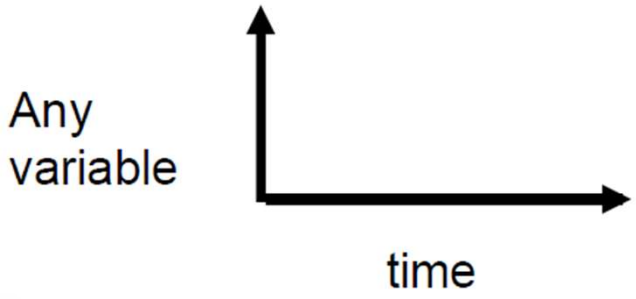
Time using Parallel Coordinates

Overview

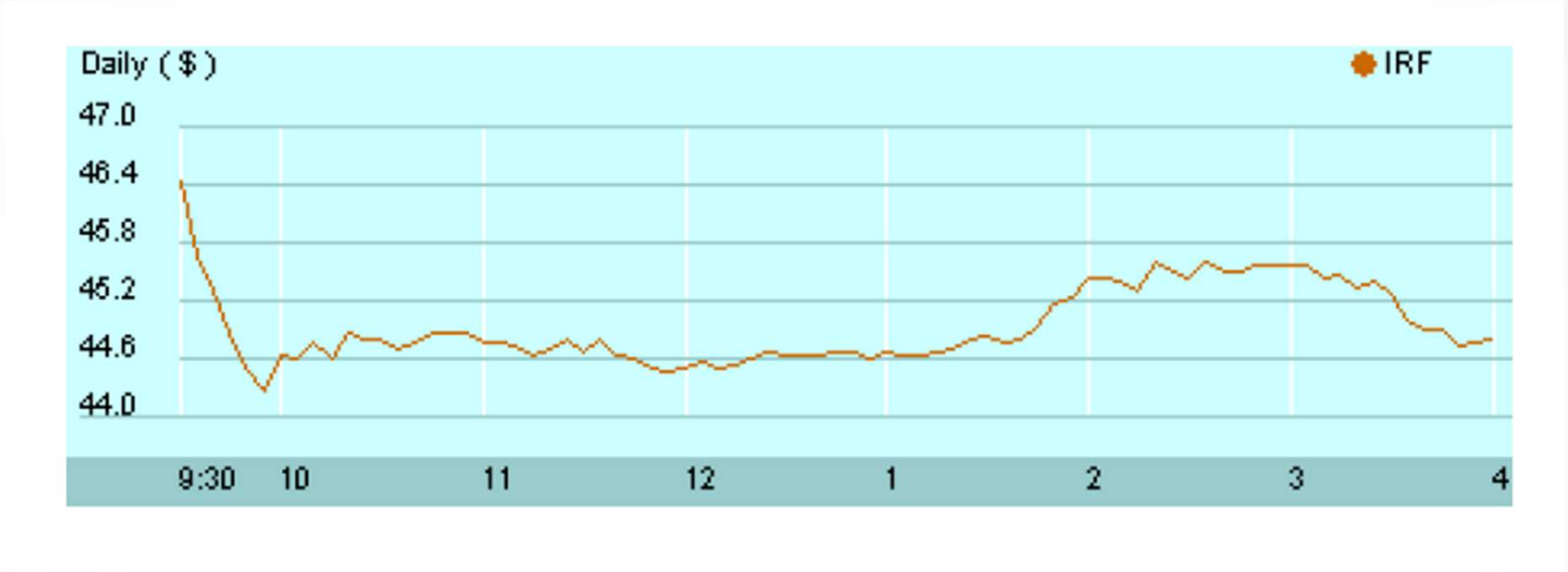
- Introduction
- Time-series Data
- Applications

Time-series data

- Time series data is a sequence of data points indexed in time order.
- These data points typically consist of successive measurements made from the same source over a fixed time interval and are used to track change over time.
- Represents continuous time
- Two variables – a value, time

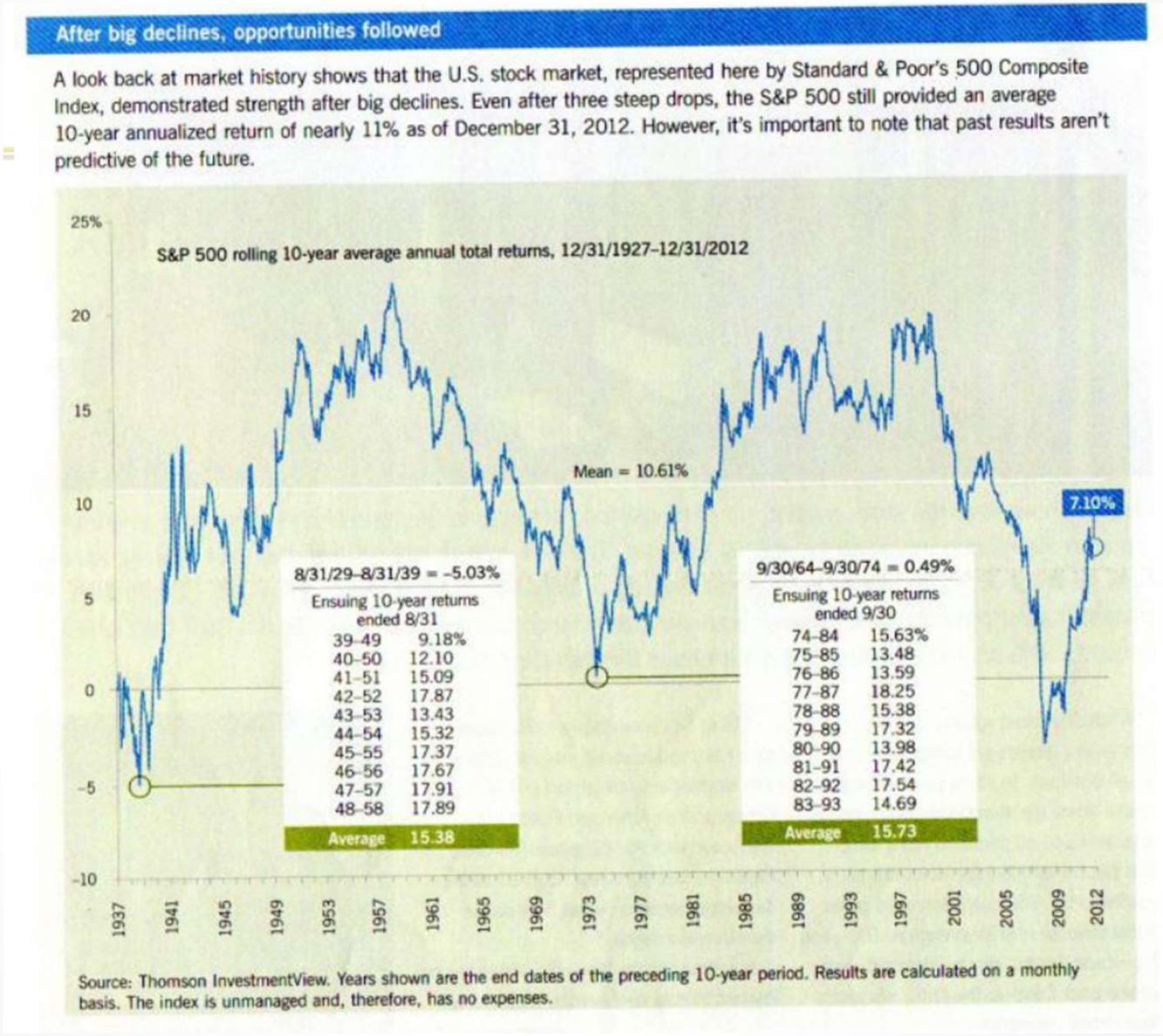


Line Graph



- One of the most common way of representing continuous time/time-series data

Example


After big declines, opportunities followed
A look back at market history shows that the U.S. stock market, represented here by Standard & Poor's 500 Composite Index, demonstrated strength after big declines. Even after three steep drops, the S&P 500 still provided an average 10-year annualized return of nearly 11% as of December 31, 2012. However, it's important to note that past results aren't predictive of the future.
S&P 500 rolling 10-year average annual total returns, 12/31/1927–12/31/2012
25%
20
15
10
5
0
-5
-10
Mean = 10.61%
7.10%
8/31/29–8/31/39 = –5.03%
Ensuing 10-year returns ended 8/31
39–49 9.18%
40–50 12.10
41–51 15.09
42–52 17.87
43–53 13.43
44–54 15.32
45–55 17.37
46–56 17.67
47–57 17.91
48–58 17.89
Average 15.38
9/30/64–9/30/74 = 0.49%
Ensuing 10-year returns ended 9/30
74–84 15.63%
75–85 13.48
76–86 13.59
77–87 18.25
78–88 15.38
79–89 17.32
80–90 13.98
81–91 17.42
82–92 17.54
83–93 14.69
Average 15.73
1937
1941
1945
1949
1953
1957
1961
1965
1969
1973
1977
1981
1985
1989
1993
1997
2001
2005
2009
2012
Source: Thomson InvestmentView. Years shown are the end dates of the preceding 10-year period. Results are calculated on a monthly basis. The index is unmanaged and, therefore, has no expenses.

Example


Compare: Enter ticker here Add Dow Jones S&P 500 ORCL more »
Zoom: 1d 5d 1m 3m 6m YTD 1y 5y 10y All
Jan 20, 2011 15:36 Price: 155.55 Vol: 46.25k
Discrete time events, no duration
159
158
157
156
155
F E D C B A
Thu Jan 20 12 pm 2 pm Fri Jan 21 12 pm 2 pm Mon Jan 24 12 pm 2 pm
Volume (thous / 2min)
450
225
2009 2010
Settings | Plot feeds | Technicals | Link to this view
Volume delayed by 15 mins.
A Nationally Known Business Author Profiles Michigan Company
Trading Markets - 30 minutes ago
B Earnings Preview: EMC Corp.
Daily Markets - 1 hour ago
C Bank of America - Worst Performing Dow Component
TradersHuddle.com - 1 hour ago
D Computer Point takes on IBM firms in Uganda
East African Business Week - 2 hours ago
E IBM rolls out virtual desktop offering
All news for International Business Machines Corp. »
Subscribe
Advertisement
FREE PENNY STOCK ALERTS

Sparklines

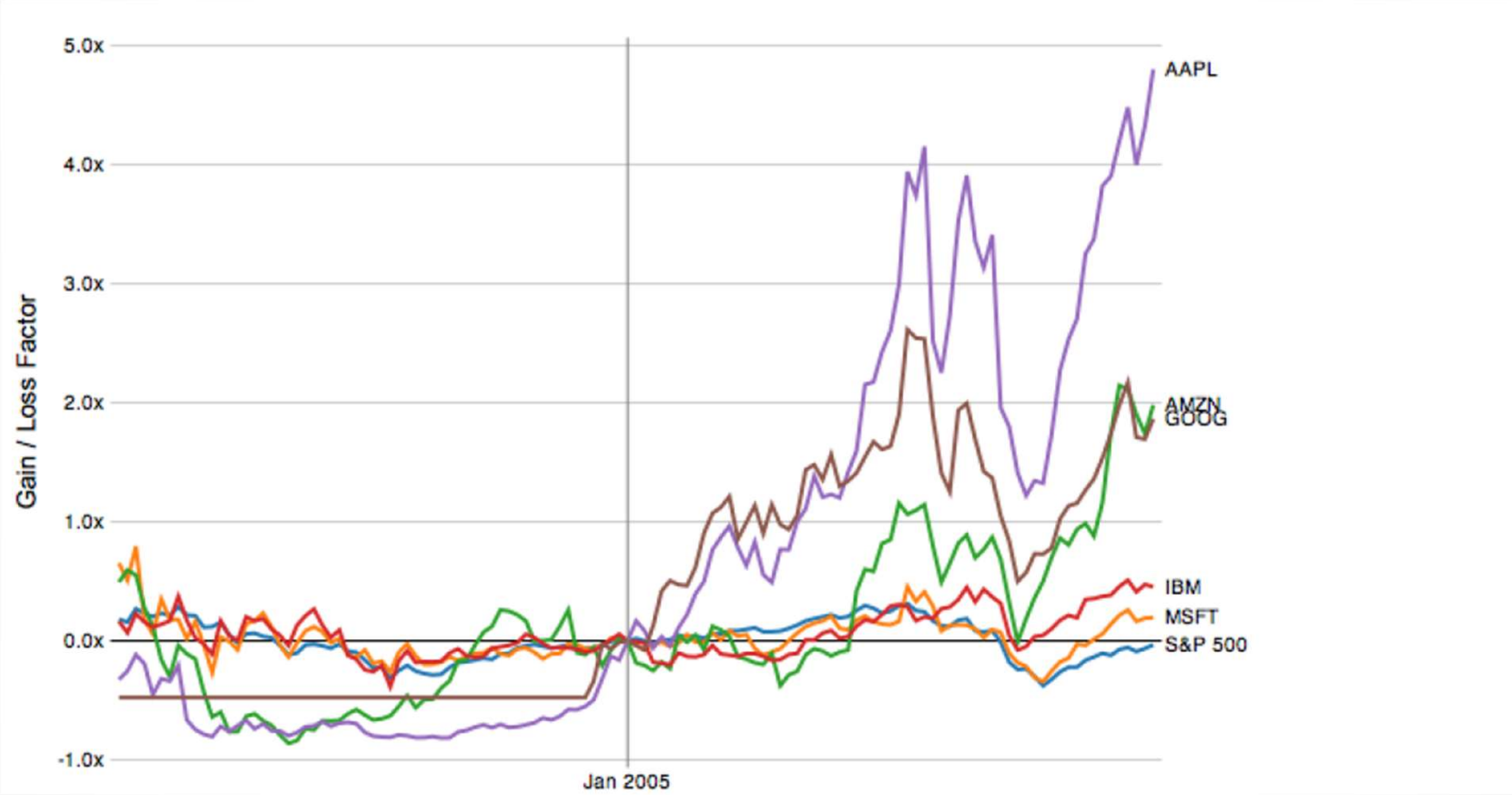
- A **sparkline** is a very small line chart, typically drawn without axes or coordinates.
- It presents the general shape of a variation (typically over time) in some measurement, such as temperature or stock market price, in a simple and highly condensed way.
- Whereas a typical chart is designed to professionally show as much data as possible, and is set off from the flow of text, sparklines are intended to be succinct, memorable, and located where they are discussed.
- Sparklines are small enough to be embedded in text, or several sparklines may be grouped together

Sparklines showing the movement of the Dow Jones Industrial Average and S&P 500 during February 7, 2006

Multivariate Time-series data

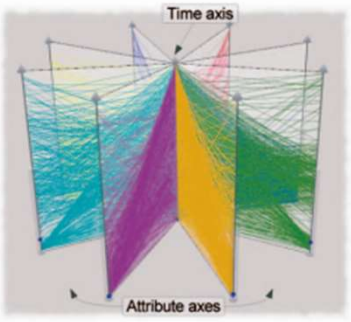
- What are these presenting?
  - One continuous quantitative value over time (time on x, variable on y)
- What if there are multiple values to track?

Multiple Lines



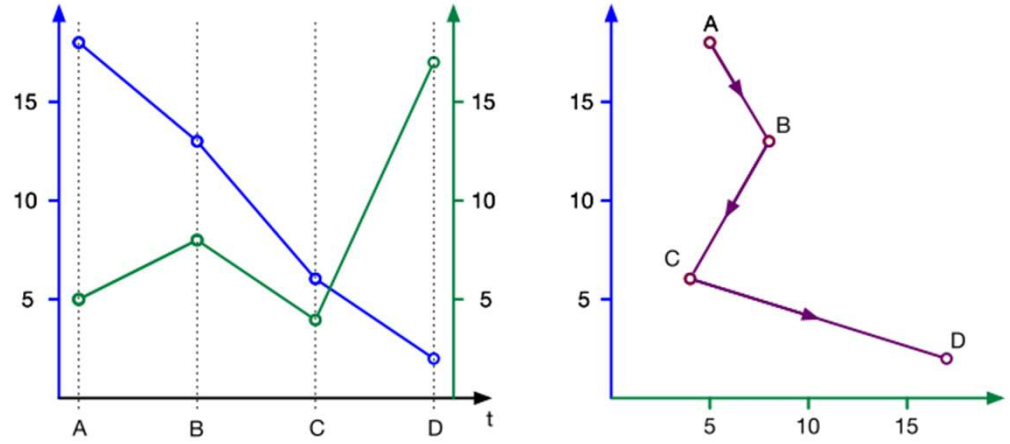
TimeWheel

- TimeWheel [Tominski et al, Infovis 2005]I is a multiaxes representation for visualizing multivariate data over time
- This is achieved by putting a time axis to a prominent position in the center of the display.
- A set of axes that encode time-dependent attributes is circularly arranged around the central time axis.
- For each time point in the considered data, lines descend from the time axis to the corresponding points on each of the attribute axes.
- The TimeWheel can be rotated to bring different attributes into
- the focus.
- Furthermore, each axis can be equipped with a slider to zoom into value ranges of interest and, in particular, to navigate the time axis.
- Interactive labels can be activated on demand to facilitate the identification of data values.
- Since the TimeWheel uses lines to represent data for each point in time, it is useful only for multivariatedata that are related to time points; data based on time intervals cannot be represented.



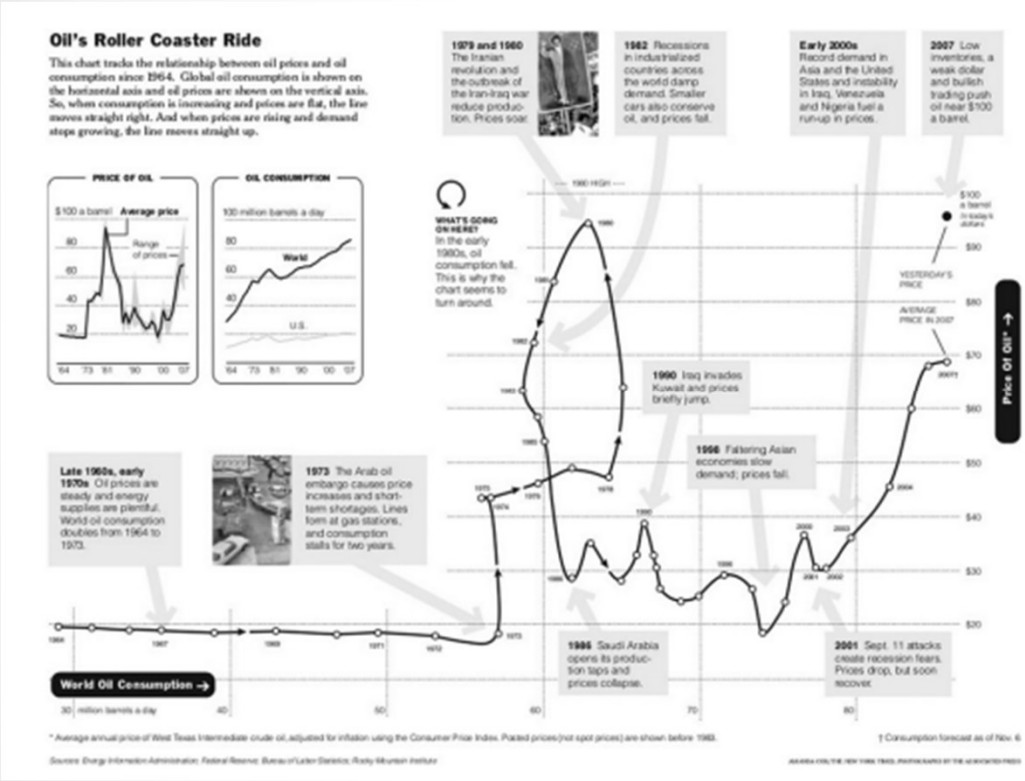
Connected Scatterplot

- Showing two variables over time
  - Use standard scatterplot
  - Plot the two values at different points in time
  - Connect those points, in order, with a line
  - Label key times (e.g., years)



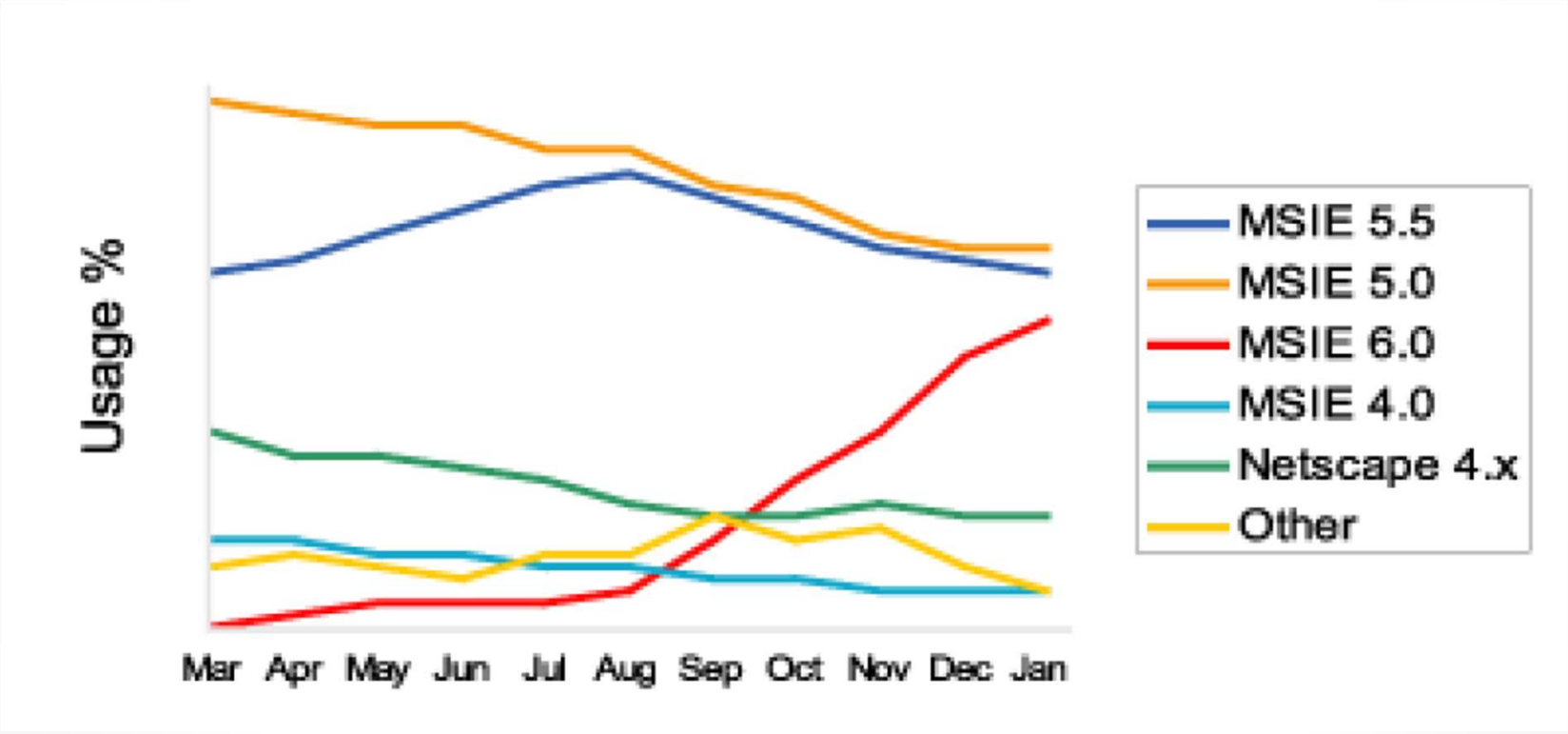
Two time series plotted in a dual-axis line chart (left) and a connected scatterplot (right).

Oil's Roller Coaster Ride

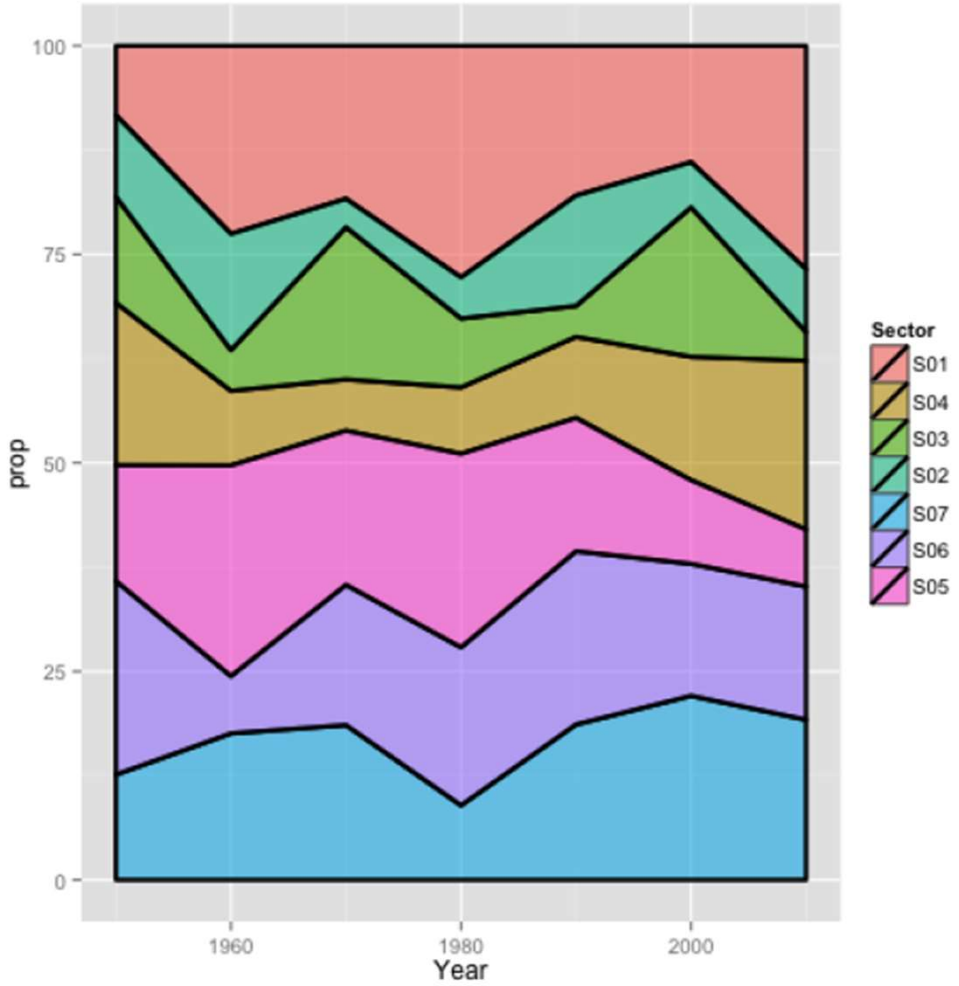


- One of the first uses of this visualization in news graphics was Oil's Roller Coaster Ride by Amanda Cox for The New York Times in February 2008
- Oil's Roller Coaster Ride uses the connected scatterplot to show the relationship between oil consumption (horizontal axis) and oil price (vertical axis) over time.
- The prominent loop draws readers' attention, annotations point out particular points of interes
- Since that article was published, many other instances of connected scatterplots have appeared in the news media

Proportions of Total

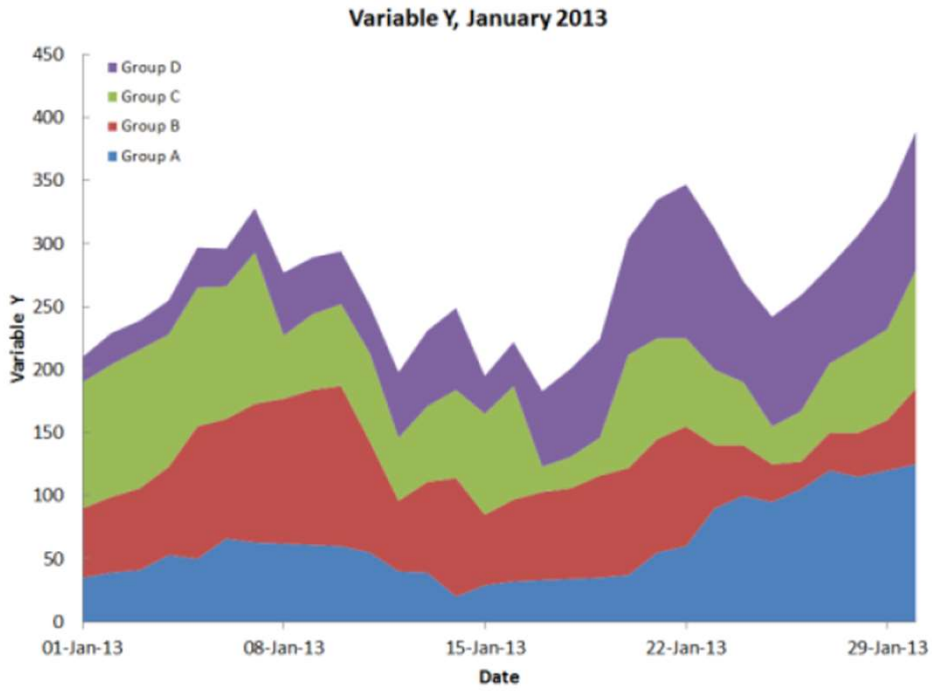


Stacked Area Graph



Stacked Area Chart

- What if the different values don't comprise a whole?
  - Don't add up to 100%

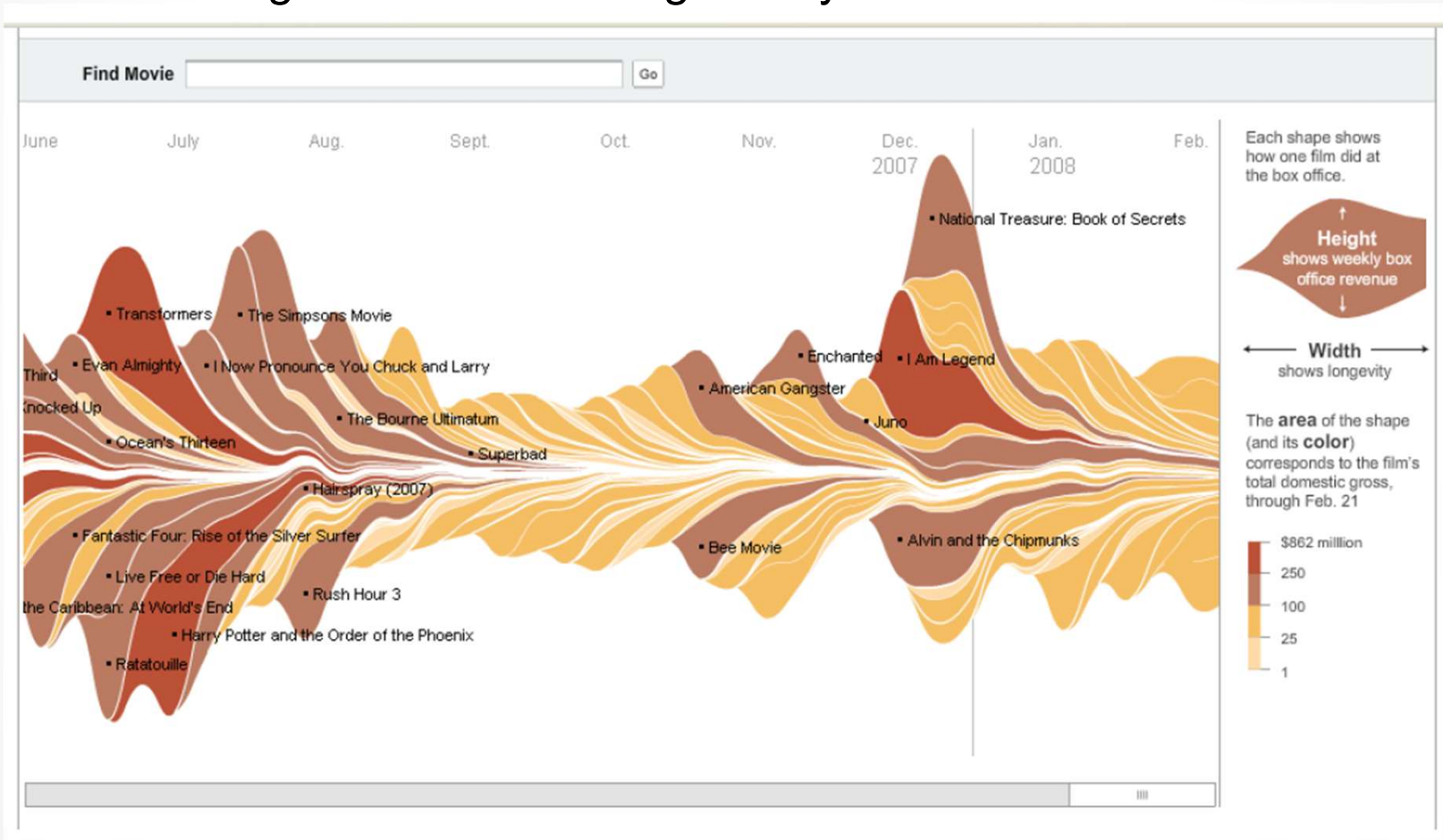


StreamGraph

- A **streamgraph**, or **stream graph**, is a type of stacked area graph which is displaced around a central axis, resulting in a flowing, organic shape.
- Unlike a traditional stacked area graph in which the layers are stacked on top of an axis, in a streamgraph the layers are positioned to minimize their "wiggle".
- More formally, the layers are displaced to minimize the sum of the squared slopes of each layer, weighted by the area of the layer.
- Streamgraphs display data with only positive values, and are not able to represent both negative and positive values
- Byron & Wattenberg TVCG'08

StreamGraph

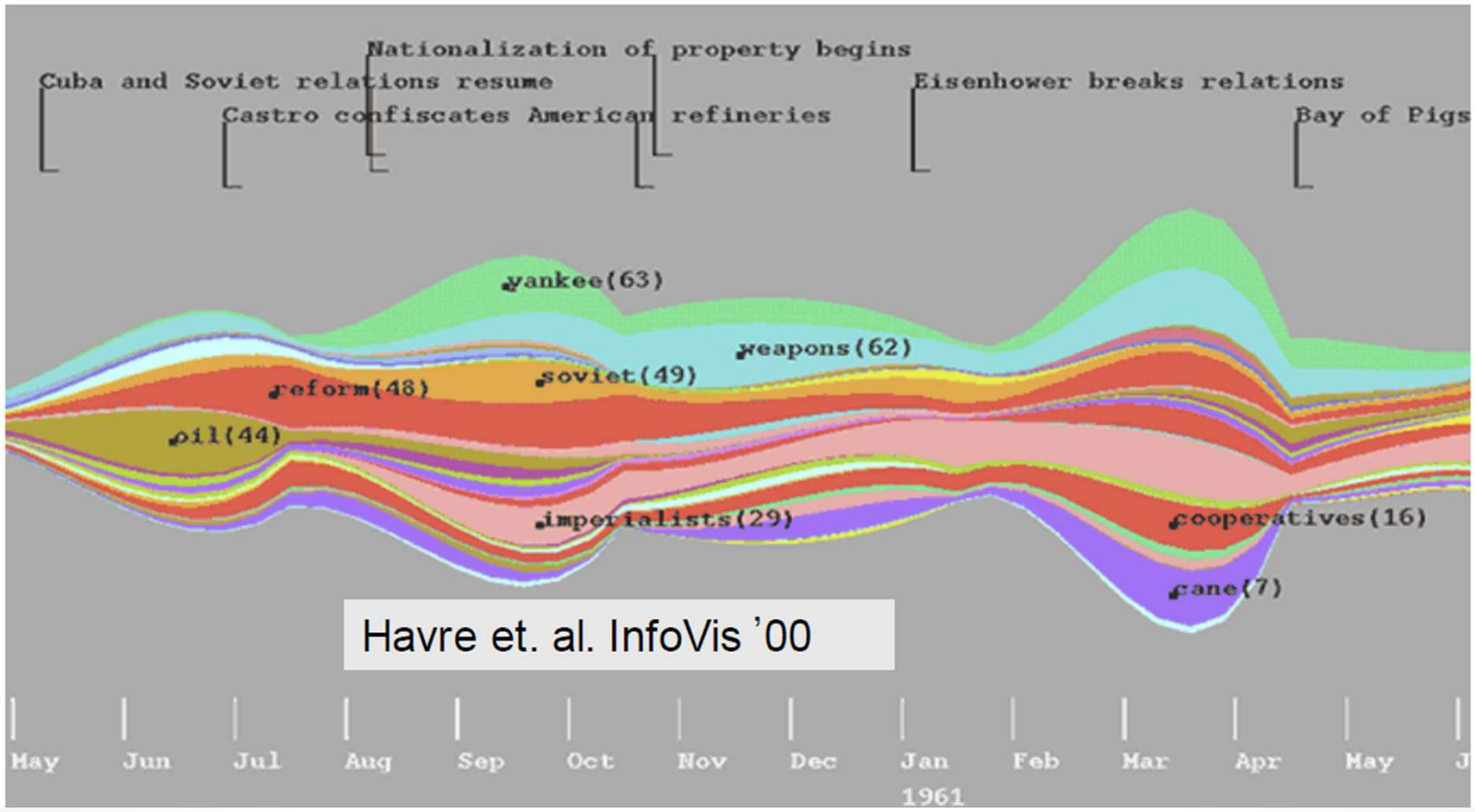
- Streamgraphs and their use were popularized by Amanda Cox in a February 2008 *New York Times* article on movie box office revenues.
- Cox got the idea from then-undergraduate Lee Byron who had used a similar method for visualizing his music listening history in CMU.



NYT article:
http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html

ThemeRiver

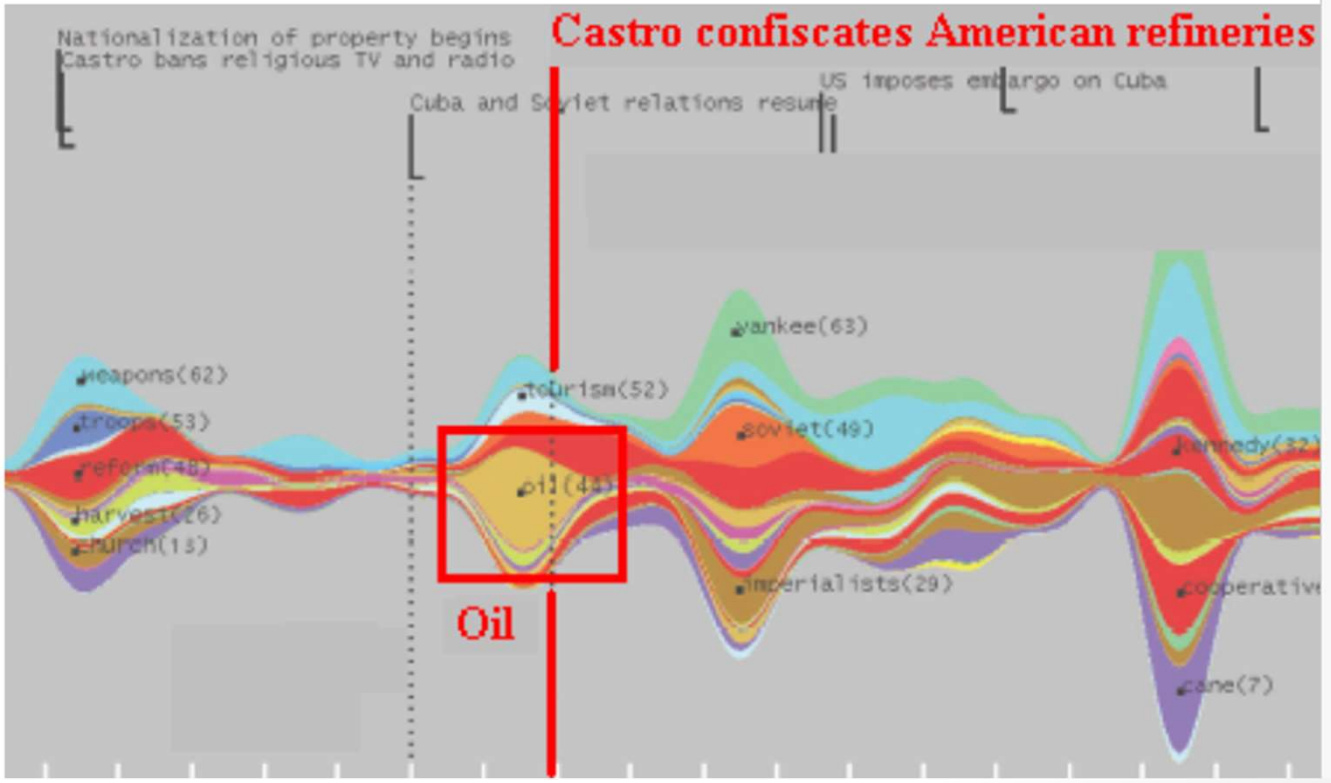
- Each 'theme' from news stories is a 'river'



Temporal Variation of Themes

ThemeRiver

- River height (thickness) encodes relative frequency of themes
- Key events overlaid

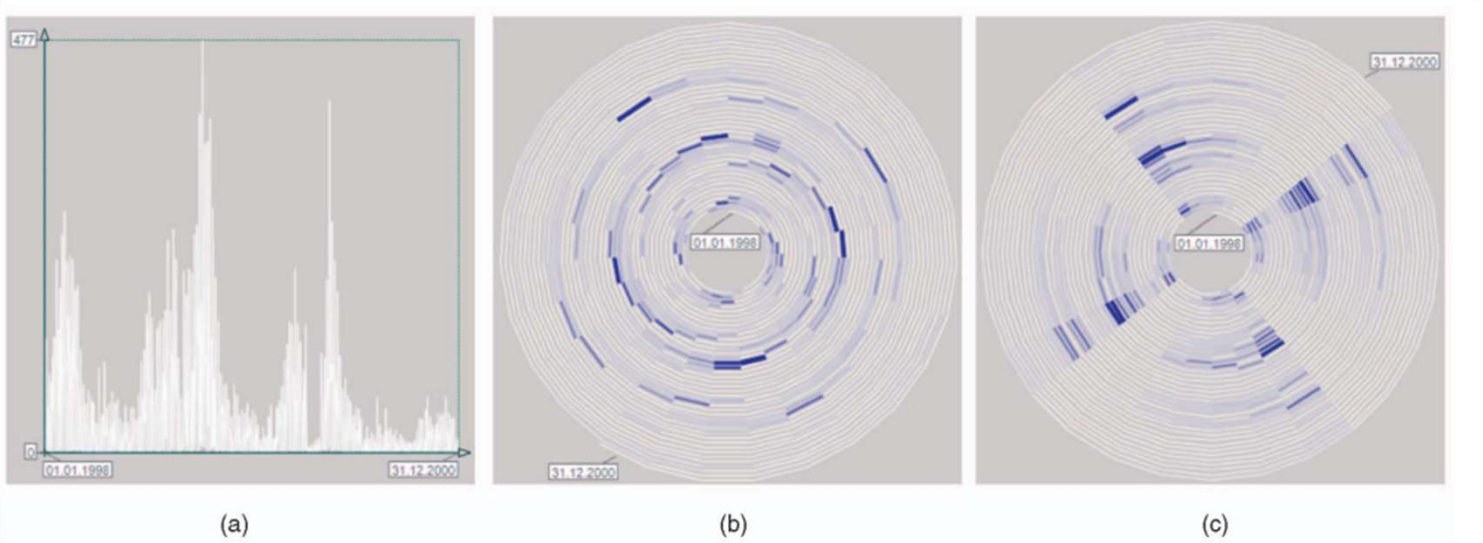


Innovations in ThemeRiver:

- A novel technique for creating a smooth interpolation from discrete data
- A layout method in which layers were not stacked starting on the x-axis, but rather in a symmetrical shape with the x-axis at the center.

Visualizing Cyclic Data

- The Spiral Graph [M. Weber et al, InfoVis 2001), is a visualization technique that focuses on cyclic characteristics of time-series data by using a spirally shaped time axis
- The main purpose of this technique is the detection of previously unknown periodic behavior of the data.



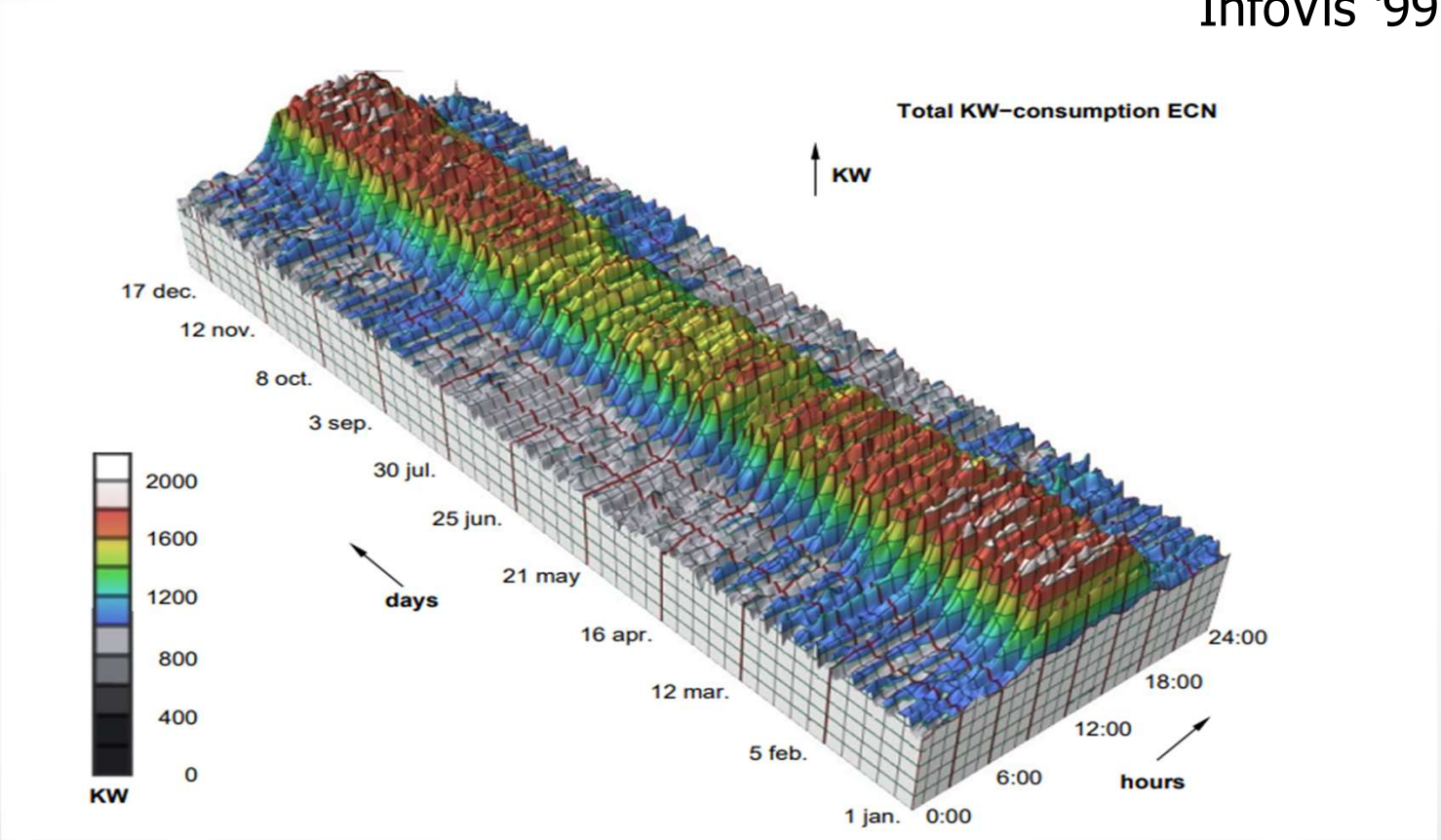
Different visual representations of a time-oriented data set describing the number of influenza cases over a period of three years.

a) Time series plot (periodic pattern is difficult to discern).
b) SpiralGraph encoding 27 days per cycle (improperly parameterized—periodic pattern is hard to see).
c) SpiralGraph encoding 28 days per cycle (properly parameterized—periodic pattern stands out)

Visualizing large Time-series data

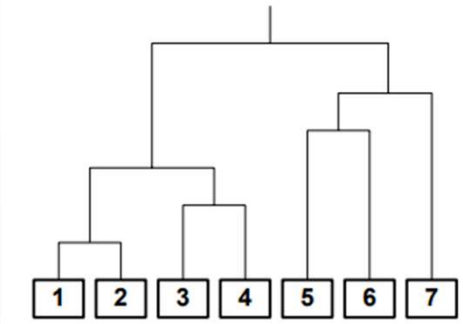
- Make the data as two-dimensional, for instance as f (day, hour)
- The days and hours are mapped on different axes
- Data is visualized via color
- The third dimension is also used to display the data, yielding a mountain landscape
- Shows: i) Typical daily pattern ii) Seasonal trends
- Can't Show: i) Weekly patterns ii) Details

van Wijk & van Selow
InfoVis '99

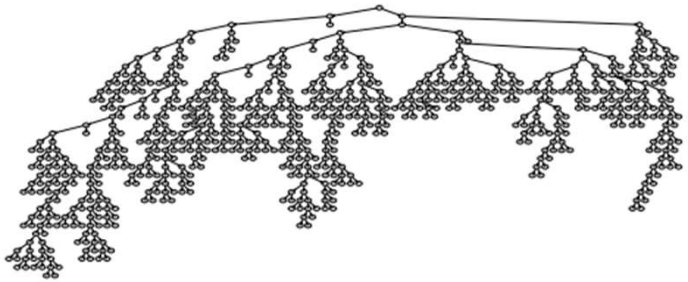


Cluster Analysis

- Goal:
  - Merge similar day patterns into clusters, such that the day patterns within a cluster are more similar than the day patterns in other clusters
- Technique:
  - Find two most similar days, make into one new composite
  - Keep repeating until some preset number left or some condition met
  - Cluster can be visualized as a dendogram

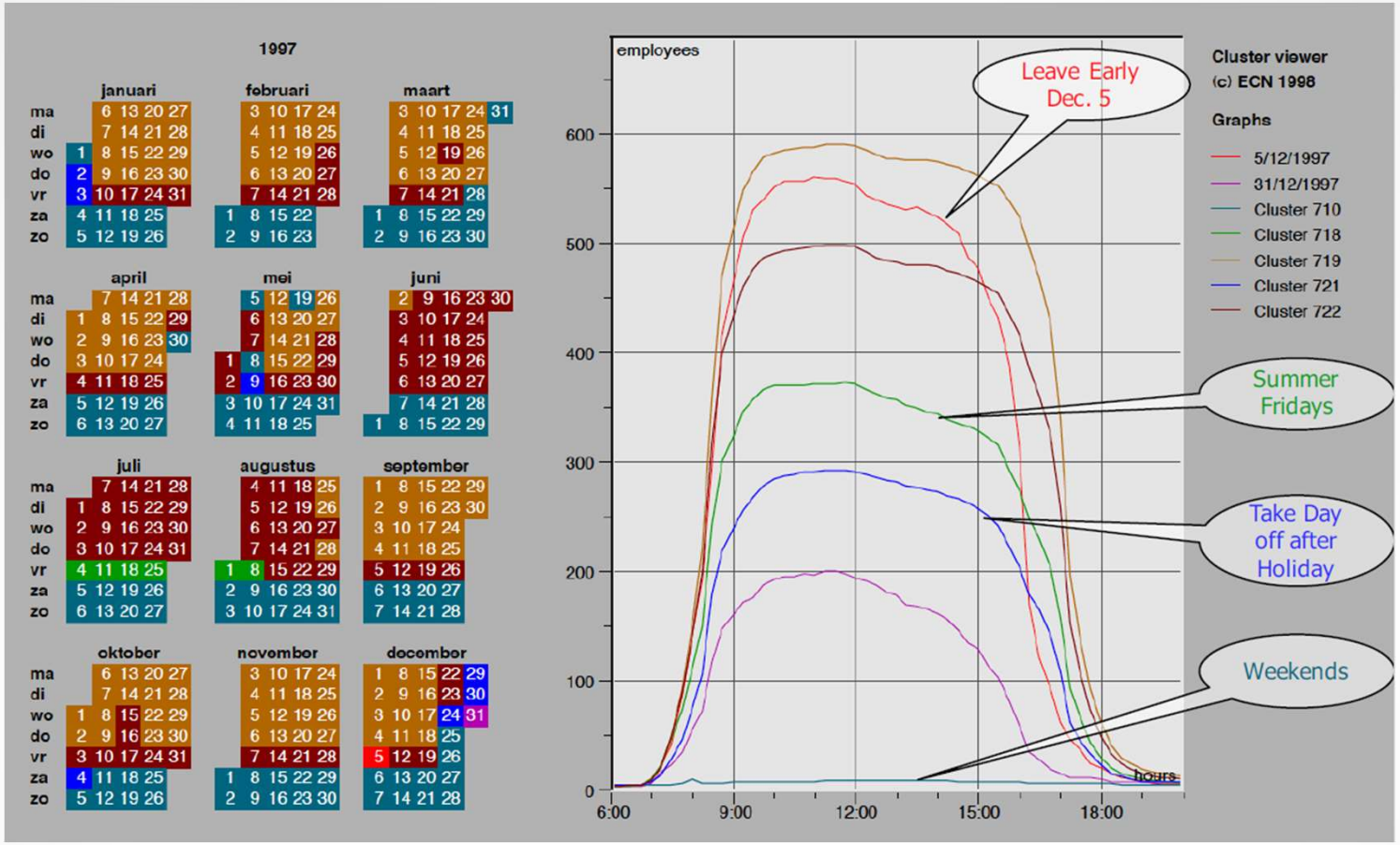


Dendrogram



Full clustering tree (for 365 days)

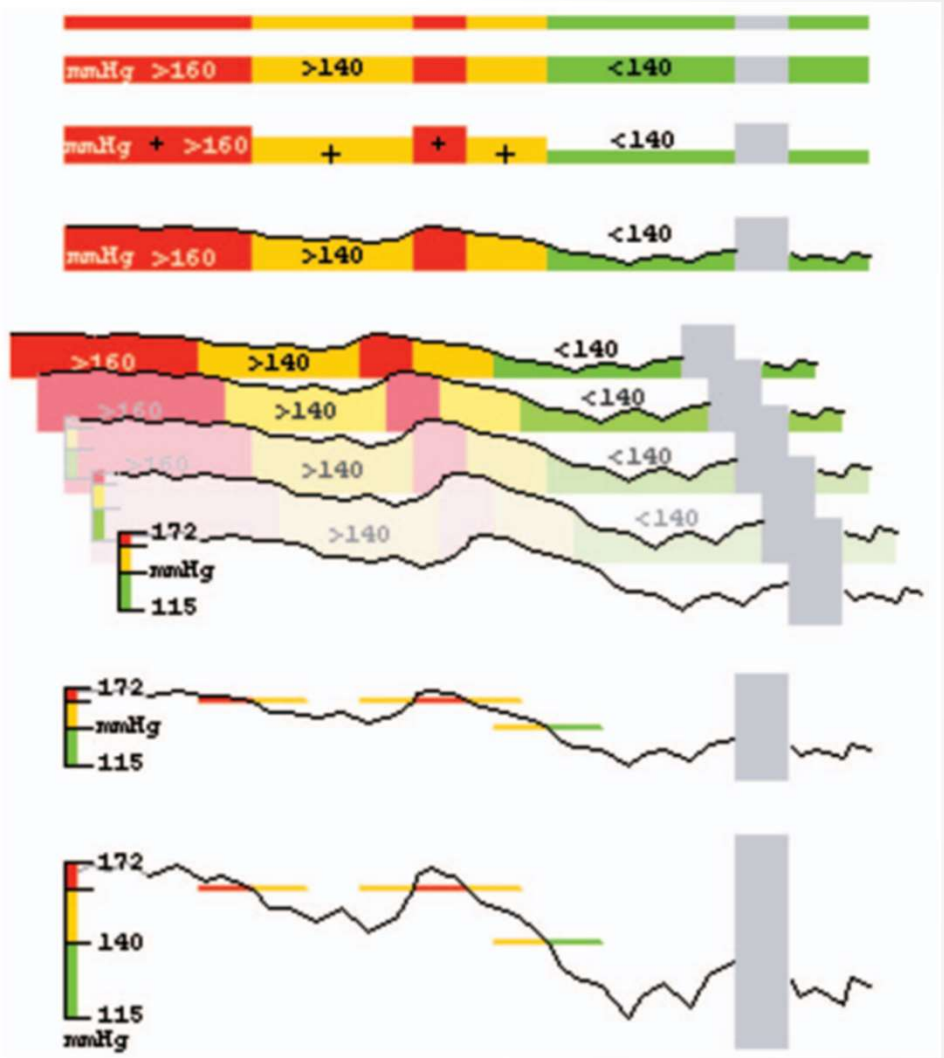
Visualization – Daily Patterns & Clusters



- A cluster analysis of time series data on the number of employees present at ECN
- Daily patterns are shown as graphs, clusters are shown on a calendar.
- Colors indicate corresponding clusters and patterns.
- The most significant seven clusters are shown.
- On the right, the average value per cluster is shown as a colored graph
- On the left, each day in the calendar is colored according to the cluster to which it belongs

Temporal Data Abstraction

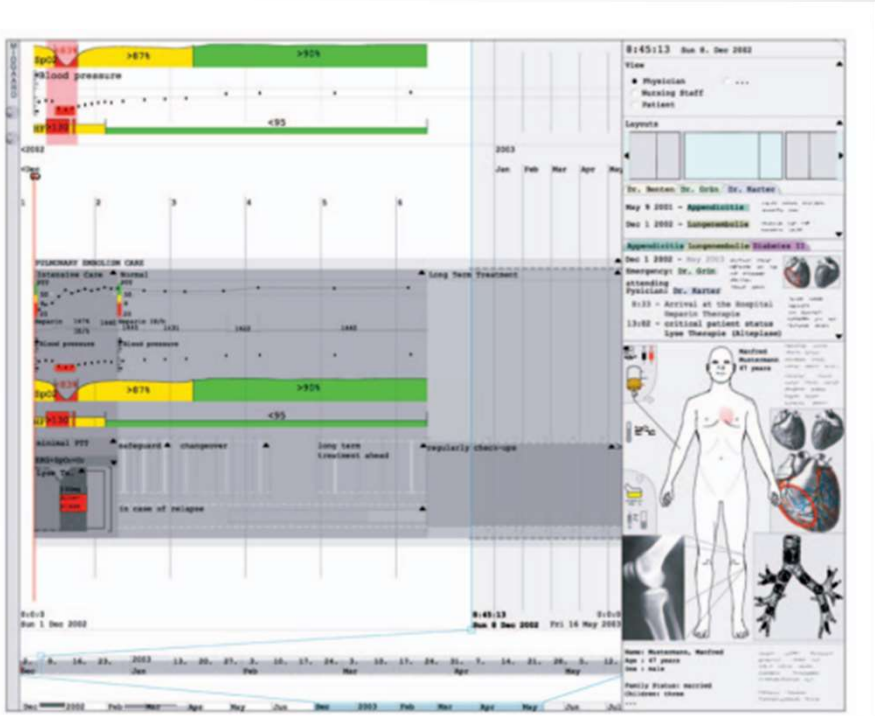
- The Midgaard project provides different levels of abstractions for time-oriented data.
- Switching between these levels results in a smoothly integrated semantic zoom functionality
- Figure shows steps of resizing/zooming the representation of a data stream from a broad overview with qualitative values to the fine structure with quantitative details



R. Bade, S. Schlechtweg, and S. Miksch, "Connecting Time-Oriented Data and Information to a Coherent InteractiveVisualization," Proc. 2004 Conf. Human Factors in Computing Systems (CHI '04), 2004

Midgaard UI

- The visualization of temporal aspects comprises three linked time axes:
  - The first one (bottom) provides a fixed overview of the underlying data and their full temporal range.
  - Selecting a subrange in that time axis defines the temporal bounds for the second (middle)
  - The third (top) time axis. By interactively adjusting the subrange, users can easily zoom and pan in time



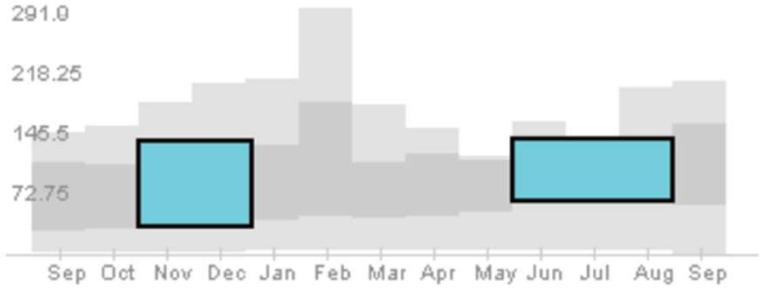
- The upper left part shows different measurements (for example, blood gas measurements and blood pressure) and their corresponding temporal abstractions.
- The right part explains the additional patient's information.
- The lower left part explains the time axis interaction
- The selected subrange at the bottom time axis can be moved and rescaled to pan and zoom the time range shown in the middle and top time axes.

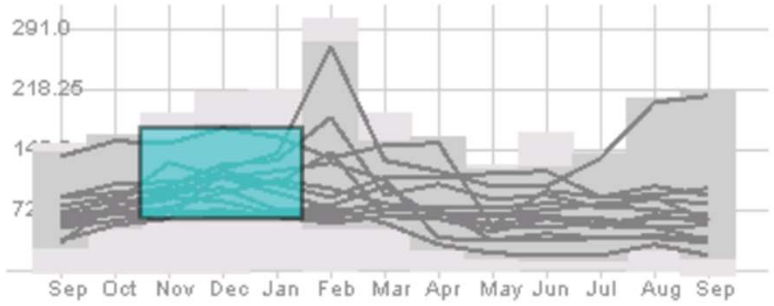
Querying Time Series Data

- Timeboxes are rectangular widgets that can be used in direct-manipulation graphical user interfaces (GUIs) to specify query constraints on time series data sets.
- Timeboxes are used to specify simultaneously two sets of constraints: given a set of N time series profiles, a timebox covering time periods $x_1 \ldots x_2$ ($x_1 < x_2$) and values $y_1 \ldots y_2$ ($y_1 < y_2$) will retrieve only those $n \in N$ that have values $y_1 \leq y_2$ during all times $x_1 \leq x_2$.
- TimeSearcher is an information visualization tool that combines timebox queries with overview displays, query-by-example facilities, and support for queries over multiple time-varying attributes

Timeboxes

Can create rectangles that function as matching regions





Multiple boxes are "anded"

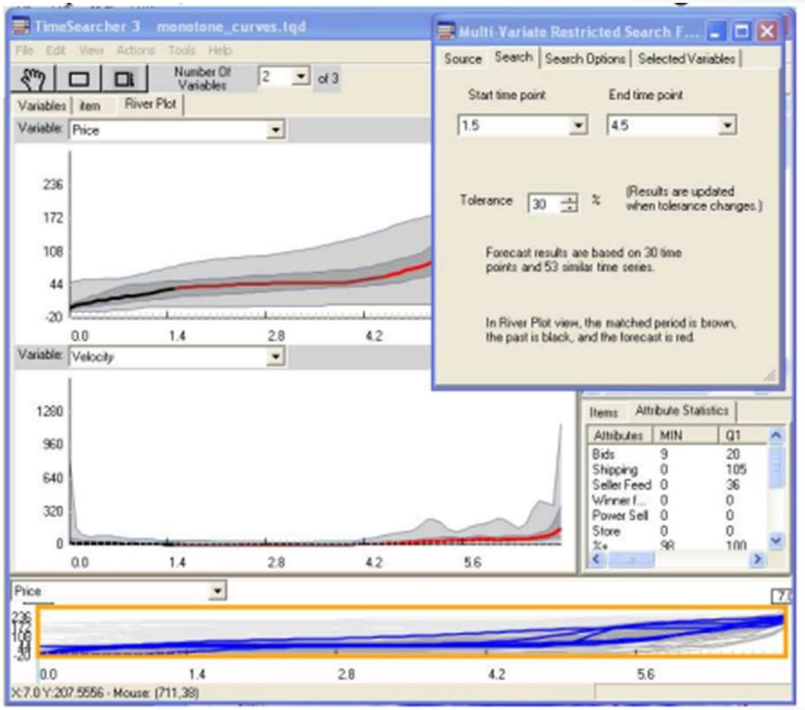
Light gray is all data's extent

Darker grayed region is data envelope that shows extreme values of queries matching criteria

Hochheiser & Shneiderman
Proc. Discovery Science '01
Info Vis '04

Time-series Forecasting

- Time-series forecasting has a large number of applications.
- Users with a partial time desire estimates of the future behavior. The paper presented a data driven forecasting method and interface called Similarity-Based Forecasting (SBF).
- A pattern matching search in an historical time series dataset produces a subset of curves similar to the partial time series.
- The forecast is displayed graphically as a river plot showing statistical information about the SBF subset.
- A forecasting preview interface allows users to interactively explore alternative pattern matching parameters and see multiple forecasts simultaneously.
- Buono, Plaisant, Simeone, Aris, Shneiderman, et.al. Infovis, 2007

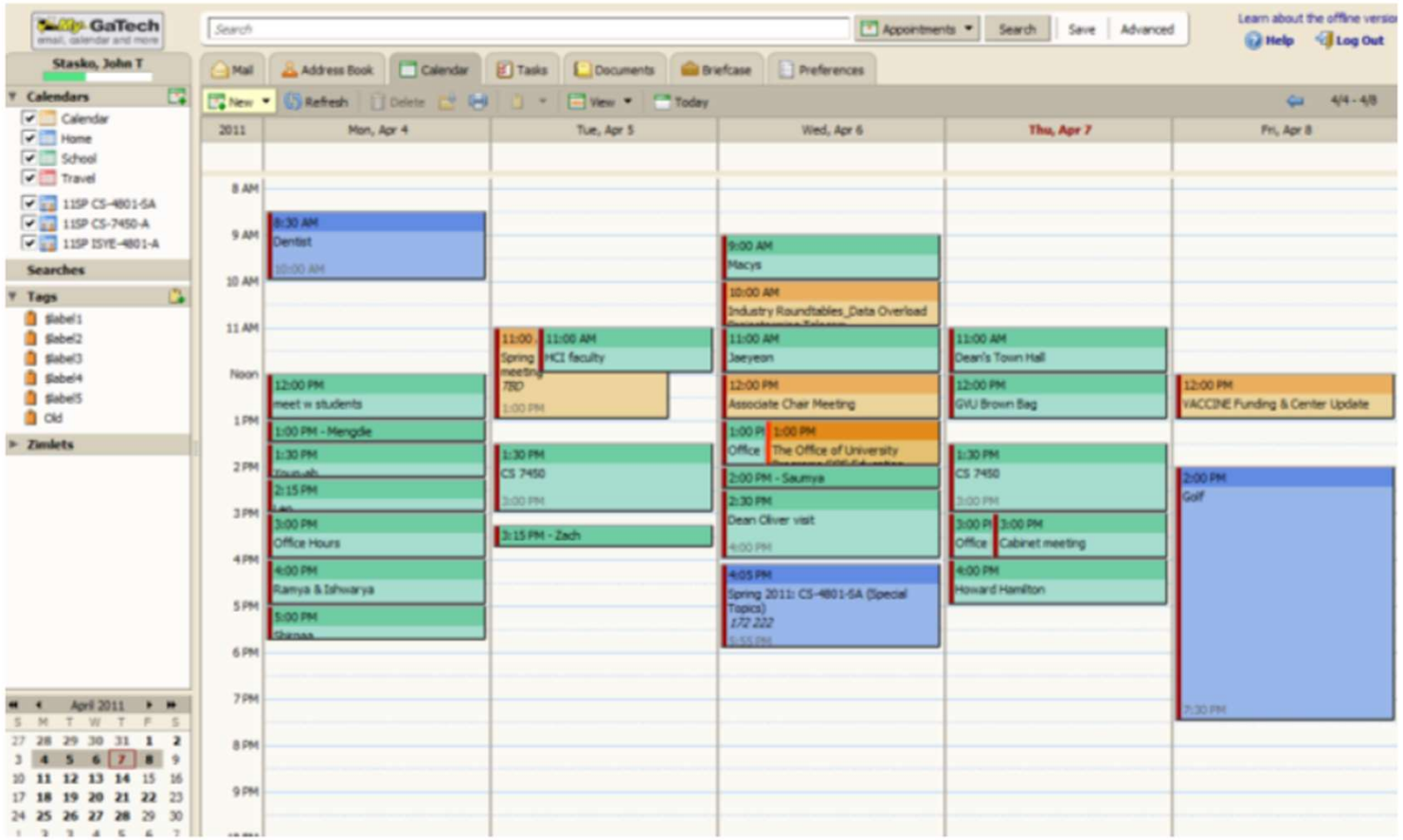


Forecasting interface: red represents the forecast. The median of the subset of matched items during the period used for matching is brown, while the median before this period is black

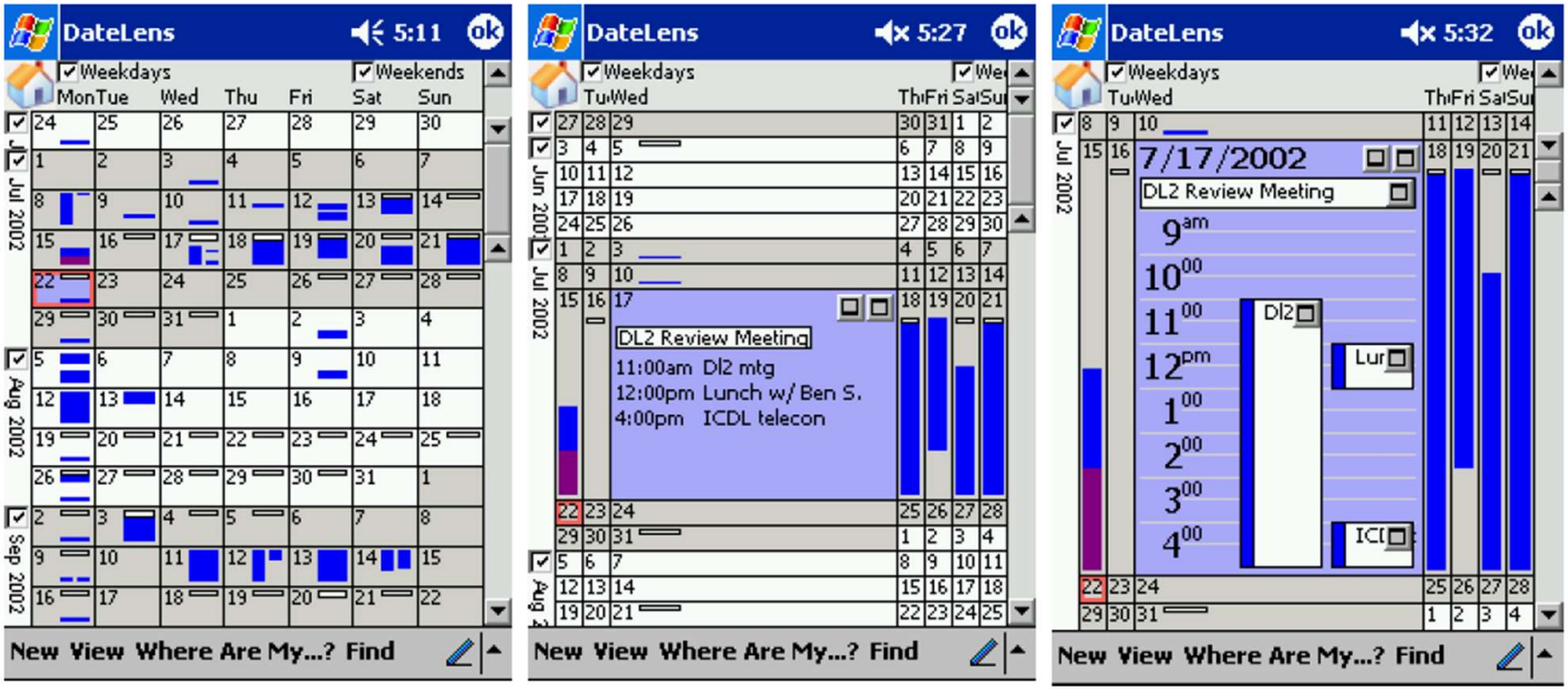
Overview

- Introduction
- Time-series Data
- Applications

Calendar View



DataLens

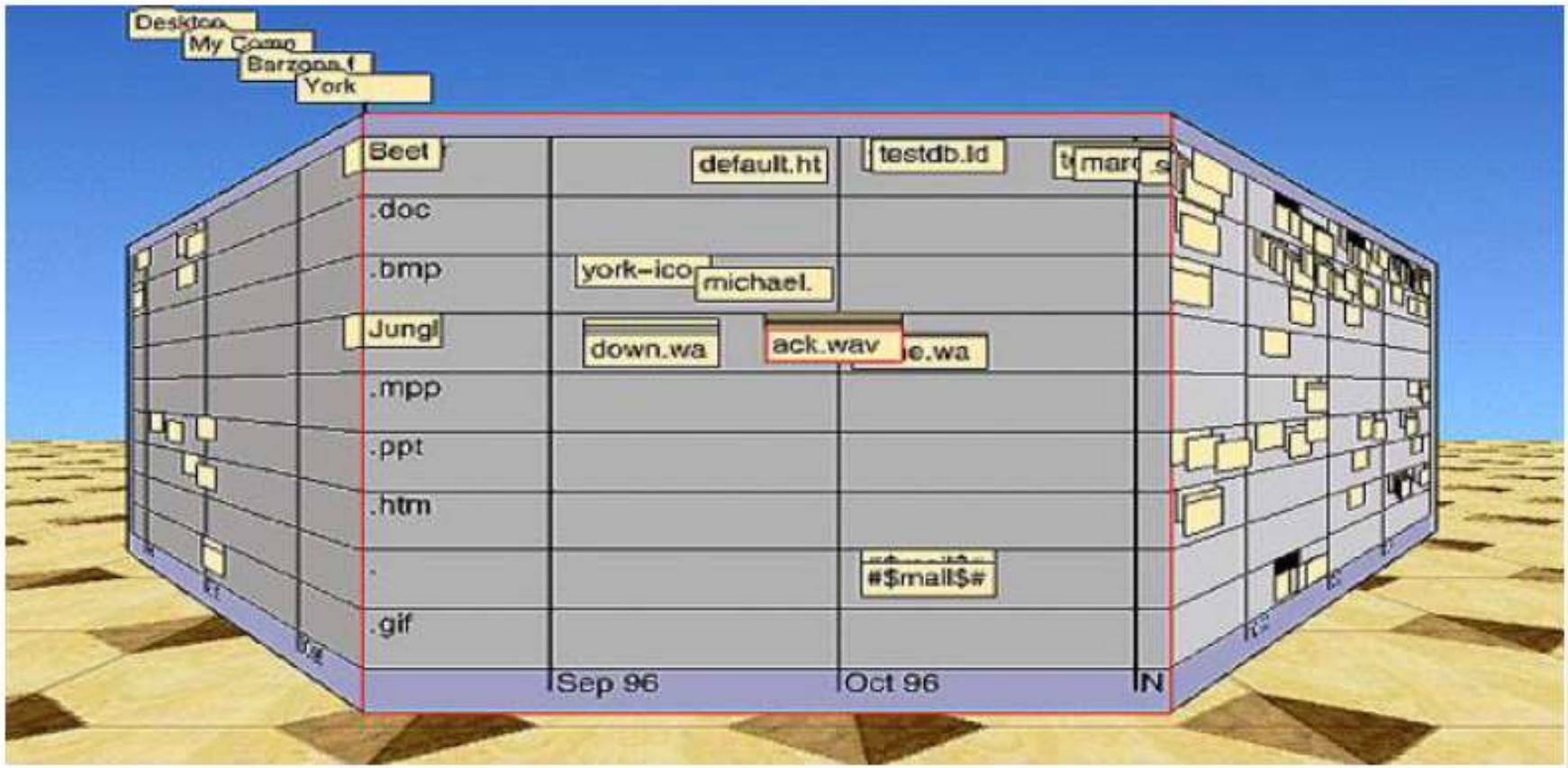


Fisheye approach

Bederson et al
ACM ToCHI '04

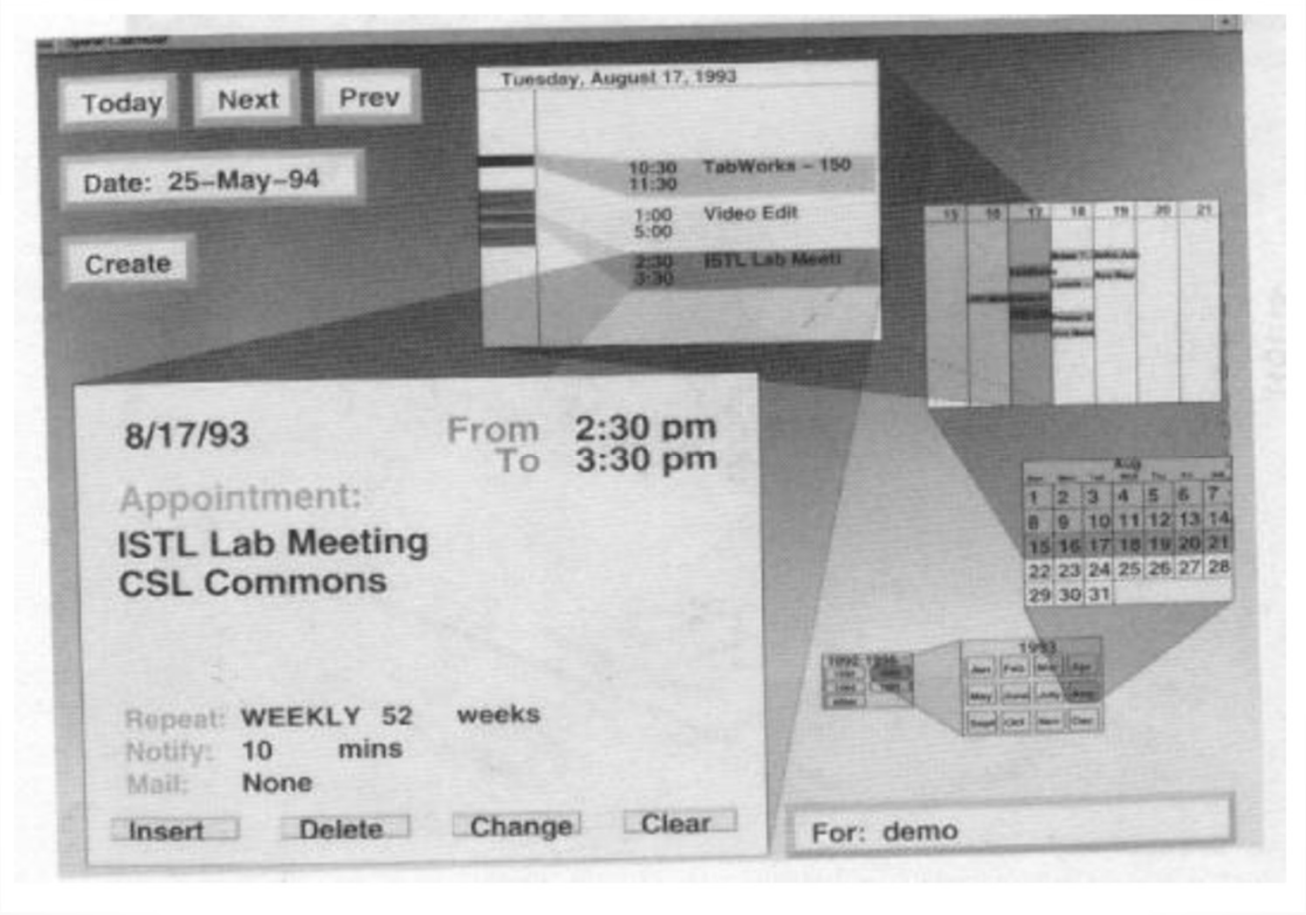
Perspective Wall

Focus+Context View of Calendars



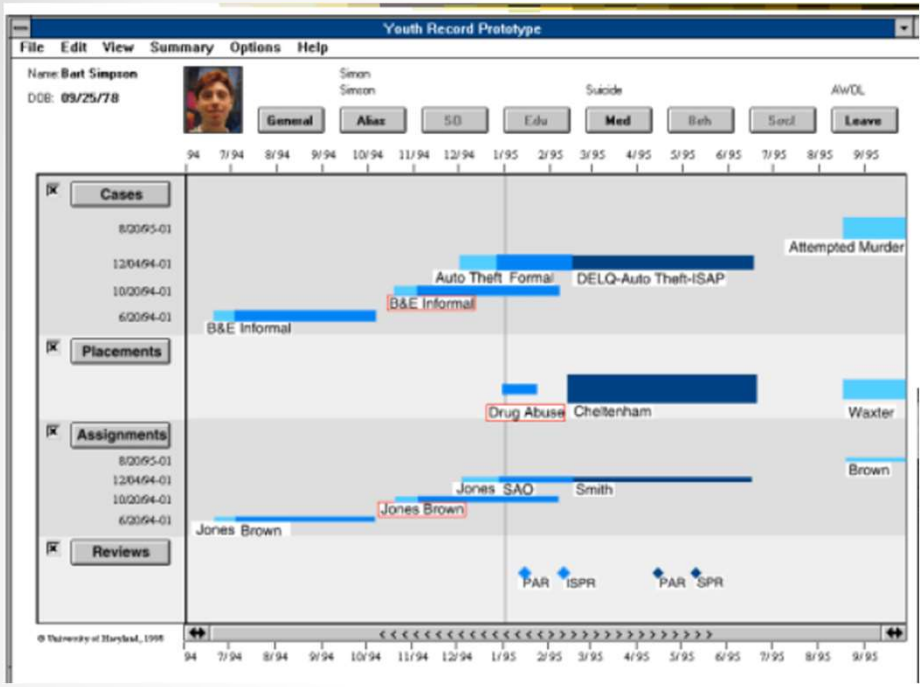
Mackinlay, Roberston and Card, 1991

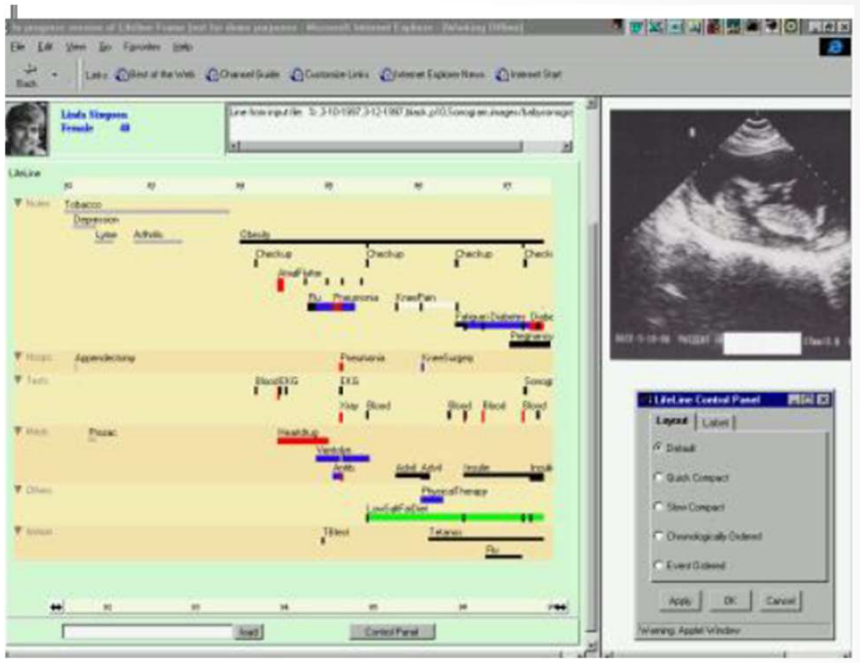
Spiral Calendar



Mackinlay, Robertson & DeLine
UIST '94

Lifelines





Visualize personal history in some domain (legal/medical)

Plaisantet al CHI ‘96

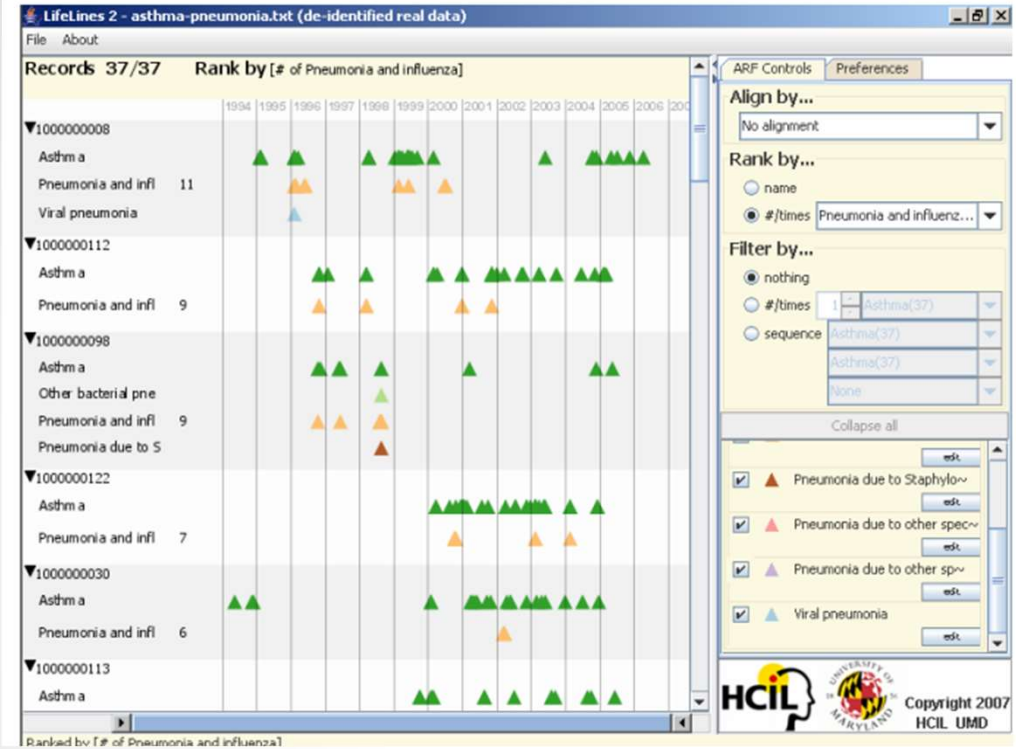
Lifelines

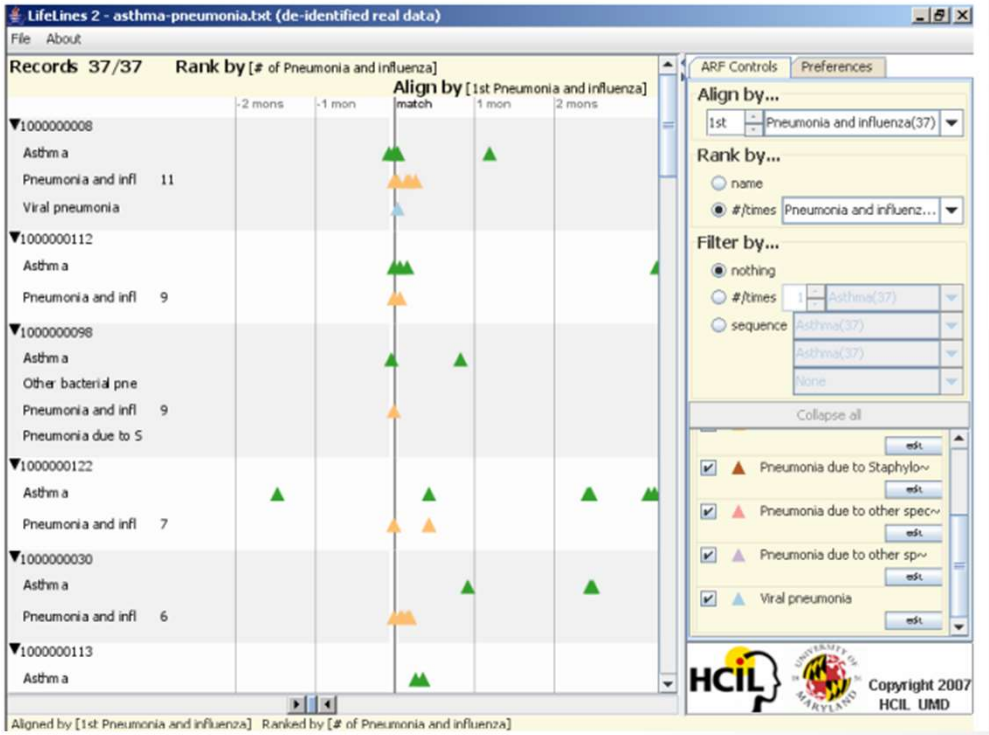
- LifeLines provide a general visualization environment for personal histories that can be applied to medical and court records, professional histories and other types of biographical data.
- A one screen overview shows multiple facets of the records.
- Aspects, for example medical conditions or legal cases, are displayed as individual time lines, while icons indicate discrete events, such as physician consultations or legal reviews.
- Line color and thickness illustrate relationships or significance.
- Rescaling tools and filters allow users to focus on part of the information.

Lifelines 2

Focus on alignment along events

Wang et al CHI '08



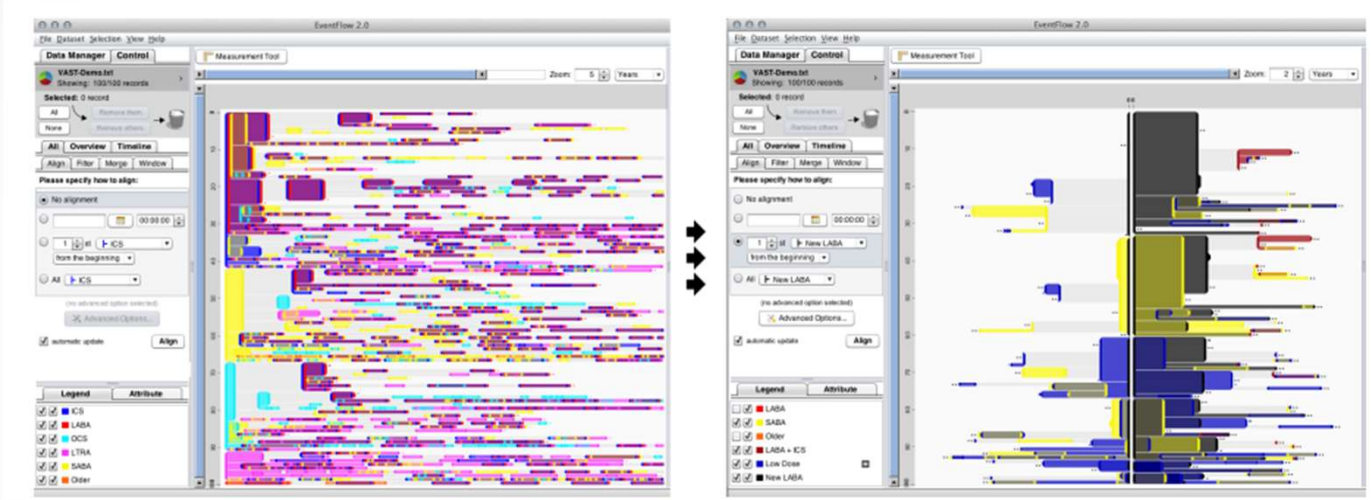


The interface without alignment. Each triangle represents an event. Note the data is presented chronologically, and the records are ranked by the number of Pneumonia and influenza events. It is easy to see the co-occurrence of Pneumonia and influenza and Asthma. However, it is not clear in patients' first Pneumonia and influenza, whether Asthma occurred before or after. Users are forced to zoom in to each first occurrence of Pneumonia and influenza for details, but each zoom can only reveal the details around a particular Pneumonia and influenza event.

Shows the same dataset. However, all patient records are aligned by the 1st Pneumonia and influenza. Note the relative time scale on the top. A single zoom had been applied to the alignment line. It is easily verifiable that the first 3 patients were diagnosed with asthma within a month prior or at the same time their pneumonia was diagnosed, while the other 2 patients in view were not.

Event Flow

- Follow on from Lifelines
- Smart aggregations to show overviews of large collections of events
- Monroe et al TVCG'13



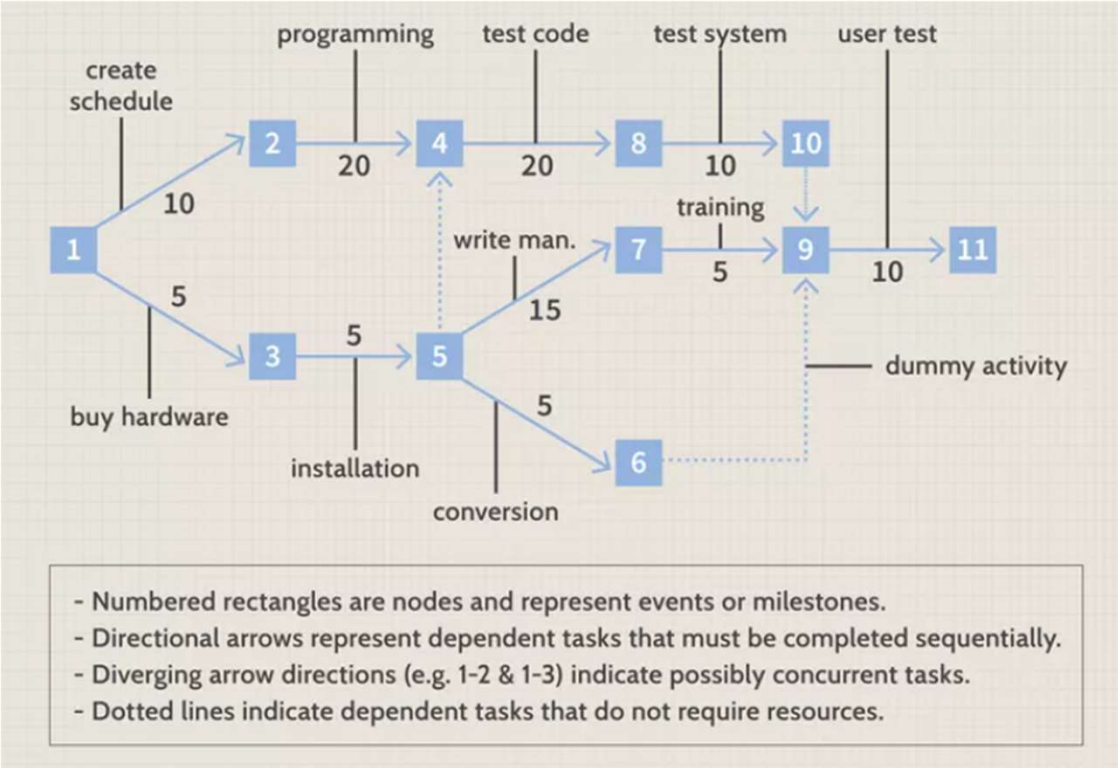
- In EventFlow, the original LABA dataset, consisting of over 2700 visual elements (left), was quickly pared down to the events most critical to the study.
- The simplified dataset (right) consists of only 492 visual elements, an 80% reduction in visual complexity.
- From this simplified figure, aligned by the patients' "new" LABA prescription, researchers were immediately able to notice the data sparsity on the left side of the alignment point, indicating that patients had not received other treatments in the months leading up to their LABA prescription (i.e. not following the recommended practices)

Project Management

- Project (write software, design/build plane, hire new person) involves
  - Multiple steps
  - Spread over time (Time Intervals)
  - Some steps depend on other steps

- How can we plan/manage project?
  - With PERT/Gnatt/CPM charts

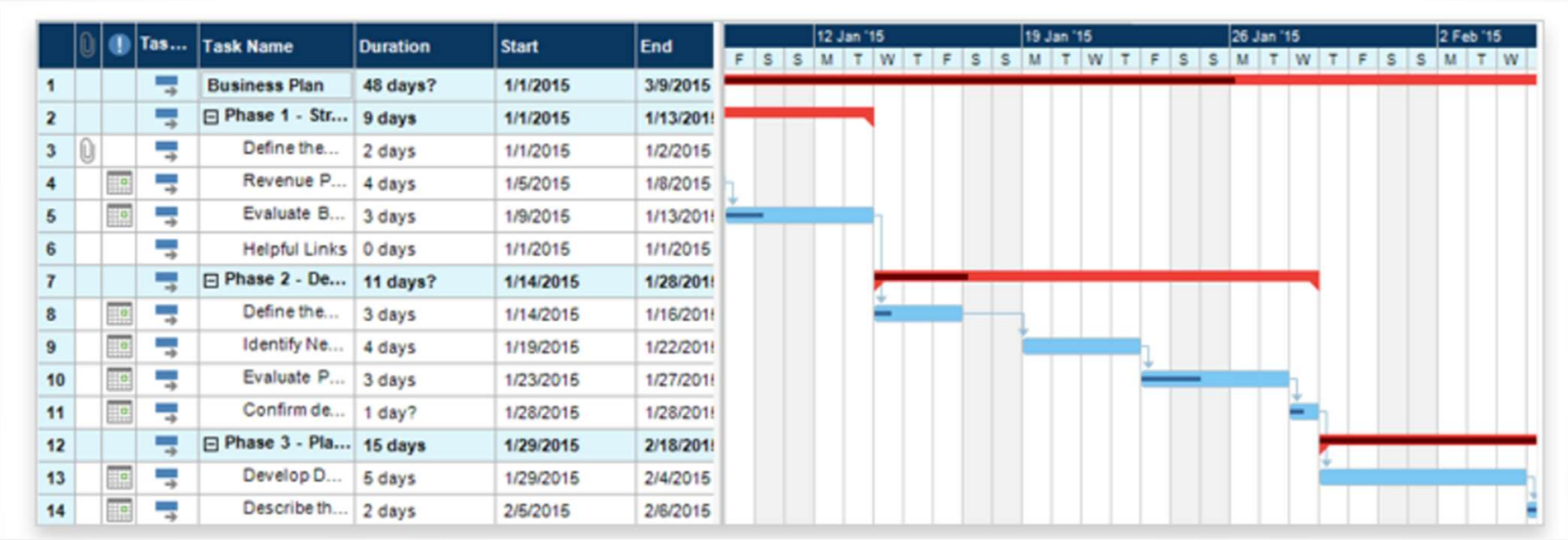
PERT Chart



- A PERT chart is a tool used to schedule, organize, and map out tasks within a project.
- PERT stands for program evaluation and review technique.
- It provides a visual representation of a project's timeline and breaks down individual tasks.
- PERT chart leads to Gnatt chart

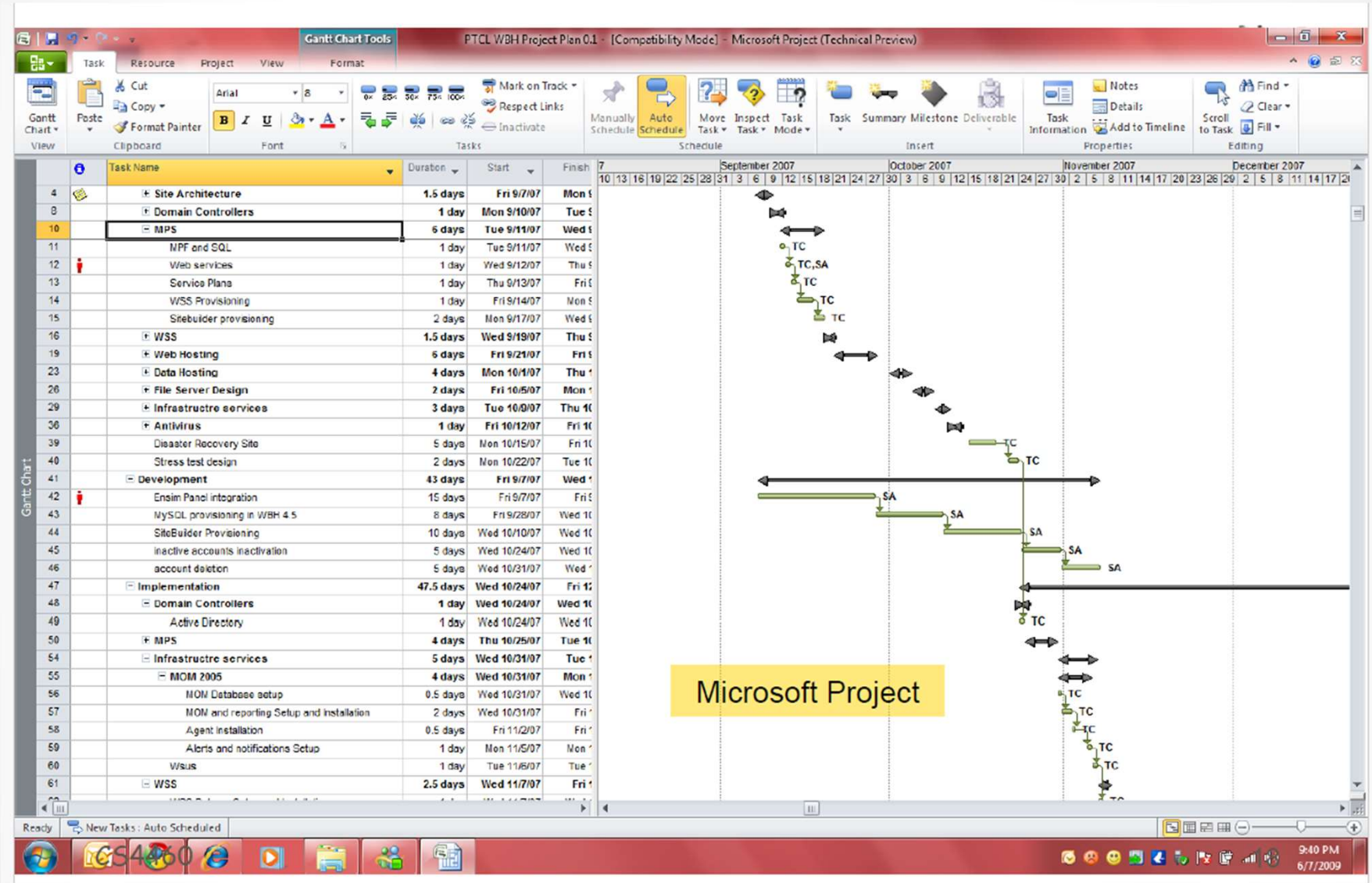
Gantt Chart

- A Gantt chart, commonly used in project management, is one of the most popular and useful ways of showing activities (tasks or events) displayed against time.
- On the left of the chart is a list of the activities and along the top is a suitable time scale.
- Each activity is represented by a bar; the position and length of the bar reflects the start date, duration and end date of the activity.

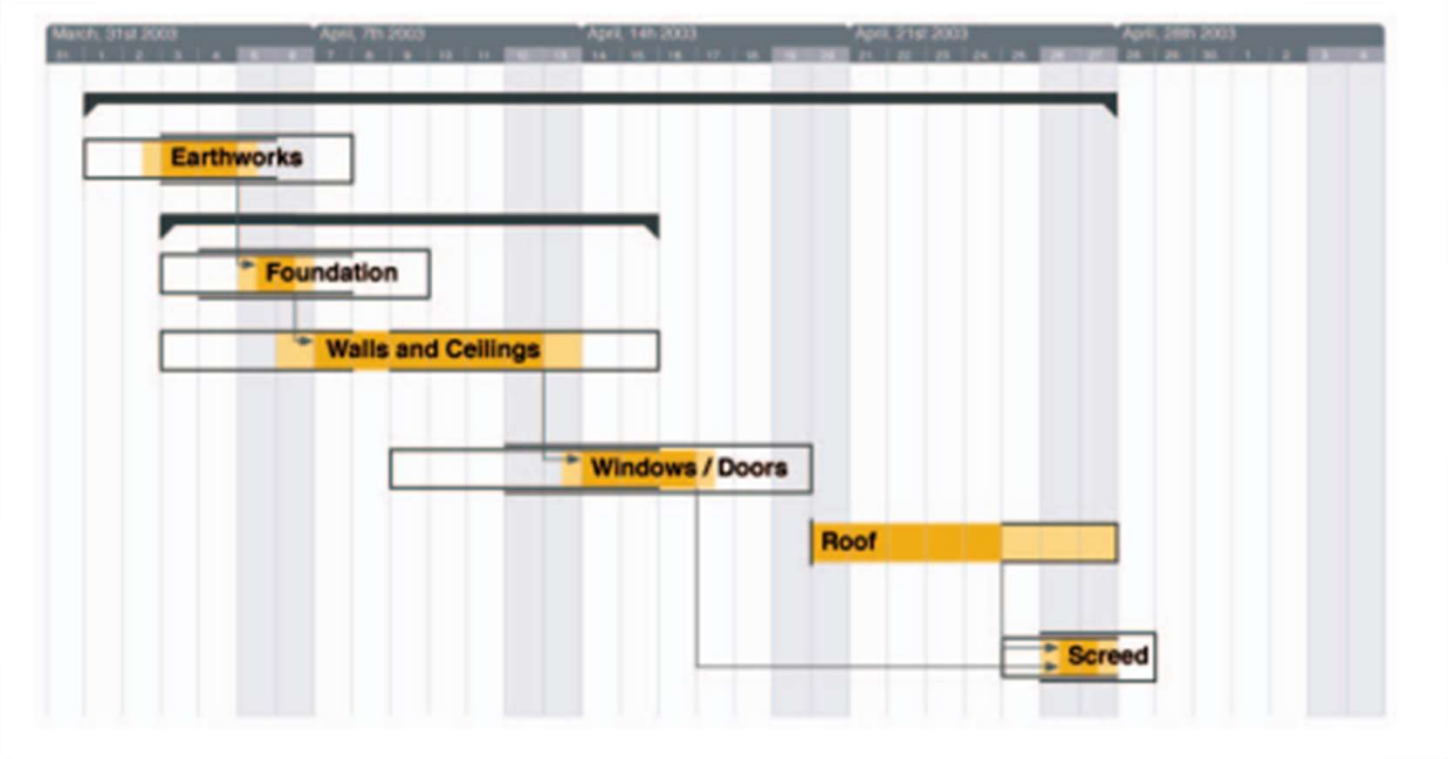


	Tas...	Task Name	Duration	Start	End
1		Business Plan	48 days?	1/1/2015	3/9/2015
2		⊟ Phase 1 - Str...	9 days	1/1/2015	1/13/201
3		Define the...	2 days	1/1/2015	1/2/2015
4		Revenue P...	4 days	1/5/2015	1/8/2015
5		Evaluate B...	3 days	1/9/2015	1/13/201
6		Helpful Links	0 days	1/1/2015	1/1/2015
7		⊟ Phase 2 - De...	11 days?	1/14/2015	1/28/201
8		Define the...	3 days	1/14/2015	1/16/201
9		Identify Ne...	4 days	1/19/2015	1/22/201
10		Evaluate P...	3 days	1/23/2015	1/27/201
11		Confirm de...	1 day?	1/28/2015	1/28/201
12		⊟ Phase 3 - Pla...	15 days	1/29/2015	2/18/201
13		Develop D...	5 days	1/29/2015	2/4/2015
14		Describe th...	2 days	2/5/2015	2/6/2015

Many Project Management Products



Mapping Uncertainty

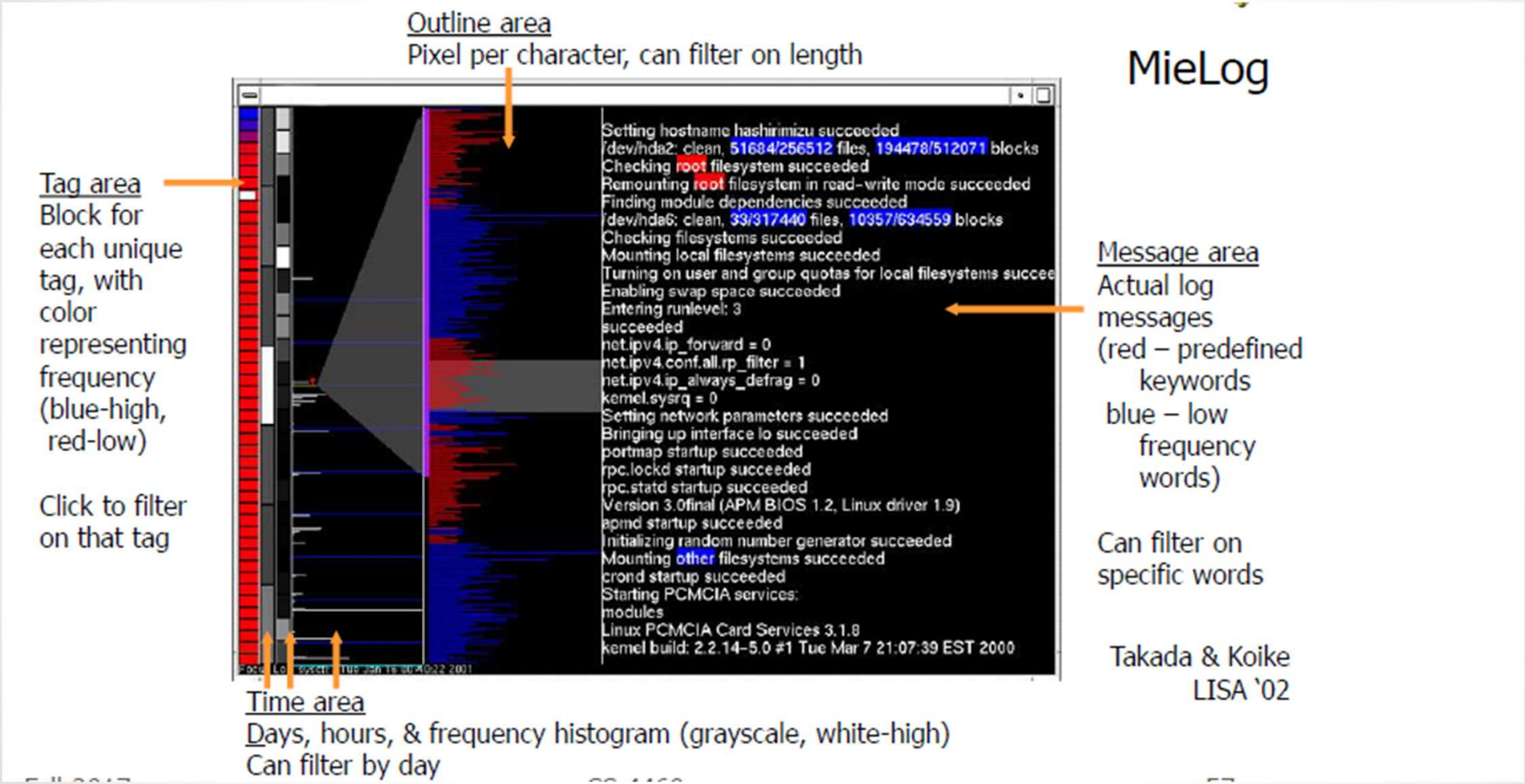


- PlanningLines (Aigner et al. Infovis 2005) consist of two encapsulated bars that represent the minimum and maximum durations
- Bounded by two caps representing the start and end intervals.
- Allows the representation of possible distributions of start, end, and duration of an activity,
- Also allows showing temporal uncertainty - usually connected with future planning.

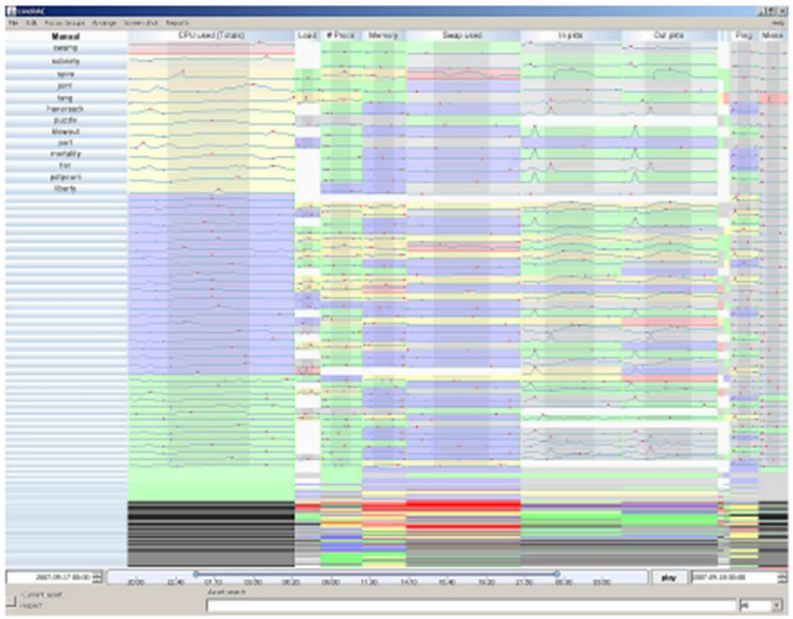
Computer Systems Management

- Computer system logs
  - Very large-scale temporal data
  - Complex: Many processes, machines
  - Potentially huge amount of data
- Tedious to examine the text
- Requirements:
  - Look for unusual circumstances, patterns, etc
  - Show more context of what else was going on at that time
  - Allow several different levels of detail
- Examples:
  - MieLog: For after-the fact inspection of discrete event data
  - LiveRAC: Supports review of continuous data and real-time monitoring
  - Splunk: Commercial System

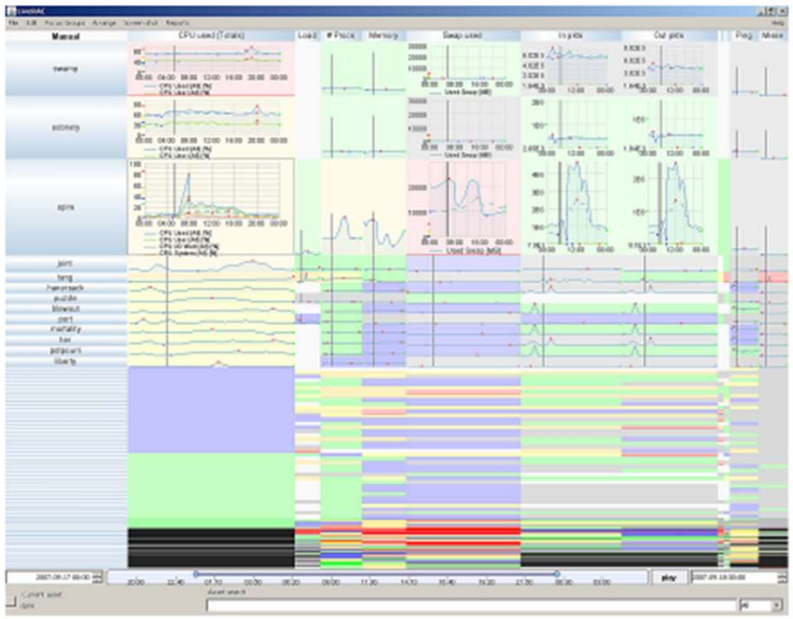
MieLog: Visual Log Browser



LiveRAC: Visual Exploration of System Management Time-Series Data



(a)



(b)

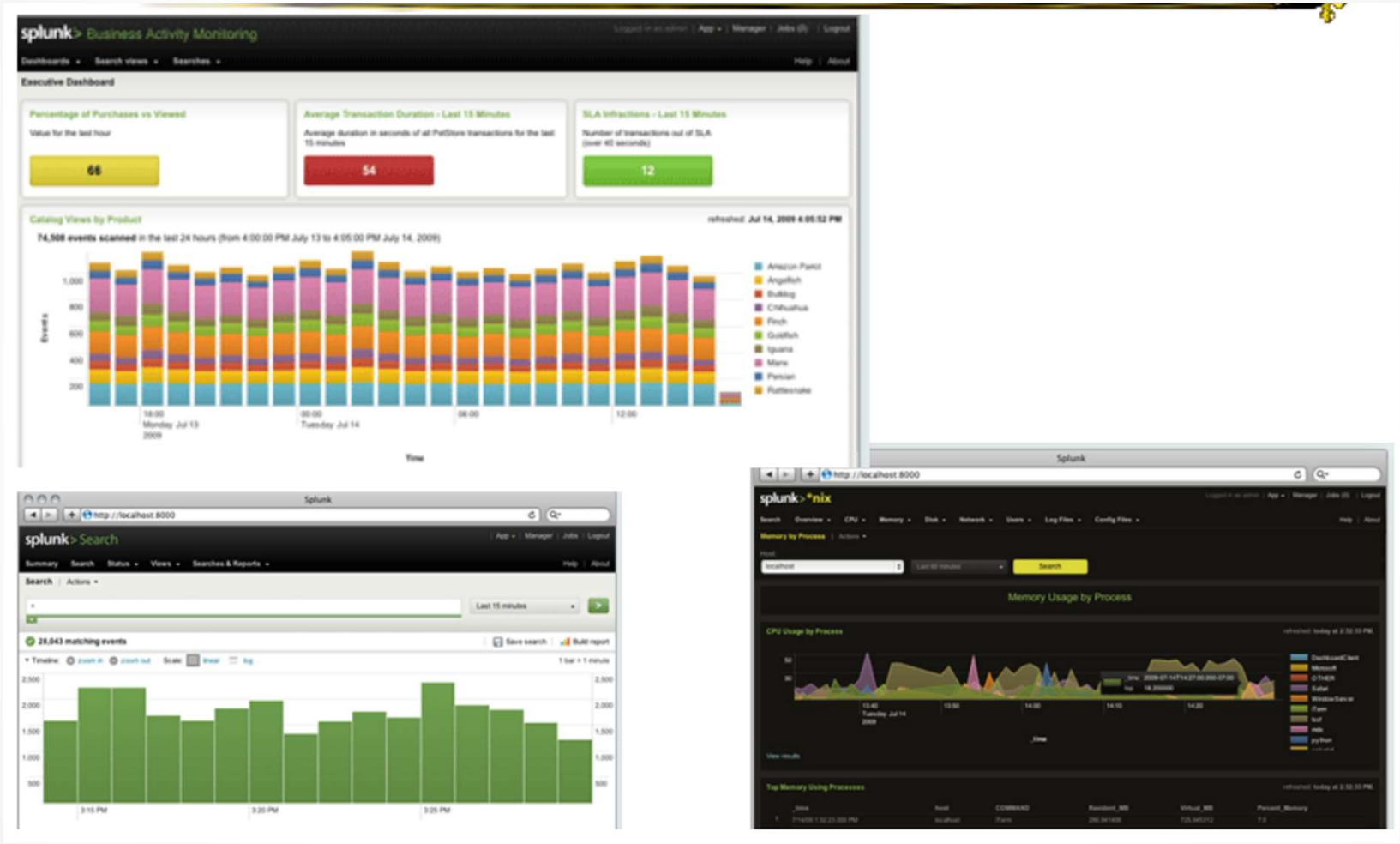
McLachlan et al CHI 2008

LiveRAC shows a full day of system management time-series data using a reorderable matrix of area-aware charts.
Over 4000 devices are shown in rows, with 11 columns representing groups of monitored parameters.
(a)The user has sorted by the maximum value in the CPU column. The first several dozen rows have been stretched to show sparklines for the devices, with the top 13 enlarged enough to display text labels. The time period of business hours has been selected, showing the increase in the In pkts parameter for many devices.
(b) The top three rows have been further enlarged to show fully detailed charts in the CPU column and partially detailed ones in Swap and two other columns.
The time marker (vertical black line on each chart) indicates the start of anomalous activity in several of spire's parameters.
Below the labeled rows, we see many blocks at the lowest semantic zoom level, and further below we see a compressed region of highly saturated blocks that aggregate information from many charts

LiveRAC Design Principles

- Involve users early
- Use spatial position as perceptual cue
  - Related items are proximate (grouped)
- Use side-by-Sparklines to minimize load on short term memory
- Link multiple views
- Animate changes – no abrupt transitions
- Immediate feedback
- Use graph styles with which users are already familiar
- Overview first, zoom and filter, details on demand
- Assertion: Several levels of detail side-by-side

Splunk



http://www.splunk.com/

Geography + Time

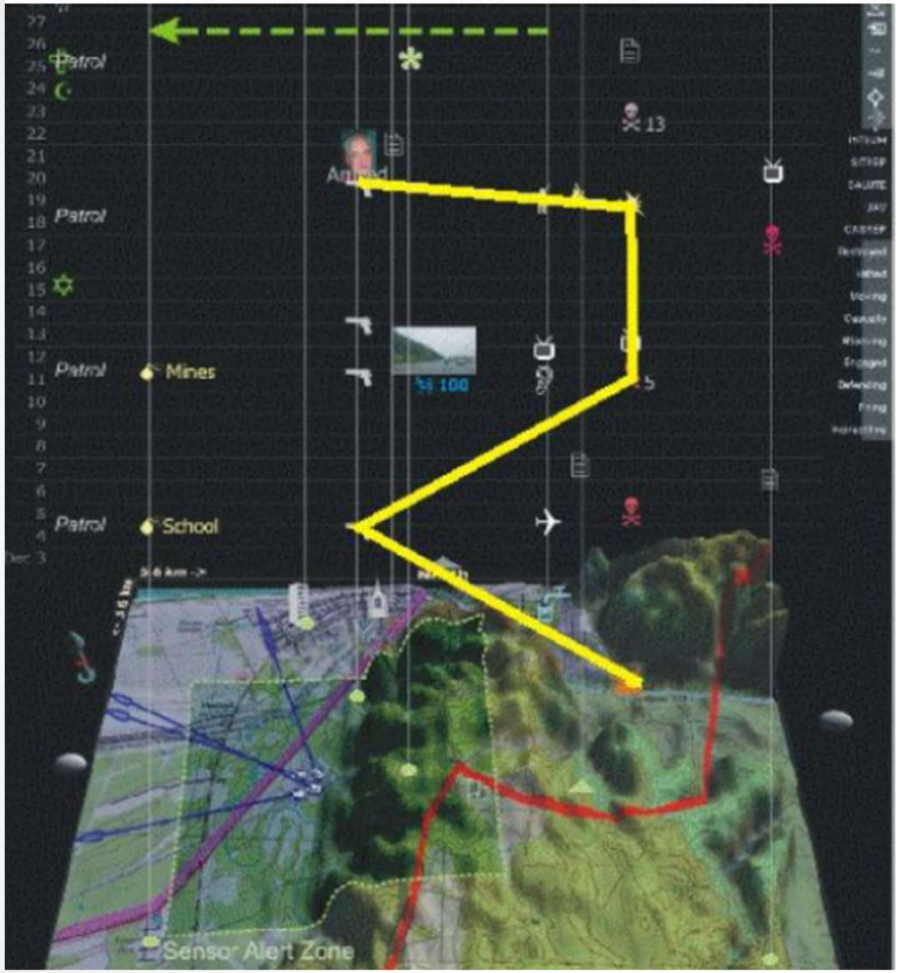
- Typically superimpose temporal events on a map
- Can be utilized for analysis of events spread over geography for a duration
  - Example: Pandemic
- Need techniques for analysis and visualization of spatio-temporal data

GeoTime

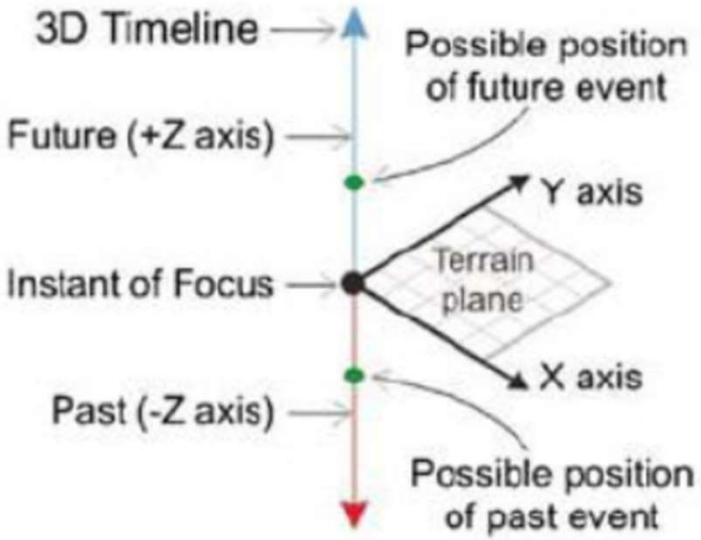
- Represent place by 2D plane (or maybe 3D topography)
- Use 3rd dimension to encode time
- Object types:
  - Entities (people or things)
  - Locations (geospatial or conceptual)
  - Events (occurrences or discovered facts)

Kapler & Wright InfoVis '04

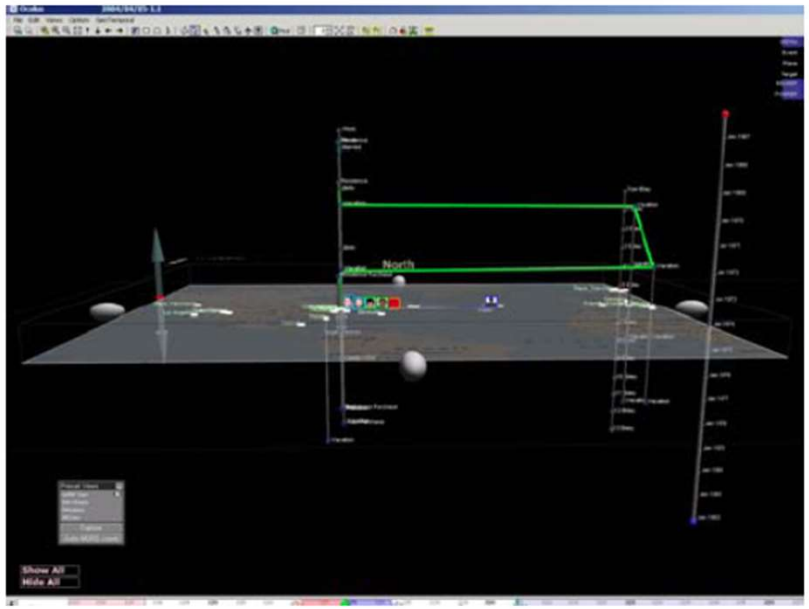
GeoTime Axis



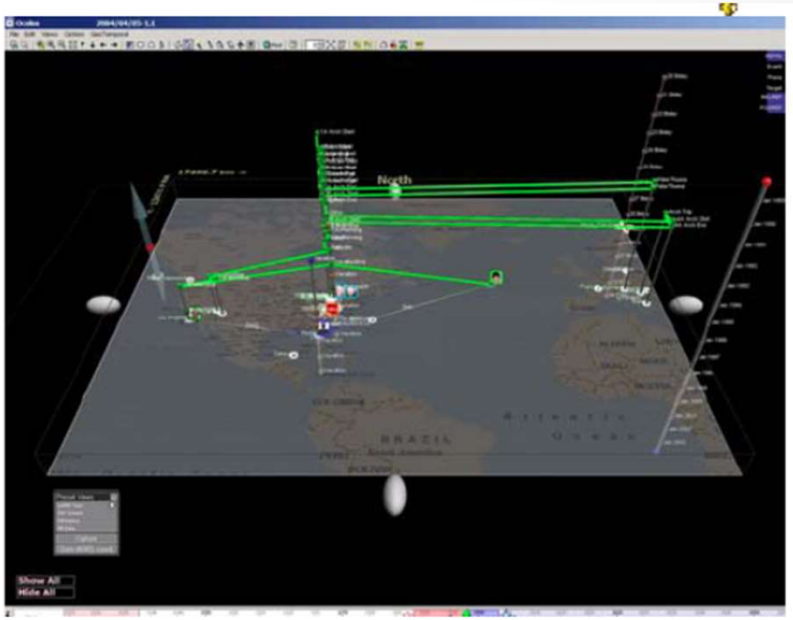
- Vertical time axis (z-axis)
- Terrain plane (xy



GeoTime Interactions



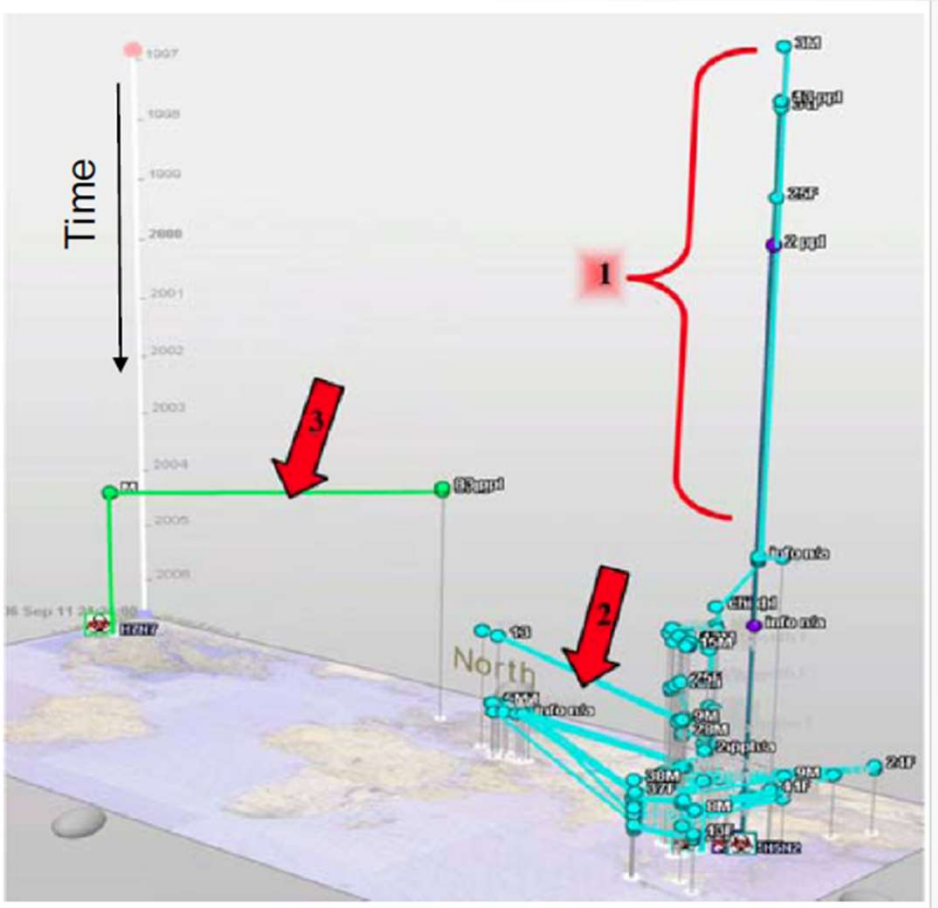
Time slider at bottom
Moveable time scale at right
Green line traces one entity's movement in time and geography



Overhead view
Time slider advanced from view on left

Avian Flu Spread

1. Worldwide human cases of avian flu. Sporadic localized activity seen in Asia until 2003.
2. H5N1 cases increased in frequency in 2004 with the bulk of disease activity in Asia.
3. Isolated incidents of strain H7N7 found in Canada and Europe



- Proulx et al, IEEE VAST 2008

Readings

- Wolfgang Aigner, Silvia Miksch, Wolfgang Mu¨ller, Heidrun Schumann, and Christian Tominski. Visual Methods for Analysing Time-Oriented Data. IEEE Transactions on Visualization and Computer Graphics (Volume: 14, Issue: 1), Jan.-Feb. 2008