Chapter 7 - Time-Based Data

Visualizing Change Over Time

Vorlesung „Informationsvisualisierung“
Prof. Dr. Florian Alt, WS 2013/14

Konzept und Folien (4th revised edition):
Thorsten Büring, Andreas Butz, Michael Burch
Outline

• Motivation and Definitions
• Categorisation for Time Dimension
  – Characteristics of the Time Axis
  – Criteria for the „What is analyzed?“ question
  – Criteria for the „How is it represented?“ question
• Examples
  – Historic time-based graphics
  – Timeline representations in Information Visualization
  – Use case: time-based computer desktops
  – Searching time-based data
Motivation and Definitions
The Crowdsourced Metropolis

What a Hundred Million Calls to 311 Reveal About New York

By Steven Johnson  November 1, 2010 | 12:00 pm | Wired November 2010

There were 54,522 complaints called in to 311 between September 8 and September 15, 2010. Here are the most common, plotted by time of day.

Illustration: Pitch Interactive
The Crowdsourced Metropolis

New Yorkers are accustomed to strong odors, but several years ago a new aroma began wafting through the city’s streets, a smell that was more unnerving than the usual offenders (trash, sweat, urine) precisely because it was so delightful: the sweet, unmistakable scent of maple syrup. It was a fickle miasma, though, draping itself over Morningside Heights one afternoon, disappearing for weeks, reemerging in Chelsea for a few passing hours before vanishing again. Fearing a chemical warfare attack, perhaps from the Aunt Jemima wing of al Qaeda, hundreds of New Yorkers reported the smell to authorities. The New York Times first wrote about it in October 2005; local blogs covered each outbreak, augmented by firsthand reports in their comment threads.

The city quickly determined that the odor was harmless, but the mystery of its origin persisted for four years. During maple syrup events, as they came to be called, operators at the city’s popular NYC311 call center—set up to field complaints and provide information on school closings and the like—were instructed to reassure callers that they could go about their business as usual.

But then city officials had an idea. Those calls into the 311 line, they realized, weren’t simply queries from an edgy populace. They were clues.

On January 29, 2009, another maple syrup event commenced in northern Manhattan. The first reports triggered a new protocol that routed all complaints to the Office of Emergency Management and Department of Environmental Protection, which took precise location data from each syrup smell. Within hours, inspectors were taking air quality samples in the affected regions. The reports were tagged by location and mapped against previous complaints. A working group gathered atmospheric data from past syrup events: temperature, humidity, wind direction, velocity.

Seen all together, the data formed a giant arrow aiming at a group of industrial plants in northeastern New Jersey. A quick bit of shoe-leather detective work led the authorities to a flavor compound manufacturer named Frutarom, which had been processing fenugreek seeds on January 29. Fenugreek is a versatile spice used in many cuisines around the world, but in American supermarkets, it’s most commonly found in the products on one shelf—the one where they sell cheap maple-syrup substitutes.

Definition

• A time series is a sequence of data points, measured typically at successive time instants
• In general, this data takes the following form:
  \[ D = \{(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)\} \]
  where \( y_i = f(t_i) \) for all \( 1 \leq i \leq n \)
• Function \( f \) maps time steps to instances of data of a special type, e.g. quantitative values, or more complex data such as graphs
• Function \( f: \) 
  Natural numbers \( \rightarrow \) Datatype
• Examples for quantitative time series data:
  – Daily closing value of the Dow Jones index
  – Annual flow volume of the Nile River at Aswan
  – Seismogram
  – Electrocardiogram
Which data types can vary over time?

• Answer: Generally, all of them!
  – Quantitative (time series)
  – Ordinal (time series)
  – Categorical (time series)
  – Relational/Hierarchical (time series)
  – Textual (time series)
  – Multivariate (time series)
  – Meta (time series)
  – ...
  – Any combination thereof

• The difficulty is: „Find a reliable visual metaphor for a given time-series value of a specific data type!“

• Further Examples: Sunspot activity, baseball games, medicines taken, cities visited, stock prices, newswires, network resource measures, water levels, …
Which data types can vary over time?

• A chronological component is a fundamental property of a time series dataset
  – Single data records of a time series dataset can be ordered by their chronological occurrences
    • Order in the dataset MUST also be visible in the visualisation
    • Number of occurrences in each possible subinterval MUST also be visible in the visualisation
  -> Watch out for the Lie Factor
Which data types can vary over time?

• Edward Tufte
  – Take a random sample of 4000 graphics from 15 of world’s newspapers and magazines from 1974 until 1980
  – Result: About 75 percent of graphics published there focus on time series data

• Today: Similar results
  – One reason for this phenomenon is that people (readers from newspapers and magazines) always want to compare the actual numbers, values, and relations with older ones to understand their current situation.
  – Second reason: Learn from the past to forecast the future!
    • A specific known dynamic pattern may indicate a consequence.
    • Earthquake, Heart attack, …

• 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?
Exercises to think about

• How would you visually encode these datasets?
  – Dataset 1:
    • Time 1, sunshine intensity, temperature
    • Time 2, sunshine intensity, temperature
    • Time m, sunshine intensity, temperature
  – Dataset 2:
    • Day 1, 5 news articles about Clinton, 7 news articles about oil, and 2 news about Iraq
    • Day 2, 3 news articles about Clinton, 3 news articles about oil, and 5 news about Iraq
    • Day m, 7 news articles about Clinton, 1 news article about oil, and 6 news about Iraq
  – Dataset 3:
    • Lisa was born in Worcester in 2002, she weighted 11 lbs at that time,
      – she went to Austin in 2004 for 1 week, she weighted 23 lbs at that time,
      – she moved to Charlotte in 2005, she weighted 28 lbs at that time,
      – she visited China in 2006 for one month, she weighted 30 lbs at that time.
  
• How would you visually encode time-varying multivariate datasets with the goal to visually explore the data for dynamic patterns? This is a real big challenge!
Exercises to think about

• Other important 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?

• Important questions about times series data (Stasko)
  – 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
Exercises to think about

• Three practical question categories:
  – What are the characteristics of the time axis?
  – What is analysed?
  – How is it represented?

• These three questions correspond to the categorisation criteria:
  – time
  – data
  – representation
Categorisation for Time Dimension
What are the characteristics of the time axis?
Temporal Primitives

- Time points vs. Time intervals
  - Time axis can be composed of time points or time intervals
  - Time point considered an instant in time
  - In contrast, a time interval is a temporal primitive with an extent. It can be specified by two time points or by a time point plus a duration.
  - When reasoning about time, the question of whether time points or time intervals are considered is decisive.
  - Different relations are possible among time points and among time intervals.
  - Accordingly, different analysis tasks or goals can be accomplished depending on the addressed temporal primitives.
Structure of Time

• Linear vs. Cyclic vs. Branching
  – Three different structures distinguished
  – Linear time corresponds to our natural perception of time as being a (totally or partially) ordered collection of temporal primitives, i.e., time proceeds from the past to the future.
  – A cyclic time axis is composed of a finite set of recurring temporal primitives (e.g., the seasons of the year).
  – On a cyclic time axis, any temporal primitive A is proceeded and succeeded at the same time by any other temporal primitive B (e.g., winter comes before summer, but winter also succeeds summer).
  – In practical applications it is often useful to unroll a cyclic time axis to a linear time axis.

(a)  
(b)  
(c)
Structure of Time

• Linear vs. Cyclic vs. Branching
  – Branching time axes are modelled as graphs.
  – Temporal primitives are the vertices of the graph.
  – Directed edges describe temporal order.
  – Vertices with more than one outgoing edge indicate a split of the time axis into alternative scenarios, which is particularly relevant for planning or prediction.
  – Apparently, linear time and cyclic time can be seen as special cases of branching time where the graph obeys certain constraints (i.e., for linear time, every vertex has no more than one outgoing edge; for cyclic time, the graph is a circle).
  – Linear and cyclic time are covered well by existing visual and analytical approaches.
  – However, methods for analyzing branching time are still rare.
  – This must be in focus of future work in Visual Analytics.
Structure of Time

• Linear vs. Cyclic vs. Branching
  – Decision to which category a time-oriented dataset belongs is not always fully determined, but depends on the interpretation of the user, on the task, or on the application.
  – If for instance a user seeks to find a general trend in the data, a linear interpretation of the time axis makes sense.
  – On the other hand, detecting seasonal effects in the data can be easier if a cyclic time axis is assumed.

(Weber et al., Visualizing Time-Series on Spirals, 2001)
Criteria for the „What is analysed?“ question
Data Context: Abstract vs. Spatial data

• By abstract data we mean data that have been collected in a non-spatial context, i.e., data that are not per se connected to some spatial layout.

• In contrast to that, spatial data contain an inherent spatial layout, which can be conditioned by natural circumstances or modelled realities.

• The distinction between abstract and spatial data reflects the crystallisation of different subfields of visualisation research in the last decade.

• Information visualisation, graph visualisation, or software visualisation are more concerned with abstract data, whereas spatial data are addressed by scientific visualisation (flow visualisation, volume visualisation) or geographic information systems.

• Each field handles time-oriented data differently, despite the fact that a unified view would be more desirable.
Data Context: Abstract vs. Spatial data

- The main reason for a distinction between abstract and spatial data is the way of how data are processed in Visual Analytics.
- For spatial data, the inherent spatial information can be exploited to find a suitable mapping of data to screen. The representation of time has to be incorporated into that mapping, where it is not always easy to achieve an emphasis of the time domain.
- For abstract data, no a priori spatial mapping is given. On the one hand, that implies, it is first of all necessary to contrive an expressive spatial layout. This requires creative thinking and experience. On the other hand, screen dimensions can be used almost exclusively to expose the time domain.
Number of variables: Univariate data vs. Multivariate data

• When speaking of variables, we do not limit our consideration to basic data types like integers, real numbers, or categorical enumerations. We also consider a vector, a matrix, or a news article as possible data variables if this is required by the application at hand.

• Obviously, it makes a difference if we have to represent data where each temporal primitive is associated with a single data value (i.e., univariate data) or if multiple data values (i.e., multivariate data) must be considered.

• With the latter case, an additional visualisation goal – the detection of correlations – is introduced.

• Approaches for single-valued data have been around for a long time.

• There are also various techniques that allow the visualisation of two or three data values (which are literally already multivariate).

• Big Challenge in Visual Analytics: handle larger numbers of variables.

• This is where analytical methods come into play. Usually, it is necessary to apply dimension reduction methods (e.g., principle component analysis) to derive major temporal trends.
Level of abstraction: Data vs. Data abstractions

• Edward Tufte claims: “Above all else, show all data.”
• The majority of visual methods follow that claim.
• Visualising data is useful in many application scenarios. However, if larger data sets must be analysed, Tufte’s postulation is hard to fulfil without introducing new problems like overcrowded and cluttered displays.
• In such cases, it makes sense to melt down the data to condensed form, i.e., to derive data abstractions that reflect interests and needs of users.
• Calculating aggregated data values is one example for deriving abstractions, which is particularly useful to drive overview+detail interfaces.
• Feature visualisation also follows the idea of computing data abstractions. Features are data portions that obey certain user-defined constraints.
• In the context of time-oriented data a third derivable information unit must be mentioned – events. Events are special situations in the development of time-oriented data. Events can be user-defined or found by methods of Artificial Intelligence.
Time dependency: Static vs. Dynamic

• Static representations visualise time-oriented data in static images (i.e., representations that do not change automatically over time).

• In contrast to that, dynamic representations utilise the physical dimension time to convey the time dependency of the data (i.e., representations that change automatically over time such as slide shows or animations).

• The presence or absence of interaction facilities has no influence on whether a visualisation approach is categorised as static or dynamic.

• Distinguishing between static and dynamic representations is crucial for Visual Analytics, because different tasks and goals are supported.

• Dynamic representations are well suited to convey the general development of the analysed data over time.

• However, there are also critical voices on animation.

• Especially when longer multivariate time series have to be visualised, animation-based approaches reach their limits.

• Users simply cannot follow all changes in the visual representation and the animation takes too long for the user to remember its course.
Time dependency: Static vs. Dynamic

• Static representations show all information on one screen, which is advantageous to fully concentrate on the data and to compare different parts of the time axis.

• However, in contrast to animations, static representations require screen real estate to represent the time axis itself.

• Therefore, it is challenging to develop representations that avoid visual clutter.
Criteria for the „How is it represented?“ question
Dimensionality: 2D vs. 3D

- The question of whether or not it makes sense to exploit three dimensions for visualisation is discussed heavily in the community.
- One camp of researchers argues that two dimensions are sufficient for effective data analysis. In their thinking the third dimension involves unnecessary difficulties like occlusion and lost information on back faces.
- The other camp of researchers sees the third dimension as a possibility to encode further information.
- Undoubtedly, certain types of data (e.g., flow data or volume data) even require the third dimension for expressive data visualisation.
- The mentioned disadvantages of a three dimensional presentation space are tackled with advanced interaction techniques or additional visual cues.
- Both options are required depending on task and data at hand.
Examples
Imports and Exports

• Playfair 1785
Train Schedule Paris - Lyon

- Marey 1885 (from Tufte 2001)
ThemeRiver

• Havre et al. 2002
• River metaphor for visualizing thematic variations over time within a document collection
  – 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?“
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
  - 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

(Revision as of 22:16 4 Jun 2003)

“Abortion” in its most commonly used sense, refers to the voluntary termination of a pregnancy. Abortion may be spontaneous, 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 a few weeks after conception, 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 medical abortion is the usual method; the drug mifepristone is usually the only legal method, although research has uncovered similar effects from methotrexate and misoprostol. Concomitant with medical abortion and ending up until around the fifteenth week suction aspiration or vacuum aspiration 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 ensure abortion in the third trimester; premature expulsion of the fetus can be induced with prostaglandins. This can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought about by the controversy intact dilation and extraction (D & X) or a hysterotomy abortion, similar to a cesarean 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 also been opposed by some institutionalized religious and governments. In the 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, a fetus (sometimes called the “unborn” by pro-life/anti-abortion advocates) is 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

(Revision as of 22/S 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 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 misoprostol is usually the only legal method. Although research has uncovered similar effects from methotrexate and mifepristone, concurrent with chemical abortion and extending up until around the fifteenth week, vacuum abortion is the most common method, 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 prostaglandins, this can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought about by the controversial intact dilation and extraction (D & I) 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 religious groups 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, a treated, 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, a fetus is a being (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/](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