Cohort Comparison of Event Sequences with Balanced Integration of Visual Analytics and Statistics

Sana Malik, Fan Du, Megan Monroe[*]
University of Maryland
College Park, MD 20740
{maliks, fan, madeyjay}@cs.umd.edu

Eberechukwu Onukwugha
University of Maryland
Baltimore, MD 21201
eonukwug@rx.umaryland.edu

Catherine Plaisant, Ben Shneiderman
University of Maryland
College Park, MD 20740
{plaisant, ben}@cs.umd.edu

ABSTRACT

Finding the differences and similarities between two datasets is a common analytics task. With temporal event sequence data, this task is complex because of the many ways single events and event sequences can differ between the two datasets (or cohorts) of records: the structure of the event sequences (e.g., event order, co-occurring events, or event frequencies), the attributes of events and records (e.g., patient gender), or metrics about the timestamps themselves (e.g., event duration). In exploratory analyses, running statistical tests to cover all cases is time-consuming and determining which results are significant becomes cumbersome. Current analytics tools for comparing groups of event sequences emphasize a purely statistical or purely visual approach for comparison. This paper presents a taxonomy of metrics for comparing cohorts of temporal event sequences, showing that the problem-space is bounded. We also present a visual analytics tool, CoCo (for "Cohort Comparison"), which implements balanced integration of automated statistics with an intelligent user interface to guide users to significant, distinguishing features between the cohorts. Lastly, we describe two early case studies: the first with a research team studying medical team performance in the emergency department and the second with pharmacy researchers.

Author Keywords

temporal data; cohort comparison; visual analytics

INTRODUCTION

Sequences of timestamped events are currently being generated across nearly every domain of data analytics. Consider a typical e-commerce site tracking each of its users through a series of search results and product pages until a purchase is made. Or consider a database of electronic health records containing the symptoms, medications, and outcomes of each patient who is treated. Every day, this data type is reviewed by humans who apply statistical tests, hoping to learn everything they can about how these processes work, why they break, and how they can be improved upon.

Human eyes and statistical tests, however, reveal very different things. Statistical tests show metrics, uncertainty, and statistical significance. Human eyes see context, accountability, and most notably, things that they may not have even been looking for.

Visualization tools strive to capitalize on these latter, human strengths. For example, the EventFlow visualization tool [39] supports exploratory, visual analyses over large datasets of temporal event sequences. This support for open-ended exploration, however, comes at a cost. The more that a visual analytics tool is designed around open-ended questions and flexible data exploration, the less it is able to effectively integrate automated, statistical analysis. Automated statistics can provide answers, but only when the questions are known.

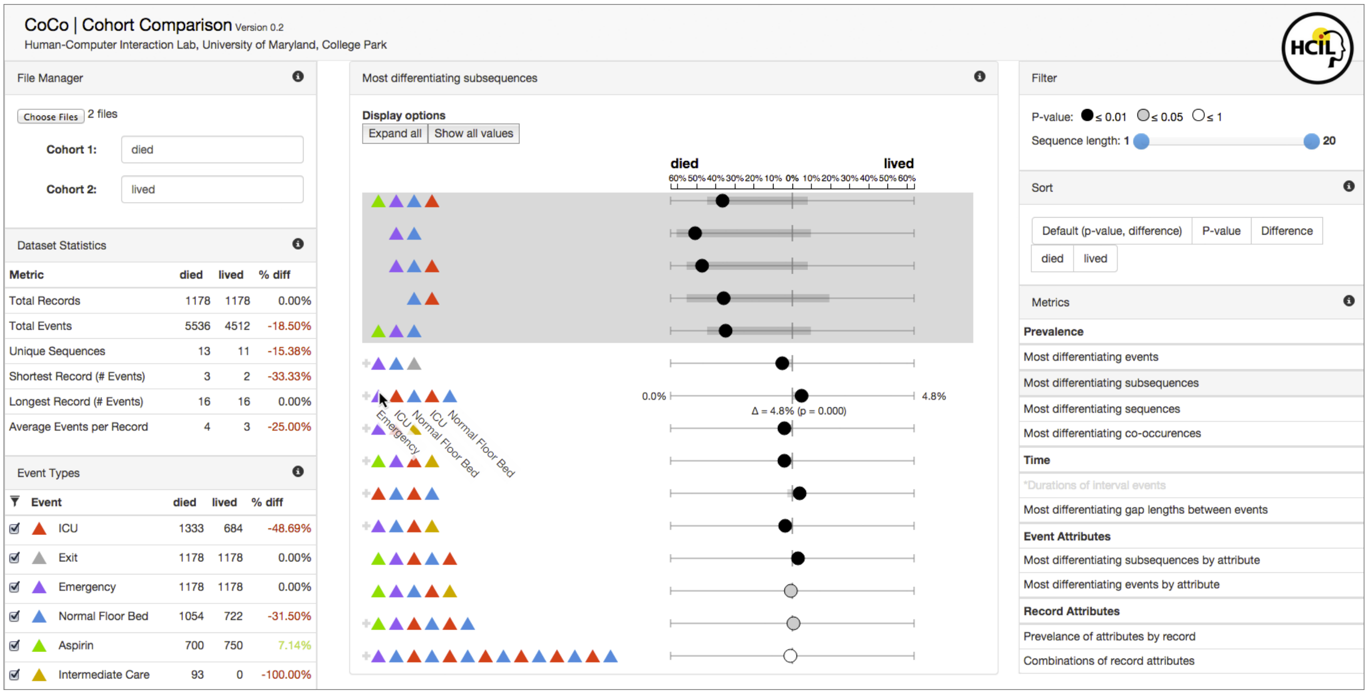
The opportunity to combine these two approaches lies in the middle ground. By all accounts, the goal of open-ended questions is to generate more concrete ones. As these questions come into focus, so too does the ability to automatically generate the answers. This paper introduces CoCo (for "Cohort Comparison", Figure 1), a visual analytics tool that is designed to capitalize on one such scenario.

Consider again the information that is tracked on an e-commerce site. From a business perspective, the users of the site fall into one of two groups: people who bought something and people who did not. If the goal is to convert more of the latter into the former, it is critical to understand how these two groups, or cohorts, are different. Did one group look at more product pages? Or spend more time on the site? Or have some clear demographic identifier such as gender, race, or age? Similar questions arise in the medical domain as well. Which patients responded well to a given medication? How are did their treatment patterns differ the patients who didn't?

Although comparing two groups of data is a common task, with temporal event sequence data in particular, the task of running many statistical tests becomes complex because of the variety of ways the cohorts, sequences (entire records), subsequences (a subset of events in a record), and events can differ. In addition to the structure of the event sequences (e.g., order, co-occurrences, or frequencies of events), the attributes about the events and records (e.g., gender of a patient), and

[*] Current Address: IBM T. J. Watson Research Center, Cambridge, MA 02142

IUI 2015, March 29–April 1, 2015, Atlanta, GA, USA.
Copyright © 2015 ACM 978-1-4503-3306-1/15/03 ...$15.00.
http://dx.doi.org/10.1145/2678025.2701407



CoCo | Cohort Comparison Version 0.2
Human-Computer Interaction Lab, University of Maryland, College Park

File Manager

Choose Files 2 files

Cohort 1: died

Cohort 2: lived

Dataset Statistics

Metric	died	lived	% diff
Total Records	1178	1178	0.00%
Total Events	5536	4512	-18.50%
Unique Sequences	13	11	-15.38%
Shortest Record (# Events)	3	2	-33.33%
Longest Record (# Events)	16	16	0.00%
Average Events per Record	4	3	-25.00%

Event Types

	Event	died	lived	% diff
☑	ICU	1333	684	-48.69%
☑	Exit	1178	1178	0.00%
☑	Emergency	1178	1178	0.00%
☑	Normal Floor Bed	1054	722	-31.50%
☑	Aspirin	700	750	7.14%
☑	Intermediate Care	93	0	-100.00%

Most differentiating subsequences

Display options
Expand all | Show all values

Filter

P-value: ● ≤ 0.01 ○ ≤ 0.05 ○ ≤ 1
Sequence length: 1 – 20

Sort

Default (p-value, difference) | P-value | Difference
died | lived

Metrics

Prevalence
Most differentiating events
Most differentiating subsequences
Most differentiating sequences
Most differentiating co-occurences
Time
*Durations of interval events
Most differentiating gap lengths between events
Event Attributes
Most differentiating subsequences by attribute
Most differentiating events by attribute
Record Attributes
Prevelance of attributes by record
Combinations of record attributes

Figure 1. CoCo combines automated statistical analysis with an intelligent user interface to enable insights, hypothesis generation, and data exploration when comparing two groups of temporal event sequences. Users are provided with pre-defined metrics (bottom right) as a starting point for their exploration and they are able to parse results with a visualization and interactions such as sorting and filtering. In this example, we have two groups of patients as they are transferred throughout a hospital: those who lived and those who died. The selected metric is the most differentiating subsequences. We can see that being transferred from the emergency room (purple) directly to the normal floor bed (blue) appears statistically significantly more in the group of patients who died ($p \leq 0.01$).

the timestamps themselves (e.g., an event's duration) can be distinguishing features between the cohorts. For this reason, running statistical tests to cover all these cases and determining which results are significant becomes cumbersome. Additionally, the factor on which the cohorts are formed may call for different types of questions to be asked about the data. For example, in a set of medical records split by date (e.g., last month's trials vs. this month's), a research may be interested in how outcomes for the patients differ between the cohorts, whereas a dataset split by the patient's outcome (e.g., patients who die vs. those who live) would ignore such a metric.

Current tools for cohort comparison of temporal event data (described in the next section) emphasize one of two strategies: 1) purely visual comparisons between groups, with no integrated statistics, or 2) purely statistical comparisons over one or more features of the dataset. By contrast, CoCo is designed to provide a more balanced integration of both human-driven and automated strategies. We begin by showing that the task of cohort comparison is specific enough to support automatic computation against a bounded set of potential questions and objectives. From this starting point, we demonstrate that the diversity of these objectives, both across and within different domains, as well as the inherent complexities of real world datasets, still require human involvement to determine meaningful insights. Through case studies, we look at how CoCo can support the task of cohort comparison more specifically than previous visualization efforts.

The direct contributions of this paper are:

1. A taxonomy of metrics for comparing groups of temporal event sequences.
2. A visual analytics tool which demonstrates balanced integration of automated analysis and user-guided analysis with an intelligent user interface.
3. Case studies that illustrate the benefits of CoCo's utility while suggesting further refinements.

On a broader level, the goal of this paper is to highlight the relationship between task specificity and the ideal balance between humans and statistical analysis, so that future efforts can better leverage the strengths of both approaches.

RELATED WORKS

Visualizing Groups of Sequential Data

Work on visualization of sequential data is described here in two parts: visualizations of a single group of event sequences and visualizations comparing two or more sequences.

Single Groups

EventFlow [39] (Figure 2) and OutFlow [56] create simplified visualizations of collections of event and interval sequences. Both tools aggregate a single cohort and the complete sequences of records, with EventFlow allowing users to view details about individual records as well. While they only support visualizing a single group of records, comparison of multiple cohorts can be facilitated by using multiple instances

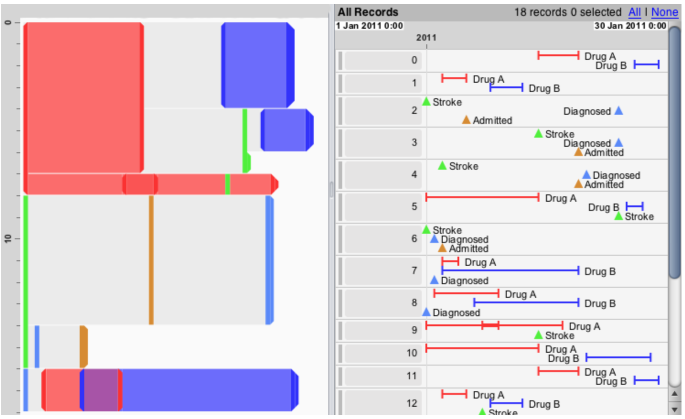


Figure 2. EventFlow visualizes event sequences in an aggregated, tree-like overview and as individual records in a timeline.

of the visualization and visual inspection by the user. These tools do not provide statistical information about the differences. CoCo borrows some event icon motifs from EventFlow (such as triangles for point events and T-shaped markers for interval events).

Event Sequence Comparison

Solutions for comparing sequential data have been explored in many different fields, including comparative genomics, text mining, and tree comparison. They are discussed here in the context of event history data and discrete-time models [3].

We draw first on methods to compare collections of general sequences without the notion of time, most notably the fields of comparative genomics and text mining, where the data is ordered with respect to some index [32].

Genome browsers [13, 21, 29, 52, 55, 12] visualize genome sequences. They compare genomes by visualizing the position of each nucleotide and consider a genome as a long and linear sequence of nucleotides. Scientists also compare genomes at the gene level. However, most existing tools are able to compare either only the similarities or only the differences of collections of gene sequences. For example, MizBee [37] measures the similarity between genomes by visualizing the regions of shared sequences, while Variant View [20], cBio [11] and MuSiC [17] only support displaying sequence variants. Further, genome sequences are often compared as a sequence of linear *positions*, which does not lend itself to distinctions between point and interval events.

Texts are often compared by extraction of frequent n-grams [9]. FeatureLens by Don et al. [18] define an n-gram as a contiguous sequence of words and use a visualization approach to compare the co-occurrences of frequent n-grams in texts. However, it only supports locating n-grams with specific features but does not find which n-grams are the most differentiating. Jankowska et al. [28] proposed the conversion of documents into vectors of frequent character n-grams and designed a relative n-gram signature to encode the distance between n-gram vectors. Viégas et al. presented history flow [54] to visually compare between versions of a document, which assumes that the later version of a document is developed based on the earlier one, which is not applicable to event history data.

Most of the techniques mentioned above (in both genomics and text mining) only provide a visual comparison between single long sequences, whereas event history data consists of large numbers of independent transactional sequences.

Temporal event sequences are often represented as trees. While many comparison techniques exist for trees, many do not take into account values or attributes of nodes and none are specifically designed for temporal data. Munzner presented the TreeJuxtaposer [42] system to help biologists explore structural details of phylogenetics, but focuses only on structural differences in the trees and not any attributes about the nodes (such as timestamps). Bremm [8] studied the comparison of phylogenetic trees in a more statistical way by extending the algorithms of TreeJuxtaposer to compare more than two trees and considers "edge length" which could be generalized to durations of gaps between sequential events. Holten [27] presented an interactive visualization method to compare different versions of hierarchically organized software systems. He proposed two methods of tree comparison: icicle plot and hierarchical sorting, but does not propose any statistical comparison technique, and focuses more on "leaf-to-leaf" matching, which considers whole paths (or sequences) only. TreeVersity2 [24] compares by tree structure and the node values. Though TreeVersity2 is general to all trees, it leaves out temporal-specific analysis such as duration of or between interval events. TreeVersity2 compares two datasets over time, but assumes these time periods are disjoint. TreeVersity2 also includes a textual reporting tool that highlights outliers in the data.

Many of these comparison techniques also lack statistical tests for the comparisons. In our work, the balanced integration supports both visual and statistical approaches.

Statistics for Comparing Cohorts

In medical cohort studies, the most prevalent approach for comparison is survival analysis, where survival time is defined as the time from a defined point to the occurrence of a given event [7]. The Kaplan-Meier method is often used to analyze the survival time of patients on different treatments and to compare their risks of death [14, 19, 23]. Based on the Kaplan-Meier estimate, survival time of two groups of patients can be visualized and compared with survival curves, which plot the cumulative proportion surviving against the survival times [7]. Also, the log-rank test is often used to statistically compare two survival curves by testing the null hypothesis. Compared with survival analysis, the event sequences data used in our work is much more complicated, and requires a more advanced analysis model.

Currently tools that combine visualization and statistics for medical cohort analysis focus on single cohorts. CAVA [57] is a visualization tool for interactively refining cohorts and performing statistics on a single group. Recently, Oracle published a visualization tool for cohort study [45]. Based on patients' clinical data, it supports interactive data exploration and provides statistics as well as visualization functionalities.

These tools similarly focus on combining visualization with automated statistics and providing an interactive interface for selecting cohorts; however, both tools aim at grouping and identifying patient cohorts for further characterization, while our work focuses on comparing two existing cohorts based on their event histories.

Temporal Data Mining

Previous work studying temporal data mining has mostly focused on discovering frequent temporal patterns [4, 6, 15, 43, 33, 22, 2, 36, 46] and computing temporal abstractions [30, 41, 5] of time-oriented data.

Pattern discovery is an open-ended problem which aims to unearth all patterns of interest [32]. Much of the literature is concerned with developing efficient algorithms to automatically discover frequent temporal patterns and extract temporal association rules [4, 6, 15, 33, 22, 2, 36]. To constrain the search procedure, some algorithms [4, 15] allow users to provide initial knowledge and rules. To show the results, Norén et al. [43] used a graphical approach to visualize temporal associations. This work can be extended to address unique temporal constraints, such as dealing with concurrent events, which Cule et al. address in the context of pattern mining [16] and association rule mining [50].

Temporal abstraction focuses on obtaining a succinct and meaningful description of a time series [41]. Various approaches have been proposed. Klimov et al. [30] developed VISITORS to visualize patient records by grouping the event attribute values at different temporal granularities. Moskovitch et al. [41] aggregated values of point data by state and trend, to obtain its interval representation. Batal et al. [5] converted time series data into vectors of frequent patterns, which can be used with standard vector-based algorithms. However, most of the work in this topic only focused on the time and changes in an event's value (a concept), which is considered as event attributes in our work. Tatti and Vreekan [51] introduce a novel algorithm for summarizing a set of a sequences by providing a descriptive and non-redundant set of sequences, accounting for long gaps.

METRICS FOR COMPARING COHORTS

Metrics for comparing cohorts are numerous and can be grouped into five main categories: summary metrics, time metrics, event sequence (both whole record sequences and subsequences thereof) metrics, event attribute metrics, and record attribute metrics. These metrics are a direct result of observing EventFlow users as they analyzed cohorts of event sequences in seven case studies performed over three years [38]. Five case studies were in the health care domain (with pharmacists and epidemiologists), one in sports analytics (basketball), and one in transportation.

Summary Metrics

Summary statistics deal with the cohorts as a whole and provide a high-level overview of the datasets.

Number of records Total number of records in each cohort.

Number of events Total number of events in each cohort.

Number of unique records Total number of unique records in each cohort based on the sequence of events (absolute times are not considered).

Number of each event Total number of occurrences for each event type per cohort.

Minimum, Maximum, and Average length of records The length of a record is considered as the number of events per that record.

Event Sequence Metrics

Event sequence metrics deal with the order and structure of event sequences. Sequences are differentiated by whether they occur as an entire *sequence* in a record or a *subsequence* of a record. Each of the following metrics can be presented as the percent of records containing the event or sequence or as the percent of all events or sequences that it occurs. The former method provides a sense of how many individual records had this sequence occur, whereas the latter method provides a sense of how events or sequences might repeat themselves within one record.

Prevalence of an event The percent of records or total number of events that a particular event occurs in.

Prevalence of a subsequence The percent of records in which the subsequence appears. For example, patients who lived are given aspirin before going to the emergency room more often than the patients who died.

Prevalence of a whole sequence Percent of records with a given sequence.

Order of sequential events in a subsequence The percent of records containing event A directly preceding event B versus B preceding A. For example, perhaps patient who go to the ICU before the floor are more likely to live than patients who have these events in the reverse order.

Commonly Co-occurring (non-consecutive) events The percent of records containing both events A and B (in any order, with any number of events between them).

Prevalence of Outcomes If a single event is prevalent as an "outcome" (i.e., the last event in the sequence). This metric in particular applies only to cohorts that are not already split on an outcome event.

Time Metrics

Time metrics deal with the timestamps at both the event and sequence levels – relative and absolute.

Absolute time of an event Prevalence of a particular timestamp of an event or multiple events (e.g., if all events in one cohort occurred on the same day).

Duration of interval events The duration of particular interval event. For example, this can be the length of exposure to a treatment or the duration of a prescription.

Duration between sequential events The time between the end of one event and the beginning of the next. For example, the average length of time between hospital patients entering the emergency room and being transferred to the ICU is under two hours in patients who lived and over two hours in those who died.

Duration between co-occurring (non-sequential) events The length of time between non-sequential events (two events with some number of other events occurring between them).

Duration of a subsequence The length of time from the beginning of the first event in a subsequence to the end of the last event in the subsequence.

Duration from a fixed point in time The length of time from a user-specified, fixed point – aligned by either a selected event or absolute date-time.

Duration of overlap in interval events The overlap (or lack thereof) of interval events. For example, the overlap of Drug A and Drug B could be more common in the cohort of patients who lived versus those who died.

Cyclic events and sequences The duration between cyclic events and sequences.

Survivor analysis How an event or sequence occurs or diminishes over time.

Statistics for each of these metrics include the minimum, maximum, and average durations or values and the distributions of the values between the cohorts.

Event Attribute Metrics

Any of the above metrics can be applied over values of an attribute of the events instead of the event type itself. This can be done by swapping an event type by the values of a particular attribute. For example, in a medical dataset, we might be interested in seeing how a particular emergency room doctor might be related to the outcome of a patient. We would then switch all events of type "Emergency" with the value of its "doctor" attribute. If there are three doctors, this would create 3 new pseudo-event types. We can use the metrics from above to see the difference in event sequences, times, or prevalence of each doctor in either cohort.

Record Attribute Metrics

Record level attributes (such as patient gender or age) compare the cohorts as population statistics. General statistics across the entire dataset is a problem already tackled by analytics tools such as Spotfire [53] or Tableau [1], however these tools look at a single attribute. For example, they might compare the number of males versus females or patients on Wednesday versus Thursday. There may be implications about the *combinations* of record attributes (e.g., the women on Wednesday versus the women on Thursday versus the men on Wednesday versus the men on Thursday). In clinical trials, it is important that all patient attributes are balanced and currently no tools exist for visually confirming that all attribute combinations are balanced.

Combining Metrics

The number of metrics is further multiplied because any combination of the above metrics is a new metric. For example, a sports analytics researcher may be interested in how a particular player (as an attribute of an event) performs within two minutes (time) after halftime (event order).

BALANCING AUTOMATION WITH HUMAN INTERACTION

Purely statistical methods of comparison would benefit from user intervention. With the sheer number of metrics, it would be computationally time consuming to run every metric ahead of time, especially when not every metric may be required for analysis. Users with domain knowledge about the datasets would ideally be able to select from the metrics and easily eliminate unnecessary metrics. Further, questions asked during cohort comparison may vary based on how the cohorts were divided. If the cohorts were divided by outcome (e.g., patients who lived versus patients who died), the sequence of events leading up to them becomes more important. Analysis might revolve around determining what factors (time or attributes) or events lead to the outcome by determining how the metrics differ between the groups. Conversely, if the cohorts were split based on an event type, questions may revolve around finding distinguishing outcomes (e.g., patients who took Drug A may result in more strokes than patients who took Drug B). Exploration of cohorts that are split by time (e.g., the same patients over two different months) may be more open-ended and require all metrics. The cohorts can be distinguished by time factors, event attributes, or events themselves (sequences of events or outcomes).

Our contribution is to enable researchers to be far more flexible in examining cohorts and facilitate human intervention where it can save time and effort. Because of the pre-defined problem space of comparing temporal event sequences, we can save users time by having answers to common questions readily available and giving them a starting point for their exploration.

Purely visual tools for temporal event sequences are a good starting point for developing analysis tools for cohort studies, but can be improved by the inclusion of the statistical tests used in automated approaches. For example, EventFlow assumes that each patient record consists of time-stamped point events (e.g. heart attack, vaccination, first occurrence of symptom), temporal interval events (e.g. medication episode, dietary regime, exercise plan), and patient attributes (e.g. gender, age, weight, ethnic background, etc.). In multiple case studies with EventFlow, the researchers repeatedly observed users visually comparing event patterns in one group of records with those in another group. In simple terms the question was: what are the sequences of events that differentiate one group from the other? A common aspiration is to find clues that lead to new hypotheses about the series of events that lead to particular outcomes, but many other simple questions also involved comparisons. Epidemiologists analyzing the patterns of drug prescriptions [40] tried to compare the patterns of different classes of drugs. Hospital administrators looking at patient journeys through the hospital compared the data of one month with the previous month. Researchers analyzing task performance during trauma resuscitation [10] wanted to compare performance between cases where the response team was alerted of the upcoming arrival of the patient or not alerted. Transportation analysts looking at highway incident responses [25] wanted to compare how an agency handled its incidents differently from another. Their observations suggest that some broad insights can be gained by visually comparing pairs of EventFlow displays (e.g., users could see if the patterns were very similar overall between one month and the next) or very different (e.g., a lot more red or the most common patterns were different) but users repeatedly expressed the desire for more systematic ways to compare cohorts of records.

COHORT COMPARISON WITH COCO

Though CoCo can be used in a variety of fields, the synthetic dataset used as an example for the remainder of the paper consists of records of patients admitted to the emergency room and follows their movement through their stay at the hospital: being administered aspirin, being admitted into the hospital room, transferring between a normal floor bed, intermediate care, and the intensive care unit (ICU), and ultimately being discharged either dead or alive. The dataset is split into two cohorts: patients who died and patients who lived.

Design

Based on the case studies which shaped the taxonomy, our three design goals for balanced integration were:

G1. Automatic, efficient computation of metrics.

G2. Guided process for reading results.

G3. Visualization and interaction techniques for parsing and sorting results.

We used an iterative design process based on feedback from on-going case study partners and a user study [35]. We now describe the first operational CoCo prototype, organized by those three goals.

G1: Computation of Metrics

In automatically applying the metrics to the datasets, we must consider (1) what are the appropriate statistical tests for each metrics, and (2) how to compute the results quickly for the user.

Towards the first goal, the user should be involved in selecting appropriate significant tests for the metrics. Currently, CoCo implements only non-parametric tests and thus does not provide methods for selecting parametric tests. Percent prevalence and attribute significances are calculated by Chi-squared and the time significance metrics use a Wilcoxon sum-rank test across the distribution of values.

To date, we implemented two methods for automatically computing these metrics on a large set of data. The first was to apply every metric after the datasets are loaded, in order to rank the metrics from potentially most meaningful to least. However, this resulted in a long wait time for users. The current implementation computes a metric as the user selects it. However, this offers less guidance than if the metrics were pre-computed. Our future goal is to minimize wait time, but give prompt feedback on which metrics might be meaningful to look at immediately, in accordance with Stolper et al.'s design guidelines for progressive visual analytics [49].

G2: Interface

To guide users to most meaningful results, we organize results by first providing high-level dataset statistics followed by specific metrics grouped by type.

CoCo consists of a file manager pane, a dataset statistics pane, an event legend, a list of available metrics, the CoCo visualization, and options for filtering and sorting the results (Figure 1).

The summary statistics panel includes high-level statistics about both datasets, including the total number of records and events in each record. Users are then shown the Event Type pane, which serves as the legend (pairing each event type with a marker and color) and filter control. When an event is checked or unchecked, it displays or hides rows containing that event from the analysis. Frequencies of each event in the two cohorts are also shown, as the raw number of occurrences of that event type.

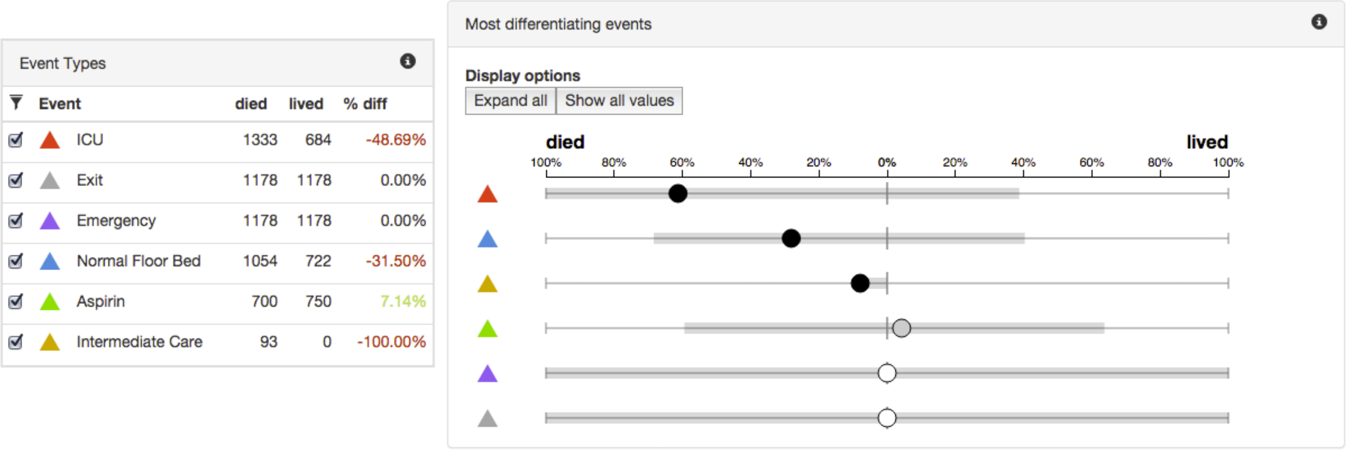
The right panel consists of filtering and sorting mechanisms and a list of metrics. The metrics panel contains the list of metrics is organized according to the taxonomy. We found that many users started with metrics dealing with prevalence to understand how events and sequence occur within the datasets before moving onto metrics dealing with time. Similarly, within each group, the metrics are organized from simplest to more complex: singular event metrics, subsequence metrics, then finally whole record metrics.

The filtering and sorting panels provide ways for users to parse the results of their selected metric. Users may filter by p-value or sequence length. The default sorting behavior for the results is first by p-value group, then by magnitude of difference. Users may also choose to sort by only p-value, only absolute difference, or by most differentiating in cohort α or by β.

Preliminary versions of CoCo introduce each panel one-by-one, so users could click-through as they finished analyzing each section. However, feedback from users suggested this was more cumbersome than helpful and we opted to display all panels at once.

G3: CoCo Visualization

The visualization needs to convey the differences themselves (e.g., proportion of records in each cohort containing a sequence), but also the result of the statistical test. Since the meaning of the results differ between results (e.g., some metrics refer to a percentage of records while others refer to a time duration), it was also important to visualize the results in a way that supports both types of results. Because users are more focused on the *difference* of a value between two cohorts, we chose to use a back-to-back bar chart in order to emphasize the magnitude and direction of the difference, so users can more easily scan across multiple rows for results they are interested in.



Event Types

		Event	died	lived	% diff
☑	▲	ICU	1333	684	-48.69%
☑	▲	Exit	1178	1178	0.00%
☑	▲	Emergency	1178	1178	0.00%
☑	▲	Normal Floor Bed	1054	722	-31.50%
☑	▲	Aspirin	700	750	7.14%
☑	▲	Intermediate Care	93	0	-100.00%

Most differentiating events

Display options

Figure 3. The CoCo visualization shown with the event legend. In this example, we see that ICU, Normal Floor Bed, and Intermediate Care events occur significantly more frequently in the "died" cohort than in the "lived." Because of the nature of the dataset, 100% of records in both cohorts contain the "Emergency" and "Exit" events, so there is no significant difference and the circle marker is placed in the middle.

The CoCo visualization (Figure 3) displays the results of significance tests in a unified form. For each event or sequence of events, CoCo displays the value of the metric in each cohort (e.g., percent of prevalence, gap duration), the difference between the two values, and the significance value of the difference.

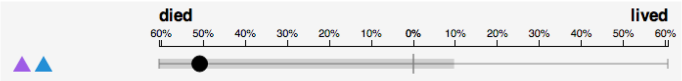


Figure 4. The sequence "Emergency" followed by a "Normal Floor Bed" event occurs in about 60% of records in "died" and only 10% of records in "lived" for a statistically significant difference of 50% ($p \leq 0.01$)

Each row (Figure 4) consists of a horizontal axis, where the left is cohort α and the right is cohort β (the labels can be renamed by users). A semi-transparent bar grows from the middle towards each direction in the respective cohort to show the value of the metric for that particular event or sequence. The axis is scaled by the maximum value for all sequences (e.g., if the maximum percentage is 60%, the maximum value on the axis will also be 60%). The axis works for both percentage and time. The circle marker is placed horizontally based on the difference between the values, in the direction of whichever cohort's value is higher. The circle marker is filled corresponding to the significance of the difference: black is used for p-values with a significance of less than or equal to 0.01, grey for less than or equal to 0.05 but higher than 0.01, and white for values over 0.05 up to 1.

Rows are ranked first by their p-value group (with records with the most significant p-value appearing first), and within each of the three significance groups, rows are ordered by the absolute percent difference between the two groups. Users can filter out records by p-values in a certain group with the legend on the top right of the visualization.

Hovering over any sequence displays an informational tooltip (Figure 5), which gives the event names for each event marker, the corresponding value in each cohort, the values' difference, and the exact p-value. Users can choose to always display these values in the display options in the top left of the visualization (Figure 3).

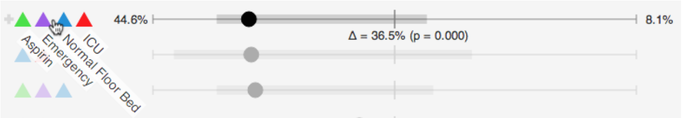


Figure 5. Exact values for each row are shown in the hover tooltip. The sequence "Aspirin," "Emergency," "Normal Bed Floor," then "ICU" appears in 44.6% of records in "died" and 8.1% of records in "lived."

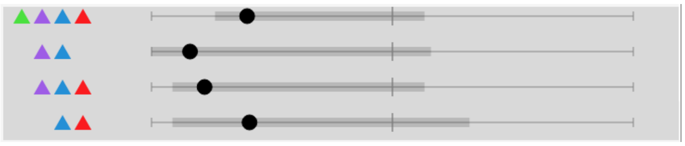
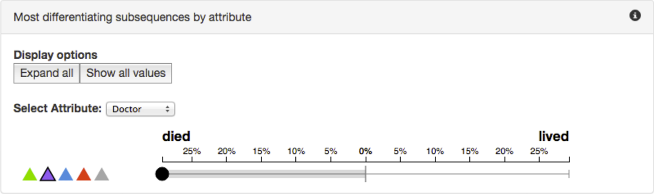


Figure 6. Rows whose sequence are a subsequence of another row are aggregated together and only the longest sequence is shown. The nested records are expanded on click and are visually grouped with the rest of the aggregated rows in the group on a dark background.

With rows that are sequences of events, there are sometimes shorter subsequences of that sequence that have a similar significance value. In these cases, CoCo aggregates rows when a sequence is a subsequence of another row and both event sequences fall within the same significance category. An aggregated row is indicated with a "+" indicator to the right of the sequence markers (Figure 5) and is expanded on when the top-level (nested) row is clicked. Nested subsequences of the row are aligned with the marker and displayed on a darker background color with the top-level record (Figure 6). Users can expand or collapse all nested records at once in the display options to the top left of the visualization (Figure 3).

In metrics dealing with attributes, users can select a particular attribute from a drop-down under the display options in the visualization. Events that contain a value with the selected attribute are outline with a black border and the value of the attribute is shown on hover as the value appended to the event type with a pipe ("|") separator (Figure 7).



Most differentiating subsequences by attribute

Display options

Select Attribute: Doctor

Figure 7. Users can select an attribute from the drop-down in the display options. Events containing an attribute value are outlined in black and its value is shown in the hover tooltip. In this example, the attribute selected is the doctor that is on call in the emergency room (purple). The event containing the attribute is highlighted in black.

G3: Interactions for Parsing Results

Though CoCo initially sorts the results based on the significance and magnitude of difference in the results, in preliminary sessions with case study partners, it became clear that researchers might have more specific questions that require displaying the results differently. CoCo allows users to sort and filter the results based on their needs. Specifically, the sorting method may be changed based on:

- Magnitude of difference and p-value group (default)
- P-value only
- Magnitude of difference only
- Most differentiating towards the alpha group
- Most differentiating towards the beta group

To further filter out potentially noisy data, we added controls for filtering by p-value or sequence length. For example, if users are interested in a particular subset of events, they can use the event type legend to uncheck events that they are not interested in. Then, only sequences containing the checked event types will appear.

PRELIMINARY CASE STUDIES

Case Study: Exploring Adherence to Advanced Trauma Life Support Protocol

To investigate the strengths and limitations of CoCo as an automated cohort comparison tool, we conducted a case study following the procedure of a Multi-Dimensional, Long-term In-depth case study (MILC) [48]. We worked with medical researchers at Children's National Medical Center who were investigating trauma teams' adherence to the Advanced Trauma Life Support (ATLS) protocol and possible reasons for deviations. In a previous study [10], they found that about 50% of resuscitations did not follow the ATLS protocol. As a follow-up, the researchers collected additional data about the resuscitation process after implementation of an ATLS compliance checklist to re-evaluate the attributes associated with protocol deviations. Specifically, the researchers' questions were:

1. What percent of patients are treated in adherence to protocol?
2. Are there distinguishing attributes (e.g., time of day, patient gender, team lead) between protocol adherence and non-adherence?
3. What are the most common deviations from the protocol?

After an initial training session and interviews to understand the researchers' goals and questions, we observed the researchers as they conducted a 3-hour session of data exploration and analysis.

The dataset consisted of 181 patient records, with event types for the five steps in the ATLS protocol: airway evaluation, listening for breath sounds, assessment of circulation, evaluation of neurological status disability, and temperature control. Patient attributes included injury severity score (ISS), the day of week, length of hospital stay, time between notification and arrival at the hospital, and if the patient was admitted to the hospital.

The dataset was stored as a single file. They used EventFlow's "group by attribute" feature to split the dataset into separate cohorts based on attributes and adherence to the protocol. Over the course of the 3 hours, the researchers split the dataset in six ways to load six different pairs of cohorts in Coco as they explored different hypotheses:

1. Patients treated in adherence to the ATLS protocol versus those that showed any deviation.
2. Patients admitted to the floor versus ICU (with discharged patients removed).
3. Patients who arrived with at least five minutes warning before arrival at the trauma bay versus those who arrived with no warning ("now" patients).
4. Patients with a high (above 25) versus low ISS.
5. Patients treated on the weekend versus on a weekday.
6. Patients treated during the day versus at night.

In every comparison group, the analysts began by looking at the prevalence of single events, to determine how often they occurred. The analysts then looked at the most differentiating *entire record* sequences, because the subsequences were less informative about how the protocol was followed. They would then make their way down the provided metrics list, in the order that they appeared: most differentiating time gaps and then prevalence of record attributes. They did not look at the prevalence of record attribute combinations for any of the datasets.

For this dataset, they expected to see that all records contained every event. This finding was not observed for two of the comparisons: correctly treated patients versus those with deviations and day versus night patients, with the latter of both groups receiving the airway check significantly less that daytime patients. In the day versus night group, the analyst also found that the "most differentiating sequence" was the *correct* order, meaning that the nighttime patients were treated in the correct order significantly less than daytime patients. Additionally, patients treated at night had more variance in the procedure, with 26 unique sequences in the 83 patients versus 20 unique sequences in the 101 daytime patients. A possible reason for this finding is that during the day, nurse practitioners perform these procedures, but at night, less experienced junior residents are on-call instead.

At times, the researchers saw that certain groups occurred only rarely in the cohorts (under 20 times), so the researchers decided not to consider the comparisons. For example, among patients admitted to the ICU or floor, only about 80 patients remained, making the sample sizes too small to run many of the significance metrics about event types. As one analyst worked to confirm her expectations and check several hypotheses, she found a surprising and potentially important result: about 25% more patients who were admitted to the floor were "now" patients ($p < 0.05$), which led to splitting the cohort into the third group: now versus not now patients.

In the closing interview, one analyst said, *"We don't need to solve everything with EventFlow and CoCo. These tools let us explore the data and narrow our hypothesis."* From these results, the analysts submitted abstracts about and presented these findings at an internal symposium on trauma care.

Additional case studies and targeted controlled studies will be necessary to characterize the effectiveness of CoCo, but this first case study suggests that CoCo can be effective for exploratory analysis and hypothesis generation.

On-going Case Study: Comparing Algorithms for Distinguishing Types of Radiation to the Bone

We are currently working with our partners at the Department of Pharmaceutical Health Services Research at the University of Maryland School of Pharmacy in Baltimore. In previous work, the researchers were interested in developing an algorithm using claims data to differentiate between radiation delivered to the bone versus radiation delivered to the prostate gland, because billing codes available in claims data do not distinguish the site of radiation. Reliable measures for identifying the receipt of radiation to the bone are important in order to avoid bias in estimating the prevalence and/or mortality impact of skeletal-related events, including radiation to the bone.

Studies using healthcare claims employ various claims-based algorithms to identify radiation to the bone and mostly condition on prior claims with a bone metastasis diagnosis (billing) code [47, 44, 31]. They developed three classification algorithms that were compared using CoCo and EventFlow to investigate the timing of possible radiation to the bone among patients diagnosed with incident metastatic and nonmetastatic prostate cancer. One algorithm was based on prior literature while the other two were based on insights gained from data visualization software. Based on clinical input regarding the duration of palliative [26, 34] versus curative radiation, the researchers investigated the length of radiation episodes and found differences between cohorts in terms of the length of radiation. As expected, patients diagnosed with metastatic disease received shorter course radiation than patients diagnosed with nonmetastatic disease.

The feedback on CoCo was positive and the team valued the opportunity to visually compare cohorts of patients using summary statistics that pertained to the timing and frequency of events. The graphical results were shared with clinicians on the research team in order to determine whether the patterns were consistent with their expectations. The researchers felt the meaning of metrics could be explained more clearly; it was sometimes unclear what the x-axis represented and what statistical tests were used. They also suggested always showing the event labels, particularly for single-event metrics, to make understanding the icons a bit easier. The researchers expressed a need to be able to sort the rows of results with different factors, including by raw percentage of values in each cohort. We implemented this feature before the formal case study.

We are also starting to work on case studies in other application domains such as transportation. For example, we started a project with the Baltimore Metropolitan Council and are currently preparing and cleaning data with EventFlow so that their analysts can then use CoCo to compare how different jurisdictions are managing highway incidents or how their incident management has changed over the years.

CONCLUSIONS AND FUTURE WORK

CoCo is a novel visual analytics tool with balanced integration of visual analytics and statistics. CoCo's benefits include: better collaboration among colleagues, easier intermediate results discussion, and meaningful outcome presentations. Though CoCo was initially designed for expert users, primarily in healthcare, the taxonomy can be extended and refined for other specific domains, and our approach for the interface and visualizations would allow extensions in many ways. First, the metrics implemented are already proving valuable, but many more metrics are possible. Our current focus has been identifying sequential, contiguous subsequences in the datasets, but generalizing to identify any co-occurring, non-contiguous events is a natural next step (e.g., did patients have more than three aspirins at any time during their treatment) which would require more research. The comments by our colleagues in the usability study and the way they compared the cohorts motivate CoCo design improvements including new visualizations. Additional data mining and statistical techniques could be added to improve insight discovery, such as anomaly detection to find unusual records or clustering find similar records between the datasets. As we continue developing CoCo, we will conduct controlled experiment to understand its strength and weaknesses, as well as long-term case studies with domain experts to demonstrate value with realistic problems and to guide our development.

We recognize that there are limitations to CoCo in terms of the complexity of datasets, current emphasis on two cohorts, and the need for more user control on which events to study. On the other hand, the fresh possibilities for statistical comparisons, supported by visual presentations and an intelligent user interface, opens many doors for further research. While we are encouraged by our initial feedback, we see a huge set of possible features to add, which will empower medical and other researchers as they conduct exploratory data analysis on temporal event sequences.

ACKNOWLEDGEMENTS

The authors wish to thank Chris Imbriano and Seth Powsner for their input during the development of CoCo. We also thank our partners James Gardner at the School of Pharmacy at University of Maryland, Baltimore and Randall Burd and Rachel Webman at Children's National Medical Center in Washington, DC. This work is supported in part by Oracle. We gratefully acknowledge funding provided by The University of Maryland/Mpowering the State through the Center for Health-related Informatics and Bioimaging.

REFERENCES

1. Tableau software. `http://www.tableausoftware.com/`, Mar 2014.
2. Agrawal, R., and Srikant, R. Mining sequential patterns. In *Proc. 11th International Conference on Data Engineering*, IEEE Comput. Soc. Press (1995), 3–14.
3. Allison, P. D. Discrete-time methods for the analysis of event histories. *Sociological Methodology 13*, 1 (1982), 61–98.
4. Álvarez, M. R., Félix, P., and Cariñena, P. Discovering metric temporal constraint networks on temporal databases. *Artificial Intelligence in Medicine 58*, 3 (July 2013), 139–54.
5. Batal, I., Sacchi, L., Bellazzi, R., and Hauskrecht, M. A temporal abstraction framework for classifying clinical temporal data. *Proc. AMIA Annual Symposium 2009* (Jan. 2009), 29–33.
6. Bellazzi, R., Sacchi, L., and Concaro, S. Methods and tools for mining multivariate temporal data in clinical and biomedical applications. *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009* (Jan. 2009), 5629–32.
7. Bewick, V., Cheek, L., and Ball, J. Statistics review 12: survival analysis. *Critical Care (London, England) 8*, 5 (Oct. 2004), 389–94.
8. Bremm, S., von Landesberger, T., Hess, M., Schreck, T., Weil, P., and Hamacherk, K. Interactive visual comparison of multiple trees. In *Proc. 2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2011), 31–40.
9. Brown, P. F., DeSouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. Class-based n-gram models of natural language. *Computational Linguistics 18*, 4 (Dec. 1992), 467–479.
10. Carter, E., Burd, R., Monroe, M., Plaisant, C., and Shneiderman, B. Using eventflow to analyze task performance during trauma resuscitation. *Proceedings of the Workshop on Interactive Systems in Healthcare (WISH 2013)* (2013).
11. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery 2*, 5 (May 2012), 401–4.
12. Chelaru, F., Smith, L., Goldstein, N., and Bravo, H. C. Epiviz: interactive visual analytics for functional genomics data. *Nat Meth 11*, 9 (Sept. 2014), 938–940.
13. Chen, Y., Cunningham, F., Rios, D., McLaren, W. M., Smith, J., Pritchard, B., Spudich, G. M., Brent, S., Kulesha, E., Marin-Garcia, P., Smedley, D., Birney, E., and Flicek, P. Ensembl variation resources. *BMC genomics 11*, 1 (Jan. 2010), 293.
14. Collett, D. *Modelling survival data in medical research*. CRC press, 2003.
15. Concaro, S., Sacchi, L., Cerra, C., Fratino, P., and Bellazzi, R. Mining health care administrative data with temporal association rules on hybrid events. *Methods of Information in Medicine 50*, 2 (Jan. 2011), 166–79.
16. Cule, B., Tatti, N., and Goethals, B. Marbles: Mining association rules buried in long event sequences. *Statistical Analysis and Data Mining 7*, 2 (2014), 93–110.
17. Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., Mooney, T. B., Callaway, M. B., Dooling, D., Mardis, E. R., Wilson, R. K., and Ding, L. MuSiC: identifying mutational significance in cancer genomes. *Genome Research 22*, 8 (Aug. 2012), 1589–98.
18. Don, A., Zheleva, E., Gregory, M., Tarkan, S., Auvil, L., Clement, T., Shneiderman, B., and Plaisant, C. Discovering interesting usage patterns in text collections. In *Proc. 16th ACM Conference on Conference on Information and Knowledge Management - CIKM '07*, ACM Press (New York, USA, Nov. 2007), 213.
19. Dupont, M., Gacouin, A., Lena, H., Lavoué, S., Brinchault, G., Delaval, P., and Thomas, R. Survival of patients with bronchiectasis after the first ICU stay for respiratory failure. *Chest 125*, 5 (May 2004), 1815–20.
20. Ferstay, J. A., Nielsen, C. B., and Munzner, T. Variant view: visualizing sequence variants in their gene context. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2546–55.
21. Fiume, M., Williams, V., Brook, A., and Brudno, M. Savant: genome browser for high-throughput sequencing data. *Bioinformatics (Oxford, England) 26*, 16 (Aug. 2010), 1938–44.
22. Fournier-Viger, P., Faghihi, U., Nkambou, R., and Nguifo, E. M. CMRules: Mining sequential rules common to several sequences. *Knowledge-Based Systems 25*, 1 (Feb. 2012), 63–76.
23. Goel, M. K., Khanna, P., and Kishore, J. Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research 1*, 4 (Oct. 2010), 274–8.

24. Guerra-gómez, J. A., Pack, M. L., Plaisant, C., and Shneiderman, B. Visualizing changes over time in datasets using dynamic hierarchies. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2566–2575.

25. Guerra-gómez, J. A., Wongsuphasawat, K., Wang, T. D., Pack, M. L., and Plaisant, C. Analyzing incident management event sequences with interactive visualization. *Proceedings of the Transportation Research Board 90th annual meeting* (2011).

26. Hartsell, W. F., Scott, C. B., Bruner, D. W., Scarantino, C. W., Ivker, R. A., Roach, M., Suh, J. H., Demas, W. F., Movsas, B., Petersen, I. A., Konski, A. A., Cleeland, C. S., Janjan, N. A., and DeSilvio, M. Randomized trial of short- versus long-course radiotherapy for palliation of painful bone metastases. *Journal of the National Cancer Institute 97*, 11 (2005), 798–804.

27. Holten, D., and van Wijk, J. J. Visual Comparison of Hierarchically Organized Data. *Computer Graphics Forum 27*, 3 (May 2008), 759–766.

28. Jankowska, M., Keselj, V., and Milios, E. Relative N-gram signatures: Document visualization at the level of character N-grams. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, IEEE (Oct. 2012), 103–112.

29. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, a. D. The Human Genome Browser at UCSC. *Genome Research 12*, 6 (May 2002), 996–1006.

30. Klimov, D., Shahar, Y., and Taieb-Maimon, M. Intelligent visualization and exploration of time-oriented data of multiple patients. *Artificial Intelligence in Medicine 49*, 1 (May 2010), 11–31.

31. Lage, M., Barber, B. L., Harrison, D. J., , and Jun, S. The Cost of Treating Skeletal-Related Events in Patients With Prostate Cancer. *Am J Manag Care 14*, 5 (2008), 317–322.

32. Laxman, S., and Sastry, P. S. A survey of temporal data mining. *Sadhana 31*, 2 (Apr. 2006), 173–198.

33. Lee, Y. J., Lee, J. W., Chai, D. J., Hwang, B. H., and Ryu, K. H. Mining temporal interval relational rules from temporal data. *Journal of Systems and Software 82*, 1 (2009), 155–167.

34. Lutz, S. T., Jones, J., and Chow, E. Role of radiation therapy in palliative care of the patient with cancer. *Journal of Clinical Oncology* (2014).

35. Malik, S., Du, F., Monroe, M., Onukwugha, E., Plaisant, C., and Shneiderman, B. An evaluation of visual analytics approaches to comparing cohorts of event sequences. In *EHRVis Workshop on Visualizing Electronic Health Record Data at VIS '14* (2014).

36. Mannila, H., Toivonen, H., and Verkamo, A. I. Discovery of Frequent Episodes in Event Sequences. *Data Mining and Knowledge Discovery 1*, 3 (Sept. 1997), 259–289.

37. Meyer, M., Munzner, T., and Pfister, H. MizBee: a multiscale synteny browser. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Jan. 2009), 897–904.

38. Monroe, M. *Interactive Event Sequence Query and Transformation*. PhD thesis, University of Maryland, "2014".

39. Monroe, M., Lan, R., Lee, H., Plaisant, C., and Shneiderman, B. Temporal event sequence simplification. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (Dec 2013), 2227–2236.

40. Monroe, M., Meyer, T. E., Plaisant, C., Lan, R., Wongsuphasawat, K., Coster, T. S., Gold, S., Millstein, J., and Shneiderman, B. Visualizing patterns of drug prescriptions with eventflow: A pilot study of asthma medications in the military health system. *Proceedings of Workshop on Visual Analytics in Healthcare (VAHC 2013)* (2013).

41. Moskovitch, R., and Shahar, Y. Medical temporal-knowledge discovery via temporal abstraction. *Proc. AMIA Annual Symposium 2009* (Jan. 2009), 452–6.

42. Munzner, T., Guimbretière, F., Tasiran, S., Zhang, L., and Zhou, Y. TreeJuxtaposer: Scalable Tree Comparison using Focus+Context with Guaranteed Visibility. In *ACM SIGGRAPH 2003*, no. 1, ACM Press (New York, USA, 2003), 453.

43. Norén, G. N., Hopstadius, J., Bate, A., Star, K., and Edwards, I. R. Temporal pattern discovery in longitudinal electronic patient records. *Data Mining and Knowledge Discovery 20*, 3 (Nov. 2009), 361–387.

44. Nørgaard, M., Jensen, A. Ø., Jacobsen, J., Cetin, K., Fryzek, J., and Sørensen, H. Skeletal related events, bone metastasis and survival of prostate cancer: a population based cohort study in Denmark (1999 to 2007). *J Urol. 184*, 1 (2010), 162–167.

45. Oracle. Oracle Health Sciences Cohort Explorer User's Guide. Tech. rep., Oracle, 2011.

46. Perer, A., and Wang, F. Frequence: Interactive mining and visualization of temporal frequent event sequences. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, ACM (New York, USA, 2014), 153–162.

47. Sathiakumar, N., Delzell, E., Morrisey, M., Falkson, C., Yong, M., Chia, V., Blackburn, J., Arora, T., and Kilgore, M. Mortality following bone metastasis and skeletal-related events among patients 65 years and above with lung cancer: A population-based analysis of U.S. Medicare beneficiaries, 1999-2006. *Lung India 30*, 1 (2013), 20–26.

48. Shneiderman, B., and Plaisant, C. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time*

and Errors: Novel Evaluation Methods for Information Visualization, BELIV '06, ACM (New York, NY, USA, 2006), 1–7.

49. Stolper, C., Perer, A., and Gotz, D. Progressive visual analytics. In *To appear in IEEE Transactions on Visualization and Computer Graphics*, TVCG (2014).

50. Tatti, N., and Cule, B. Mining closed episodes with simultaneous events. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, ACM (New York, NY, USA, 2011), 1172–1180.

51. Tatti, N., and Vreeken, J. The long and the short of it: Summarising event sequences with serial episodes. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, ACM (New York, NY, USA, 2012), 462–470.

52. Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics 14*, 2 (Mar. 2013), 178–92.

53. TIBCO. Spotfire. `http://spotfire.tibco.com/`, Mar 2014.

54. Viégas, F. B., Wattenberg, M., and Dave, K. Studying cooperation and conflict between authors with history flow visualizations. In *Proc. 2004 Conference on Human Factors in Computing Systems - CHI '04*, ACM Press (New York, USA, Apr. 2004), 575–582.

55. Wang, J., Kong, L., Gao, G., and Luo, J. A brief introduction to web-based genome browsers. *Briefings in Bioinformatics 14*, 2 (Mar. 2013), 131–43.

56. Wongsuphasawat, K., and Gotz, D. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2659–2668.

57. Zhang, Z., Gotz, D., and Perer, A. Interactive Cohort Analysis and Exploration. *Journal of Information Visualization (IVS), to appear.* (2014).