Seeing Sequences: How Current Temporal Visualization Techniques Fail in Complex Domains

Katie Minardo
The MITRE Corporation
202 Burlington Road, Bedford, MA, 01730
kminardo@mitre.org
781-271-3244

ABSTRACT
While temporal reasoning is a key component of situational awareness, most visualization and HCI research focuses on time-series rather than time-sequence datasets. This paper outlines the unique challenges in sequence visualization and defines the gap in existing visualization and workflow techniques.

Author Keywords
Information visualization, complexity, timeline analysis, temporal data, time-series data, sequences.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

INTRODUCTION

Timestamps are a ubiquitous component of datasets. When linked together, these individual timestamps provide critical sequencing and trend information that is otherwise invisible. Sequencing information is used to facilitate planning, to compare courses of action, track dynamic situations and provide forensic analysis.

Temporal reasoning is a critical factor in maintaining situational awareness and making optimal decisions. Temporal sequencing improves user comprehension and projection [6] by allowing users to extend their understanding of the current situation into the past (*How did we get to this point?)* and into the future (*Where are we headed?).*

UNDERSTANDING THE TWO TYPES OF TIME DATA

Temporal data is grouped into two major categories: series and sequence. Time series data is continuous, tracking changes in particular value over time (e.g., a person's heart rate, or a company stock price). Sequences depict the ordering of events, with durations of time in between (e.g., a schedule of meetings for the day). In terms of visualization, time series data is simpler to visualize, because the set of actors is constant (i.e., the stock price for a particular company), and we can reasonably assume a numerical value exists for every timestep (i.e., the value of the stock). There are many advances for time-series data, to address high volumes of data [8, 11, 13], outline new visualizations [1, 5, 18] or improve the analysis workflow [2, 9, 15, 16].

Few of these innovations, however, translate to non-continuous temporal datasets like sequences. Time sequences, because they are discrete, are much more difficult to visualize. For example, a police investigation is an example of a sequence problem. Different actors are involved at different times, from the initial officers who respond, to the crime-scene investigators who later examine the scene, to the detectives who interview suspects days or weeks later. Subsequences for each actor start and end at different times, the relevant list of actors is dynamic and the "values" are categorical rather than numeric. A visual representation of the actors and their tasks could be useful in a court case to define the steps and timing of the overall investigation.

There has been significant work in characterizing temporal data [4] and categorizing temporal-analysis tasks (i.e., identifying correlations or comparing dimensions) [17, 19]. Sequence visualization, however, is often through a simple Gantt chart, showing a line for each actor and duration for each singular event. This framework is used in both experimental systems [3, 12, 14] and commercial systems [20, 21].

As operational datasets grow in detail complexity, they quickly outpace the existing visualization techniques for sequences. Certainly, the current techniques are adequate for a small problem set [4] or an exploration task [14]. Larger datasets or more complex tasks, however, quickly become intractable with the current tools. Managing an airline fleet is a clear example of detail complexity, where hundreds of planes must be scheduled, serviced and tracked through a battery of different tasks. While Butler [3] addressed the schedule complexity with an automation solution, there was no refinement of the visualization for decision-makers.

Just as there have been minimal advances in visualizing sequences, there are limited efforts at improving the sequence-analysis workflow. Current sequence analysis is often completed via "hunt-and-peck" methods, where users must explore the dataset, identify patterns and synthesize data manually. This one-by-one method quickly becomes intractable for datasets with a high level of detail complexity or dynamic complexity. Analysis and decision-making are reliant on the operator's ability to "select, sort, and organize [the available] information," [10] but the available organizing mechanism is inadequate for complex temporal sequences.

DEFINING THE GAP

Improved sequence visualization will have an immediate impact on complex situations. Problems like tracking a shipping fleet, managing a swarm of unmanned vehicles, or analyzing a corpus of intrusion events requires better sequence visualization than what is currently available. Three complexity challenges that sequences visualizations must address are:

(1) Shift the focus to sequences rather events. Current tools depict the temporal sequences as a series of singular events, without aggregation, clustering or highlighting techniques to depict higher-level patterns. For example, in this type of framework, a truck-delivery route from the warehouse to five delivery points and back again, would be plotted as seven distinct events, each with a start time, end time and duration. Visualization and encoding frameworks must be defined for showing the data of the overall sequence (e.g. the truck), in addition to the individual event level (e.g., the delivery events). Sequence visualization could help managers identify follow-on effects or unexpected consequences at the actor rather than the event level.

(2) Depict meaningful aggregates and sub-aggregates. Currently, only two types of aggregates are available: simple roll-up aggregates (akin to a Gantt chart in MS Project) that provide the start and end time for the subordinate sequences, and volume aggregates that show a level of activity (i.e., there are more deliveries today than yesterday). These are often shown as a summary histogram.

Visualizations are needed that can provide a crisp depiction of *both* the individual actor (i.e., the schedule for an individual truck) and relevant sub-aggregates and aggregates (i.e., schedule constraints for fleet of trucks that share a few shipping bays). The key characteristics of each aggregate will be context-sensitive, but may include: detailed items (i.e, route duration for each truck), summary data (i.e., the number of trucks enroute route at a specific time), statistical summaries from quartiles or averages (i.e., percentage of trucks on schedule). Developing and visualizing these aggregates, at multiple levels, will help operators make sense of the detail complexity.

(3) Require less exploration and synthesis. Although visualizations can ingest larger datasets technically, their workflow continues to rely solely on operator-exploration for the analysis and synthesis, as noted earlier. The current tools assume a workflow where each individual item can be explored and the scope of the overall problem can be easily grasped visually. Users don't have the time to "visually [inspect] an extremely high-resolution dataset"[7]. Additional tools are needed to highlight key areas or scope the exploration. While the overload problem isn't unique to the temporal visualization, it is particularly vexing for visualizations outside the common mapping plane.

CONCLUSIONS

More research on sequence visualization is needed, to identify key sequence features and define an initial framework for visualizing complex sequences and their aggregates. Solutions may require revising techniques from related domains, like time-series or geographic visualization, to increase the volume of data that can be understood without increasing operator workload. Similarly, representative tasks and datasets must be defined for a realistic level of operational complexity. These complex tasks and high-scale data must be tested to ensure the usefulness and scalability of the proposed techniques.

ACKNOWLEDGEMENTS:

The author would like to thank collaborators and reviewers of this work including: Curtis Brown, Kevin Cabana, Erika Darling, Neal Rothleder, and Beth Yost.

This paper was approved for public release; distribution unlimited (07-1225).

REFERENCES

1. Bottger, J., Balzer, M., Deussen, O. "Complex Logarithmic View for Small Details in Large Contexts," IEEE Transactions on Visualization and Computer Graphics, Vol 12, No 5, Sept/Oct 2006.
2. Buring, T., Gerken, J., Reiterer, H. "User Interaction with Scatterplots on Small Screens, A comparative evaluation of Geometric-Semantic Zoom and Fisheye Distortion," IEEE Transactions on Visualization and Computer Graphics, Vol 12, No 5, Sept / Oct 2006.

3. Butler, K., et al. "Work-Centered Design of Mixed Initiative Scheduler," In *Proc. CHI 2007*, ACM Press (2007).
4. Chittaro, L. & Combi, C., "Representation of Temporal Intervals and Relations: Information Visualization Aspects and their Evaluation," IEEE Symposium on Temporal Representation and Reasoning, 2001.
5. Craig, P., Kennedy, J. "Coordinated Graph and ScatterPlot Views for the Visual Exploration of Microarray Time-Series Data," IEEE Symposium on Information Visualization, 2003.
6. Endsley, M., Bolte, B., and Jones, D. *Designing for Situation Awareness: An Approach to User-Centered Design*, Taylor & Francis (2003), pp. 19.
7. Hamann, Bern. Comments at SIGGRAPH 1999 Panel on *"Visualization of Large Scale Data; Do we have a solution yet?"* funded by NSF and the Department of Energy.
8. Hao, M.C., Dayal, U., Keim, D.A., and Schreck, T. "Importance-Driven Visualization Layouts for Large Time Series Data," IEEE Symposium on Information Visualization, 2005.
9. Heer, J., and Viegas, F., "Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization," presented at CHI 2007.
10. Heuer, R. *The Psychology of Intelligence Analysis*. Government Printing Office (1999), pp. 41.
11. Kumar, Garland, "Visual Exploration of Complex Time-Varying Graphs," IEEE Transactions on Visualization and Computer Graphics, Vol 12, No 5, Sept / Oct 2006.
12. Newman, A., Emory, J., Bennett, W., "Synchronized GMTI Radar Collection Management in a Coalition Environment," IEEE Information Fusion, Volume 2, 8-11 July 2002.
13. Novotny, M., Hauser, H., "Outlier preserving Focus and Context Visualization in Parallel Coordinates," IEEE Transactions on Visualization and Computer Graphic, Vol 12, No 5, Sept / Oct 2006.
14. Plaisant, C., Mushlin, R., & Snyder, A., (1998), "LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records." IBM T.J. Watson, White Paper.
15. Saraiya, P., Lee, P., and North, C., "Visualization of Graphs with Associated Timeseries Data," IEEE Symposium on Information Visualization, 2005.
16. Shanbhag, P., Rheingans, P., and desJardins, M. "Temporal Visualization of Planning Polygons for Efficient Partitioning of Geo-Spatial Data," IEEE Symposium on Information Visualization, 2005.
17. Shneiderman, B. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization," IEEE Symposium on Visual Languages, 1996.
18. Wong, Mackey, Perrine, Eagan, Foote and Thomas, "Dynamic Visualization of Graphs with Extended Labels," IEEE Symposium on Information Visualization, 2005.
19. Valiati, Pimenta, and Freitas, "A Taxonomy of Tasks for Guiding the Evaluation of Multidimensional Visualizations," Proceedings of the 2006 AVI workshop (BELIV '06).
20. Simile Timeline, MIT. http://simile.mit.edu/timeline/
21. TimeWall, InXight Software. http://www.inxight.com/products/vizserver/tw/