The Event Quartet:
How Visual Analytics Works for Temporal Data

Ben Shneiderman

Abstract— Francis Anscombe's Quartet of bi-variate numerical datasets (1973), each with only 12 data points, were hard to interpret in tabular form. He designed his datasets so that statistical methods such as mean, standard deviations, and correlation coefficients produced the same values, but when viewed as four scattergrams, the dataset patterns were immediately recognizable. Temporal categorical datasets provide new challenges for designers of visual presentations and statistical algorithms. Existing spreadsheet, database, and statistical tools are poorly matched with the needs of temporal data. Visual presentations of even a single timeline help clean up troubling data errors and reveal meaningful patterns in the data. More complex patterns can be found when multiple timelines can be seen simultaneously. Adding many event types generates still richer patterns. This paper shows examples of the power of visual presentations to enable users to make important insights.

INTRODUCTION

Temporal data from electronic health records, e-commerce web logs, social media posts, sensors, financial trades, and many others sources has been growing exponentially. Analysts struggle with these vast data resources to extract valuable insights, which might guide future decisions. When appropriate visualizations and statistical methods are well-integrated, analysts are more likely to find expected patterns, identify novel configurations, and detect surprising anomalies (Van Wijk, 2005).

In the case of multi-dimensional numerical data, visualizations and statistical methods have been refined for hundreds of years and analysts are experienced with their use (Cleaveland, 1994). Scattergrams and scatterplot matrices have become familiar visual representations enabling users to spot correlations, clusters, outliers, and other features. Similarly, statistical methods such as means, standard deviations, correlation coefficients, and hierarchical clustering can provide other insights. Beyond static presentations, interactive visualizations support richer forms of rapid exploration that further amplify analyst abilities when dealing with multi-dimensional data.

For geospatial data there is a long history of map making to highlight key features, such as borders, relative sizes of regions, positions of rivers and mountains, and routes connecting key locations. Modern maps include regional socio-economic-political information such as populations, farm productivity, economic development measures, or voting results, whose design is guided by well-understood perceptual and spatial cognition theories (MacEachren, 2004).

The rapid growth of interest in other data types, such as networks, has led to dramatic maturation of visualizations and statistical methods tuned to the relevant patterns such as connected components, cliques, common motifs, significant nodes with high centrality metrics, or tightly connected neighborhoods.

Now there is a rapid growth of interest in visualization and statistical methods to analyze categorical time-stamped event data (Aigner et al, 2011; Du et al., 2016; Gotz & Stavropoulos, 2014; Monroe et al., 2013; Perer et al., 2015; Plaisant et al., 1998; Shneiderman & Plaisant, 2016; Vrotsou et al. 2013; Wongsuphasawat & Gotz, 2012; Zgraggen et al., 2015). To demonstrate the power of visualization for temporal data, this paper presents a set of progressively more complex examples of categorical event data. My hope is to inspire discussion of what patterns, features, or events are important for designers to expose through interactive visual interfaces and statistical methods. Then interface designers, guided by perceptual and cognitive theories, can explore use of spatial attributes, color, shape, size, and other visual properties (Ware, 2012).

Since Anscombe's Quartet (1973) had such a profound impact, I sought to emulate his approach, which shows how easily patterns, features, and anomalies are discerned in graphs, yet very difficult to find in tables. I also followed his lead by choosing examples for which visualizations would show patterns that would be difficult to detect with general purpose algorithmic and statistical methods.

SINGLE TIMELINE, SINGLE EVENT TYPE

Example 1: Let's start with the simple case of a single timeline with only 12 occurrences of a single event type (Fig. 1). Times flows from left to right. In this case seeing the pattern in the table is not too difficult, but can you see the anomaly in the data? A simple visual presentation makes the pattern of a regular Green event with one missing point, which could be an absence date for student attendance or a skipped medication (Fig. 2).

Record ID	Event	Timestamp
Alpha	Green	8/7/2016 12:39
Alpha	Green	8/8/2016 12:35
Alpha	Green	8/9/2016 12:39
Alpha	Green	8/10/2016 12:37
Alpha	Green	8/11/2016 12:47
Alpha	Green	8/12/2016 12:31
Alpha	Green	8/13/2016 12:34
Alpha	Green	8/14/2016 12:29
Alpha	Green	8/15/2016 12:29
Alpha	Green	8/17/2016 12:34
Alpha	Green	8/18/2016 12:33
Alpha	Green	8/19/2016 12:29

Figure 2: Single timeline, single event type

Figure 2: Regular events with one missing event

- *Ben Shneiderman is with the Department of Computer Science and the Human-Computer Interaction lab at the University of Maryland.*

  Email: ben@cs.umd.edu

Proceedings of the IEEE VIS 2016 Workshop on Temporal & Sequential Event Analysis. Available online at: http://eventevent.github.io

Example 2: This is a slightly more complex pattern, difficult to spot in the tabular representation (Fig. 3) and even difficult to detect with a general purpose algorithm, yet clear to viewers in a simple timeline (Fig. 4). This visual pattern shows website statistics that indicate high weekday visitation with gaps on the weekends. Repeating patterns can be arbitrarily complexity, thus difficult to find with query languages or algorithms.

Record ID	Event	Timestamp
Alpha	Orange	8/8/2016 12:00
Alpha	Orange	8/9/2016 12:00
Alpha	Orange	8/10/2016 12:00
Alpha	Orange	8/11/2016 12:00
Alpha	Orange	8/12/2016 12:00
Alpha	Orange	8/15/2016 12:00
Alpha	Orange	8/16/2016 12:00
Alpha	Orange	8/17/2016 12:00
Alpha	Orange	8/18/2016 12:00
Alpha	Orange	8/19/2016 12:00
Alpha	Orange	8/22/2016 12:00
Alpha	Orange	8/23/2016 12:00

Figure 3: A slightly more complex repeated pattern

Figure 4: Website statistics show high weekday visitation with gaps on weekends

Example 3: This example is yet more difficult to identify from the table (Fig. 5), yet clear to viewers in a simple timeline (Fig. 6). This pattern of fewer events per unit time might indicate the decaying frequency of aftershocks of an earthquake. In this case the decline is quite regular, but in many datasets the irregular pattern makes it difficult for authors of statistical algorithms to detect such patterns.

Record ID	Event	Timestamp
Alpha	Purple	8/8/2016 12:39
Alpha	Purple	8/8/2016 13:35
Alpha	Purple	8/8/2016 15:39
Alpha	Purple	8/8/2016 21:37
Alpha	Purple	8/9/2016 10:47
Alpha	Purple	8/9/2016 23:31
Alpha	Purple	8/10/2016 22:34
Alpha	Purple	8/12/2016 12:29
Alpha	Purple	8/14/2016 3:44
Alpha	Purple	8/18/2016 12:33
Alpha	Purple	8/26/2016 12:29
Alpha	Purple	9/7/2016 7:03

Figure 5: Fewer events per unit time, e.g. earthquake aftershocks

Figure 6: Slowdown: fewer events per unit time, e.g. earthquake aftershocks

Example 4: This irregular pattern has random events with two clusters of 4 and 3 events, which are typical of tweeting patterns for a given hashtag (Fig. 7). Irregular patterns are elusive, but the clusters might be found if viewers were looking for it or had appropriate algorithms, but once again a visual presentation helps analysts to detect this pattern.

Record ID	Event	Timestamp
Alpha	Gray	8/8/2016 12:39
Alpha	Gray	8/9/2016 13:35
Alpha	Gray	8/10/2016 15:39
Alpha	Gray	8/10/2016 16:57
Alpha	Gray	8/10/2016 17:27
Alpha	Gray	8/10/2016 19:31
Alpha	Gray	8/13/2016 22:34
Alpha	Gray	8/14/2016 12:29
Alpha	Gray	8/16/2016 3:44
Alpha	Gray	8/16/2016 4:33
Alpha	Gray	8/16/2016 7:29
Alpha	Gray	8/18/2016 7:03

Figure 7: Irregular pattern, which is difficult to decipher in tabular form

Figure 8: Irregular pattern: Occasional clusters and random events, e.g. tweets on a given hashtag

Overall, these examples are meant to show how helpful a visual presentation can be. They are also meant to suggest new challenges for algorithm designers to be able to identify these and many other patterns even in the simple case of a single timeline and one event type.

SINGLE TIMELINE, MULTIPLE EVENT TYPES

The next step in complexity is to have a single timeline with multiple event types. The tabular versions would be complex, so the remaining examples show only the visual versions with explanations below (Fig. 9). Each event type is at the same level, to help analysts spot patterns.

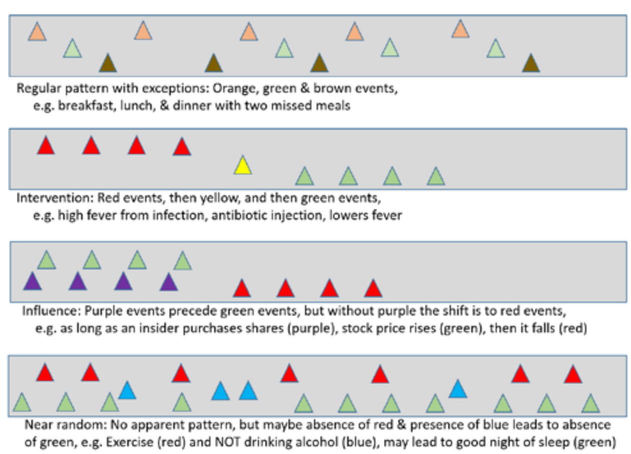


Figure 9: Four examples of single timeline with multiple event types. Each event type is on a separate horizontal line.

MULTIPLE TIMELINES, SINGLE EVENT TYPE

Many applications of temporal data require working with multiple timelines, such as multiple patients, but only a single event type, such as high blood pressure readings over time. Still, datasets following this simple description may vary in many different ways. These four examples all have four timelines, sometimes three have the same pattern, with one timeline showing a different pattern (Fig. 10).

MULTIPLE TIMELINES, MULTIPLE EVENT TYPES

Most of the applications of temporal data come with multiple timelines, such as multiple patients, and multiple event types, such as symptoms, medications, lab tests, and diagnoses. These four examples all have four timelines with multiple event types showing the kinds of features analysts seek to understand (Fig. 11).

CONCLUSION

The White House press release that describes a national effort to deal with Big Data (2012) put visual reasoning methods on an equal basis with algorithmic methods as it described two challenges:

- Developing scalable algorithms for processing imperfect data in distributed data stores
- Creating effective **human-computer interaction tools** for facilitating rapidly customizable **visual reasoning** for diverse missions.

This paper demonstrates the power of visual approaches in dealing with temporal data that has categorical event types. It suggests features that would be helpful to analysts in their visual exploration. It also suggests a set of challenges for designers of statistical algorithms to detect patterns that are apparent to visual analysts. The combination of visual and statistical approaches amplifies the ability of analysts to make important insights. Further discussion of this and related event analytics issues will be on:

http://hcil.umd.edu/eventanalytics

ACKNOWLEDGMENTS

I thank all the HCIL students who have worked on event analytics, our dozens of case study partners, the many sponsors of our work, and my long-term collaborator Catherine Plaisant. I appreciated supportive comments on drafts from Alan MacEachren, Sana Malik, Megan Monroe, Seth Powsner, Jarke van Wijk, and Colin Ware.

REFERENCES

W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-oriented Data*. Springer, 2011.

F. J. Anscombe, Graphs in statistical analysis, *The American Statistician 27*, 1, 1973, 17-21.

W. S. Cleveland, *The Elements of Graphing Data*, Hobart Press 1994

F. Du, B. Shneiderman, C. Plaisant, S. Malik, and A. Perer, Coping with volume and variety in temporal event sequences: Strategies for sharpening analytic focus, *IEEE Transactions on Visualization and Computer Graphics*, 2016 (to appear)

D. Gotz and H. Stavropoulos, DecisionFlow: Visual analytics for high-dimensional temporal event sequence data, *IEEE Transactions on Visualization and Computer Graphics 20*, 12, 2014, 1783–1792.

T. Gschwandtner, J. Gartner, W. Aigner, and S. Miksch, A taxonomy of dirty time-oriented data. In *Multidisciplinary Research and Practice for Information Systems*, Springer, 2012, 58–72.

A. M. MacEachren, *How Maps Work: Representation, Visualization, and Design*, Guilford Press, 2004.

M. Monroe, R. Lan, J. del Olmo, B. Shneiderman, C. Plaisant, and J. Millstein. The challenges of specifying intervals and absences in temporal queries: A graphical language approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, 2013, 2349–2358, 2013.

A. Perer, F. Wang, and J. Hu, Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics 56*, 2015, 369-378.

C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller and B. Shneiderman, LifeLines: using visualization to enhance navigation and analysis of patient records. In Proceedings of the AMIA Symposium, American Medical Informatics Association (1998) 76

B., Shneiderman and C., Plaisant, Tick, Tick, Tick: The Vitality of Temporal Data (2016 under review)

J. Van Wijk, The value of visualization, *Proc IEEE Visualization Conference*, 2005, 79-86.

K. Vrotsou, A. Ynnerman, and M. Cooper, Are we what we do? Exploring group behaviour through user-defined event-sequence similarity, *Information Visualization*, 2013, 232–247.

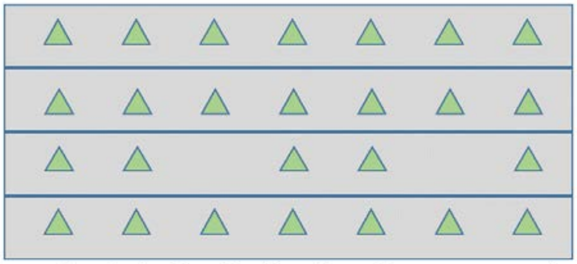
C. Ware, *Information Visualization: Perception for Design: Third Edition*, Morgan Kaufmann Publ., San Francisco, CA (2012).

White House Press Release, Obama Administration unveils "Big Data" initiative, Office of Science and Technology Policy, 2012, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

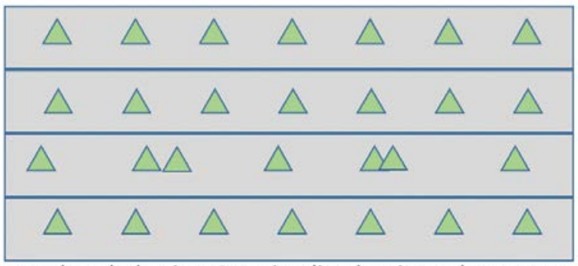
K. Wongsuphasawat, J. A. Guerra Gomez, C. Plaisant, T. D. Wang, M. Taieb-Maimon and B. Shneiderman, Lifeflow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2011, 1747–1756.

K. Wongsuphasawat and D. Gotz, 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

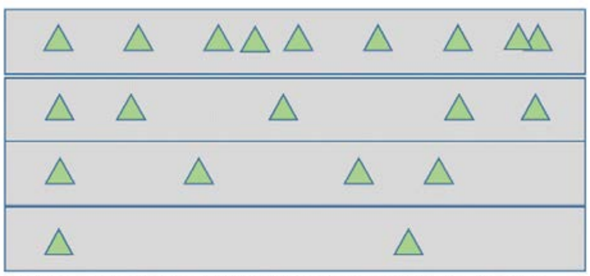
E. Zgraggen, S. M. Drucker, D. Fisher, and R. DeLine, (s|qu)eries: Visual regular expressions for querying and exploring event sequences. In *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, CHI '15, 2015, 2683–2692.



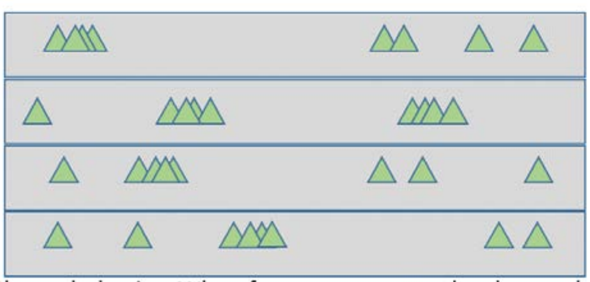
Anomalous behavior: One timeline misses green events.
e.g. Each employee attends weekly meetings, except one



Anomalous behavior: One timeline has irregular green events.
e.g. Each patient has regular heartbeats, one has arrhythmia

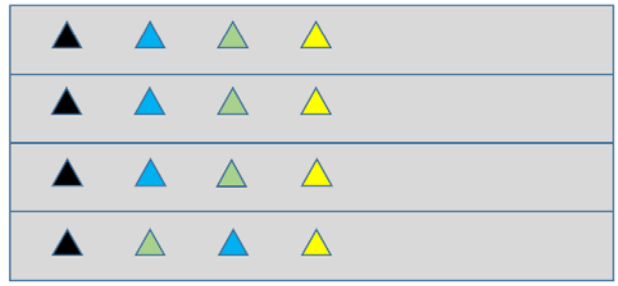


Varying behavior: Varying numbers of green events.
e.g. Patients vary by number of physician visits

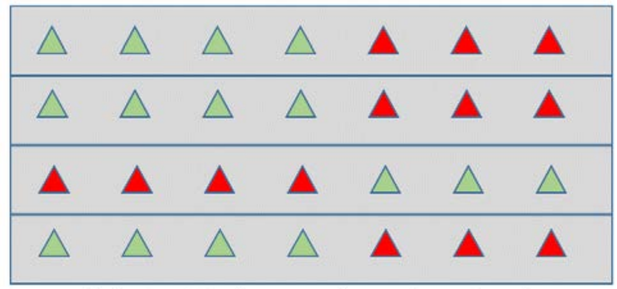


Cluster behavior: When four events occur closely together, there is a hiatus for some time, e.g. binge drinking`

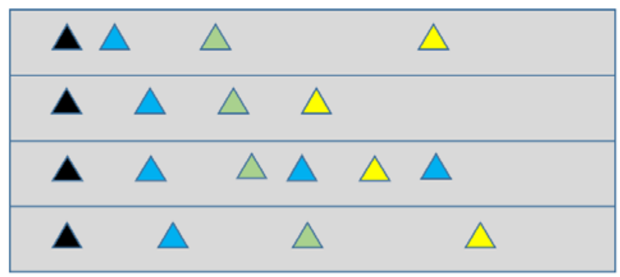
Figure 10: Four examples of four timelines each, all with single event types



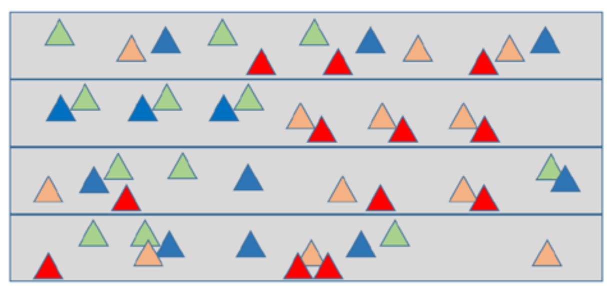
Following protocol: 1, 2, 3 follow order, 4 with inverted order, e.g. medical treatment protocols



Inverse behavior: Mostly green, then red, e.g. Growth stocks rise (1,2,4), while utility stock fall (3), then inverse



Variant behaviors: 1, 2 & 4 follow protocol with varying times, 3 follow protocol, but does blue three times, e.g. medical exam



Influence: Random, but timeline (2) has regular pattern (blue → green, then orange → red), e.g. stock purchases (blue) → rise (green), stock sales (orange) → fall (red)

Figure 11: Four examples of four timelines each with multiple event types