Assignment 2: Literature Review

Chris Storey

UMBC

CMSC 636: Data Visualization

29 September 2017

# Literature Review: Critical Analysis of Two Visualization Examples in Cyber Security

The application of visualization techniques to the field of cyber security is an active area of research given the complexity, large volumes of data, and speed with which security professions have to sift through the hundreds of signals present in typical network traffic. For a security practitioner to be successful, one has to readily identify actionable patterns within the data. Over the years, many techniques have been proposed, but effective idioms to reliably visualize network security data continues to elude the community. This paper attempts to review the effectiveness of two visualization examples from the literature that illustrate good and bad approaches to this problem.

## Overview

(Munzner, 2015) proposes eight rules of thumb for effective use of visualization techniques to convey meaning to views of various forms of data. These rules of thumb come from the author’s personal experience, as well as from empirical human factors studies of idioms in a variety of use cases. The eight rules of thumb are: 1) No Unjustified 3D; 2) No Unjustified 2D; 3) Eyes Beat Memory; 4) Resolution over Immersion; 5) Overview First, Zoom and Filter, Detail on Demand; 6) Responsiveness is Required; 7) Get It Right in Black and White; and 8) Function First, Form Next. Often there are trade offs in the use of each form of representation that have been identified by researchers. This review will examine each example in terms of its effectiveness and the tradeoffs relative to the use case.

## Comparative Review

(Milos Krstajic, 2011) proposes the novel use of CloudLines to visualize time-series data. The key benefit identified by this approach is the ability to display large amounts of data in a limited space. CloudLine plots are constructed of individual circles, representing event, that are then placed temporally along a timescale. An obvious application of this technique is to visualize network or cyber security data. The author’s attempt to overcome many of the tradeoffs that could potentially sideline this approach as an effective means to display large amounts of data. For example, in early versions of this technique, they recognized the difficulty that some users have distinguishing between events that are depicted very close together. Munzer calls this potential pitfall occlusion – where some objects cannot be seen because they are hidden behind others. As a result of this challenge, a lens-based interaction was added to allow direct access to these events. Likewise, opacity was added to represent the relative density of plots along the linear line. This density translates into areas of interest that might be of interest to the reader.

All of the above design choices make CloudLine plots effective representations of time-series data. Figure 1 was taken from (Milos Krstajic, 2011). The most noticeable elements of the visualization are the contrasts that are inherent between the temporal lines representing each political candidate. There does not appear to be any essential information missing, nor elements, which carry no information. Colors are used effectively to distinguish between the political candidates, while opacity is used to highlight areas of intensity. Because all of the elements used within this chart have been tailored to a particular purpose, this visualization does not encourage any misperceptions.

The CloudLine plots as depicted in this article could possibly be improved if they were integrated hierarchically with other idioms. For example, drill down functionality could be added provide a layered view of the data. This would augment the existing lens-based interaction, while improving the users ability to overcome occlusion. The implementation of this effect must guard against creating a cognitive load on the viewer when switching from one view to the other.



Figure 1: CloudLines visualization showing the relative number of appearances in the news media by sixteen different politicians during the month of February 2011.

(Florian Mansmann, 2007) extend concepts from logistics and resource planning to the visual analysis of network traffic. This article explores the use of various types of Treemaps – space-filling layouts of nested rectangles. The authors conclude that a geographically based Treemap called a HistoMap should be used to depicted the top layers of the network and provide an overview for the user. Additionally this top-level view provides scalability, squareness, and stability benefits. For lower levels of the network hierarchy, the author’s believe that Strip Treemaps are most appropriate due to their maintaining the input order of nodes (Florian Mansmann, 2007, p. 1111).

(Munzner, 2015) discusses the pros and cons of Treemap extensively. On the one hand, the idiom of treemaps can effectively show hierarchy relationships. However, they are most effective for hierarchies that are shallow rather than deep (Munzner, 2015, p. 214). The application of Treemaps to the representation of internet traffic at the TCP/IP level is therefore a mistake, given the many layers of abstraction from Automated Systems at the top layers, down to the individual IP addresses near host systems. The most noticeable elements of the visualization depicted in Figure 2 are the boundary lines that are allocated to each node. These relationships are the most important. There is a lot of information missing in this view because of the scale. It is extremely difficult for the view to see detail down to the granular level of IP addresses in this view. Most of the diagram provides no information. Color is used effectively to show intensity or the number of connections from low (blue) to high (red). It is possible that a viewer could misperceive the importance of a particular node or activity based upon the scale of the hierarchy. In the domain of internet security, even the small nodes are important.

The Treemap plots as depicted in the article could possibly be improved if they were combined with some additional algorithms that would assist in lessening the cognitive load. According to Munzner, it might be helpful, given the size of the depicted dataset, to create a useful overview for top-down exploration. This would allow the user to find context within the data and starting points for further analysis.



Figure 2: A Treemap visualization of outgoing traffic connections through a university gateway on November 29th, 2015, showing 197427 IP prefixes.

## Conclusions

By grouping rules of thumb about the current state of knowledge on visualization techniques, Munzner provides a service. The literature discusses a broad listing of idioms that could be applied to cyber security data. Exploring the trades that can be made between different design elements are useful excursions and contain suggestions for future research.

# Bibliography

Florian Mansmann, D. A. (2007). Visual Analysis of Network Traffic for Resource Planning, Interactive Monitoring, and Interpretation of Security Threats. *IEEE Transactions on Visualization and Computer Graphics* *, 13* (6), 1105-1112.

Milos Krstajic, E. B. (2011). CloudLines: Compact Display of Event Episodes in Multiple Time-Series. *IEEE Transactions on Visualization and Computer Graphics* *, 17* (12), 2432-2439.

Munzner, T. (2015). *Visualization Analysis & Design.* Boca Raton, FL: CRC Press, Taylor & Francis Group.