A Visual Analytics Approach to Understanding Spatiotemporal Hotspots
Ross Maciejewski, Student Member, IEEE, Stephen Rudolph, Ryan Hafen, Ahmad M. Abusalah, Mohamed Yakout, Mourad Ouzzani, William S. Cleveland, Shaun J. Grannis, and David S. Ebert, Fellow, IEEE

In this paper, the authors present a suite of tools designed to facilitate the exploration of spatiotemporal data sets. Their system allows users to search for hotspots in both space and time, combining linked views and interactive filtering to provide users with contextual information about their data and allow the user to develop and explore their hypotheses. Statistical data models and alert detection algorithms are provided to help draw user attention to critical areas. Demographic filtering can then be further applied as hypotheses generated become fine tuned. This paper demonstrates the use of such tools on multiple geospatiotemporal data sets, including law enforcement data and syndromic surveillance data.

Figure 1. The visual analytics system. (a) The conceptual diagram of our visual analytics system. Observe the interaction between the analyst and the system as well as the modeling components of the system. (b) Our visual analytics system. The left portion of the screen represents the interactive temporal tools. We include time aggregation tools, pull down menus for data classifier selections, keyword searches for chief complaint text, and demographic filtering for age and gender. The main viewing area is a geospatiotemporal view that has pan and zoom controls in the upper left corner. Hospitals and regions of the map may be selected with a circular query tool for interactive time series generation. The rightmost windows are the temporal views, showing selected time series plots broken down into their relevant components. Users may select points or regions of time to interactively manipulate the geospatial temporal window. For analyzing crime data, the interface is modified only slightly to reflect the relevant categories.
Figure 2. Data aggregation and privacy preservation. (a) Georeferenced syndromic surveillance data as small additive opacity circles. (b) Georeferenced data overlaid with red circles representing syndromic patients. (c) Data aggregation for enhanced visualization. (d) High-resolution zoom of an area of interest. (e) Actual patient locations at a high-resolution zoom overlayed with our data aggregation method.

Figure 3. Data aggregation and privacy preservation visualized as a percentage of syndromic population over the total population seen. (a) Data aggregated by county. (b) Data aggregated through nearest neighbor groupings. (c) A combination of data aggregation to enhance contextual visualization.
Figure 4. Kernel density estimate (KDE) heatmaps visualized as a percentage of syndromic population over the total population seen. (a) KDE heatmap. (b) Contextualizing the KDE heatmap by overlaying patient data aggregated through nearest neighbor groupings. (c) A zoomed in view of a local hotspot. (d) Contextualizing a hotspot through interactive coloring.

Figure 5. Kernel density estimate (KDE) heatmaps visualized as a percentage of syndromic population over the total population seen. (a) KDE heatmap. (b) Contextualizing the KDE heatmap by overlaying patient data aggregated through nearest neighbor groupings. (c) A zoomed in view of a local hotspot. (d) Contextualizing a hotspot through interactive coloring.
Figure 6. Contour mapping for contextual cues. (a) The analyst has created a heatmap of one year’s worth of crime data in West Lafayette Indiana with the green polygon representing Purdue University Campus. The contours overlaid represent the noise complaints aggregated over the past 100 days. (b) The analyst plots only the noise complaints aggregated over the past 50 days as a heatmap. (c) The analyst plots the noise complaints aggregated over the next 50 day period. The contours in this figure represent the previous noise complaint map outline.

Figure 7. Multivariate views. (a) The analyst has created a heatmap of shock/coma cases overlaid with contours for rash. (b) The analyst adds the category respiratory as the height dimension.
Figure 8. Interactive thresholding. (a) The analyst searches for gastrointestinal hotspots. (b) The analyst uses the thresholding capability to filter the data. (c) The analyst moves forward in time viewing movement trends among the data.

Figure 9. Using visual analytics for hypothesis exploration in syndromic surveillance. (a) The user observes a heatmap for a given syndrome, in this case, gastrointestinal. (b) Next, the user selects an area of interest, generating a time series plot for that region. Note that in the time series plot generated, an alert is occurring on the day of interest. (c) The user then drills down to the hospital level by selecting the neighboring hospital and generating a time series plot for that emergency department. Here, we see that there is no hospital level alert for gastrointestinal syndromes. (d) Finally, the user looks for correlating symptoms and filters by the keyword fever. New time series plots are generated. While an alert still exists for the selected area, the user can now see that this alert was generated by only one individual, meaning an outbreak is unlikely.
Their current work demonstrates the benefits of visual analytics for understanding syndromic hotspots. By linking a variety of data sources and models, we are able to enhance the hypothesis generation and exploration abilities of our state partners. Their initial results show the benefits of linking traditional time series views with geospatiotemporal views for enhanced exploration and data analysis. Their system also moves away from traditional spatial histogram visualizations, providing a finer granularity of heatmap for more accurate hotspot detection. Other future work includes advanced modeling of geospatiotemporal data for enhanced data exploration and hotspot detection. Furthermore, they plan to include a suite of aberration detection algorithms and their corresponding control charts for enhanced alert detection in the temporal domain. They also plan on employing spatiotemporal clustering algorithms.

Figure 10. Using visual analytics for hypothesis testing in crime analysis. The user is analyzing thefts (as contours) versus all crimes (as color) for a given school year (2007-2008). (a) The user analyzes fall semester thefts (contours) compared with the overall school year crimes. (b) The user analyzes the last 20 days of the fall semester for thefts. (c) The user analyzes spring semester thefts. (d) The user analyzes the last 20 days of spring semester.
Furthermore, they plan to enhance our system from a visual analytics system to a predictive analytics system, creating views to allow for event planning, prediction and interdiction. Once these features are implemented, they plan to deploy our system with their state partners for further evaluation.

**URL:**
http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5226628

**Citation:**
19

**Venue's Impact Factor of IEEE Transactions on Visualization and Computer Graphics**
Year 2010: 1.922