Chapter 5

DATA VISUALIZATION, INFORMATION DISSEMINATION, AND ALERTING

Syndromic surveillance systems are critical for public health surveillance because they often provide epidemiologists and public health officials the visual analytics tools and techniques to synthesize information and detect the data anomalies (possible outbreaks) from massive, dynamic, and often ambiguous surveillance data. Represented visually, the assessments of public health status are better understood and also more effectively communicated for action. The geographic or spatial components of the surveillance data enable the natural application of visualization techniques for computerized assistance for decision making in spatial (and often spatial-temporal) analytics (e.g., clustering detection and resource logistics). In addition, the interplay between simulation and visualization provides a powerful combination. Visualization techniques can be used to analyze simulation output and analysis results, and can drastically improve the understandability and accessibility of the model to both technical and nontechnical audiences. Virtually all simulation software packages have some level of visualization, ranging from basic diagrams to full animation.

This chapter provides a systematic summarization of data visualization techniques that are employed in the surveyed syndromic surveillance systems. Taxonomy of the visualization techniques precedes the discussion of the two classes of visualization technologies: visual information display and interactive visual data exploration. A number of example screenshots from the surveyed syndromic surveillance systems visualizations are shown along with the text.
1. SCOPE AND TAXONOMY

Visualization technology and user interface design involve a vast literature in HCI, computing graphics, psychology, database organization, dynamic query, and display algorithms, as well as screen management algorithms. But we are not targeting a complete scientific discussion of visualization technologies, but instead, we identify the different visual representations of the syndromic surveillance data in the public health context and examine how to advance the relevant applications of user interface design methods.

Shneiderman (1996) identified two aspects of visualization technology that can be directly applied to a given structure. One focuses on mapping abstract information to a visual representation and the other provides user-interface interactions for effective navigation over displays on a screen. Our focus of the discussion accordingly includes two pieces of visualization techniques: visual information display and interactive visual data exploration. By and large, the visual information display includes temporal, spatial, and temporal-spatial information display exploring different dimensions of the information.

The readers should also be reminded that data analysis technologies serve as back-end support for visualization functions. As noted by Chen (1999), data analysis technology as a “third-dimension” of the information visualization technology serves to create structures that characterize the data set, abstract the unstructured or high-dimensional information. The data analysis technologies as discussed in the previous chapter are important back-end methods that drive many of the visualization approaches discussed in this chapter.

2. VISUAL INFORMATION DISPLAY

Visual information display techniques aim to present visually either raw surveillance data or analysis results (e.g., from the data anomaly detection algorithms) (Zhu and Chen, 2005). Visual representation techniques are applied to either time-series data or spatial/geographical data. The traditional methods of information display are multidimensional tables (line listing), and various static statistical graphics, such as line graphs, scatter plots, bar charts, and pie charts. Color-coded maps are often used to represent disease cases and clusters with case locations. Geographical Information Systems (GIS) are now being widely used for spatial information representation and cluster detection. Graphs with nodes and links, such as trees and networks, are not seen in the surveillance information display, but they might be viewed as promising tools for disease modeling based on spreading patterns.
2.1 Visualization of Time-Series Data

Line chart is a popular method to visualize time-series data as it can help identify temporal patterns such as spikes or clusters. Usually one curve represents the observed data, and the other curve is the normal pattern plotted by the temporal analysis algorithms. Line charts and other plotting methods for time-series analysis are supported by most statistical analysis packages (e.g., SAS and SPSS). The example view of the interface of BioSense application shown in Figure 5-1 is a line chart and a line listing of the fictional time-series syndromic data in a metropolitan area. Figure 5-2 shows a screenshot from the EARS system (Hutwagner et al., 2003), visualizing daily data feed from a hospital and the results of applying the CUSUM algorithm.

Other types of plots such as candlestick plot and density ratio map are also seen in syndromic surveillance applications. Figure 5-3 shows a density ratio map visualizing data aggregated by patient age in several influenza seasons (DIMACS, 2006).

Figure 5-1. Example views available in the BioSense application (source: Biosence Website).
Figure 5-2. Line charts plotting temporal patterns of disease cases (EARS system).

Figure 5-3. Density ratio maps visualizing data aggregated by patient age (DIMACS, 2006).
2.2 **Visualization of Spatial Information**

Visualizing disease cases or surveillance-related events on a map can help identify case clusters (typically indicative of outbreaks), investigate possible causes of a disease or an outbreak, and study an outbreak’s dissemination and evolutionary patterns. One major objective of visualization is to identify geographical areas with unusually high numbers of cases or events to serve surveillance purposes and inform outbreak response decisions. Another objective is to determine high-risk areas for a disease under investigation and help analyze correlations between disease occurrences, various types of environmental factors, and social-demographic variables.

There are several techniques for displaying spatial information contained in syndromic data. Printed maps are often used to identify geographic clusters or hotspots (Figure 5–4). CDC and the National Center for Health Statistics support research to investigate the design and display for disease atlases (Lawson and Kleinman, 2005). Geographical display of disease statistics in real time is also widely used for situation awareness and incident response (Kulldorff, 2001).
Techniques also exist to smooth the borders of identified regions of interest and display overlapping clusters. Boscoe et al. (2003) proposed an approach for visualizing spatial scan statistic analysis results using nested circles, which displays both the relative risk and statistical significance of identified hotspots. They show that the mapped clusters typically do not have precise boundaries. Rather they consist of relatively well-defined cores and fuzzy boundaries.

Another study presents the health statistics on a map with both geographical information and the reliability of the displayed data indicated by a texture overlay (MacEachren et al., 1998). A screenshot from their work is shown in Figure 5-5.

Color is an effective visual display property, and it can be an important aid for fast and accurate decision making. Color encoding is a traditional visualization technique to display indirectly standard deviations by which the observed data (e.g., the number of cases of a particular syndrome category in a zip code) deviate from the expected counts. The idea is to use different colors or shadings to illustrate clusters of high or low rates of disease incidence. The screenshot in Figure 5-5 employs such a color encoding technique.

Figure 5-5. A screenshot from (MacEachren et al., 1998) showing both geographical information and data reliability.
2.3 GIS for Disease Event Visualization

Geographic Information System (GIS) is a powerful spatial information visualization tool and has found important applications in public health surveillance (2003c; Hurt-Mullen and Coberly, 2005; Lombardo et al., 2003). Most of the individual records registered into syndromic surveillance systems can often be georeferenced to a range of geographic areas such as blocks, tracts, county subdivisions, and other geographic units. Many syndromic surveillance systems (e.g., BioSense, RODS, ESSENCE, BioPortal, and RSVP) in the survey interface with GIS for disease visualization and spatial analytics.

The strength of GIS lies in its ability to integrate different types of data onto a common spatial platform. The integration of the environmental factors (e.g., groundwater contamination), demographical data and remote sensing data (e.g., satellite data) such as vegetation, land-use patterns and soil types, climatic changes, and so on, helps to identify and track the environmental characteristics and risks for epidemiological studies. First, GIS is a powerful tool for disease mapping and spatial visualization of environmental factors. In addition to visualization, disease outbreak detection and prediction based on the GIS analytical tools has been studied widely.

Geostatistical functions are provided in many statistical software packages. To date, the GIS softwares are capable of disease mapping, geographical correlation studies, disease clustering, spatial-temporal analysis, disease data visualization. S+SpatialStats, available from Mathsoft, implements lattice model estimators. Matlab has a Mapping Toolbox (Matlab), a collection of Matlab functions, user interfaces, sample data sets, and demos that read, write, display, and manipulate geospatial data, that contains Kriging functions and SpaceStat (TerraSeer) provides tools for exploratory spatial data analysis such as Moran’s I, Geary’s C and spatial regression methods including trend surface regression, spatial analysis of variance among others. The SAS Bridge from SAS bridges SAS and ESRI’s ArcGIS9 by linking spatial, numeric, and textual data through a single interface, saving the efforts of customizing data transformation and transfer. In addition, GSLIB and GEOEAS among others are also serving the market. GEOEAS is a collection of interactive software tools for geostatistical analysis. The principal functions of the package are the production of grids and contour maps of interpolated (Kriged) estimates from sample data. GEOEAS can produce data maps, univariate statistics, scatter plots/linear regression, and variogram computation and model fitting. GSLIB (Geostatistical Software Library and related software) maintains a collection of geostatistical programs developed at Stanford University over the years. GSLIB provides variogram analysis and Kriging techniques. It also analyzes three-dimensional data sets.
ESRI is a leading GIS modeling and mapping software and technology provider. The Spatial Analyst, 3D analyst with ArcGIS, Geostatistical Analyst and Tracking Analyst tools have found their applications in disease monitoring, tracking and outbreak detection for various types of disease data [e.g., West Nile Virus (WNV)] (ESRI). Geostatistical Analyst provides ESDA (Exploratory Spatial Data Analysis), Deterministic interpolation methods and Kriging interpolation methods. Tracking Analyst of ArcGIS can map objects that move or change status through time. Figure 5-6 shows the screenshot when executing spatial-temporal Analysis using tracking analyst in ArcGIS, illustrating the evolvement of Hepatitis B in China during 1999–2001 (Zhong et al., 2005). In another application, the Missouri Department of Health and Senior Services employs ArcGIS for disease and bioterrorism surveillance by tracking syndromic information.

GeoMedStat is another GIS application developed at ESRI. Using GeoMedStat, real-time syndrome data (typically visits for each syndrome) can be mapped at the ZIP Code level within the state over a Web-based interface (Li et al., 2006).

By integrating GIS and the city’s standardized location data with various agency-wide databases, the New York City Department of Health and Mental Hygiene (DOHMH) is able to analyze a range of health data and evaluate disease trends and their relationship with environmental conditions.

Figure 5-6. GIS application for disease incidence tracking (Zhong et al., 2005).
After West Nile Virus first appeared in the United States in the summer of 1999, DOHMH developed a vector and avian (mosquito and bird) tracking system using a GIS-enabled database and Intranet application. “The database is implemented with both spatial (ArcSDE) and relational (SQL Server) database management system software that allows staff to collect incoming information from the public through an ArcIMS software-enabled Web site” (Mostashari, 2002). The city’s Bioterrorism Response Geographic Information System (BTRGIS) is also based on GIS technology.

GIS maps are also supported by Spatial Temporal Visualizer (STV) available from BioPortal disease surveillance system. The STV GIS view displays cases and sightings on a map, allowing the user to select multiple data sets (e.g., disease cases, natural land features, land-use elements) to be shown on the map in a layered manner using the checkboxes. It also supports dynamically generated views, zooming, brushing, and animation. In addition, it allows the user to invoke advanced spatial temporal analysis methods such as Prospective Support Vector Clustering (PSVC) (see Chapter 4 for details) and visually inspect their results through STV. A screenshot of GIS views from STV is shown in Figure 5-8.

Figure 5-7. NYC disease surveillance system GIS view (2003a).
In addition, GIS is also used in generating simulated cluster data that can be used as artificial outbreaks for evaluating the detection capability of the outbreak detection algorithms. The artificially generated clusters are customized for desired cluster radius, density, distance, relative location from a reference point, and temporal epidemiological growth pattern to explore a variety of the uncertainties for disease detection algorithm to test (Cassa et al., 2005; Watkins et al., 2005). For instance, based on user-specified parameters describing the location, properties, and temporal pattern of simulated clusters, the AEGIS Cluster Creation Tool (AEGIS-CCT) enables users to create simulated clusters with controlled feature sets.

Internet-based GIS technology and mobile GIS technology provide innovative mechanisms to facilitate flow of information. They allow the instant availability and accessibility of the information across the globe. We expect that the technologies can further facilitate the field data collection, real-time information sharing, and event investigations in the domain of disease surveillance.

![Figure 5-8. Visualization of dead bird cases distributed along populated areas near Hudson River by BioPortal STV (2006a).](image-url)
2.4 Spatial-Temporal Disease Modeling and Other Visualization Examples

As an ongoing research project, IBM has developed a Spatio-temporal Epidemiological Modeler (STEM) to model and visualize the spatial and temporal models of emerging infectious diseases. The tool has built in GIS data and it integrates with Susceptible/Infectious/Recovered (SIR) and Susceptible/Exposed/Infectious/Recovered (SEIR) models.

The STEM model is one of the few works on visualizing the infectious disease spreading models. An example from STEM is shown in Figure 5-9.

“Policymakers responsible for creating strategies to contain diseases and prevent epidemics need an accurate understanding of disease dynamics and the likely outcomes of preventive actions. In an increasingly connected world with extremely efficient global transportation links, the vectors of infection can be quite complex. STEM facilitates the development of advanced mathematical models, the creation of flexible models involving multiple populations (species) and interactions between diseases, and a better understanding of epidemiology. The STEM application has built in GIS data for every county in the United States. It comes with data about county borders, populations, shared borders (neighbors), interstate highways, state highways, and airports” (Ford et al., 2005).

![Figure 5-9. Visualization using IBM STEM (source: http://www.alphaworks.ibm.com/tech/stem).](http://www.alphaworks.ibm.com/tech/stem)
The integrated visualization and analysis environment of BioPortal system also supports a sequence-based phylogenetic tree visualization of infectious disease when gene sequence information is available. The sequence-based phylogenetic tree visualizer has been recently developed for diseases such as the Foot-and-Mouth disease as shown in Figure 5-10.

3. INTERACTIVE VISUAL DATA EXPLORATION

Interactive visual data exploration entails a wide range of techniques and operations for effective navigation on computer screens, the process of information query and, if needed, close examination of individual cases or patterns (Shneiderman, 1998). In particular, the operations and methods are expected to provide support for flexibility and interactivity, which allow the users to explore the information (e.g., a database) dynamically by specifying a year, a county, and the demographic querying criteria such as age and gender. Rapid, smooth screen changes on users’ demand are essential for the perception of patterns, facilitating the early detection of changes in disease incidence rate over time and in correlation with demographic variables.
There are generally six types of interface functionality in syndromic surveillance applications: overview, zoom, filtering, details on demand, relate, and history (MacEachren et al., 1998). A typical surveillance task always involves a continuous combination of a set of functionalities of the six types.

As an example, the interactive visual data exploration environment from the BioPortal project, called the Spatial-Temporal Visualizer, supports all six elements to display disease hotspots (see Figure 5-11). This environment consists of a GIS display, a Gantt-chart temporal display, statistical plottings, and a time-range filter, which are all user controllable and synchronized.

In summary, we found that very few systems (e.g., BioPortal) support dynamic GIS functions or a full-blown interactive visual data exploration environment. Systems including RODS, ESSENCE, and BioSense provide limited support for interactive data exploration. Most syndromic surveillance systems support geographic displays of a local region with vector maps. All systems offer time-series plottings, arranged or aggregated by syndrome categories, ages, and other covariates.
There are several challenges with data visualization in syndromic surveillance. First, the number of maps generated daily for review is often large (Wagner et al., 2004b). For example, if there are 8 syndrome categories and 10 geographical regions, at least 80 maps need to be generated for daily review. If other parameters such as age and gender are also included in the analysis, the number of the maps generated quickly becomes unmanageable.

Therefore, automatic screening of the maps (e.g., based on anomaly detection algorithms) is critical.

Next, when the current research is focused around the best methods for automating the visual presentation and interpretation of the data, a major problem with spatial data analysis is data normalization. There is a large amount of both temporal and spatial variability that must be taken into account. For example, a known temporal variability is the seasonal variation in respiratory diseases with increases during the winter months. Spatial variability is even more problematic. A certain healthcare facility is centrally located and draws patients from the entire state. However, the number of patients seen and the severity of their illnesses are associated with the distance the patient must travel to reach the hospital. Rural areas also have large variations in population density that must be considered. These normalization issues are a complex topic.

In addition, although the surveyed visualization tools used in biosurveillance present a wide application of visualization on disease incidence clustering, we notice that there is a lack of research on infectious disease modeling. Research on disease modeling with visual model presentation is critical for enhancing the understanding of the nature of infectious disease and its causes, processes, development, and consequences, so as to facilitate the surveillance process.

In general, we note that interactive, user-controlled, and real-time renderable data visualization can be leveraged to enable effective surveillance and decision support, and represents an important research direction.

5. INFORMATION DISSEMINATION AND REPORTING

We summarize below some existing work on information dissemination channels for real time alerting and investigation process triggering. Information dissemination and alerting are concerned with managing and distributing daily or weekly public health updates and outbreak alerts for involved parties
such as public health officials, analysts, primary care providers, and possible public safety and homeland security officials.

Existing syndromic surveillance information dissemination approaches include email, FAX, pager, phone calls, Web, and dedicated communication networks. These approaches differ greatly in their level of security, labor, and resources involved in the procedure, and delay in processing time.

A few nation-wide secure networks have been built for public health information dissemination and alerting. The CDC’s Health Alert Network (HAN) serves as a communication backbone, linking public health departments in 37 states to CDC headquarters in Atlanta, and now is being expanded nationwide (2004b). The Epidemic Information Exchange (Epi-X) system is the CDC’s secure, Web-based communications network that serves as an exchange between the CDC, state and local health departments, poison control centers, and other public health professionals (CDC, 2006b). Epi-X provides rapid reporting, immediate notification, and coordination of health investigations. The Public Health Information Network Messaging System (PHINMS) provides a secure and reliable messaging system for the PHIN (2003b; Barry and Kailar, 2005). PHINMS implements ebXML standard (Kotok, 2003) for bidirectional data transport, which offers high-quality encryption and authentication. An implementation of HAN- and PHINMS-based syndromic surveillance is described in (Daniel et al., 2005).

Most syndromic surveillance systems support multiple dissemination channels. The most commonly used methods, such as Email notification and voice communications, are relatively fast. Web-based messages and alerting networks are used less frequently. Secure network alerting with automatic role-based personnel directory access can be very useful in automatic and real-time alert distribution and is increasingly gaining acceptance.