CPSC 533C : Topics in Computer Graphics Visualization

Submission Date :

30th October, 2009

Project Title :

Diseases Data Correlator (DDC)

Project Group Members :

Shama Rashid; e-mail: shama@cs.ubc.ca

Domain, Task, and Dataset :

This tool is intended to facilitate exploration of data related to population infected by a particular disease in order to find data trends and to figure out correlation among different data attributes (i.e spread of disease 1 vs disease 2, age group and rate of infection etc.). I shall be using the data available from the website statsci.org and/or CDC (CDC provides the information for easy access by general public, so some processing has to be done to format it as statistical data). The main challenge here is in designing a representation that can show the overview of the general situation overlaid on a map while showing a details view for a particular exploratory task. The dataset to work on are simple tuples like <state, year, disease, count or percentage data>.

Personal Expertise:

I do not have any particular expertise in the area of health science but geo-spatial data representation is an area I have taken an interest to for a while now. I have worked for a telecom operator the last few years. The presence of different types of equipments manufactured by different companies, each having their own proprietory protocol for performance and fault management, had posed a challenge to our network monitoring team. To facilitate maintenance of network quality we have developed integrated tools for alarm management and parameter auditing but none of these tools were proper GUIs considering all information were shown in the classic tabular format. To provide a more intuitive representation of these equipment data, positioned all across the country, a geo-spatial solution was required. I have chosen to work on correlation of disease information due to the availability of a sizable data set accessible publicly.

Infovis Solution and Scenario:

Demographic distribution of a disease:

This is a map-based display to show the extent of spread of a disease in different regions (US states in my case). When a user chooses the disease name from a menu the representation will be color coded state wise according to the intensity of the spread for the chosen timeframe. Additionally all related data will be available in tabular format for further review.

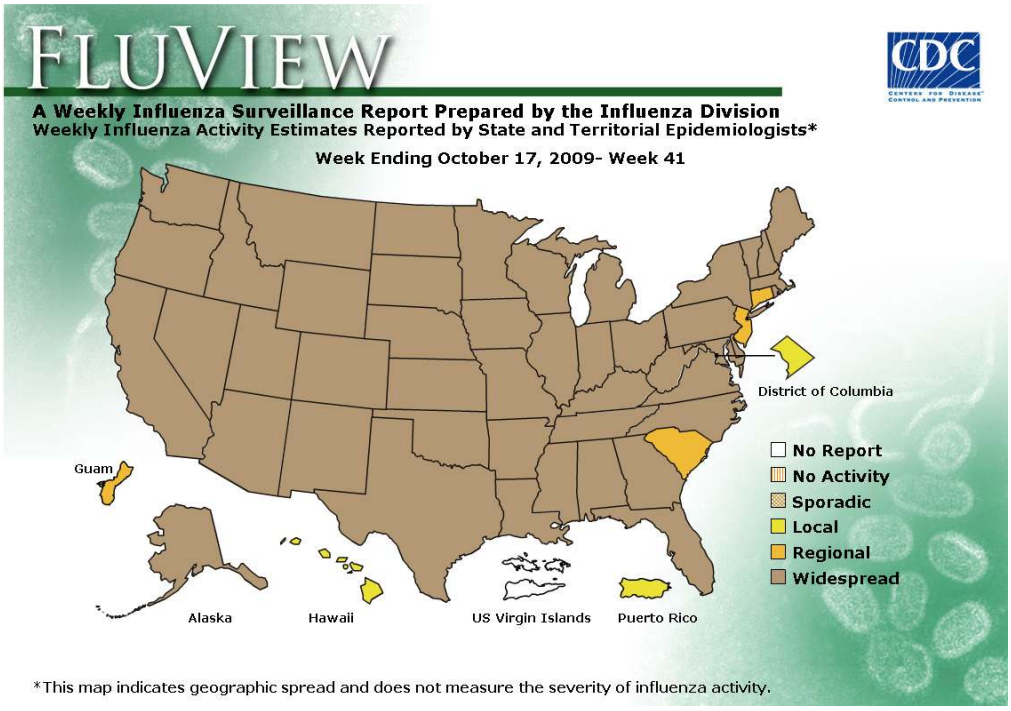

Image 01 : Sketch for a particular disease extent

Source URL : http://www.cdc.gov/healthmarketing/ehm/data-visualization.html

Disease trend graphs:

This is a simple trend graph based on time series data for a particular disease. This will provide a sense whether the disease is affecting more or less population with time.

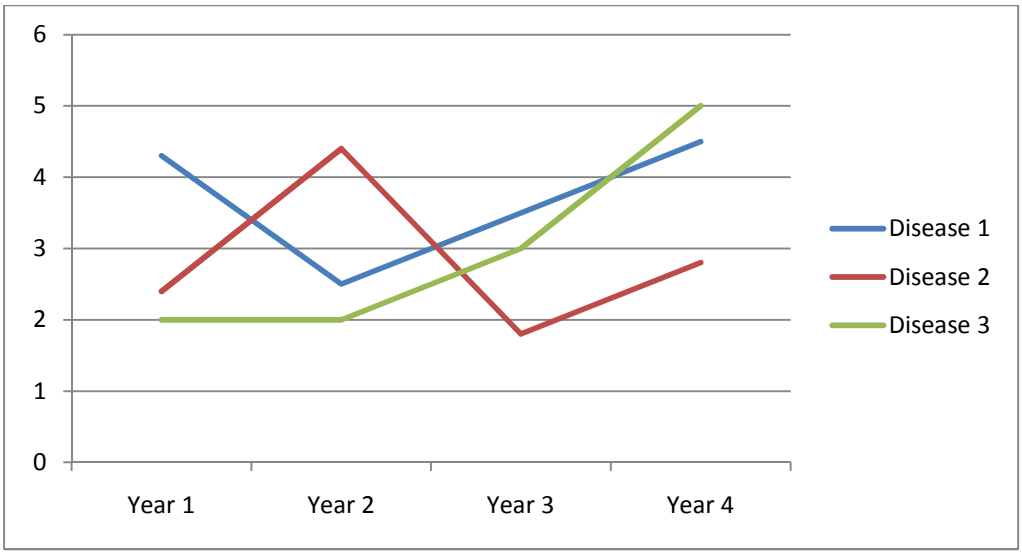

Image 02: Disease trend graph

Disease correlation analyzer:

Pair-wise disease data for each state can be correlated using a scatter-plot to find out whether these two diseases have a common contributing factor. In the overview, a state-wise bar-chart can be used. Since manually selecting all possible combinations for the scatter-plot graph one at a time is a daunting task a scagnostic-like representation could be used to provide the user an interface to choose the dataset. (This would be analogous to correlating network equipment faults to find out whether failure in one node has led to the performance degradation in another).

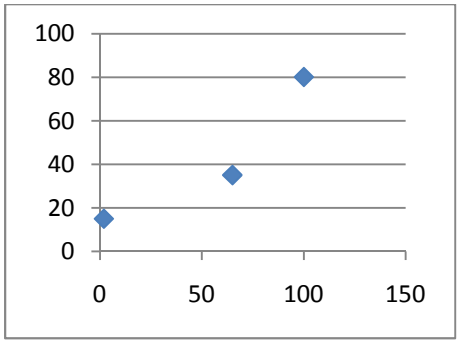

Image 03: Disease correlation graph

The correlation can also be done for disease data and other attributes like age group, income etc.

Implementation Approach:

GeoVISTA, Java

Target Project Milestones :

1. Project proposal 30-Oct-09
2. Assemble data from CDC site (format as statistical data) 03-Nov-09
3. Overlay single disease demographic data 07-Nov-09
4. Demo correlation charts or graphs (details view) 14-Nov-09
5. Project update presentation 16-Nov-09 or 18-Nov-09
6. Overview charts 30-Nov-09
7. Final presentation 14-Dec-09
8. Final Report 16-Dec-09

Previous Works:

The area of visualizing health threat data is well explored. The basic spread of disease data can be seen nowadays on live media or websites as a part of programmes conducted by governments or institutions to increase public awareness. For Example, CDC (Centre for Disease Control) is working to develop interactive tools to improve visualizing health data. This tool provides an easy to comprehend representation of regional penetration of common health threats like seasonal flu based on dynamic data. Although, I couldn`t find any commercial software for particularly this domain, statistical data analysis tools are readily available. Similarly, there are numerous works on demographic data trend, geospatial data representation etc. Some of the readings we have done for this course that are related to this project are as follows:

1. Geographically weighted visualization – interactive graphics for scale-varying exploratory analysis - Chris Brunsdon, Jason Dykes
2. Graph-Theoretic Scagnostics - Leland Wilkinson, Anushka Anand, and Robert Grossman