

# T-Cube Web Interface Software Requirement Specifications

Auton Lab

Carnegie Mellon University

Version 0.1

Table 1: Revision history

<b>Date</b>	<b>Version</b>	<b>Description</b>	<b>Author</b>
June 4, 2009	0.1	Initial Proposal	Robin Sabhnani, Artur Dubrawski

# Contents

<b>1</b>	<b>Overview</b>	<b>5</b>
1.1	Product perspective . . . . .	5
1.2	User characteristics . . . . .	5
1.3	Requirements . . . . .	5
1.4	Constraints . . . . .	6
<b>2</b>	<b>Definitions, Abbreviations</b>	<b>7</b>
2.1	Definitions . . . . .	7
2.2	Abbreviations . . . . .	7
<b>3</b>	<b>Functionality</b>	<b>7</b>
3.1	Semi-synthetic data . . . . .	8
3.2	Time series analysis . . . . .	8
3.2.1	Panels . . . . .	9
3.2.2	Sample outbreaks . . . . .	12
3.3	Spatial data analysis . . . . .	13
3.3.1	Spatial data visualization . . . . .	13
3.3.2	Multivariate Bayesian spatial scan (MBSS) . . . . .	14
3.3.3	Sample outbreaks . . . . .	15
3.4	Pivot tables . . . . .	16
<b>4</b>	<b>Ongoing User Interface Work</b>	<b>17</b>
<b>5</b>	<b>Tentative Schedule</b>	<b>19</b>

## List of Tables

1	Revision history . . . . .	1
2	Abbreviations . . . . .	7
3	Daily distribution . . . . .	9
4	Cities present in the semi-synthetic data . . . . .	10
5	Diseases present in the semi-synthetic data . . . . .	11
6	Tentative schedule . . . . .	19

## List of Figures

1	TCube web interface	8
2	File upload/clear panel	9
3	Query selection panel	9
4	Time series visualization panel	12
5	Temporal scan analysis	13
6	Time series operations panel	14
7	Alerts showing increased activity (upper tail tests)	14
8	Time series massive screening panel	15
9	Food poisoning outbreak in Monaragela	16
10	Leptospirosis outbreak in Colombo	16
11	Viral hepatitis outbreak in Kandy	17
12	Spatial data visualization of all diseases on the map	18
13	Attribute selection panel under spatial analysis	19
14	Spatial data visualization for few cities and diseases	20
15	MBSS results for disease Dengue fever on July 3, 2008	21
16	Food poisoning outbreak around Kurunegala on Aug 13, 2008	22
17	Leptospirosis outbreak around Colombo on Mar 26, 2008	23
18	Dysentery outbreak in central Srilanka on Apr 22, 2008	24
19	Pivot table for semi-synthetic data	25
20	Leptospirosis disease counts in Colombo using pivot table	26

# 1 Overview

This project, titled “Evaluating a Real-Time Biosurveillance Project: A Pilot Program”, is aimed at testing the feasibility of setting up a bio-surveillance system in the Kurunegala district (Srilanka) and Sivaganga district (Tamil Nadu, India). Auton Lab at Carnegie Mellon University (Pittsburgh, USA) will collaborate with teams in Srilanka (LIRNEasia) and India to provide statistical analysis interface for the data collected during this project. The statistical tools will include both time series analysis and spatial analysis with a user interface accessible over the WWW.

Auton Lab has worked on a wide variety of statistical projects in the past. The lab has many years of experience working with bio-surveillance algorithms. The goal of this project is to provide tools for prospective bio-surveillance that will generate timely alerts resulting in efficient response to real disease outbreaks.

## 1.1 Product perspective

Auton lab will provide disease outbreak detection algorithms. Both temporal analysis and spatial analysis will be performed on the data collected during the project. Auton Lab will use efficient data structures such as T-Cubes [2] to quickly retrieve data in response to the user queries during investigation and data mining.

Using the interface, called as the TCube Web Interface (TCWI), users will perform a wide variety of activities including data visualization via time series and maps, statistical analysis, automated analysis views during logins, etc. We intend to show that the front end system can be scaled up to the national level bio-surveillance system without any problems. We also intend to evaluate the product using synthetic outbreaks injected in the real data collected during the project.

## 1.2 User characteristics

For statistical data mining via TCWI, users must have basic statistical data mining training. Users must be familiar with moving-average, significance testing, p-values, CUSUM, etc. Note users that intend to only visualize time series data or maps may not have any experience with statistical analysis. We hope the interface will provide helpful information to a wide range of users. Also, the users training time to get acquainted with the interface is expected to be only a few hours.

## 1.3 Requirements

Here are the list of requirements for the project.

- TCWI works in server-client configuration. The server must run on a Linux machine with Sun Java<sup>TM</sup> and Apache Tomcat web server installed. It is required that the server is active at all times when the users are interacting with the TCWI.
- The TCWI client can run under Linux or Windows operating system but must be accessed only with either Internet Explorer or Mozilla Firefox browsers.

- The data loaded by TCWI must be both temporal and spatial in nature. For temporal analysis, the data must contain *Date* attribute with reasonable history to learn the model baselines. For spatial analysis, the data must contain location attribute such as city or village of each patient.
- TCWI internally converts each location into its corresponding latitude and longitude for map plotting purposes. Hence, all information required to plot the map (location boundaries, latitude, longitude, etc.) will be required by the TCWI so that the server can be successfully build.

## 1.4 Constraints

Since this is a prototype project, here are a few constraints of the TCWI that will not be relaxed during the scope of this project.

- **Software constraints:** The TCWI server must run on Linux 32/64-bit machine. Packages required for the TCWI server are Apache Tomcat server (Ver. 6), and Sun Java<sup>TM</sup> runtime (Ver. 1.5 or better). The TCWI client runs under Internet Explorer (Ver. 6 and 7) or Mozilla Firefox (Ver. 2.0, 3.0). The client will also need Sun Java<sup>TM</sup> (Ver. 1.6 or better).
- **Hardware constraints:** The TCWI server must have atleast 2.0 GB RAM and Pentium 4 or better PC or processor. The TCWI client should have atleast 128 MB free memory available for the application.
- **Data constraints:** Except an optional *Count* attribute, all attributes in the data must be categorical (see Section 2.1) in nature. Since most of the data analysis is temporal in nature, the data must contain a *Date* attribute.
- **Data loading constraints:** The project data is intended to be on a MySQL database. TCWI will dump the data daily at a pre-specified time from the database and then load it into the temporary cache called TCube. During this process the interface will not allow any server access. All users looged in at this time will have to re-login after the server restarts. We are expecting the data loading time to be less than 30 mins.
- **User constraints:** There are no constraints for users who want to use the interface for data visualization. For data mining and statistical analysis, users must know basic statistical algorithms and terminology (a few listed under Section 1.2).
- **Response constraints:** For a few months all alerts generated by the system must not be treated as real alerts. Auton lab, with the help of local group of data analysts, will configure various input parameters for the algorithm to increase trust in the algorithm response. Once there is reasonable confidence in the algorithm results further action (response teams, investigation procedures, etc.) could be initiated.
- **Interface constraints:** Since this is a prototype, the user interface will have limited functionality catered towards addressing the major goals of the project: is the system feasible?

is it helpful in detecting outbreaks? can the system scale upto country level? The interface has limited functionality such as no alert reports, no database requirement for storing past alerts, etc. These features along with other user feedback will be collected in response to the project for future development. For now, the interface features will be limited to the ones described in this report.

## 2 Definitions, Abbreviations

### 2.1 Definitions

1. **Demographic Attribute:** A column or attribute in data that defines the population characteristics. For a patient data set, the demographic attributes include: *Village, Disease, Symptoms, Signs, Gender, Age group, etc.*<sup>1</sup>
2. **Categorical Attribute:** For each record in the data, a categorical attribute can take exactly one value from a finite set of distinct values. E.g., For each patient, *Disease* can be one of the 10 possible diseases present in the data and hence is a categorical attribute. *Height* of a patient is a non-categorical attribute as it has infinite values.

### 2.2 Abbreviations

Table 2 lists various acronyms in this report.

Table 2: Abbreviations

Name	Description
AD-Tree	All-Dimensional tree
CUSUM	CUMulative SUM
FDR	False Discovery Rate
MBSS	Multivariate Bayseian Spatial Scan
T-Cube	Time series Cube
TCWI	TCube Web Interface

## 3 Functionality

This section describes the functionality of the TCWI. There are three main data analysis or investigation components: Time series analysis, Spatial analysis, and Pivot tables. Figure 1 shows these components on the interface. Each component is described briefly below. A few outbreak examples using the data described in Section 3.1 are also presented in this section.

<sup>1</sup>Apart from demographic attributes, TCWI requires a *Date* attribute and an optional *Count* attribute.

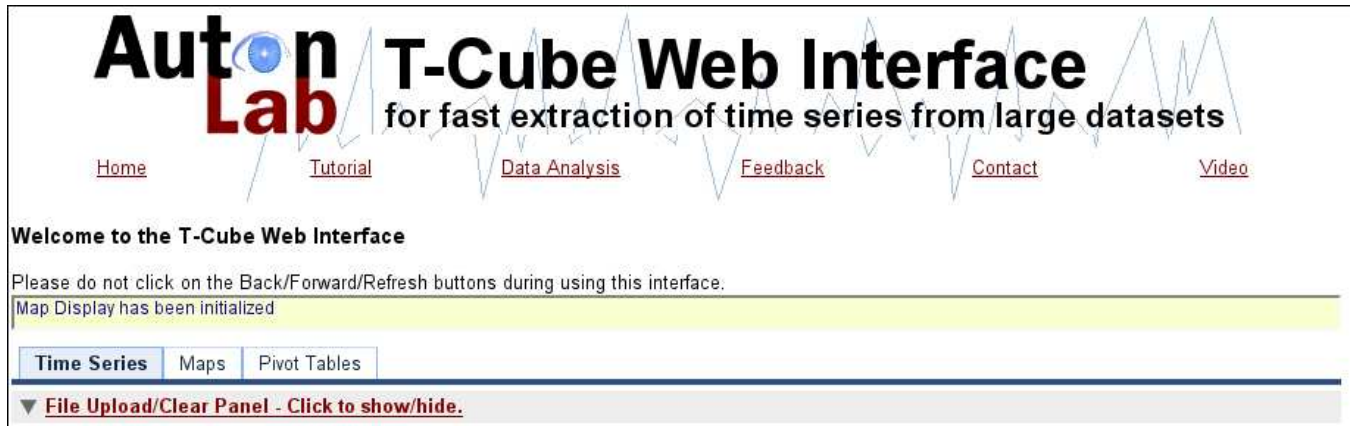


Figure 1: TCube web interface

### 3.1 Semi-synthetic data

We analyzed the weekly epidermeological data [3] using the TCWI. The data time period started from *Dec 23, 2006* and ended on *Feb 6, 2009*. As TCWI works on daily data, we converted the real weekly epidermeological data into a daily semi-synthetic data assuming the daily distribution as shown in the Table 3. We randomly assigned each patient in a specific city having a specific disease to one of 7 week day's according to the Table 3 distribution. The list of 26 cities and 9 diseases present in the data are shown in Table 4 and Table 5 respectively. The post-processed data contained 68,390 records. Here are a few example records present in the data.

```

date,lk_city,disease,count
DEC-23-2006,Colombo,Dengue_fever,7
DEC-23-2006,Kandy,Dysentery,2
...
...
JAN-30-2008,Kegalle,Viral_Hepatitis,5
JAN-31-2008,Colombo,Dengue_fever,6
...
...
FEB-06-2009,Kegalle,Dengue_fever,2
FEB-06-2009,Ratnapura,Viral_Hepatitis,1

```

### 3.2 Time series analysis

This section describes the time series component of the TCWI. First we describe the various panels available (along with their brief usage) under time series analysis. We then provide a few outbreaks found in the semi-synthetic data (Section 3.1) using the time series component.



Table 3: Daily distribution

Day-of-Week	Proportion
Monday	12%
Tuesday	8%
Wednesday	18%
Thursday	22%
Friday	17%
Saturday	13%
Sunday	10%

### 3.2.1 Panels

1. Figure 2 shows the **File Upload/Clear** panel. This panel can be used to select the data file to be loaded by the TCWI. For our case the file contains the semi-synthetic data.

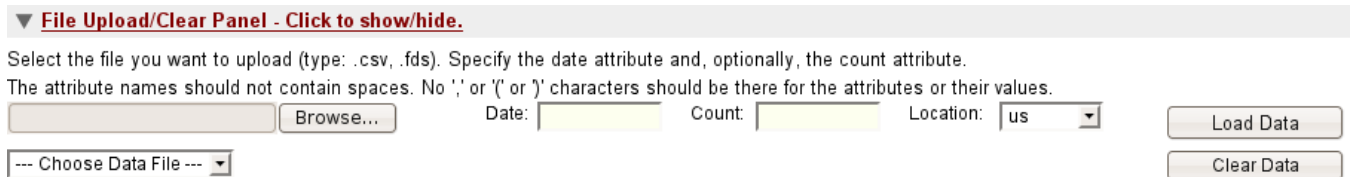


Figure 2: File upload/clear panel

2. Figure 3 shows the **Query Selection** panel. The two demographic attributes (see definition in 2.1) present in the data are patient *lk\_city* corresponding to the city and patient *disease*. This panel select the query used to create the time series in the time series panel. The user can select multiple values for either attributes and the corresponding time series will be displayed on the time series panel instantaneously. In Figure 3 we have selected two cities (*Colombo* and *Ampara*) and one disease (*Dengue fever*).

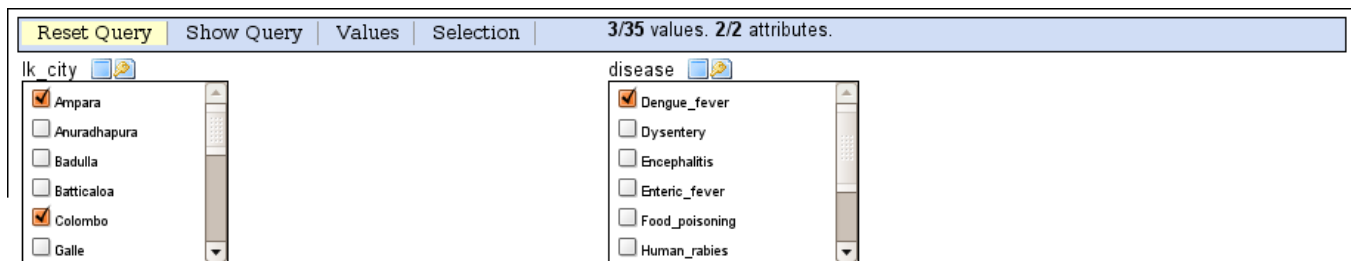


Figure 3: Query selection panel

3. Figure 4 shows the **Time Series Visualization** panel. This panel shows the time series corresponding to the query selected in query selection panel. The alerts generated by the time

Table 4: Cities present in the semi-synthetic data

City
Ampara
Anuradhapura
Badulla
Batticaloa
Colombo
Galle
Gampaha
Hambantota
Jaffna
Kalmunai
Kalutara
Kandy
Kegalle
Kilinochchi
Kurunegala
Mannar
Matale
Matara
Monaragela
Mullaitivu
Nuwara Eliya
Polonnaruwa
Puttalam
Ratnapura
Trincomalee
Vavuniya

series analysis algorithms will also be shown in this panel. Simple trend analysis such as moving average, CUSUM are available to the user for quick trend analysis. In Figure 4 the *blue* time series shows the counts on each day and *purple* time series shows the corresponding 7-day moving average. The top two scrolling bars can be used to change the start or end date of the time series seen in the panel and the middle scrolling bar (right below the top two scrollers) can be used to pane the time series shown in the panel.

This panel can also be used to perform temporal scan that detects significant changes in distributions suggesting potential disease outbreaks in the data. A temporal scan is performed by moving a window (scan window size parameter) along the time series and compare the distribution of counts inside the window and outside the window versus the distribution of counts for some baseline time series data. These distributions can be compared using hypothesis testing by setting a standard 2x2 fisher-exact test in case of low count values and

Table 5: Diseases present in the semi-synthetic data

Diseases
Dengue fever
Dysentery
Encephalitis
Enteric fever
Food poisoning
Human rabies
Leptospirosis
Typhus fever
Viral hepatitis

chi-square test in case of high values. Tests resulting in high significant values (i.e. low p-values) mean there has been a change in distribution. There are many methods to estimate the baseline: Retrospective (default), Prospective, Moving-average (day of week), Regression (day of week), Regression (day of week, last week mean) and univariate.

Figure 5 shows the result of running the temporal scan (scan window = 7) in *orange* time series on the *blue* time series. All peaks obtained after FDR correction are shown in the *orange* time series (Y-2 axis) and are significant at  $\alpha = 0.01$ .

4. Operations Panel Figure 6 shows the **Operations** panel under time series analysis. This panel can be used to perform investigation such as peak analysis and range analysis. Peak analysis can be used to find the subset of data that best explains a peak on a particular day. Range analysis can be used to compare distributions of two time durations: today versus same day last week, this week versus last week, this week versus last month, etc. This panel also allows users to only look at instances in the time series where the counts increased or decreased drastically as compared to expected (baseline). Sometimes the analysts are only interested in increases as increase in number of patients is of more concern than decrease. As an example, Figure 7 shows the alerts (FDR corrected) when there were increase in counts compared to expected (*red* time series).
5. Massive Screening Panel Figure 8 shows the massive screening panel. There are many possible queries over the demographic attributes each of which results in a unique time series. Usually an analyst working with the TCWI does not know the query that contains a potential disease outbreak. Since there are exponentially many queries, running and analyzing each time series manually might be very time consuming. This panel automates the process of massively screening lots of time series and reporting a list of alerts for the analyst to investigate.

Figure 8 shows the results of running temporal scan on all possible queries containing exactly one city and one disease. For each query, the algorithm first runs temporal scan. The algorithm then finds if there is any increased activity at 0.01 significance and finally sorts

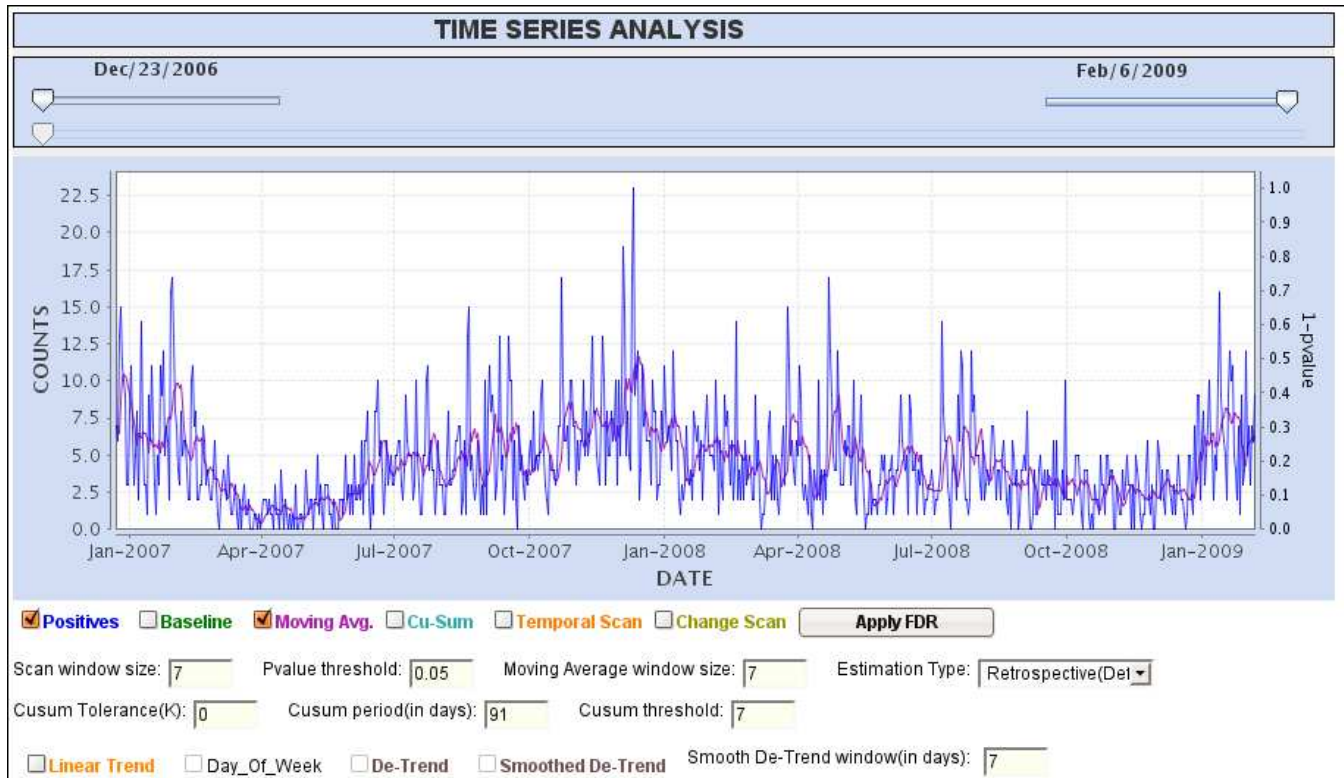


Figure 4: Time series visualization panel

all queries in the order of significance. The list at the bottom of the Figure 8 shows that there is a potential disease outbreak of *Dysentery* in city *Kilinochchi* on May 22, 2008 with significance  $\alpha < 1e^{-10}$ .

### 3.2.2 Sample outbreaks

Here are a few sample outbreaks that were found in the semi-synthetic data using massive screening panel. The temporal window was set to 7 days and prospective method was used for baseline estimation. We also applied FDR correction to plot the figures in this section.

1. Figure 9 shows a food poisoning outbreak in Monaragela on June 12, 2008. Note that the *orange* time series acting as an alert indicates that the algorithm quickly detects an emerging outbreak and could potentially save lot of lives if acted on quickly.
2. Figure 10 shows a Leptospirosis outbreak in Colombo on Sep 12, 2008.
3. Figure 11 shows a Viral hepatitis outbreak in Kandy on June 1, 2007.

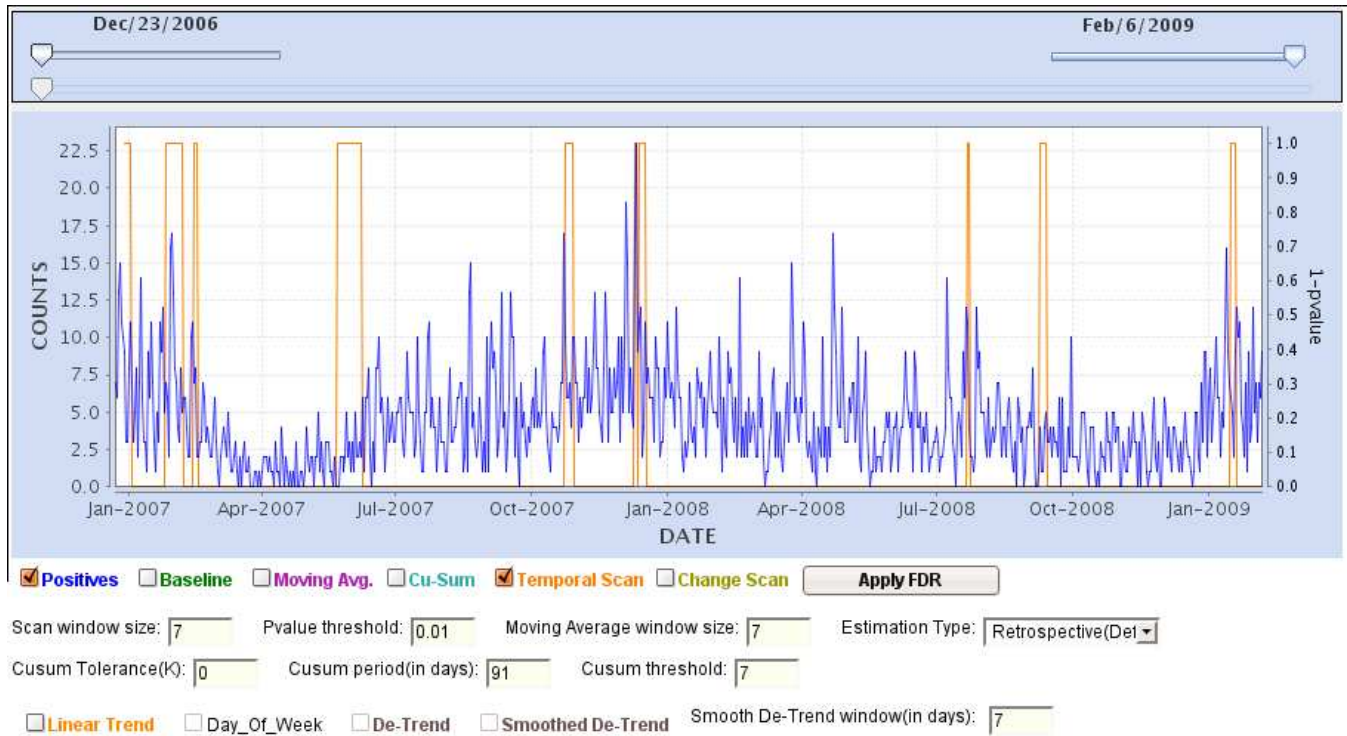


Figure 5: Temporal scan analysis

### 3.3 Spatial data analysis

The semi-synthetic data is both temporal and spatial in nature. Similar to time series analysis, TCWI can also assist in spatial analysis using **Map** component shown in Figure 1. This component helps both visualize the spatial data on the map (Section 3.3.1) as well as perform spatial scan statistics (Section 3.3.2) to determine spatially correlated disease outbreaks.

#### 3.3.1 Spatial data visualization

Given a set of diseases  $D$ , this feature of TCWI draws circles centered around each city present in data with radius proportional to the total number of patients having any disease in  $D$ . Figure 12 shows the total number of patients reported in each city for the total time duration in the data. The legend for the circles drawn on the map is shown below the map. Also, the time series drawn with *blue* bars at the bottom of the Figure 12 shows the total number of reported patients on each day in all cities and all diseases. The *green* color trend on the time series plot represents the 28-day moving average. The time-period of data to be visualized can be changed by using the scrolling bars to the right of time series plot.

It is also possible to subselect the spatial locations and the diseases using the attribute/panel. The attribute panel is shown in Figure 13 and the corresponding spatio-temporal plots are shown in Figure 14. Note that the data represented in Figure 14 is for the year 2008.

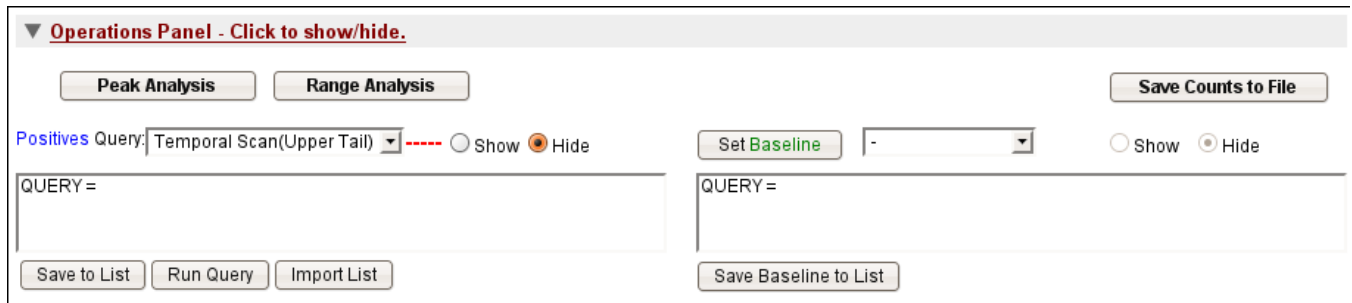


Figure 6: Time series operations panel

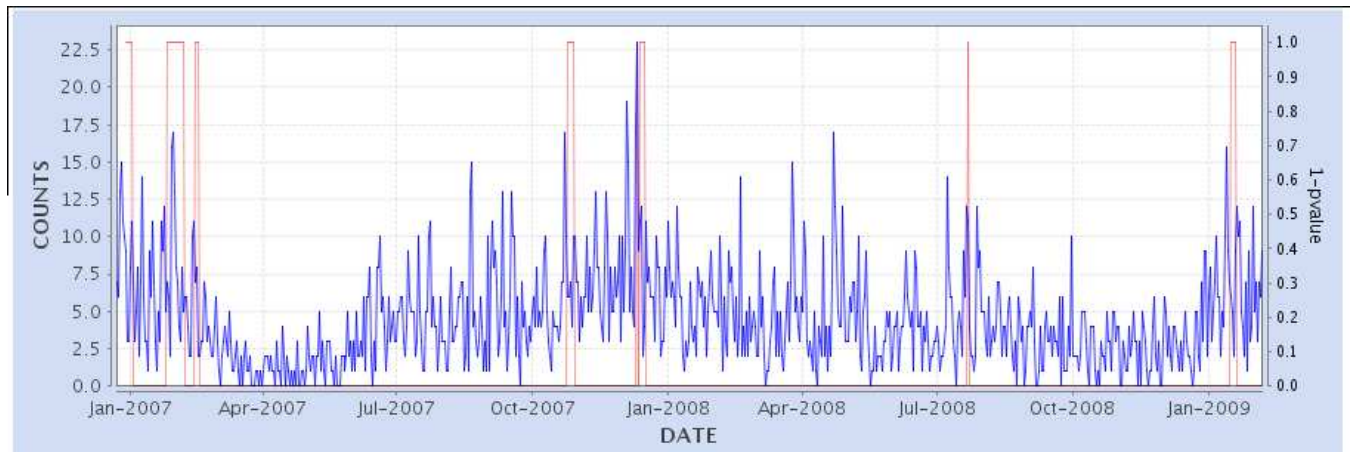


Figure 7: Alerts showing increased activity (upper tail tests)

### 3.3.2 Multivariate Bayesian spatial scan (MBSS)

Usually disease outbreaks are often restricted to one disease but rarely to one city. Multivariate Bayesian spatial scan [1] algorithms performs the spatial analysis and finds potential outbreaks spread over various cities. The algorithm requires two input parameter: temporal window size (duration of window for baseline estimation; default: 1) and maximum group size (maximum number of cities that can be affected during an outbreak; default: 20). The complete description of the algorithm can be found in the paper [1].

Figure 15 shows the result of running MBSS for disease Dengue fever on July 3, 2008 with temporal windows size = 7 and maximum group size = 20. The spatial scan global score (right top), 0.9934, represents the total probability of an outbreak on the day of analysis shown as the last day visible in the time series. The *red* circles on the map indicate the regions that are affected during the outbreak. The legend for colored circles is shown in the left (more *red* means higher probability of outbreak and more *yellow* means no outbreak). The *blue* circles on the map still indicate the total number of patients having Dengue fever in each city.

The *red* time series on the time series plot indicates the spatial scan global scores (Y-2 axis) for each day. It is evident from this time series that there are a few more outbreaks with high

**▼ Massive Screening Panel - Click to show/hide.**

Select attributes (upto3) and specific values(optional) to run massive screening. You will be able to select specific values only if the attribute is selected.)

Temporal Scan   
 Change Analysis   
Scan window sizes:

Start Date(mm/dd/yyyy):    
End Date(mm/dd/yyyy):    

Use Hierarchical Attribute   
 Rank Pvalues

ik_city	disease
Ampara	Dengue_fever
Anuradhapura	Dysentery
Badulla	Encephalitis
Batticaloa	Enteric_fever
Colombo	Food_poisoning

SAVED QUERY LISTS   

```

ik_city=(Kilinochchi),disease=(Dysentery) --- PValue=0.000000(MAY-22-2008),Window=7,0.005148, once in 194 days
ik_city=(Batticaloa),disease=(Typhus_fever) --- PValue=0.000000(JUN-04-2007),Window=7,0.009009, once in 111 days
ik_city=(Ratnapura),disease=(Food_poisoning) --- PValue=0.000000(JAN-04-2008),Window=7,0.002574, once in 389 days
ik_city=(Monaragela),disease=(Food_poisoning) --- PValue=0.000000(JUN-12-2008),Window=7,0.002574, once in 389 days

```

Figure 8: Time series massive screening panel

probability in the past.

### 3.3.3 Sample outbreaks

Here are a few examples of outbreaks found using MBSS. The temporal size window was set to 7 and maximum group size was set to 20 for all examples.

- Figure 16 shows a food poisoning outbreak around Kurunegala on Aug 13, 2008. It looks like 6 cities were affected with food poisoning cases.
- Figure 17 shows a Leptospirosis outbreak around Colombo on Mar 26, 2008. The whole southwest Srilanka seems to be affected by this outbreak.
- Figure 18 shows a Dysentery outbreak in central Srilanka on April 22, 2008.

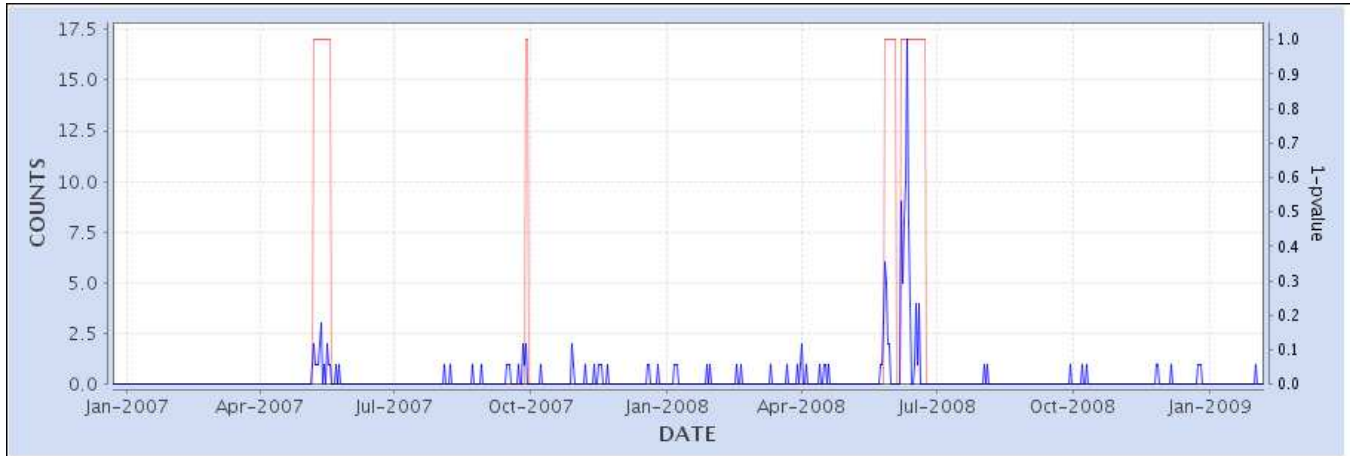


Figure 9: Food poisoning outbreak in Monaragela

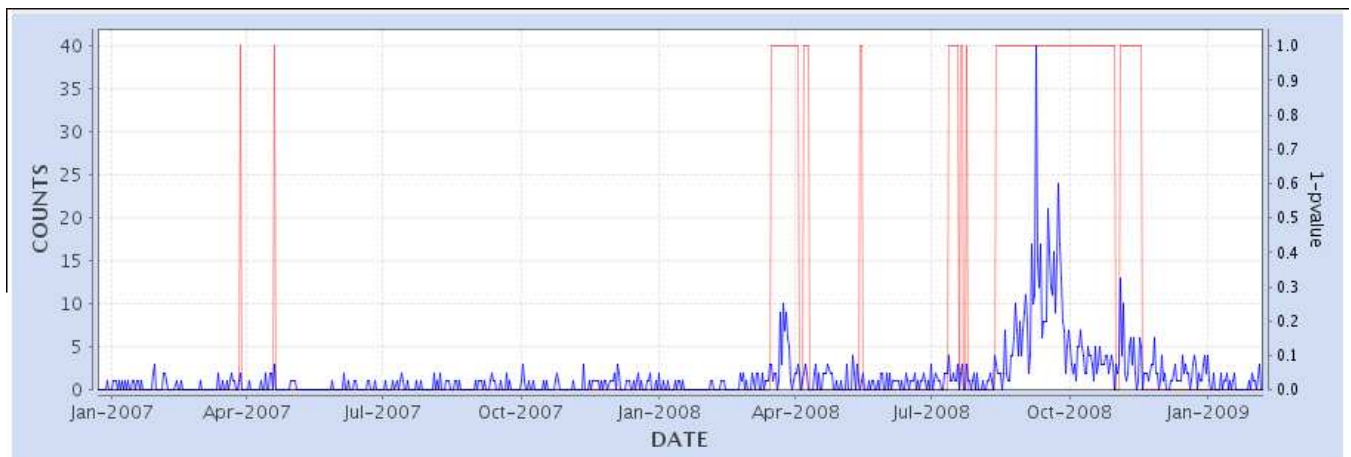


Figure 10: Leptospirosis outbreak in Colombo

### 3.4 Pivot tables

Pivot tables component can display the data in the form of a table or spreadsheet. Since semi-synthetic data only has two demographic attributes, we can only put one attribute along one dimension and other along the second dimension. Figure 19 shows the pivot table for the semi-synthetic data. Note that each cell in the table contains the total number of patients visited in the time range (shown at the bottom) selected by the user.

The pivot table also shows the marginal counts both for cities (last column) and diseases (last row). The user can also plot the time series corresponding to any cell in the table by clicking on the cell. Figure 20 shows the time series of Leptospirosis disease counts in Colombo.



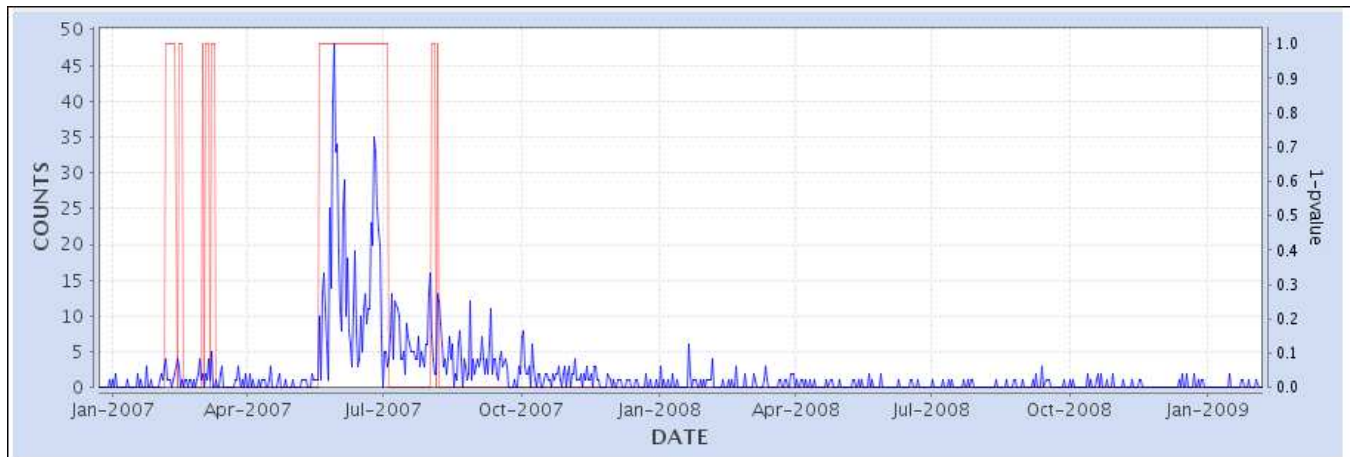


Figure 11: Viral hepatitis outbreak in Kandy

## 4 Ongoing User Interface Work

This section describes various ideas related to the project; some of them will be part of final TCWI interface.

- **Interfacing with MySQL Database**

TCWI loads data into the TCube using comma-separated file. The project data resides on the MySQL database. We propose to dump the relevant MySQL data for TCWI daily at a pre-defined time and reload the data into TCWI server cache.

- **User-level Password Access**

Currently, the TCWI has global password access control. We plan to implement user-level password protection. This will allow user management, user usage tracking, system-level performance management, load-balancing, data loading management, etc.

- **User Configuration**

We envision each user running similar kind of analysis everyday on new set of data. Even though the kind of analysis (time series analysis, spatial scan) may be the same, the input parameters to the algorithms will be most likely different for different users. This suggests saving user configuration so that they could re-run their favorite analysis during login. We hope to provide this feature in the final interface.

- **Unified Dashboard for Time Series and Spatial Scan Results**

The current TCWI interface has separate tabs for time series and spatial scan analysis even though the process of analysis has similar flavors. Both search through a set of queries and report the unusual findings. This motivates the idea for having the interface execute both scans and report the results as soon as the user logs into the system. This would help users prioritize the work by investigating the most significant alerts without manually running the scans.

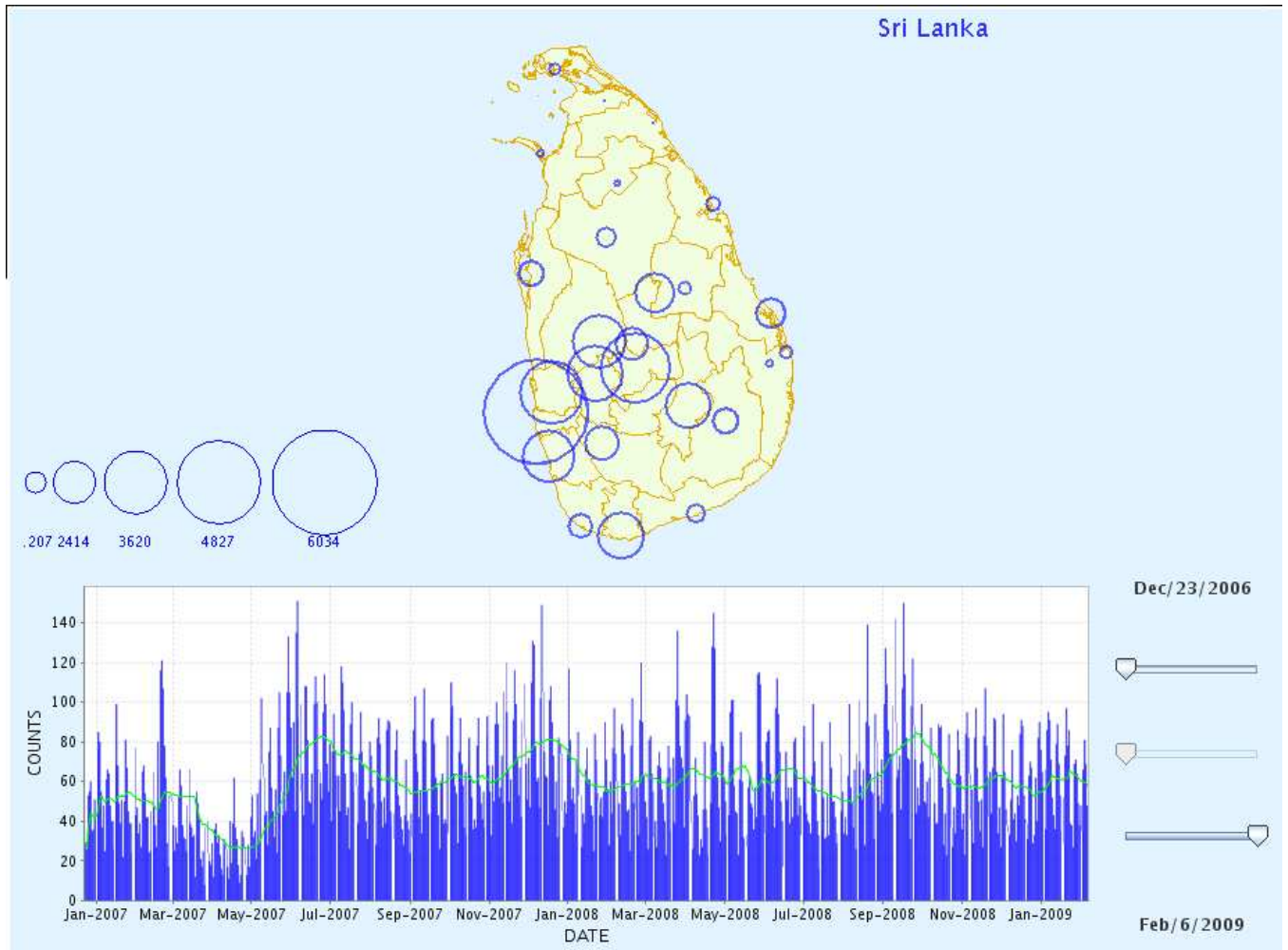


Figure 12: Spatial data visualization of all diseases on the map

- **Massive Spatial Scan**

The spatial scan can only be executed on a single disease. We propose having massive spatial scan similar to massive time series analysis that reports a prioritized list of disease alerts by scanning over all diseases individually.

- **Restricted Spatial Scan**

MBSS generates alerts for each day in the time series starting from first day in the data. Generally users are only interested in alerts for recent time (week, month, etc.). This feature will make the MBSS much more responsive and interactive as it will only need to scan a few days instead of multiple years.



Figure 13: Attribute selection panel under spatial analysis

## 5 Tentative Schedule

Table 6 details the proposed tentative schedule until Sep 15, 2009.

Table 6: Tentative schedule

Deadline	Task description
June 15, 2009	Build Interface to MySQL database and test daily server launch
June 30, 2009	User-level configuration and password access
July 15, 2009	Finalize GUI design
July 15, 2009	User training manuals
Aug 01, 2009	Go live with real data from May - July
Aug 01, 2009	Finish GUI development
Sep 01, 2009	Use real or synthetic data to visualize and investigate disease outbreaks using TCWI
Sep 15, 2009	Start running simulations

## References

- [1] Daniel Neill and Gregory F. Cooper. A multivariate spatial scan statistic for early event detection and characterization. *Machine Learning*, 2009. In Press.
- [2] Maheshkumar R. Sabhnani, Andrew W. Moore, and Artur W. Dubrawski. T-cube: A data structure for fast extraction of time series from large datasets. Technical Report CMU-ML-07-114, Carnegie Mellon University, April 2007.
- [3] Weekly srilanka epidermeological data. <http://www.epid.gov.lk/wer.htm>.

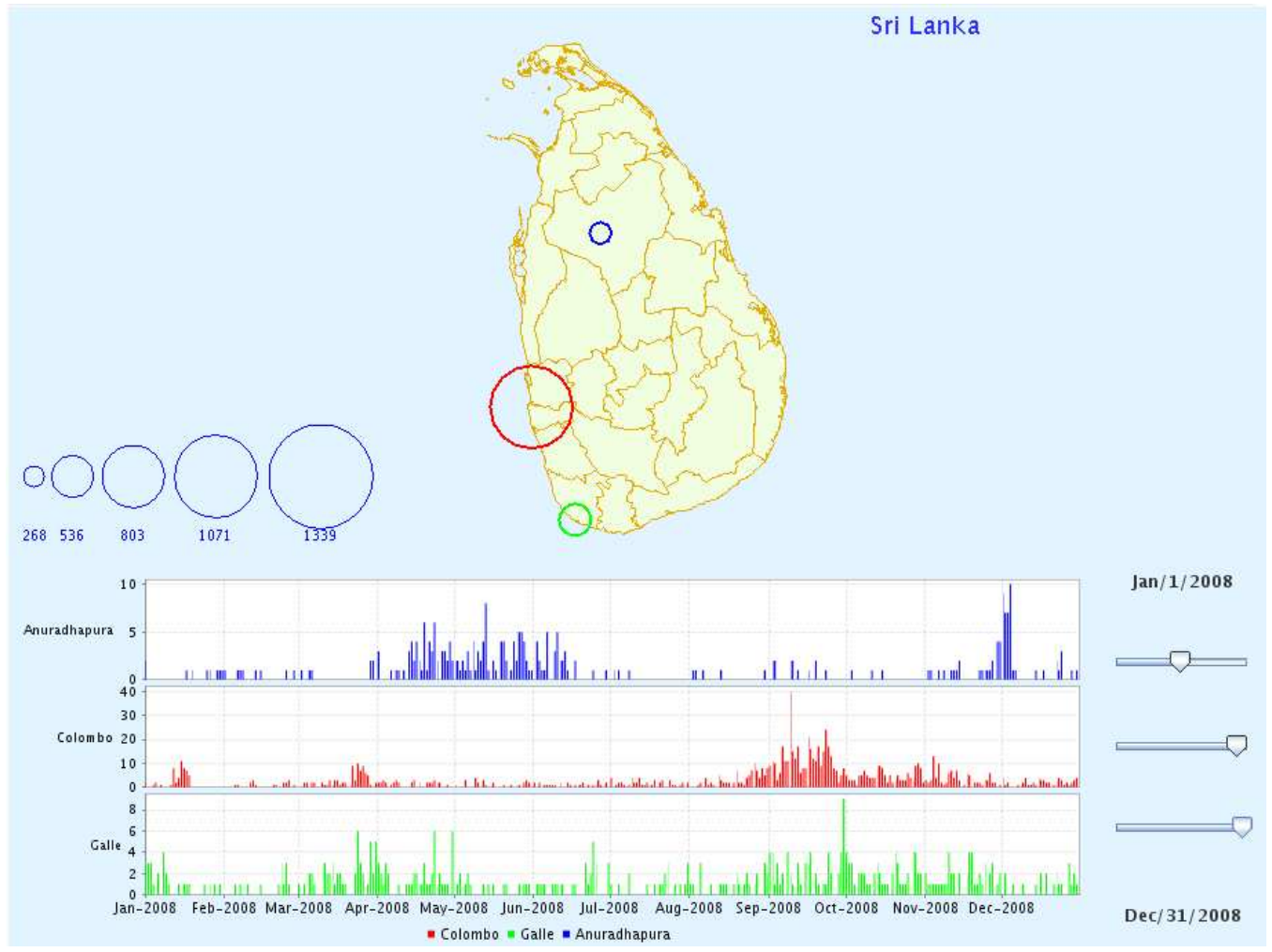


Figure 14: Spatial data visualization for few cities and diseases

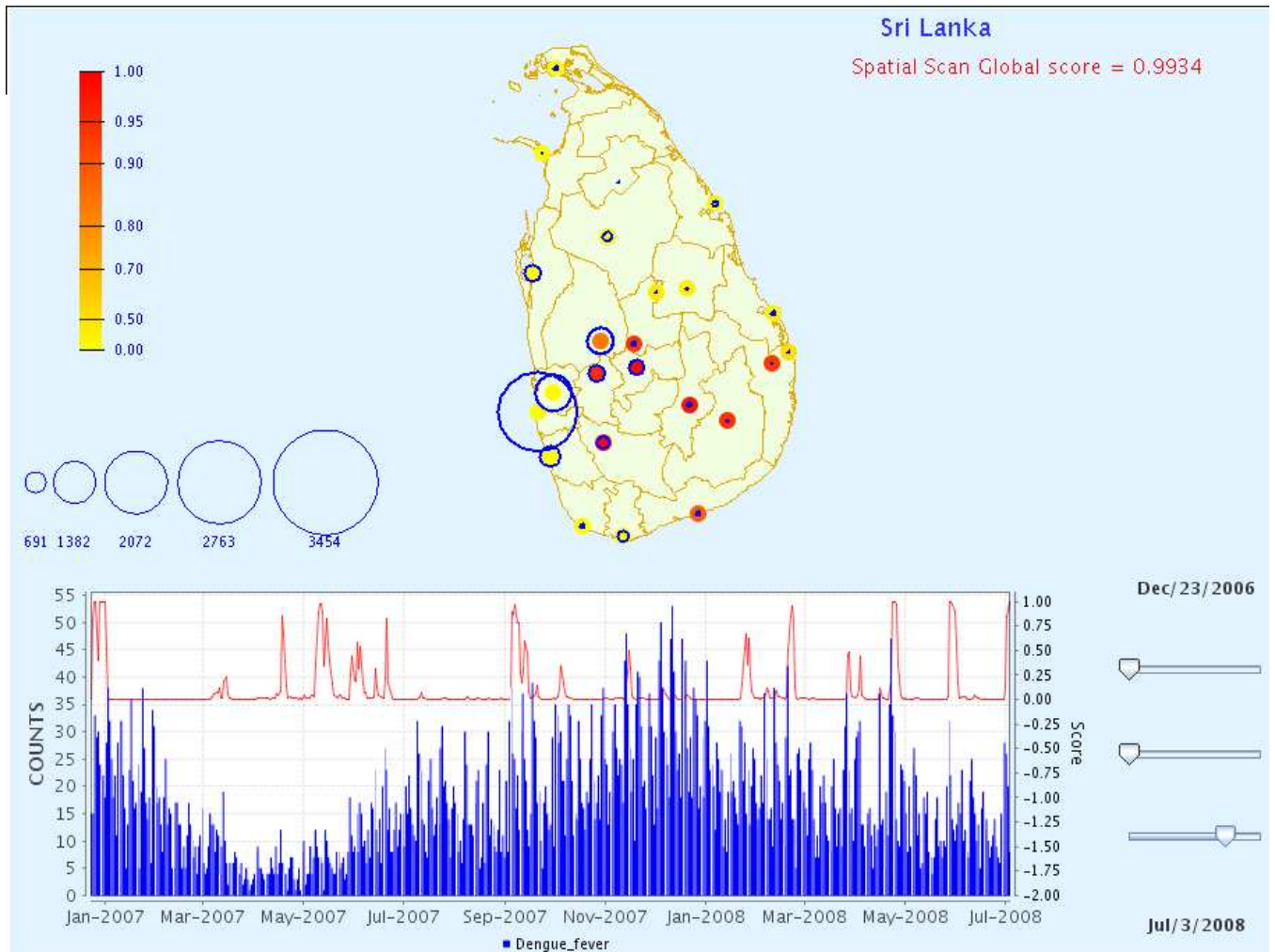


Figure 15: MBSS results for disease Dengue fever on July 3, 2008

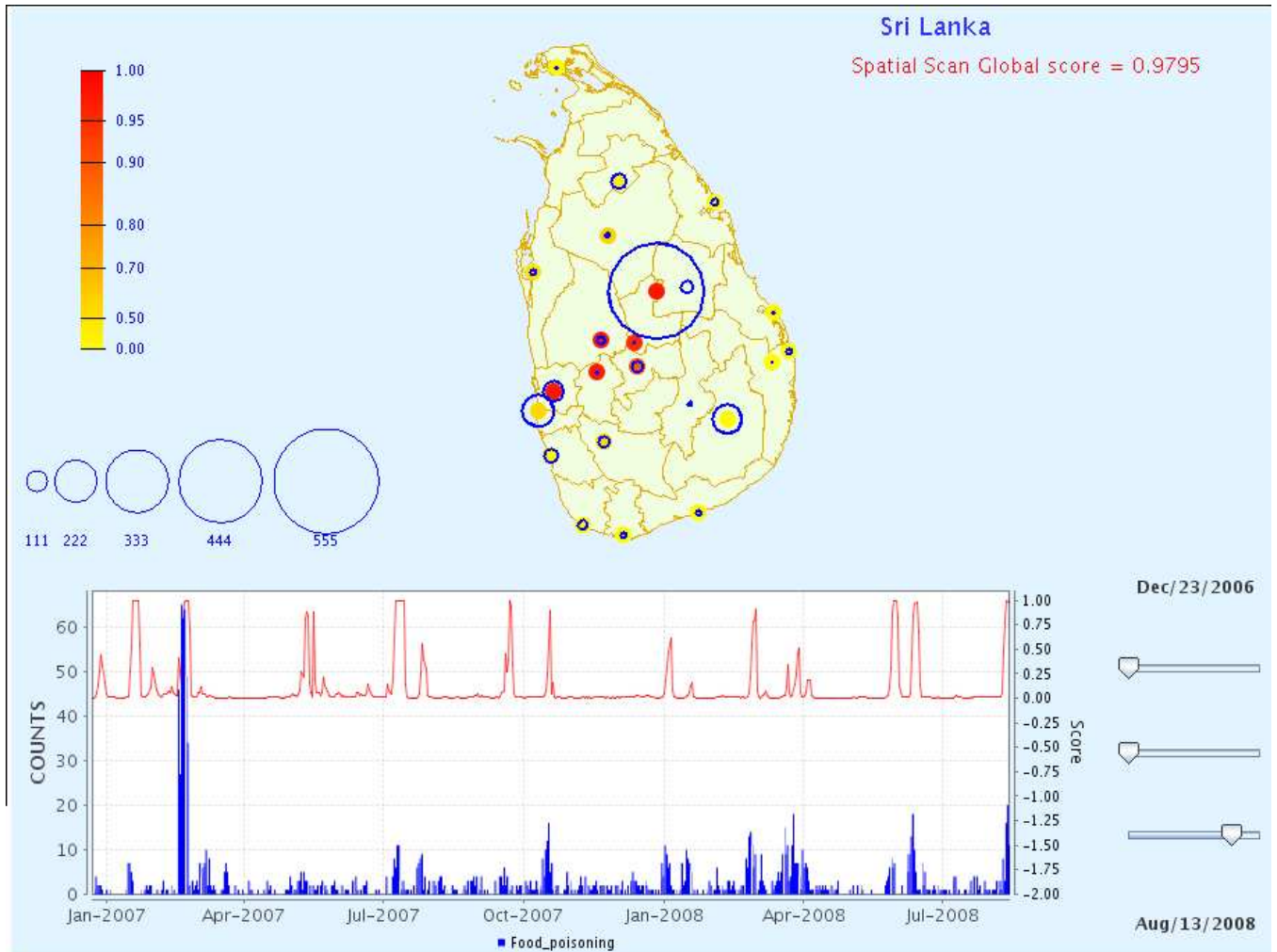


Figure 16: Food poisoning outbreak around Kurunegala on Aug 13, 2008

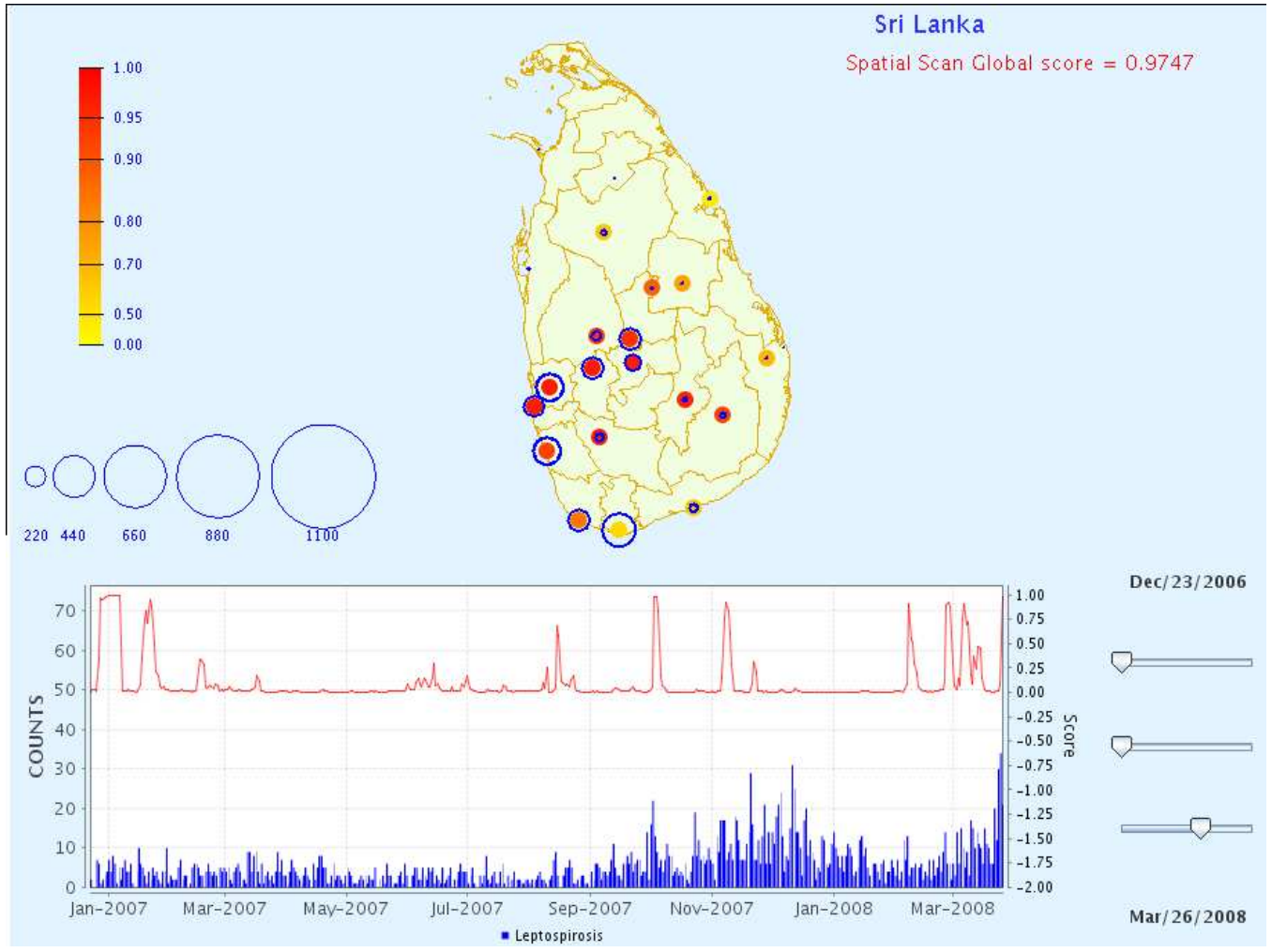


Figure 17: Leptospirosis outbreak around Colombo on Mar 26, 2008

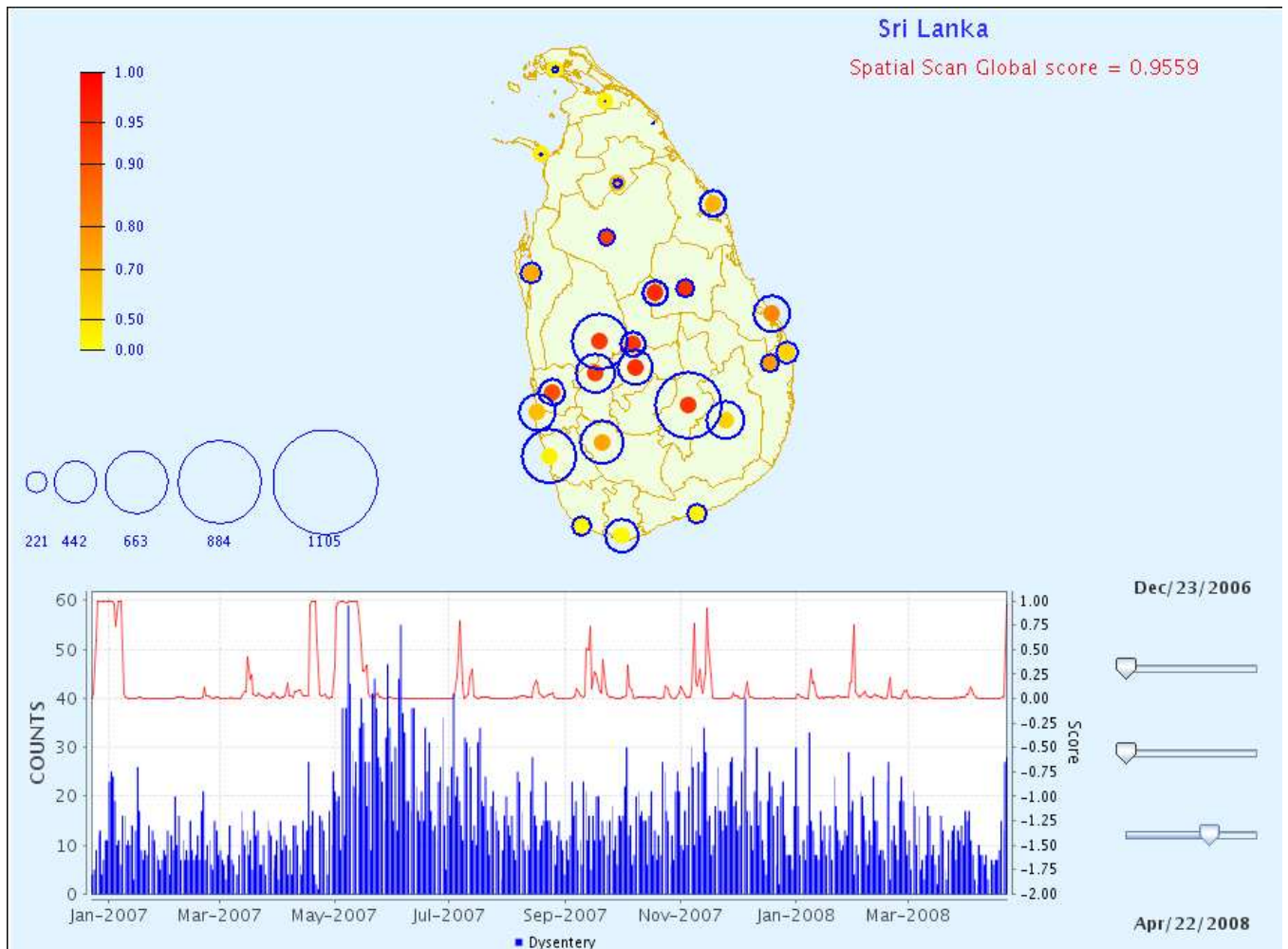


Figure 18: Dysentery outbreak in central Srilanka on Apr 22, 2008



Attributes:

Rows: lk\_city

Columns: disease

	Dengue_fever	Dysentery	Encephalitis	Enteric_fever	Food_poisoning	Human_rabies	Leptospirosis	Typhus_fever	Viral_Hepatitis	Total
Ampara	25	308	0	11	7	0	26	2	28	407
Anuradhapura	335	298	18	22	71	1	281	31	55	1112
Badulla	185	1105	18	219	112	1	125	269	577	2611
Batticaloa	163	562	21	47	20	12	8	23	837	1693
Colombo	3454	611	29	329	239	3	1100	12	257	6034
Galle	198	354	35	41	56	3	575	44	32	1338
Gampaha	1624	428	43	137	145	4	906	26	319	3632
Hambantota	242	327	19	33	42	1	197	176	45	1082
Jaffna	138	131	2	200	37	0	0	134	33	675
Kalmunai	67	472	6	21	31	0	6	4	128	735
Kalutara	912	835	22	135	98	2	841	6	100	2951
Kandy	865	639	16	121	103	1	644	199	1463	4051
Kegalle	859	560	39	142	29	1	737	113	724	3204
Kilinochchi	0	11	0	5	0	0	1	2	3	22
Kurunegala	1061	758	27	122	62	14	705	95	183	3027
Mannar	36	50	0	249	0	0	1	1	39	376
Matale	332	384	12	79	24	0	838	7	150	1826
Matara	662	555	27	95	45	2	750	475	43	2654
Monaragela	103	663	4	112	165	1	142	181	101	1472
Mullaitivu	0	13	4	9	0	0	0	0	12	38
Nuwara_Eliya	77	519	6	336	555	0	95	73	595	2256
Polonnaruwa	125	305	4	42	80	0	135	1	66	758
Puttalam	609	358	32	170	47	3	100	50	89	1458
Ratnapura	549	747	35	113	78	0	222	67	127	1938
Trincomalee	205	363	4	35	38	0	40	33	118	836
Vavuniya	52	144	6	34	91	0	10	1	17	355
<b>Total</b>	<b>12878</b>	<b>11500</b>	<b>429</b>	<b>2859</b>	<b>2175</b>	<b>49</b>	<b>8485</b>	<b>2025</b>	<b>6141</b>	<b>46541</b>

Start Date << 12/23/2006 >> End Date << 02/06/2009 >>

Figure 19: Pivot table for semi-synthetic data

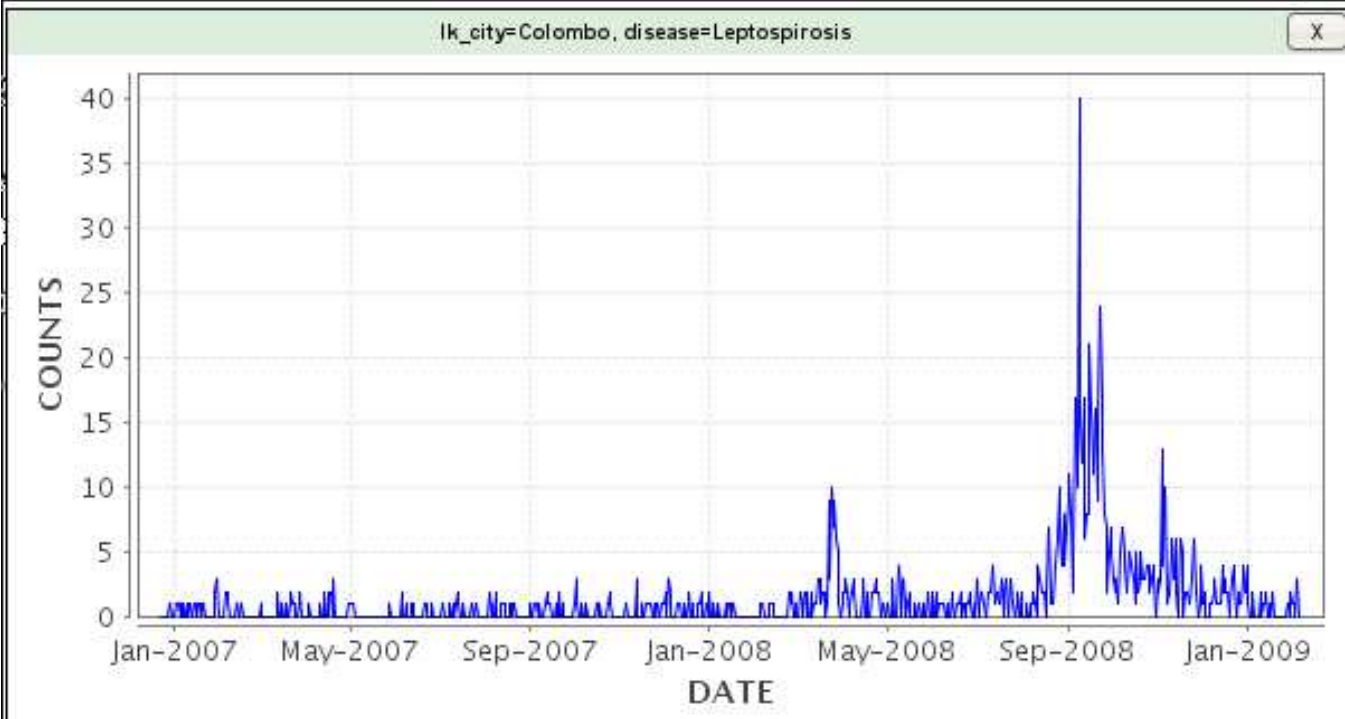


Figure 20: Leptospirosis disease counts in Colombo using pivot table