NSTD-11-0935

ESSENCE Desktop Edition (EDE) User's Guide

Compatible with 03 June 2011 EDE Release or later

June 2011

APL
JOHNS HOPKINS
UNIVERSITY

Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel MD 20723-6099

National Security Technology Department

Operating Under Contract Number: N00024-03-D-6606, Task Order: 1427

APL NATIONAL SECURITY TECHNOLOGY DEPARTMENT

NSTD-11-0935

ESSENCE Desktop Edition (EDE) User's Guide

Compatible with 03 June 2011 EDE Release or later

June 2011

This page intentionally left blank.

Table of Contents

List of Figures

1 INSTALLATION INSTRUCTIONS

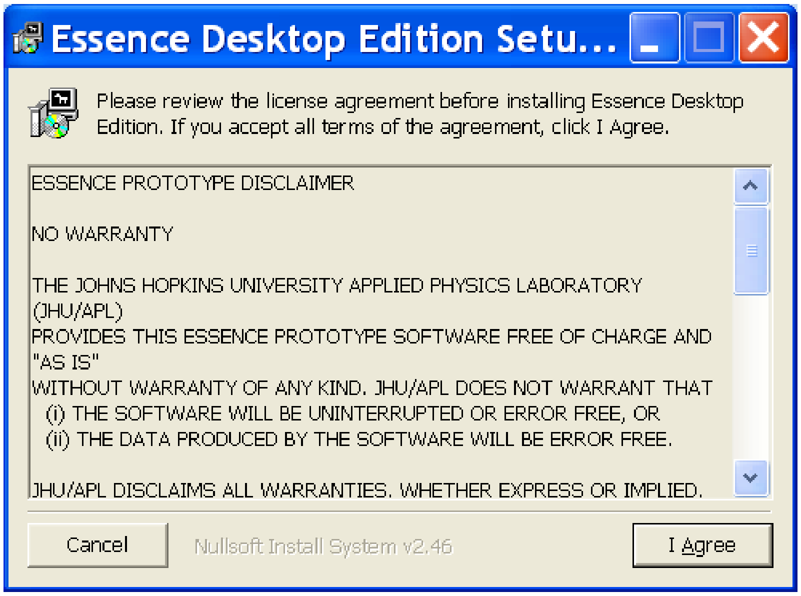
Download the most recent version of the EDE Setup .exe file, and the Cityville Example Data/Map Files from the Tools section of the SAGES website. (http://www.jhuapl.edu/Sages/) Place the files on the **C:\** drive of the computer.

1.1 EDE INSTALLATION

Download the most recent version of the EDE Setup .exe file from the Tools section of the SAGES website. (http://www.jhuapl.edu/Sages/) Place the file on the **C:\** drive of the computer.

If there is already an earlier version of EDE installed on the computer it is not necessary to uninstall the older version, especially if there were saved Data Sources and/or Saved Queries in the older version. When the June 2011 and later EDE version installs it will look for earlier versions of EDE and keep the saved Data Sources and Queries so they are available for analysis in the newly installed version.

Double click on the file name to begin installation. The disclaimer shown in Figure 1 will appear. To proceed with installation, click the **<I Agree>** button. To cancel installation click the **<Cancel>** button.



Please review the license agreement before installing Essence Desktop Edition. If you accept all terms of the agreement, click I Agree.

ESSENCE PROTOTYPE DISCLAIMER

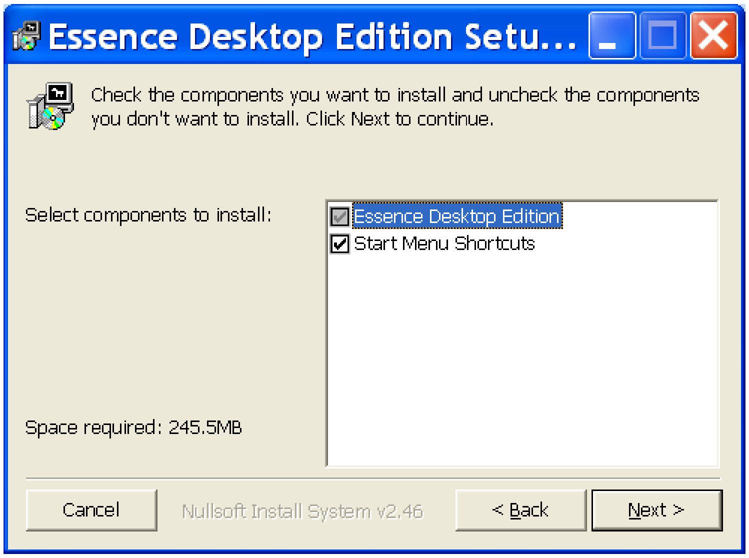
NO WARRANTY

THE JOHNS HOPKINS UNIVERSITY APPLIED PHYSICS LABORATORY (JHU/APL)
PROVIDES THIS ESSENCE PROTOTYPE SOFTWARE FREE OF CHARGE AND "AS IS"
WITHOUT WARRANTY OF ANY KIND. JHU/APL DOES NOT WARRANT THAT
(i) THE SOFTWARE WILL BE UNINTERRUPTED OR ERROR FREE, OR
(ii) THE DATA PRODUCED BY THE SOFTWARE WILL BE ERROR FREE.

JHU/APL DISCLAIMS ALL WARRANTIES, WHETHER EXPRESS OR IMPLIED,

Figure 1 ESSENCE Use Agreement Window

After clicking **<I Agree>**, the license agreement window is replaced with the window shown in Figure 2. This window gives the user the option of installing an EDE shortcut on the computer's Start Menu. Click on the <**Start Menu Shortcuts**> to select or de-select this option, think click **<Next>** to continue.


Essence Desktop Edition Setu...
Cancel
Nullsoft Install System v2.46
< Back
Next >

Check the components you want to install and uncheck the components you don't want to install. Click Next to continue.

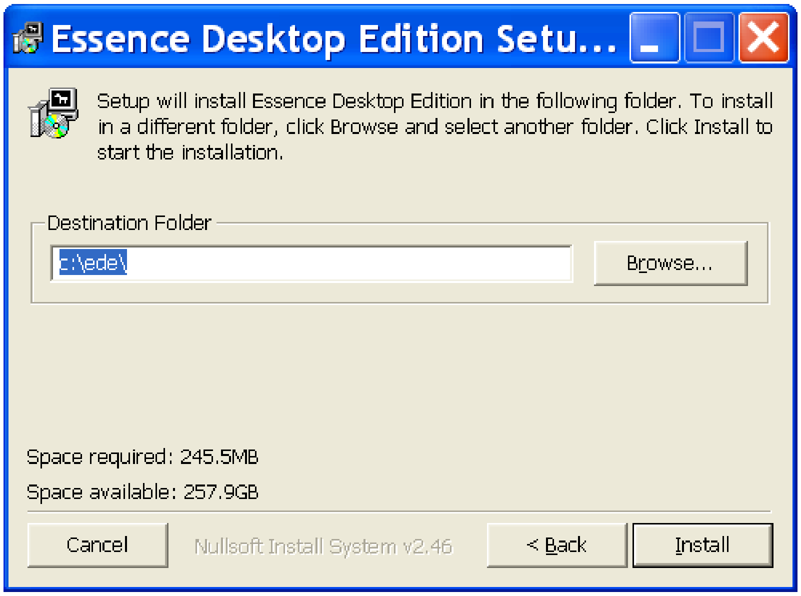
Select components to install:

☑ Essence Desktop Edition
☑ Start Menu Shortcuts

Space required: 245.5MB

Figure 2 ESSENCE Use Agreement Window

The next window that opens allows the user to select placement of the **\EDE** folder on the computer. Figure 3) To select the default, **C:\EDE\**, click the **<Install>** button at the bottom right corner of the window.   If a different directory is preferred use the **<Browse>** buton to scroll thru the directory and locate the desired location, then click the **<Install>** button at the bottom right corner of the window.


Essence Desktop Edition Setu...
Browse...
Cancel
Nullsoft Install System v2.46
< Back
Install

Setup will install Essence Desktop Edition in the following folder. To install in a different folder, click Browse and select another folder. Click Install to start the installation.

Destination Folder: c:\ede\

Space required: 245.5MB

Space available: 257.9GB

Figure 3 Identification of C:\EDE\ Sub-directory computer location

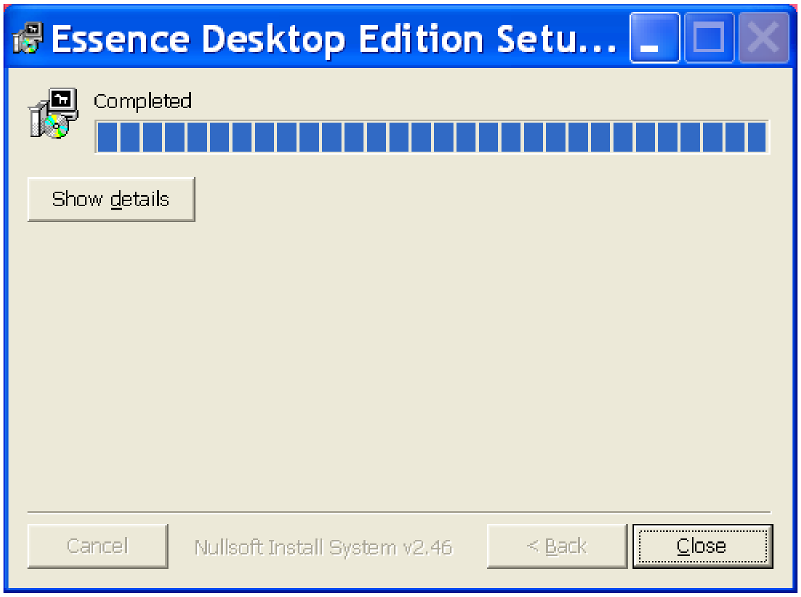

Essence Desktop Edition Setu...
Completed
Show details
Cancel
Nullsoft Install System v2.46
< Back
Close

Figure 4 EDE Installation Counter Window

The window in Figure 4 appears after the <Install> button is clicked. The bar in this window shows the progress of the installation. When it reaches 100%, click the **<Close>** button at the bottom right of the window to complete the installation process.

1.2 INSTALLING THE CITYVILLE EXAMPLE MAP/DATA FILES

Use WinZip to unzip the Cityville data and map files to **<C:\EDE>**. Once unzipped, place the files in the map files (Demo_Region5.*) in the sub-directory: **<C:\EDE\workspace\workspace\MapFiles>**. All other map files to be used with EDE should also be copied into this sub-directory. The Cityville data files (.mdb, xls, and .csv files) may be left in the **<C:\EDE>** directory or moved as desired.

2 INTRODUCTION

ESSENCE is the acronym for the Electronic Surveillance System for the Early Notification of Community-based Epidemics. It is a Web-based system used by state and city public health authorities throughout the United States to monitor disease trends in their communities. The system was conceived to help public health authorities rapidly identify disease outbreaks possibly associated with bio-terrorist attacks. Epidemiologists using the system quickly determined, however, that the technology was as useful—or more useful—in monitoring trends and outbreaks of naturally occurring diseases.

The Web-based system reads computerized health care data originally collected for other purposes, such as emergency room admission data, and categorizes the data into disease syndromes, such as fever and gastrointestinal illness. ESSENCE then runs statistical algorithms to identify “alert days” or “alerts,” which are days when the daily count (the number of records or cases reported) of a disease syndrome is statistically significantly higher than counts on previous days. The data are then presented as a time series with alerts marked as red or yellow dots. A line listing of all records included in the time series, graphs, and maps is also produced.

It became clear over time that an ESSENCE-type system would be useful in geographic areas and situations where Web access is limited and standard ESSENCE cannot be used. The ESSENCE Desktop Edition (EDE) was designed for this purpose. EDE is designed to run independently on a single desk or laptop computer. The application can read case-based data stored in a variety of different database programs. Once pointed to the appropriate machine address of a data file, EDE will automatically read the data into its own datasource file using the labels from the original data file. The user can then edit the EDE datasource to include only the desired variables. Standard ESSENCE processes syndromic data only, but EDE will process any type of case-based or aggregated temporal data, whether it is grouped by syndrome, disease name, disease type, or some other variable. EDE can be used to evaluate temporal trends in reportable disease and monthly morbidity data at the national, regional, and municipal level. It may also be useful at the clinic level depending on the volume of data collected. Like ESSENCE, EDE can create disease location maps for any geographic region for which Environmental Systems Research Institute (ESRI) shape files are available.

In summary, EDE allows a user to read case-by-case and aggregate temporal disease/event surveillance data, produce a time-series representation of that data, and run a statistical algorithm to determine whether the number of cases of disease/event reported on a given day is significantly greater than expected based on the preceding daily counts. It also produces pie and bar graphs of demographic characteristics and maps of disease counts and alerts by geographic region.

3 OVERVIEW OF EDE COMPONENTS

3.1 WINDOWS TOOLBAR OPTIONS

The **Windows Toolbar** has four options:

3.1.1 File & Window

The File and Window choices provide no options that are necessary in EDE, and should generally be ignored.

3.1.2 Perspectives

Perspectives provide the user with four possible views or perspectives of the EDE information:

3.1.2.1 EDE Perspective

The **EDE Perspective** presents the data in the standard EDE format. It should be used the perspective used most of the time. The only exception is when the user is working with an UDig map.

3.1.2.2 UDig Perspective

The **UDig Perspective** is not useful in EDE and should be ignored

3.1.2.3 Map Perspective

The **Map Perspective** should be used when an UDig map is being examined or modified.

3.1.2.4 Style Perspective

The **Style Perspective** is not useful in EDE and should be ignored.

3.1.3 Help.

Help allows the user to view two types of information: the JHU/APL EDE prototype disclaimer and the EDE version information.

3.2 WINDOWING FEATURES

The EDE main screen has multiple windows. These windows can be moved and resized within the main screen to suit the individual user. Windows can be closed by clicking on the X at the top right of the window label.

To move or resize a window within the main screen, left-click on the window's name tab and drag the window to a new position. A hatched outline of the new window location will appear as the label is dragged. When the outline is in the proper position, the window is "dropped" into place by releasing the left-click mouse button. Windows can also be dragged off onto the desktop, but they cannot be dragged back into EDE. To resize a window, click on a side or corner and pull the window to the desired size.

3.3 WINDOW OPTIONS

The basic windows needed to run EDE open automatically when the application is opened and when a query is processed. The list of available windows is found by left-clicking **Window** and then **Show View** on the main toolbar. A pop-up menu appears that shows all available windows. To open a window, left-click on the desired window name in the pull-down list. The default windows appearing on the screen include the following, indicated by circled letters in Figures 1 and 2:

Available Data View (Figure 5, A) (Section 2.3)

DataSet Manager (Figure 5, B & C) (Section 2.4)

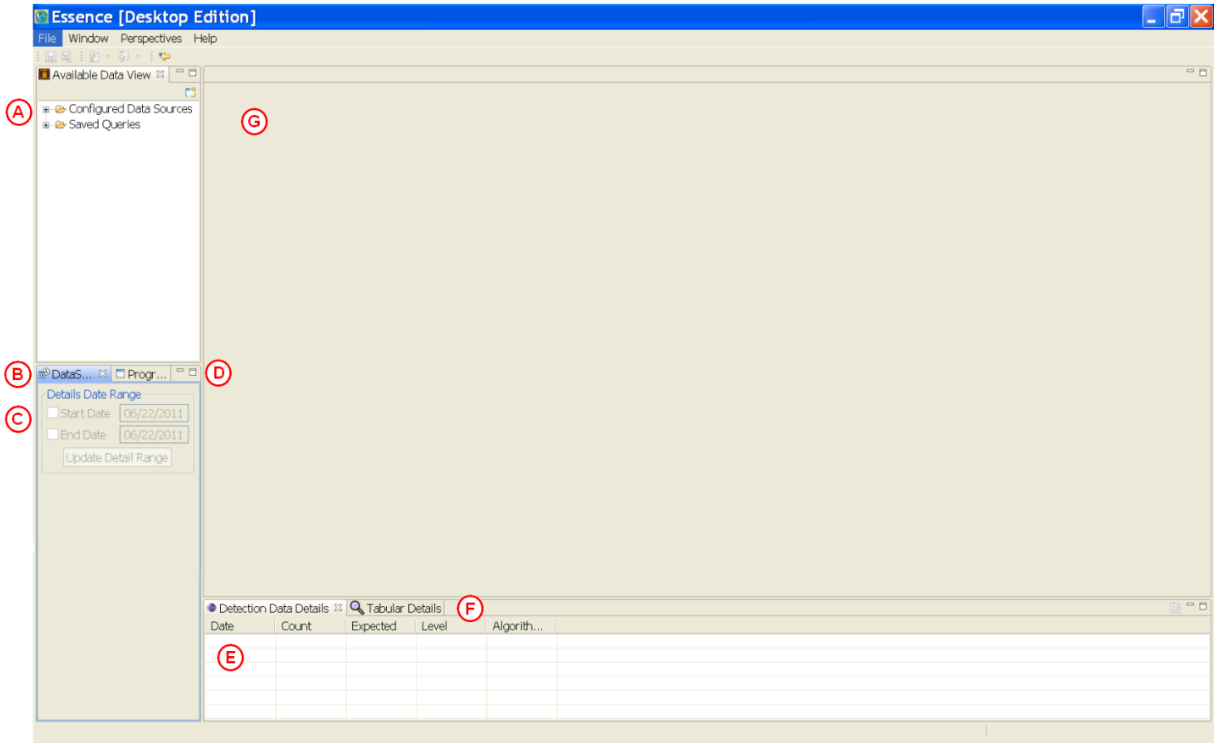
Progress (Figure 5, D) (Section 2.5)

Detection Data Details (Figure 5, E) (Section 2.6)

Tabular Details (Figure 5, F) (Section 2.7)

Query (Figure 5, G) (Section 2.8)

These windows are described in the sections noted above.



□ Start Date 06/22/2011

□ End Date 06/22/2011

Date	Count	Expected	Level	Algorith...

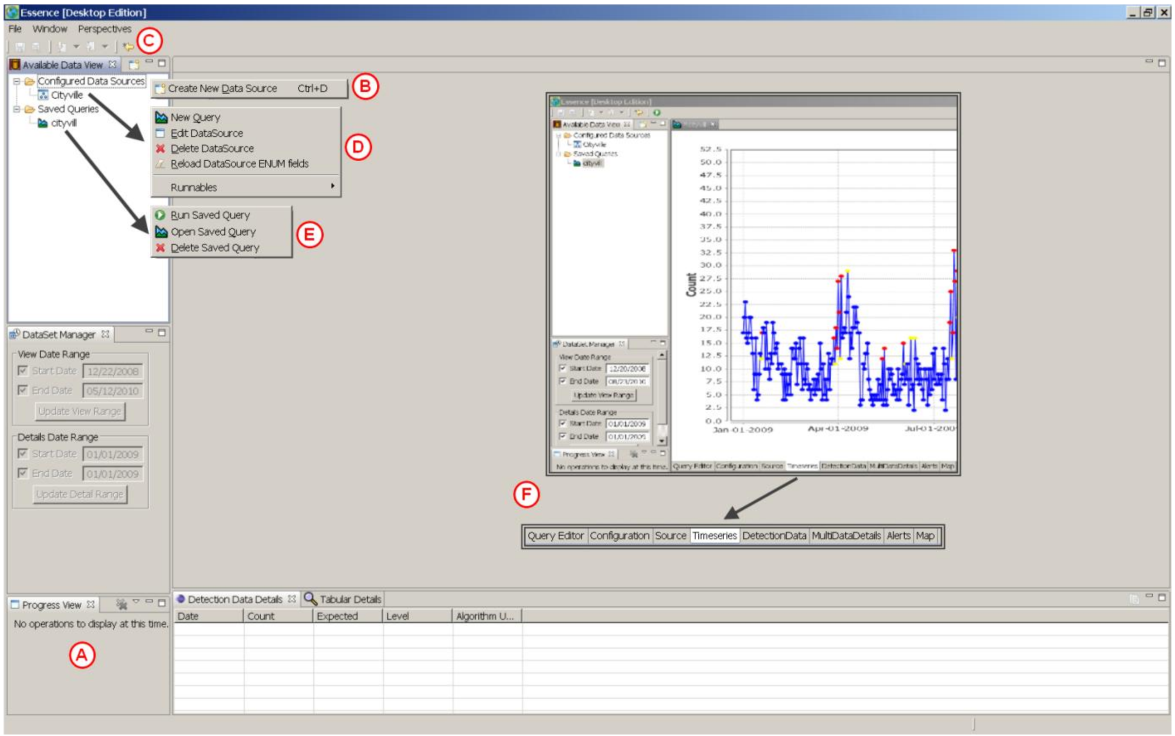
Figure 5 Opening Screen of EDE

The **Available Data View** window (Figure 5, A) shows a directory of the saved EDE data sources and queries.

3.3.1 Configured Data Sources

Right-clicking on **Configured Data Sources** produces a pop-up window that leads to a wizard (Figure 6, B) that facilitates creation of a new EDE datasource, definition, and configuration file (DSDC file). This is described in detail in section 4.2.1. The data source creation wizard can also be opened by left-clicking on the window icon that is located above **Configured Data Sources** and immediately to the right of the window name (**Available Data View**). (Figure 6, C)

Newly made EDE DSDCs can be saved and reused. Once saved, the names appear under **Configured Data Sources** when the folder is opened. (Figure 6) Right-clicking on a DSDC name in the list opens a pop-up window. (Figure 6, D) that allows the user to create a new query, edit or delete the data source, or update a DSDC that draws data from a simple text or a Microsoft Office Excel file. There is a fifth option in the pop-up window, **Runnables**, which is not necessary in EDE and should be ignored by users.



View Date Range: Start Date 12/22/2008, End Date 05/12/2010

Details Date Range: Start Date 01/01/2009, End Date 01/01/2009

No operations to display at this time.

Figure 6 EDE Main Screen Showing the Expanded Progress Window and Query Results Tabs Available Data View Window

The **Available Data View** window (Figure 5, A) shows a directory of the saved EDE data sources and queries.

3.3.2 Saved Queries

A data query is a logical statement that defines a specific subset of the DSDC and produces a time series and corresponding analysis on that subset. Queries can be saved and reused to simplify daily analysis. When available, the names of saved queries are listed in the **Saved Queries** folder in the **Available Data View** window.

Right-clicking on a saved query name produces a pop-up window that allows the user to: run the saved query on the current EDE DSDC file, edit (i.e. Open) the saved query, or delete the saved query. The option to save new queries is offered when a new query is closed.

3.4 DATASET MANAGER WINDOW

The **DataSet Manager** window (Figure 5, B) shows the data ranges for the data details [**Details Data Range** (Figure 5, C)]. This data range can be changed by clicking on the appropriate checkbox and editing the date to the right of the box. Dates must be entered as MM/DD/YYYY.

3.5 PROGRESS WINDOW

The **Progress** window (Figure 6, A) shows text that reports the progress of the current operation. In the default view of EDE, this window and the **DataSet Manager** window are stacked, with the **Progress** window located in the back. Only the named tab is visible in this view. (Figure 5, D) Users may wish to drag the **Progress** window below the **DataSet Manager** window, as shown in Figure 6, A, so that it is more easily viewed.

3.6 DETECTION DATA DETAILS

The **Detection Data Details** window (Figure 5, E) shows specific detailed information for each day for which the detection algorithm was run, regardless of the alert color (red, yellow, or white). The date, record count, expected record count, p-value generated by the algorithm, and name of the algorithm used for detection are shown for each day listed.

3.7 TABULAR DETAILS

The **Tabular Details** window (Figure 5, F) shows specific detailed information for each record included in a query. The date and all active variables included in the EDE data file are listed for each record in the query.

3.8 QUERY WINDOW

The **Query** window (Figure 5, G) is used when a query is created. The parts of the logical statement are added sequentially until the query is complete; then, the query is run. After the query is run and saved, the window is labeled with the name of the query.

In addition, after a query is run, a number of tabs appear at the bottom of the window that open the results generated by the query. Left-clicking on the tab name opens a window containing specific results of the query such as a time series. (Figure 6, F Inset) The tabbed windows produced by a query are described briefly below.

3.8.1 Query Editor

This tab shows the query that produced the current time series.

3.8.2 Configuration

The **Configuration** tab shows options that can be changed by the user, including general options to change the appearance of the time series, pie chart(s) and bar graph(s), the detection threshold values for yellow and red alerts, the detection algorithm, and the variables that can be displayed in a pie chart or bar graph.

3.8.3 Source

The Source tab shows the pseudo-Standard Query Language (pseudo-SQL) command used to create the query specified in the Query Editor window. This information is for reference only, and the query cannot be modified by changing the pseudo-SQL commands shown.

3.8.4 Time series

This tab displays the time series created by the current query. (Figure 6, Inset)

3.8.5 DetectionData

Specific detailed information for each day for which the detection algorithm produced an alert regardless of the alert level (red, yellow, or green) is listed in table format in the **DectionData** tab. The date, record count, expected record count, p-value, and name of the algorithm used for detection are listed for each alert. The same information is also shown in the Detection Data window on the default EDE Main Screen. (Figure 5, F)

3.8.6 MultiDataDetails

This tab shows a sortable list of the records (cases) included in the query and the pie/bar graphs requested for the query.

3.8.7 Alerts

The **Alerts** tab shows a sortable list of alerts by geographic region and can be set to include red, yellow, or no alert days.

3.8.8 Map

EDE exports count and alert data to the UDig (the default program) or EpiMap application. Clicking the **Map** tab opens access to the specified mapping program and presentation of the query data on a map. EDE can use the data for any geographic region (location) variable in the DSDC for which ESRI shape files are available to map query results.

4 USING EDE: THREE EXAMPLES

4.1 INTRODUCTION

The best way to learn how to use EDE is to do a directed analysis of data. Three datasets are included with this manual: CITYVILLE1.CSV, CITYVILLE1.MDB, and CityvilleAgg.xls. The first two datasets are identical and represent illness data for an imaginary town called Cityville. Only the format of the two files varies, so the user can try creating an EDE DSDC using both a text data file and an ACCESS database file. The third dataset is an aggregated version of the other datasets that contains a single record per day with the number of cases observed on that day. It is strongly suggested that the user work with the test data while going through the remainder of this manual. It is designed to take a user step by step through creating DSDC files, analyzing the data, and saving files and queries.

4.2 USE CASE NO. 1 – FEVER IN CITYVILLE USING CASE-LEVEL DATA

4.2.1 Creating an EDE Datasource

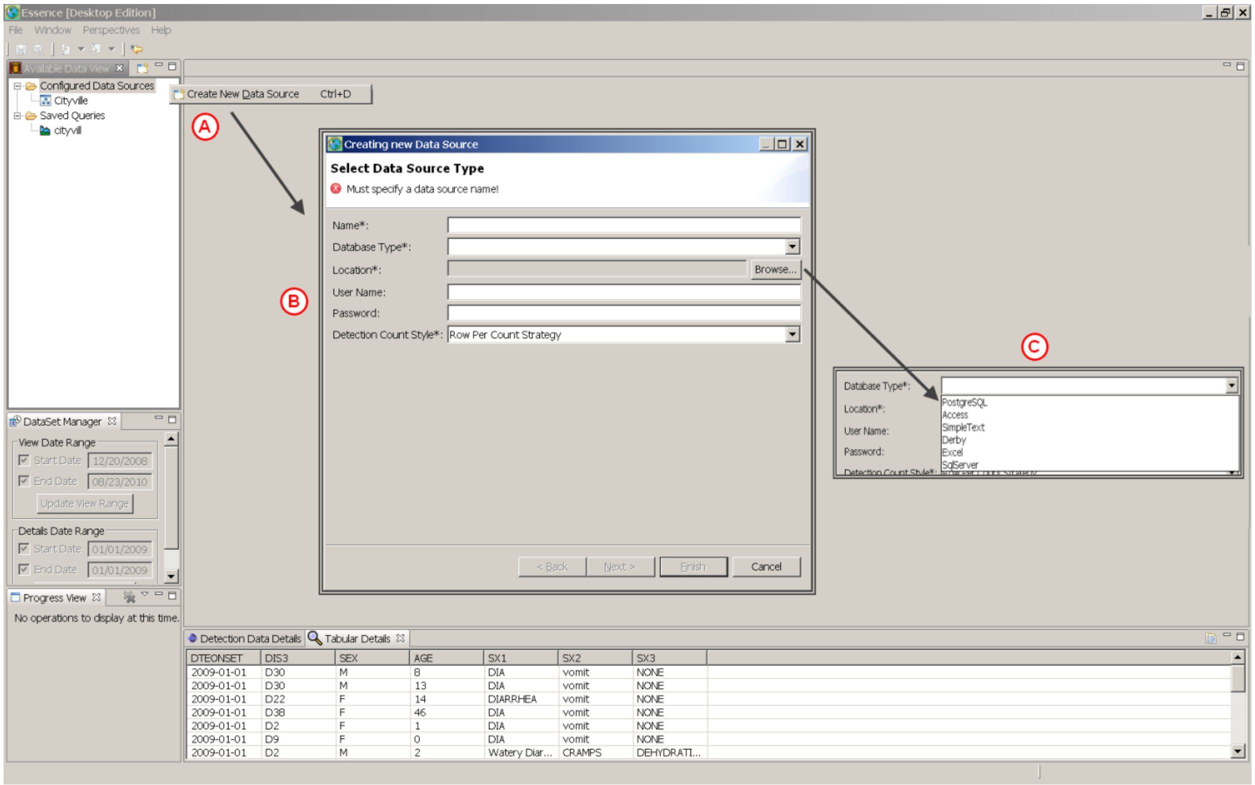
The following subsections describe how to create an EDE datasource, including how to define the name and type of data, set delimiters and column names, configure columns, save DSDC files, and edit DSDC files.

4.2.1.1 Creating an EDE Datasource – Datasource Type

The EDE default screen as it appears when the application is opened is shown in Figure 5. The **Available Data View**, **DataSet Manager**, **Progress**, **Detection Data Details**, and **Tabular Details** windows are all visible. The **Query** window (Figure 5, H) is also visible, but not yet labeled because the query has yet to be created.

Before creating a query, data must be read into EDE and configured for the application. To create an EDE DSDC file, right-click on **Configured Data Sources** in the **Available Data View**, then left-click on the pop-up window that appears. (Figure 7, A) Alternatively, click on the **New Data Source** icon on the EDE toolbar to open the datasource creation wizard.

Clicking on **Create a New Data Source** opens a pop-up window entitled **Creating New Data Source.** (Figure 7, B) This pop-up has six fields: <Name>, <Database Type>, <Location>, <User Name>, <Password>, and Password, which are defined below.



Creating new Data Source

Select Data Source Type

Must specify a data source name!

Name*:

Database Type*:

Location*:

User Name:

Password:

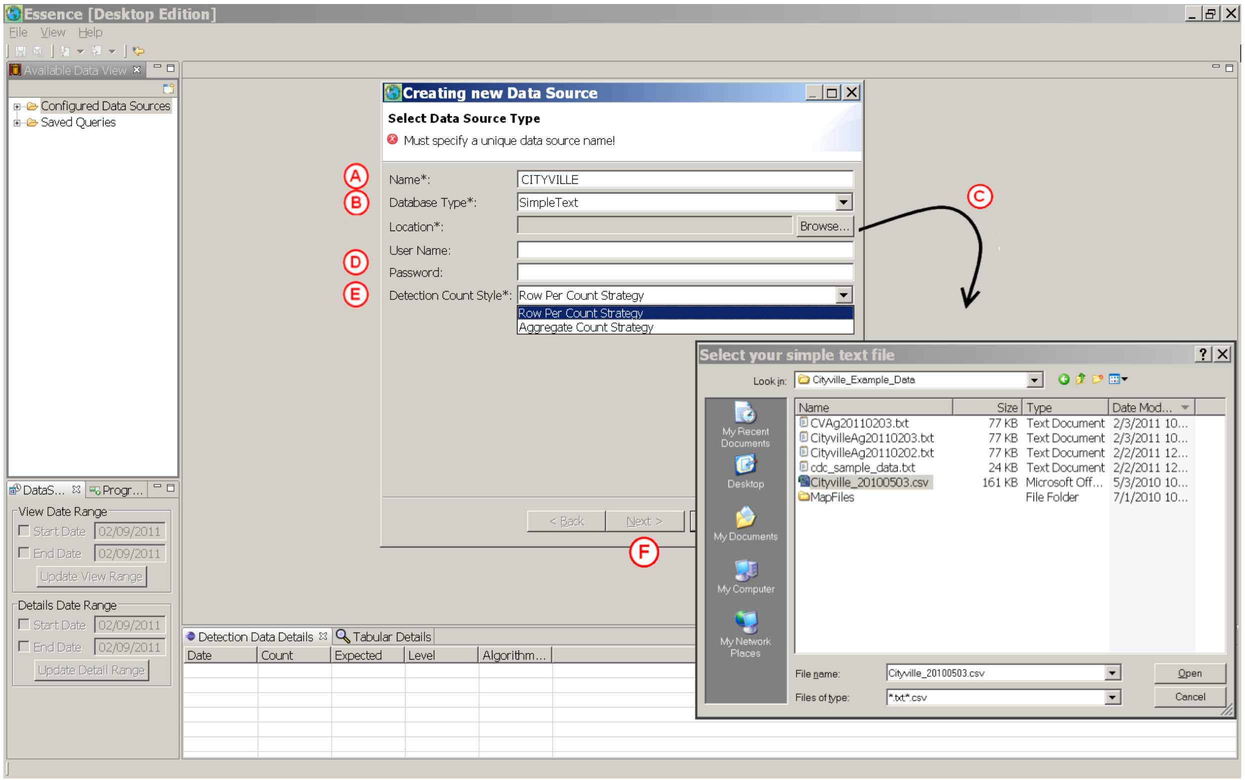
Detection Count Style*: Row Per Count Strategy

No operations to display at this time.

Database Type*:
PostgreSQL
Access
SimpleText
Derby
Excel
SqlServer

DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2009-01-01	D30	M	8	DIA	vomit	NONE
2009-01-01	D30	M	13	DIA	vomit	NONE
2009-01-01	D22	F	14	DIARRHEA	vomit	NONE
2009-01-01	D38	F	46	DIA	vomit	NONE
2009-01-01	D2	F	1	DIA	vomit	NONE
2009-01-01	D9	F	0	DIA	vomit	NONE
2009-01-01	D2	M	2	Watery Diar...	CRAMPS	DEHYDRATI...

Figure 7 Create New EDE Datasource – Selecting Database Type



Creating new Data Source

Select Data Source Type

Must specify a unique data source name!

Name*: CITYVILLE

Database Type*: SimpleText

Location*:

User Name:

Password:

Detection Count Style*: Row Per Count Strategy
Row Per Count Strategy
Aggregate Count Strategy

Select your simple text file

Look in: Cityville_Example_Data

Name	Size	Type	Date Mod...
CVAg20110203.txt	77 KB	Text Document	2/3/2011 10...
CityvilleAg20110203.txt	77 KB	Text Document	2/3/2011 10...
CityvilleAg20110202.txt	77 KB	Text Document	2/2/2011 12...
cdc_sample_data.txt	24 KB	Text Document	2/2/2011 12...
Cityville_20100503.csv	161 KB	Microsoft Off...	5/3/2010 10...
MapFiles		File Folder	7/1/2010 10...

File name: Cityville_20100503.csv

Files of type: *.txt;*.csv

Figure 8 Create New EDE DSDC – Location

1. **Name** – This refers to the name the user gives to the EDE DSDC file being created. (Figure 8, A) EDE does not accept spaces or dashes in the DSDC name, but underscores may be used. In the <Name> field, type in a name for the new DSDC. For the Cityville example, type in “Cityville.”
2. **Database Type** – This specifies the type of database being used. (Figure 7, C) EDE can currently read data from Simple Text files, Excel files, and Access, Derby, SqlServer and PostgreSQL database files. (Figure 7, C Insert) The Simple Text files should have the extension ‘.CSV’ or ‘.TXT’. Records in the text files should be delimited. A comma between fields is preferable, but any delimiter, may be used. The delimiter is specified during DSDC creation. In this Cityville example, a <Simple Text> file was selected. (Figure 8, C Insert)
3. **Location** – This refers to the location of the source text file or. Clicking the Browse Button opens a browser window to search the computer directory for the data file. For this exercise, locate the file called <CITYVILLE1.CSV> on your computer and double-click on the name, or click once on the name and click again on the <OPEN> button on the bottom of the pop-up screen. Figure 8 shows the selection of the file from the Browser window. (Figure 8, C Inset)
4. **User Name and Password** – These are needed only if an SQL database file is used. (Figure 8, D)
5. **Detection Count Style** – This refers to whether the data are case-level or aggregate data. In a case-level data file, each record contains information on a single visit/event for a single person, while in an aggregate data file, each record lists the number of people presenting on a given date with specific illness/event. If the data file contains records for individual patients/events, that is case-level data, select **Row per Count Strategy**. Alternatively, if the file contains records with a count of the number of patients/events seen per day, aggregate data, select **Aggregate Count Strategy**. (Figure 8, E)
6. **<Next>** – Once the Location and, if needed, the User Name and Password are selected, click **Next>** (Figure 8, F) to open the next pop-up window

4.2.1.2 Creating an EDE Datasource – Datasource Settings

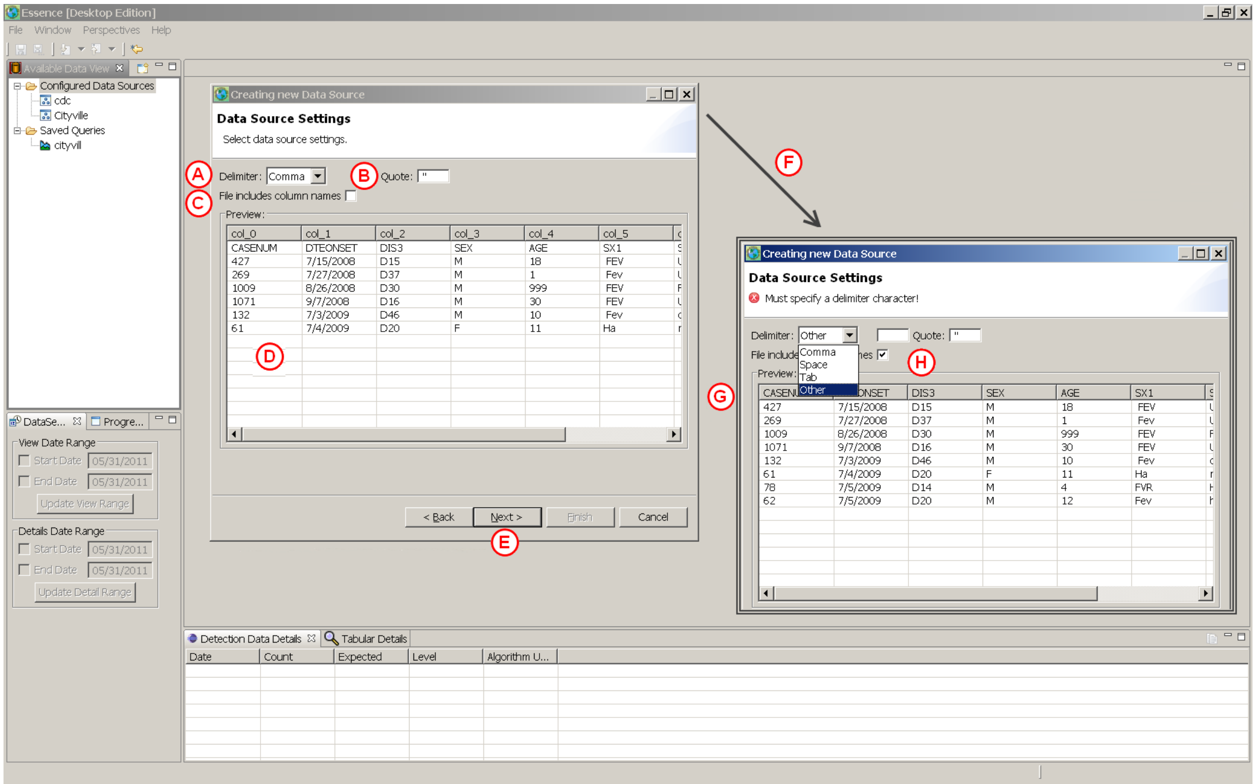
The next pop-up window is labeled **Data Source Settings**. It contains the data preview window at the bottom, and three fields to complete at the top: <Delimiter>, <Quotes>, and <File includes column names>. (Figure 9) Database files, such as Access, have the data source settings integrated into the file and EDE uses that information to automatically pre-fill these settings.

4.2.1.2.1 Delimiter Field

A delimiter is a character placed between variable values in a simple text file. An example of a delimiter is a comma. The default in EDE is a comma. To specify a different delimiting character, click the arrow on the right edge of the box next to delimiter to access a pull-down list of common delimiters. (Figure 9, F, Inset) Select the delimiter used in the current dataset from the pull-down list. If the delimiter used in the current file is not in the pull-down list, click on ***Other*** and type the new delimiter in the box that appears to the right of the delimiter box. (Figure 9, H)

4.2.1.2.2 Quotes

Refers to the text characters, such as quotes, used to enclose character (string) variables in the dataset. For example, single or double quotes are often used to enclose the value of character variables in a dataset. If the dataset uses a character to enclose the value of a string variable, type that character in the box next to **Quotes.** If the dataset has no such character, ignore the box. (Figure 9, B)



Creating new Data Source

Data Source Settings

Select data source settings.

Delimiter: Comma Quote: "

File includes column names □

Preview:

col_0	col_1	col_2	col_3	col_4	col_5
CASENUM	DTEONSET	DIS3	SEX	AGE	SX1
427	7/15/2008	D15	M	18	FEV
269	7/27/2008	D37	M	1	Fev
1009	8/26/2008	D30	M	999	FEV
1071	9/7/2008	D16	M	30	FEV
132	7/3/2009	D46	M	10	Fev
61	7/4/2009	D20	F	11	Ha

Creating new Data Source

Data Source Settings

Must specify a delimiter character!

Delimiter: Other Quote: "

File include[...]es ☑

Preview:

CASENU	ONSET	DIS3	SEX	AGE	SX1
427	7/15/2008	D15	M	18	FEV
269	7/27/2008	D37	M	1	Fev
1009	8/26/2008	D30	M	999	FEV
1071	9/7/2008	D16	M	30	FEV
132	7/3/2009	D46	M	10	Fev
61	7/4/2009	D20	F	11	Ha
78	7/5/2009	D14	M	4	FVR
62	7/5/2009	D20	M	12	Fev

Figure 9 Create New EDE DSDC – Select Data Source Settings

4.2.1.2.3 File Includes Column Names

Just under the delimiter pull-down box is a check box for **File Includes Column Names** which refers to whether the first row of data in the file contains variable names. This is applicable for text & Excel files; variable names are already incorporated into database files. If the first line of the text/Excel file does contain variable names, check the box by clicking on it. (Figure 9, C) A sample of the data can be examined in the **Data Preview Window** to verify whether there are variable names in the file. If there are then the names will appear as the first row of data in the preview window. (Figure 9, D) After clicking on the check box (Figure 9, C), the variable names contained in the file will replace the generic variable names generated by EDE. (Figure 9, G) If the file does not contain variable names leave the box unchecked The Cityville data set does contain variable names, so click on the box.

4.2.1.2.4 Data Preview Window

This refers to the box at the bottom of the **Data Source Settings** window. (Figure 9, D) The first few lines of data in the database appear in this window as the datasource settings are added. Review the data in this window to make sure the variable values appear correctly. Then press the **<Next>** button at the bottom of the screen (Figure 9, E) to open the **Configure Columns** window.

4.2.1.3 Creating an EDE Datasource – Configure Columns

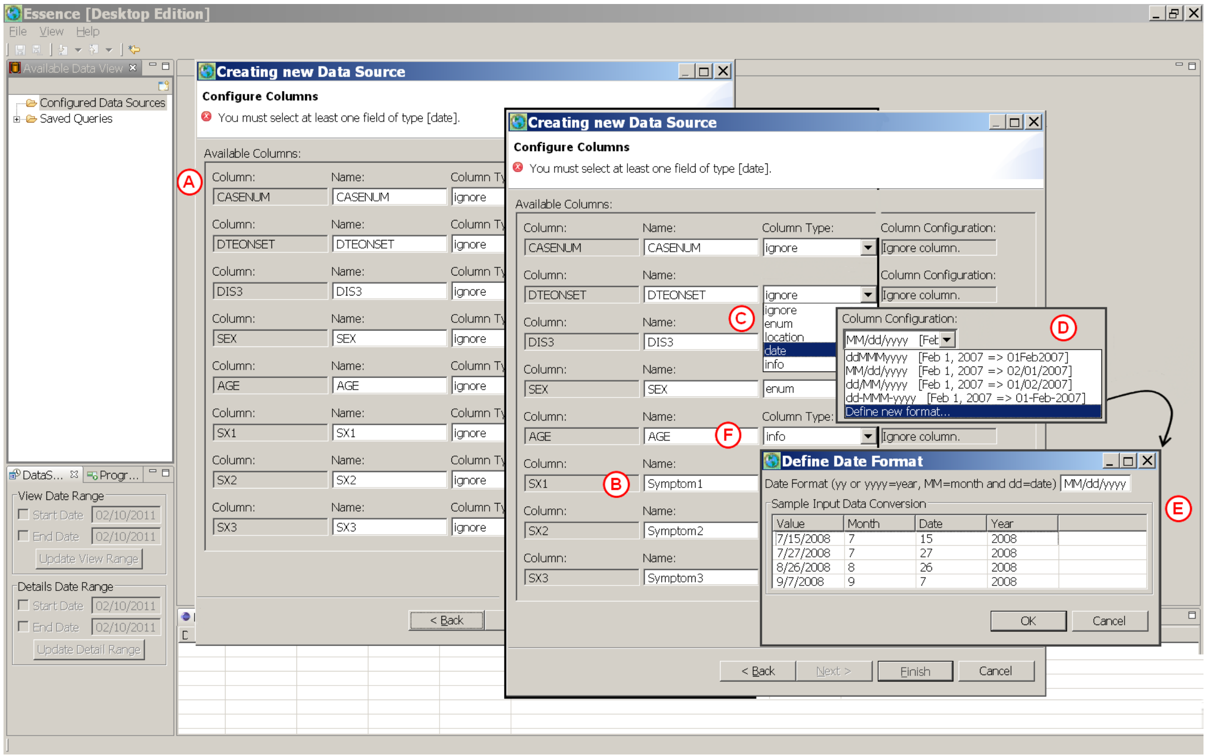
The **Configure Columns** window opens small, so it is useful to stretch this box to the left so that all of the columns are visible. (Figure 10) The fields in this pop-up are in four columns with the headings **Column**, **Name**, **Column Type**, and **Column Configuration**.

4.2.1.3.1 Column

Column refers to the variable names in the original data file. If the data set being used lists the variable names in the first row of data or if a database is used these names will appear in the **Column** fields. (Figure 10, A) If the dataset does not have variable names, the variables will be named sequentially. In this Cityville example, the variable names were included in the data set and appear in the fields under **Column**, because the **File includes column name** box was checked on the last screen. These values cannot be changed. They are internal to EDE.

4.2.1.3.2 Name

Name refers to the name given to the variable in the EDE DSDC file. By default, these fields will be the same as the fields under **Column** (Figure 10, A), but they can be changed. To change the field value, click on the box, erase the current name, and type in the name you prefer. In Figure 10, the three symptom variables, SX1, SX2, SX3, have been given new names in the datasource file, Symptom1, Symptom2, and Symptom3 respectively. (Figure 10, B)



View Date Range
□ Start Date 02/10/2011
□ End Date 02/10/2011

Details Date Range
□ Start Date 02/10/2011
□ End Date 02/10/2011

Configure Columns

You must select at least one field of type [date].

Available Columns:

Column:	Name:	Column Type:	Column Configuration:
CASENUM	CASENUM	ignore	Ignore column.
DTEONSET	DTEONSET	ignore ignore enum location date info	Ignore column.
DIS3	DIS3		MM/dd/yyyy [Feb
SEX	SEX	enum	
AGE	AGE	info	Ignore column.
SX1	Symptom1		
SX2	Symptom2		
SX3	Symptom3		

Column Configuration:

MM/dd/yyyy [Feb
ddMMMyyyy [Feb 1, 2007 => 01Feb2007]
MM/dd/yyyy [Feb 1, 2007 => 02/01/2007]
dd/MM/yyyy [Feb 1, 2007 => 01/02/2007]
dd-MMM-yyyy [Feb 1, 2007 => 01-Feb-2007]
Define new format...

Date Format (yy or yyyy=year, MM=month and dd=date) MM/dd/yyyy

Sample Input Data Conversion

Value	Month	Date	Year
7/15/2008	7	15	2008
7/27/2008	7	27	2008
8/26/2008	8	26	2008
9/7/2008	9	7	2008

Figure 10 Create New EDE DSDC – Configuring Columns

4.2.1.3.3 Column Type

Column Type does not refer to variable type, i.e., whether the variable is numeric or character (string) data. Both string and numeric data can be analyzed in EDE. **Column Type** refers instead to how the variable will be used in EDE. **Column Type** is selected from the pull-down menu that appears when you click on a box. (Figure 10, C) The five choices are: ignore, enum, location, date, and info.

Ignore – This is the default value for **Column Type**. Assign this value if the variable should *not* appear in the EDE DSDC. In this Cityville example, the variable CASENUM is listed as ignored. (Figure 10, C)

Enum – Refers to variables for which each unique variable value will be counted (enumerated) and/or used to categorize the dataset for analysis, graphing, and mapping. Use this for non-date and non-location variables that will be used to subdivide data during analyses and/or graphing. In this example, “Symptom1,” the patient’s first listed symptom, is specified as “enum” to create a time series showing the number of cases with a specific symptom reported by the date of onset. (Figure 10, C) Note again, enum does not refer to whether the variable is numeric or character/string. EDE can analyze either type of data because it enumerates, or counts, the number of times each variable value occurs in the data for each enum variable.

Location – Refers to variables that will be used to map the data in EDE. Using UDig or EpiMap, EDE can map counts or alert levels for any geographic location for which shape files are available. In this example, DIS3 is specified as “location.” DIS3 refers to the 47 districts in Cityville.

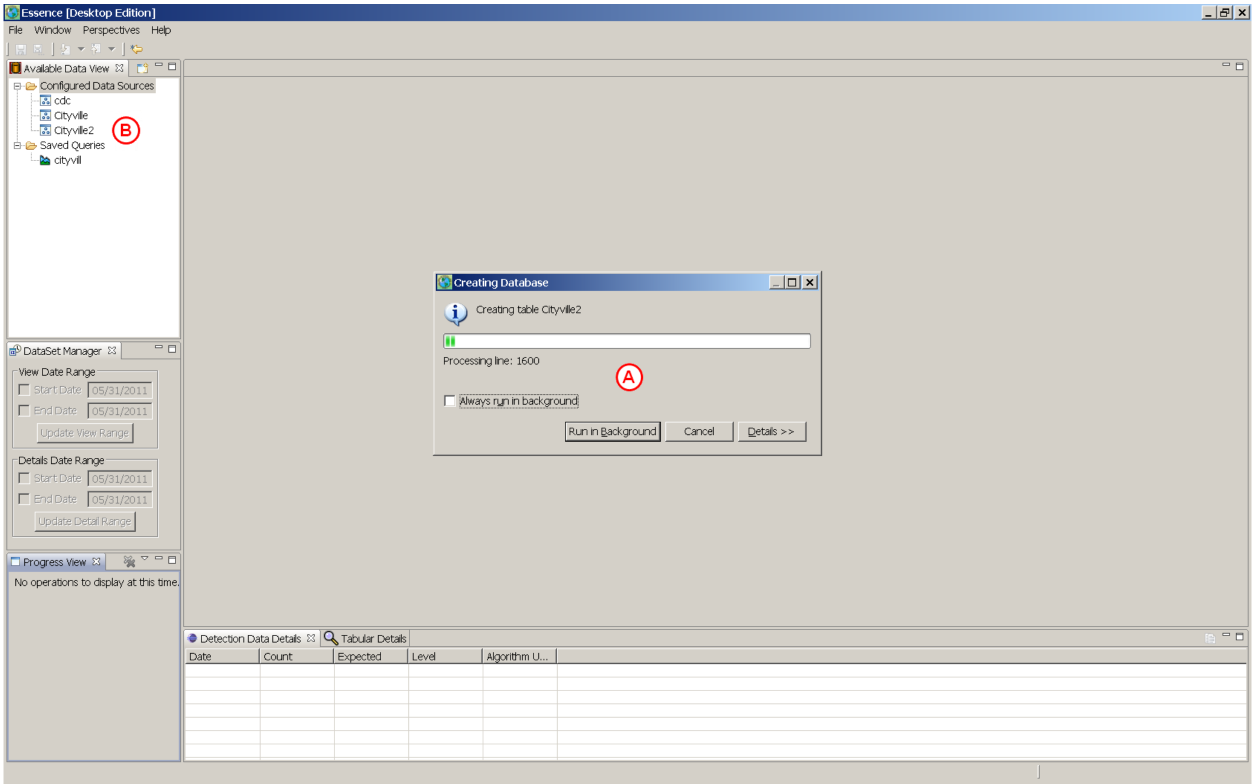
Date – Refers to date variables. This column type should be assigned when the variable is a date that will be used in an analysis. The date format used in the dataset is selected from the pull-down menu under **Column Configuration.** (Figure 10, D) If the date format needed is not available click in the pull-down window, click on **<Define new format>** and type the new format in the **Define Date Format** window. (Figure 10, E) Then click **<OK>** and continue configuring columns

Info – Refers to variables that should appear in line listings and data details, but will not be used in queries or to graph data. In t**he** Cityville example, age is listed as **Info** because it is a continuous variable with many values and is not appropriate for graphing or sub-grouping during analysis. It is useful when examining alerts, however, and by designating it **Info** it will appear in line listings. (Figure 10, F)

4.2.1.3.4 **Column Configuration** – Is used only for **Column Type** equal **Date**, so that the date format can be specified. For other data types the column will self-fill and should be ignored.

4.2.1.4 Creating an EDE Datasource – Save DSDC

When all of the variables on the **Configure Columns** are complete, click the **Finish** button. The datasource pop-up window will disappear and a timer box entitled **Creating Database** (Figure 7, A) appears and remains until the DSDC has been completed. If there is an error during DSDC creation, a message will appear in a pop-up box.



View Date Range: □ Start Date 05/31/2011; □ End Date 05/31/2011

Details Date Range: □ Start Date 05/31/2011; □ End Date 05/31/2011

No operations to display at this time.

Creating table Cityville2

Processing line: 1600

□ Always run in background

Figure 11 Create a New EDE DSDC – Save DSDC

Once a DSDC has been created, it will be listed under **Configured Data Sources** in the **Available Data View** window to the left side of the screen. The datasource "CITYVILLE" appears in this spot once it is complete. (Figure 7, B)

4.2.1.5 Creating an EDE Datasource – Editing a DSDC

Saved EDE DSDC files can also be edited. (Section 2.4.1) To edit a file, right-click on the name of the EDE DSDC file. A pop-up window appears offering four choices: New Query, Edit Datasource, and Delete Datasource and Reload DataSource ENUM Fields. Click **Edit** and the **Creating DataSource** Wizard, now labeled **Editing (*DSDC name*),** appears showing the values selected when the DSDC file was created. These values can be modified as needed. The name of the saved datasource, the location of the data file, the database table, or configuration specifications can be changed, and the file then resaved.

The other choices allow the user to create a new query, delete the datasource, or update a DSDC that draws data from a simple text or a Microsoft Office Excel file. The last two options, **Manage Data**, and **Runnables**, are artifacts of the ECLIPSE programming language used to create EDE and should be ignored by users.

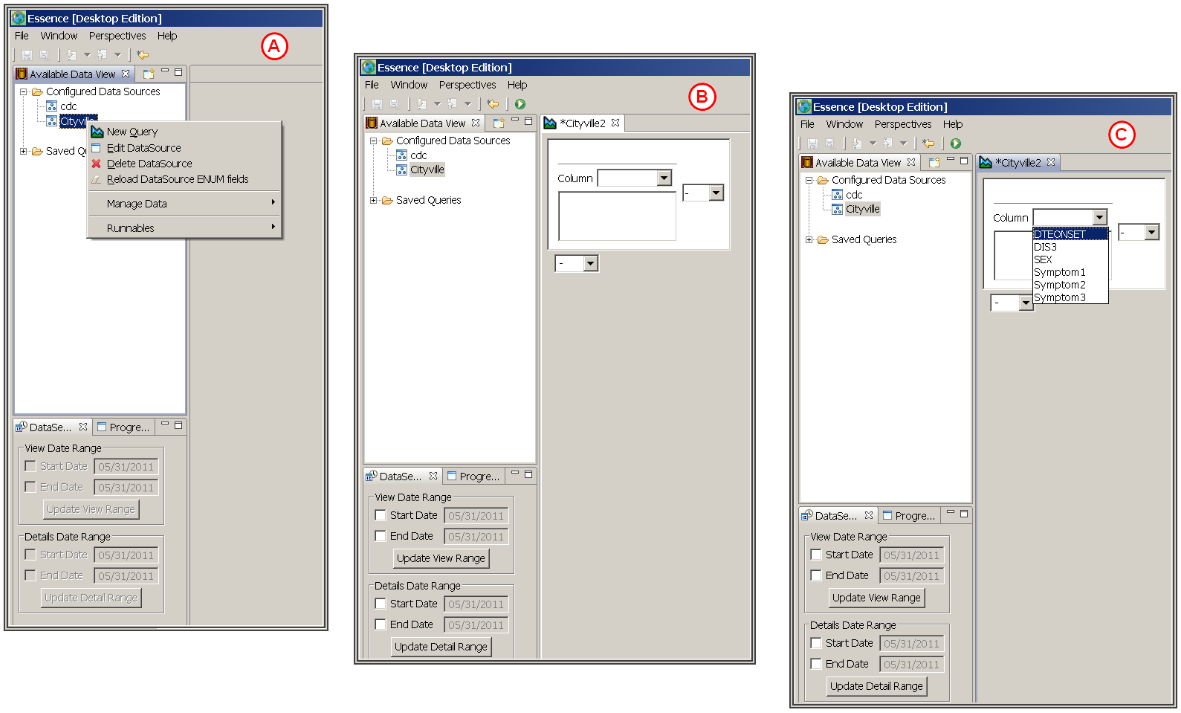


Figure 12 Creating a New Query, the First Variable

4.2.2 Creating a Query

Once the EDE datasource is created, it can be queried. Creating a query is the first step in examining the data in an EDE DSDC. A query is a logical statement used to create a subset of data from a specified data set. It defines the data you want to examine. For example, for a time series of fever cases in Cityville, two variables need to be defined: time of onset of fever, and records where Symptom1 equals fever. In English, the query would be something like this: “From all of the records in the Cityville dataset give me those that occurred between *start date* and *end date*, and also have Symptom1 equal to any phrase that might mean fever.” To create this with the Query Wizard in EDE, specify each of the variables you want to include in the statement, (i.e., date and symptom) and how they should be joined (“And.”)

To create the first query in this Cityville exercise, right-click on the name of the saved datasource (i.e., CITYVILLE) under **Configured Data Sources** in the **Available Data View** window. A pop-up window will appear offering four valid choices: create a **New Query**, **Edit Datasource**, **Delete Datasource**, and **Reload Datasource ENUM fields. (**Figure 8, A) Click on the first choice, **New Query**. A tabbed box will appear in the query window. It will be labeled at the top with the name of the EDE datasource being used. In this example, the tab is labeled CITYVILLE because that is the datasource being used. (Figure 8, B) A variable definition box (labeled Column) will also appear in the query window.

To select the first variable in the query statement, click on the pull-down arrow in the variable definition box to the right of **Column**. The pull-down list that appears contains the variables requested as date, enum, location or info when the datasource was created. Select the variable named **<DTEONSET>** by clicking on the name. (Figure 8, C)

Whenever a date field is opened in EDE, a Date Selection box opens. (Figure 9, A) This contains space to enter start and end dates. By default, EDE shows the current date for both the **<Start>** and **<End>** dates. If, however, these dates are not altered, when the query is run EDE will automatically select the earliest and latest dates in the DSDC as the **<Start>** and **<End>** dates. If different dates are preferred, click on the checkbox(s) and type the preferred **<Start>** and/or **<End>** dates. In this CITYVILLE example, EDE will automatically select the start and end dates. (Figure 9, A)

To select how the first and second variables are joined, click on the pull-down arrow in the small box to the right of the variable definition box. (Figure 9, B) Three choices appear in the pull-down, **<And>**, <**Or>**, and **<—>** or null which means no connecter is selected. Either **<And>** or **<Or>** must be selected when more than one variable is used in the query. In this Cityville example, click **<And>** to join the date and symptom variables.

Note

> There is also a pull-down box below the first variable. (Figure 9, C) This allows construction of a complex query statement that includes brackets. Selecting the box to the right allows you to continue with the logical statement started with the first box. Selecting the pull-down below the variable box begins creating a second phrase in the logical statement that will be placed within parentheses and joined, by an **<And>** or an **<Or>** to the statement created above it (also placed within brackets).

Selecting a join value (**<And>** or **<Or>**) opens another variable definition box beside (or below) the first one. (Figure 9, D) Select the second variable by clicking the arrow in the box next to **Column** and highlighting the desired variable name in the pull-down list. Select variable **<Symptom1>** for the Cityville example.

Once Symptom1 is selected from the pull-down list, EDE creates a list of all of the values of that variable and displays them in a pick list below the variable name. (Figure 9, E) Select a single value by clicking on it. Select multiple values by CRTL-clicking (press the CTRL key and click simultaneously) on additional variable values in the pick list. The ability to select multiple values is useful. Perhaps two diagnoses are equally likely in patients with a particular illness; both diagnoses can be selected for analysis. Having the option to select multiple values is useful for another reason in this Cityville example. Scroll all the way through the values for Symptom1 in the Cityville dataset. Notice that there are many variations of the word ”fever” in the value list. To complete the query successfully, all of the fever values must be included in the query. To do so, CTRL-click on all of the permutations of fever in the value pick list. There are 46 different values, 18 of these are for fever. Four of the 12 fever values are shown in Figure 9, E.

There is also a search feature that can aid identification of similar or specific values of a variable. Above the list of variable values there is a **<Search>** check box and text box. (Figure 9, F) By default, the **<Search>** box is checked, and the value **<%>** appears in the text box. This **<%>** is the search ‘wild card’, and indicates that all values of the variable should be displayed in the list. To identify all values beginning with an **‘f’**, the phrase **<%f%>** would be typed in the text box. This yields a list of values that have an ‘f’ (either lower or upper case) anywhere in their length. In this Cityville example this identifies all fever values, because some of the fever values do not begin with ‘f’ but all have ‘f’ somewhere in their length. Note that if other values for this variable contained an ‘f’ somewhere in their length, ex. **<fainted>**, they too would be included in the value list called by the phrase **<%f%>**.

The complete query needed for the current example is shown in Figure 9, inset E. To run the query, click on the **<Run Query>** button at the bottom of the Query window. (Figure 9, G) EDE will run the detection algorithm on the selected data and display the results in the tabbed boxes in the Query window. The **<Time Series>** tab will open automatically when the query has run.

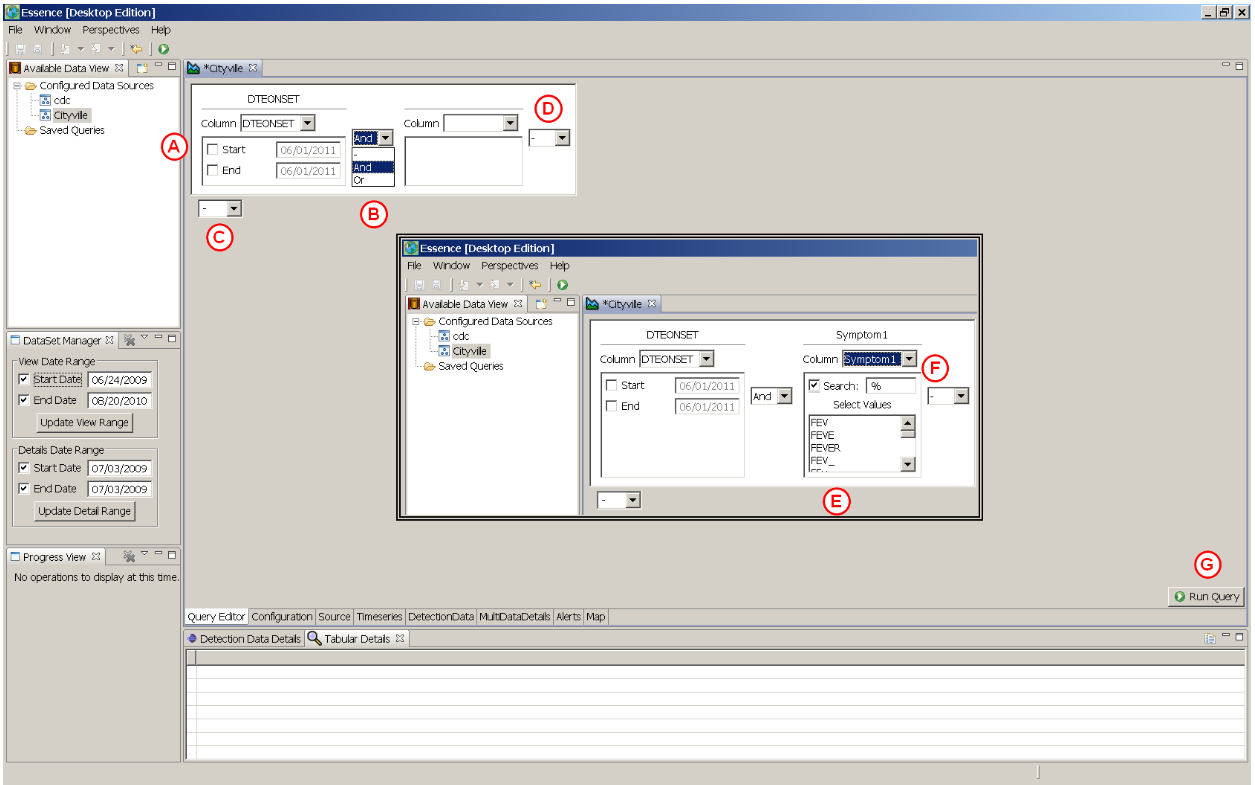
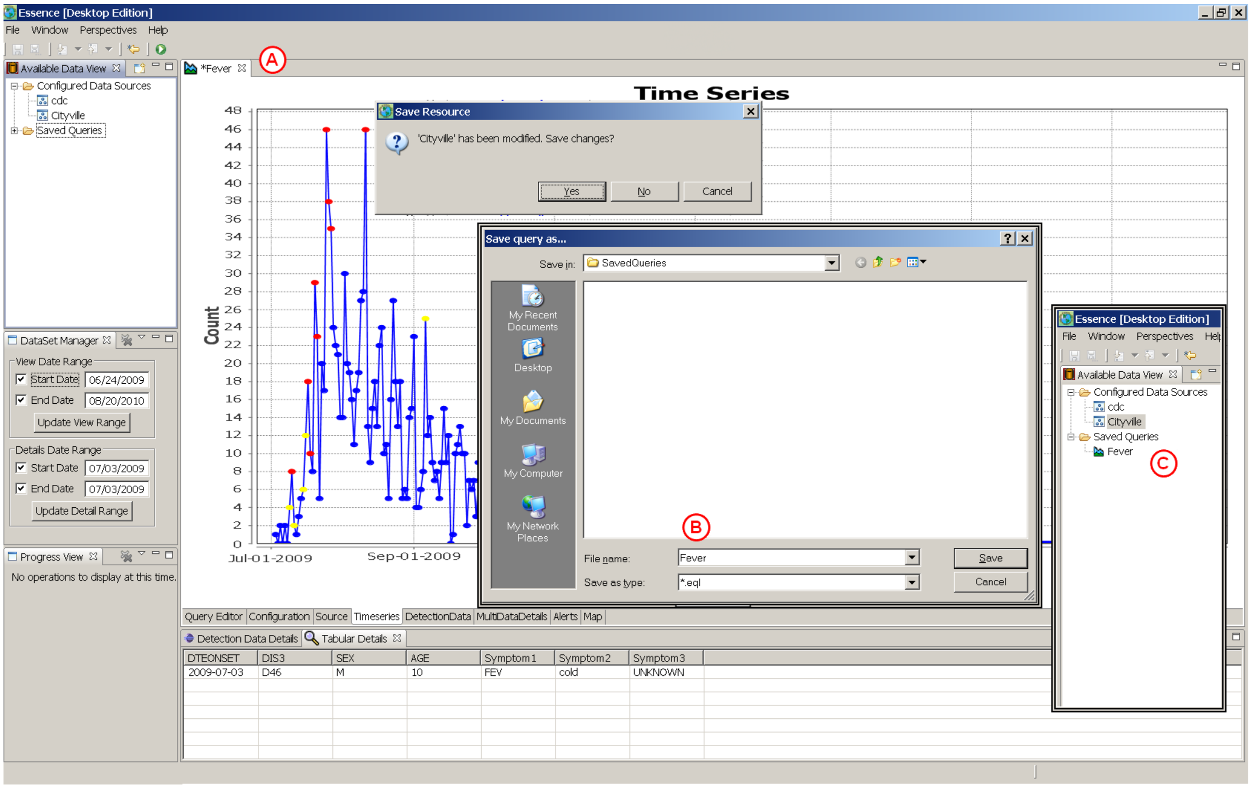


Figure 13 Creating a New Query, the Second Variable

4.2.3 Saving a Query

Queries may be saved for later use. When the query is closed by left-clicking on the **X** at the top right of the query tab (Figure 14, A) the **<Save Resource>** window opens with the question, **<'Cityville' has been modified. Save changes?>**. To save the query click **<Yes>**, to discard it click **<No>**, or to return to the query click **<Cancel>**. Left-clicking on **<Yes>** opens a **<Save Query As>** window appears. Type the query name in the **File Name** box (Figure 14, B), and click **Save**. In this Cityville example the query was saved as **<Fever>**. Once the query is saved, the query name will appear under **Saved Queries** in the **Available Data View** window. (Figure 14, C)



View Date Range

☑ Start Date 06/24/2009

☑ End Date 08/20/2010

Details Date Range

☑ Start Date 07/03/2009

☑ End Date 07/03/2009

No operations to display at this time.

'Cityville' has been modified. Save changes?

Save in: SavedQueries

File name: Fever

Save as type: *.eql

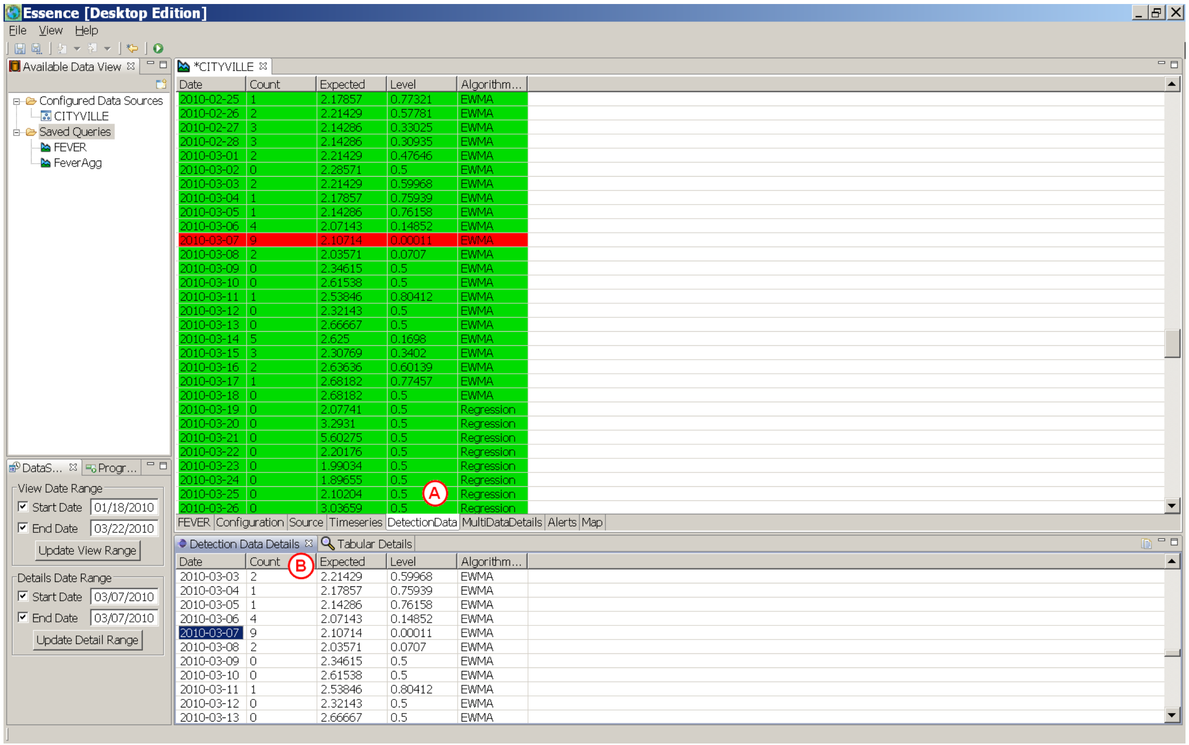
DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2009-07-03	D46	M	10	FEV	cold	UNKNOWN

Figure 14 Saving a Query

4.2.4 Time Series Tab

Once a query is run, the **Time series** tab is displayed automatically in the **Query** window in place of the **Query Editor** tab. This tabbed window opens to a time series of the query data. (Figure 15) Alerts, or days when the detection algorithm indicates that the record count for that day is significantly greater than expected based on recent data, are marked on the time series as red and yellow dots that correspond to specific p-values. The user can set the values of red and yellow alerts. The default values are p-value equals 0.01 for red and 0.05 for yellow. When the cursor is placed on the line, a pop-up window shows the p-value, the count, and the date, of the point on which the cursor sits. (Figure 15, B)

It is also possible to zoom in on a specific section of the time series by dragging a box open over the selected area. (Figure 15, C) This causes the selected section to be expanded in the **Time Series** tab. (Figure 15, Inset D)



View Date Range

☑ Start Date 01/18/2010

☑ End Date 03/22/2010

Details Date Range

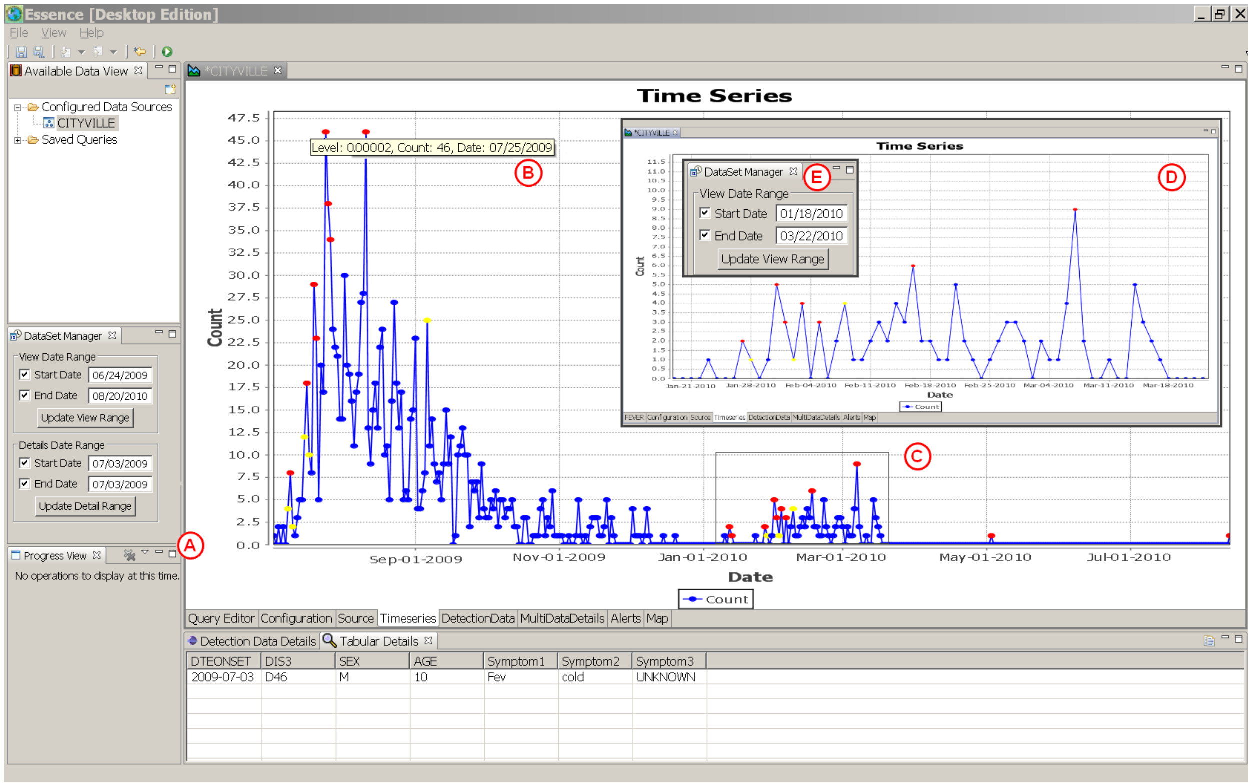
☑ Start Date 03/07/2010

☑ End Date 03/07/2010

Date	Count	Expected	Level	Algorithm...
2010-02-25	1	2.17857	0.77321	EWMA
2010-02-26	2	2.21429	0.57781	EWMA
2010-02-27	3	2.14286	0.33025	EWMA
2010-02-28	3	2.14286	0.30935	EWMA
2010-03-01	2	2.21429	0.47646	EWMA
2010-03-02	0	2.28571	0.5	EWMA
2010-03-03	2	2.21429	0.59968	EWMA
2010-03-04	1	2.17857	0.75939	EWMA
2010-03-05	1	2.14286	0.76158	EWMA
2010-03-06	4	2.07143	0.14852	EWMA
2010-03-07	9	2.10714	0.00011	EWMA
2010-03-08	2	2.03571	0.0707	EWMA
2010-03-09	0	2.34615	0.5	EWMA
2010-03-10	0	2.61538	0.5	EWMA
2010-03-11	1	2.53846	0.80412	EWMA
2010-03-12	0	2.32143	0.5	EWMA
2010-03-13	0	2.66667	0.5	EWMA
2010-03-14	5	2.625	0.1698	EWMA
2010-03-15	3	2.30769	0.3402	EWMA
2010-03-16	2	2.63636	0.60139	EWMA
2010-03-17	1	2.68182	0.77457	EWMA
2010-03-18	0	2.68182	0.5	EWMA
2010-03-19	0	2.07741	0.5	Regression
2010-03-20	0	3.2931	0.5	Regression
2010-03-21	0	5.60275	0.5	Regression
2010-03-22	0	2.20176	0.5	Regression
2010-03-23	0	1.99034	0.5	Regression
2010-03-24	0	1.89655	0.5	Regression
2010-03-25	0	2.10204	0.5	Regression
2010-03-26	0	3.03659	0.5	Regression

Date	Count	Expected	Level	Algorithm...
2010-03-03	2	2.21429	0.59968	EWMA
2010-03-04	1	2.17857	0.75939	EWMA
2010-03-05	1	2.14286	0.76158	EWMA
2010-03-06	4	2.07143	0.14852	EWMA
2010-03-07	9	2.10714	0.00011	EWMA
2010-03-08	2	2.03571	0.0707	EWMA
2010-03-09	0	2.34615	0.5	EWMA
2010-03-10	0	2.61538	0.5	EWMA
2010-03-11	1	2.53846	0.80412	EWMA
2010-03-12	0	2.32143	0.5	EWMA
2010-03-13	0	2.66667	0.5	EWMA

Figure 15 Timeseries Tab Display



Level: 0.00002, Count: 46, Date: 07/25/2009

View Date Range

☑ Start Date 06/24/2009

☑ End Date 08/20/2010

Details Date Range

☑ Start Date 07/03/2009

☑ End Date 07/03/2009

Inset: View Date Range

☑ Start Date 01/18/2010

☑ End Date 03/22/2010

No operations to display at this time.

DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2009-07-03	D46	M	10	Fev	cold	UNKNOWN

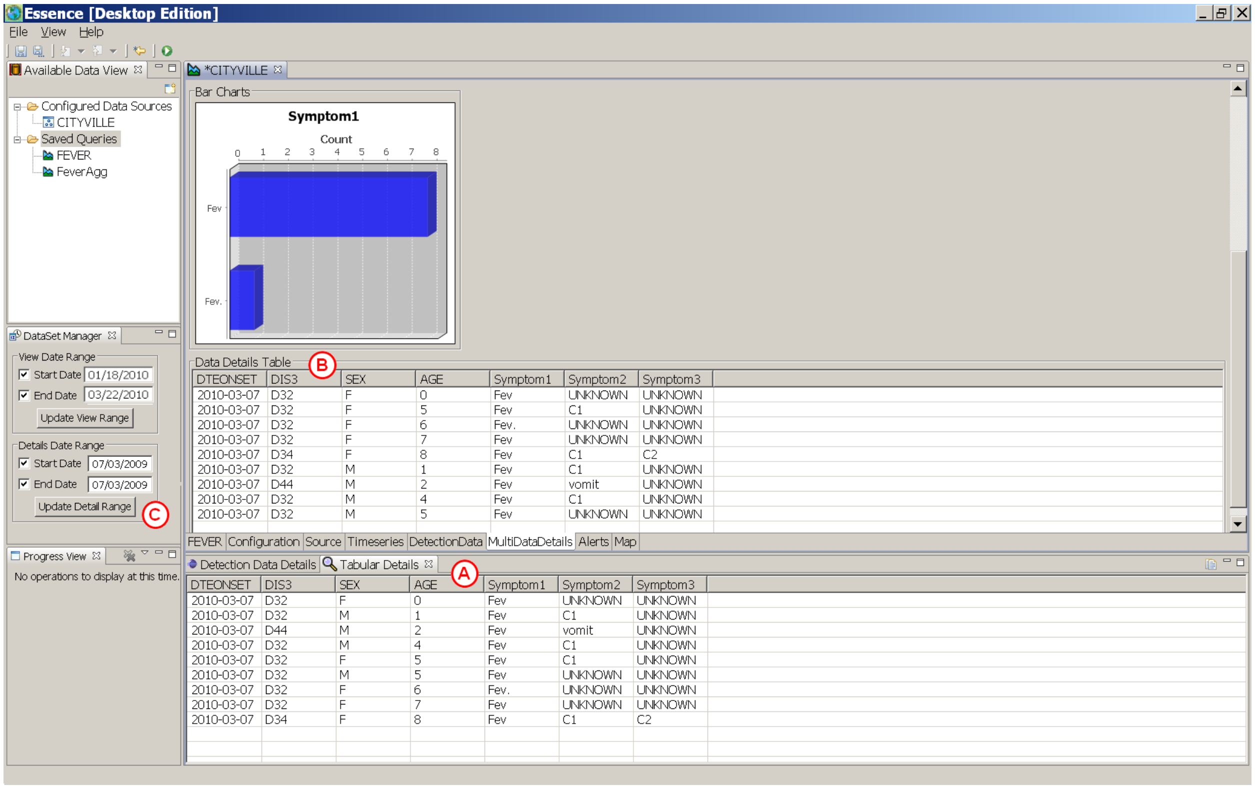
Figure 16 Expanded Time Series Showing Detection Data Details

4.2.5 Detection Data Window and Tab

The alerts found in the time series, and data useful in assessing the validity of the alerts are listed in the **DetectionData** tab of the **Query** window (Figure 16, A), and in the window at the bottom of the screen marked **DetectionDataDetails**. (Figure 16, B) The date, record count, expected record count, p-value, and the algorithm used are listed in these windows. Many people like to view the details of the detection and the time series together. Therefore, the default EDE main screen always has a window available that displays the **DetectionDataDetails** for the active **Query** window. (Figure 16, B) The advantage of the **DetectionData** tab in the **Query** window is that the red and yellow alerts are colored and easier to identify. Note that two or more query windows may be open at the same time. (see Use Case No. 3, Section 3.4) The active **Query** window is the one with the tab at the top of the window highlighted in blue. The other windows open on the screen (**Detection Data Details**, **Tabular Details**, etc.) also display results from the active query.

4.2.6 Data Details Window and Tab

The demographic details of the records included in an alert are useful in assessing its validity and importance. All variables in the EDE dataset, except those categorized as 'ignore', are listed, case by case, in the **Tabular Details** window (Figure 17, A) and in the **MultiDataDetails** tab. (Figure 17, B) The demographic details in either list can be sorted on any single column (variable) in the list by clicking on the variable name at the top of the column. As with the **DetectionData**, many people like to view the time series



View Date Range
☑ Start Date 01/18/2010
☑ End Date 03/22/2010

Details Date Range
☑ Start Date 07/03/2009
☑ End Date 07/03/2009

No operations to display at this time.

Data Details Table

DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2010-03-07	D32	F	0	Fev	UNKNOWN	UNKNOWN
2010-03-07	D32	F	5	Fev	C1	UNKNOWN
2010-03-07	D32	F	6	Fev.	UNKNOWN	UNKNOWN
2010-03-07	D32	F	7	Fev	UNKNOWN	UNKNOWN
2010-03-07	D34	F	8	Fev	C1	C2
2010-03-07	D32	M	1	Fev	C1	UNKNOWN
2010-03-07	D44	M	2	Fev	vomit	UNKNOWN
2010-03-07	D32	M	4	Fev	C1	UNKNOWN
2010-03-07	D32	M	5	Fev	UNKNOWN	UNKNOWN

Tabular Details

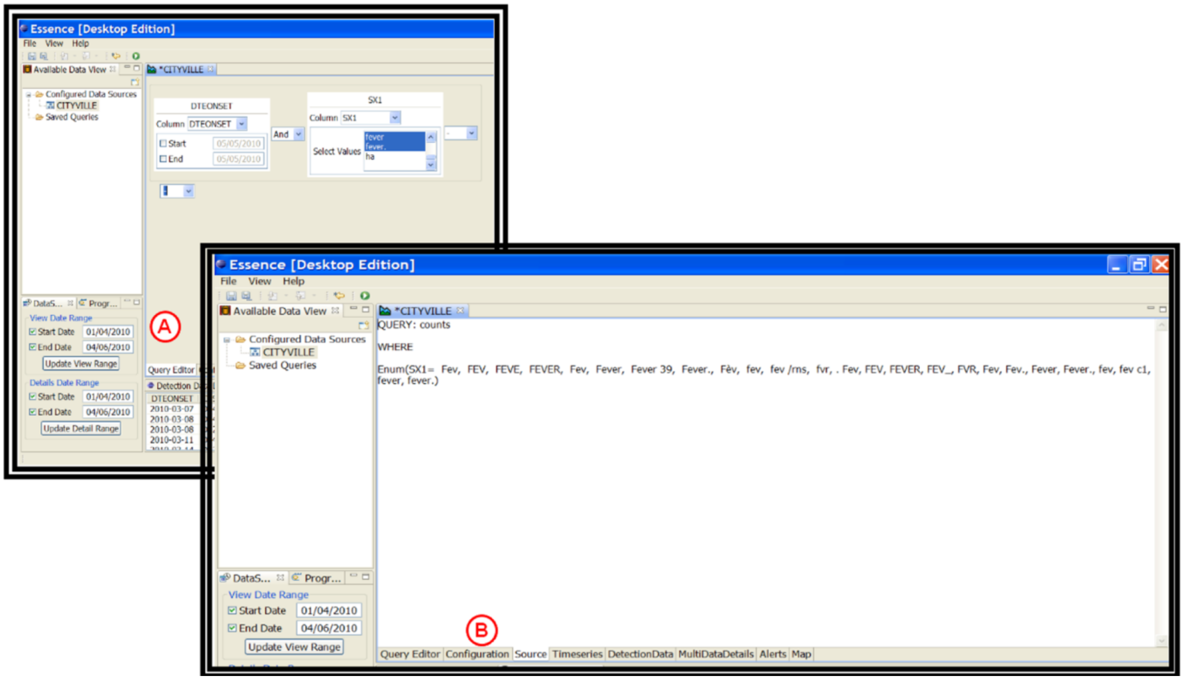
DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2010-03-07	D32	F	0	Fev	UNKNOWN	UNKNOWN
2010-03-07	D32	M	1	Fev	C1	UNKNOWN
2010-03-07	D44	M	2	Fev	vomit	UNKNOWN
2010-03-07	D32	M	4	Fev	C1	UNKNOWN
2010-03-07	D32	F	5	Fev	C1	UNKNOWN
2010-03-07	D32	M	5	Fev	UNKNOWN	UNKNOWN
2010-03-07	D32	F	6	Fev.	UNKNOWN	UNKNOWN
2010-03-07	D32	F	7	Fev	UNKNOWN	UNKNOWN
2010-03-07	D34	F	8	Fev	C1	C2

Figure 17 Data Details Window & Tab

and the demographic data details together. So, the EDE default main screen for the active query has a window open for the time series and another at the bottom for **Tabular Details** The time period displayed in the **Tabular Details** window and the **MultiDataDetails** tab corresponds to the date range selected under **Details Date Range** in the **DataSet Manager** window. (Figure 17, C) Changing the dates in this window will change the time period represented in the different data details tables, as well as in all other non-time series windows/tabs.

4.2.7 Query Editor Tab

The **Query Editor** tab displays the query wizard used to create the current query. (Figure 18, A) Remember this tab shows the query name after a query is saved. This tab can also be used to edit a query. It is often convenient to create a new query by modifying an existing one. The old query is opened by double-clicking on the name under **Saved Queries** in the **Available Data View** window. Then the query is edited, run, and saved as a new query with a different name.



Column DTEONSET

□ Start 05/05/2010

□ End 05/05/2010

Column SX1

☑ Start Date 01/04/2010

☑ End Date 04/06/2010

DTEONSET
2010-03-07
2010-03-08
2010-03-08
2010-03-11

```
QUERY: counts

WHERE

Enum(SX1= Fev, FEV, FEVE, FEVER, Fev, Fever, Fever 39, Fever., Fèv, fev, fev /rns, fvr, . Fev, FEV, FEVER, FEV_, FVR, Fev, Fev., Fever, Fever., fev, fev c1,
fever, fever.)
```

Figure 18 Query Editor and Source Tab Displays

4.2.8 The Source Tab

The **Source** tab (Figure 18, B) displays the pseudo-SQL command used to create the query specified in the **Query Editor** window. The pseudo-SQL statement is for reference only for users familiar with SQL. The statement cannot be edited or used to run a query.

4.2.9 The Configuration Tab

4.2.9.1 General Configuration

There are three options under General Configuration (Figure 19, A) at the top of the Configuration Tab. The first, **<Group data by>**, at this time only allows the user to analyze the data day by day. Future versions of EDE will also allow analysis of data grouped into CDC weeks.

The second option, **<Show Timeseries Date Marks (Blue)>** controls whether a blue dot is placed on the time series graph for dates without alerts. When checked, dots are placed on the time series graph. When the box is not checked, dots are not placed on the graph. Note that the blue dots are so close together on many time series that they are not visible and the time series line simply looks bolder when the **<Show Timeseries Date Marks (Blue)>** box is checked. The blue dots are visible when the time series is magnified. (Figure 16, D)

The third option controls whether the y-axis on charts and graphs is shown as a whole or real number. Checking the box forces the values to be whole numbers.

4.2.9.2 Detection Configuration

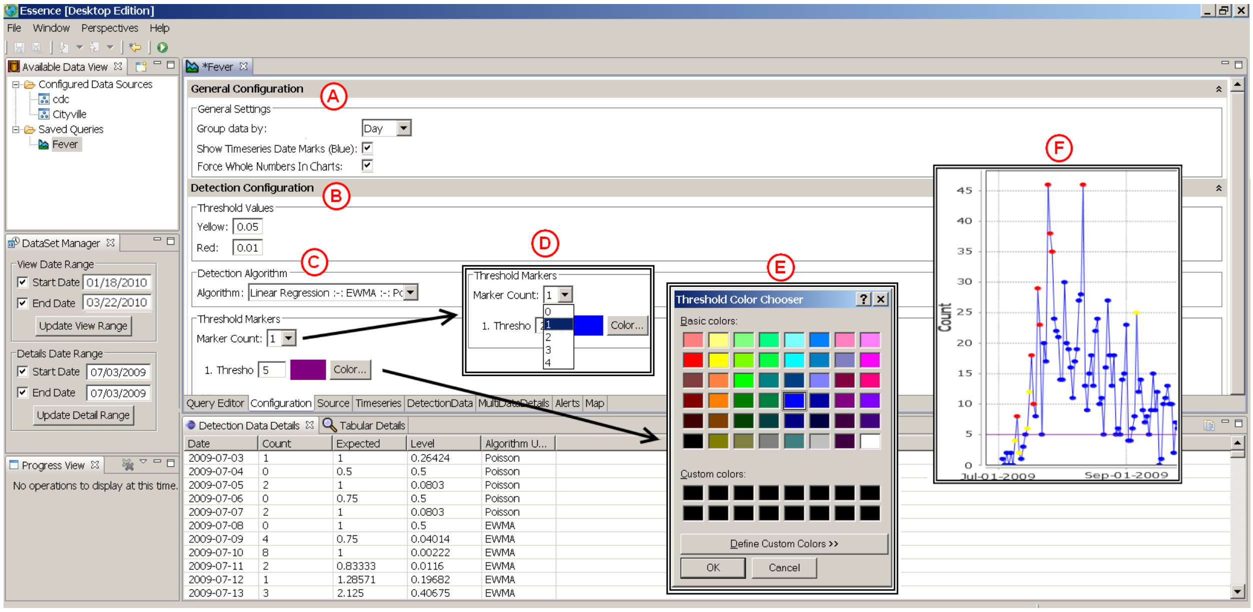
There are three groups of options under the **Detection Configuration** heading on the **Configuration** tab. (Figure 19, B)

4.2.9.2.1 Threshold Values

This option allows the user to set the p-value thresholds for Yellow and Red alerts. (Figure 19, B) A p-value is calculated for each day the detection algorithm is run. It reflects the probability that the data for that day are the same as the data in preceding days. The p-value obtained for a specific day is compared to the threshold p-values. If the p-value is lower than the threshold value for a Red or Yellow alert, that the count for that day is statistically greater than seen in the preceding days, and the day is flagged. The number of preceding days and other factors considered in calculating the p-value vary by the type of algorithm used. (Appendix I) The default threshold p-values in EDE are, 0.05 for Yellow and 0.01 for Red alerts. These are values commonly used by operators of the web-based ESSENCE system, but they can be adjusted by typing the preferred value in the appropriate box. Decreasing the threshold p-values generally reduces the number of.

4.2.9.2.2 Detection Algorithm

This feature allows the user to select the detection algorithm used by EDE. (Figure 19, C) Eight algorithms are currently available for use in EDE. To select an algorithm, left-click on the arrow in the box, and left-click the desired algorithm on the pull-down list. The default detection algorithm in EDE is the **<Linear Regression:-:EWMA:-:Poisson Switch>** algorithm. It examines the data to be analyzed and determines whether a regression, EWMA (estimated weighted moving average), or Poisson algorithm is most appropriate and uses that algorithm for the day being analyzed. It repeats the process for each day being analyzed. This is a good overall algorithm for novice users. It is also the one used in the Cityville exercises. The other algorithms available are all commonly used in the public-health community in the United States. All of the algorithms are good in some situations and not as good in others. A brief discussion of the pros and cons of the algorithms is available in Appendix I, with references to recent literature which discuss the issue in more depth.



DataSet Manager

View Date Range
☑ Start Date 01/18/2010
☑ End Date 03/22/2010
Update View Range

Details Date Range
☑ Start Date 07/03/2009
☑ End Date 07/03/2009
Update Detail Range

No operations to display at this time.

General Configuration

General Settings
Group data by: Day
Show Timeseries Date Marks (Blue): ☑
Force Whole Numbers In Charts: ☑

Detection Configuration

Threshold Values
Yellow: 0.05
Red: 0.01

Detection Algorithm
Algorithm: Linear Regression :-: EWMA :-: Pc

Threshold Markers
Marker Count: 1
1. Thresho 5 Color...

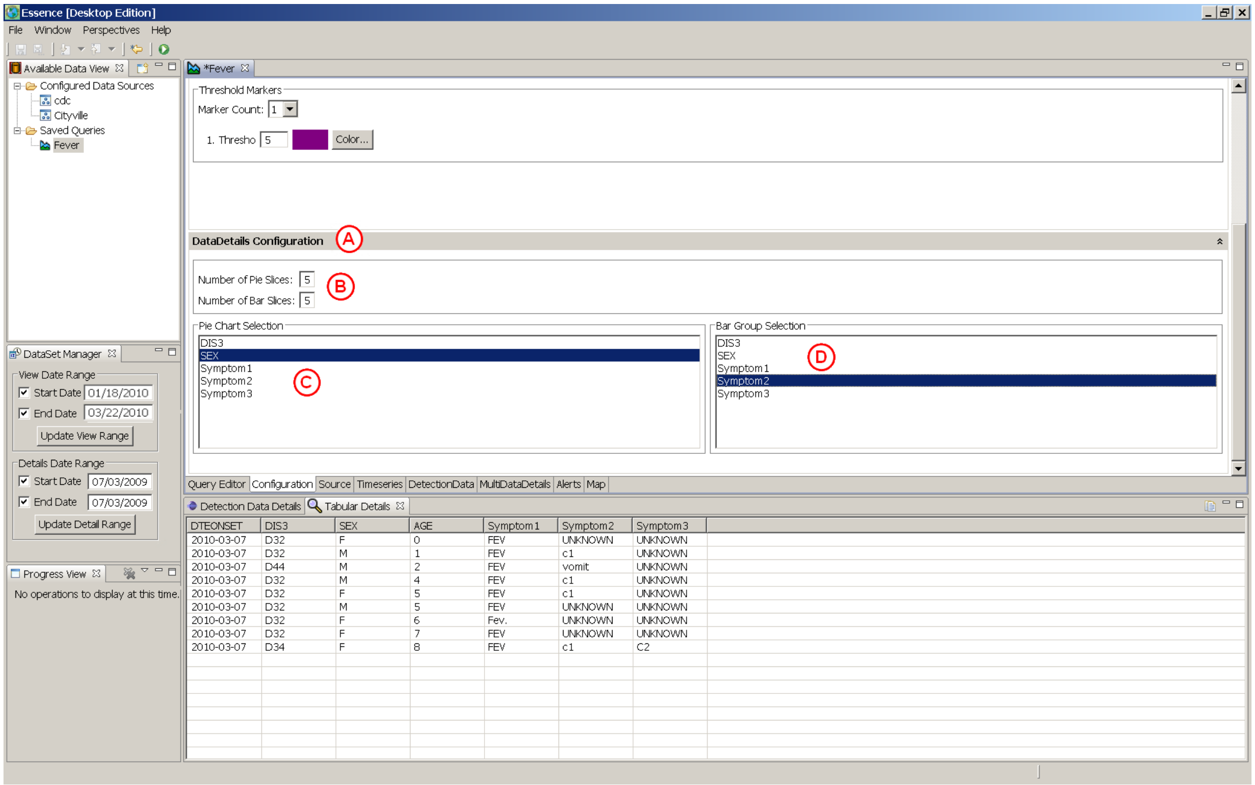
Threshold Markers
Marker Count: 1
0
1
2
3
4
1. Thresho 2 Color...

Date	Count	Expected	Level	Algorithm U...
2009-07-03	1	1	0.26424	Poisson
2009-07-04	0	0.5	0.5	Poisson
2009-07-05	2	1	0.0803	Poisson
2009-07-06	0	0.75	0.5	Poisson
2009-07-07	2	1	0.0803	Poisson
2009-07-08	0	1	0.5	EWMA
2009-07-09	4	0.75	0.04014	EWMA
2009-07-10	8	1	0.00222	EWMA
2009-07-11	2	0.83333	0.0116	EWMA
2009-07-12	1	1.28571	0.19682	EWMA
2009-07-13	3	2.125	0.40675	EWMA

Figure 19 Configuration Tab and MultiDataDetails Tab

4.2.9.2.3 Threshold Markers

For some illnesses/events, users may simply want to know whether the current number of cases has exceeded a previously determined threshold. For example, in an imaginary health district, public health officers do not care whether there is a red or yellow alert for rash unless there are a minimum of 20 cases of rash reported, To facilitate visual analysis of the data by threshold, the user can select up to four threshold lines that will be placed horizontally on the time series graph to quickly determine whether the number of cases on any given day has exceeded the threshold. (Figure 19, D) To create a threshold marker, left-click on the arrow in the box to the right of **<Marker Count>** and select the number of threshold lines to be created. For each line, type the threshold value in the box to the right of **<Threshold>**, and select the line color by left-clicking on the **<Color>** button to the right of the threshold box and left-clicking on the desired color. (Figure 19, E) In this Cityville example, a threshold line of 5 was set for the fever query. (Figure 19, F)



Marker Count: 1

1. Thresho 5

Number of Pie Slices: 5

Number of Bar Slices: 5

View Date Range: ☑ Start Date 01/18/2010 ☑ End Date 03/22/2010

Details Date Range: ☑ Start Date 07/03/2009 ☑ End Date 07/03/2009

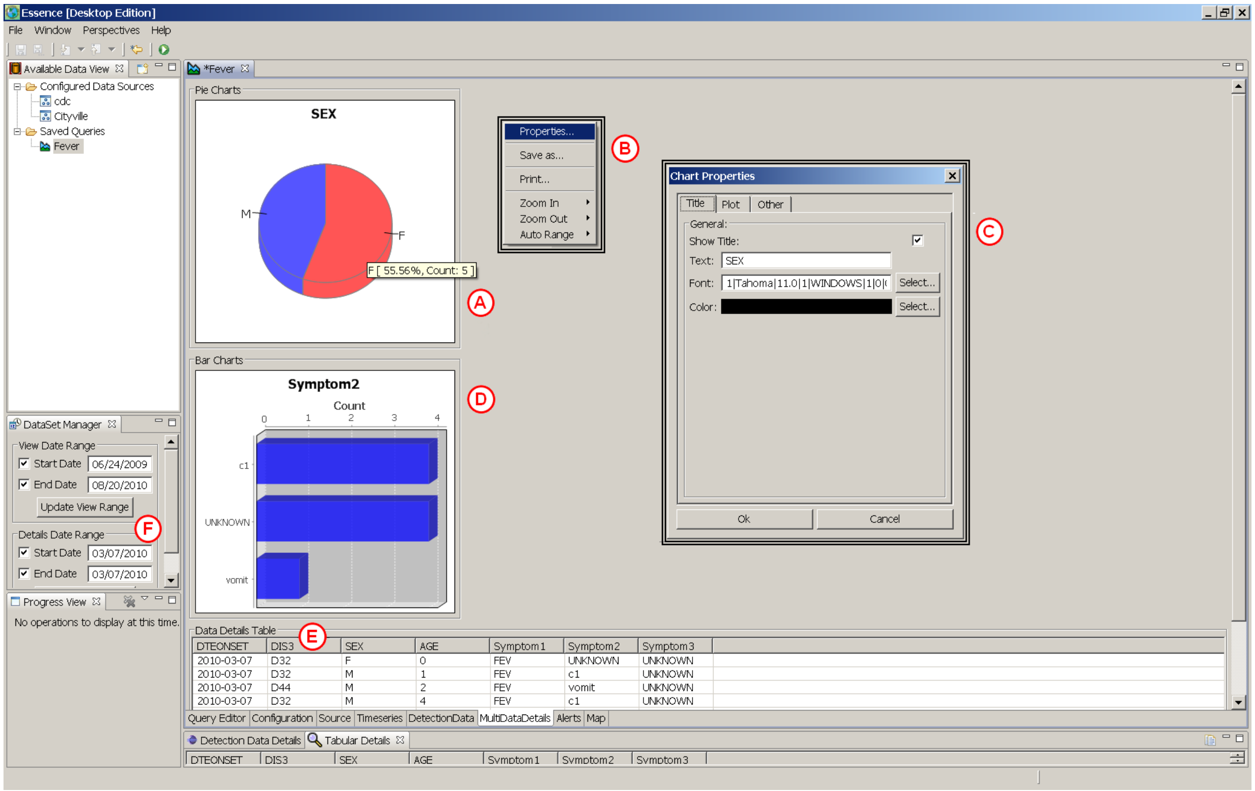
No operations to display at this time.

DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2010-03-07	D32	F	0	FEV	UNKNOWN	UNKNOWN
2010-03-07	D32	M	1	FEV	c1	UNKNOWN
2010-03-07	D44	M	2	FEV	vomit	UNKNOWN
2010-03-07	D32	M	4	FEV	c1	UNKNOWN
2010-03-07	D32	F	5	FEV	c1	UNKNOWN
2010-03-07	D32	M	5	FEV	UNKNOWN	UNKNOWN
2010-03-07	D32	F	6	Fev.	UNKNOWN	UNKNOWN
2010-03-07	D32	F	7	FEV	UNKNOWN	UNKNOWN
2010-03-07	D34	F	8	FEV	c1	C2

Figure 20 Data Details Configuration, Configuration Tab

4.2.9.3 DataDetails Configuration

The **<DataDetails Configuration>** display (Figure 20, A) at the bottom of the **Configuration** tab allows the user to select variables to be graphed (pie or bar chart) and displayed in the **MultiDataDetails** tab. (Figure 20, C-D) The user can also specify the number of slices/bars displayed on the pie charts and/or bar graphs. (Figure 20, B) In the default state, no variables are selected. To select variables for graphing, click on the desired variable name in the pie or bar chart list. (Figure 20, C-D) Multiple variables may be selected for each type of chart by highlighting all desired variables. CTRL-Click is also used to de-select a variable. To change the number of slices or bars in a chart, type the desired number in the box beside **Number of Pie Slices** or **Number of Bar Slices**. The variables Symptom2 and SEX (District) were selected for bar and pie charts, respectively. (Figure 21)



F [55.56%, Count: 5]

General:
Show Title: ☑
Text: SEX
Font: 1|Tahoma|11.0|1|WINDOWS|1|0|(
Color:

View Date Range
☑ Start Date 06/24/2009
☑ End Date 08/20/2010

Details Date Range
☑ Start Date 03/07/2010
☑ End Date 03/07/2010

DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2010-03-07	D32	F	0	FEV	UNKNOWN	UNKNOWN
2010-03-07	D32	M	1	FEV	c1	UNKNOWN
2010-03-07	D44	M	2	FEV	vomit	UNKNOWN
2010-03-07	D32	M	4	FEV	c1	UNKNOWN

DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3

Figure 21 Features of the MultiData Details Tab in the Query Window

4.2.10 The MultiDataDetails Tab

The variables selected for graphing in the **Data Details Configuration** in the **Configuration** tab of the **Query** window are displayed in the **MultiDataDetails** tab of the **Query** window. (Figure 21, A & D) The count and percent of total for a specific slice of a pie chart or a bar in a bar graph are displayed over the chart when the cursor hovers over it. (Figure 21, A)

The appearance of pie and bar chart properties can be modified by right-clicking on the chart/graph in the **MultiDataDetails** tab in the **Query** window. This opens a window with six options. (Figure 21, B) Left-clicking on the first option, **Properties**, opens a **Chart Properties** box that allows the user to change a variety of display factors for the chart (Figure 21, C) In Figure 21, the Chart Properties box (G) is open for the SEX variable bar chart, and the bar orientation is being changed from vertical, the default, to horizontal.

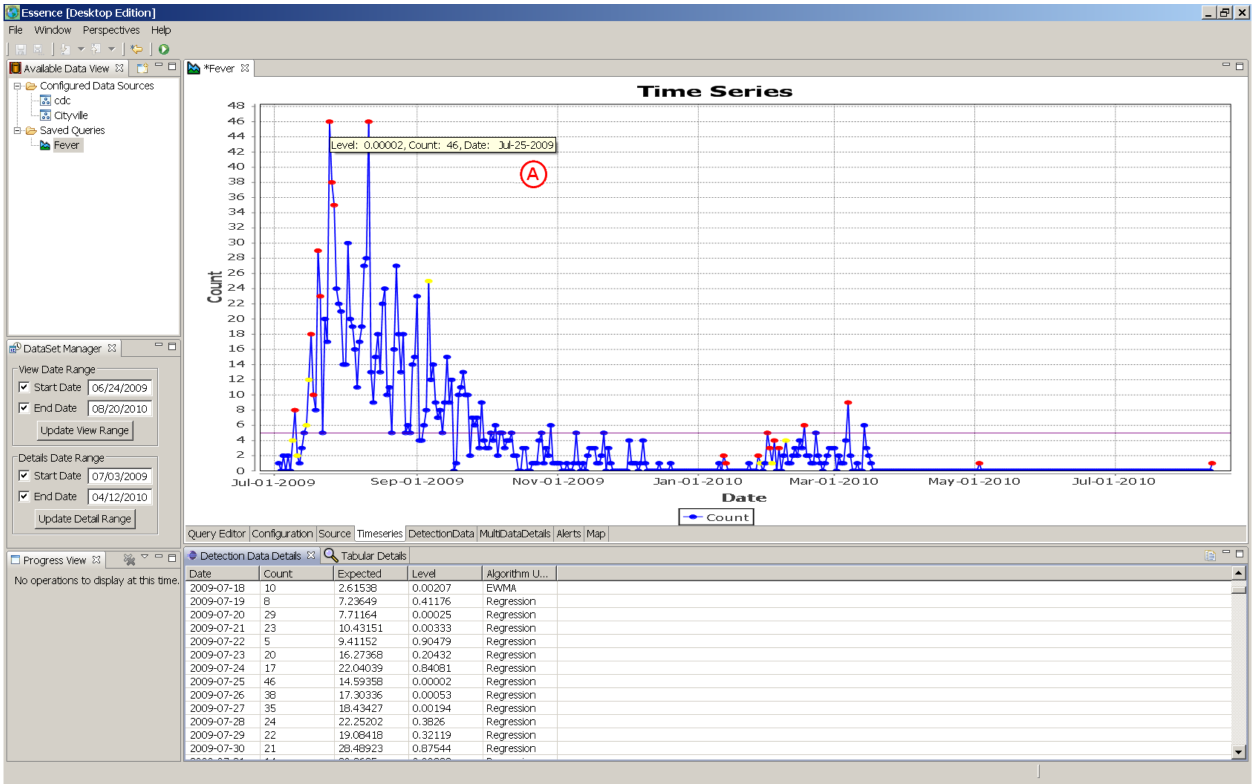
The window opened by right-clicking on a chart/graph has several other useful options. Left-clicking on **Save As…** allows the user to save the chart/graph as a picture (.png) file for easy incorporation into reports, and **Print** allows the user to print a copy of the chart. The other options listed - **Zoom In**, **Zoom Out**, and **Auto Range -** are not enabled in for pie charts, but can be used on bar graphs although their utility is limited.

A sortable list of the cases/records making up the query is displayed in the **Data Details Table** at the bottom of the **MultiDataDetails** tab. (Figure 21, E) This same data is displayed in the **Tabular Details** tab which opens in the default EDE main screen below the **Query** window. (Figure 17, A) The **Data Details Table** opens in the default EDE screen below the time series because the information in it is useful in determining whether an alert on the time series is due to an artifact or may represent a true disease outbreak.

The data displayed in the **MultiDataDetails** tab correspond to the time period selected under **Details Date Range** in the **DataSet Manager** window. (Figure 21, F) In this Cityville example, only a single day 7/3/2010, is included in the **Data Details Table** and **Tabular Details** window. (Figure 21, E & F)

4.2.11 Alert Configuration Tab

The **DetectionData** tab, the **Detection Data Details** window and the time series in the current Cityville example show alerts for the selected time period for all districts combined because no specific region was included in the query. A Red alert on 7/25/2009 on the time series (Figure 22 A) means on that day the number of cases in all districts combined was higher than expected.



View Date Range

☑ Start Date 06/24/2009

☑ End Date 08/20/2010

Details Date Range

☑ Start Date 07/03/2009

☑ End Date 04/12/2010

No operations to display at this time.

Date	Count	Expected	Level	Algorithm U...
2009-07-18	10	2.61538	0.00207	EWMA
2009-07-19	8	7.23649	0.41176	Regression
2009-07-20	29	7.71164	0.00025	Regression
2009-07-21	23	10.43151	0.00333	Regression
2009-07-22	5	9.41152	0.90479	Regression
2009-07-23	20	16.27368	0.20432	Regression
2009-07-24	17	22.04039	0.84081	Regression
2009-07-25	46	14.59358	0.00002	Regression
2009-07-26	38	17.30336	0.00053	Regression
2009-07-27	35	18.43427	0.00194	Regression
2009-07-28	24	22.25202	0.3826	Regression
2009-07-29	22	19.08418	0.32119	Regression
2009-07-30	21	28.48923	0.87544	Regression

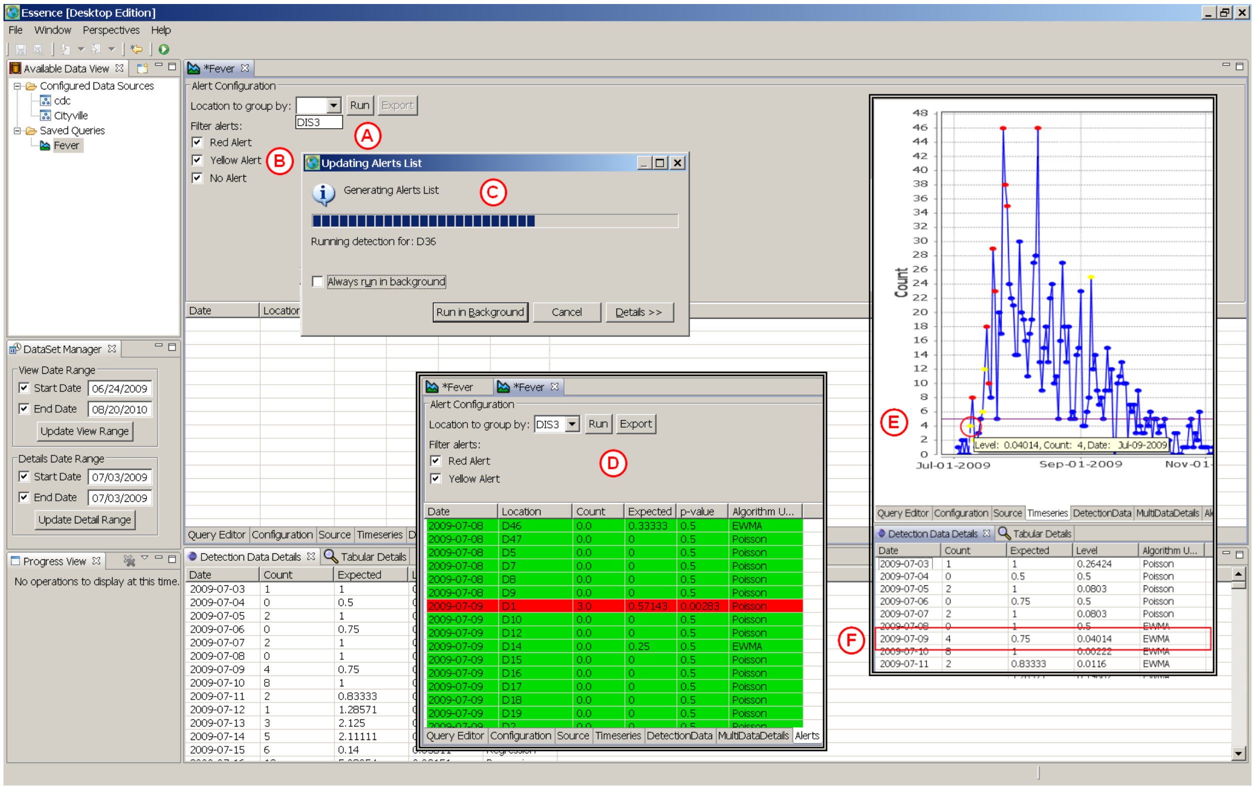
Figure 22 Truncated Fever Time Series

EDE can also compute alerts for specific geographic regions if they are specified as locations during DSDC creation. The **Alert Configuration** box on the **Alerts** tab of the Query window allows the user to specify a location variable included in the DSDC so that detection can be run at the level of that variable. Click the **Alerts** tab in the **Query** window to open the tab, and then click the pull-down arrow next to **Location to group by:** in the box at the top of the window. In this Cityville example, the location variable is

DIS3. (Figure 23, A) The values of this variable are the different districts of Cityville. Click on DIS3 to select it. The user can choose to see **Red**, **Yellow** and/or **No Alerts** by clicking the appropriate checkbox(es). (Figure 23, B) No alerts are null days, when the p-value calculated by the algorithm is greater than either the yellow or red threshold p-values, i.e. days when the observed count is not statistically different than the expected count.

After selecting the location variable, left-click on **Run** button next to the location variable box. A progress box entitled **Updating Alerts List** appears to indicate that the alerts are being calculated. (Figure 23, C) This box is replaced by a revised screen containing the location and date specific alerts list at the of the **Alert Configuration** tab. (Figure 23, D) **No Alert** days are highlighted in green on the **Alert Configuration** results table, and red and yellow alerts are highlighted in red and yellow respectively. (Figure 23, D)

After the alerts are run, the **Alerts List** can be exported to a simple text file (.csv) by clicking on the **Export** button at the top of the **Alert Configuration** box. (Figure 23, A) This opens a **Data Export** window where the user can browse to a location in the computer and save the list as a simple text file.



Alert Configuration

Location to group by: DIS3

Filter alerts:

☑ Red Alert

☑ Yellow Alert

☑ No Alert

Updating Alerts List

Generating Alerts List

Running detection for: D36

☐ Always run in background

Run in Background | Cancel | Details >>

DataSet Manager

View Date Range

☑ Start Date 06/24/2009

☑ End Date 08/20/2010

Update View Range

Details Date Range

☑ Start Date 07/03/2009

☑ End Date 07/03/2009

Update Detail Range

Progress View

No operations to display at this time.

Date	Count	Expected
2009-07-03	1	1
2009-07-04	0	0.5
2009-07-05	2	1
2009-07-06	0	0.75
2009-07-07	2	1
2009-07-08	0	1
2009-07-09	4	0.75
2009-07-10	8	1
2009-07-11	2	0.83333
2009-07-12	1	1.28571
2009-07-13	3	2.125
2009-07-14	5	2.11111
2009-07-15	6	0.14

Date	Location	Count	Expected	p-value	Algorithm U...
2009-07-08	D46	0.0	0.33333	0.5	EWMA
2009-07-08	D47	0.0	0	0.5	Poisson
2009-07-08	D5	0.0	0	0.5	Poisson
2009-07-08	D7	0.0	0	0.5	Poisson
2009-07-08	D8	0.0	0	0.5	Poisson
2009-07-08	D9	0.0	0	0.5	Poisson
2009-07-09	D1	3.0	0.57143	0.00283	Poisson
2009-07-09	D10	0.0	0	0.5	Poisson
2009-07-09	D12	0.0	0	0.5	Poisson
2009-07-09	D14	0.0	0.25	0.5	EWMA
2009-07-09	D15	0.0	0	0.5	Poisson
2009-07-09	D16	0.0	0	0.5	Poisson
2009-07-09	D17	0.0	0	0.5	Poisson
2009-07-09	D18	0.0	0	0.5	Poisson
2009-07-09	D19	0.0	0	0.5	Poisson

Date	Count	Expected	Level	Algorithm U...
2009-07-03	1	1	0.26424	Poisson
2009-07-04	0	0.5	0.5	Poisson
2009-07-05	2	1	0.0803	Poisson
2009-07-06	0	0.75	0.5	Poisson
2009-07-07	2	1	0.0803	Poisson
2009-07-08	0	1	0.5	EWMA
2009-07-09	4	0.75	0.04014	EWMA
2009-07-10	8	1	0.00222	EWMA
2009-07-11	2	0.83333	0.0116	EWMA

Figure 23 Alerts Tab –Configuration and Results

The dates in the **Alerts** tab correspond to those under **Details Date Range** in the **DataSet Manager** window. The date defaults to the current day or the first day of the time period specified in the query. To select a different date range, type over the default Start and End dates under the **Details Date Range**, and click **Update Detail Range**. In the Cityville example, replace the Start date with 07/03/2009, and the End date with 4/12/2010.

When the location-based alert calculations are finished, the alerts appear in the bottom of the **Alert Configuration** window. (Figure 23, D) These results may not be the same as those appearing on the **Time Series** display, because the alerts on the time series are for the entire region, while those in the **Alerts List** are stratified by location (DIS3). For example, the time series (Figure 23, E & F) shows a yellow alert for 7/9/2009 for all locations, while in the **Alerts List** (Figure 23, D) only District 1 shows a red alert and the other districts show no alerts...

4.2.12 The Map Tab

The **Map** tab enables exports of the EDE data and calculations to UDig or EpiMap. The count, percent, and/or alert level can be displayed by any geographic region specified as a location in the DSDC, and for which map shape files are available. The UDig software is included in the EDE program files and does not need to be installed separately onto your computer. To create a map from EDE using EpiMap, however, EpiMap must be installed on your computer and be part of the computer's path. For either mapping program, EDE assumes the map files are in the EDE subdirectory (C:\EDE\workspace\workspace\MapFiles\). It is recommended that a copy of the shape files, not the original files, be used because EDE modifies the .dbf shape files during processing. Copy the shape files to C:\EDE\WORKSPACE\WORKSPACE\MAPFILES.

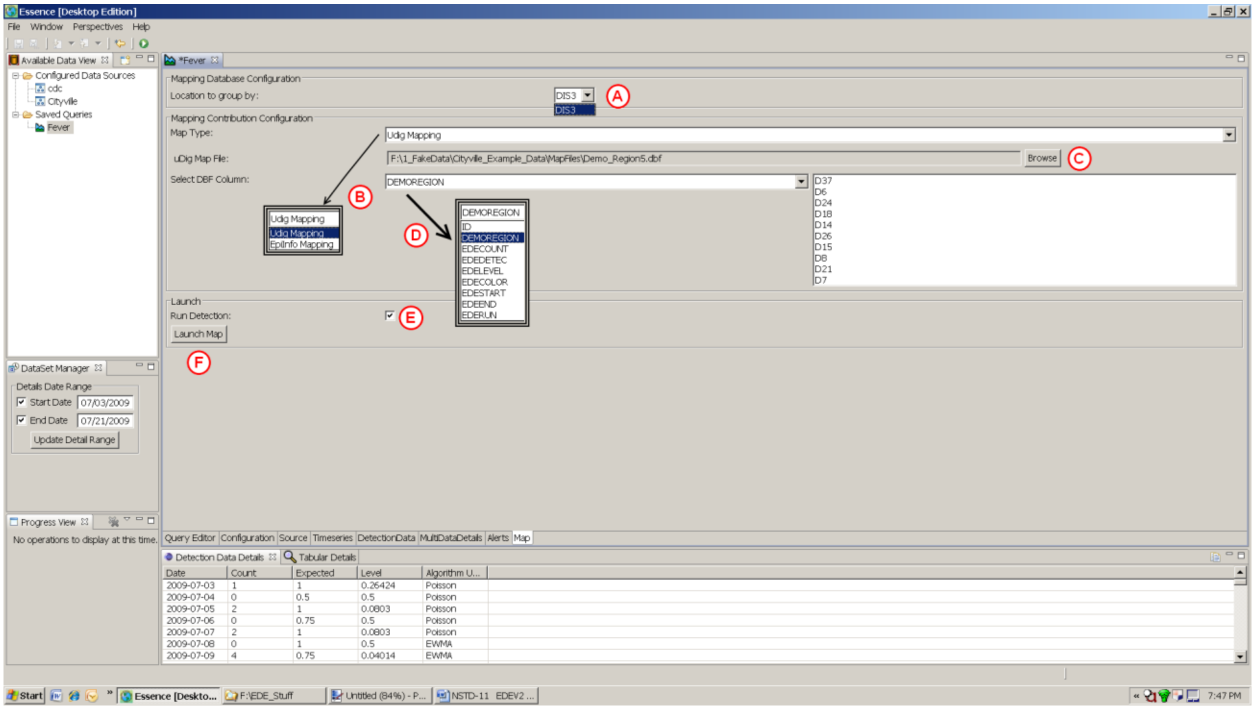
Both the UDig and EpiMap programs are fairly complex and take time to learn. The instructions here will help the user set up simple maps, but the full extent of the UDig and EpiMap programs are not addressed in this document. Tutorials for both programs can be found on the web.

EDE calculates several variables for each location value and adds them to the .DBF file being used. The variables **EDESTART** and **EDEEND** contain the start and end dates listed in the **Details Date Range** box. **EDECOUNT** is the number of cases observed in the region during the specified time period. **EDECOLOR** is a string variable coding the alert level for each district, GREEN (no alert), YELLOW, or RED. **EDELEVEL** is a numeric variable that codes the alert level, GREEN=0, YELLOW=1, and RED=2. Finally, **EDEDETEC** lists the probability (p-value) that the observed number of cases for the district was no greater than expected.

Maps can be created for single or multiple day periods by changing the Start and End dates in the Details Date Range box. However, the detection algorithms work by comparing the number of observed cases on a given day with the number expected in a set preceding time period. This means that detection can only be run on a single day. If multiple days are selected in the Details Date Range box, detection will be run and presented in the map only for the End date of the period. If any of the detection variables are mapped, the alerts mapped will be for the End date only. However, if the case count (EDECOUNT) is mapped all of the cases in the multiple day time period will be mapped.

To begin mapping open the **Map** tab in the **Query** window. At the top of the screen under **Mapping Database Configuration** click on the pull down arrow beside **Location to group by** and select the **Location** variable to be used. (Figure 24, A) In the Cityville example, there is only one location variable, DIS3 which is the Cityville district number.

Next select the mapping program. Click on the pull-down arrow in the box next to **Map Type** under **Mapping Contribution Configuration** and click on UDig or EpiMap. (Figure 24, B)



Mapping Database Configuration

Location to group by: DIS3

Mapping Contribution Configuration

Map Type: Udig Mapping

Udig Mapping
Udig Mapping
EpiInfo Mapping

uDig Map File: F:\1_FakeData\Cityvile_Example_Data\MapFiles\Demo_Region5.dbf

Select DBF Column: DEMOREGION

DEMOREGION
ID
DEMOREGION
EDECOUNT
EDEDETEC
EDELEVEL
EDECOLOR
EDESTART
EDEEND
EDERUN

D37
D6
D24
D18
D14
D26
D15
D8
D21
D7

Launch

Run Detection: ☑

Details Date Range

☑ Start Date 07/03/2009

☑ End Date 07/21/2009

No operations to display at this time.

Date	Count	Expected	Level	Algorithm U...
2009-07-03	1	1	0.26424	Poisson
2009-07-04	0	0.5	0.5	Poisson
2009-07-05	2	1	0.0803	Poisson
2009-07-06	0	0.75	0.5	Poisson
2009-07-07	2	1	0.0803	Poisson
2009-07-08	0	1	0.5	EWMA
2009-07-09	4	0.75	0.04014	EWMA

Figure 24 Introduction to the Map Tab of the Query Window

4.2.12.1 Using UDig with EDE

4.2.12.1.1 UDig: Select the DBF file

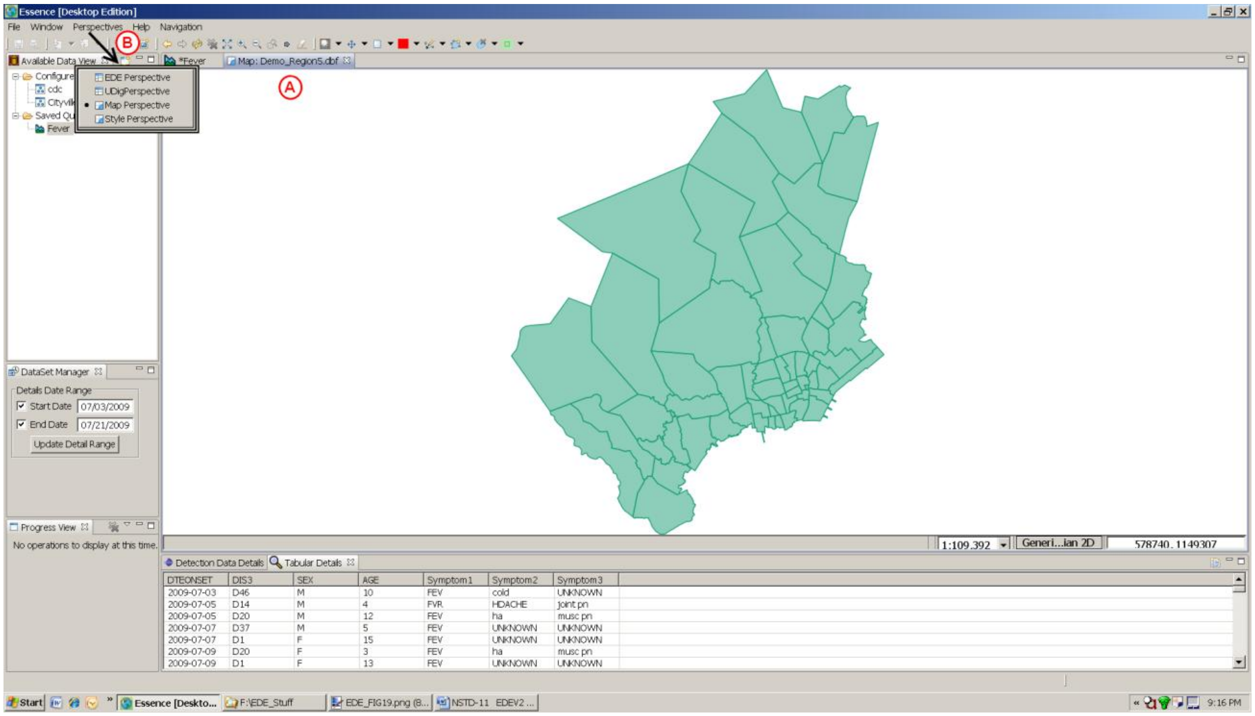
Use the **Browse** button next to **UDig Map File** (Figure 24, C) to select the location of the .dbf shape file to be used. For the Cityville example the file should be **Demo_Region5.dbf** that was placed in C:\EDE\WORKSPACE\WORKSPACE\MAPFILES. In the pull-down box next to **Select DBF Column**, pick the variable that contains the region names for the location variable being used (Figure 24, D) In the Cityville example, select DEMOREGION. The values of DEMOREGION appear in the box on the far right. Decide if detection should be run on the various levels of DEMOREGION. If so, then click the checkbox next to **Run Detection**. In this Cityville example, **Run Detection** was selected (Figure 24, E) In the Cityville example, the alerts and map will be created for a single day, 7/21/2009 (Figure 24)

4.2.12.1.2 UDig: Launch the Map

Once the appropriate date range is selected, click the **Launch Map** button (Figure 24, F) to create the map. A pop-up window displaying the status of the map generation appears in the **Query** window when the launch button is pressed. UDig will open automatically in a new tab in the Query Window and display the specified map. (Figure 25, A)

Reminder

The map displays data from the time period selected under **Details Date Range** in the **DataSet Manager** window. The map will fail if the Details Start Date is earlier than the first record in the DSDC or the Start date is later than the End date. Correct the Start and/or End date(s) and launch the map again.



Details Date Range
☑ Start Date 07/03/2009
☑ End Date 07/21/2009
Update Detail Range

No operations to display at this time.

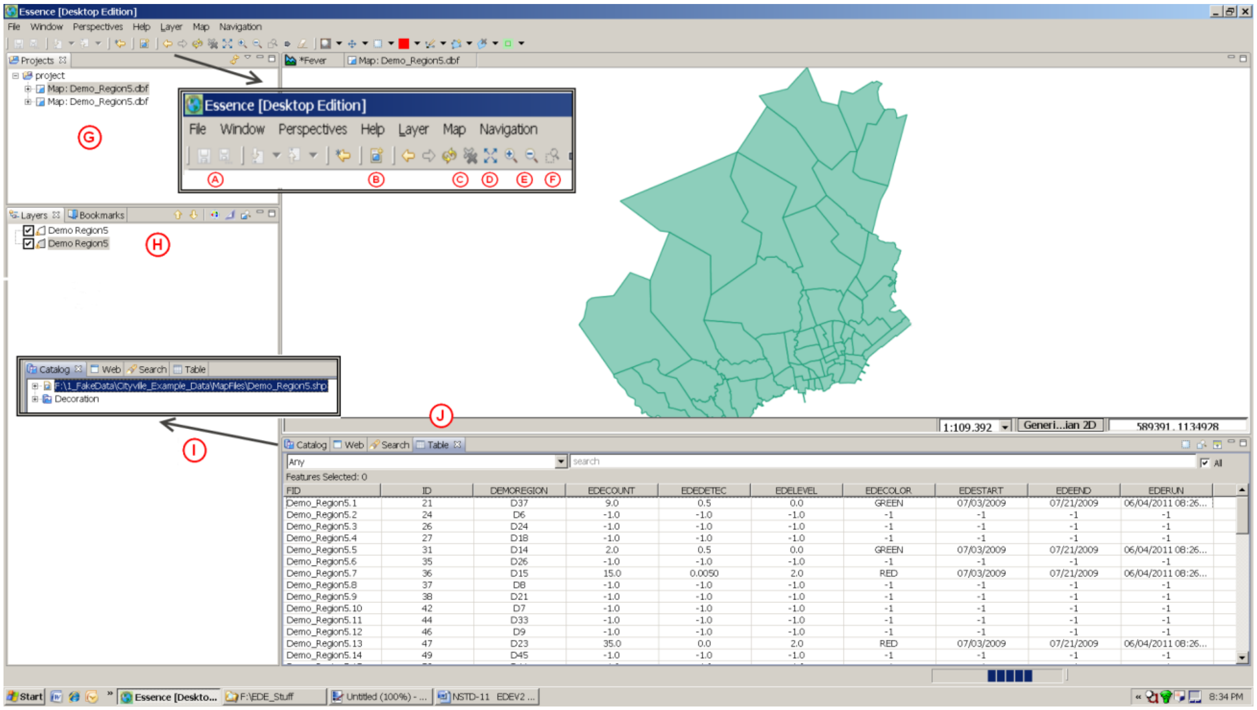
DTEONSET	DIS3	SEX	AGE	Symptom1	Symptom2	Symptom3
2009-07-03	D46	M	10	FEV	cold	UNKNOWN
2009-07-05	D14	M	4	FVR	HDACHE	joint pn
2009-07-05	D20	M	12	FEV	ha	musc pn
2009-07-07	D37	M	5	FEV	UNKNOWN	UNKNOWN
2009-07-07	D1	F	15	FEV	UNKNOWN	UNKNOWN
2009-07-09	D20	F	3	FEV	ha	musc pn
2009-07-09	D1	F	13	FEV	UNKNOWN	UNKNOWN

Figure 25 UDig Cityville Map

4.2.12.1.3 UDig: A Quick Tour

When the map tab opens in UDig, click on **Perspectives** in the Windows bar at the top of the screen (Figure 25, B) and change from the **EDE Perspective** to the **Map Perspective**. This opens an UDig view of the map tab. This perspective is divided into a number of tabbed boxes like the **EDE Perspective**. In the **UDig Perspective** the **Project** tab (Figure 26, G) shows the project name for the map being created (at the bottom of the list) and maps previously created. Below this is the **Layers** tab (Figure 26, H), which shows the different layers of the map being created. The map itself opens in a layer labeled with the name of the .dbf file related to the shape files being used. The EDE query tab shares a window with the map tab. Below the map tab is a window with four tabs: **Catalog**, **Web**, **Search**, and **Table**. The novice user need only be concerned with the **Catalog** and **Table** tabs. The **Catalog** tab lists the .dbf files used to create the map layers listed under the Layers tab. (Figure 26, I) The **Table** tab contains a view of the .dbf file being used to create the map. (Figure 26, J)The user can check the values in the **Table** tab if the map is failing to see if they correspond to the expected values.

Below the Windows bar at the top of the **Map Perspective** is a row of icons. An inset in Figure 26 shows a blow-up of some useful icons. (Figure 26, A) These include: **Create a Page to be Printed** (Figure 26, B), **Redraw & Stop Drawing Map** (Figure 26, C), **Show All Data** (Figure 26, D) which returns the view to the entire map, **Zoom In & Out**, (Figure 26, E), and **Show Selected Data** (Figure 26, F) which allows the user to drag a box over the map to select a specific area for viewing.



Any

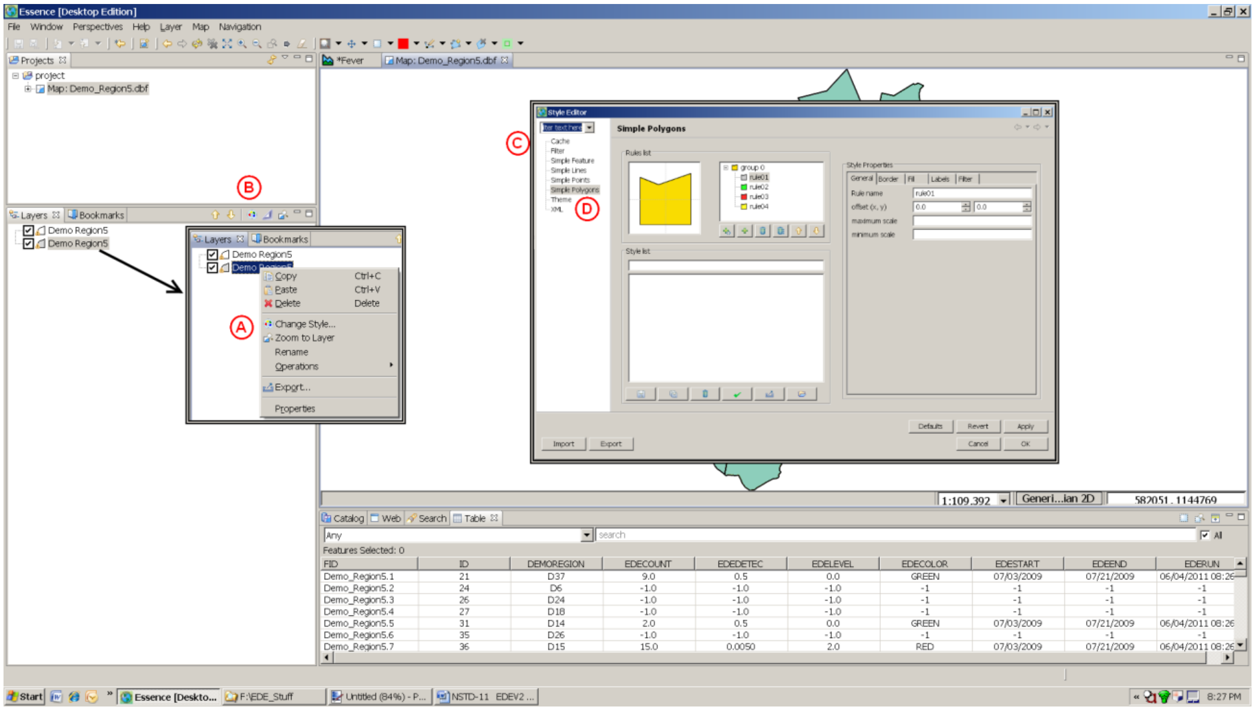
search

☑ All

Features Selected: 0

FID	ID	DEMOREGION	EDECOUNT	EDEDETEC	EDELEVEL	EDECOLOR	EDESTART	EDEEND	EDERUN
Demo_Region5.1	21	D37	9.0	0.5	0.0	GREEN	07/03/2009	07/21/2009	06/04/2011 08:26...
Demo_Region5.2	24	D6	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.3	26	D24	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.4	27	D18	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.5	31	D14	2.0	0.5	0.0	GREEN	07/03/2009	07/21/2009	06/04/2011 08:26...
Demo_Region5.6	35	D26	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.7	36	D15	15.0	0.0050	2.0	RED	07/03/2009	07/21/2009	06/04/2011 08:26...
Demo_Region5.8	37	D8	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.9	38	D21	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.10	42	D7	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.11	44	D33	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.12	46	D9	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.13	47	D23	35.0	0.0	2.0	RED	07/03/2009	07/21/2009	06/04/2011 08:26...
Demo_Region5.14	49	D45	-1.0	-1.0	-1.0	-1	-1	-1	-1

Figure 26 UDig, The Map Perspective



Rule name: rule01

offset (x, y): 0.0 0.0

Any

search

☑ All

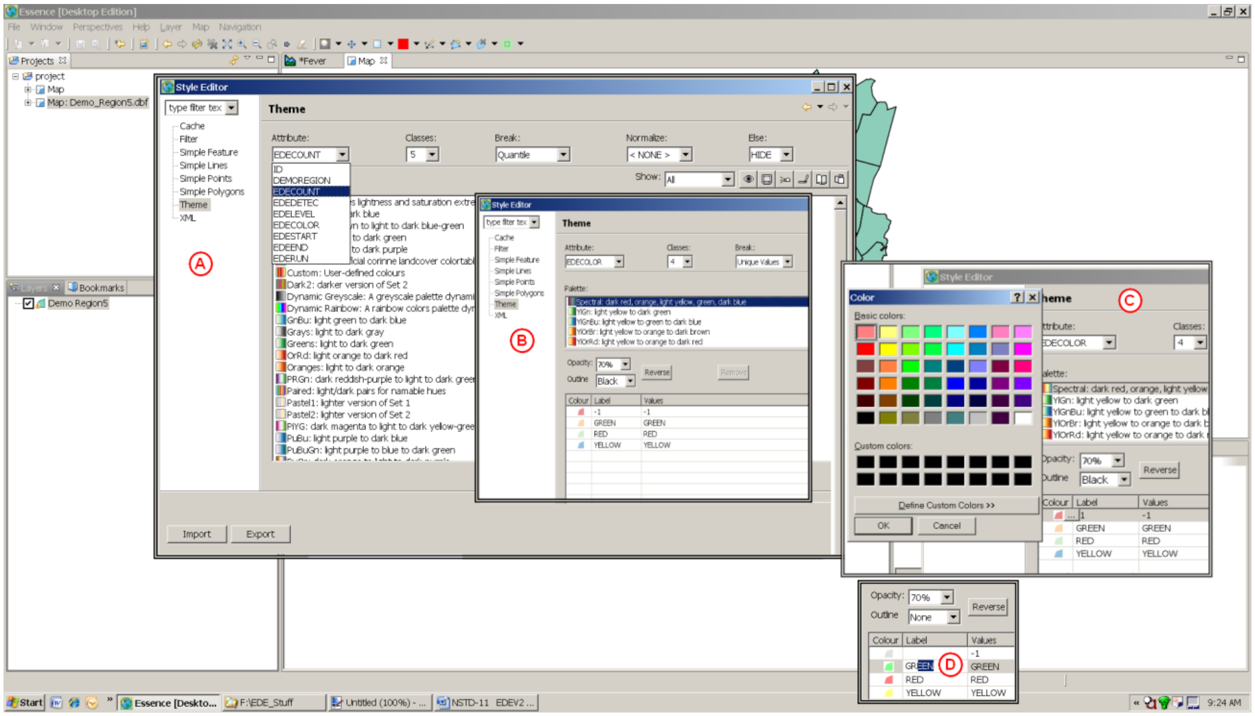
Features Selected: 0

FID	ID	DEMOREGION	EDECOUNT	EDEDETEC	EDELEVEL	EDECOLOR	EDESTART	EDEEND	EDERUN
Demo_Region5.1	21	D37	9.0	0.5	0.0	GREEN	07/03/2009	07/21/2009	06/04/2011 08:26
Demo_Region5.2	24	D6	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.3	26	D24	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.4	27	D18	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.5	31	D14	2.0	0.5	0.0	GREEN	07/03/2009	07/21/2009	06/04/2011 08:26
Demo_Region5.6	35	D26	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.7	36	D15	15.0	0.0050	2.0	RED	07/03/2009	07/21/2009	06/04/2011 08:26

Figure 27 UDig: Creating an Alert Map, Style Editor

4.2.12.1.4 UDig: Creating an Alert Map

In the 06/03/2011 version of EDE the UDig map will appear as in Figure 25. To create a map showing EDE alert colors (red, yellow & green) the layers need to be modified. To modify a layer first right-click on the layer name. (Figure 27) A pop-up box will appear with nine options. Left-click on **Change Style…** (Figure 27, A); alternatively highlight the layer to be modified and left-click on the **Change Style** icon to the right of the **Layers** tab name. (Figure 27, B) This opens a pop-up box labeled **Style Editor**. (Figure 27, C)



Theme

Attribute: EDECOUNT Classes: 5 Break: Quantile Normalize: < NONE > Else: HIDE

Show: All

Attribute options: ID, DEMOREGION, EDECOUNT, EDEDETEC, EDELEVEL, EDECOLOR, EDESTART, EDEEND, EDERUN

Palette:
- Custom: User-defined colours
- Dark2: darker version of Set 2
- Dynamic Greyscale: A greyscale palette dynami
- Dynamic Rainbow: A rainbow colors palette dy
- GnBu: light green to dark blue
- Grays: light to dark gray
- Greens: light to dark green
- OrRd: light orange to dark red
- Oranges: light to dark orange
- PRGn: dark reddish-purple to light to dark gree
- Paired: light/dark pairs for namable hues
- Pastel1: lighter version of Set 1
- Pastel2: lighter version of Set 2
- PiYG: dark magenta to light to dark yellow-gree
- PuBu: light purple to dark blue
- PuBuGn: light purple to blue to dark green

Theme

Attribute: EDECOLOR Classes: 4 Break: Unique Values

Palette:
- Spectral: dark red, orange, light yellow, green, dark blue
- YlGn: light yellow to dark green
- YlGnBu: light yellow to green to dark blue
- YlOrBr: light yellow to orange to dark brown
- YlOrRd: light yellow to orange to dark red

Opacity: 70% Outline: Black

Colour	Label	Values
	-1	-1
	GREEN	GREEN
	RED	RED
	YELLOW	YELLOW

Opacity: 70% Outline: None

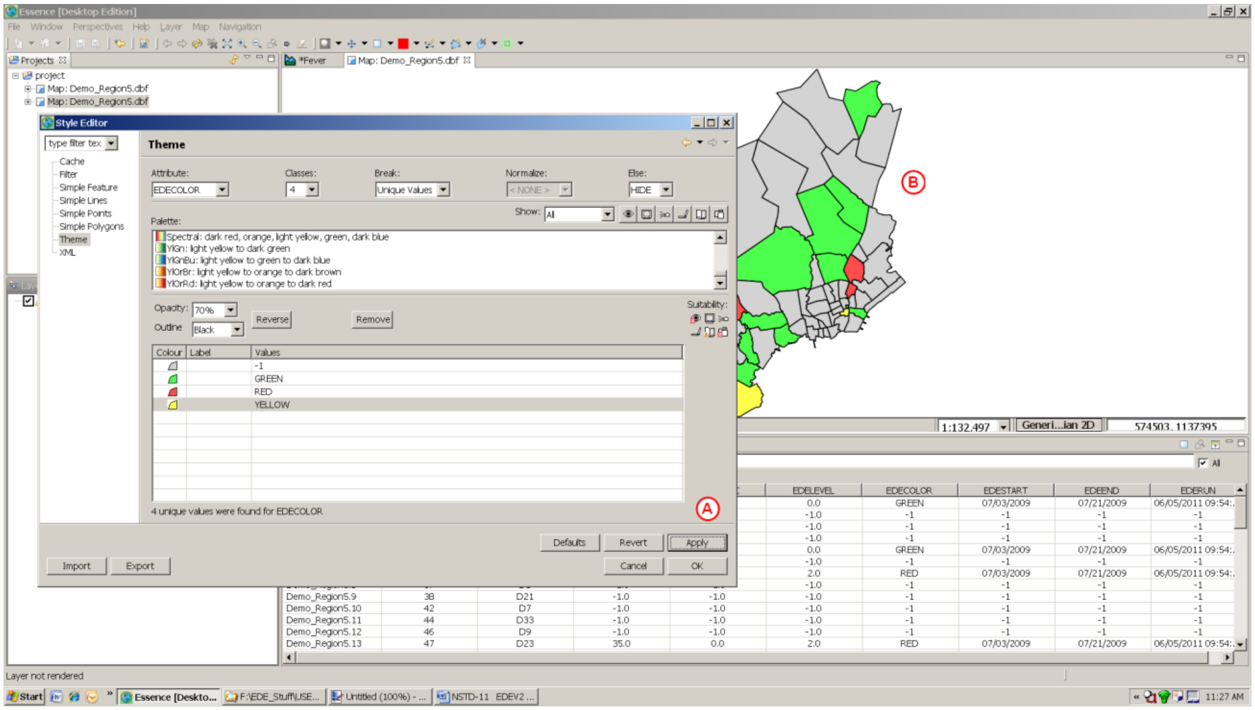
Colour	Label	Values
		-1
	GREEN	GREEN
	RED	RED
	YELLOW	YELLOW

Figure 28 UDig: Creating an Alert Map, Theme Box

In the **Style Editor** left click on **Theme** (Figure 27, D) to configure the EDECOLOR variable. In the **Theme** box that opens click on the box next to **Attribute** and select the variable **EDECOLOR**. (Figure 28, A) This will change the values of **Classes** and **Break**. **Classes** becomes equal to the total number of values in **EDECOLOR**, and because **EDECOLOR** is discrete, **Break** changes from **Quantiles** to **Unique Values**. (Figure 28, B) Next under Palette select the desired color palette. In this Cityville example the Spectral palette was selected. (Figure 28, B) Below the palette selection are several other configuration variables. In this example O**pacity** is set at 70%, instead of the initial 50%, and Outline is set to Black to add black outlines around each district. Figure 28, C)

As mentioned previously, EDECOLOR has 4 values, **-1** means there is no information for that district, **GREEN** means the alert is null, and **RED** and **YELLOW** means the alert is red or yellow respectively. Remember these values can be seen in the **Table** tab at the bottom of the **Map Perspective**. (Figure 26, J) UDig applies colors to variable values automatically. (Figure 28, B) To make the map colors reflect the alert colors left-click on the **Color** box to the left of the value to be changed. A small box appears to the left of the color block. Left-click on the box to open a **Color** selection box. (Figure 28 C)

Then left-click the desired color and left-click **OK**. In this Cityville example the grey-blue color (third from the right on the bottom of the box) was selected for the **-1** value (i.e. no information). Click on the **Color** box for each value of **EDECOLOR** and select the color corresponding to the value. (Figure 28, D) If left as seen in Figure 28, C, UDig will write the EDECOLOR values as a label on each district. This is redundant since the value is the same as the color just applied. To remove the labels click on the text under label for each of the four colors and erase the label. Leave the label blank. (Figure 28, D)



Attribute: EDECOLOR

Classes: 4

Break: Unique Values

Normalize: < NONE >

Else: HIDE

Show: All

Palette:

- Spectral: dark red, orange, light yellow, green, dark blue
- YlGn: light yellow to dark green
- YlGnBu: light yellow to green to dark blue
- YlOrBr: light yellow to orange to dark brown
- YlOrRd: light yellow to orange to dark red

Opacity: 70%

Outline: Black

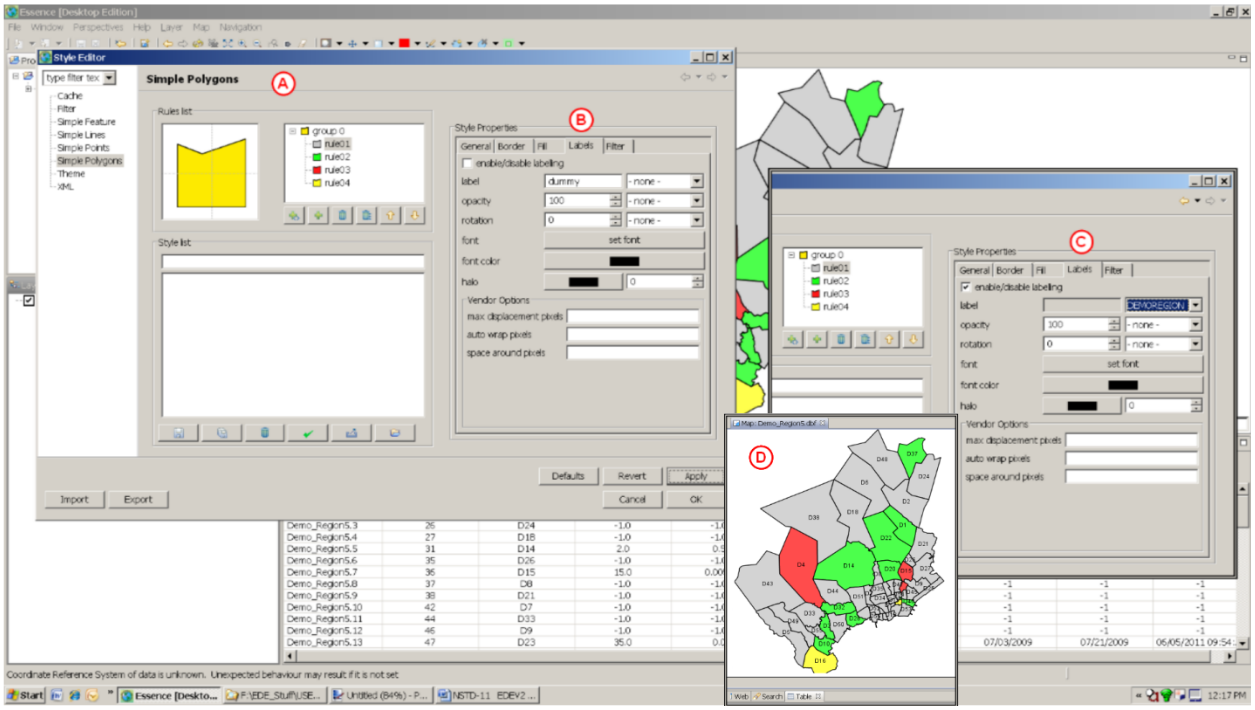
Colour	Label	Values
▱		-1
🟩		GREEN
🟥		RED
🟨		YELLOW

4 unique values were found for EDECOLOR

					EDELEVEL	EDECOLOR	EDESTART	EDEEND	EDERUN
					0.0	GREEN	07/03/2009	07/21/2009	06/05/2011 09:54:.
					-1.0	-1	-1	-1	-1
					-1.0	-1	-1	-1	-1
					-1.0	-1	-1	-1	-1
					0.0	GREEN	07/03/2009	07/21/2009	06/05/2011 09:54:.
					-1.0	-1	-1	-1	-1
					2.0	RED	07/03/2009	07/21/2009	06/05/2011 09:54:.
					-1.0	-1	-1	-1	-1
Demo_Region5.9	38	D21	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.10	42	D7	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.11	44	D33	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.12	46	D9	-1.0	-1.0	-1.0	-1	-1	-1	-1
Demo_Region5.13	47	D23	35.0	0.0	2.0	RED	07/03/2009	07/21/2009	06/05/2011 09:54:.

Figure 29 UDig: Creating an Alert Map, Theme Box 2

Once the desired changes are made left-click **Apply** (Figure 29, A) at the bottom right of the **Style Editor** box. The map will redraw showing the selections made. (Figure 29, B) The last thing to do is to add the district numbers as labels on the map. Left-click on **Simple Polygons** in the left had box of the **Style Editor** window. A new window, labeled **Simple Polygons**, opens. (Figure 30, A) There are three sections in this box: **Rule List**, **Style List**, **and Style Properties**. In **Rule List** there is a diagram of a polygon and a list of groups in the map. The groups correspond to those created in the **Theme** box. So in Figure 30, group 0 with the grey-blue color corresponds to level **-1**, or no info, created in the **Theme** box; the green box is level green, etc. To add labels to the map, the **Label** properties under **Style Properties** have to be changed individually for each group, and they must be changed in a specific order. To do this highlight group 1 under the Rule List and left-click on the Label tab at the top of the Style Properties box. (Figure 30, B) Make sure the **enable/disable mapping** box is unchecked, the word 'dummy' appears in the first box to the right of **label**, and **–none-** appears in the box to the right of that. (Figure 30, B) Next click the enable/disable mapping box, delete the word 'dummy' in the first box and left-click on **DEMOREGION** in the pull-down box to the far right of **label.** If the word 'dummy' appears in the window again repeat the process until it disappears. Repeat these steps for each group. Finally, left-click **Apply** at the bottom of the **Simple Polygons** window. The map should redraw with each district labeled with district number. (Figure 30, D)



Style Properties (B): ☐ enable/disable labeling; label: dummy, - none -; opacity: 100, - none -; rotation: 0, - none -; font: set font; halo: 0

Style Properties (C): ☑ enable/disable labeling; label: DEMOREGION; opacity: 100, - none -; rotation: 0, - none -; font: set font; halo: 0

Demo_Region5.3	26	D24	-1.0	-1.0
Demo_Region5.4	27	D18	-1.0	-1.0
Demo_Region5.5	31	D14	2.0	0.5
Demo_Region5.6	35	D26	-1.0	-1.0
Demo_Region5.7	36	D15	15.0	0.00
Demo_Region5.8	37	D8	-1.0	-1.0
Demo_Region5.9	38	D21	-1.0	-1.0
Demo_Region5.10	42	D7	-1.0	-1.0
Demo_Region5.11	44	D33	-1.0	-1.0
Demo_Region5.12	46	D9	-1.0	-1.0
Demo_Region5.13	47	D23	35.0	0.0

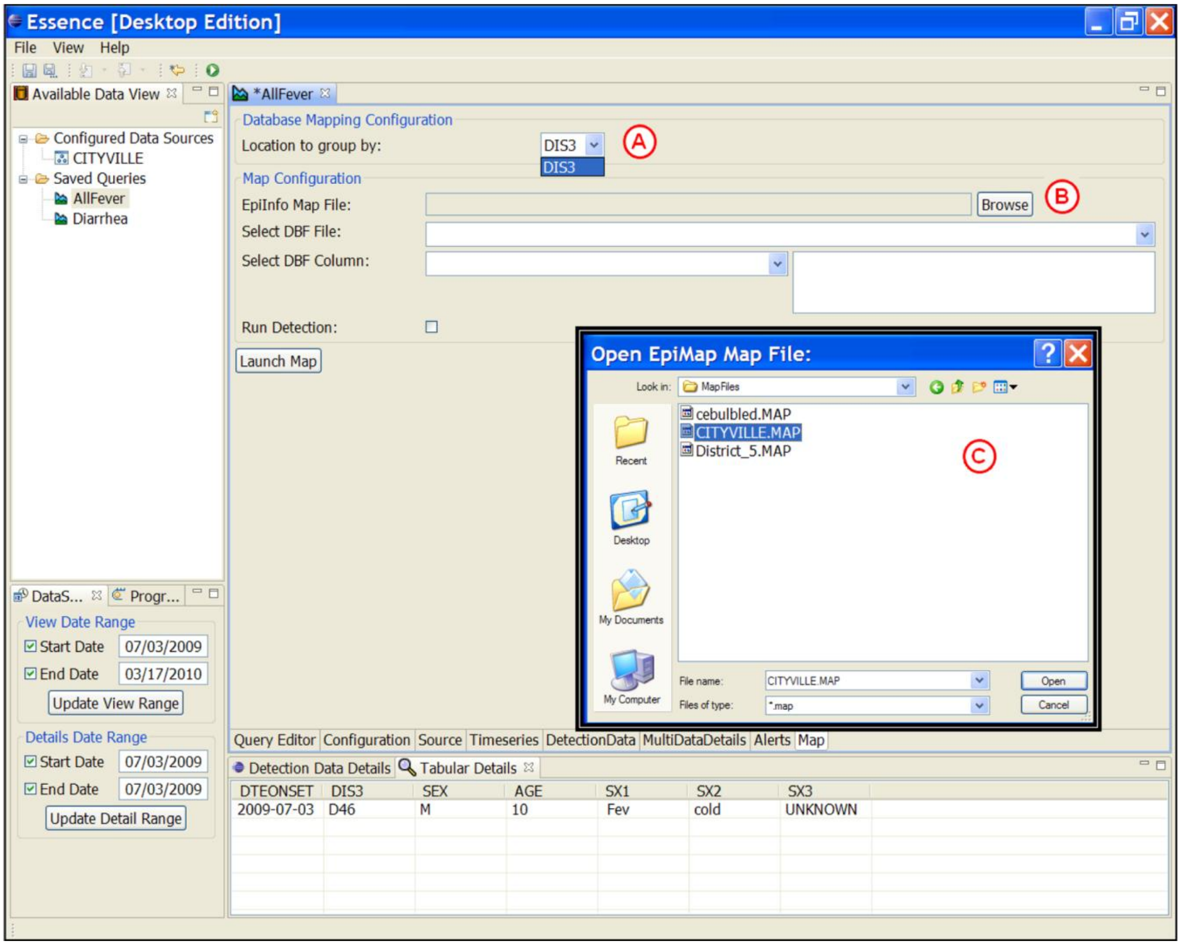
-1	-1	-1
-1	-1	-1
-1	-1	-1
-1	-1	-1
-1	-1	-1
07/03/2009	07/21/2009	06/05/2011 09:54:...

Coordinate Reference System of data is unknown. Unexpected behaviour may result if it is not set

Figure 30 UDig: Adding Labels to the Alert Map

4.2.12.1.5EpiMap: Creating a '.map' File

To map the data, first open EpiMap and create a map file for the geographic region interest. Return to EDE, and select the **Location** variable (Figure 31, A) to be used. A .map file (Cityville1.map) is included with the other shape files for the Cityville example. The location variable in the example is DIS3. Next, specify the location of the map file created in EpiMap in the **EpiInfo Map File** box under **Map Configuration** on the **Map** tab. (Figure 31, B) The **Browse** button next to the box that opens a pop-up window that allows the user to search for the folder in which the map file is located (Figure 31, C) EDE assumes the map files are in the EDE subdirectory (C:\EDE\workspace\workspace\MapFiles\). They can, however, be placed elsewhere as long as the location is specified on the **Map** tab. It is recommended that a copy of shape files, not the original file, be used with EDE because it modifies the .dbf shape files during processing. Copy the shape files to C:\EDE\WORKSPACE\WORKSPACE\MAPFILES.



Location to group by: DIS3

EpiInfo Map File:

Select DBF File:

Select DBF Column:

Run Detection: ☐

Look in: MapFiles

File name: CITYVILLE.MAP

Files of type: *.map

View Date Range: ☑ Start Date 07/03/2009 ☑ End Date 03/17/2010

Details Date Range: ☑ Start Date 07/03/2009 ☑ End Date 07/03/2009

DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2009-07-03	D46	M	10	Fev	cold	UNKNOWN

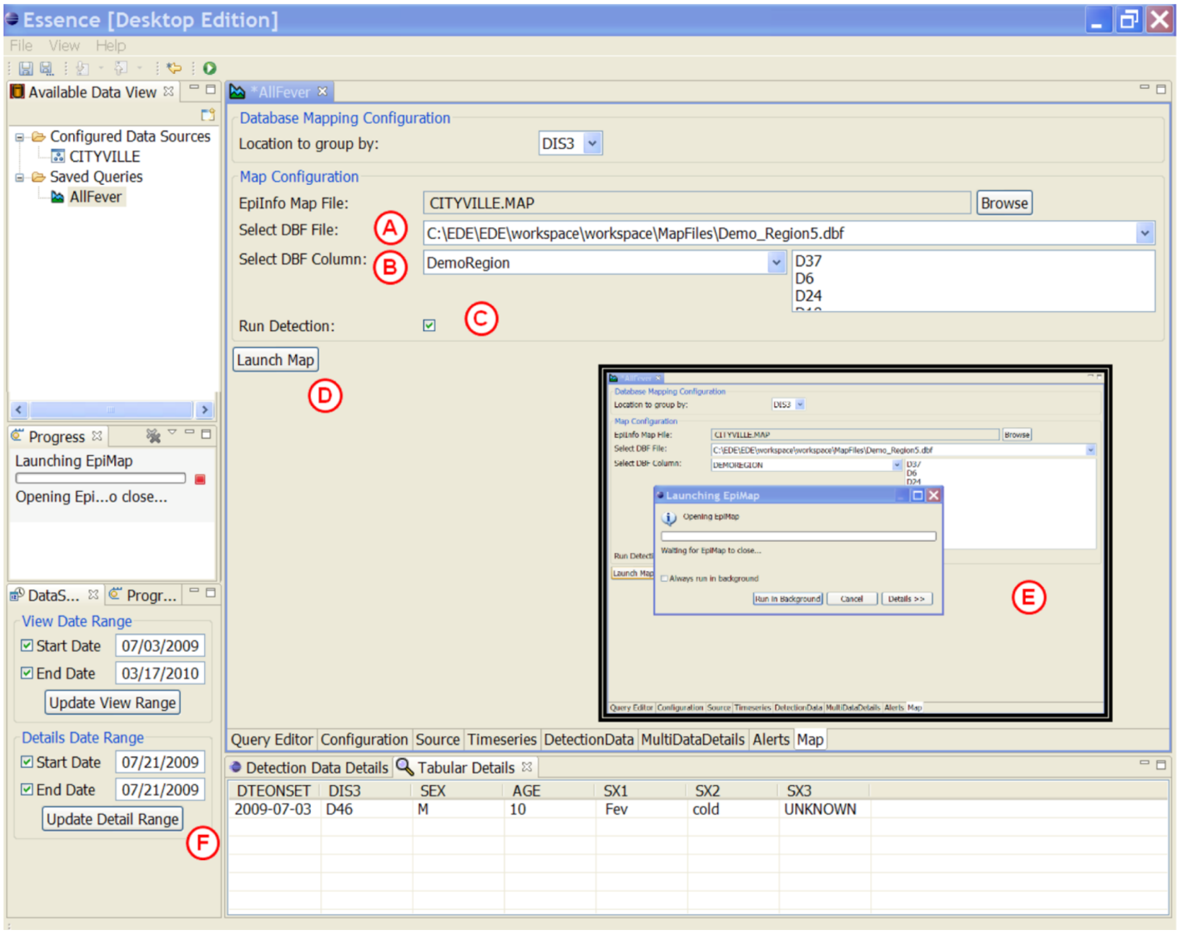
Figure 31 The Map Tab – Location Variable and Locating the Map File

4.2.12.2 Select DBF File

Selecting the file location automatically fills the next box, **Select DBF File**. (Figure 32, A) In the pull-down box next to **Select DBF Column**, pick the variable that contains the region names for the location variable that will be used for detection. (Figure 32, B) In the Cityville example, select DEMOREGION. The values of DEMOREGION appear in the box on the far right. Decide if detection should be run on the various levels of DEMOREGION. If so, then click the checkbox next to **Run Detection** under **Map Configuration** on the **Map** tab. In this Cityville example, **Run Detection** was selected. (Figure 32, C)

REMINDER

> The dates for which the map and alerts are created correspond with the Start and End dates in the **Details Date Range** box.



Location to group by: DIS3

EpiInfo Map File: CITYVILLE.MAP

Select DBF File: C:\EDE\EDE\workspace\workspace\MapFiles\Demo_Region5.dbf

Select DBF Column: DemoRegion (D37, D6, D24)

Run Detection: ☑

View Date Range: ☑ Start Date 07/03/2009 ☑ End Date 03/17/2010

Details Date Range: ☑ Start Date 07/21/2009 ☑ End Date 07/21/2009

DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2009-07-03	D46	M	10	Fev	cold	UNKNOWN

Figure 32 The Map Tab – DBF Files and Column and Launching the Map

In the Cityville example, the alerts and map will be created for a single day, 7/21/2009. (Figure 32, F) Maps can be created for multiple day periods by changing the Start and End dates in the **Details Date Range** box. However, the detection algorithms run by comparing the number of observed cases on a given day with the number expected in a set preceding time period. This means that detection can only be run on a single day. If multiple days are selected in the **Details Date Range** box, detection will be run and presented in the map only for the End date of the period. If any of the detection variables are mapped, the alerts mapped will be for the End date only. However, if a case (spot) map is created, all of the cases in the multiple day time period will be mapped.

4.2.12.3 Launch the Map

Once the appropriate date range is selected, click the **Launch Map** button (Figure 32, D) to create the map. A pop-up window displaying the status of the map generation appears in the **Query** window when the launch button is pressed. (Figure 32, E) That is the file Cityville.map in the current example. EpiMap will open automatically and display EpiInfo Map File specified. Note, the **Launching EpiMap** pop-up window will remain open in EDE and appear to be launching EpiMap until EpiMap is closed.

Reminder

> The map displays data from the time period selected under **Details Date Range** in the **DataSet Manager** window. If the Details Start Date is earlier than the first record in the DSDC, the map will fail. Correct the Start Date to run the map.

EDE calculates several variables for each location value and adds them to the .DBF file associated with the map file being used. The variables EDESTART and EDEEND contain the start and end dates listed in the **Details Date Range** box. EDECOUNT is the number of cases observed in the region during the specified time period. EDECOLOR is a string variable coding the alert level for each district, GREEN (no alert), YELLOW, or RED. EDELEVEL is a numeric variable that codes the alert level, GREEN=0, YELLOW=1, and RED=2. Finally, EDEDETEC lists the probability (p-value) that the observed number of cases for the district was no greater than expected.

The initial map for the Cityville AllFever example is displayed in Figure 33. The only saved property on the map is DEMOREGION, so the district names are the only items that appear on the map when it is opened.

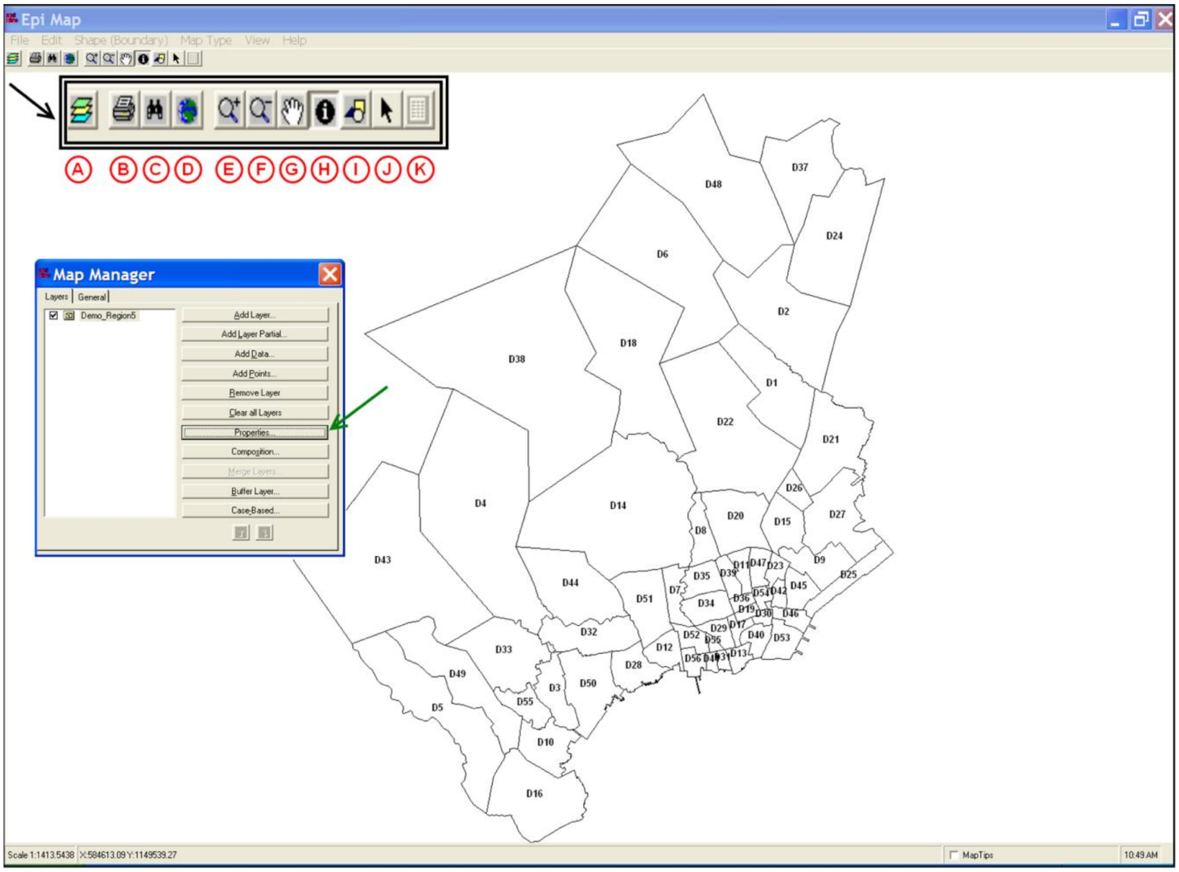


Figure 33 EpiMap – Initial Cityville Map Screen

4.2.12.4 Some EpiMap Features

To facilitate map creation a limited number of the EpiMap features are described here. The reader is referred to the EpiMap help function for further information. The **File**,

Edit, **View**, and **Help** functions on the Windows bar at the top of the EpiMap screen are similar to those found in other Windows programs. From **File** the user can create a new map file from saved shape files, save the current file, export the current map as a bitmap, and print the screen. **Edit** offers only one choice, to copy the file to the clipboard as a bit map. **View** allows some limited manipulation of the map. Titles can be added and formatted here, the cursor can be modified, and the user can select to zoom in or out on the map or pan across it. **Help** (<F1>) opens a help window for EpiMap. Please see the **Help** feature in EpiMap for further explanations of these functions.

The buttons in the second row (toolbar) at the top of the EpiMap screen are used to modify the map file. Each function on the toolbar is described in the list below (the letters next to the button name refer to the inset in Figure 33)

A: **Map Manager** – Allows the user to change the layers and presentation included on the map.

B: **Print Map Now** – Prints the map as it appears on the screen.

C: **Find** – Lets the user search for a specific character string in the different map layers.

D: **Full Extent** – Returns the full map after zooming.

E: **Zoom In** – Allows the user to select and enlarge a specific area on the map by dragging a box around the desired area.

F: **Zoom Out** – Allows the user to return to the original map after zooming in.

G: **Pan** – Allows the user to pan across an enlarged map.

H: **Identify** – Produces a pop-up box with information on the selected region contained in the .dbf file associated with the map.

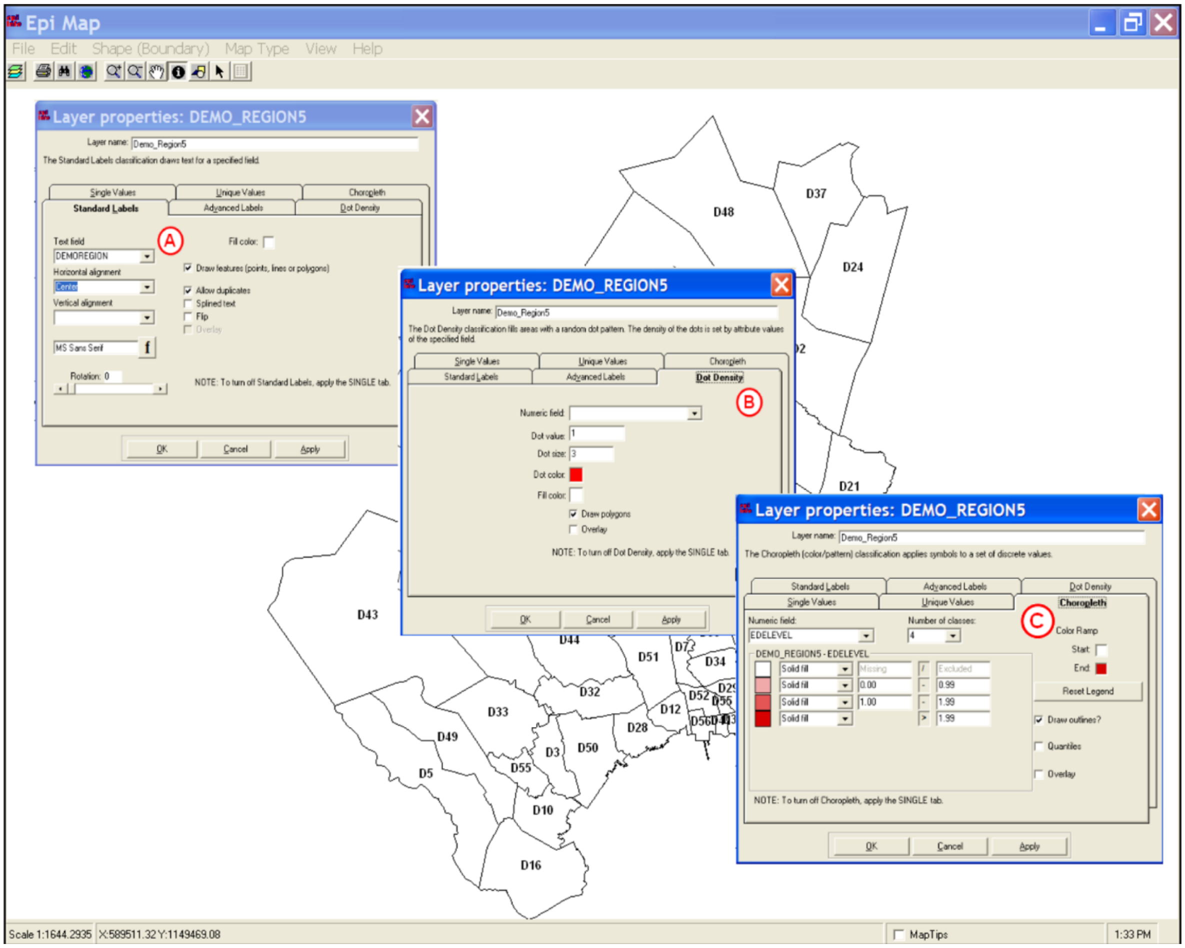
I: **Graphics** – Provides a few tools to add graphics to a map.

J: **Return Cursor to Arrow** – See EpiMap help for further information.

K: **Records for Feature** – See EpiMap help for further information.

4.2.12.5 Map Properties

To change what is displayed on the map, click on the **Map Manager** button. (Figure 33, A) This causes the **Map Manager** window to appear. (Figure 33, Inset) For simple changes, click the **<Properties…>** button in the **Layers** tab of the window. (Figure 33 Inset, Green Arrow) This causes the **Layer Properties** window to open. (Figure 26, A) This window has six tabs: **Standard Labels**, **Advanced Labels**, **Dot Density**, **Choropleth**, **Unique Values**, and **Single Values**.



Layer name: Demo_Region5

The Standard Labels classification draws text for a specified field.

Text field: DEMOREGION

Horizontal alignment: Center

Vertical alignment:

MS Sans Serif

Rotation: 0

Fill color: □

☑ Draw features (points, lines or polygons)

☑ Allow duplicates

□ Splined text

□ Flip

□ Overlay

NOTE: To turn off Standard Labels, apply the SINGLE tab.

Layer name: Demo_Region5

The Dot Density classification fills areas with a random dot pattern. The density of the dots is set by attribute values of the specified field.

Numeric field:

Dot value: 1

Dot size: 3

Dot color: 🟥

Fill color: □

☑ Draw polygons

□ Overlay

NOTE: To turn off Dot Density, apply the SINGLE tab.

Layer name: Demo_Region5

The Choropleth (color/pattern) classification applies symbols to a set of discrete values.

Numeric field: EDELEVEL

Number of classes: 4

DEMO_REGION5 - EDELEVEL

□	Solid fill	Missing	/	Excluded
🟥	Solid fill	0.00	-	0.99
🟥	Solid fill	1.00	-	1.99
🟥	Solid fill		>	1.99

Color Ramp

Start: □

End: 🟥

Reset Legend

☑ Draw outlines?

□ Quantiles

□ Overlay

NOTE: To turn off Choropleth, apply the SINGLE tab.

Scale 1:1644.2935 X:589511.32 Y:1149469.08

Figure 34 EpiMap Showing Different Tabs in the Layer Properties Window

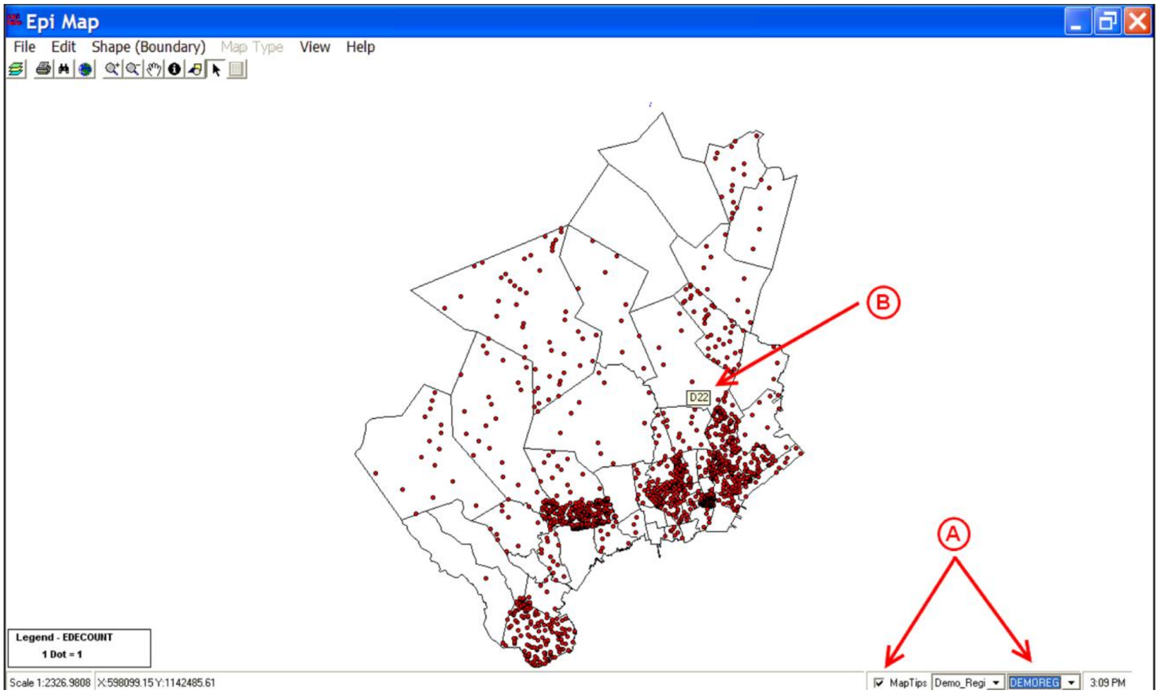
4.2.12.5.1Standard Labels Tab

The **Standard Labels** tab usually opens first. Under the **Text Field** box on that tab, a pull-down button lists the variables in the .dbf file associated with the map. (Figure 34, A) For labels, select the variable that contains the region name values. In the Cityville example, the variable is DEMOREGION. The placement and font of the labels can also be modified in this tab. The **Advanced Labels** tab provides a few more options.

4.2.12.5.2Dot Density Tab

The **Dot Density** tab creates a spot map. It places a dot in the appropriate region for each case/record. The cases are plotted randomly within the region. This can be misleading because most would assume that the cases are plotted at a specific address. Still, it provides a spot map at the region level which can be informative. To create the spot map, first select the **Numeric Field** variable from the pull-down list. (Figure 34, B) The value to use for this will always be EDECOUNT [the number of cases observed on the day(s) being plotted]. Dot Value refers to the number of cases plotted per dot. The default is one (1). Dot size and color and background region color can also be modified. Once selections are made, click **Apply** or **OK** at the bottom of the **Layer Properties** window to plot the cases. Note that when the spots appear on the map, the district names disappear. (Figure 34) However, another feature in EpiMap allows the user to select a variable and have its values appear when the cursor is placed on the region. On the bottom-right of the window is a box labeled **Map Tips**.

(Figure 35, A) Clicking on this box causes two pop-up windows to be displayed. The first pop-up lists the layers of the map and allows the user to select which layer to use for the **Map Tips**. The second pull-down lists the variables whose value can be displayed on the map when the cursor is placed over a specific region. In the Cityville example in Figure 35, the variable DIS3 was selected so the specific districts can be easily identified. (Figure 35, B)



Legend - EDECOUNT
1 Dot = 1

Scale 1:2326.9808 X:598099.15 Y:1142485.61

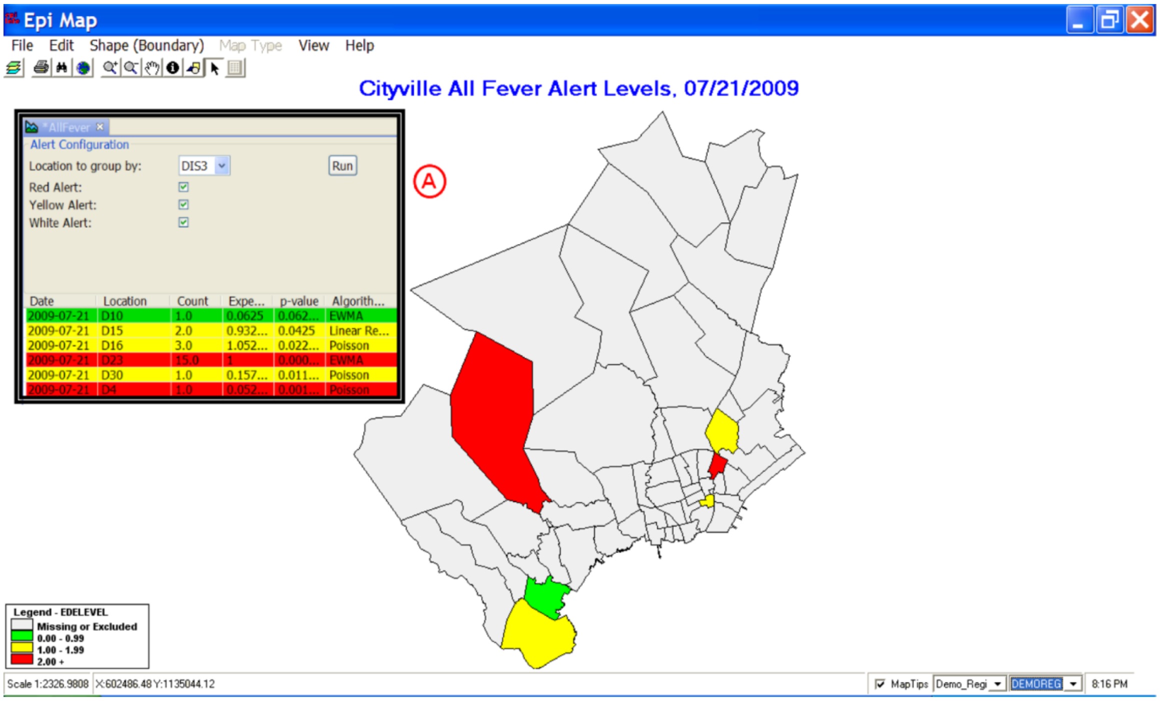
☑ MapTips Demo_Regi DEMOREG 3:09 PM

Figure 35 All Fever Spot Map of Cityville

4.2.12.5.3Choropleth Tab

Finally, the **Choropleth** tab allows users to categorize a variable and plot the regions by category color. As above, first select the variable to be plotted from the pull-down in the **Numeric Field** box. (Figure 34, C) Three variables can be used here: EDECOUNT, EDEDETEC and EDELEVEL. EDECOUNT and EDEDETEC are continuous variables; EDELEVEL has three discrete values. EDELEVEL, with only three values, is easiest to plot and conveys the same information as EDEDETEC categorizing and plotting EDEDETEC. Pick EDELEVL in the **Numeric Field** pull-down. EDELEVEL equals zero if there is no alert, 1 if there is a Yellow alert in the region, and 2 if there is a Red alert. Including a category for missing/no information, there are four possible values for the variable, so select 4 in the pull-down menu for the **Number of Classes. (**Figure 34, C) Next, modify the values for each level to match the values of EDELEVEL. This can be done by hand or the program can automatically select a color gradient when the **Reset Legend** button is clicked. (Figure 34, C)

With a discrete variable like EDELEVEL, it is generally better to select the values manually. In the Cityville example, gray was picked for regions with no information, green for those without an alert, yellow for Yellow alerts, and red for Red alerts. (Figure 34, C) The final map of EDELEVEL is shown in Figure 36. The inset (A) in Figure 36 shows a section of the **Alerts** tab for the map displayed, which shows the five districts that had data available to be analyzed.



Alert Configuration

Location to group by: DIS3

Red Alert: ☑

Yellow Alert: ☑

White Alert: ☑

Date	Location	Count	Expe...	p-value	Algorith...
2009-07-21	D10	1.0	0.0625	0.062...	EWMA
2009-07-21	D15	2.0	0.932...	0.0425	Linear Re...
2009-07-21	D16	3.0	1.052...	0.022...	Poisson
2009-07-21	D23	15.0	1	0.000...	EWMA
2009-07-21	D30	1.0	0.157...	0.011...	Poisson
2009-07-21	D4	1.0	0.052...	0.001...	Poisson

Scale 1:2326.9808 X:602486.48 Y:1135044.12 ☑ MapTips Demo_Regi DEMOREG 8:16 PM

Figure 36 EDELEVELS for All Fever Query for 7/21/2009

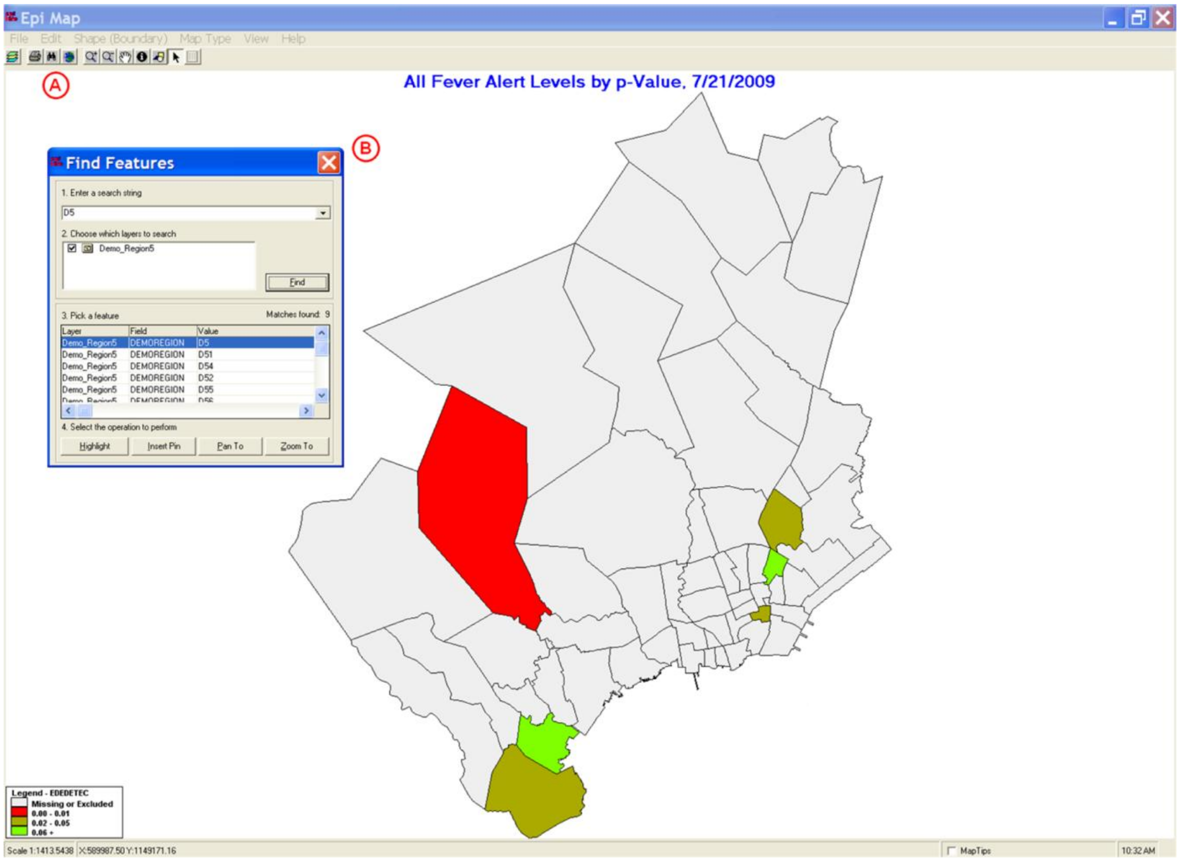
4.2.12.5.4 Title Placement

A title can be added to a map. Make sure the **Map Manager** window is closed and place the cursor in the top center of the map page. A vertical cursor will appear. Type the title, only a single line is allowed, and click elsewhere on the map. To customize the title, click **View** and then **Title Properties** to find the title font properties. Try this for the Cityville example. Add the title "Cityville All Fever Alert Levels, 07/21/2009."

4.2.12.5.5Other EpiMap Features

Clicking the **Magnify** button (Figure 33, E) on the toolbar allows the user to zoom in on a specific area. To do so, click on the **Magnify** button and drag open a box over the selected area. To return the map to the original version, click on the **Full Extent** button on the toolbar. (Figure 33, D)

The **Find** Button on the toolbar can be used to locate specific geographic areas on the map. Clicking on the **Find** button (Figure 37, A) opens the **Find Features** window. (Figure 37, B) A full or partial name of a place can be typed in the top box; the layer in which the feature resides is selected when the **Find** button is pressed. After that, if the **Highlight** button at the bottom left of the window is clicked, the place selected on the map will flash blue. In Figure 37, D5 (District 5) was selected, but notice how all areas beginning with the phrase "D5" were included. (Figure 37, B)



All Fever Alert Levels by p-Value, 7/21/2009

1. Enter a search string

D5

2. Choose which layers to search

☑ Demo_Region5

3. Pick a feature

Matches found: 9

Layer	Field	Value
Demo_Region5	DEMOREGION	D5
Demo_Region5	DEMOREGION	D51
Demo_Region5	DEMOREGION	D54
Demo_Region5	DEMOREGION	D52
Demo_Region5	DEMOREGION	D55
Demo_Region5	DEMOREGION	D56

4. Select the operation to perform

Legend - EDEDETEC

⬜	Missing or Excluded
🟥	0.00 - 0.01
🟫	0.02 - 0.05
🟩	0.06 +

Scale 1:1413.5438 X:589987.50 Y:1149171.16

Figure 37 EDEDETEC Levels by p-Value

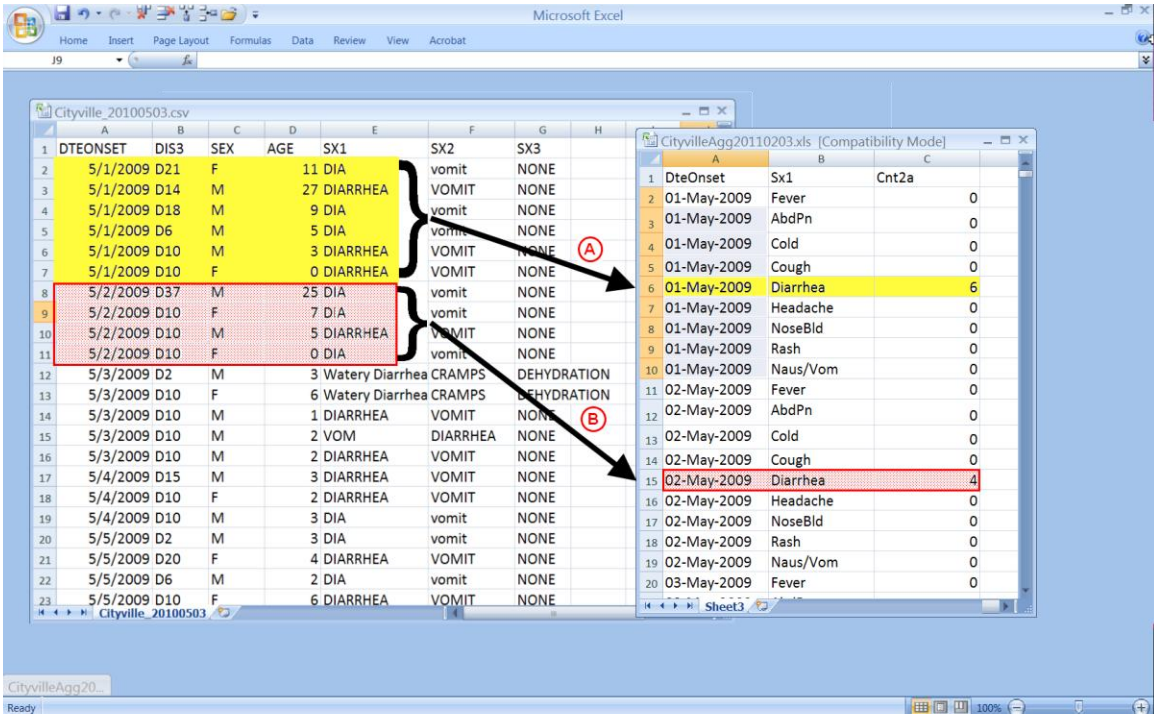
If the **Insert Pin** button is clicked, a blue cross will be placed on the map at the location selected. Similarly, clicking the **Pan To** button multiple times zooms in on the selected area in incremental steps, while the **Zoom To** button immediately focuses down on the selected area.

Refer to the EpiMap documentation for information on other mapping options.

4.3 Use Case No. 2 – FEVER IN CITYVILLE USING AGGREGATED DATA

4.3.1 Aggregated Data – Definition

Unaggregated data files contain a record for each individual person, medical visit, or event in the file. Since the records are generally for individuals they often contain demographic information such as age, sex, and address as well as disease diagnoses and/or symptoms. In contrast, aggregate data, sometimes called count data, contains a single record per day for everyone experiencing a specific disease diagnosis, symptom, or event. An aggregate record generally contains the date, the event description, and the total number of people presenting with that event on the specified date. While the aggregate data may have some demographic data, that information pertains to all patients in that count. For example, there may be a record of the number of males presenting with disease X on day Z, and a record for females presenting on the same day with the same disease.



Cityville_20100503.csv

	A	B	C	D	E	F	G
1	DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2	5/1/2009	D21	F	11	DIA	vomit	NONE
3	5/1/2009	D14	M	27	DIARRHEA	VOMIT	NONE
4	5/1/2009	D18	M	9	DIA	vomit	NONE
5	5/1/2009	D6	M	5	DIA	vomit	NONE
6	5/1/2009	D10	M	3	DIARRHEA	VOMIT	NONE
7	5/1/2009	D10	F	0	DIARRHEA	VOMIT	NONE
8	5/2/2009	D37	M	25	DIA	vomit	NONE
9	5/2/2009	D10	F	7	DIA	vomit	NONE
10	5/2/2009	D10	M	5	DIARRHEA	VOMIT	NONE
11	5/2/2009	D10	F	0	DIA	vomit	NONE
12	5/3/2009	D2	M	3	Watery Diarrhea	CRAMPS	DEHYDRATION
13	5/3/2009	D10	F	6	Watery Diarrhea	CRAMPS	DEHYDRATION
14	5/3/2009	D10	M	1	DIARRHEA	VOMIT	NONE
15	5/3/2009	D10	M	2	VOM	DIARRHEA	NONE
16	5/3/2009	D10	M	2	DIARRHEA	VOMIT	NONE
17	5/4/2009	D15	M	3	DIARRHEA	VOMIT	NONE
18	5/4/2009	D10	F	2	DIARRHEA	VOMIT	NONE
19	5/4/2009	D10	M	3	DIA	vomit	NONE
20	5/5/2009	D2	M	3	DIA	vomit	NONE
21	5/5/2009	D20	F	4	DIARRHEA	VOMIT	NONE
22	5/5/2009	D6	M	2	DIA	vomit	NONE
23	5/5/2009	D10	F	6	DIARRHEA	VOMIT	NONE

Cityville_20100503

CityvilleAgg20110203.xls [Compatibility Mode]

	A	B	C
1	DteOnset	Sx1	Cnt2a
2	01-May-2009	Fever	0
3	01-May-2009	AbdPn	0
4	01-May-2009	Cold	0
5	01-May-2009	Cough	0
6	01-May-2009	Diarrhea	6
7	01-May-2009	Headache	0
8	01-May-2009	NoseBld	0
9	01-May-2009	Rash	0
10	01-May-2009	Naus/Vom	0
11	02-May-2009	Fever	0
12	02-May-2009	AbdPn	0
13	02-May-2009	Cold	0
14	02-May-2009	Cough	0
15	02-May-2009	Diarrhea	4
16	02-May-2009	Headache	0
17	02-May-2009	NoseBld	0
18	02-May-2009	Rash	0
19	02-May-2009	Naus/Vom	0
20	03-May-2009	Fever	0

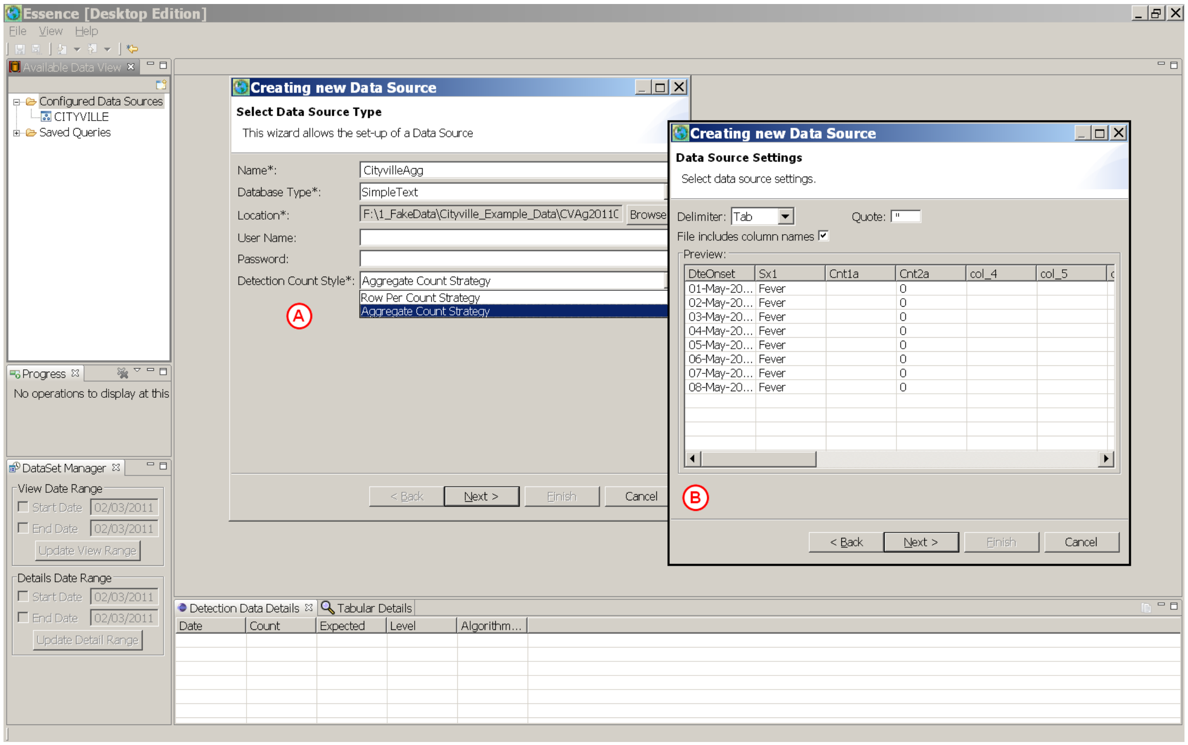
Sheet3

Figure 38 Comparison of Unaggregated and Aggregated Cityville Data Files

Figure 38 compares the unaggregated (on the left) and aggregated (on the right) data files for the Cityville data. The six lines highlighted in yellow in the unaggregated data file represent six people presenting with diarrhea listed as their primary symptom on 5/1/2009. In the aggregated file these same people are represented as a single record (Figure 38, A) that contains the date, the symptom (diarrhea) and the number of people presenting with diarrhea on that day. Similarly, the areas highlighted in red compare records for people presenting with diarrhea on 5/2/2009 in the unaggregated and aggregated format.

4.3.2 Creating an EDE Datasource from Aggregated Data

As described in the previous section, before creating a query, data must be read into EDE and configured for the application. Selection of the datasource for an aggregated data file is very similar to that described for an unaggregated file. Right-click on **Configured Data Sources** in the **Available Data View**, then left-click on the pop-up window that appears to open the **Create a New Data Source** window. Select a DSDC **Name**, **Database Type**, **Location**, and, if needed, **User Name**, and **Password** (Figure 31, A) as described for an unaggregated data file. The figures in this section show creation and use of the aggregated Cityville data, CityvilleAgg20110203.xls. Following the steps below will create the same files and queries described here. The **Detection Count Style** is different for aggregated files. The second choice in this box is highlighted for this option, i.e. **Aggregate Count Strategy**. After highlighting that choice click on the **Next>** button to continue.



Creating new Data Source

Select Data Source Type

This wizard allows the set-up of a Data Source

Name*: CityvilleAgg
Database Type*: SimpleText
Location*: F:\1_FakeData\Cityville_Example_Data\CVAg2011C
User Name:
Password:
Detection Count Style*: Aggregate Count Strategy
Row Per Count Strategy
Aggregate Count Strategy

Creating new Data Source

Data Source Settings

Select data source settings.

Delimiter: Tab   Quote: "
File includes column names ☑

Preview:

DteOnset	Sx1	Cnt1a	Cnt2a	col_4	col_5
01-May-20...	Fever		0		
02-May-20...	Fever		0		
03-May-20...	Fever		0		
04-May-20...	Fever		0		
05-May-20...	Fever		0		
06-May-20...	Fever		0		
07-May-20...	Fever		0		
08-May-20...	Fever		0		

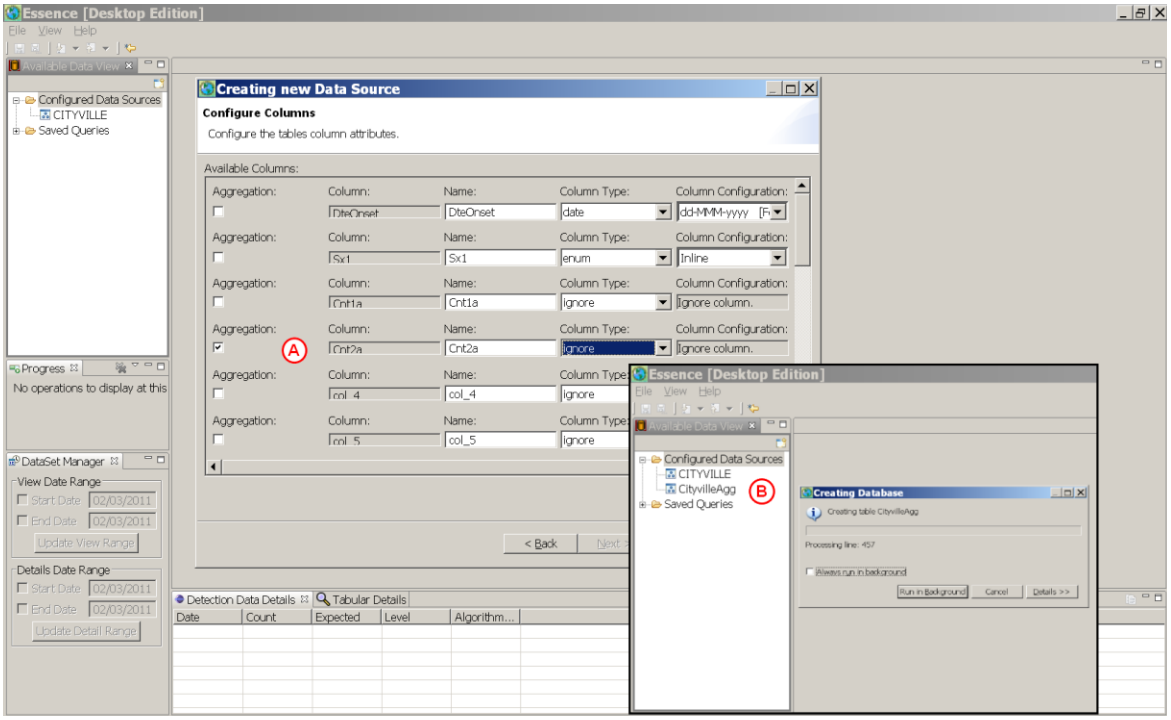
Figure 39 Create New EDE Datasource from an Aggregated Data File

4.3.2.1 Creating an EDE Datasource – Data Source Settings

The **Data Source Settings** screen appears next. (Figure 39, B) The options for this screen are identical to those for an unaggregated data file. Clicking on the **Next>** button when the options have been selected opens **Configure Columns** screen.

4.3.2.2 Creating an EDE Datasource – Configure Columns

The **Configure Columns** screen opens small, so it may be useful to stretch this box to the left so that all of the columns are visible. This screen has the same variable options described before: **Column**, **Name**, **Column Type**, and **Column Configuration**. In addition, the screen now has a column labeled, **Aggregation**. (Figure 40, A) Check the box in this column for the 'count' variable in the data file, i.e. the variable whose value is the number of people presenting with a specific symptom on a specific day. Simply click the **Aggregation** check box for the count variable, and change the DSDC variable name if desired. The other columns, **Column Type** and **Column Configuration**, do not need to be modified for the 'count' variable. When the **Column Configuration** screen is completed, click **Finish**, and EDE will create the DSDC for the specified file, and add the datasource name to **Configured Data Sources** in the **Available Data View** window. (Figure 40, B)



Creating new Data Source

Configure Columns

Configure the tables column attributes.

Available Columns:

Aggregation:	Column:	Name:	Column Type:	Column Configuration:
☐	DteOnset	DteOnset	date	dd-MMM-yyyy [F
☐	Sx1	Sx1	enum	Inline
☐	Cnt1a	Cnt1a	ignore	Ignore column.
☑	Cnt2a	Cnt2a	ignore	Ignore column.
☐	col_4	col_4	ignore	
☐	col_5	col_5	ignore	

Creating Database

Creating table CityvilleAgg

Processing line: 457

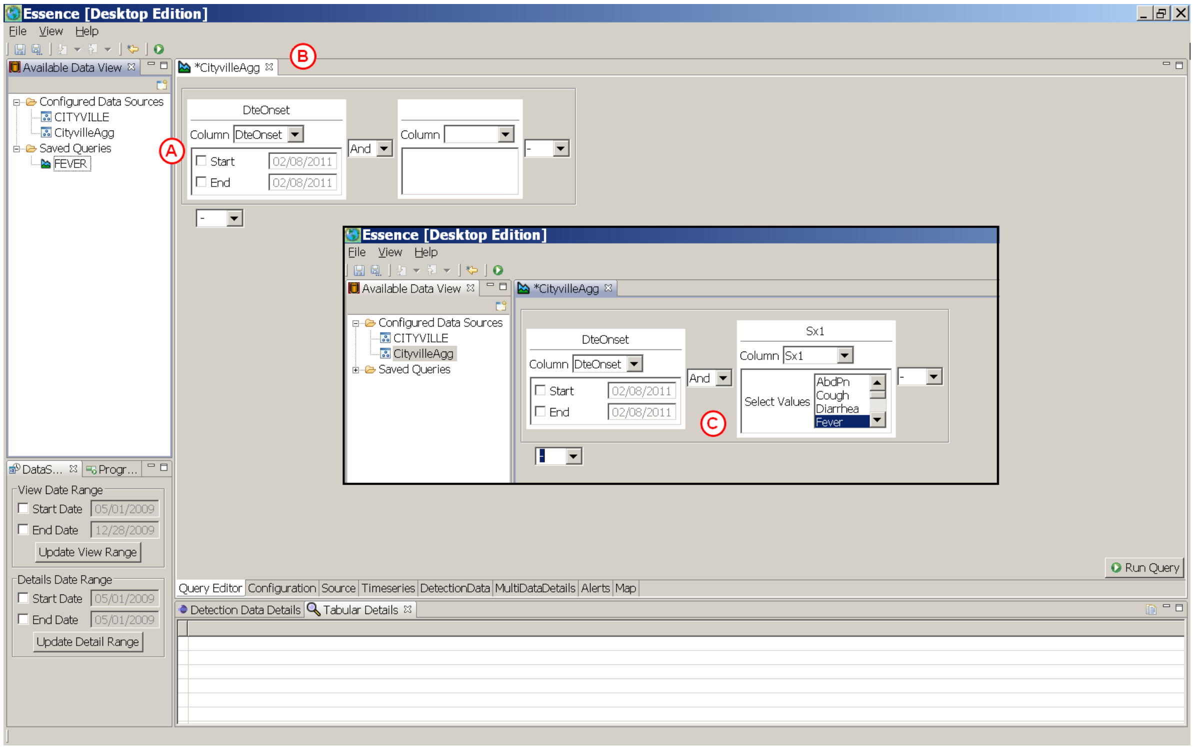
☐ Always run in background

Figure 40 Configuring Columns for an New EDE DSDC from an Aggregated Data File

Once the DSDC is created it is treated like any other datasource. It is saved the same way and queries are created and run the same way.

4.3.3 Creating a Query

To create a fever query for the aggregate Cityville datasource, right-click on the name of the saved datasource (i.e., CityvilleAgg) under **Configured Data Sources** in the **Available Data View** window. A pop-up window appears. (Figure 41, A) Click on the first choice, **New Query**. A tabbed box, labeled at the top with datasource name, ‘CityvilleAgg’ in this example, appears in the query window. (Figure 41, B) A variable definition box will also appear in the query window. (Figure 41, C) For this example, the query will produce a time series of fever data. So, select DTEONSET as the first variable (without specifying dates) and join it to the variable Sx1, value equal ’Fever’ with an **And**. Note that in the aggregated data there is a single value for fever in SX1. Click **Run Query**. EDE will run the detection algorithm on the selected data and display the results in the tabbed boxes in the Query window. (Figure 42)



DteOnset
Column DteOnset
□ Start 02/08/2011
□ End 02/08/2011
And

Sx1
Column Sx1
Select Values AbdPn, Cough, Diarrhea, Fever

View Date Range
□ Start Date 05/01/2009
□ End Date 12/28/2009
Update View Range

Details Date Range
□ Start Date 05/01/2009
□ End Date 05/01/2009
Update Detail Range

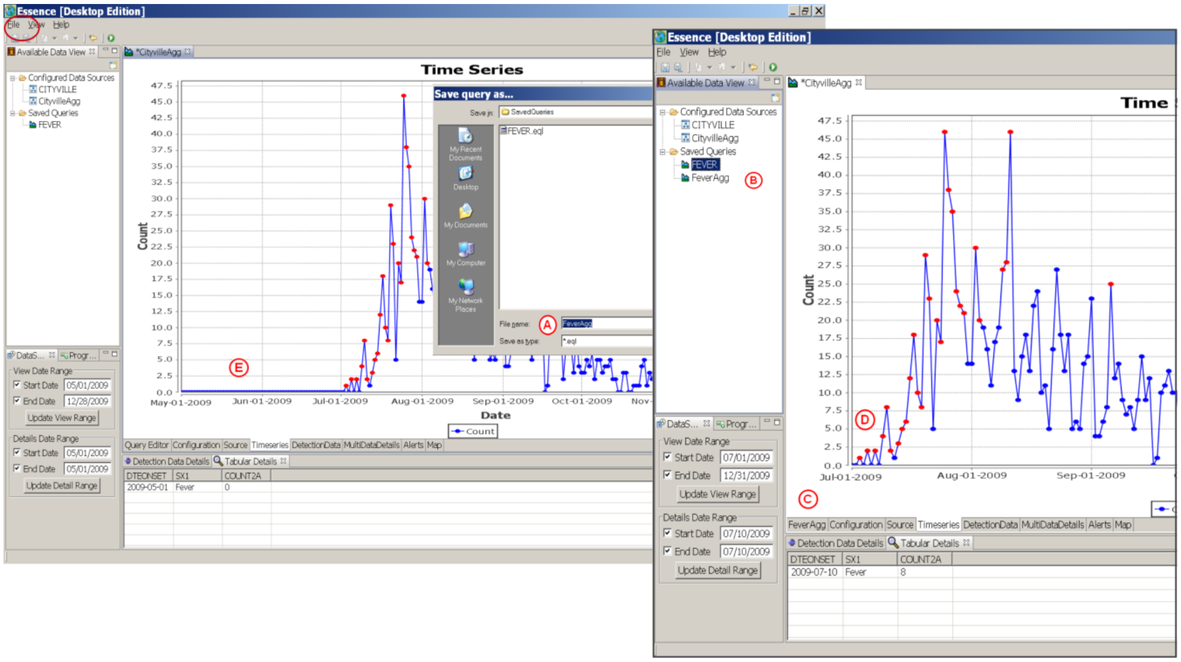
Figure 41 Creating a New Query, the Second Variable

4.3.4 Saving a Query

Queries may be saved for later use as previously described. For this example, click the **Save As** icon on the toolbar, or click **File** on the toolbar and then **Save As** in the pop-up window. Type the query name in the **File Name** box and click **Save**. In Figure 42, A, the query is saved as 'FeverAgg'. Once the query is saved, the name will appear under **Saved Queries** in the **Available Data View Window** (Figure 42, B) and replace the phrase 'Query Editor' on the Query Editor tab at the bottom of the **Query Window**. (Figure 42, C)

The aggregate Cityville dataset is not as large as the unaggregated one. The aggregate file has data from 1 May 2009 through 31 December 2009, and the symptom Fever doesn't appear until 30 June 2009. Figure 42, E, shows the time series returned by the original query that covers the entire time period in the data set. The time series has been right truncated in the second inset in Figure 42 (D), to begin with the date 1 July 2009.

The remaining tabs in the **Query Window, Configuration, Source, DetectionData, MultiDataDetails, Alerts, and Map,** are the same as those created for the unaggregated data. The results seen in the windows will be similar or identical to the results for a similar query on the unaggregated data.



File name: FeverAgg
Save as type: *.eql

View Date Range
☑ Start Date 05/01/2009
☑ End Date 12/28/2009

Details Date Range
☑ Start Date 05/01/2009
☑ End Date 05/01/2009

DTEONSET	SX1	COUNT2A
2009-05-01	Fever	0

View Date Range
☑ Start Date 07/01/2009
☑ End Date 12/31/2009

Details Date Range
☑ Start Date 07/10/2009
☑ End Date 07/10/2009

DTEONSET	SX1	COUNT2A
2009-07-10	Fever	8

Figure 42 Saving a Query

4.4 USE CASE NO. 3 – COMPARING FEVER AND DIARRHEA QUERIES

4.4.1 Opening Multiple Query Windows at the Same Time

Query windows for different queries may be open at the same time. A simple way to create a new query is to open a saved query similar to what is needed, modify it, and save it with the new parameters. In Figure 43, a second query window has been opened for the **AllFever** query by right-clicking **AllFever** under **Saved Queries** in the **Available Data View** window and selecting **Run Query**. When multiple windows are open, whether query windows or some other type of window, the active window is the one highlighted in blue. (Figure 43, B) In Figure 43, the second of the two **AllFever** query tabs is active.

When multiple query windows are open, the other tabs within the query window refer specifically to that query. However, the **Detection Data** and **Tabular Details** windows correspond to whichever window is active.

4.4.2 Creating a Second Query Window

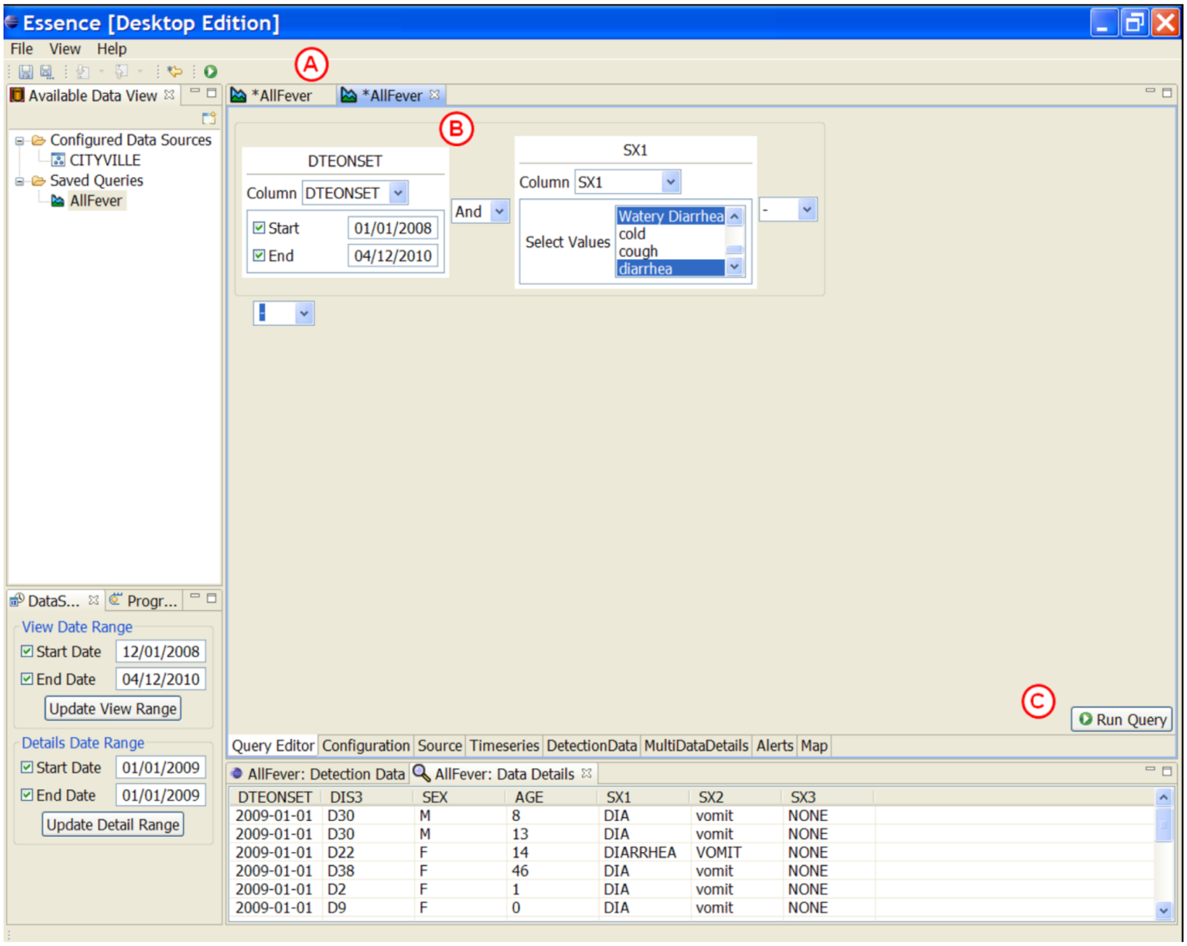
In Figure 43, the **AllFever** query was changed to a query for all patients with diarrhea. This was done by changing the selected values for SX1. All five permutations of DIARRHEA were selected for the new query. Then the query was executed by clicking the **Run Query** button (Figure 43, C) at the bottom of the **Query Editor** display.

The second **AllFever** window now shows a Time Series for diarrhea. (Figure 44, A) The **Detection Data** and **Tabular Details** windows correspond to this window because it is active. This can be checked by looking at the SX1 column under **Tabular Details**.

Note that the first symptom listed for all cases/records is now diarrhea, (Figure 43, B), but the query is still named **AllFever**. (Figure 44, D) To save it with a more appropriate name, click the **Save As** icon on the toolbar (Figure 44, C) and save the query with a new name. (Figure 44, Inset E)

CAUTION

> Do not use the **Save** button, or you will overwrite the original **AllFever** query with the new **Diarrhea** query.



DTEONSET
Column DTEONSET
☑ Start 01/01/2008
☑ End 04/12/2010

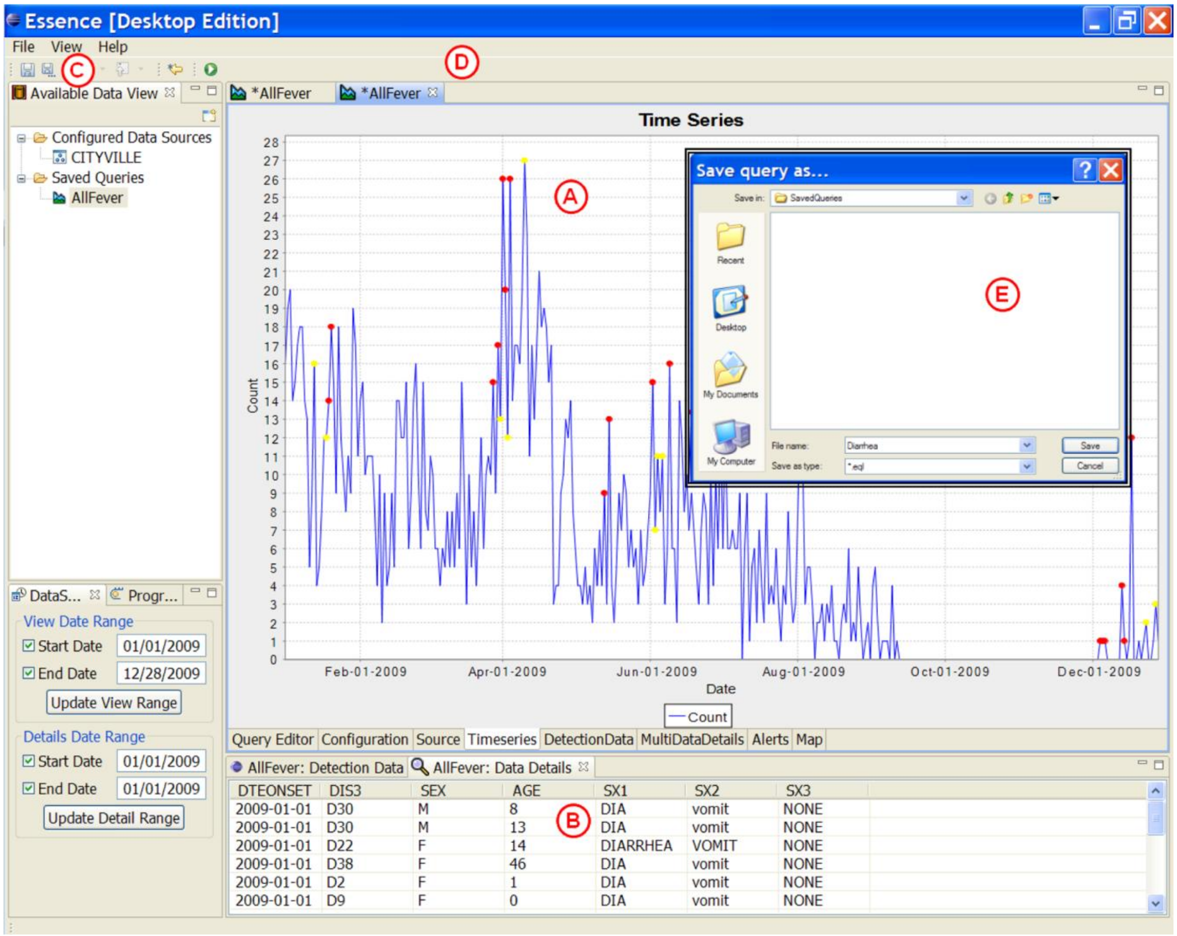
SX1
Column SX1
Select Values: Watery Diarrhea, cold, cough, diarrhea

View Date Range
☑ Start Date 12/01/2008
☑ End Date 04/12/2010

Details Date Range
☑ Start Date 01/01/2009
☑ End Date 01/01/2009

DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2009-01-01	D30	M	8	DIA	vomit	NONE
2009-01-01	D30	M	13	DIA	vomit	NONE
2009-01-01	D22	F	14	DIARRHEA	VOMIT	NONE
2009-01-01	D38	F	46	DIA	vomit	NONE
2009-01-01	D2	F	1	DIA	vomit	NONE
2009-01-01	D9	F	0	DIA	vomit	NONE

Figure 43 Opening a Second Query Window Using Saved Query as a Template



Save in: SavedQueries

File name: Diarrhea

Save as type: *.eql

View Date Range

☑ Start Date 01/01/2009

☑ End Date 12/28/2009

Details Date Range

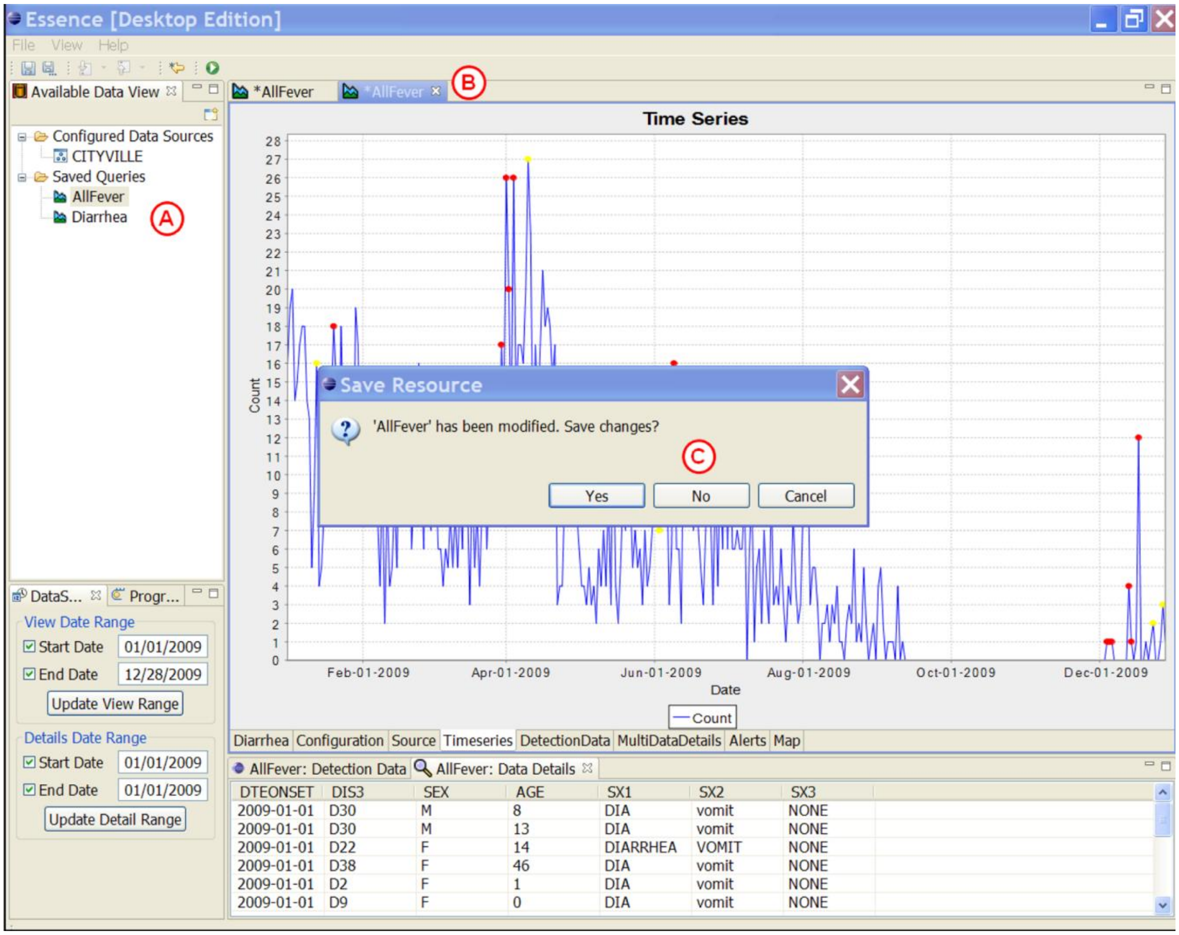
☑ Start Date 01/01/2009

☑ End Date 01/01/2009

DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2009-01-01	D30	M	8	DIA	vomit	NONE
2009-01-01	D30	M	13	DIA	vomit	NONE
2009-01-01	D22	F	14	DIARRHEA	VOMIT	NONE
2009-01-01	D38	F	46	DIA	vomit	NONE
2009-01-01	D2	F	1	DIA	vomit	NONE
2009-01-01	D9	F	0	DIA	vomit	NONE

Figure 44 Saving a Modified Query

Once the **Diarrhea** query is saved, it will appear under the list of **Saved Queries** in the **Available Data View** window. (Figure 45, A) However, the name on the **Query** window tab will not change until the new, saved query is opened. (Figure 45, B) To avoid confusion, it is best to shut the second AllFever (i.e., the **Diarrhea** query) by clicking on the X in **Query** window **Name** tab. (Figure 45, B) A pop-up window will appear asking if you want to save the AllFever window that has been modified. (Figure 45, C) Click **No**, and run the newly saved Diarrhea query again to open it.



'AllFever' has been modified. Save changes?

View Date Range

☑ Start Date 01/01/2009

☑ End Date 12/28/2009

Details Date Range

☑ Start Date 01/01/2009

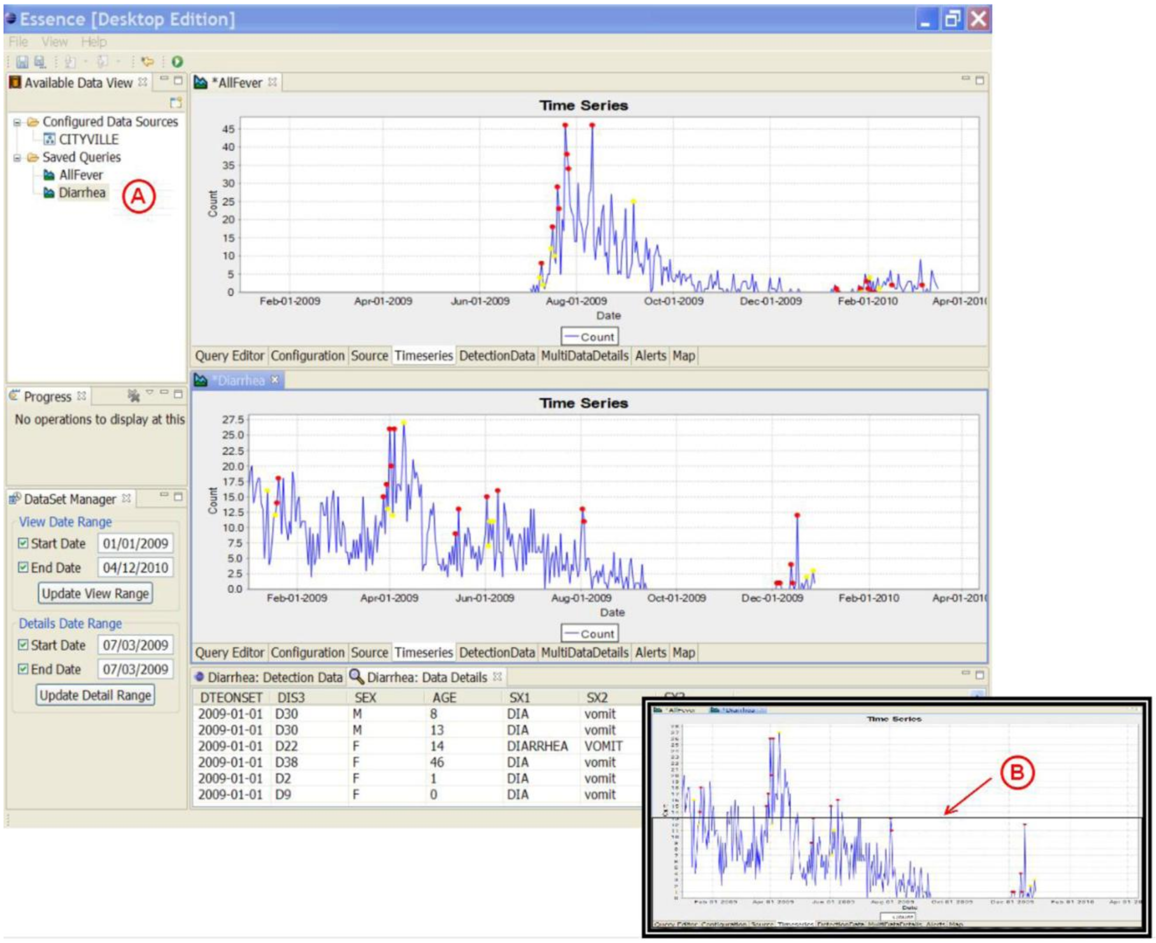
☑ End Date 01/01/2009

DTEONSET	DIS3	SEX	AGE	SX1	SX2	SX3
2009-01-01	D30	M	8	DIA	vomit	NONE
2009-01-01	D30	M	13	DIA	vomit	NONE
2009-01-01	D22	F	14	DIARRHEA	VOMIT	NONE
2009-01-01	D38	F	46	DIA	vomit	NONE
2009-01-01	D2	F	1	DIA	vomit	NONE
2009-01-01	D9	F	0	DIA	vomit	NONE

Figure 45 Opening the New Diarrhea Query

4.4.3 Dragging the Second Query Window

Once the two queries are run, one of the two queries can be dragged so that the two Query windows appear together, one above the other. (Figure 46) This allows easy comparison of the distribution over time of the two symptoms being examined.



View Date Range

☑ Start Date 01/01/2009

☑ End Date 04/12/2010

Update View Range

Details Date Range

☑ Start Date 07/03/2009

☑ End Date 07/03/2009

Update Detail Range

DTEONSET	DIS3	SEX	AGE	SX1	SX2
2009-01-01	D30	M	8	DIA	vomit
2009-01-01	D30	M	13	DIA	vomit
2009-01-01	D22	F	14	DIARRHEA	VOMIT
2009-01-01	D38	F	46	DIA	vomit
2009-01-01	D2	F	1	DIA	vomit
2009-01-01	D9	F	0	DIA	vomit

Figure 46 Multiple Query Windows Fever and Diarrhea

To drag the **Diarrhea** query window, click and hold the name tab at the top of the **Diarrhea** query window. While continuing to hold down the mouse button on the name tab, drag the cursor to where you want to place the window. As the window moves, an outline box appears showing the final position the query window will take. . (Figure 46, Arrow in Inset B) Note in Figure 46 the **Diarrhea** query is active (the name tab is highlighted in blue). That means that the other windows—the **DataSet Manager**, **DetectionData**, and the **Data Details**—show information about the **Diarrhea** query. If **AllFever** is highlighted, those windows show information about the **AllFever** query.

5 ABBREVIATIONS AND ACRONYMS

DSDC	Datasource, Definition and Configuration
EDE	ESSENCE Desktop Edition
ESRI	Environmental Systems Research Institute
ESSENCE	Electronic Surveillance System for the Early Notification of Community-based Epidemics
EWMA	Estimated Weighted Moving Average
SQL	Standard Query Language

APPENDIX I

ESSENCE Algorithms for Univariate Temporal Alerting

Howard Burkom, PhD

The following descriptions are written to explain the statistical alerting methods in ESSENCE that operate on single time series, i.e. the univariate temporal algorithms.

1 GENERAL CONSIDERATIONS

These methods are not intended to positively identify outbreaks without supporting evidence. Their purpose is to direct the attention of a limited monitoring staff with increasingly complex data streams to data features that merit further investigation. They have also been useful for corroboration of clinical suspicions, rumor control, tracking of known or suspected outbreaks, monitoring of special events and health effects of severe weather, and other locally important aspects of situational awareness. Successful users value these methods more for the latter purposes and do not base public health responses solely on algorithm alerts.

All of these algorithms are one-sided tests that monitor only for unusually high counts, not low ones. Low counts could result from an emergency situation because data reporting could be interrupted, but there are many more common reasons for low counts (such as unscheduled closings or system problems), so the algorithms do not test for abnormally low counts.

In addition to data- and disease-specific considerations below, algorithm selection was also driven by system considerations. Users need to monitor many types of data rapidly. External covariates such as climate data or clinic schedules are not available for prompt analysis. Many methods in the literature, armed with substantial retrospective data of a certain type, depend on analysis of substantial history. Day-to-day users, often with only a small fraction of time available for monitoring, will not wait several minutes for each query. In the absence of data history and data-specific analysis time for each stream, ESSENCE methods have been adapted from the literature and engineered to system requirements.

In addition to methods based on single time series, some ESSENCE implementations also include space-time cluster detection based on scan statistics, not discussed in this document. Appropriate use of these methods requires data records with reliable and relevant spatial information such as patient zip code of likely exposure (residence or workplace), and concepts of the clusters of practical interest and of how to investigate them. This feature is not yet available in EDE.

The default algorithm in the EDE and other ESSENCE systems is an automated selection between data modeling (adaptive multiple regression) and control-chart-based (adaptive EWMA) algorithms, resorting to a simplistic (Poisson) method if only a few days of recent data are available. The primary regression and EWMA methods are discussed first separately.

Each description below gives a method category, purposes of the method, a brief technical description, key benefits, limitations, and literature sources.

2 ALGORITHM, LINEAR REGRESSION

2.1 CATEGORIZATION

Adaptive Multiple Regression Model

2.2 PURPOSES:

This model is an adaptive regression model applied to remove the systematic behavior often seen in time series of daily, syndromic, clinical visit counts and in other surveillance data. The reason for removing these common effects is to avoid bias in identifying unusual behavior. For example, there is a customary jump in visits on Mondays because many clinics resume normal hours, and this expected jump should not automatically increase the possibility of an alarm. Similarly, alarms should be possible on weekends even though visit counts drop off from weekday levels.

2.3 TECHNICAL DETAILS:

This adaptive, multiple, least-squares regression algorithm contains terms to account for linear trends, day-of-week effects, and holidays. Multipliers for these terms are calculated using 4 weeks of recent counts as a training period. This training period is separated from the date of the test data by a 2-day buffer intended to keep early outbreak effects from contaminating the training. Extreme data values in the training period are reduced to reasonable values in order to avoid inappropriate predictions. This outlier correction for model inference avoids loss of sensitivity in the weeks after either data problems or true outbreaks.

The regression multipliers are recomputed each day for calculation of a predicted count based on the expected data trends. The algorithm then subtracts this prediction from the observed visit count, scales the excess by the standard error of regression, and applies a statistical hypothesis test to determine whether to signal an alert. The test is a Student’s t distribution at significance levels of 1% for red alerts and 5% for yellow alerts, with the number of degrees of freedom determined by the number of regression covariates and the baseline length.

2.4 BENEFITS:

The main benefit is avoiding alerting bias resulting from expected data trends. The length for the training baseline is critical. Based on performance comparisons among multiple baseline lengths, it was chosen to be short and recent enough to capture seasonal time series behavior but long enough to smooth out daily fluctuations. Separate multipliers are updated so that a datasource with regular but unusual patterns such as high weekend counts will be modeled correctly. While a better fit may often be obtained with a more complex model for a given data stream with a certain syndromic filter for a certain sub=region and analysis of sufficient data history, the current regression approach is relatively robust across recent ESSENCE time series.

2.5 LIMITATIONS

If this algorithm is applied to a data series without the baseline weekly and seasonal behavior, the model will not explain the data well, and the detection sensitivity and specificity will be decreased. The automated switch in the default method is applied for this reason. There is no claim of optimal modeling for a given time series.

2.6 SOURCES:

Brillman JC, Burr T, Forslund D, Joyce E, Picard R and Umland E. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance, BMC Medical Informatics and Decision Making 2005, 5:4, pp 1-14 http://www.biomedcentral.com/content/pdf/1472-6947-5-4.pdf.

Burkom, H.S., Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance, Johns Hopkins APL Technical Digest 24 (2007), 4: 335-342

3 ALGORITHM, EWMA (EXPONENTIALLY-WEIGHTED MOVING AVERAGE)

3.1 CATEGORIZATION

Adaptive Control Chart

3.2 PURPOSES:

This algorithm is appropriate for daily counts that do not have the characteristic features modeled in the regression algorithm. It is more applicable for Emergency Department data from certain hospital groups and for time series with small counts (daily average below 10) because of the limited case definition or chosen geographic region. \

3.3 TECHNICAL DETAILS:

This algorithm compares a weighted average of the most recent visit counts to a baseline expectation. For the weighted average to be tested, an exponential weighting gives the most influence to the most recent observations. Two weightings are applied: the first gives negligible weight to observations over 3 days old and is designed to detect sudden events where most outbreak cases affect data within a few days. The second weighting distributes influence further over the past week for sensitivity to more gradual outbreaks.

The monitored weighted averages are the Sk values given by:

$$Sk = \omega S\ k\text{-}1 + (1\text{-}\omega)\ Xk,$$

for a constant smoothing coefficient ω, with 0 < ω < 1 and Xk as the successive data counts, with X0 = 0 and S0 = half the alerting threshold for prompt sensitivity. Occasionally a useful starting value for X0 is known, but restarts may occur for many reasons, so the conservative initialization to 0 is used. For separate monitoring of sudden and gradual events, smoothing coefficients ω= 0.9 and 0.4 are used.

For both weighted averages, the 4-week baseline mean is subtracted, with a 2-day buffer period to separate the baseline from the counts being tested. The rationale for the baseline length was the same as described above for the regression method above. The test statistic is then (Sk – μk) / αk, where μk , αk are baseline mean and standard deviation. As in the regression method, the hypothesis applied to determine alerting is a Student's t distribution at significance levels of 1% for red alerts and 5% for yellow alerts. The number of degrees of freedom is the baseline length + 1.

This algorithm is designed for any series that does not fit the characteristic trends, so safeguards are included for rapid adjustment to and recovery from data dropouts and catch-ups and for avoiding excessive alerts when counts are sparse.

3.4 BENEFITS:

This method gives sensitivity to both sudden and gradual outbreaks and has demonstrated prompt alerting capability. It is less susceptible than the EARS methods C1, C2, and C3 to trends and to day-of-week effects. The added recovery features handle common problems in the data acquisition chain. Alerting is indirectly adjusted for the data distribution via the standardized residual test statistic, which provides a safeguard against excessive alerting when counts are small.

3.5 LIMITATIONS

This algorithm applied to pure daily counts does not control for expected trends or cyclic effects as in the regression method.

3.6 SOURCES

Ryan TP. Statistical Methods for Quality Improvement. New York: John Wiley & Sons: New York, 1989

EWMA-Shewhart charts in Morton AP, Whitby M, McLaws M-L, Dobson A, McElwain S, Looke D, Stackelroth J, Sartor A; The application of statistical process control charts to the detection and monitoring of hospital-acquired infections; J Qual Clin Prac 2001; 21:112-117

4 ALGORITHM, POISSON/REGRESSION/EWMA (DEFAULT)

4.1 CATEGORIZATION

Automated switch between data model and control chart

4.2 PURPOSE

Many researchers and developers have applied complex statistical models to surveillance data for prediction and detection. However, the predictive capability of a model varies according to the specific data stream and how it is filtered and aggregated. This capability may also be affected by data behavior changes that result from seasonal variations, population shifts, and changes in the informatics. To account for such day-to-day changes, ESSENCE automatically monitors its predictive capability of its regression model each day. When this test fails, indicating that the model is not helpful for explaining the data, the system switches to the EWMA adaptation described above. The result is that the regression model is usually applied for the common respiratory and gastrointestinal syndrome classifications applied to county-level data, but EWMA is more commonly applied to rare syndrome data.

For situations where less than a week of recent baseline data exists, a simple Poisson detector is applied. Such situations include new start-ups and more common restarts after long (several-week) intervals of missing data.

4.3 TECHNICAL DETAILS

Details for the separate regression and EWMA methods are given in the preceding pages. The adjusted R2 coefficient for the regression is tested each day. This coefficient does not give the quality of regression but is employed here specifically as a measure of daily predictive capability using an empirically derived threshold criterion. When the data pass this test, the model is assumed to have explanatory value, and the regression algorithm is applied. When the data fail this test, the EWMA algorithm is used.

The Poisson distribution test is applied when less than a week (3-6 days) of recent data is available. A Poisson distribution is assumed with mean and variance equal to the mean of the recent counts. An alert is issued if the current count exceeds this mean and if its probability is less than 1% (red alert) or 5% (yellow alert) according to the Poisson assumption. For additional features engineered to meet the needs and requests of epidemiologist users, see the reference below.

4.4 BENEFITS

This algorithm is the default because it is designed to avoid mismatching the method to the data. The regression model accounts for the expected data trends when they are seen in the baseline. When they are absent because of the case definition used to filter the data, because of the size of the monitored region, or because of data problems, alerting is based on the EWMA algorithm.

4.5 LIMITATIONS

The goodness-of-fit test occasionally misclassifies the data. The test is set to err toward the more conservative EWMA to avoid mis-fitting the data model.

4.6 SOURCES

Burkom HS, Elbert Y, Magruder SF, Najmi AH, Peter W, Thompson MW. Developments in the roles, features, and evaluation of alerting algorithms for disease outbreak monitoring. Johns Hopkins APL Technical Digest 2008;27:313.

5 ALGORITHMS, C1, C2, AND C3

5.1 CATEGORIZATION

Adaptive Control Chart

5.2 PURPOSE

To purpose is to detect general data aberrations. Algorithms C1, C2, and C3 of the Early Aberration Reporting System (EARS) developed at the Centers for Disease Control and Prevention are used in many U.S. states and in numerous foreign countries. They are included in the ESSENCE suite because of their wide application. While they lack many of the features described above, their simplicity has both benefits and limitations.

5.3 TECHNICAL DETAILS

The C1 algorithm subtracts the daily count from the mean of a moving baseline ending the previous day. In effect, it then divides this difference by the standard deviation of counts in that baseline. If the result exceeds 3, indicating an increase above the mean of more than 3 standard deviations, an alert is issued.

The C2 algorithm does the same calculation but imposes a 2-day buffer between the test day and the baseline.

The C3 algorithm is a more sensitive version of C2 that adds the values from the 2 previous days if they do not exceed the threshold. All three algorithms use the same criterion of an increase of at least 3 baseline standard deviations above the sliding baseline mean.

An important implementation detail is that ESSENCE does not use the standard 7-day baseline because substantial experience has shown that for many time series, such a short baseline gives an unstable statistic that can lead to a loss of confidence in the results. The implemented baseline is 28 days as in the EWMA and regression methods. There are no other changes to the standard EARS methods, including retention of the flat 3-standard-deviation threshold regardless of the data stream.

5.4 BENEFITS

The methods are easy to understand and widely known.

5.5 LIMITATIONS

Like the EWMA, the methods take no account of systematic data behavior such as day-of-week effects or seasonal trends. C3 is the only one of these methods with sensitivity to gradual outbreak effects, but it is known to produce high alarm rates. For all three methods, threshold data values for alerting may fluctuate noticeably from day to day.

5.6 SOURCES

Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM. Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. Emerg Infect Dis 1997; 3:395–400

Tokars JI, Burkom HS, Xing J, English R, Bloom S, Cox K, and Pavlin JA, Enhancing Time-Series Detection Algorithms for Automated Biosurveillance, Emerg Infect Dis. 2009 Apr;15(4):533-9.

6 ALGORITHM, GSTAT

6.1 CATEGORIZATION

Temporal Scan Statistic

6.2 PURPOSE:

GStat was added to ESSENCE for applications that require sensitivity to brief signals with relatively few cases in sparse data series, but at a manageable false alarm rate. For example, the problem of interest may be to monitor daily counts of visits with one of the rarer syndromes such as rash or to monitor cases from a small geographic region, both of which can produce time series with many zeros. For prompt sensitivity to the beginning of an outbreak, a simple case threshold (alert whenever there are more than 3 cases in a day) may produce excessive false alarms

6.3 TECHNICAL DETAILS

GStat is similar to a moving average (MA) chart in that it tests the sum of cases in the most recent time window of a specified length, such as the sum of cases in the most recent 7 days. It differs from an MA chart in the test statistic, adapted from the G-Surveillance temporal scan statistic of Wallenstein and Naus. The statistic is based on a generalized likelihood ratio test from work of Martin Kulldorff for spatiotemporal scan statistics, widely used in SaTScan software.

For a moving time window of length w days, the statistic is:

$$G_t(w) = Y_t(w) \ln[Y_t(w) / E_t(w)] – [Y_t(w) - E_t(w)],$$
if observed $Y_t(w) > \text{expected } E_t(w)$, and
$= 0$, if $Y_t(w) < E_t(w)$

where:
$Y_t(w)$ = observed number of events in last w days, on day t
$E_t(w)$ = expected number of events in last w days, on day t

Scan statistics generally do not obey known distributions to allow analytic thresholds. Wallenstein and Naus determined thresholds by simulation. Users of GStat in ESSENCE have obtained empirical thresholds using a modest amount (less than a year of daily counts) of data history.

User-settable options are:
Length (days) of fixed test window
Length (days) of buffer between test window and baseline
Length (days) of baseline period
Alerting threshold

6.4 SOURCES

Wallenstein S and Naus J, Scan Statistics for Temporal Surveillance for Biologic Terrorism, MMWR Supplement, v.53, Sept. 24, 2004, pp.74-78

Wallenstein S and Naus J. Temporal Surveillance Using Scan Statistics, Statistics in Medicine, 2005.

This page intentionally left blank.

APL