Getting to know ESTAT: 
The Exploratory Spatio-Temporal Analysis Toolkit

Before you Start:
Click here to launch a full version of GeoVISTA Studio that has the ESTAT design pre-loaded:
http://www.geovistastudio.psu.edu/autobuild/gvstudio-ESTAT.jnlp

1: Using the Data Wizard:
After you’ve launched the application, click on the file icon in the upper left corner to launch the Data Loader Wizard. A small popup will appear to ask if you’d like to create a new project or use an existing one. Select the ‘New Project’ option and click ‘OK’ to proceed.

Next, enter a name to describe your project.
Now you will be prompted to choose the kinds of data you’d like to load. Since you have both PCP data and Time Series data, make sure both checkboxes are selected.

![Image of data selection dialog]

After you click ‘Next’, you will see the path selection dialog. The top half of this dialog is where data for the PCP/Map/Scatterplot are selected. The bottom portion is dedicated to selection of time series data.

![Image of path selection dialog]

Now, navigate to the directory in which you saved the sample data files and load each path so that your screen looks like the one below.
Make sure that you have selected the ‘Projected Data’ option so that the geographic data will display correctly.

When you click ‘Next,’ an error message will appear regarding your ObMeta file. Ignore this message and continue by clicking ‘OK.’

The next (and final) dialog will allow you to select which variables you’d like to display in the PCP/Map/Scatterplot and Time Series graph application elements. By default, all of the available variables are selected to be displayed, and for this tutorial you should leave it as is.

It’s easy to add and remove variables by using the arrow buttons and using standard shift/control clicking to select and deselect multiple categories.
To complete the data loading procedure, click the ‘Finish’ button and the project will be automatically saved while your data is being loaded. The application may appear to hesitate for a moment while loading data – this is not unusual.

2: An Introductory Task using the Scatter Plot

Now, we will explore ESTAT by moving through each of its elements one-by-one. For these tasks, you will be looking at county level data for the United States related to Breast Cancer mortality and a host of socio-economic variables. A list of variables and their data descriptions is provided as a supplement to this tutorial for reference.

We will begin with the Scatter Plot element of ESTAT. By default it is located in the top-left quadrant of the program.

First, change the data displayed so that ‘brcamort9397’ is plotted versus ‘unemploy’. Use the drag-down boxes to select these variables as shown below:
Now that you have changed the plotted data, select the points representing places above 20% unemployment by clicking and dragging a window over that region of the plot, as shown below:

Since ESTAT elements are all dynamically linked, this selection will cause the map window to update showing the points you just selected in the scatter plot. The map should look like the one below:
The map at this point doesn’t look very exciting, and it’s also displaying different variables than the ones you’re looking at in the scatter plot.

First, change the variables displayed to ‘brcmort9397’ and ‘unemploy’ to match the scatter plot. Use the drag down menus like you did before. After you’ve done that, try zooming in to the southwestern portion of the map using the +zoom button and click-dragging over the map:

The updated view will show county outlines and provide a better sense of geographical context. Notice where the counties you selected in the scatter plot are located. They are primarily along the border with Mexico, and in particular
are host to major border towns such as El Paso, Yuma, etc... To see the specific rates for both variables you're visualizing, click on the eccentric label icon (please pardon the misspelling in our application) as shown below:

This tool lets you roll over the map and see the values for the variables you've selected, in addition to the names of adjacent spatial units. To examine one county in particular, simply click on it and it will hold the variable display for that place. An example is shown here:
3: TheParallel Coordinate Plot

The Parallel Coordinate Plot (PCP) is a tool designed to display multi-dimensional data in a visually accessible format. It works by converting data categories into axes and then drawing lines from one category to another based on the values at each intercept.

An example of a simplified PCP is shown below:

In this example, each line represents a range of variables and their values for an individual type of automobile. The first two axes demonstrate a positive correlation – as horsepower increases, so does weight. If you examine the second pair of axes, you’ll notice a negative correlation. As weight increases, MPG decreases. Finally, if you follow the lines out to the final axes, you can see what happens when categories are visualized in a PCP. An obvious hypothesis from this data display would be that high-horsepower, heavy vehicles with low MPG rating tend to have 8 cylinder engines.

To get started with the PCP, let’s examine the same two categories of information we’ve been looking at with the scatter plot and map. By default, the PCP window is in the bottom-right corner of the application, and should look like this:
The PCP supports brushing, which enables you to mouse over the lines and see values at each axes dynamically. With a large data set, there is some latency involved with this procedure, and you will notice that it's difficult to pick out a particular single line in a large grouping. Later we’ll look at a few of the tools available for modifying the display so that you can analyze relationships in detail. Also, notice that the brushing you do on the PCP populates to the map, scatter plot, and time series windows in the application, enabling you to see multiple representations at a time.

So, to begin looking at the two variables you’ve examined before, you need to reorder the axes so that you are looking at them side-by-side. This can be done in two different ways, either by dragging the square at the bottom of the 'unemploy' axis and pulling it over so that it is adjacent to the 'rate_brcamort9397' axis, or by using the configuration panel. By default, the 'unemploy' axis is somewhat far away from the 'rate_brcamort9397' axis, so let’s try using the configuration panel. Click on the 'Configure' tab at the left margin of the PCP display. It will bring up this window:
This panel provides access to all of the controls for the PCP’s display. For now, the tab called ‘Location’ is the one you should click. This panel allows you to decide which categories to display, decide their ordering, and control whether or not the axis line for each category is drawn or not. To move the ‘unemploy’ category next to the ‘rate_brcamort9397’ first click on the row labeled ‘unemploy’ in the table labeled ‘Visible.’ Then click on the ‘Up’ button until the row is adjacent to the ‘rate_brcamort9397’ row:

Once you’ve moved the axis in the ‘Visible’ table, click on the ‘Apply’ button and the change will happen in the PCP.
Besides this method for ordering axes, you can also click and drag on the large boxes at the bottom of each axis to reorder. This method is most useful for fine tuning the order of a subset of axes on-the-fly.

Now, turn off the brushing feature by clicking on the ‘highlight data’ icon in the upper left portion of the PCP toolbar:

![PCP toolbar with highlight data icon highlighted]

Turning off brushing allows you to see the entire visualization of data a little more clearly, as the application won’t constantly highlight lines. Notice that there are several trends appearing between the rate of breast cancer mortality and the percentage of unemployment in those places. The outliers are always the most noticeable, but there are also places in the middle range that appear to draw straight across to the unemployment axis.

To begin understanding these relationships, first add boxplots to the view by clicking on the boxplot icon:

![Boxplot added to PCP]

If you like, you can click on the icon to the right of the ‘show boxplot’ button and align these boxplots automatically. To reset all the axes again, click on the green arrow icon to the right of the ‘align boxplots’ button and the PCP will revert to the default configuration.

To select a range of lines, simply click and drag across them on the display:
Now that you've selected a range of lines, you can see how this subset behaves across the rest of the axes. If you wish you can turn the interactive brushing back on and look at specific values within this group.

It may be helpful to expand the 'rate_brcamort9397' axis in order to examine the subset you've created more effectively. To start doing this, open the configuration panel again. Click on the 'Range' tab if it isn't already selected and adjust the axis range by dragging on the blue sliders. Arrange the configuration panel so that you can see the PCP window, and you can then watch it update in real-time. This also works for the data range sliders, which are represented with green handles. To change the axis you are focusing, simply select it from the list shown on the left side of the panel:
4: Working with Time Series

The Time Series plot is functionally identical to the PCP you were just working with. By default, it is located in the upper-right corner of the application. This element is intended to display temporal data for a single category. In this example, you are looking at the age-adjusted rate for Breast Cancer mortality across several time periods from 1973 to 1998. All of the configuration controls, brushing, and analysis tools that work in the PCP work identically in the Time Series plot.

Notice that the range you selected in the PCP is also selected in the Time Series plot. This allows you to examine temporal trends related to the subset you’ve made:

5: Explore!

ESTAT is designed to enable interactive geovisualization with complex datasets. Take some time to try some of the buttons and configuration panels that you did not explore in this tutorial. We hope most of them are self-explanatory.

Your feedback to us is extremely valuable. We want to create tools that help you explore and uncover new hypotheses. In the focus session at the end of the series of tasks you will tack next, we will ask you to tell us what you think of the basic layout and functionality of this software. Think about tools and options you think should be included that aren’t, or about features in the existing package that you don’t understand or are awkward to use.

Questions or Comments? Contact us – arobinson@psu.edu