PLTW COMPUTER SCIENCE

Activity 3.2.4

Geographic Data

Introduction

Many data sets include information about location. These data are naturally visualized with a map. Other information in the data set can be displayed with

- point markers (with color and shape to indicate categories)
- direction markers (with arrow width, length, and color to indicate quantitative variables)
- shading and contour lines to show a location-dependent value like temperature or mean income
- a color to indicate a location-dependent frequency or density like population
- distortions of the areas of map regions to display values, frequencies, or densities.

What would you want to know about that might vary by location?

Image courtesy NASA ©2011

Equipment

- Computer with Enthought Canopy distribution of *Python*® programming language, Microsoft® Office®, and access to the Internet

Resources

3.2.4 source Files

Procedure

Part I: Getting Data

In this activity you will use *Python* to create a data visualization of the worldwide locations of seismic events.

1. Form pairs as directed by your teacher. Meet or greet each other to practice professional skills. Set team norms.
2. View earthquake data from the US Geological Survey at http://earthquake.usgs.gov/. We want to collect a list of seismic events from this site and then use *Python* to create our own visualization.
3. Click on the **Earthquakes** link and then click on **Search Earthquake Catgalog.**
4. The USGS's search function can search thousands and thousands of earthquake records. We want the search to return a large but manageable number of records with a good set of sample data. So, set the Basic Options as stated.
   - Select the **Magnitude** radio button that indicates **4.5+**
   - In **Date & Time**, choose **Custom** and specify a 5 year date range. You may change just the year or you may change year, day and month. You do not need to specify a time of day.
   - For **Geographic Region**, select **World.**
5. Expand the **Output Options** and choose the following.
   - For **Format**, select **CSV**. A comma-separated value (CSV) file will be generated.
   - For **Order By**, select **Time - Newest First**.
   - **Limit Results** by setting **Number of Events** to 20000.
6. Click on the **Search** button. When the search finishes processing, the CSV file will be automatically downloaded to your computer.

Part II: Inspecting Data

7. To open the CSV file, select it from the Downloads page in your browser, or go to your Downloads folder. Double-click the file to open it in Excel®.
8. Discuss with your partner what data is displayed in each column in the CSV.
   - What data is recorded in the columns that you understand? Refresh your memory on the meaning of longitude and latitude.
   - Which columns do you not understand?
   - We are going to be plotting the locations of the seismic events. Which columns of data are the most important for that activity?
9. Use Excel to find the seismic event with the greatest magnitude.
   - Use Google Maps to locate the event on a map. Capture a view of the map (using Snipping Tool) that shows the location and paste it in your answer document.
   - Research the date of the seismic event and briefly describe the impact of the event.

Part III: Visualizing Data with *Python*

10. In Canopy, open the program `world_map.py`. Save or copy the `query.csv` file so that it is in the same folder as `world_map.py`.
11. Execute `world_map.py`. Save an image of the plot that is created.
12. Discuss the plot with your partner. Record some observations about the data that you can make by using the visualization.
13. The intensity of an earthquake might depend on how deep below the surface it originates, since surface rock is cooler and more brittle. Modify the code so that earthquakes originating less than 10 km below the surface are magenta, while other points are colored yellow.
   - Save your code and plot as directed by your teacher.

○ Record observations about the data that you can make from the visualization.

14. You downloaded only selected data from the USGS web site. How would changing your search parameters change the data visualization plot? Generate a plot from a new, modified search. Explain what you observe from your visualization.

15. How could you present more information in the data visualization or present the information in another way?
    ○ Brainstorm ideas with your partner.
    ○ Choose one of your ideas and explain how you might implement the idea in code.

Part IV: Considering Geographic Visualizations

In this part you will consider visualizations of three different data sets.

16. Examine the two maps showing the most popular boy and girl name for each state. http://bigthink.com/strange-maps/610-would-smell-as-sweet-geo-popularity-of-given-names
    ○ Describe some part of the pattern that you found most interesting.
    ○ Give an example of a question that would have been easy to answer without a geographic visualization by using the SSA baby names data set.
    ○ Give an example of a question that would have been difficult or impossible to answer without a geographic visualization of the dataset.

17. Examine the maps at http://www.businessinsider.com/22-maps-that-show-the-deepest-linguistic-conflicts-in-america-2013-6. Select **Click here to see the maps**. Coordinate your selections so students in the class choose different maps.
    ○ Describe the pattern apparent from the visualization.
    ○ Give an example of a question that could have been answered from the researcher's survey data in raw form without a geographic visualization.
    ○ Give an example of a pattern that would have been difficult or impossible to detect without a geographic visualization of the dataset.

18. Examine the maps at http://www.businessinsider.com/most-segregated-cities-census-maps-2013-4. Select **Click here to see the most segregated cities**.
    ○ Discuss with your partner how this data visualization could be used in helpful or harmful ways by decision makers locating a factory, park, or school.
    ○ Could the media or the public use the same visualization to influence the decision maker you described in part a?
    ○ How does computational power affect our ability to make decisions?

Conclusion

1. Why are geographic data such an important category of data to visualize?
2. What sort of patterns in data can only be discovered by a visualization that uses a map?
3. For now, the human brain is better than a computer at picking out objects in a picture and identifying them. Explain why a combination of computer and human intelligence is necessary to discover the patterns you described in this activity.