

Visualizing Spread of H1N1 Virus

Data appropriate to tracking the spread of H1N1 are surprisingly hard to find. The CDC seems dedicated to a particularly obstructionist presentation of its data, preferring to offer it in pdf tables. Never the less, some enterprising websites have compiled data from CDC, WHO, and other such sources and made them available in a more general format.

<http://www.mapcruzin.com/free-download-h1n1-swine-flu-arcgis-shapefile.htm>

The data consists of daily reports of "confirmed" cases of swine flu, including the country, city, state (in the US) and county (when appropriate), date, whether the case was fatal, a "description", and some other information. There are just under 9000 cases.

The data are messy, but lead themselves to the following educational opportunities:

- 1) Working with dates (probably the most common "non traditional" data form students routinely encounter that are routinely ignored in undergrad stats classes.
- 2) Text analysis. The "descriptions" are sometimes quite detailed, and one might ask if there are common features to the descriptions, uncommon features, if these vary by location, etc.
- 3) Data Cleaning. Even the indicator variable seems to have some odd features. The size of the data set requires some strategizing in order to identify oddities.
- 4) Shape files are available for integrating the data into a GIS package.

Potential Analyses:

- 1) Can you visualize the spread of the virus within the US? Can you prepare a visualization that compares the spread in neighboring countries?
- 2) Can you quantify the rate of increase? Can you identify plateaus?