Raul Martinez - Ivan Ulloa

Exercise 7 Report: West Nile Virus Geospatial and Temporal Visualization

Motivation

The West Nile Virus is a single stranded RNA virus that causes the West Nile Fever, it was first observed in Uganda in 1937 and first detected in Morocco in 1996. The natural hosts are birds and it is generally transmitted by mosquitoes to humans. Thousands of cases are reported in the United States every year, with most of them occurring in the months of August and September. Those at high risk are people over the age of 60 and others with health related problems, no specific treatment is available but pain medications might be useful. Due to the nature of the virus transmission and propagation it is important to have tools for spatial and temporal traceability of human infection. With that in mind, the visualization presented in this report provides a geospatial representation for the number of positive cases encountered within the counties of the state of California, it displays the number of cases per multiple weeks for the years of 2006 up to 2015.

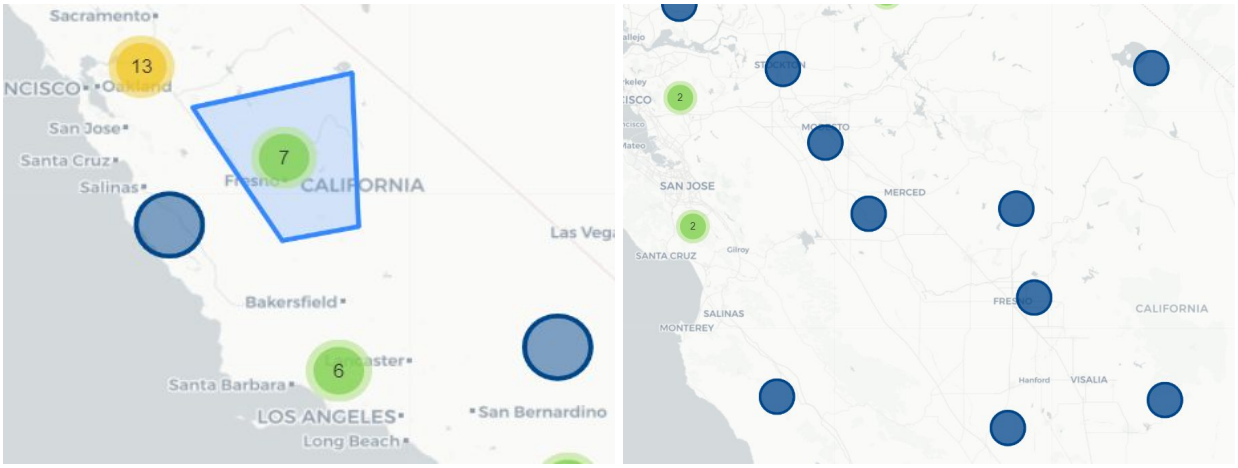

Figure 1. Geoshape used to hover and select regions at a high-level (left), when clicked the markers are zoomed in and expanded (right).

Data Augmentation

The original dataset comprised county information for year, week reported, and number of positive cases. To consolidate temporal information among the visual encodings and allow the data to be distributed spatially a series of steps were taken. First, the latitude and longitude coordinates were merged into the dataset for each county location, these coordinates were then transformed slightly with an offset in order to allow each year marker to be placed in a circled pattern (for a maximum of ten), then a center marker was defined for each county to show the sum of all positive cases for all years. Lastly, the cases per week reported were added together to report a single value per year, and to present the aggregated information a new column was added to report a string of week numbers with the number of cases in parenthesis; for the circle centers pertaining to all years a value of "Not Applicable" is shown (*see Figure 2*).

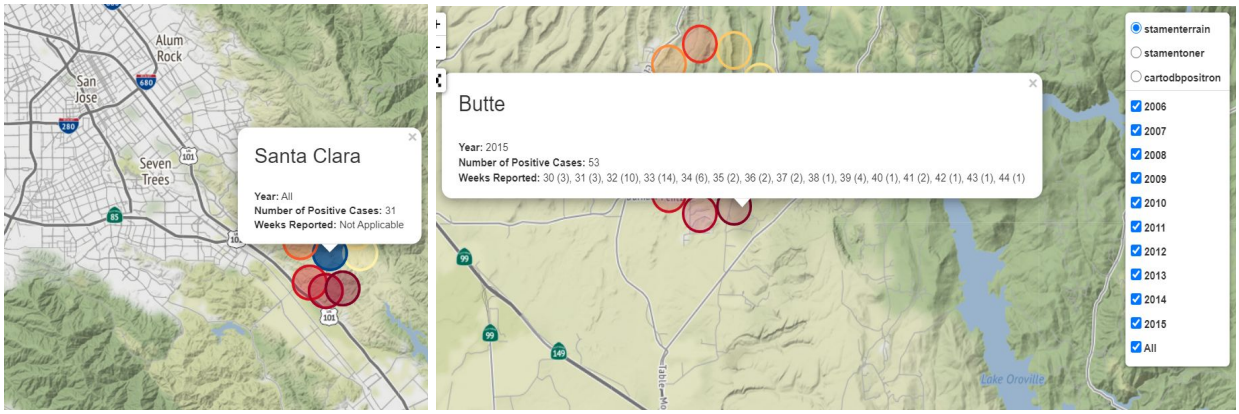

Figure 2. Dark blue markers represent all positive cases for the county (left), while the rest colors organized in a circle represent cases for each year (right).

Tasks

The geospatial map of West Nile Virus cases for the state of California aims to explore the following questions: Is there a relationship or pattern between number of reported cases and proximity to water sources? Which county has the highest number of reported cases? Can we attribute geographical location, for example northern counties vs southern counties to the number of cases reported?

Expressiveness of design

The visualization makes use of circled markers to indicate geospatially county level data. It starts by loading the entire state of California to provide a high-level view of all visual elements, then the user can zoom in and out within the interface to explore regions of interest as shown in *Figure 1*. Each year is represented with an individual marker by using a sequential multi-hue color map (yellow to red), with an additional dark blue marker created to display the sum of all years; this color was selected to provide high contrast. Markers were designed to be in a circled pattern for easy identification of each year, and checkboxes on the right were placed to allow the user to add or remove individual years or to include a summary for all years. With this design both temporal and spatial data attributes are displayed in the visual system, and the visual encodings are structured to guide the user and avoid misinterpretation of non-existing data trends.

Effectiveness of the solution

The importance of the information presented was encoded primarily based on a combination of color and density of the circular markers. The hierarchy begins with high density and dark color as the summary of all cases for one county. This method is effective at discriminating between the year by year cases and the total summation by emphasizing the summary as a distinctive measurement. The map provides additional information depending on the type of map selected as shown in *Figure 3*, the three different maps included “stamenterrain”, “stamentoner”, and “cartodbpositron” all provided the viewer with different features that could be used in the analysis of the reported cases, for example, the terrain such as mountains, lakes, rivers, and coast regions can be easily be identified using the “stamenterrain” map. The other two maps focus on providing information about street names and main routes using high and low contrast.

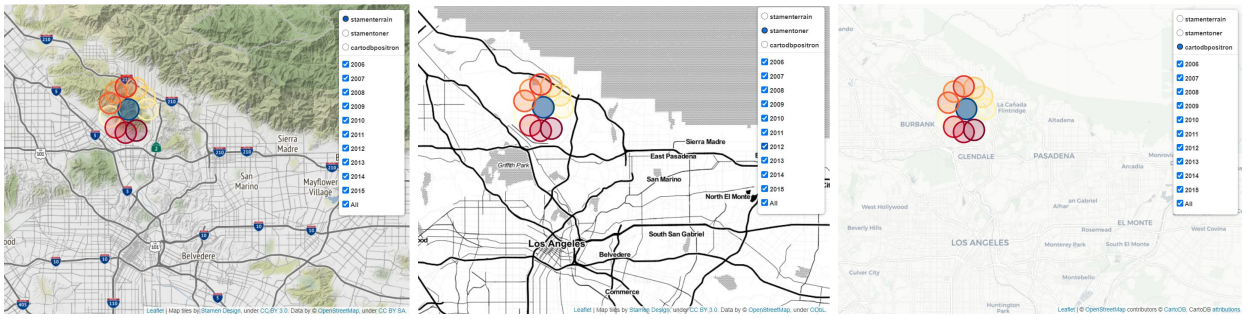

Figure 3. Three different topographies are available for visualization named as stamenterrain, stamentoner, cartodbpositron; in order left to right.

Interaction

There are various interactive elements included with the visualization, the first is the ability to select solid markers that contain a number at the center as shown in *Figure 1*, this number represents the number of counties that are included within the region that this marker covers. This feature is used to reduce the amount of clutter when the map is zoomed out and many markers could appear. The user can select this marker which will zoom into the region and expand the counties. Another interactive feature is the ability to select each of the yearly or summary markers to show the name of the county, the year, and the number of cases. Another interactive feature is the ability to toggle between different types of maps as well as to toggle on and off data for the years and also the "All" feature.

Conclusion

A geospatial and temporal visualization for the number of positive cases for the West Nile Virus in the state of California was discussed. The design proposed provides the option to select from three topologies to display spatial information at a county level. Temporal data attributes are encoded with multiple ordered circled markers and by displaying metadata with interactive elements. The most number of cases observed in the north belong to the counties of Butte, Sacramento, and Yolo with 170, 123, and 70 respectively; while the highest numbers for the south are in Los Angeles, Orange County, and Riverside with 1112, 500, and 286. In summary, the design provides an accurate representation for displaying spatial information on large surfaces and facilitates data exploration for comparing across counties and years, additional channels of information can be used to provide better differentiation for the number of cases such as marker size or colormaps.

References

- https://github.com/NoahChristiansen/folium-FireMap
- https://en.wikipedia.org/wiki/West_Nile_virus
- https://public.opendatasoft.com/explore/dataset/us-county-boundaries