Akshit Bhatnagar Jared Goldsmith Jillian Jarrett March 1, 2018

Exercise 8: Geospatial and Temporal Nile-ism

Motivation: In 2015, there were 783 reported cases of West Nile virus (WNV) in California. Of those, 53 included fatalities. Exploring patterns in the temporal and geospatial distribution of West Nile virus cases in California would help illuminate trends and inform mitigation efforts.

What: The West Nile virus dataset includes weekly reports from California counties documenting the number of positive cases of the virus.

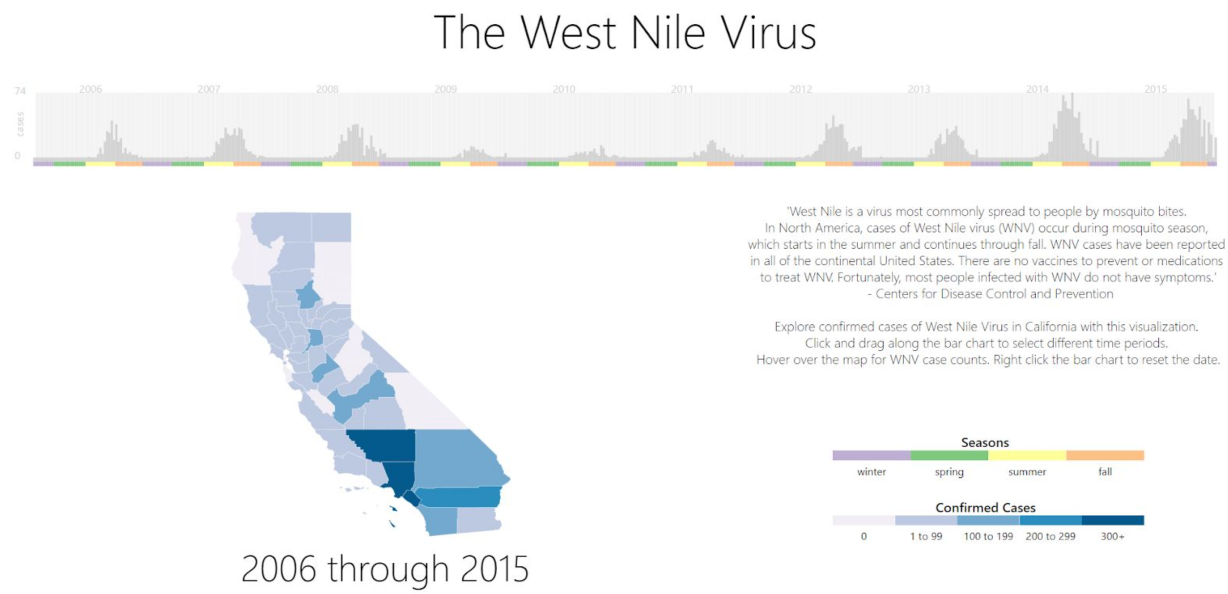
Attributes & Types

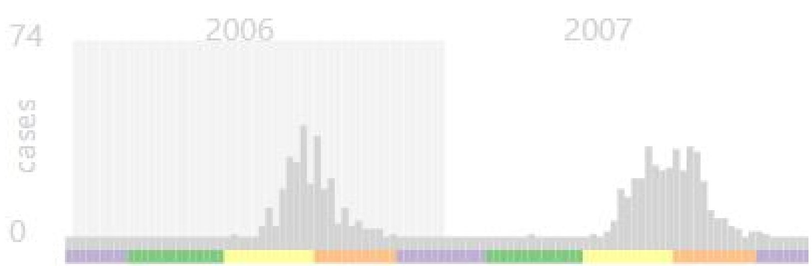
County	Categorical, Geographical
Year	Quantitative Sequential
Week_Reported	Quantitative Sequential
Positive_Cases	Quantitative Sequential
Season	Derived Attribute, Ordinal

Why: The targets for this visualization were temporal trends and geospatial distributions of West Nile virus cases in California. Analysis was expected to be discovery and general consumption of raw data. Search with the visualization would include looking through known attributes (temporal and geospatial) for unknown patterns. And temporal queries should summarize the selected subset data.

Actions
Analyze ➜ Consume ➜ Discover
Search ➜ Browse
Query ➜ Summarize

How: All 5 attributes of the dataset (county, year, week, case count, and season) are encoded in the visualization, summarizing case counts across time and space. Counties are encoded as geographic spatial regions, the most effective categorical channel. Years and weeks are encoded as the X-axis on a bar chart, using position on a common scale, the most effective ordered channel. Seasons are encoded both as categories using colored rectangles and ordered along the common time scale as years and weeks. Case counts are aggregated across time and encoded using vertical bar length and aggregated across geographic region and encoded using luminance. The case count luminance scale was a sequential monochromatic scale from colorbrewer with 5 bins, the most recommended by Munzer to retain discriminability. The top two most effective categorical channels are used as well as 2 of the top 3 most effective ordered channels. Data can be reduced by timespan selection. Selections are marked with highlighting.





Results & Conclusions: Combining a selectable timeline with a choropleth map proved to be a useful display idiom for exploring patterns across the time and space axes. Using the derived season attribute, we were able to see the peak WNV season push further into the fall over the encompassed years. Using the animated selection functionality we were able to see the WNV propagation pattern each season. Typically the first cases are around Kern or Tulare county, both home to lots of rivers and ponds which are breeding grounds for mosquitos, and then the cases spread out across the central valley.

There were several difficulties with the design choices presented. First, comparing different time ranges spatially required the user remember the last view. Allowing multiple ranges with multiple small maps could alleviate this, though the shrinking map size could interfere with the luminance channel. Another challenge was appropriate binning of the data for the luminance channel. Using only 5 bins limited comparisons on smaller time ranges, but we decided the increased discriminability was worth the trade off. Another challenge was our attempt to utilize the maximum resolution provided by the client. We were able to code most of the elements to scale dynamically with screen size, however text sizing and centering proved difficult. The final challenge was providing accurate representation of cases per capita or per square mile. Without secondary data sources for population or county size, this was not an option.