DSE 241 – Data Visualization

Submitted by Alejandro Hohmann
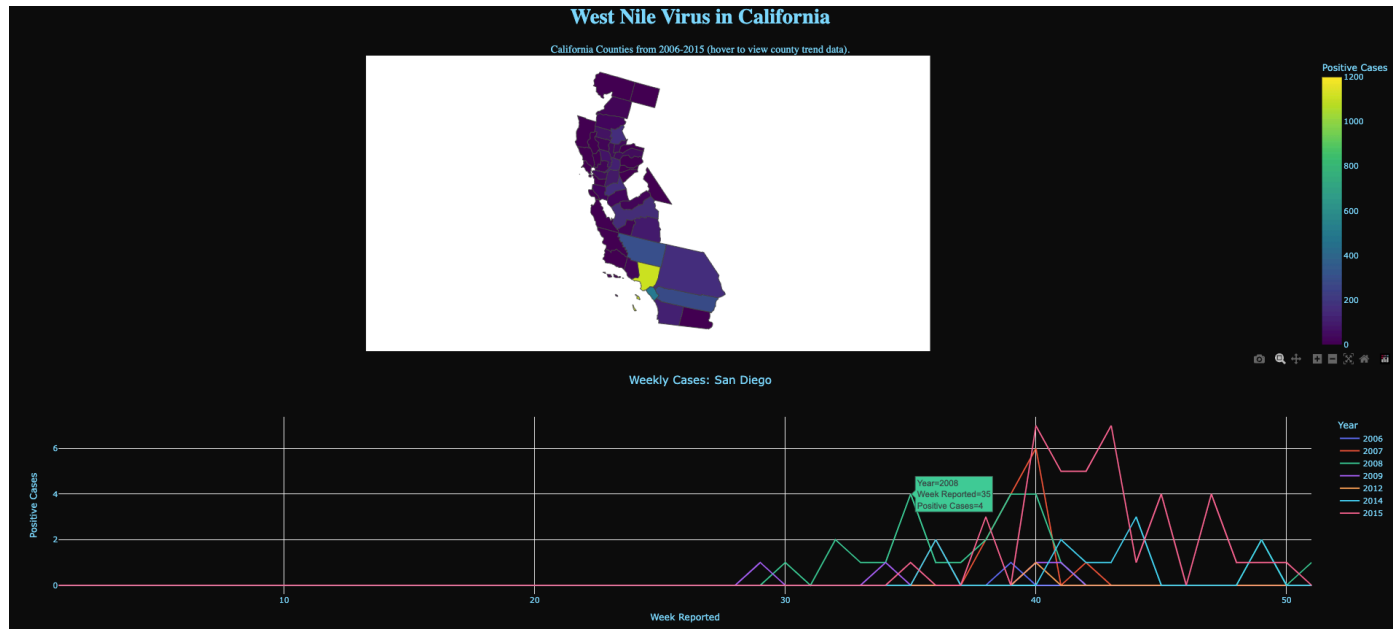
Exercise 6 Report: *West Nile Virus in California*

Report Contents
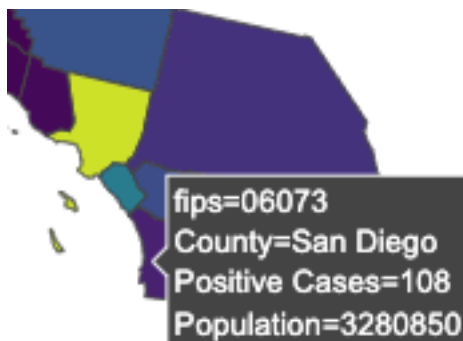
The Viz

Actual viz requires interaction (run the accompanying jupyter notebook to display in browser).

Choropleth Tool tip example:

Motivation

This visualization is intended to quickly edify users on:
- which counties have the most cases of West Nile Virus
- the seasonality/trend of the virus
- comparison of years of the virus

Data augmentation (if any)

- querying the polygon bounds for each the of individual counties to plot in geospatial format
- formatting FIPS codes for the mapping of counties to geospatial data
- filling in missing weeks so they are not imputed on the line chart
- renaming columns so they are chart friendly

Tasks

The user is required to hover over individual counties to see the total number of cases in an individual county during this time frame. The user can also ascertain trend data by looking at the corresponding county data by year by week. The user should also think critically about population size embedded in the choropleth tooltip, and what effect it may have on the case counts.

Expressiveness of design

The color channel of the choropleth does a good job of highlighting LA County as having significantly more cases than the rest of California. The diverging color palette pleasantly differentiates high case counts from low case counts.

Effectiveness of the solution

There are two different color channels:

- Choropleth color channel
  - The color channel in the choropleth clearly shows the separation of counties, especially LA County.
  - One drawback of the choropleth is that it does not readily encode any information for the total population. The tooltip does include the total population (as of 2015) but this requires more analysis by the user. In a separate chart that shows normalized population (cases per 100k), you can see that Los Angeles is actually in-line with the rest of the counties, but their large population brings the total case number up. A much smaller county (Glenn) now stands out.
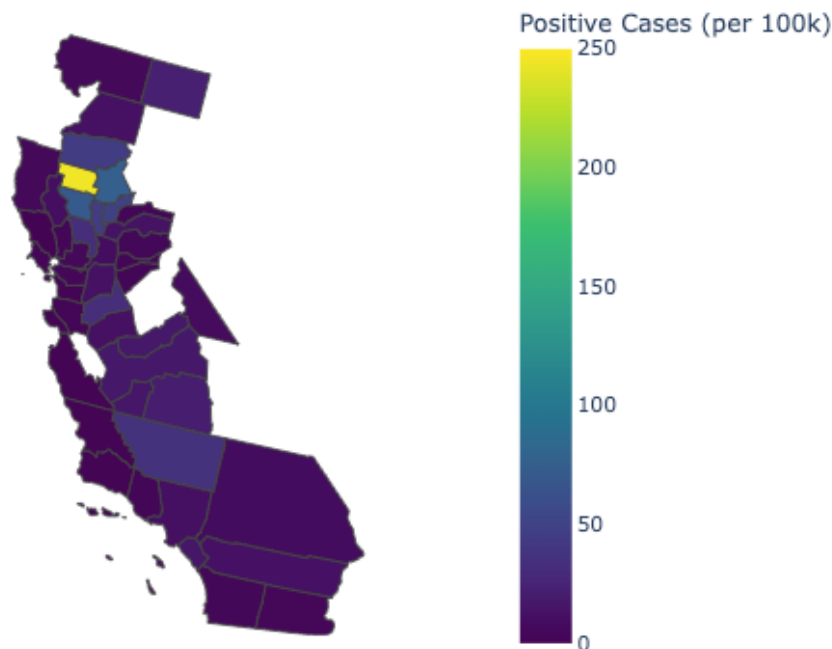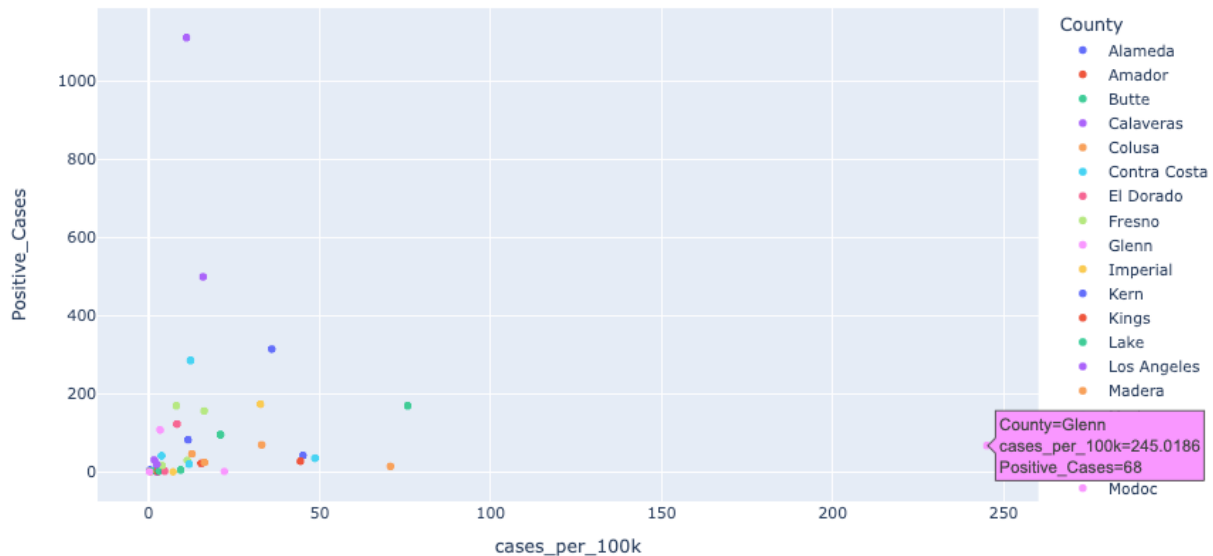

- An alternative idiom like a scatter plot would do a better job of showing the relationship between population and cases. The population in this

example is from the 2015 Census estimate and clearly shows LA is in-line with the rest of California.
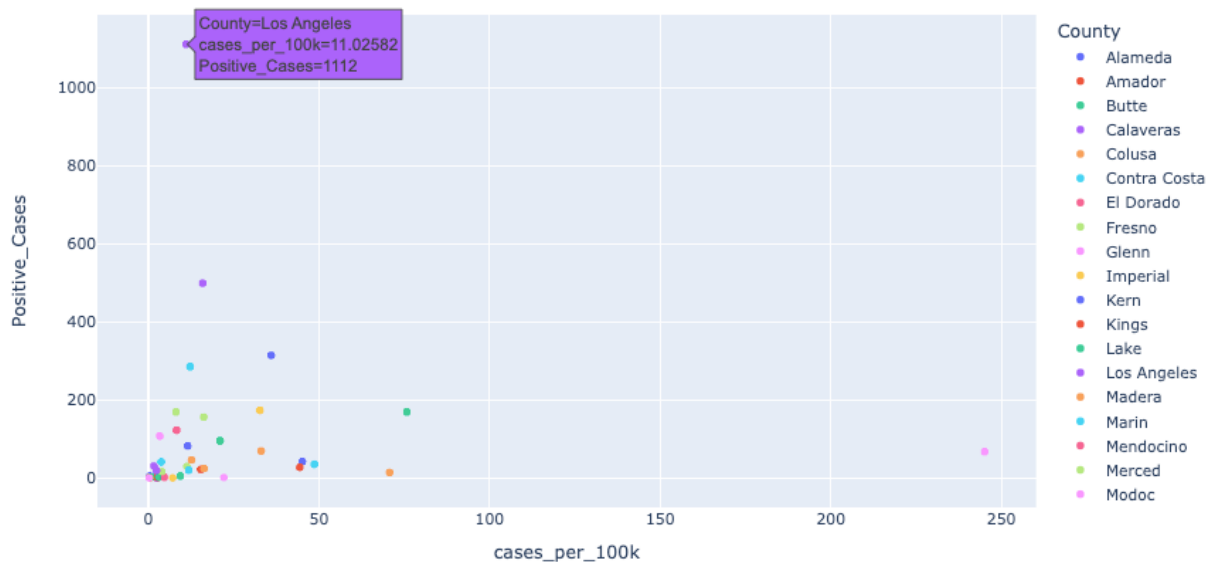
It is Glenn county that stands out as having far more cases per 100k than any other county.

- Line Chart color channel
  - o Does not do a great job of separating the years in all cases because in some of the counties there are many overlapping lines. In counties that have missing years, perhaps with no cases in those years, the color channel does a better job.
  - o However, where there is ambiguity the user can sort it out by interacting with the chart. Hovering over points on a line displays the year, week number, and the number of positive cases in that given week.
- Line Chart Size channel
  - o The lines clearly show that the data is bunched up in the later part of the year (generally weeks 30-52). This may be due to missing data or may be due to

seasonality. Assuming seasonality, it is pretty stark and well represented by the trend chart.

- o The lines also show that there are not generally increasing/decreasing cases year-on-year as this has fluctuated up and down.

Interaction

The user must hover over individual counties to get the trend data. One drawback of this viz is that the trend data does not display until the interaction occurs. Hovering over individual counties in choropleth allows the user to dive into the data more. This interaction may ultimately keep the user more engaged and thinking critically in a way that they internalize the data, but the user has to get past orienting themselves to the viz's functionality.

Conclusions

This viz is effective at conveying total West Nile Virus cases for an individual county, with Los Angeles having the highest number of cases due to its population size. The trend data by county by week by year effectively tells the story of cases generally occurring during the second half of the year. Overall this small dashboard does a decent job of displaying the information embedded in the raw data.