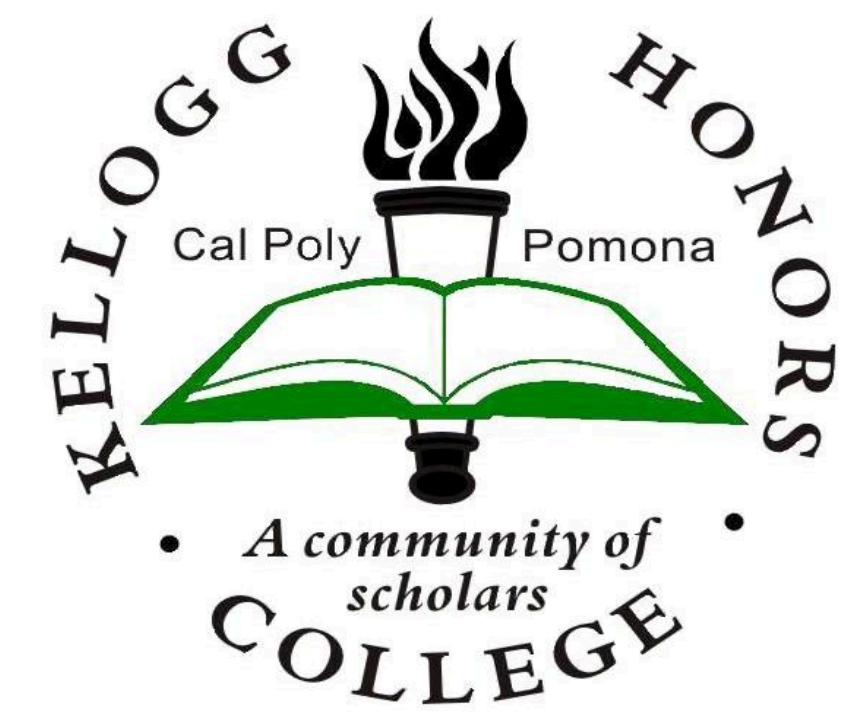


California Counties and Covid-19: A Descriptive Analysis



Patrick Ogaz
Computer Information Systems
 Mentor: Dr. Rita Kumar
 Kellogg Honors College Capstone Project



ABSTRACT

With the growth of data and tracking, Covid-19 has been distinct in its ability to be one of the first fully tracked Pandemics in U.S. history. As a result, many companies and Government agencies ask us what the data tells us about the pandemic and how we understand it to help make decisions. This study analyzes a dataset of Covid-19 data points provided from data.ca.gov and the U.S. Census. We utilize descriptive analytics on the Tableau software to visualize trends and patterns within the 58 California counties. The goal was to understand better what happened during this pandemic and observing how the pandemic impacted different communities. Included in this analysis are data about demographics, population, and socioeconomic factors, and their potential effect. In this study, we also deploy a time series dashboard that allows users to interact with the data to better understand trends and the various factors of Covid-19 in California.

Objectives

- Examine and analyze Covid-19 cases and deaths by county
- Get a better understanding of how Covid-19 changed regionally over time
- Better understand the demographics of those most affected
- Visualize and understand the differences in health and socioeconomics of California counties and its possible connection to Covid-19
- Cluster counties to better understand performance

Key Data Fields

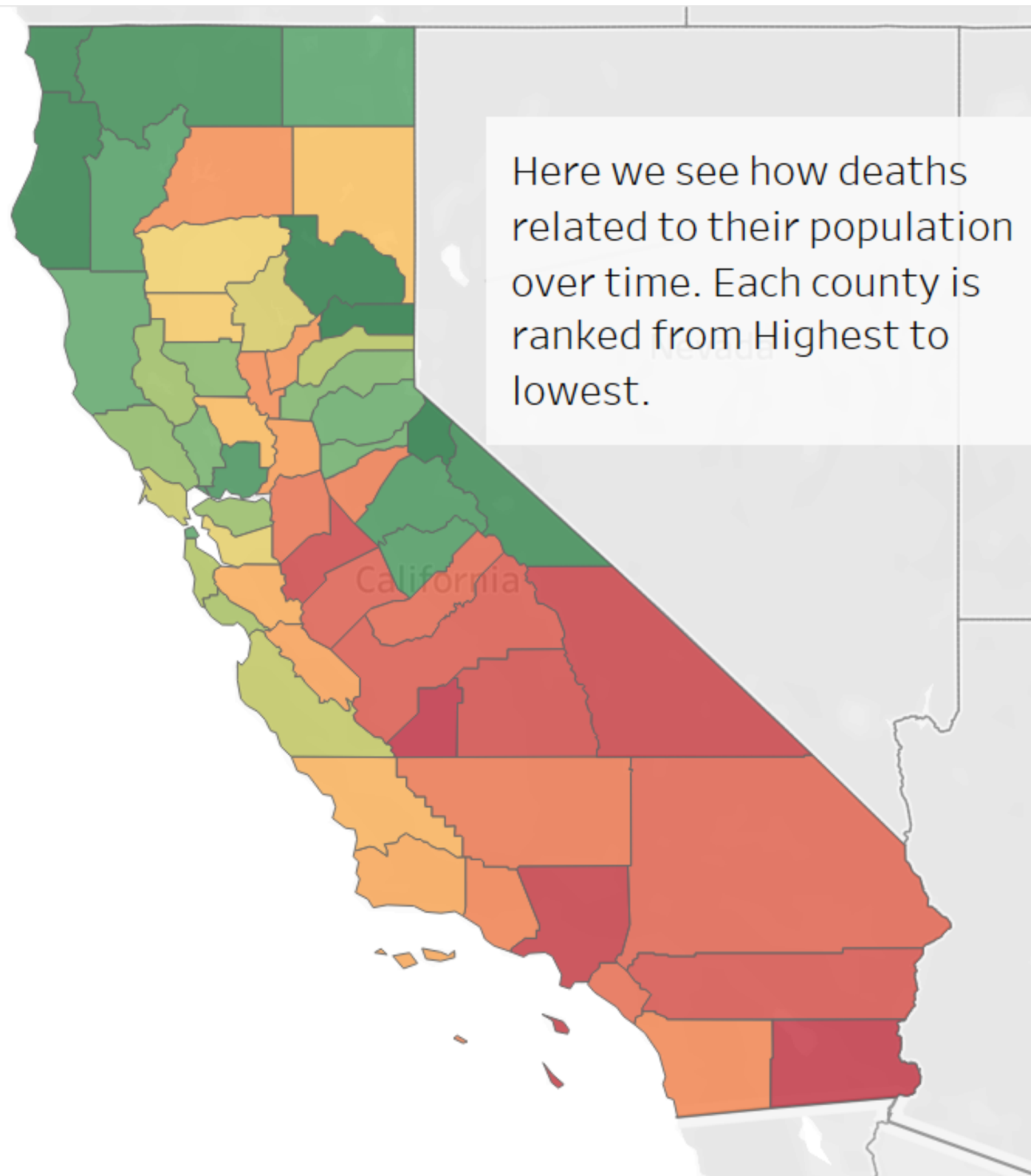
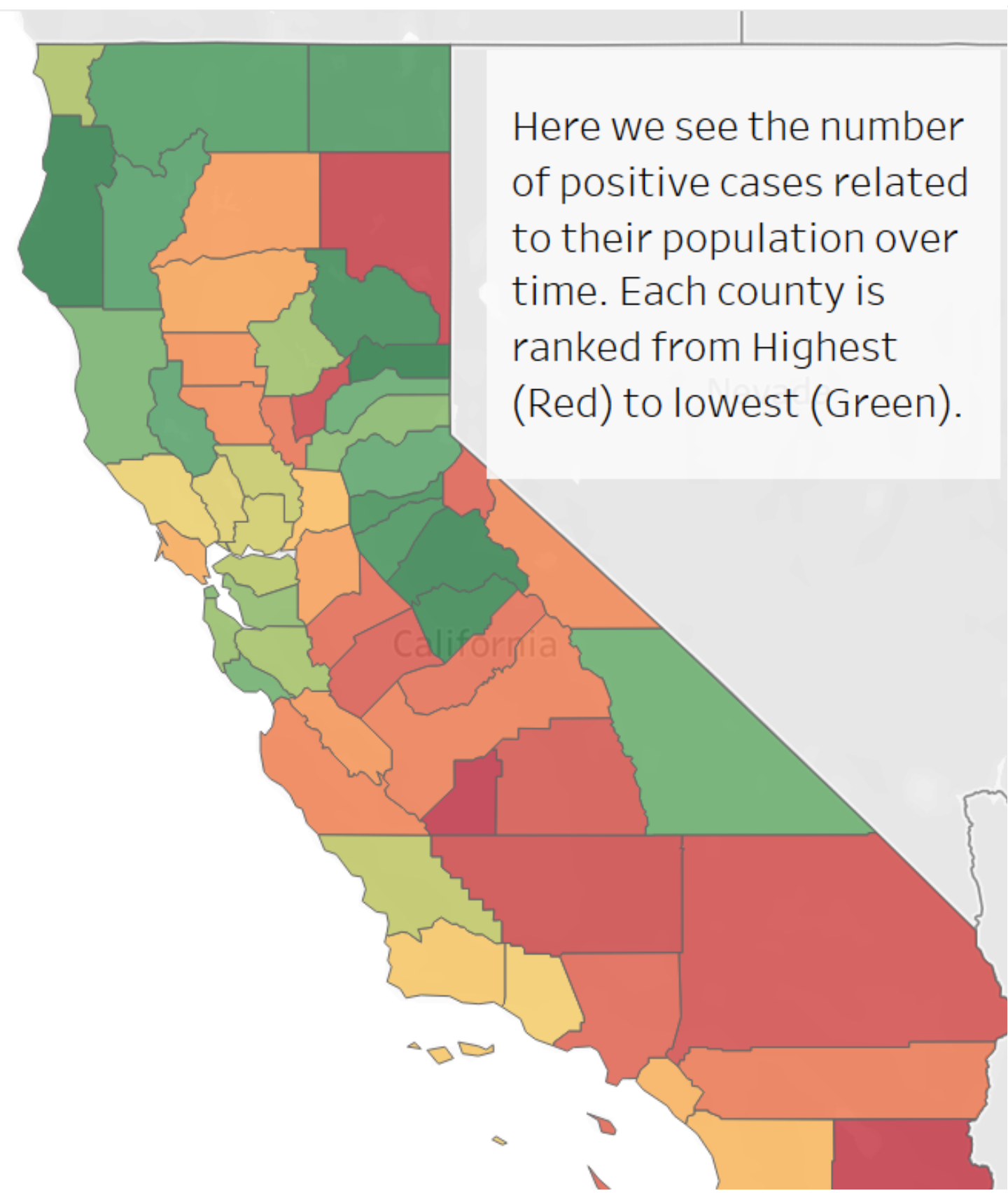
- Daily Positive cases by county timeseries
- Daily Deaths timeseries by county
- Deaths and Cases demographics (Gender, Age)
- County Demographics
- Vaccine timeseries data

Methodology

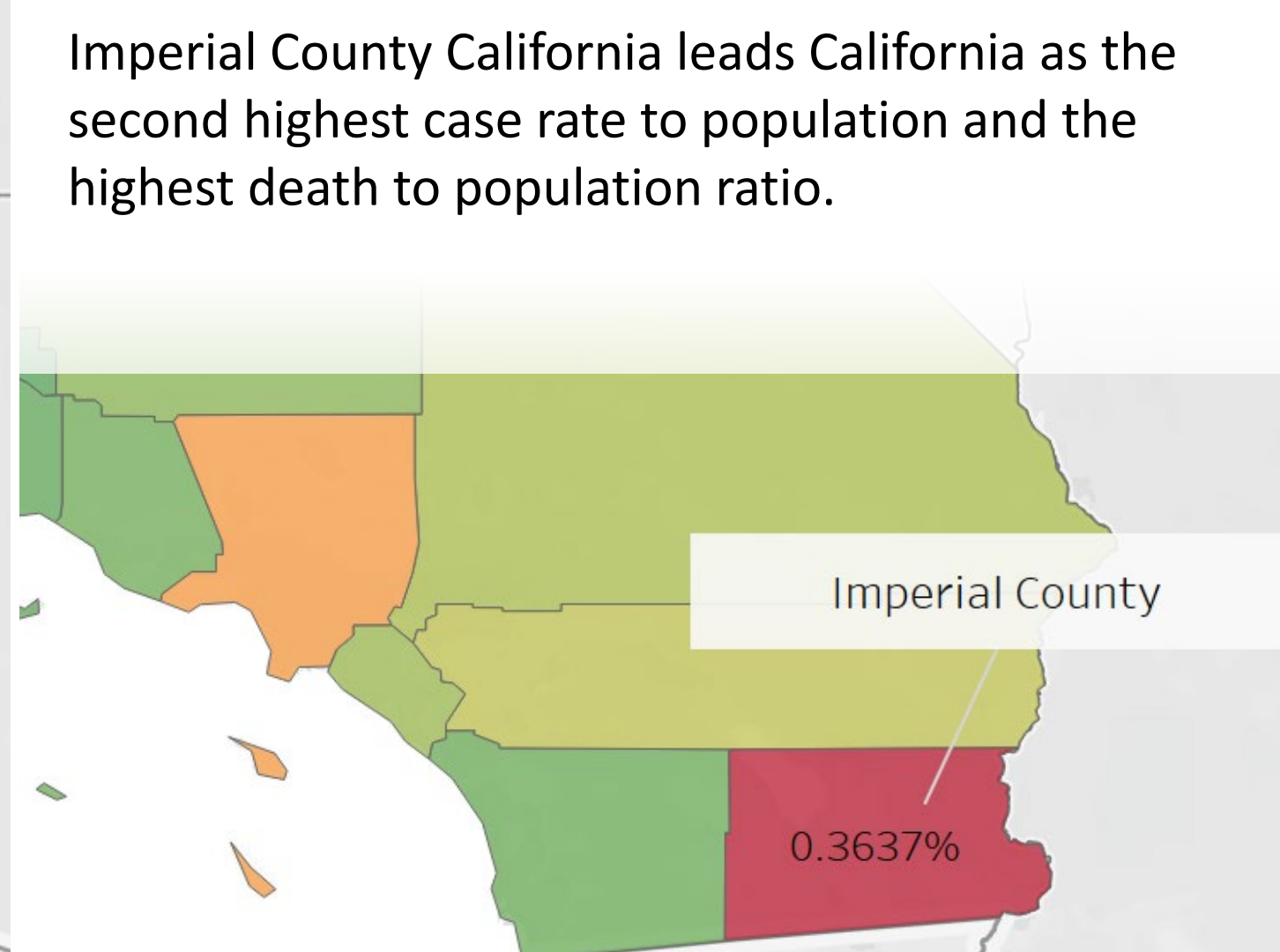
- I used Excel to compile many of the data files that I found from sources like data.ca.gov, the CDC, and other resources. I also used Excel to create aggregations by county.
- I also took medians for most values to get a central measure for what was happening in each county and for much of the demographic data.
- I utilized Tableau to Visualize the data as a timeseries.
- I then created visualizations utilizing demographic information and merged (blended) the files in tableau.
- I then ran trend lines (linear regression) through scatter plots with demographic data to see if there were any trends with demographics.
- I then published the Tableau file via Tableau public.



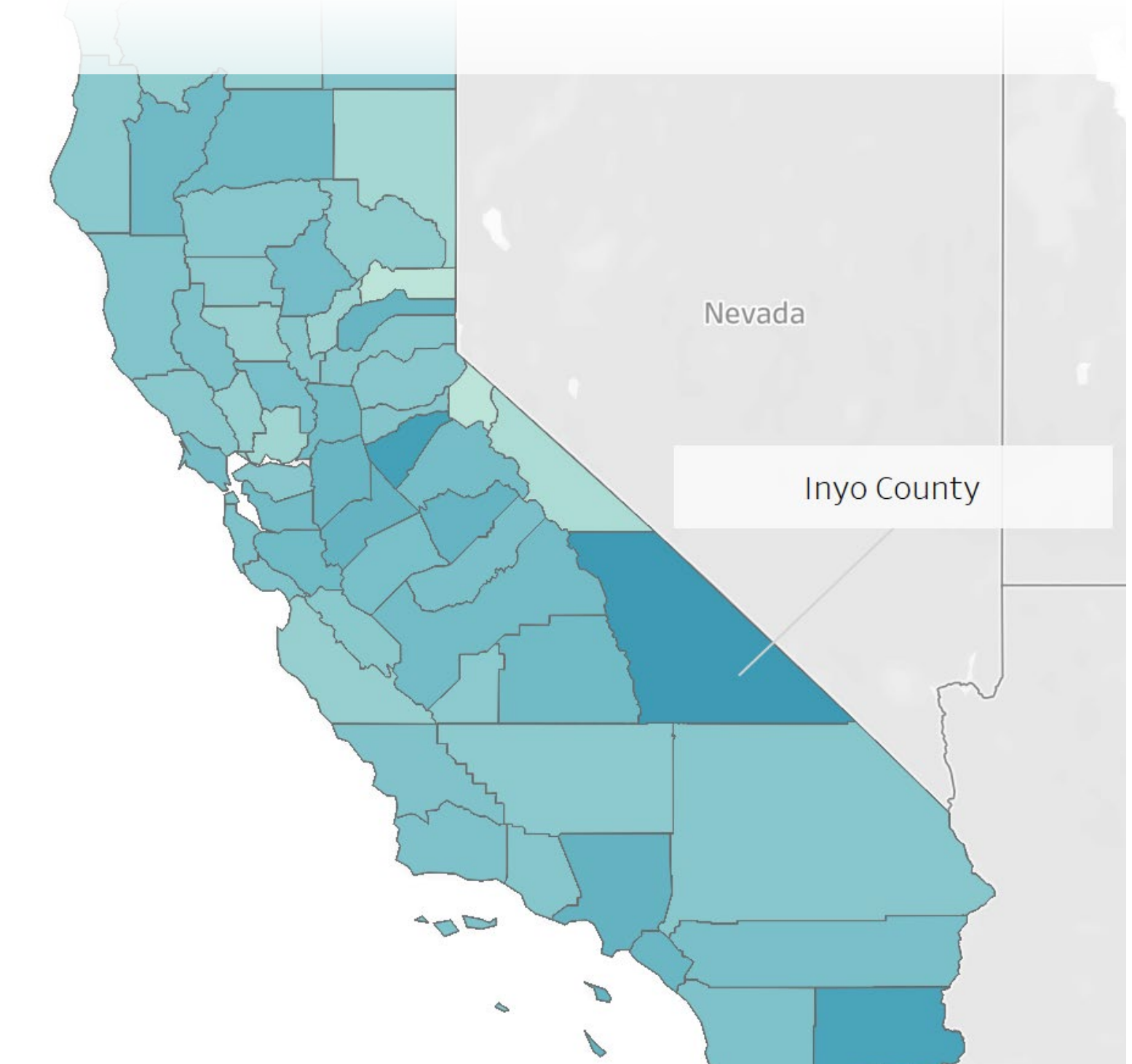
Visualizations and Findings



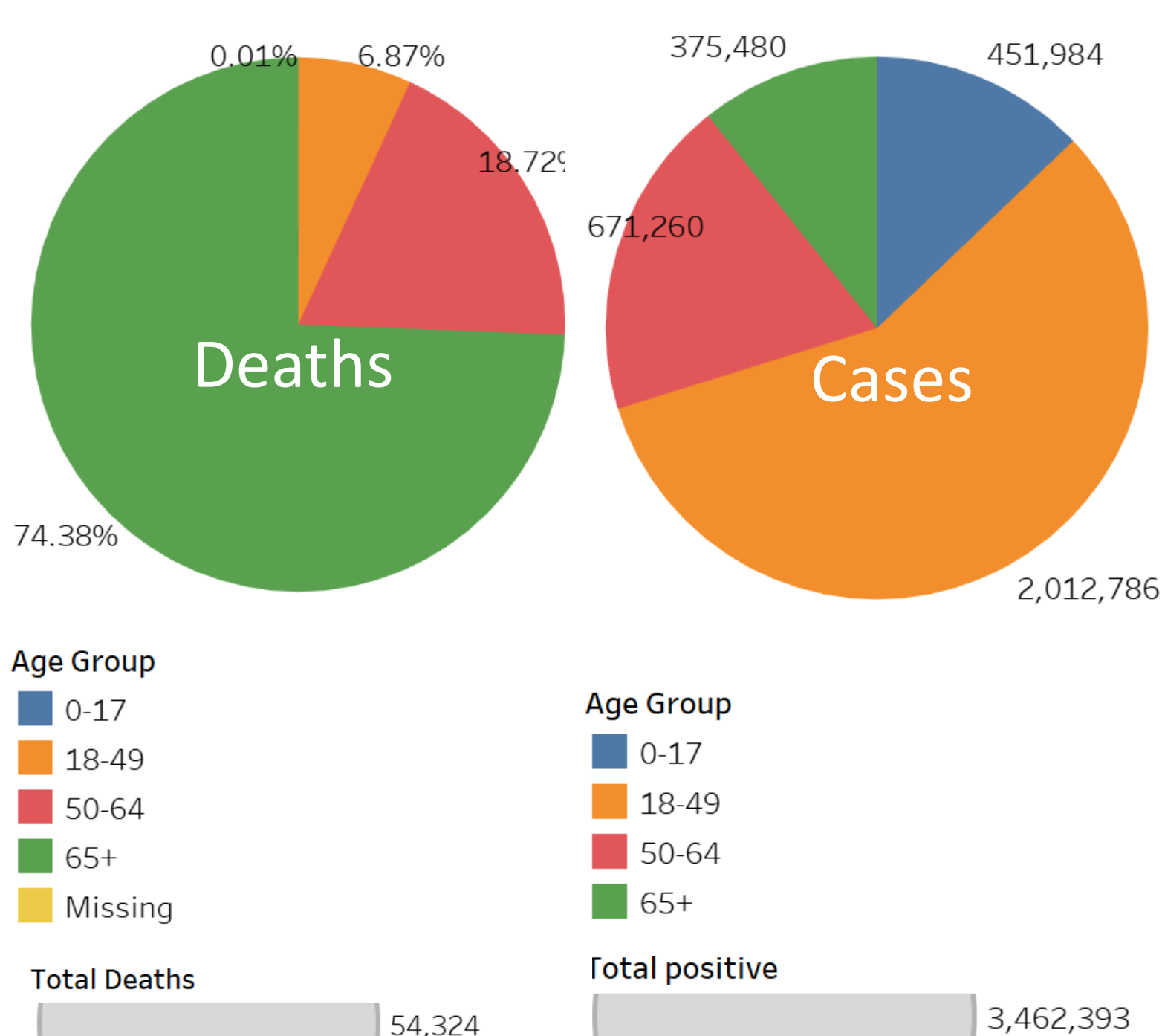
County	Rank (1-58)	Deaths to Po..	Total Deaths	Population
Imperial	1	0.003637	659	181,215
Los Angeles	2	0.002201	22,099	10,039,107
Inyo	3	0.002051	37	18,039
Stanislaus	4	0.001742	959	550,660
Tulare	5	0.001665	776	466,195
Riverside	6	0.001595	3,940	2,470,546
Fresno	7	0.001510	1,509	999,101
Merced	8	0.001498	416	277,680
Kings	9	0.001497	229	152,940
San Bernardino	10	0.001459	3,180	2,180,085



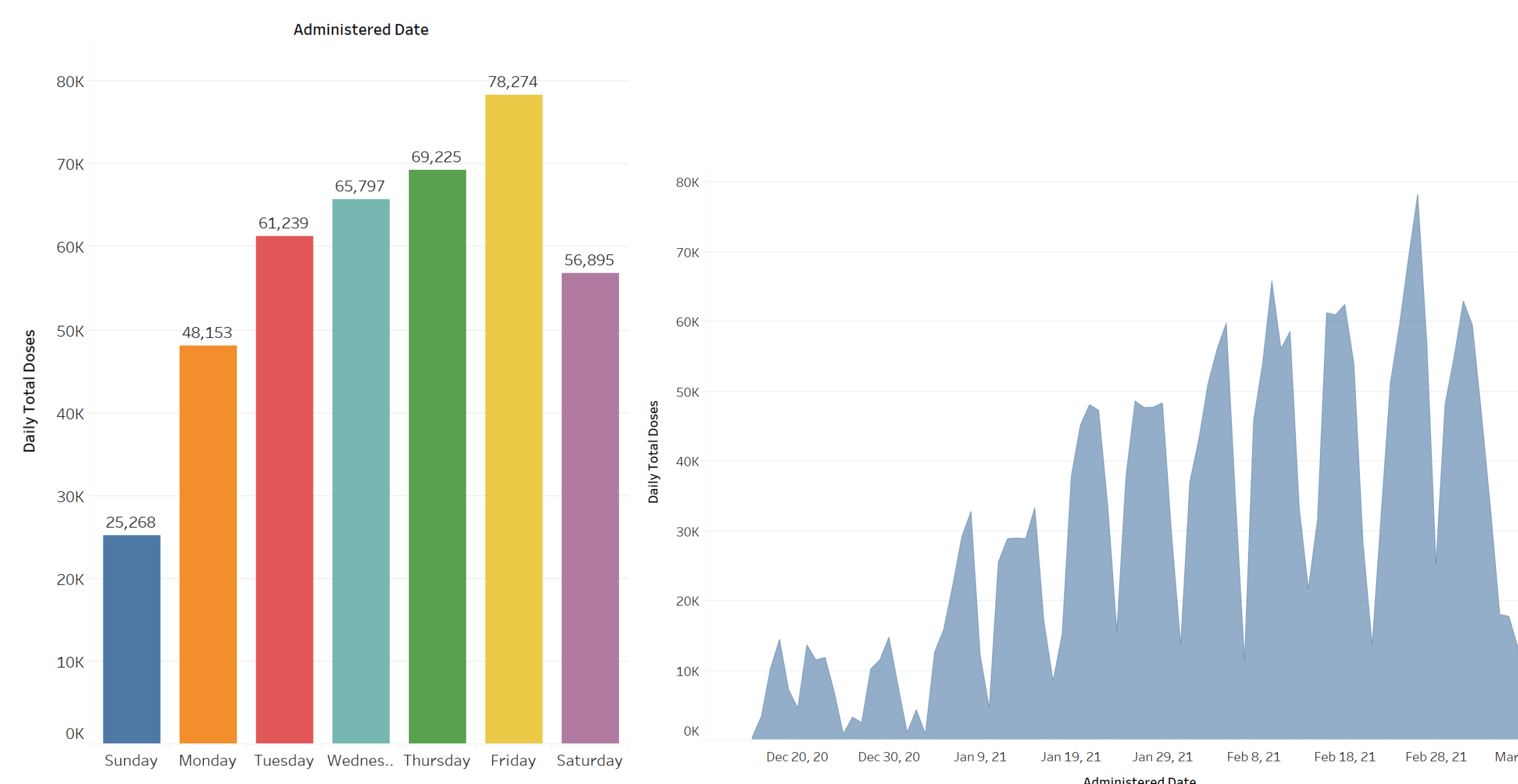
As of March 2021, Inyo county has the highest deaths to positive cases ratio. Meaning that of the people that tested positive for Covid, the percent who died was higher in Inyo county than other counties.



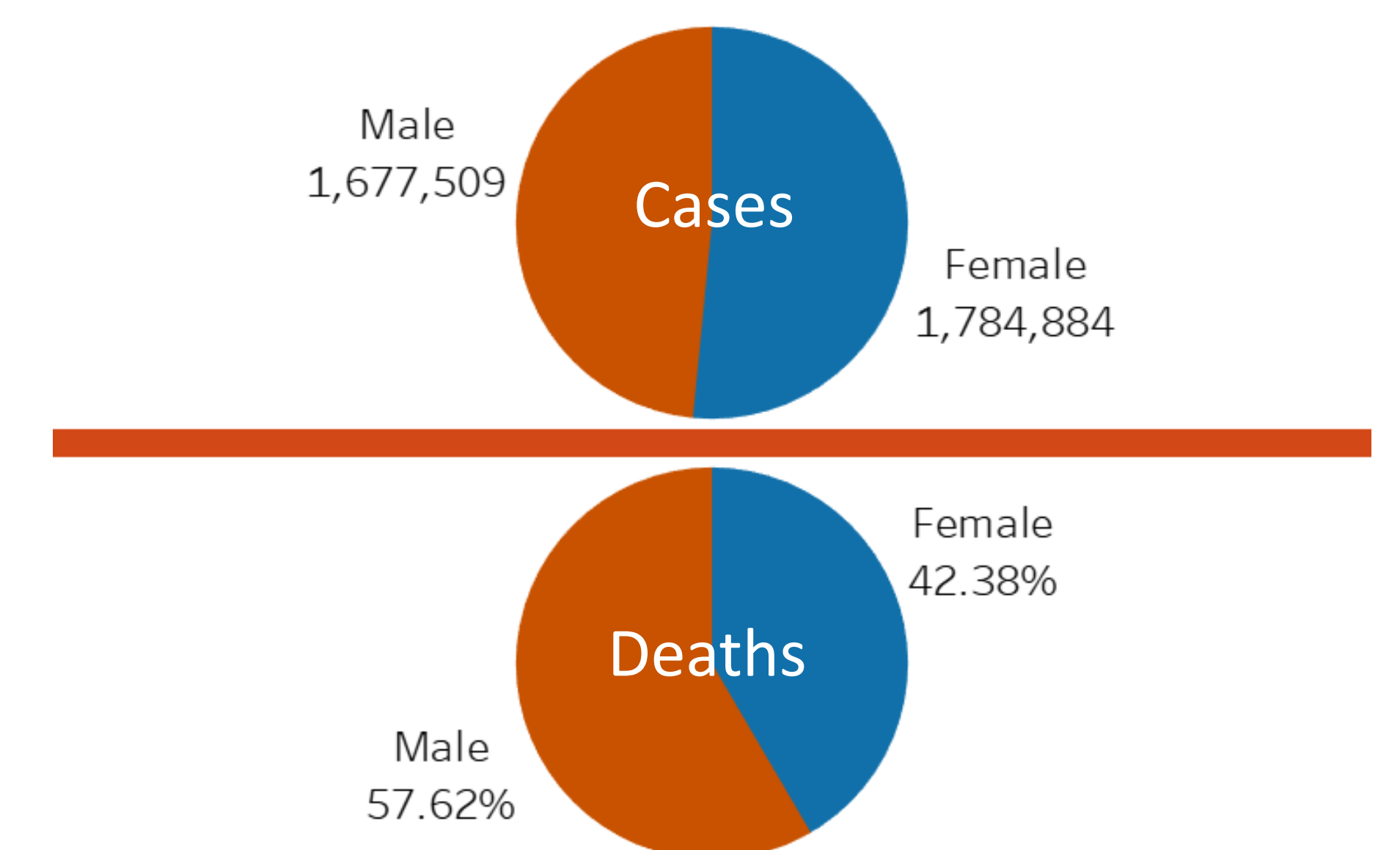
While the elderly had the least number of cases, they had the highest percent of deaths. Meaning the elderly were at much higher risk of death than others.



As of March 2021, Friday has become the most popular day for vaccine administration. We see weekends were not initially popular as we see large valleys during weekends followed by spikes on weekdays. This may be due to fewer sites open on weekends.



While Women lead in cases, Men lead in deaths. Meaning while women were more likely to get Covid-19, they were much less likely to die from it.



Implications

Within these descriptives we were better able to understand what happened within these counties over time. I was able to find counties that did well, and other counties that struggled. I was also able to establish how much more dangerous Covid-19 was to elderly people. I was also able to establish that men had a higher mortality rate over women, despite women's higher rate of testing positive for Covid-19. In addition, I was able to test socioeconomic and other demographic information against the Covid-19 performance of counties. While the results were not conclusive, I believe that the correlations can be better explored in the future.