Group 10

Group Project Report

Dhanu Saswanth Anandhakumar
Kabir Mehra
Zhongqi Hu (Zoe)
Shaun Benjamin Justin Solomon
Pranav Panchal

<u>Table of content</u>

Executive Summary

In the past three years, the one word people do not want to hear is "Covid-19". It has caused so many loss of lives worldwide. Not only were people suffering, but many animals were also hunted because of it. The global economy, public health system, food system, and work system were also severely impacted by Covid-19, and many industries are still recovering from the pandemic.

Although Covid-19 brought us so much pain and sad memories, the detailed recorded data give us so much information we can analyze, which can help us to understand it better. If similar events happen in the future, we can face them better.

For this project, our group took three different datasets from trustworthy websites to analyze COVID-19. Those datasets gave us some critical information about Covid-19, and we were able to make many assumptions and create visualizations using those data. Those assumptions are related to death rates, different age groups, underlying conditions, vaccines, and geography density. Some of the findings were completely out of our expectations, and others matched our daily knowledge about Covid-19. Overall, with this project, each member of our group was able to understand Covid-19 more, and we learned that we have to find supporting data before making a conclusion, otherwise, our assumption could be wrong.

Data Description

Data is the foundation of any kind of data analysis, and it has played an indispensable role during the Covid-19 pandemic. Researchers used different data collected during Covid-19 to make predictions and produce vaccines; governments use their predictions to take necessary steps for provision; people use different Covid-19 data to know cases, number of deaths, etc. For our project, we use trustworthy websites like the Centers for Disease Control and Prevention (CDC) as our main source of datasets.

We have used a total of three different datasets in our project. They are Provisional COVID-19 Death Counts by Week Ending Date and State, COVID-19 Vaccinations in the United States, Jurisdiction, and California coronavirus cases and deaths. Those datasets include information about Covid-19 deaths by age range, Covid-19 vaccination, number of deaths caused by Covid-19, etc. We used all three datasets to develop our hypothesis and then tested them out individually.

Data Cleaning

For all the datasets, we don't need to do any kind of data cleaning to get what we want from the data. All the data in the original dataset were organized the way we wanted. The basic filter functions in Tableau were the only tool we used to organize and clean the data.

General Introduction

This project aims to comprehensively analyze and interpret the COVID-19 death rate and vaccination data for the United States. Our analysis will cover various aspects, including the death rate between different states, comparison for different age groups, underlying condition effects on Covid-19, vaccination rate, impact on density, etc.

We came up with five different hypotheses using the three datasets we had. Those hypotheses are:

1. The percentage of deaths of people (death rate) in the state of California is higher than the percentage of deaths of people (death rate) in the state of Texas during the years 2020-2023.
2. In the state of California, people over the age of 65 have a higher number of deaths due to Covid-19 compared to people aged 65 and less from 2020 to 2023, from the first to the fourth quarter of each year.
3. In the state of California, the number of people having underlying conditions (people who have both Pneumonia and Covid-19) has a higher number of deaths compared to people who were only affected by Covid-19 in Quarter 1 of 2020.
4. In the state of California, the vaccination rate helped decrease the death rate of people from 2021 to 2023.
5. In the state of California, there was a positive correlation between population density and the number of deaths due to Covid-19, during the years 2020 to 2023.

Insights and Findings

1. To filter out which states in the United States our group should focus on, we first want to see each state's death rate and then determine our focus based on the result. We did that by using our first dataset, putting either the total number of deaths or just deaths caused by Covid-19 into Tableau. The visualization below compares the death rate only caused by Covid-19 in each state.
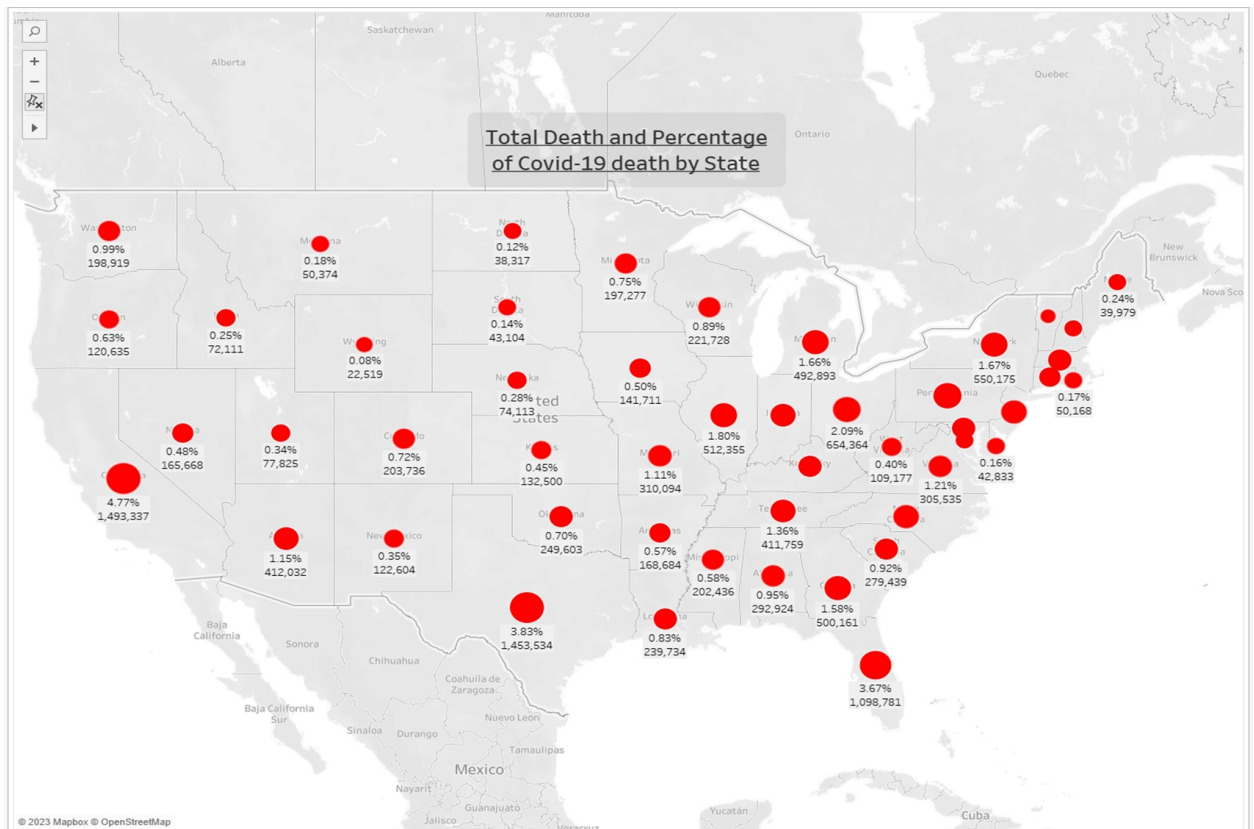
The above visualization shows each state's death rate (in percentage). The larger the circle is, the higher the rate is. We can see that some states (like California, Texas, Florida, etc.) have a pretty high death rate compared to other states, which tells us that they have been more severely affected than other states. We can see that Texas only has a

death rate of 3.83%, which is high compared to other states but not as high as California, which has an even higher death rate of 4.77%.
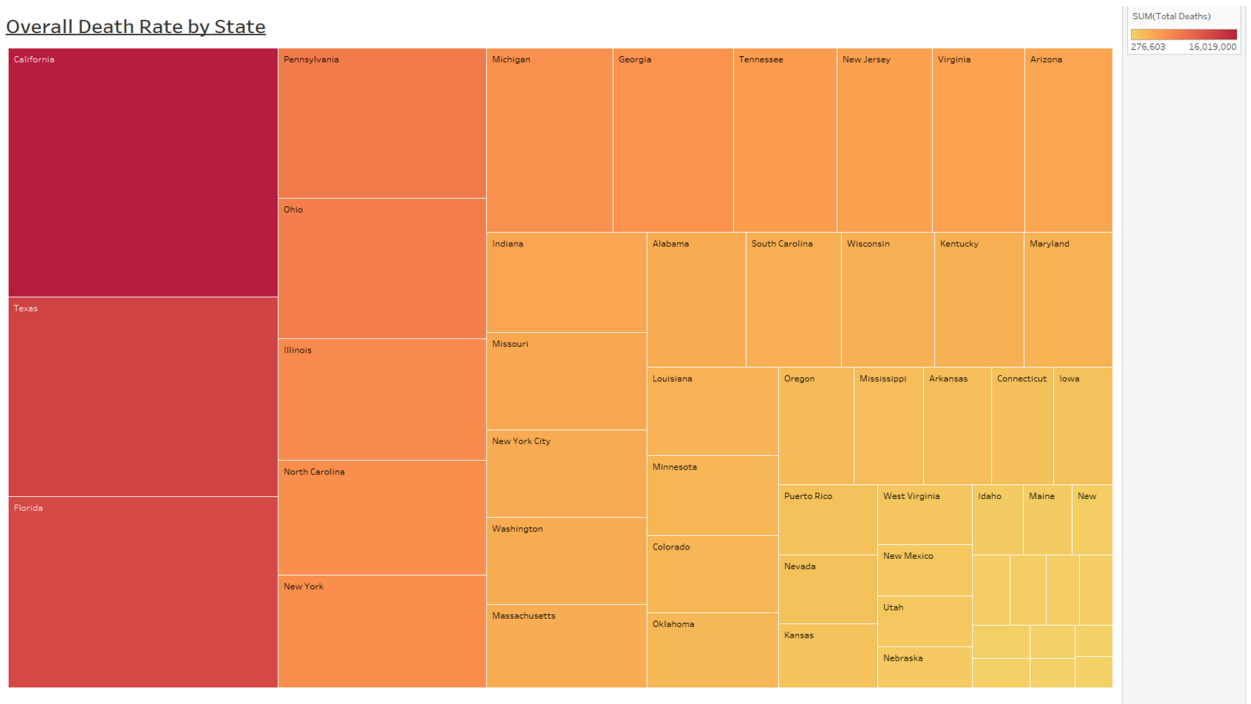

In our first data set, different records were collected about people who already had Pneumonia and got Covid-19 later. There are records about the number of people who passed away with only Covid-19 or only Pneumonia and records about the number of people who passed away with both Covid-19 and Pneumonia. So, we also want to extend our research to all deaths and see if California still has the highest death rate. So, we did a state heat map for all deaths (including all underlying conditions). In this visualization, the bigger the box size, the higher the death rate. We can see that California ranked top on the map, and Texas ranked second. This tells us that these two states' death rate rankings stay the same, even with underlying conditions counted.

2. The first visualization talks about the COVID-19 death rates across different age groups in California from 2020 to 2023. In 2020, the state witnessed its highest death rates, with the elderly population aged 65 and above getting affected the most. However, through the years 2021,2022 and 2023 there was a gradual decline in death rates across all age groups. Despite this trend, the age group 65 and above still had the highest death rate. The insights drawn from this chart showcases the vulnerability of individuals aged 65 and above, and asks for the need for more protection for this age group.
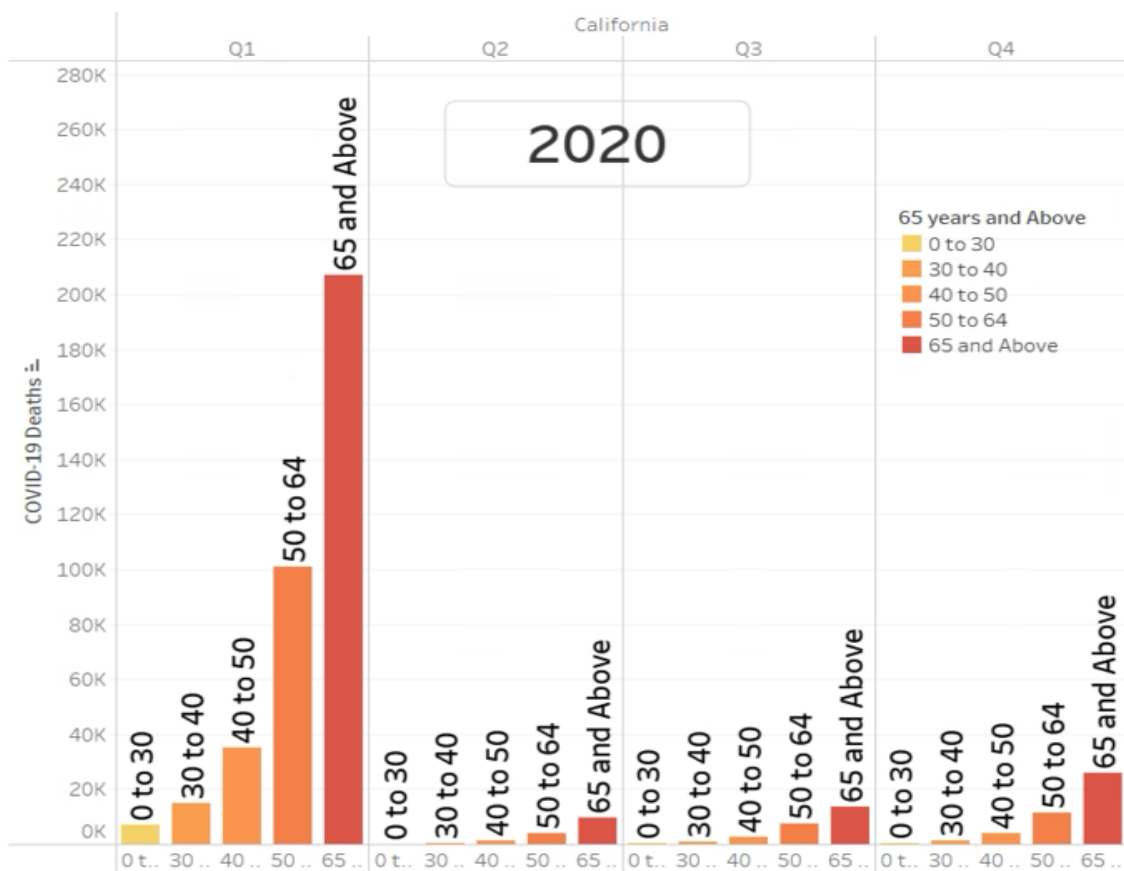

The second visualization is a pie chart, offering a picture of COVID-19 death distribution between individuals aged 65 and above versus those aged 64 and below. Ages 65 and

above constitute 56.44% of the total COVID-19 deaths in California. This highlights a big contrast in effects of Covid-19 between the elderly and younger populations. The fact that more than half of the deaths occurred in the 65 and above age group shows that there is a big amount of burden and vulnerability on the older generation. These findings collectively show the need for prioritizing the elderly, and supporting initiatives to make sure that the most vulnerable can be protected.

Covid Death Rate - 65 years and Above (Vs) 64 years and Lower

Age Range
- 64 and Lower
- 65 and Above

California   Q1

294,984
43.56%

382,184
56.44%
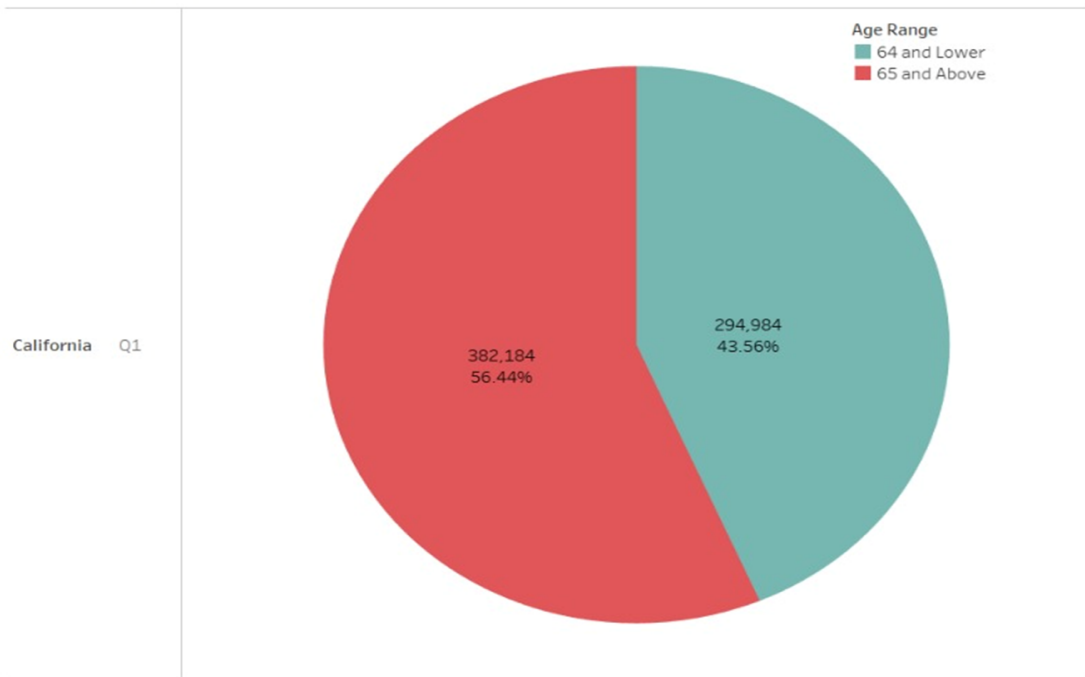
3. To visualize this hypothesis, we plotted an animated bar chart. Each bar represents different age groups, starting from ages 0-17 till ages 85 and above. The expectation was that as the age range increased, the bars for individuals with both COVID-19 and Pneumonia would grow taller, indicating a higher number of deaths.
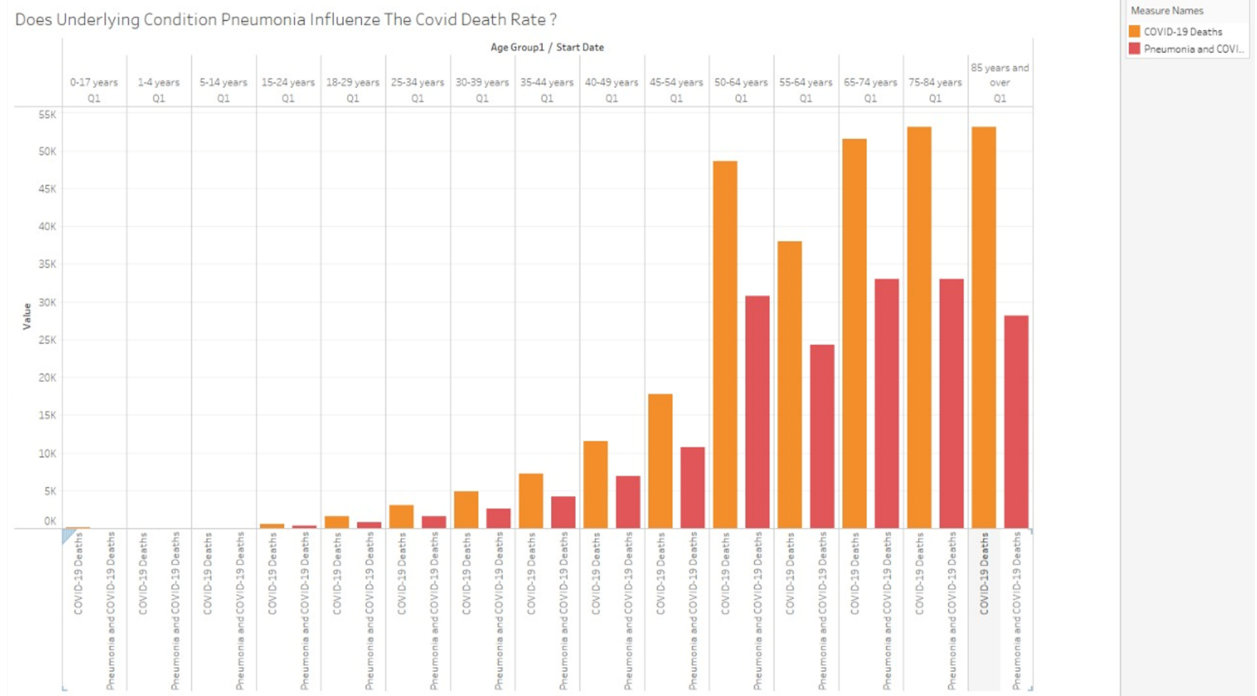
However, the animation of this visualization tells a different story. The numbers and bars don't align with the expectation in the years 2020 through 2023. The animation clearly depicts that individuals with underlying conditions, particularly those with both Pneumonia and COVID-19, do not have a higher number of deaths compared to individuals affected by COVID-19 only

4. For the visualization below, we took a count of vaccines and the Covid-19 death count. We used both line and bar charts together. In 2021, the death rate was high at first. But as the years went by and more people had their shots and immunization campaigns advanced, a striking pattern became apparent. The death rate was shown to be gradually declining over time in the line chart. This decrease can be attributed to the growing number of individuals developing COVID-19 immunity as a result of immunization. Growing vaccination rates are positively correlated with declining death rates, indicating that immunization efforts are effective in reducing the impact of the virus.

The positive correlation between increasing vaccination rates and decreasing death rates provides compelling evidence to support the notion that immunization campaigns are a key factor in shaping the trajectory of the pandemic.
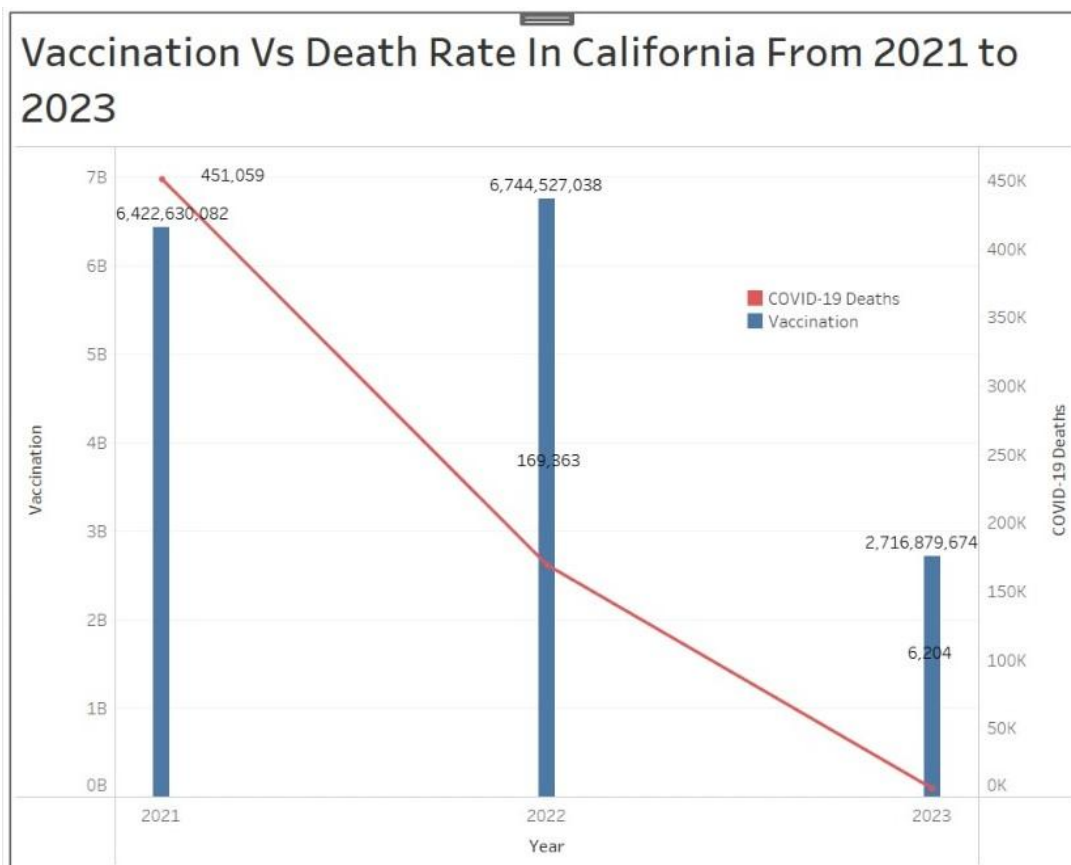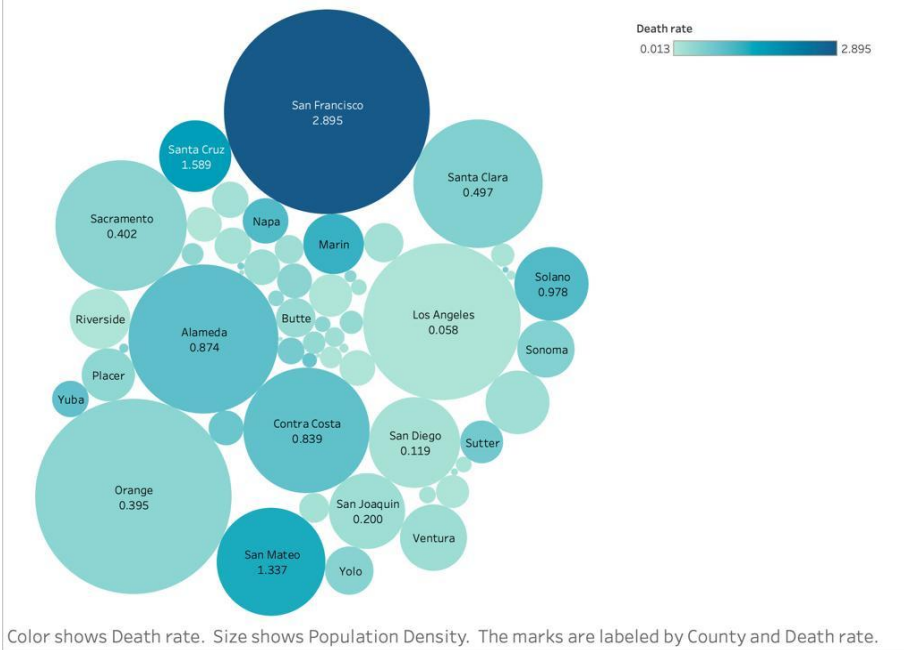
5. Our investigation involved a visual analysis wherein we plotted the death rate of California counties against their respective population densities. Initially, we hypothesized that densely populated areas, potentially having lower adherence to social distancing, might exhibit higher death rates. However, the visualized data challenges this hypothesis.

Although San Francisco, the most densely populated county, does have the highest death rate, a mixed pattern in death rates is observed among other counties with higher population density. Additionally, several counties with lower population density also display low death rates.
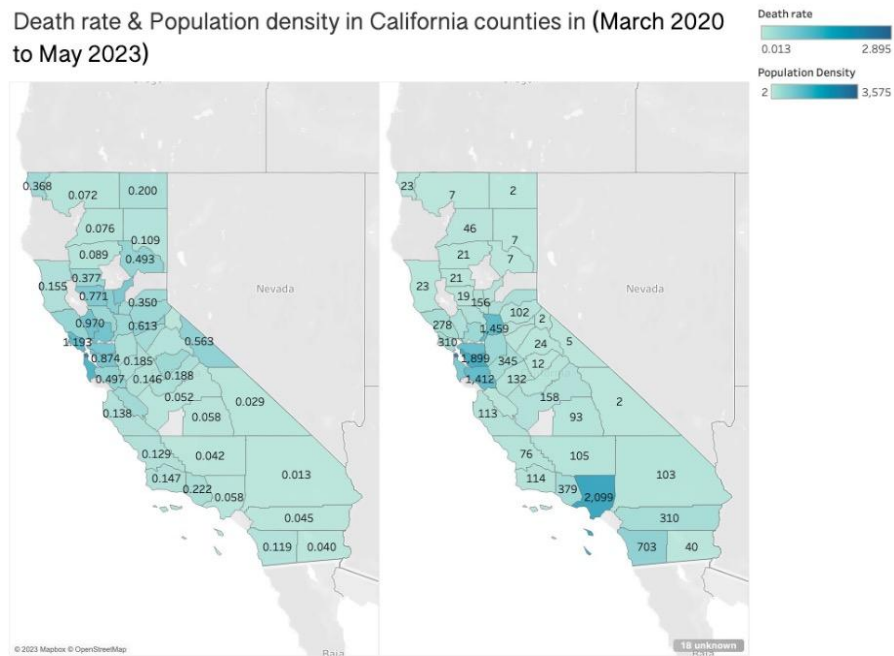
The absence of a consistent positive correlation between population density and death rate suggests that other variables may play a significant role in influencing outcomes. It is clear that a more comprehensive analysis, incorporating factors such as healthcare infrastructure, vaccination rates, and socioeconomic conditions, is essential for a nuanced understanding of the observed patterns.

Consequently, our initial hypothesis is not conclusively supported. This underscores the necessity for further investigation and the inclusion of additional variables to ensure a more accurate interpretation.

Death rate & Population density in California counties in (March 2020 to May 2023)

Color shows Death rate. Size shows Population Density. The marks are labeled by County and Death rate.

Death rate & Population density in California counties in (March 2020 to May 2023)

Map based on Longitude and Longitude and Latitude . Details are shown for County. For pane Longitude : Color shows sum of Death rate. For pane Longitude (2): Color shows sum of Population Density.

Conclusion

In our investigation of COVID-19 data, we first looked at death rates in different U.S. states and found that California and Texas had higher rates. Yet, when we considered all deaths, including those with underlying conditions, California consistently ranked highest. Shifting focus to age groups, we discovered that individuals aged 65 and above consistently experienced the highest death rates, highlighting the importance of specific protection measures for this group. Our analysis challenged the common belief that the presence of both COVID-19 and Pneumonia would lead to more deaths, as animated data showed this wasn't the case at least during our study period. Additionally, we noted a positive link between increasing vaccination rates and decreasing death rates, signaling the effectiveness of immunization efforts in shaping the pandemic's course.

In another study, we delved into the connection between death rates and population density in California counties. Although our initial assumption that densely populated areas would have higher death rates was somewhat supported by San Francisco, the overall pattern was inconsistent. Many densely populated counties displayed varied death rates, and some less populated areas had low death rates. This complexity suggested that factors beyond population density, such as healthcare infrastructure and socioeconomic conditions, played crucial roles. In conclusion, our findings highlight the need for a more thorough analysis, including additional variables, to achieve a nuanced understanding of the observed patterns and guide effective public health strategies.