Vaccination rates and COVID-19 cases and deaths in California

Carmen Chen

12/2/2021

This is my PM566 Final Project website home.

The website is online **here** (https://carmenchenepi.github.io/PM566-Final-Project)

Link to **download the PDF report** (https://raw.githubusercontent.com/CarmenChenEpi/PM566-Final-Project/main/report.pd)

Introduction

The vaccinations have begun almost a year since its distribution. The fully vaccinated percentage in California was increasing and stagnated around 60%. However, the herd immunity didn't seem to be achieved and the COVID-19 daily new cases is still very high in California. Therefore, this project aims to investigate the relationship of vaccination rates on COVID-19 daily new cases and deaths in California by using the data from California Health & Human Services Agency and US Center for Disease Control and Prevention.

Methods

Data of COVID-19 cases and vaccination rates were acquired from California Health & Human Services Agency and US Center for Disease Control and Prevention, respectively. Data in California and with variables of interest (i.e., date, county, percentage of first dose, percentage of second dose, daily new cases, cumulative cases, daily new deaths, and cumulative deaths) were extracted. The date variables were formatted in both datasets.

Dimensions, headers, and footers of the two datasets were checked. There are 311 observations and 3 variables in the "vaccinations" dataset, as well as 628 observations and 5 variables in the "cases" dataset. Implausible data (e.g., 0 cases increase) was found in the date variable on "2021-10-20" in the "cases" dataset. Considering the 14-day incubation period of the COVID-19 disease, the data from 2021-10-06 to 2021-10-20 were not the final accurate number of cases and deaths since there are still many cases and deaths were not reported timely. Thus, these data were removed from the "cases" dataset.

The two datasets were merged into one dataset by date variable. Final dataset has 296 observations and 7 variables. Exploratory data analysis was conducted in the merged dataset. No missing value, implaussible vaule or data error was found. The data includes COVID-19 partial and fully vaccination rates, daily new cases, cumulative cases, daily new deaths, as well cumulative deaths from 2020/12/14 to 2021/10/05. Both univariate and bivariate summary statistics was analyzed. Exploratory graphs were generated between vaccination rates and cases and deaths.

Preliminary Results

Table 1 presented the summary statistics of the data, including range of the date and median (IQR) of first dose, second dose, daily new cases, cumulative cases, daily new deaths, and cumulative deaths. There are a total of 296 observations collected from 2020-12-14 to 2021-10-05. The maximum partial and fully vaccination rates in California are 52% and 64%, respectively. The medians (IQRs) of the daily new cases and deaths are 4275 (1810, 11520) and 64 (23, 159), respectively.

Table 1. Characteristics of the COVID-19 data

Characteristic	N = 296[1]
Date	2020-12-14 to 2021-10-05
First dose, %	52 (15, 64)
Second dose, %	36 (0, 52)
Daily new cases	4,275 (1,810, 11,520)
Cumulative cases	3,669,089 (3,512,726, 3,833,183)
Daily new deaths	64 (23, 159)
Cumulative deaths	62,710 (58,891, 63,905)

[1] Range; Median (IQR)

Table 2 presented the correlation coefficients of vaccination rates with daily new cases and deaths. Vaccination rates were negatively associated with daily new cases and deaths. Such negative association was stronger in the correlation between vaccination rates and daily new deaths (R: -0.78 for first dose, -0.68 for second dose).

Table 2. Correlation coefficients of vacciantion rates and cases/deaths

	First dose	Second dose
Cases	-0.4990338	-0.3931054

	First dose	Second dose
Deaths	-0.7814445	-0.6816794

Exploratory graphs were presented in figure 1 and figure 2a-2e. Vaccinations in California started in February and the rates continued to increase. The increase of vaccination rates became more slowly when it achieved around 60% of first dose vaccination rate. The daily new cases started to decrease drastically around February and the cases remained in a stable small number until July. A small break out in daily new cases occurred in July and it achieved its peak in September. The trend pattern of the daily new deaths is similar to daily new cases.

Figure 1. Vaccination rates from 2020-12-14 to 2021-10-05

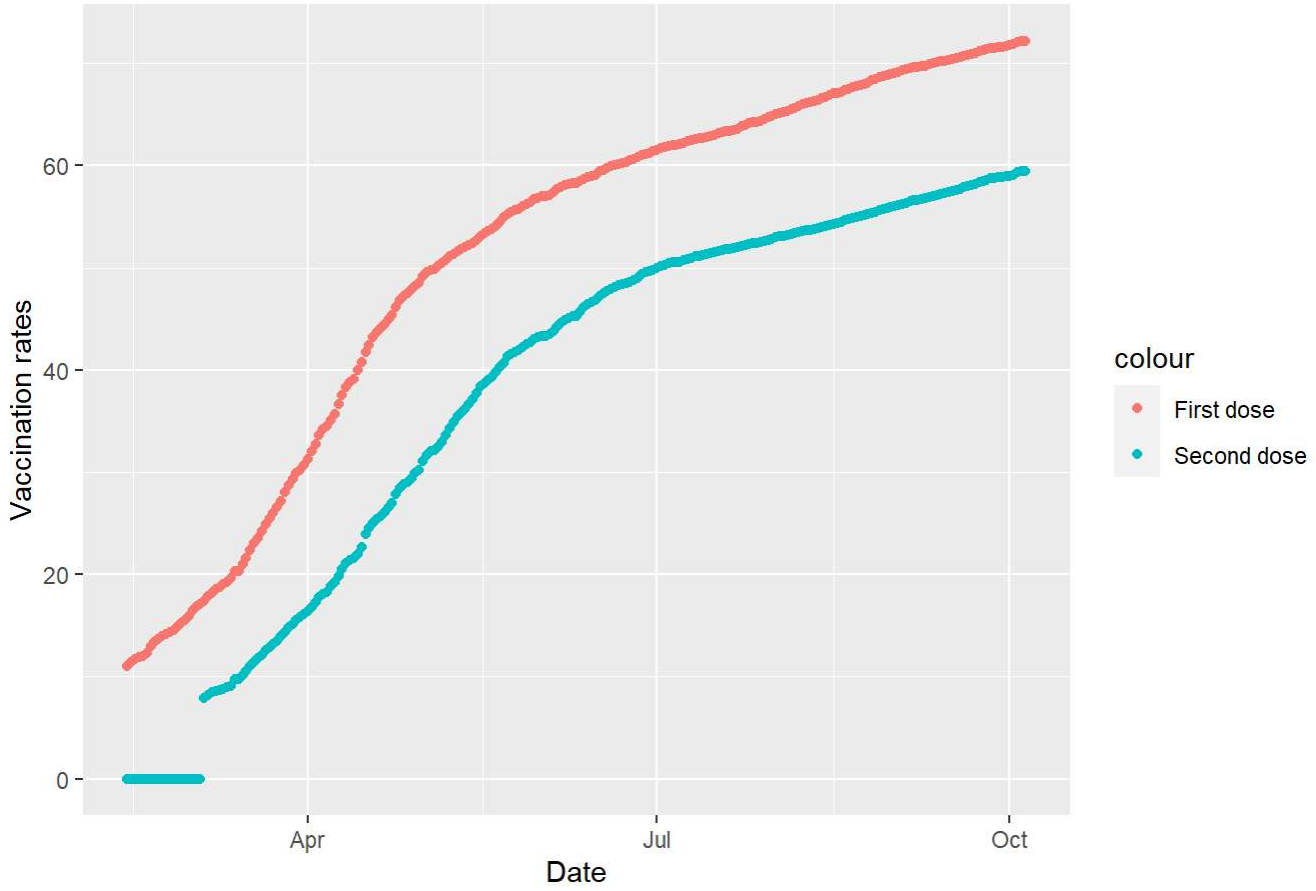


Figure 2a. Daily new cases from 2020-12-14 to 2021-10-05

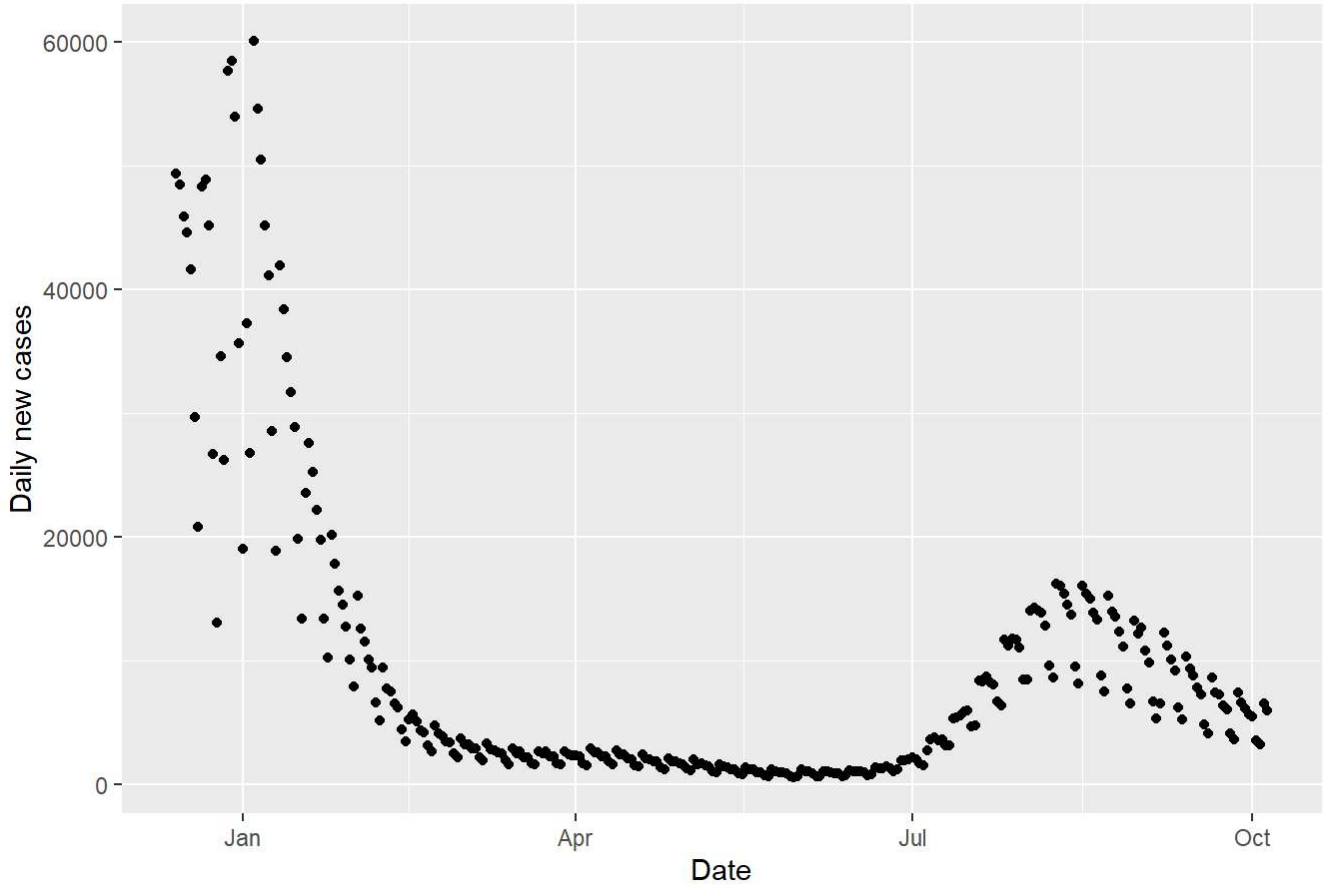


Figure 2b. Daily new deaths from 2020-12-14 to 2021-10-05

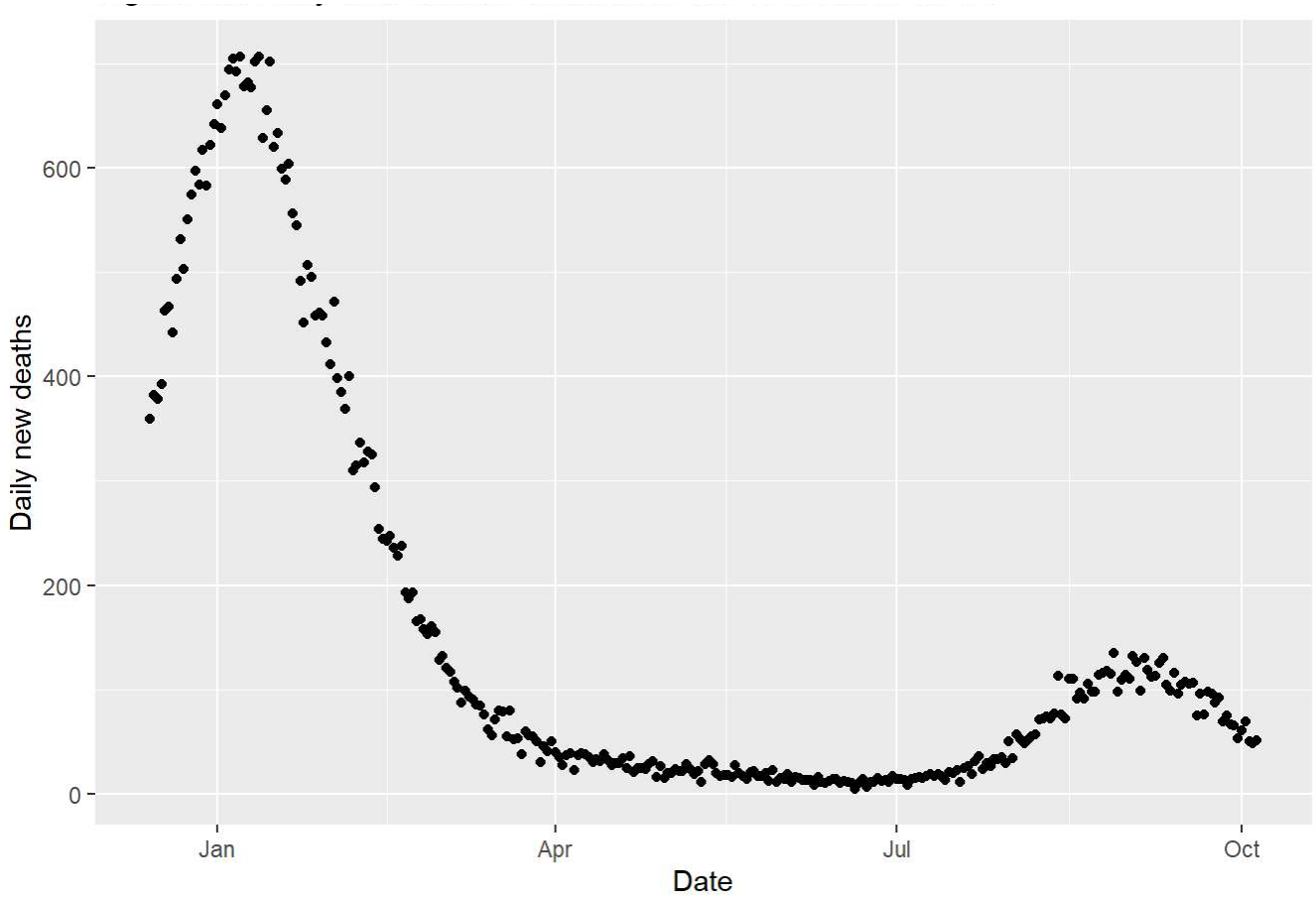

Daily new deaths
600
400
200
0
Jan
Apr
Jul
Oct
Date

Figure 2c. Cumulative cases from 2020-12-14 to 2021-10-05

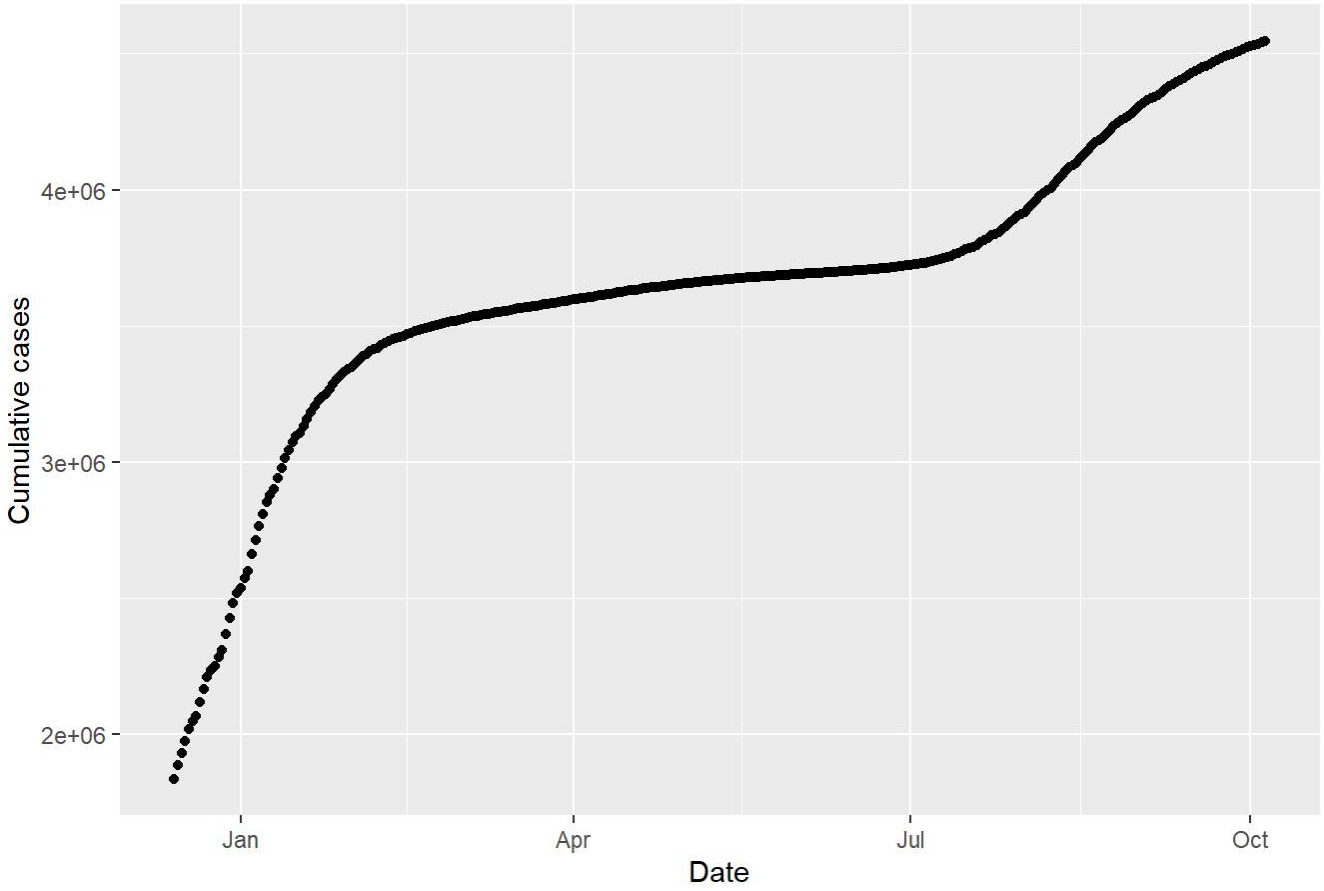
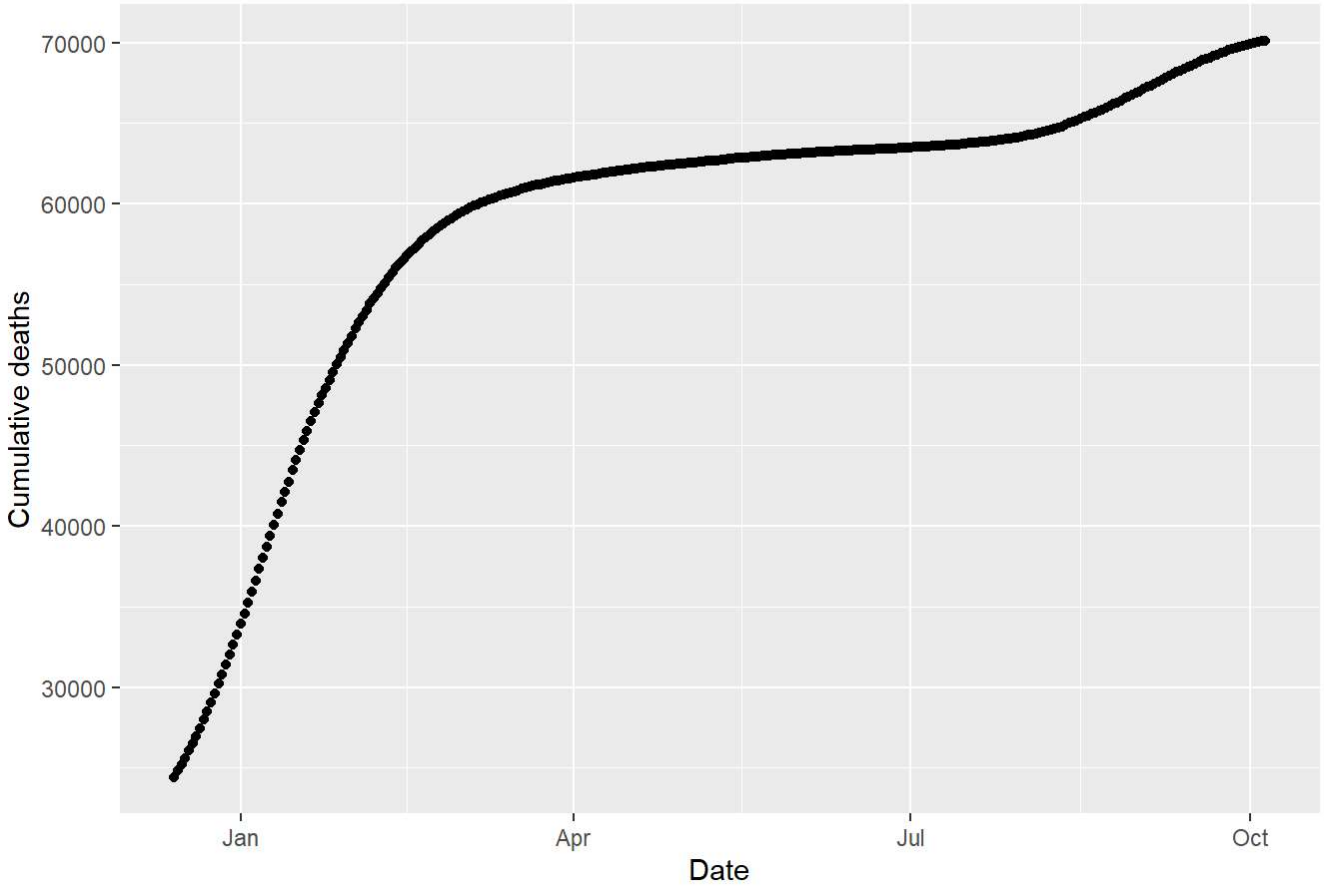


Figure 2e. Cumulative deaths from 2020-12-14 to 2021-10-05



Data visualization graphs were generated in Figure 3a & 3b to visualize the relationship between vaccination rates and daily new cases and deaths. There was an non-symmetrical inverse pattern in the relationship of vaccination rates and daily new cases in the beginning. Daily new cases was negatively associated with vaccination rates until the first dose vaccination rates achieved around 50%. After that, daily new cases increase drastically with the increase of vaccination rates. This same pattern was also found in the relationship of vaccination rates and daily new deaths. However, a stronger negative association was observed in the daily new deaths figure in the beginning of the data.

Figure 3a. Vaccination rates and cases

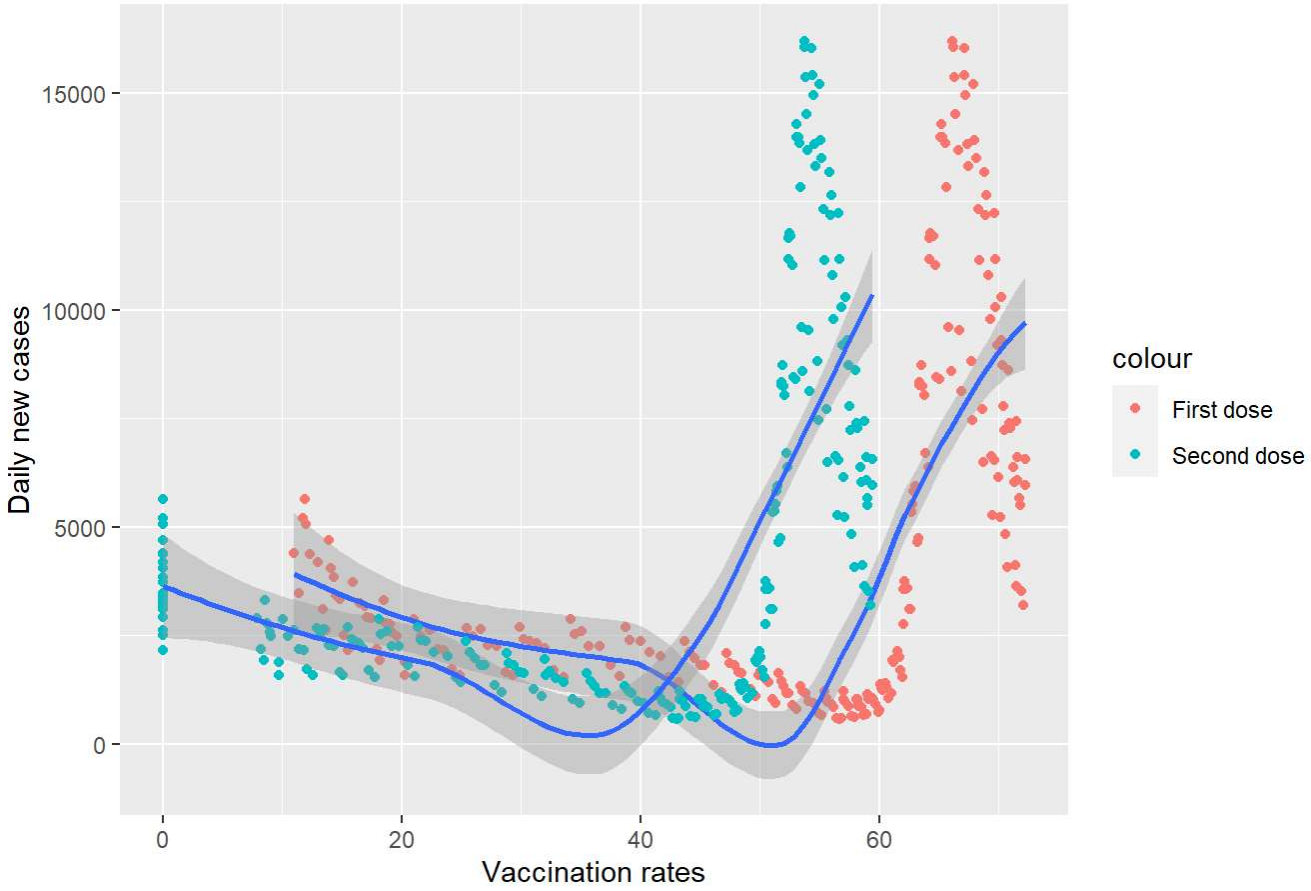
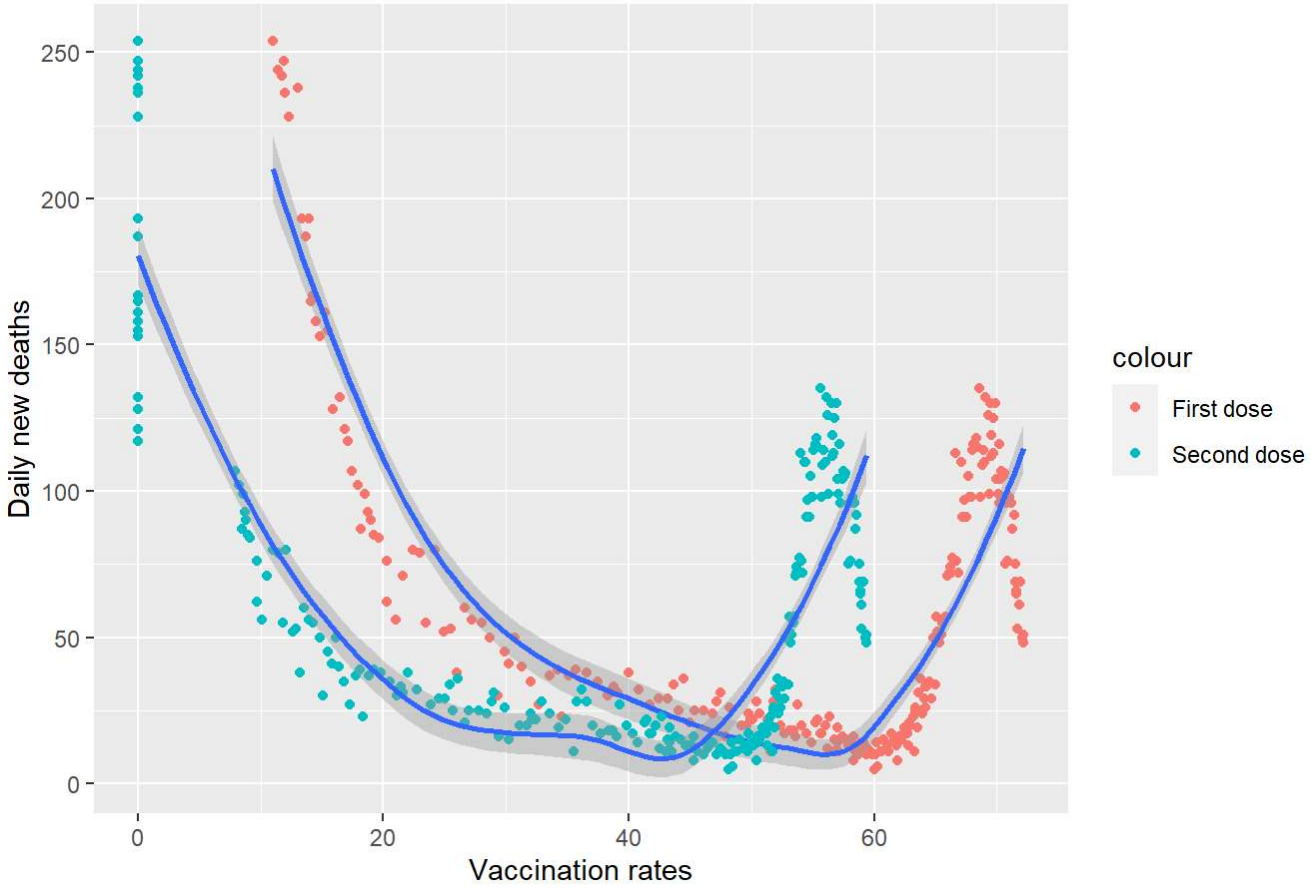
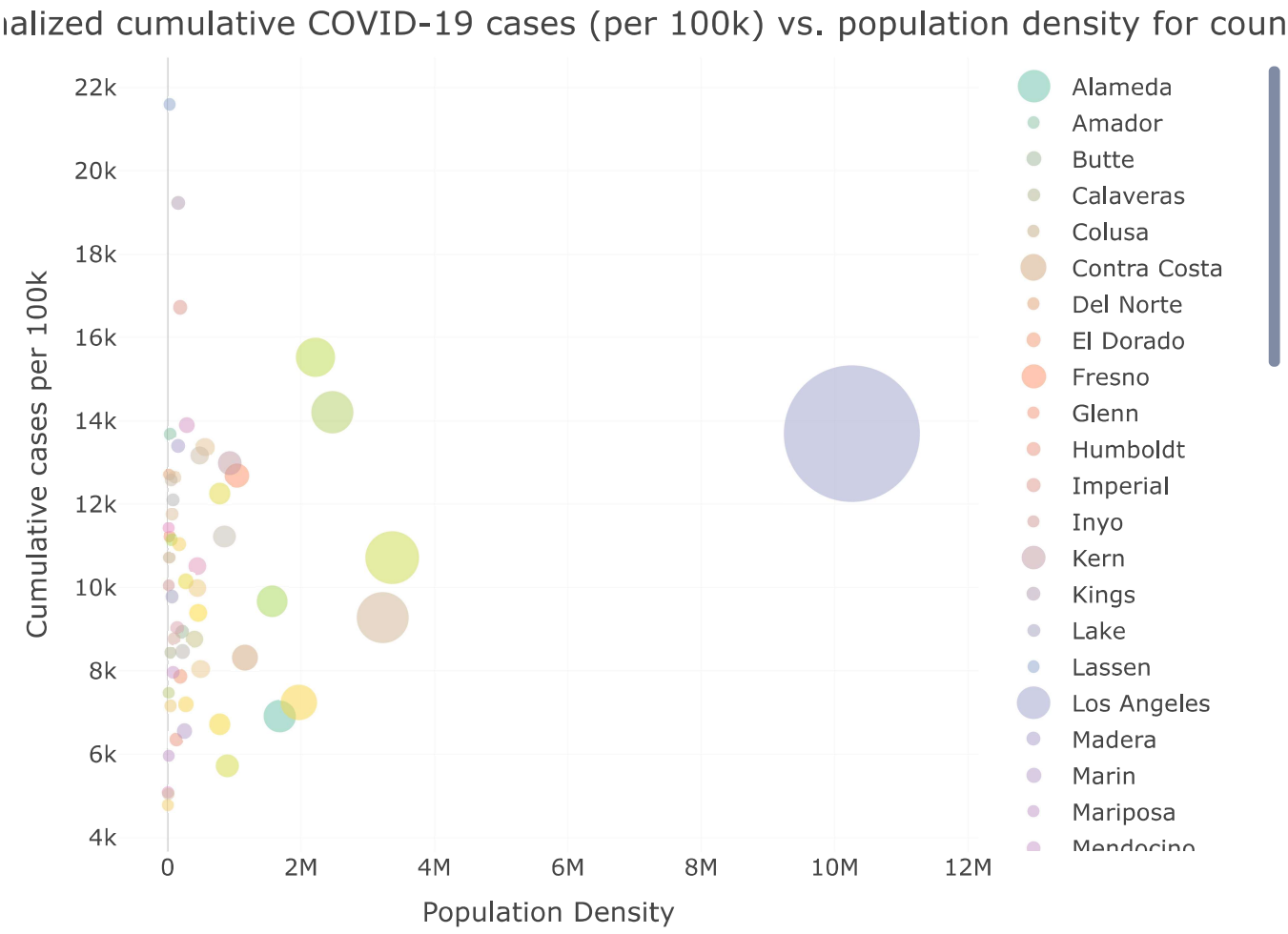


Figure 3b. Vaccination rates and deaths

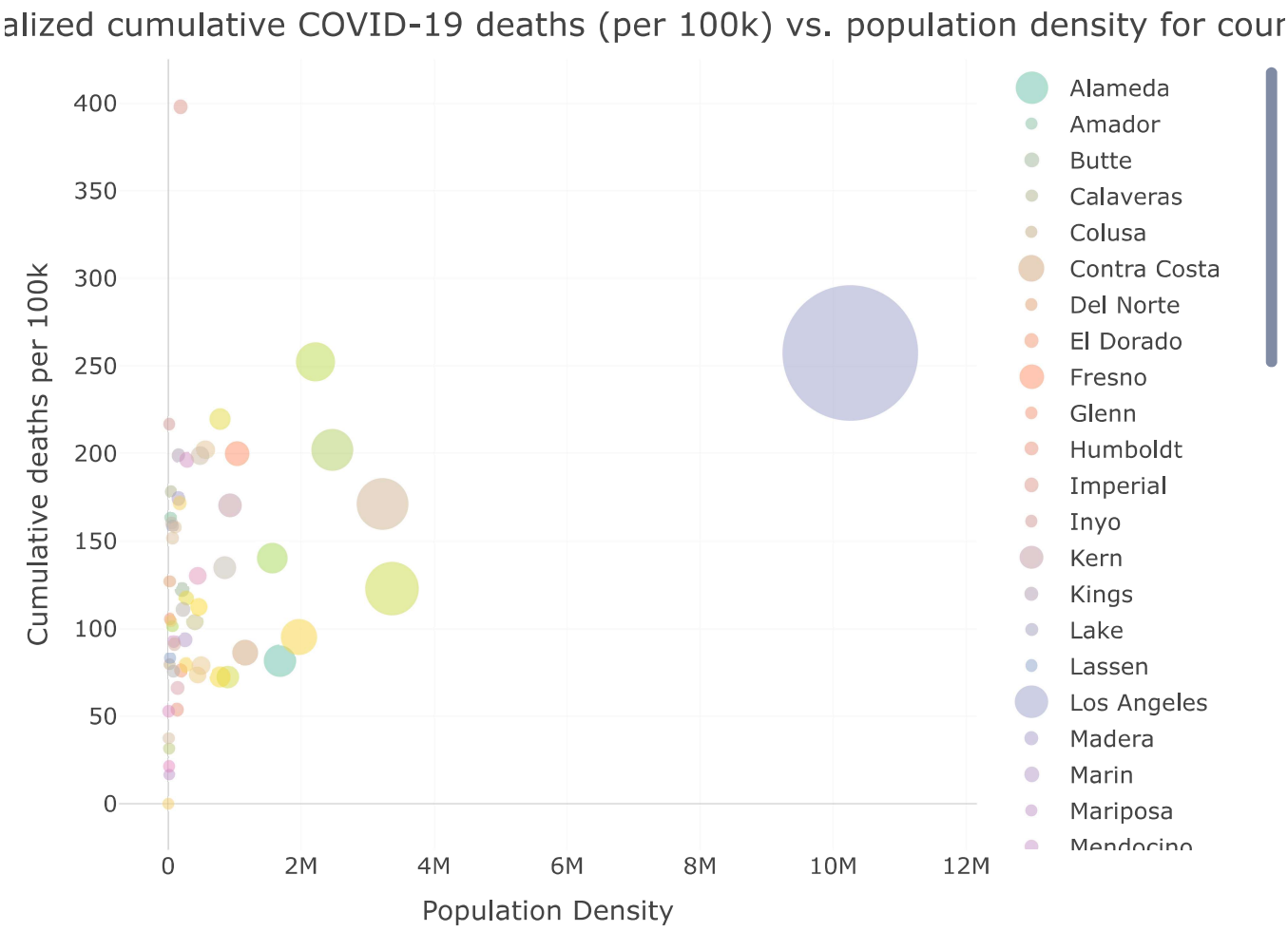


Interactive visualization plots were created to observe the cumulative COVID-19 cases and deaths, as well as vaccination rates across different counties in California. Accourding to the interactive plot, there is no association between the cumulative COVID-19 cases per 100k population and population density observed. Lassen, Kings, and Imperial county are the three counties with the most cumulative COVID-19 cases per 100k population, following with San Berndino, Riverside, and Los Angeles. Imperial, San Berdino, and Los Angeles are the three counties with the most cumulative COVID-19 deaths per 100k population, following with San Joaquin, Inyo, and Riverside. Lassen, Tehama, and Del Norte are the three counties with the least vaccination rates of less than 40% fully vaccinated people, following with Kings and Shasta.

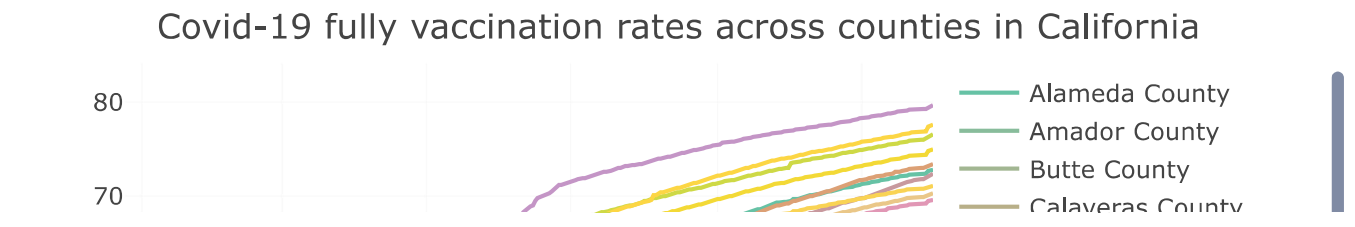
```
## Warning in RColorBrewer::brewer.pal(N, "Set2"): n too large, allowed maximum for palette Set2
is 8
## Returning the palette you asked for with that many colors

## Warning in RColorBrewer::brewer.pal(N, "Set2"): n too large, allowed maximum for palette Set2
is 8
## Returning the palette you asked for with that many colors
```

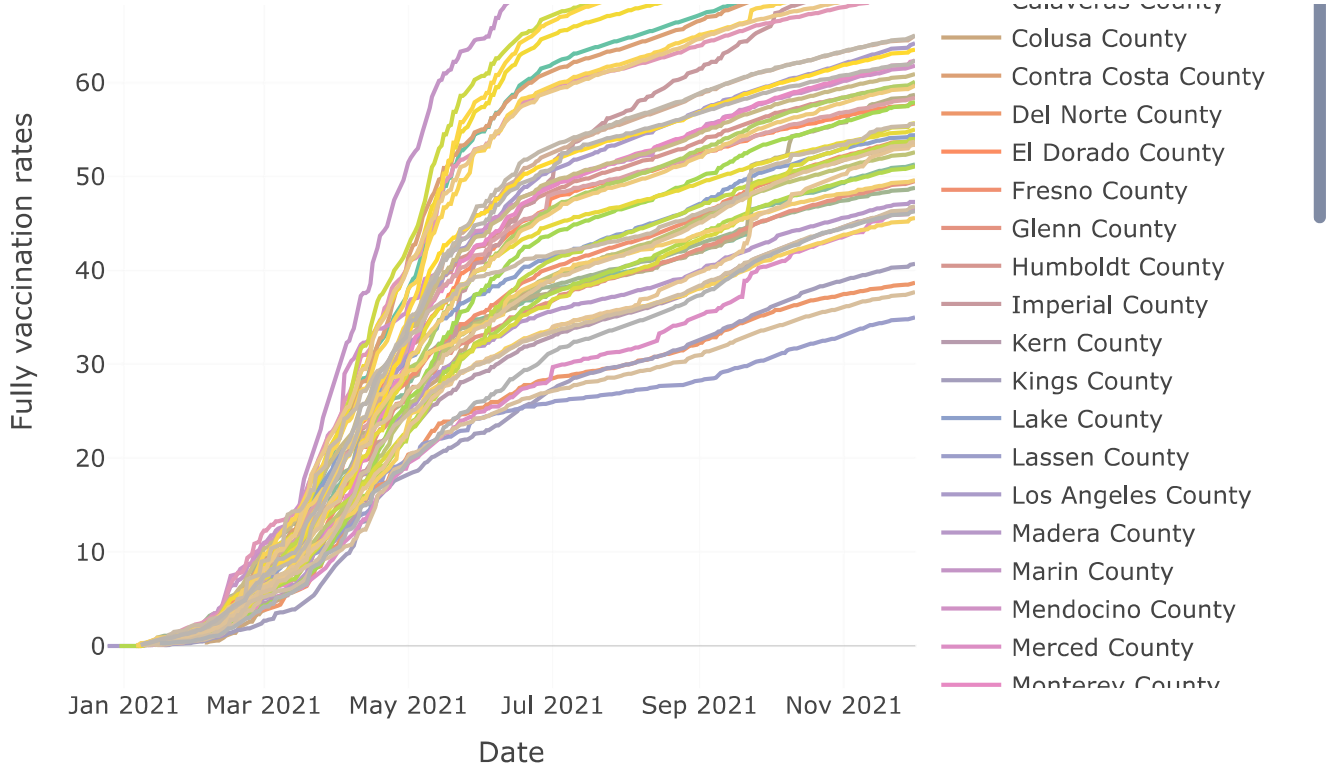


```
## Warning in RColorBrewer::brewer.pal(N, "Set2"): n too large, allowed maximum for palette Set2
is 8
## Returning the palette you asked for with that many colors

## Warning in RColorBrewer::brewer.pal(N, "Set2"): n too large, allowed maximum for palette Set2
is 8
## Returning the palette you asked for with that many colors
```




```
## Warning in RColorBrewer::brewer.pal(N, "Set2"): n too large, allowed maximum for palette Set2
is 8
## Returning the palette you asked for with that many colors

## Warning in RColorBrewer::brewer.pal(N, "Set2"): n too large, allowed maximum for palette Set2
is 8
## Returning the palette you asked for with that many colors
```





Conclusion

There is a positive association of vaccination rates and daily new cases and deaths when the first dose of vaccination rate achieved around 50%. This may be due to the re-opening of the economic and lift of mask mandate during that time. Overall, we could see the protective effect of vaccine towards infection and death according to the data in the beginning. A stronger negative association in the beginning of vaccination rates and daily new deaths compared to daily new cases may be due to a stronger efficacy of the vaccine towards preventing mortality.