**Average COVID Cases in California and Florida Throughout the COVID Pandemic and In Response to Key Policy Decisions**

**Introduction:**

In our assignment, we explored the question how does the seven-day average of COVID cases vary over time in California and Florida in response to key mitigation policies enacted at the state level? We chose to compare California and Florida as they are states that had seemingly different legislative responses to the COVID pandemic.

**Data Sources:**

In order to explore this information, we utilized five different datasets: one including COVID cases and deaths data by state throughout the COVID pandemic, one including 2019 population size by state, one including the dates when each state enacted and ended stay-at-home orders, one including the dates when each state enacted and ended mask mandates, and one which included vaccination rates by state throughout the pandemic. The COVID cases and deaths data as well as the vaccination data was retrieved from the CDC COVID Data Center. The population data was retrieved from the U.S. Census Bureau, and the stay-at-home order data was retrieved from the COVID-19 U.S. State Policies Database.

**Variables:**

In our analysis, we utilized the following variables:

* New COVID cases per day in California and Florida from the COVID dataset
* Stay-at-home beginning dates and ending dates in California and Florida from the stay-at-home dataset
* Mask mandate beginning and ending dates in California and Florida from the mask dataset
* Number of complete series vaccinations and the corresponding dates in California and Florida in the vaccine dataset
* Total population counts for California and Florida from the population dataset

We focused on these variables as they were necessary for our research question of understanding how average cases fluctuated in California versus Florida throughout the pandemic and in response to different mitigation strategies. Specifically, we wanted to understand how average cases fluctuated in response to policy decisions, such as instituting and lifting stay at home orders, instituting and lifting mask use mandates, state-wide vaccination campaigns, and vaccine requirements for employee groups; all decisions which differed between California and Florida.

**Methods:**

**Datasets and Variables**

Our first step in our project was to decide on a research question, mentioned above, and then to pick out relevant datasets. As explained, we chose the aforementioned datasets focusing on vaccinations, stay-at-home orders, and mask mandates because they focus on policies in states with contrasting COVID mitigation strategies. We felt these policy-related variables encompassed the full array of policy decisions from which a state government could choose to prevent further spread of COVID.

**Analysis**

In our code, we first imported the relevant datasets and sorted each so they could be properly merged. We then calculated the 7-day average of COVID cases. We chose to report COVID cases instead of COVID deaths because we felt cases tell a better story about the state of the COVID pandemic. Cases fluctuate immediately depending on how contagious the variant is whereas deaths are often a lagging indicator and tend to be concentrated within certain population groups. Additionally, we chose to report a 7-day average of cases versus any other statistic because we felt that the 7-day average was a more accurate measure of the rapid change in infections. While using an average makes our data susceptible to outliers, we felt it was a more representative option than reporting daily or cumulative cases. Finally, choosing a larger time period, such as a monthly average, would have distorted the rapidly changing infection rate and would not allow for visualization of small peaks in cases over time.

After calculating our 7-day average in the code, we then merged this metric with the dataset and populated within the dataset the dates when both California and Florida began and ended stay-at-home mandates and mask mandates, as well as when each state achieved a 25% vaccination rate and a 75% vaccination rate among state residents. We chose to report on the 25th and 75th percentiles of vaccination rate because the 25th and 75th percentiles are well-established statistical measures when analyzing the spread of data. We felt that reporting on the 25th and 75th percentiles would allow us to encompass the majority of the population within our visualization.

Then, we plotted the 7-day average of cases on scatter plots to visualize just the fluctuations in average cases in each state over time. This allowed us to gain an understanding of how the states differed in trends of increasing or decreasing COVID cases and how many times each state experienced a pronounced spike or dip in average cases.

Finally, we plotted our 7-day cases average and reference lines for vaccinations and policy decisions on a LOESS curve for each state from January 1, 2020 through January 31, 2022. This was the final output of our project and is further explained below in our results section. We chose to display our analysis via a LOESS curve because these curves tend to be easier to read than a scatter plot by muting noisy data and are useful when plotting data over time.

**Results:**

Our results are displayed below in the figures titled *COVID-19 7-Day Average Cases in California Y20-22* and *COVID-19 7-Day Average Cases in Florida Y20-22*.  As can be seen in the figures, overall, there wasn’t a marked difference observed between California and Florida across the 2 years. There were, however, some peaks identified in each of the states which we will try to describe and explain.

**Summer 2020**

Florida experienced a bit of a higher spike in average cases in late summer 2020 than California and we hypothesize this is because Florida had ended the state’s stay-at-home mandate by then and there may have been an increased travel to and within the state during the summer months. In contrast, California was still under a stay-at-home mandate at this point in time and experienced a more subtle spike in cases. It is important to note that during this time of summer 2020 both states had mask mandates in place and the spike in Florida may be suggestive that stay-at-home mandates were a more effective COVID prevention measure than mask mandates.

**Fall 2021**

Both states saw somewhat of a spike in cases in fall of 2021 after mask mandates in each state had ended and when the highly contagious Delta variant was surging. Florida saw a bigger spike than California at this time and we hypothesize this may have been because Florida ended the state’s mask mandate earlier than California. Florida may also have experienced an increase in cases because of increased travel to and within the state during the late summer months.

**Late 2021-Early 2022**

Both states saw a large spike in cases around the end of 2021 and the start of 2022 when the highly contagious Omicron variant was widely spreading. The spike in Florida occurred a bit earlier than that of California. The peak of cases at this time occurred in Florida closer to December 2021 while it occurred in California closer to January 2022. Despite California having already reached a vaccination rate of 75%, both states experienced this spike, we hypothesize, because vaccines were deemed less effective against the Omicron variant. Also, this peak in cases occurred around the holiday season which is a pattern seen in most states throughout the pandemic.

Additionally, California’s peak may have been a bit more delayed than Florida’s because fall 2021 is when the first booster dose was widely available for the population. It is possible that California saw a greater number of booster doses administered which were more effective than the original vaccine on the Omicron variant for at least a period of a couple months. This may have delayed the worst of California’s Omicron-related spike.

Finally, it is important to note that cases displayed on the curves for both states may be hindered because by late 2021 at-home COVID testing was more widely available. This means that fewer cases were being reported through official testing sites to state health departments and to the Centers for Disease Control and Prevention which will have affected the average cases displayed.

**Vaccination Rates**

Both states achieved a rate of 25% of the population having a completed vaccine series around the same time in April of 2021. However, California reached the threshold of 75% of the population being fully vaccinated earlier than Florida. California achieved 75% vaccinated around December 2021 and Florida achieved it around February 2022. California may have been able to reach 75% vaccinated more quickly because the state utilized numerous incentives to encourage vaccination, such as lotteries and prize giveaways. Vaccines in California were also mandated for all state employees and most major school districts required vaccines for eligible students by the start of the 2021-2022 school year. This may have helped California reach 75% vaccinated sooner than Florida because Florida did not have these vaccine mandates nor incentives.

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

**Conclusion:**

California and Florida took different approaches to COVID prevention in terms of stay-at-home orders, mask mandates, and vaccination requirements. Both states experienced various spikes in average case rates and suffered a burden from average cases throughout the time period of January 2020 through January 2022. Even though California and Florida took very different approaches to COVID prevention strategies, both states overall saw a similar average case rate over time.