PM566 Final Project

Home Page

AUTHOR
Stella Zhou

This is a final project dedicated to PM566 Introduction to Health Data Science course at USC.

1. Introduction

Background

On May 11, 2023, the federal Public Health Emergency (PHE) for COVID-19 had officially been expired under section 319 of the Public Health Service (PHS) Act. Over the last two years, The COVID-19 pandemic, which swept across the globe in early 2020, has posed an unprecedented challenge to public health systems, economies, and daily life. As the world grappled with the profound impact of the virus, the development and deployment of effective vaccines emerged as a critical pillar in the fight against this novel coronavirus. Within the United States, each state faced unique challenges and opportunities in administering vaccines, reflecting variations in population density, healthcare infrastructure, and sociodemographic factors. California, as one of the most populous states in the nation, has been a focal point of the pandemic response. With urban centers, rural areas, and diverse communities, California's experience with COVID-19 and its vaccination efforts provide a rich case study for understanding the dynamics of vaccine distribution, uptake, and public health outcomes.

Questions of Interest

The question of interest in this project are:

1. How have COVID-19 vaccination rates varied by county in California since the initial implementation?
2. Is population density associated with cumulative COVID-19 cases?
3. Is income associated with COVID-19 vaccination rates?

This inquiry is not only relevant for evaluating the effectiveness of vaccination campaigns but also for identifying disparities and potential areas for improvement in public health strategies. Analyzing vaccination rates across California's counties can shed light on the impact of demographic, socioeconomic, and healthcare-related variables on vaccine coverage. In this data analysis, county-level data is explored to understand the overall trend of cases, deaths, and vaccination rates over the two years, the extent to which communities have been reached, and the determinants of their vaccination rates.

2. Methods

Downloading Data

The data sets used in this project are obtained from California Health and Human Services. The three main datasets include the “COVID-19 Post-Vaccination Infection Data (ARCHIVED)” (1), “COVID-19 Cases Deaths Tests Data Dictionary (ARCHIVED)” (2) , “Income Limits by County” (3), and Population of Counties in California (4).

Data Wrangling & EDA

By looking at the two data tables, the cases and deaths datasets have more observations than the vaccine dataset. The entry for the vaccine set starts on 2020-01-05, much earlier than the first case&death entry, however, the real first dose initiated 2020-07-27. Additionally, date and county is named different in both sets and are unified. Dates were converted to date format. The following section shows summary statistics from the data exploratory process, time series visualization

The datasets were merged by date and county and missing variables are filtered. The data is further subseted into coundty level and state level for further analysis. New variables are created for further analysis. The average number of doses per person is calculated by cumulative total dose divided by the population in each county. and Vaccination rate was calculated by cumulative fully vaccinated divided by population.

3. Preliminary Results

Part 1. Summary tables

Table 1. provides summary statistics for cases, deaths, and daily doses for counties in California. (Note that the minimum values were all 0 therefore not included in the summary.) Based on the table, counties with higher population have more cases, deaths, and daily doses in general, which is expected due to the nature of infectious diseases. One limitation to this analysis is that the population for each county had remained the same over the 2 years in the data set.

Table 2a. Top 5 vaccinated counties in California (as of 2023-05-09)

County	Population	Cumulative Cases	Cumulative Deaths	Cumulative Total Doses	Percent vaccinated	AMI
Imperial	191649	71545	982	545286	0.9	80300
Marin	260800	49972	317	811487	0.9	166000
San Mateo	778001	186050	742	2262832	0.9	166000
Santa Clara	1967585	482934	2770	5476285	0.9	168500
San Francisco	892280	199735	1211	2616570	0.8	166000

Table 2b. Top 5 least vaccinated counties in California (as of 2023-05-09)

County	Population	Cumulative Cases	Cumulative Deaths	Cumulative Total Doses	Percent vaccinated	AMI
Lassen	30065	10294	64	24355	0.3	80300

County	Population	Cumulative Cases	Cumulative Deaths	Cumulative Total Doses	Percent vaccinated	AMI
Modoc	9475	982	9	10695	0.4	80300
Tehama	65885	14566	236	81526	0.4	80300
Trinity	13354	1481	23	18386	0.5	80300
Kings	156444	63267	486	194017	0.5	80300

Table 2a. and 2b. provide detailed statistics on the 5 counties with the highest vaccination rates and the lowest rates. The top five counties are Imperial, Marin, San Mateo, Santa Clara, and San Fransisco. The lowest are Lassen, Modoc, Tehama, Trinity, and Kings. Figure 1. shows the fully vaccinated percent for all counties in detail. This brings out the question of what are some potential factors are associated with the difference in vaccination rates.

Part 2. Visualization of cumulative dose over time by counties

Figure 2. Cumulative Total Doses per Population

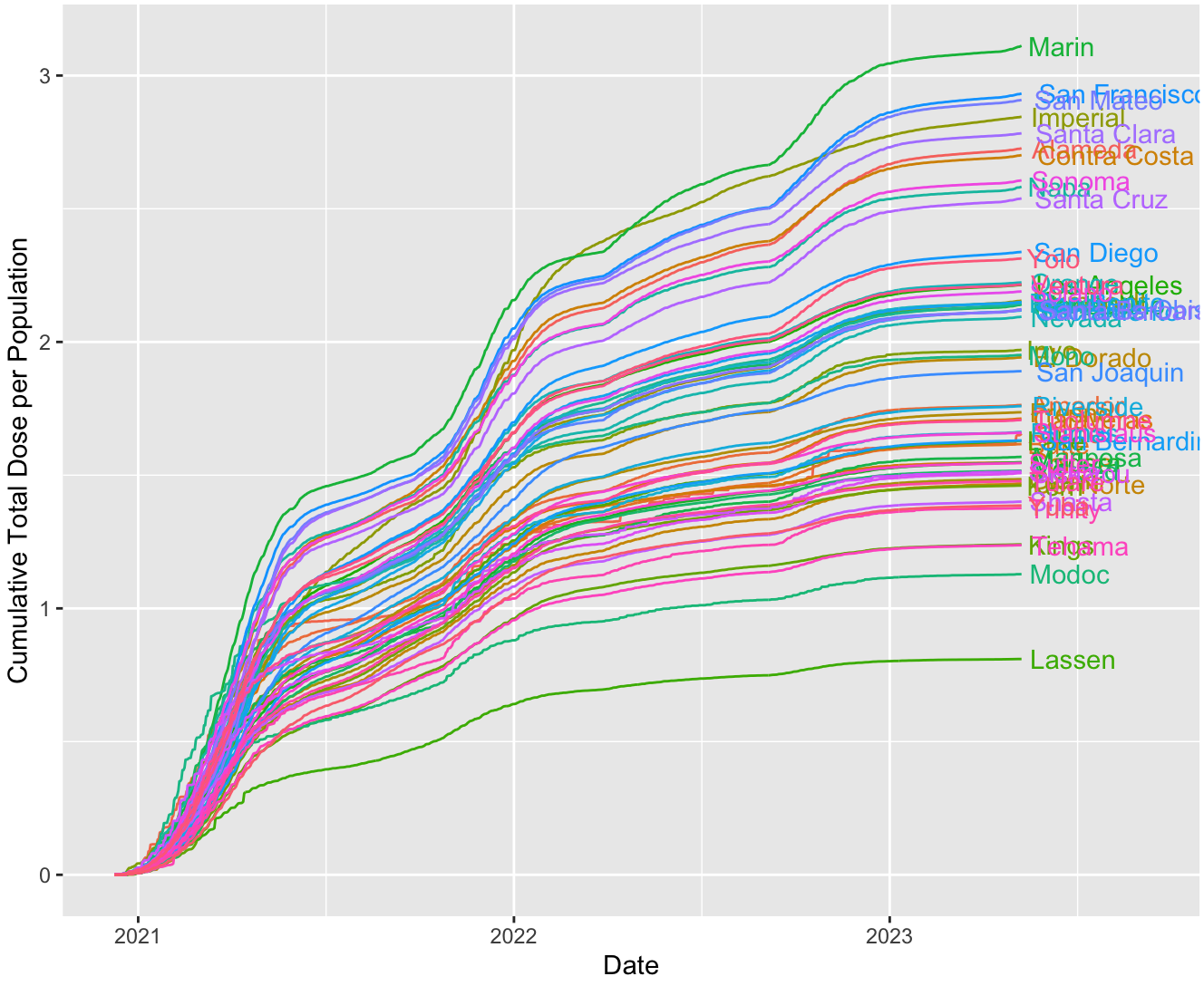


Figure 2. shows the cumulative dose (standardized by population) over time by counties. We can observe that there are three visible inclines in cumulative doses, and approached plateau around April 2021, January 2022, and December 2022. The three sections probably corresponded to the initial outbreak for the alpha variant, the following Delta variant in July 2021 and Omicron BA.1 variant in December 2021, and Omicron BA.2, XBB variants

in summer 2022. Shortly following the new variants outbreaks, new vaccine boosters were administered and thus the rapid incline in cumulative doses. The rapid incline may also be related to admininstration of flu shots during flu season in winters. We can observe similar rankings of cumulative total dose, by counties, to the results in Figure 1.

Part 3. Time series comparison of cases, deaths, and total doses in California

Figure 3. Daily Cases, Deaths, Doses from 2020-01-01 to 2023-05-13

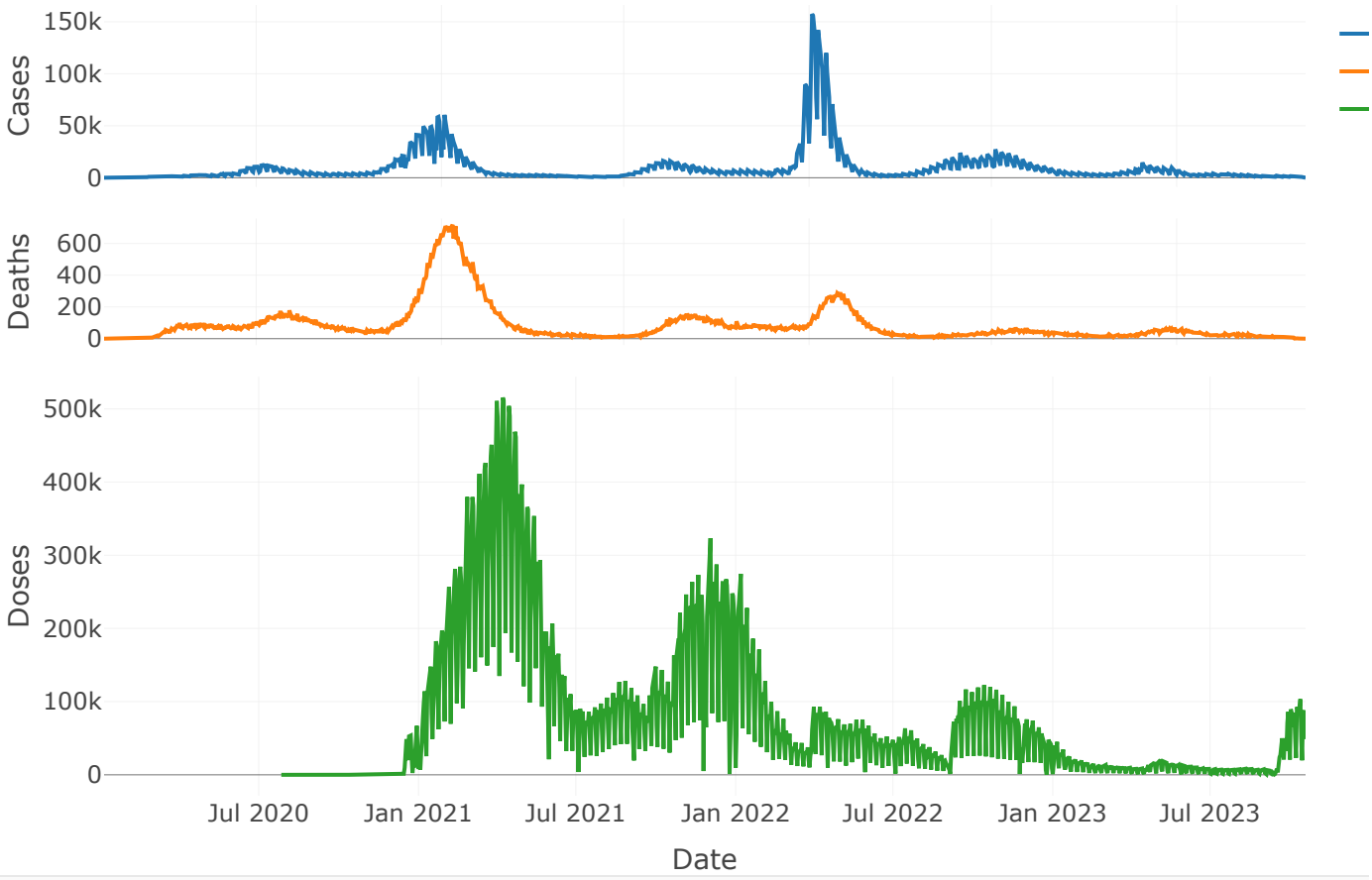
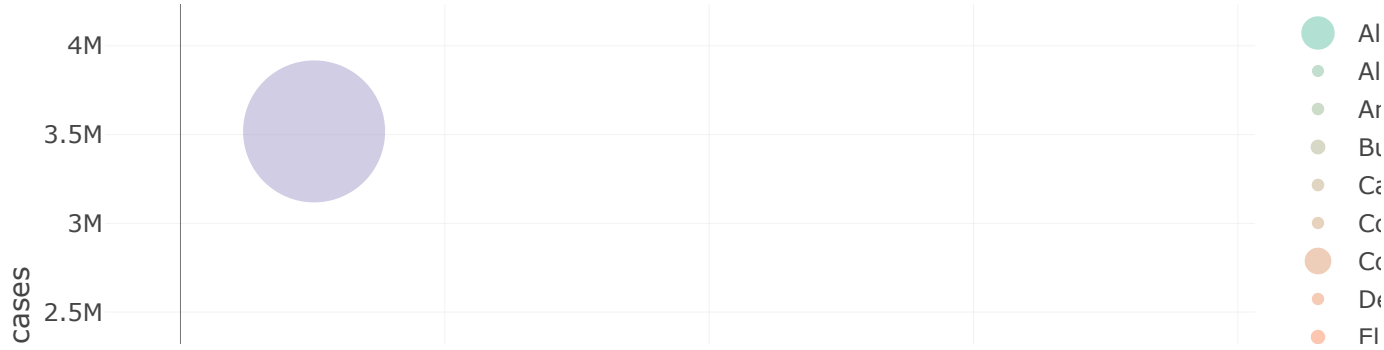


Figure 3. shows a clearer picture of cases, deaths, and daily doses in California as a whole. we can observe that vaccination efforts have played a significant role in reducing both cases and deaths, but the emergence of variants has periodically led to case surges. Note that there is huge daily fluctuation because there are limited testing and vaccine service on weekends.

Part 4. Association between cumulative cases and population density

Figure 4. Population Density vs. Cummulative Cases by 2023-05-09



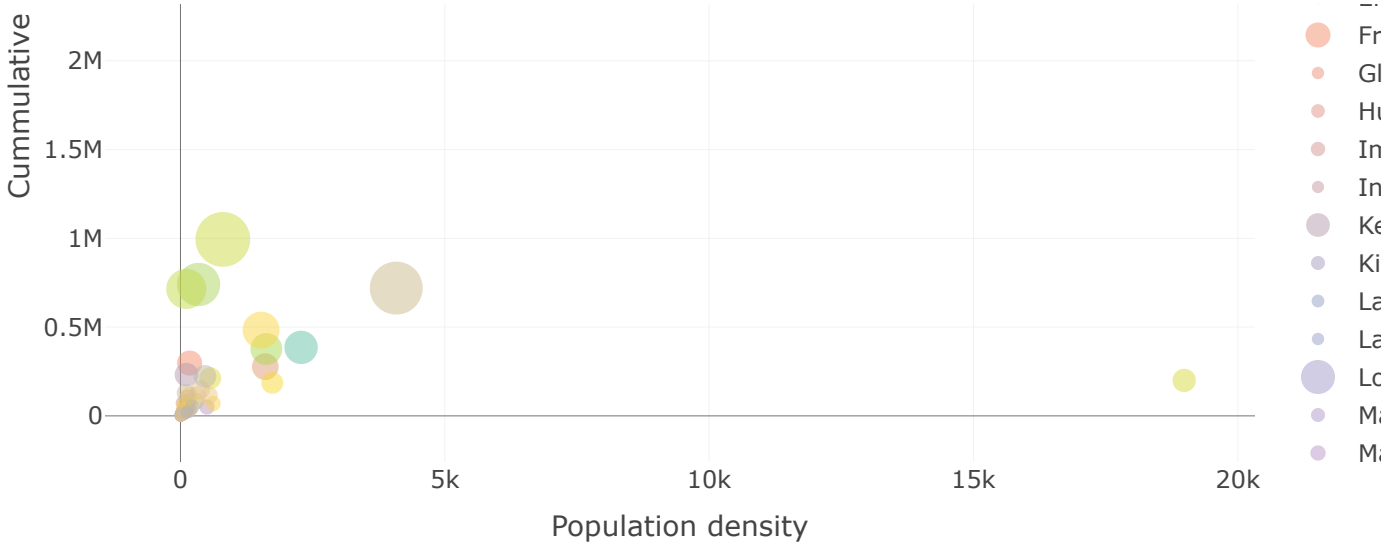


Figure 4. shows the relationship between population density and the percent vaccinated in Counties in California, by 2023-05-09. In general, counties with a higher population density tend to have higher cumulative COVID-19 cases compared to counties with lower population density. Some interesting data points are Los angeles and San Fransisco. (You may click on the name of the county in the legend to hide its data point for scaling purposes). San Fransisco being the most densely populated county in California, has a relatively low cumulative cases. Orange county, following Los Angeles, has also high population densities yet much higher cumulative cases. One factor that influences the association is population. Despite San Fransisco being the most densely populated county, it has a much smaller population compared to LA and Orange county.

Part 5. Association between vaccination rate and median income

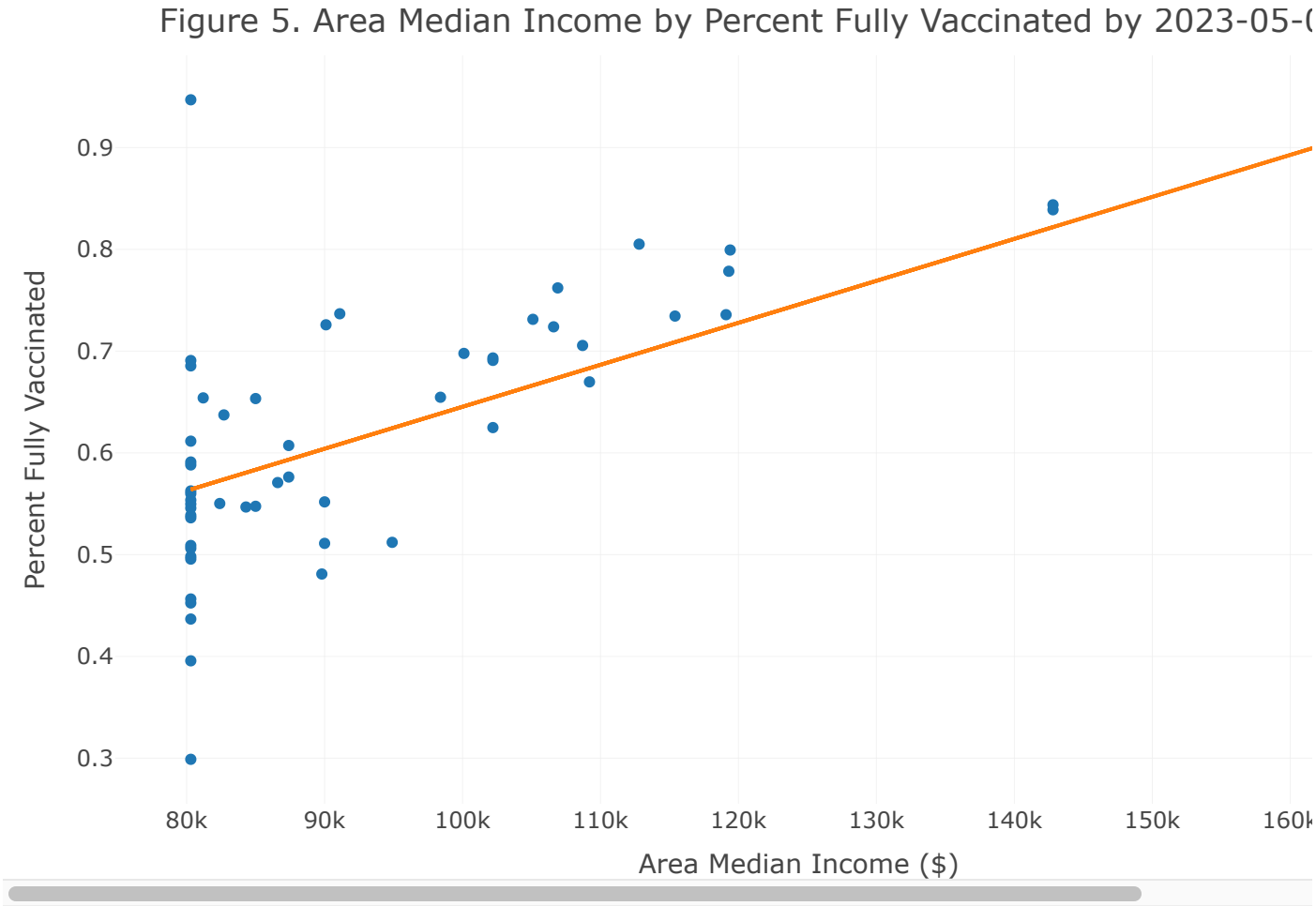


Figure 5. shows the association between median income and percent fully vaccinated in all counties. Overall, higher median income is associated with higher vaccination rates. Notably, the influential point at the top left corner is Imperial County.

4. Discussion

From the preliminary results, I have witnessed the difference in the impact of COVID-19 pandemic on communities in California. From Tabel 1., counties with densely populated urban centers, such as Los Angeles County, have borne the brunt of the virus's rapid transmission. When looking at the California as a whole, Figure 3. demonstrates that vaccination efforts have played a significant role in reducing both cases and deaths, but the emergence of variants has periodically led to case surges.

From Table 2a. and 2b., it is evident counties with the highest vaccination rates share a similar trait – they generally the wealthiest counties in California. And not surprisingly, some counties on the bottom of the list also have the lowest incomes, and are more rural. Noticeably, although Imperial County ranks as the second highest on the vaccination rate list, it also has the much lower income compared to other counties. Some articles cautioned that vaccination rate may have been boosted by Americans who live in Mexico crossing the border to get their shots, but the testing of this hypothesis is beyong the scope of this project. It is also worth noticing that Lassen County, despite having the lowest vaccination rate, was especially lower than the second least vaccinated Modoc county. Some articles suggest that it is due to it high incarcerated population and having no jurisdiction over their COVID-19 mitigation decisions, actions, and record keeping.

The varying cumulative COVID-19 cases in Counties in California shown in Figure 4. can be attributed to different factors, with population size being a significant influence. The relatively low cumulative cases in San Francisco despite its high population density may be attributed to effective public health measures, robust healthcare infrastructure, and a proactive community response. The city has likely implemented stringent preventive measures early on, such as social distancing protocols, mask mandates, and widespread testing, which could have contributed to controlling the spread of the virus.

On the other hand, Los Angeles and Orange County, with higher cumulative cases, might be facing challenges due to their larger populations. The sheer number of people residing in these areas could lead to increased interpersonal interactions, making it more challenging to implement and enforce preventive measures consistently. Factors such as urban density, public transportation usage, and community behaviors can also play a role in the higher case numbers in these counties.

Moreover, disparities in healthcare access, socio-economic factors, and demographic differences between the counties may contribute to variations in the impact of the pandemic. For instance, issues like crowded living conditions, essential workers' exposure, and variations in healthcare infrastructure can affect the spread and management of the virus differently in each region.

In contrast to densely populated urban centers, rural and less densely populated areas have faced unique challenges, possibly due to limited access to healthcare facilities and vaccine distribution. Therefore, further analysis investigates whether median income is associated with vaccination rates. Figure 5.. provides evidence and shows that higher median income is associated with higher vaccination rates

In summary, while population density is a crucial factor in understanding the dynamics of COVID-19 transmission, it is equally important to consider the overall population size, public health measures, and socio-economic factors to comprehend the variations in cumulative cases among different counties in California.

In moving forward, it is imperative that the lessons learned from the differential impact of COVID-19 on California's counties inform future public health planning. This means addressing the unique needs and vulnerabilities of each region and ensuring that healthcare resources are distributed equitably. It also necessitates continued efforts to promote vaccine access, education, and outreach to address disparities in vaccination rates.

Download my full report here.