COVID-19 IN CALIFORNIA: HOSPITALIZATIONS, VACCINE ROLL-OUT, ECONOMIC RECOVERY, DEMOGRAPHICS

STA 260 FINAL PROJECT

Sabrina Enriquez **Brittany Lemmon**

Llasmin Lopez **Katy Shea** **Zheyuan Yu**

June 10, 2021

1 Introduction

In spring 2020 while many states like New York and Washington were facing COVID-19 devastation, California was surprisingly successful at repressing a surge in cases. But by the 2020 holiday season, California quickly became an epicenter of the virus [1]. The holiday spike of November, December, and January more than doubled the state total death toll [2]. Though 2021 started grim for the state, California now has much more hopeful outlook with some regions of the state achieving the highest vaccination levels in the country [3].

In this project we are concerned with investigating the impact of the COVID-19 vaccine roll out on California Counties. We are motivated by the devastation the COVID-19 pandemic caused throughout California and hope to shed light on any relationship the vaccine has had on peoples' lives. We organize our efforts by researching in the following format:

- Vaccination and Case Count over Time: Section 2
- Hospitalization and Vaccine Roll-Out: Section 3
- Geography of Vaccination Patterns: Section 4
- Vaccination Ratio amount Different Demography Groups: Section 5
- Vaccine Equity: Section 6

2 Vaccination and Case Count over Time

We begin by investigating the relationship between positive cases of COVID-19 and vaccination rates, county by county. This relationship is evaluated with respect to the date, since this data's temporal perspective is particularly indicative of a possible relationship. We are interested in seeing if there is an observable relationship between the rise in vaccination in counties and daily positive case count.

Given that clinical data for the Pfizer-BioNTech, Moderna, and Janssen/ Johnson & Johnson vaccines show that "mRNA COVID-19 vaccines offer similar protection in real-world conditions as they have in clinical trial settings, reducing the risk of COVID-19, including severe illness, among people who are fully vaccinated by 90 percent or more" [4], we have motivation to explore how the vaccination campaigns throughout California are impacting the spread of disease. Given such information, a reasonable person could expect that as vaccination rates in California Counties increase, infection rate should decrease. We pursue this question by analyzing and exploring cumulative vaccination data from the beginning of vaccination efforts in California on 12-15-2020 until 05-08-2021. We utilize California COVID-19 data which can be freely accessed at [5] and [6].

2.1 Methods

We analyze the data retrieved from [5] and [6] to assess the vaccine's relationship with case count using two primary methods:

1. Computing the correlation between the proportion of daily positive case count over county population and the proportion of fully vaccinated people over total county population for each day and for each county.
2. Categorical exploratory data analysis methods outlined in [7].

We choose these methods because they allow us to explore the structure of the data without making the many assumptions necessary for model fitting. Moreover, we purposefully avoid computing a commonly used epidemiology statistic R_0, the "basic reproduction number" for this data, because that method requires that everyone in the population be susceptible, but that is not the case with our data [8]. Since our data does not fit the base assumptions of that method of analysis, we exclude it from this report.

2.1.1 Correlation method

For each county we have a daily count of positive cases and the daily cumulative count of fully vaccinated people, as well as the county's population. To normalize the data, we look at proportions of positive cases and the daily cumulative count of fully vaccinated people to avoid biasing toward larger counties. From now on we will refer to these proportions as "infection rate" and "vaccination rate" respectively.

Since the first reported vaccination was administered on 12-15-2020, we look at the subset of data from 12-15-2020 until 5-8-2021. Then, we split the data into each of California's 58 counties and for each county do the following:

In chronological order we compute the correlation of the "vaccination rate" and "infection rate" data for all dates up to that point in time. For example, if we are considering the 16th day since vaccination began, we take the correlation of the vaccination and infection data from 12-15-2020 until 16 days after that. So, as we move closer to present, the data which contributes to the correlation coefficient ρ is getting larger and our error bounds are decreasing. On the other hand, toward the beginning of the data, the error increases and is greatest for the correlation coefficient corresponding to 12-16-2020 (the first ρ) since it only has 2 pairs of data points. Then, we plot those ρ values over the dates that they correspond to for each county.

Given the clinical data and efficacy studies of the vaccines [4], a reasonable person might expect that as vaccination goes up, infection rates go down. Using this method, that expected relationship translates to the correlation coefficient, ρ, approaching -1 showing an inverted relationship between these factors as time and thereby vaccination increases. In Figure 1 we see that by and large, most counties cross the $\rho = 0$ threshold around March and continue to decline into May, supporting the claim that infection goes down as vaccination goes up.

Of course, there are countless confounders to this data, so we can't make any leaps in logic and imply that the vaccine rollout is directly responsible for this relationship, but there is definitely a clear trend amongst the majority of the 58 counties across California. We discuss this further in subsection 2.2.

2.1.2 Categorical Exploratory Data Analysis Method

Next we use methods outlined in [7] to explore the relationship between infection rate and vaccination rate across (County, Date) pairs. We begin by using the hierarchical clustering algorithm (HC) with method "ward.D2" on the distance matrix produced by finding the Euclidean distance between each (County,Date) observation consisting of its corresponding positive test and cumulative vaccination proportion. We see the resulting dendrogram tree in Figure 2 and using colorful rectangles show the natural choice for 6 clusters.

Interested to see if these clusters reveal any underlying structure, we further analyze the clusters in section 2.2.

2.2 Results & Discussion

With our preliminary results shown in figures 1 and 2, we dive a bit deeper and attempt to discover what these results can tell us.

2.2.1 Correlation Method

Using the correlation method we see that toward the beginning of the data points in Figure 1, there is much more noise, which is consistent with our understanding of the law of large numbers. Since the size of our data increases with time,

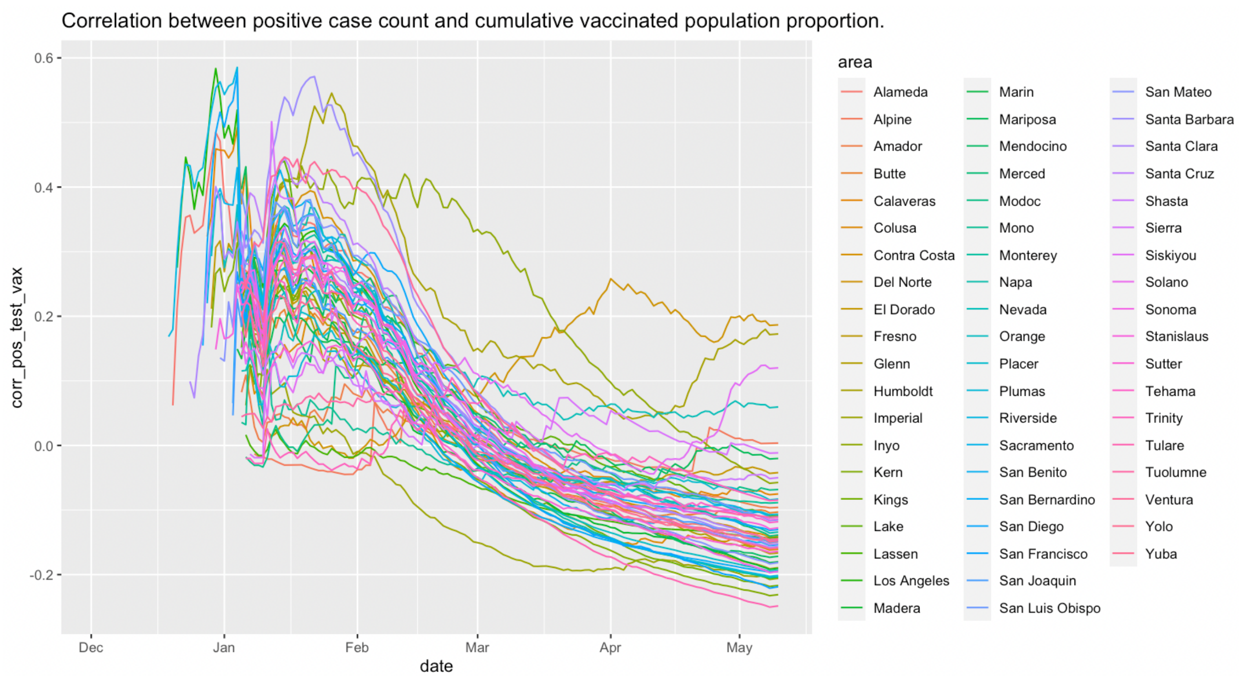

Figure 1: The cumulative correlation of vaccination rate and infection rate across all counties computed independently from one another at each date between 12-16-2020 and 05-08-2021.

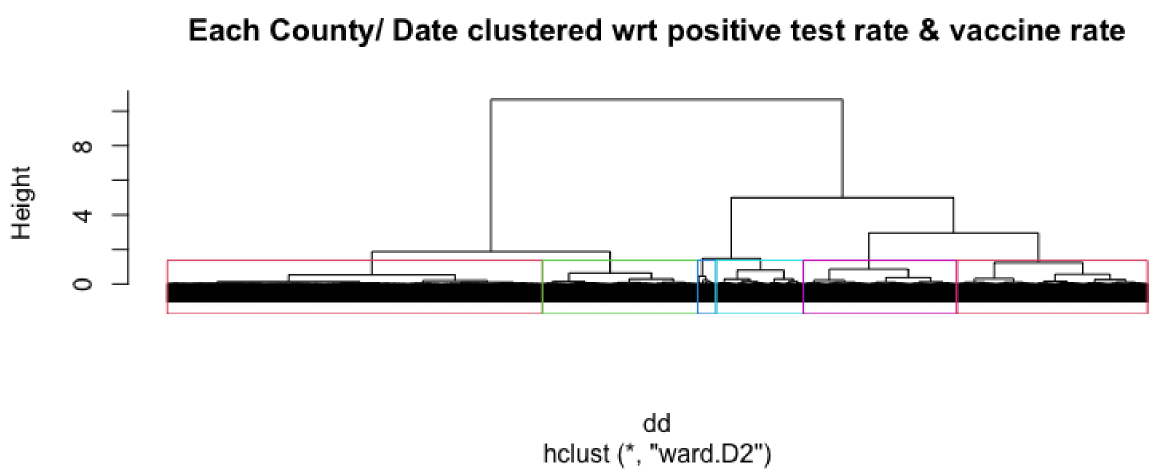

Figure 2: Dendrogram of (County, date) pairs evaluated over proportion of positive cases and proportion of cumulative fully vaccinated using the hierarchical clustering algorithm with the "Ward.D2" method.

the noise decreases and our computations are more reliable toward the final date of May 8th. As such, the clear trend for most counties approaching -1 with time is promising and aligns with our expectation.

Since the COVID-19 coronavirus is primarily spread from person-to-person and traveling has been relatively scarce since the beginning of the pandemic, it makes sense to compare nearby counties to one another. We split our 58 counties into 7 geographically close regions: the Bay Area, Central Coast, North, Northern San Joaquin Valley, Sacramento Area, South, and San Joaquin Valley. The counties within each region is listed in the legends found in Figures 3 and 4.

Observing Firgure 1 it is quickly apparent that the Bay Area and Central Coast plots look very similar because they both show sharp decline and the counties all perform relatively similarly. However, the North and Northern San Joaquin Valley plots have significantly more variation between counties and do not decline as quickly as the Bay Area and

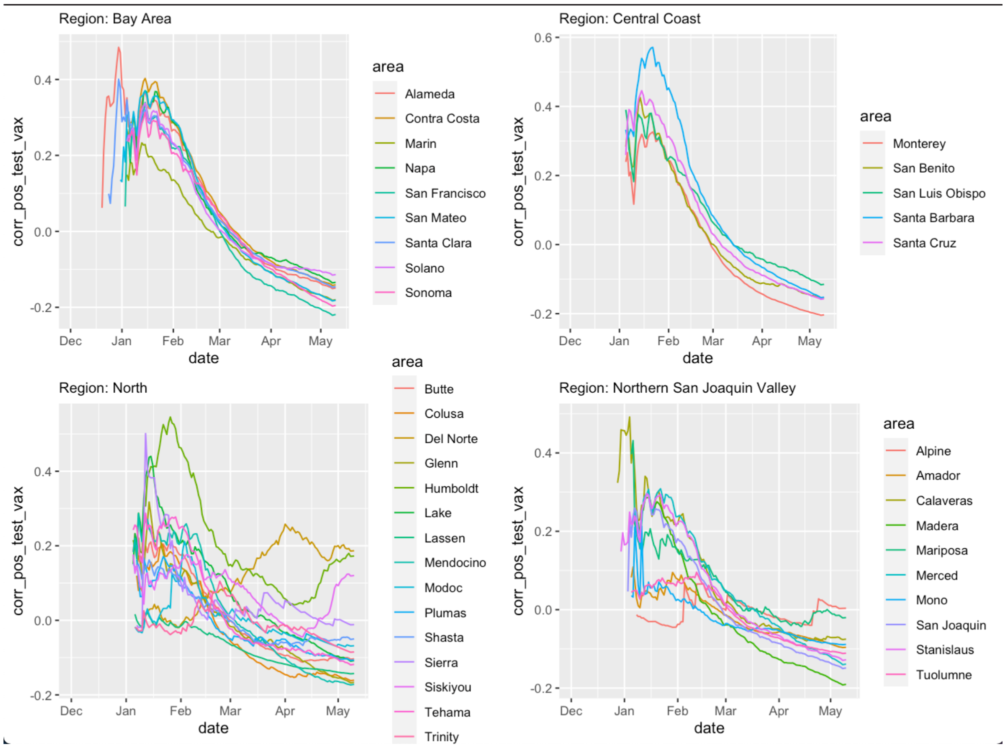

Figure 3: Cumulative correlation between positive case count proportion and vaccination proportion for the Bay Area, Central Coast, North, and Northern San Joaquin Valley regions.

Central Coast regions. This means that there is less eveidence of correlation in these regions and that the counties within those regions are less uniform in their experiences with the vaccine rollout. Notably, in Del Norte, Lake, and Siskiyou counties they experienced the opposite trend, and as vaccination increased, so did infection rate. These stand-out counties are important to further investigate through other lenses, because they are unfortunately the exceptions to the rule.

In Figure 4 we see that those regions all have downward trend as we might expect to see, but there are again a few stand out counties. In the Sacramento Area, Nevada county's correlation seems to approach zero and remain relatively constant slightly above zero, indicating no clear correlation or observable numeric relationship between vaccination and infection. Morever, Placer and El Dorado counties have a local minimum around April and then their correlation's both approach zero from below. The rest of the counties however follow the common trend and continue to decrease with time and increased vaccination.

On the other hand, in the South region, Imperial county is notable because it remains significantly below the rest of the counties in correlation, crossing 0 in February and then leveling out at around 0.2 along with other counties in mid April. This means there was more evidence of positive impact from the vaccine much more immediately in Imperial county than in its neighboring areas.

Lastly, in the Southern San Joaquin Valley we see that Inyo county had relatively constant positive correlation until about March while it's neighboring counties started to have positive impact from the vaccine roll-out. Moreover, Inyo county did get sharp decline indicating positive impact from the vaccine roll-out after March, but unfortunately $\rho = 0$ did not happen for them until almost May, while the rest of the counties in the region crossed zero in mid-February. This is a tragic indication that the people of Inyo county suffering for longer than their neighbors and it is important we find out why.

2.2.2 CEDA method

Now we switch gears and explore this data using machine learning methods. Expanding on the HC tree shown in 2, we plot the heatmap with that HC tree on the left axis in Figure 5. On this plot it becomes apparent that the clusters were largely influence by the right column corresponding to vaccination rate. To make the plot more clear we show the heatmap generated by running HC on log(1+ proportion of vaccinated) and log(1+ proportion of infection) for all

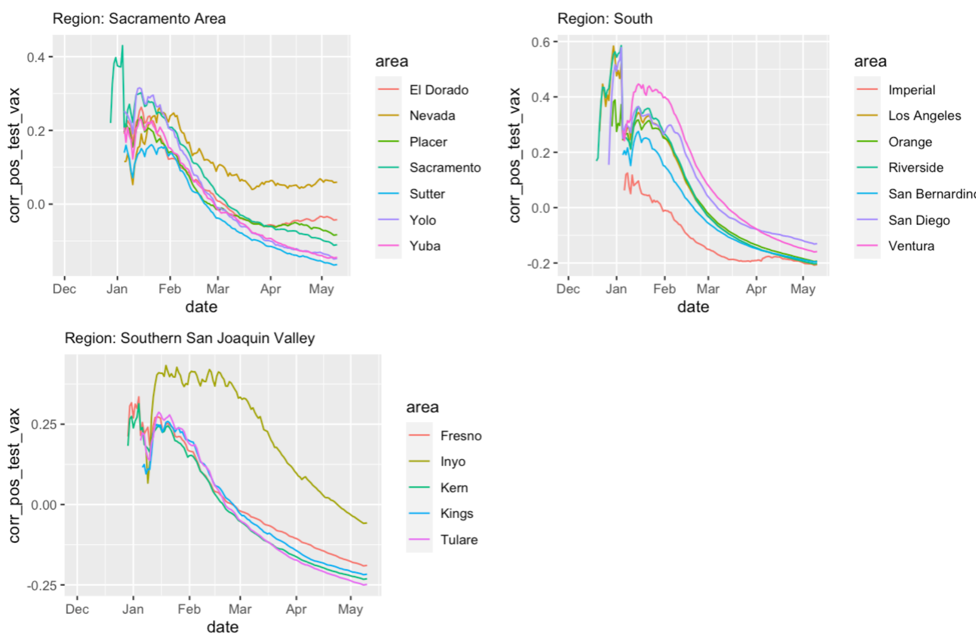

Figure 4: Cumulative correlation between positive case count proportion and vaccination proportion for the Sacramento Area, South, and San Joaquin Valley regions.

(county, date) observations. Although the date is only a label here and was not included in the clustering algorithm, upon further inspection we see in Figure 6 that the clusters are largely determined by the date. This coincidence makes sense

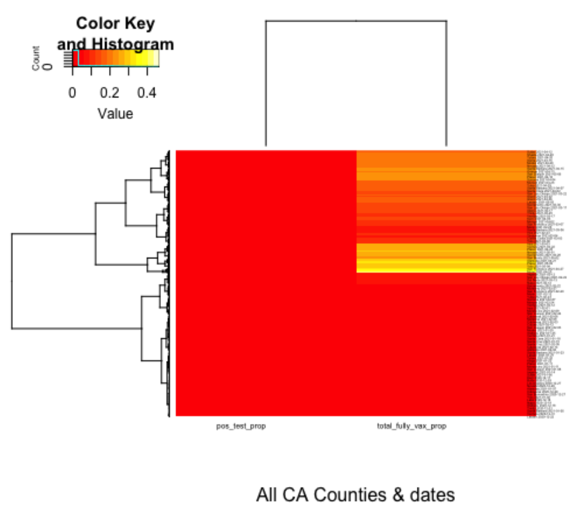

Figure 5: Heatmap of (County, date) pairs over log(1+ proportion of positive cases) and log(1+ proportion of cumulative fully vaccinated).

since vaccine roll-out largely determined the clusters and date is a confounding variable since as time goes forward, cumulative vaccination increases.

Moreover, in Figure 7, we plot how the clusters distribute over counties and recognize that all of them spend some time in each cluster, showing that the data was not very different across counties from this computational perspective. In the following sections we will further discuss how the different counties were impacted via different lenses and attempt to better understand how different communities throughout the state are being impacted by vaccine roll-out.

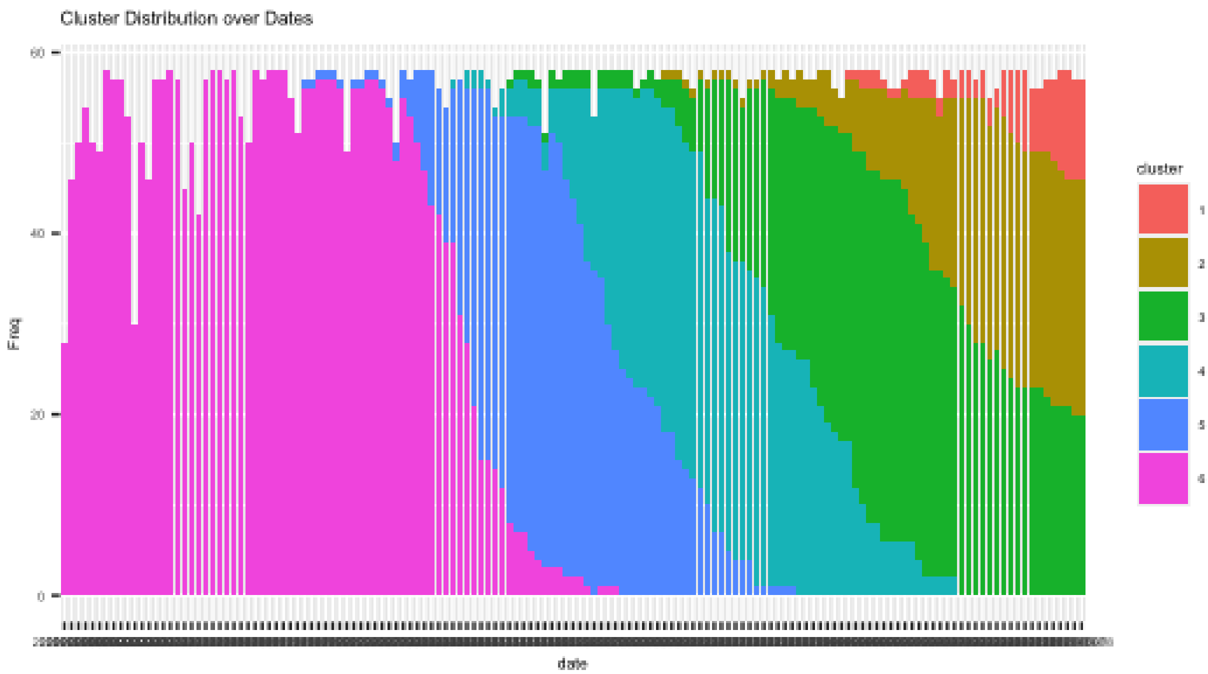

Figure 6: Distribution of cluster classification over dates.

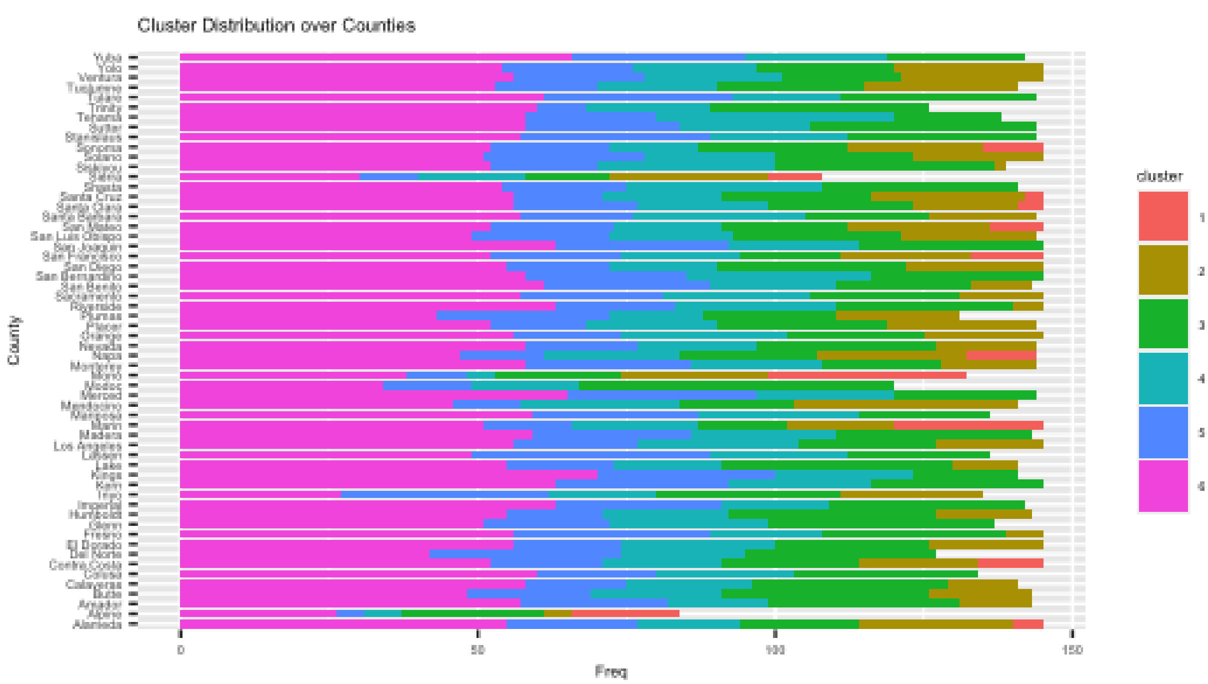

Figure 7: Distribution of cluster classification over Counties.

3 Hospitalization and Vaccine Roll-Out

Here, we explore the relationship of hospital bed availability and use with the roll-out of vaccines against SARS-CoV-2 in California. Bed availability is considered as a metric for availability of healthcare in a county. One aim is to understand if higher bed counts per-capita in a county reflects the county's vaccine administration, measured by proportion of fully vaccinated people in the county. Additionally, we study the relationship between vaccine roll-out and occupation of hospital beds by COVID-19 patients to determine if higher proportion of vaccinated population relates to lower hospitalization rates.

3.1 Methods

To answer the questions regarding the relationship between hospital beds and vaccine roll-out, we used data from California Open Data Portal. Namely, the COVID-19 Hospitalizations By County [9], Statewide COVID-19 Vaccines Administered By County [5], and Statewide COVID-19 Cases Deaths Tests [6] datasets.

As a preliminary step, the hospitalization and vaccine data were cleaned by removing any locations outside California or locations that did not indicate a specific county (i.e. "All CA Counties", "Outside California", or "Unknown"). The Hospitalization dataset contains variables referring to confirmed and suspected COVID-19 hospitalizations and another referring to total COVID-19 hospitalized cases which should be a sum of the previous two. In some counties, the total was not reported, but the first two variables were, so in these cases, the total variable was set equal to the sum of the suspected and confirmed hospitalized cases. There were three counties (Alpine, Sutter, and Sierra) for which no hospitalization data were reported, so these three counties were left out if this portion of the analysis.

After the preliminary cleaning, the hospitalization [9] and vaccination [5] datasets were merged by county and date. In some few cases, `NA`s were introduced where dates in one dataset did not exist in the other. In these cases, variables which represented daily totals (i.e. total hospitalizations) were set to zero and variables which represented cumulative totals (i.e. cumulative total vaccinations) were set equal to the previous date's value. This choice was made because the intended method of analysis is to use heatmaps, and the data must be completely numeric to generate heatmaps.

Finally, the data were assessed using heatmaps and matrix plots. Each variable analyzed was reshaped into a matrix with the county names on the rows and the dates on the columns. Because not all counties reported data on all dates, some additional `NA`s were introduced. As in the previous step, variables which represented daily totals were set to zero and variables which represented cumulative totals were set to the previous non-missing value. Using the `heatmap.2` function from the `gplots` package in R, heatmaps were generated for each variable of interest and a corresponding dendrogram generated using complete-linkage clustering to determine the similarity of the counties. Matrix plots were generated to compare the hospitalization data with vaccinations using `matplot`.

Throughout this analysis, several dates are highlighted which indicate when new groups of the population became eligible for vaccination[10]. December 15, 2020 is the earliest date in the vaccine dataset and presumably the first vaccination date in California, outside of clinical trials. From December 15, eligible groups included (in phases) healthcare workers, workers and residents at nursing homes, people aged 65 or older, education and childcare workers, emergency workers, workers in food, grocery and agriculture, and persons with disabilities. On March 15, 2021, new groups became eligible, including persons living and working in congregate care facilities like jails or homeless shelters, transportation and logistics workers and persons age 16-64 with underlying health conditions. On April 1, all persons aged 50 or older became eligible, and on April 15, persons aged 16 or older became eligible.

3.2 Results & Discussion

3.2.1 Bed Availability

Healthcare availability during a pandemic is a matter of serious concern. We are particularly interested in understanding how the availability of healthcare in a county relates to the distribution of the vaccine against SARS-CoV-2. Hospital bed availability here is measured as hospital beds in a county per 100,000 people in the population [6]. It is important to recognize that hospital bed counts alone do not capture the complexity of healthcare access in the United States [11]. However, given the public access data available, hospital bed per capita was deemed the best option for this project. Hospital bed counts were also chosen because they are based on appropriate staffing numbers to adequately serve the number of beds. This is useful to gauge approximate numbers of healthcare workers who would be qualified to administer vaccines in a county and could perhaps partially explain vaccination rates.

In Figure 8, notice that most counties have similar per-capita bed availability. Yuba and Shasta have higher counts than others throughout and Modoc county has high bed counts in earlier dates and then decreases similar availability to the majority of counties in December 2020 and into 2021. Modoc has a small population, under 9,000, so it is easy to see

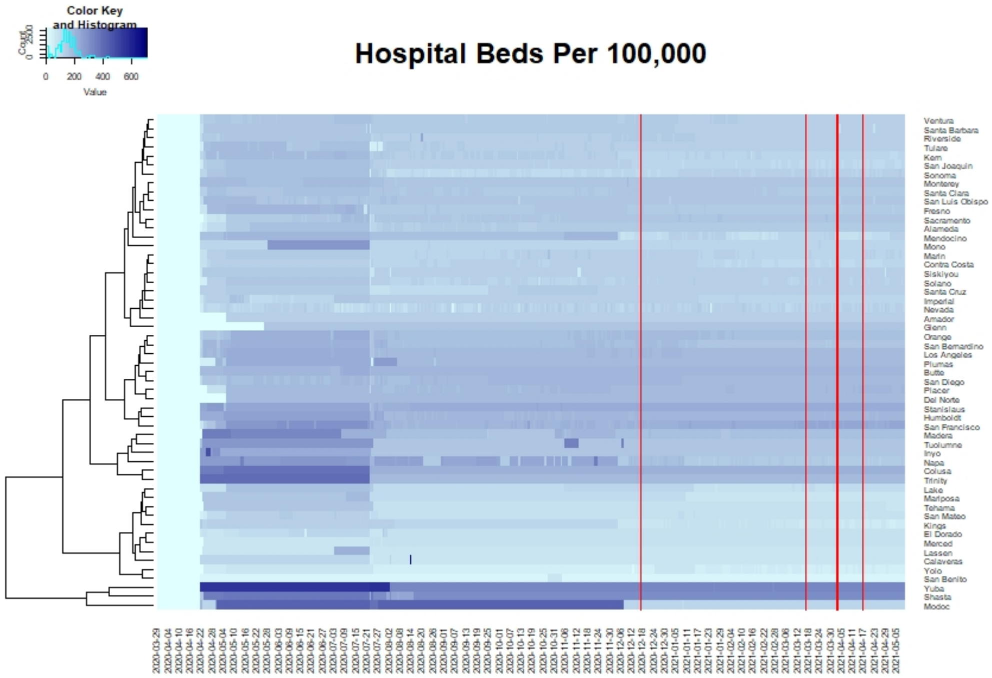

Figure 8: Hospital beds available per 100,000 persons in each county over time. Beds available indicates all beds in hospitals excluding emergency departments. Red lines (left to right): 15December2020, 15March2021, 1April2021, and 15April2021 indicate dates when new groups became eligible for vaccination.

how their availability could be high early in the pandemic if they had higher staffing numbers to prepare for the worst outcome. The hospital in Yuba county serves the populations of both Yuba and Sutter counties (note Sutter county had no hospital data for this reason), and so it requires enough beds to serve multiple populations [12]. It could perhaps be reasonable to reconstruct this heatmap with the per-capita counts for Yuba representing both counties populations, but this was not done in the interest of not making too many assumptions. It is unclear what sets Shasta county apart from the rest of the counties and why it would have such high per capita bed availability. The bed availability in Shasta county appears to remain constant throughout the pandemic, so it is possible that they simply have higher bed availability regardless of the pandemic. Among the lowest per-capita bed rates are Lake, Mariposa, Tehama, San Mateo, Kings, El Dorado, Merced, Lassen, Calaveras, Yolo, and San Benito counties. Unlike the highest bed count counties, these counties do not have easy explanations for why they have relatively low bed counts. If bed availability is a good indicator of vaccine distribution, then we could expect to see Shasta county among the highest vaccination rates and Yolo and San Benito among the the lowest rates.

Figure 9 shows a heatmap of cumulative proportion vaccinated against SARS-CoV-2 from December 15, 2020 to May 8, 2021. Referring back to the findings from the bed counts heatmap, it seems that the counties with extreme high or low bed counts like Yolo, San Benito, and Shasta are not the same counties that have extreme vaccination rates. Specifically, Yolo and San Benito counties have higher proportion population vaccinated by May 8th than Shasta county. It seems that at the extremes, hospital bed counts do not positively indicate a county's vaccination rate. To look at the relationship over all counties in Figure 10, it seems that there is no obvious relationship between hospital bed counts and proportion population vaccinated. This is perhaps not a surprise given the complexities of health care in the united states and the specific logistics of vaccine distribution (discussed in section 4).

3.2.2 Hospitalizations

In this section, we examine COVID-19 hospitalizations per 100,000 persons and the relationship with vaccination rates in each county in California. Figure 11 shows COVID-19 hospitalizations per 100,000 for each county throughout the pandemic. We see that early in the pandemic, Inyo county experienced a very high number of cases but for the remainder of the collected data, the case counts remain fairly low. The remaining counties all follow a fairly similar pattern to each other with a peak during summer 2020 for many counties and the majority of counties experiencing a peak in December of 2020 into January 2021. The increase in case counts in the winter reaches its peak a few weeks

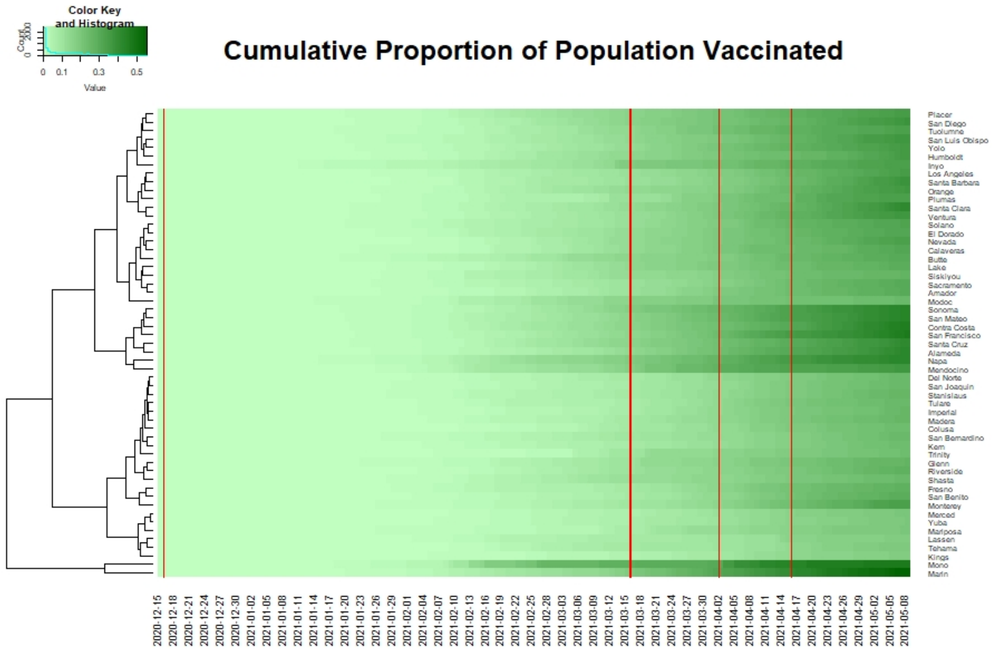

Figure 9: Cumulative proportion of persons fully vaccinated in each county over time from December 12, 2020 through May 8, 2021. Red lines (left to right): 15December2020, 15March2021, 1April2021, and 15April2021 indicate dates when new groups became eligible for vaccination.

after vaccinations begin and continues to fall until May 8th (the end of the analyzed data). The winter 2020 peak can be attributed to some obvious causes like the holiday season and the propensity of *coronaviridae* to survive and be transmitted more efficiently in cooler weather.

As noted above, the hospitalizations from the winter begin to decline a few weeks after vaccinations began and fall steadily until the present. Since case counts fall as vaccinations rise (Figure 12), it is tempting to attribute the case decline to the vaccine. However, it is important to recognize the mountain of confounding factors like season change, the end of winter holidays, and public policy (like tiers, section 6), to name a few. Further analysis would ideally include metrics like average daily temperature, but this data was not readily accessible and could not be easily obtained as of the time of this report.

Ultimately, while improvement in COVID-19 hospitalization numbers is encouraging, it is likely not a sufficient metric for studying the effectiveness of the vaccine against SARS-CoV-2 at the present. Hospitalizations would likely be more useful in a longer-term study through next winter where hospitalization rates could be compared with this past winter once maximal vaccination rates are reached.

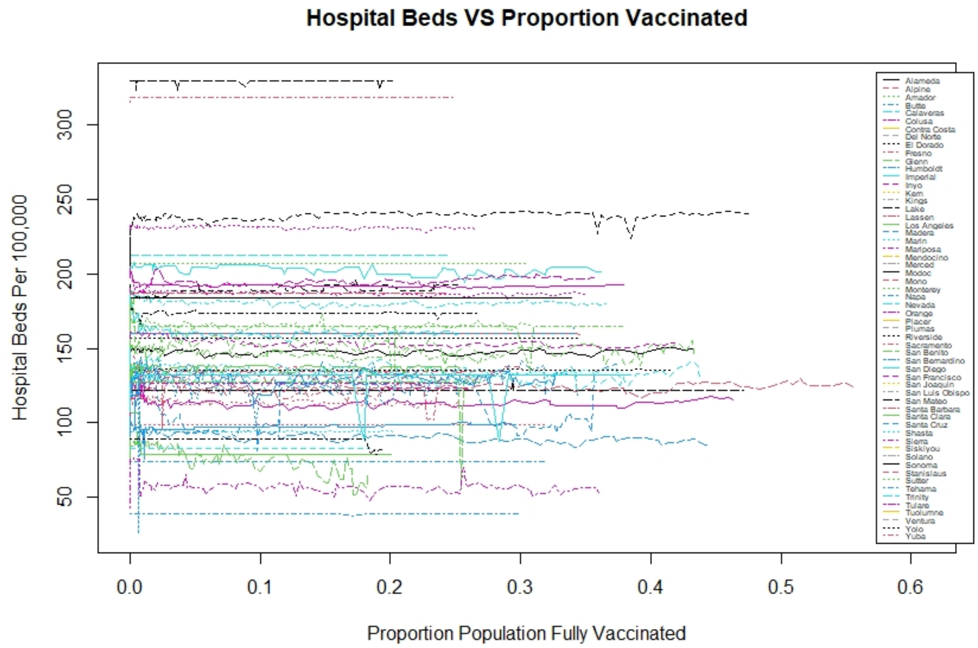

Figure 10: Proportion of population fully vaccinated compared to hospital beds available over time.

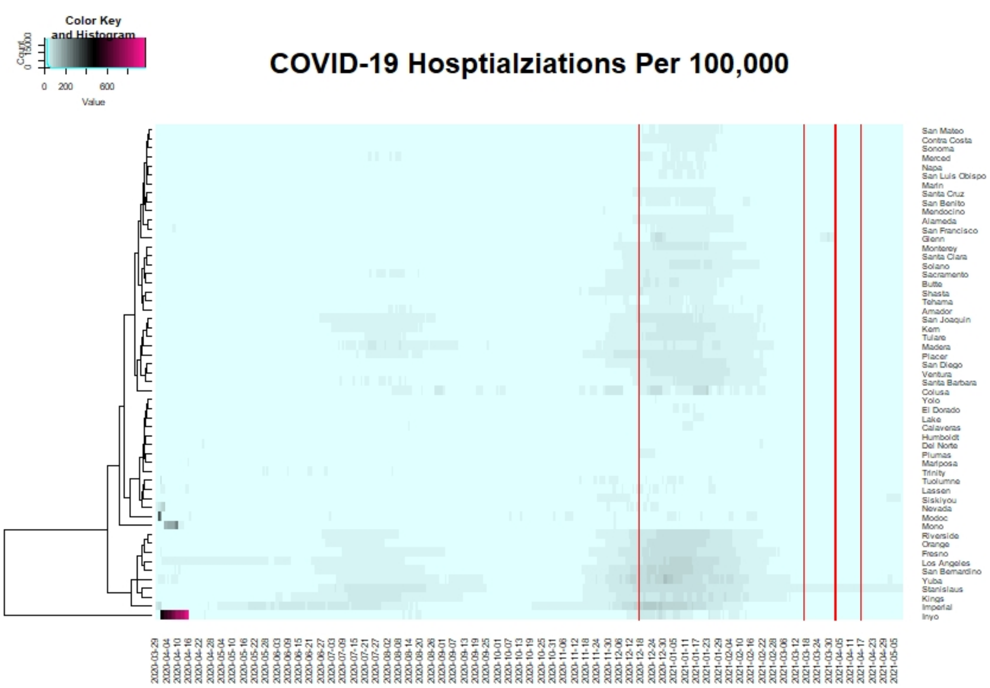

Figure 11: Total hospitalizations from COVID-19 over time in each county. Red lines (left to right): 15December2020, 15March2021, 1April2021, and 15April2021 indicate dates when new groups became eligible for vaccination.

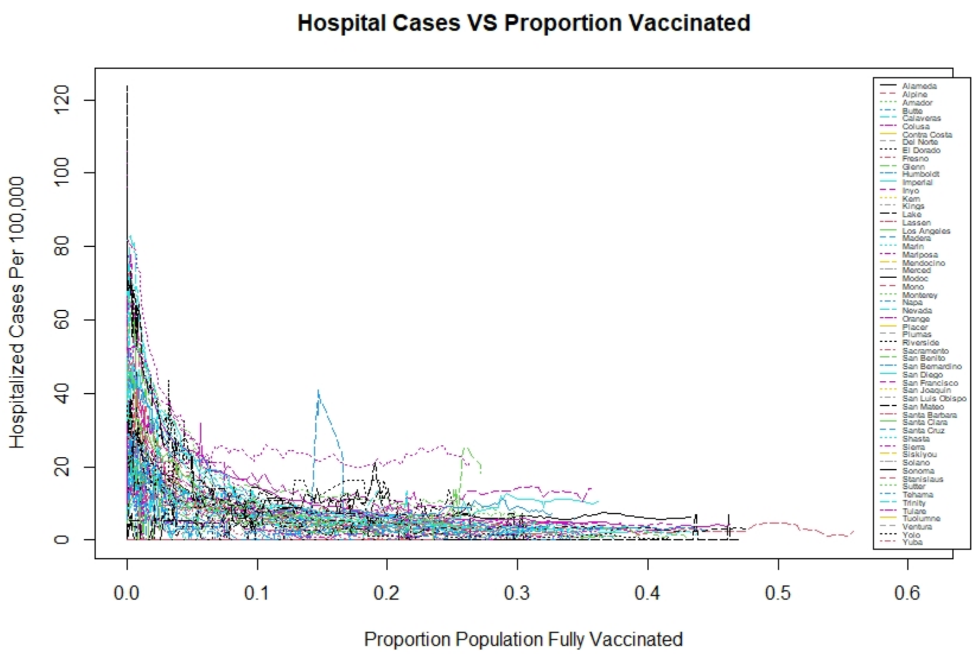

Figure 12: Proportion of population fully vaccinated compared to total hospitalizations from COVID-19

4 Geography of Vaccination Patterns

4.1 Methods

The cumulative fully vaccinated count is defined as the total number of people who have either had two doses of Pfizer, two doses of Moderna, or one dose of Johnson & Johnson on or before the specified date. The population for a county was defined as residents 16 years of age or older on May 8th [6]. The proportion of a county fully vaccinated was defined as the count of cumulative fully vaccinated individuals divided by the 16+ year old population in that county.

4.2 Results & Discussion

The earliest doses of vaccines were available to healthcare workers and workers/residents in nursing homes or other long-term care ("Phase 1A"). By mid February, "Phase 1B" eligibility opened which included people 65 years old and older, education workers, emergency service workers, and food/agriculture workers [13]. After March 15th, employees and residents of congregate care facilities, transportation and logistic workers, and individuals 16-64 years old with disabilities or other health conditions were eligible for the vaccine. Starting April 1st, the vaccine became available to adults 50 years old and older. Lastly, on April 15th the vaccine became open to all adults 16 and older [10].

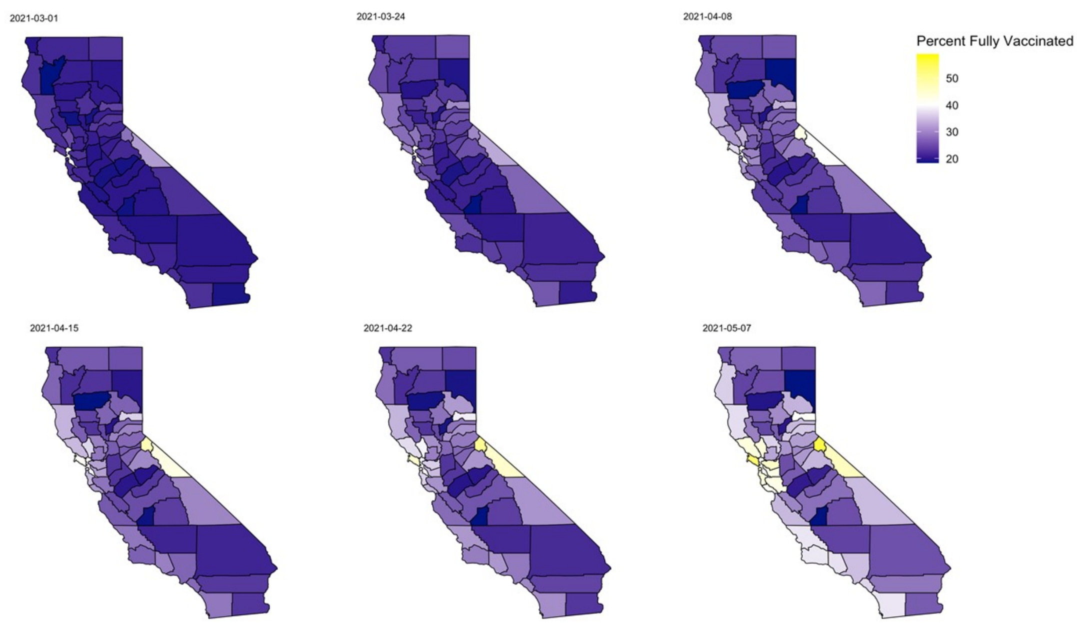

Figure 13: Proportion of fully vaccinated individuals (out of 16+ year old population) in California counties over vaccine roll-out period. Link to animated version of the map: https://github.com/blemmz/Files/blob/main/animatedCAmap.gif

From the maps we see that the Tahoe area (Alpine and Mono) and the Bay Area (Marin, San Francisco, San Mateo, Alameda) reach high levels of vaccination earlier than other counties. Most counties, especially highly populated and urban counties, have a steady trajectory and achieve around 30% fully vaccinated by May 8th. Using proportion fully vaccinated by May 8th as a ranking metric, Alpine achieved the highest level with 58.99% of adults fully vaccinated while Kings county ranked the lowest with only 18.28% fully vaccinated.

The average level of vaccination proportion on May 8th across all counties was 0.3280 (median 0.3198). The two highest ranked counties (Alpine and Marin) achieved more than 50% fully vaccinated while the two lowest ranked (Lassen and Kings) did not have more than 19% fully vaccinated by May 8th.

Population size of a county was the first concern when comparing vaccinated proportion, for example Alpine county achieved 59% vaccinated but only has 1117 residents. The most populous counties face coordination and logistic obstacles in order to administer vaccines to such a large population causing a much slower rate in vaccinations. The largest counties could eventually reach that same level of vaccination as the smallest counties in the future if the slow but steady rate is maintained.

Rank	County	Percent Vaccinated	Population Size
1	Alpine	58.99%	1117
2	Marin	55.55%	260800
3	Mono	47.12%	13961
4	San Francisco	47.03%	892280
⋮	⋮	⋮	⋮
22	Los Angeles	34.82%	10257557
⋮	⋮	⋮	⋮
56	Tehama	19.58%	65885
57	Lassen	18.37%	30065
58	Kings	18.28%	156444

Table 1: Ranked counties based on percent of fully vaccinated individuals by May 8th out of the 16+ year old population.

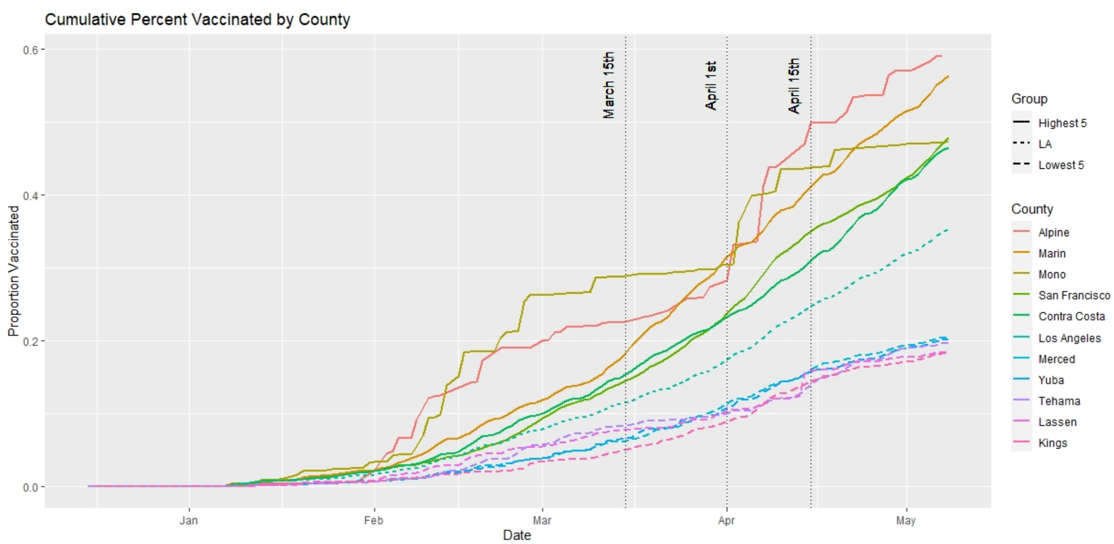

Figure 14: Cumulative vaccinated proportion from December 15th 2020 to May 8th 2021. Highest five counties and lowest five counties are based on vaccination proportion by May 8th.

In terms of population size, the five largest counties are Los Angeles, San Diego, Orange, Riverside, and San Bernardino and the five smallest are Alpine, Sierra, Modoc, Trinity, and Mono. In Figure 15, we see that Alpine, Mono, and Sierra achieve higher levels of vaccination than all of the largest counties while Modoc and Trinity end at a lower level of vaccination than all of the largest counties. This suggests that population size is not playing a highly influential role in proportion vaccinated. The slow rate in the largest counties are as expected with such large populations. On the other hand, slow rates in small counties are likely due to several other factors such as shortage in individuals to administer vaccines or unwillingness of the population to receive the vaccine.

Another concern with county population, is the amount of vaccines being used by a county. We expect the proportion of doses used in a county out of all doses used in California should be about equal to the proportion of the population of a county out of the entire state population. For example, Los Angeles county makes up 25% of the state population and therefore should be responsible for 25% of the doses administered in the state. In Figure 16 we see that the relationship between population proportion and dose proportion is linear with an estimated slope of 1.009 (95% CI: $[0.9794, 1.0386]$, p-value $< 2.2 \times 10^{-16}$). The linear regression gives $R^2 = 0.99$ which suggests that variation in population sizes accounts for nearly all of the variation in number of doses used. This slope also suggests that counties are using the appropriate number of doses relative to population size and so again this suggests that low vaccination rates are likely due to other demographic factors like race, gender, and age.

A highly interesting component of vaccination willingness is political ideology. Political ideology often encompasses several other demographics such as age, race, and education and often provides a metric for understanding behavior. Studies early in the pandemic found that self-identified conservatives reported lower willingness to receive the vaccine

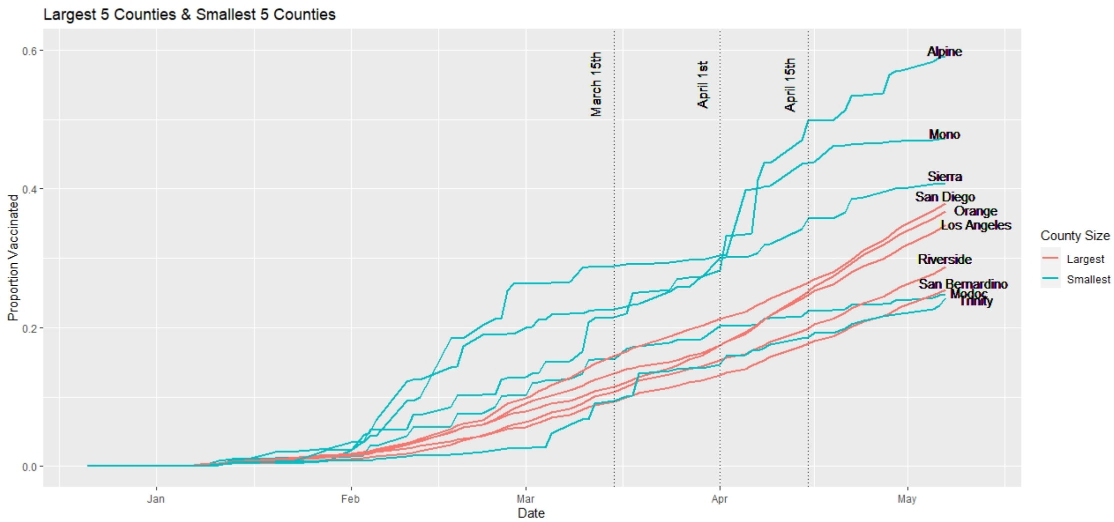

Figure 15: Cumulative fully vaccinated proportion over time in the largest five and smallest five counties.

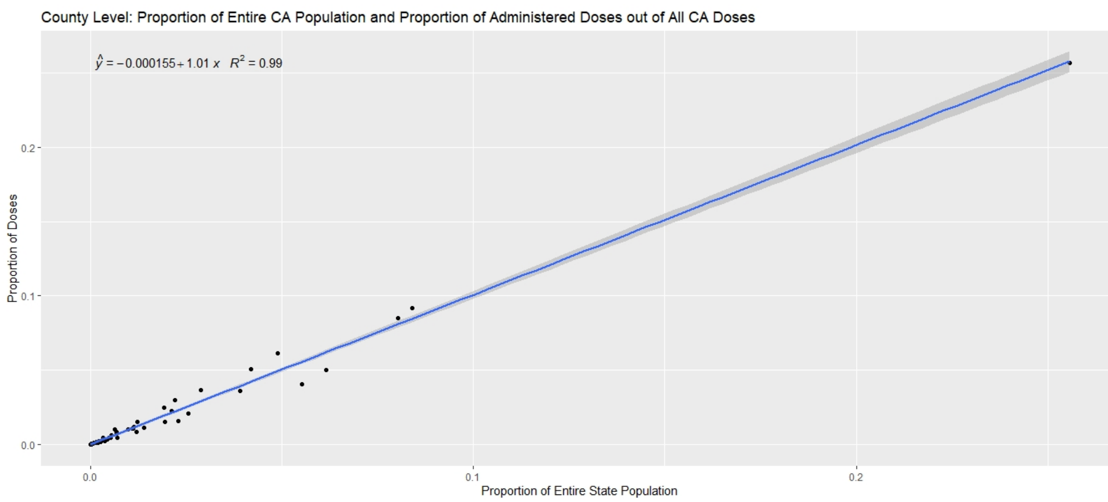

Figure 16: Counties as proportion of entire CA population vs the proportion of doses used by the county out of all doses in CA.

compared to self-identified liberals [14]. Studies also showed that supporters of former President Donald Trump were more susceptible to believing conspiracies and had more vaccination concerns compared to other Americans [15]. For this analysis, political ideology of a county was defined as either "More than 50% Vote for Biden" or "Less than 50% for Biden" based on the 2020 election results [16].

From Figure 17 we see that almost all of the counties follow the same trajectory until March 15th. After mid March, we see an increase in between-county variation with some counties showing an increase rate of vaccination while others slow down. We see that the counties that continue with high trajectories tend to have >50% Biden vote while the slower moving counties have <50% Biden vote. In Figure 18 we see the relationship between proportion voted for Biden in a county and the proportion fully vaccinated by May 8th. A linear regression gives a slope of 0.43941 (95% CI: $[0.3302, 0.5486]$, p-value $< 6.188 \times 10^{-11}$). This significant and positive slope suggests political preference in the 2020 election is an important predictor of vaccination. While this slope is significant, the $R^2 = 0.537$ suggests that there are more factors influencing the differences in vaccination rates across counties. The differences in vaccination rates can be better understood by disaggregating results by demographic groups.

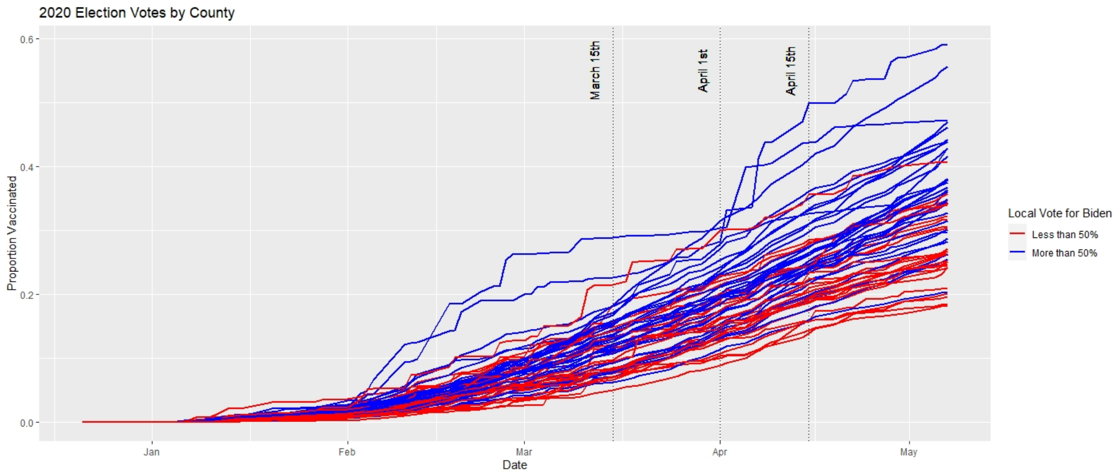

Figure 17: Proportion of a county vaccinated by May 8th and the proportion of a county that voted for Biden in 2020 election.

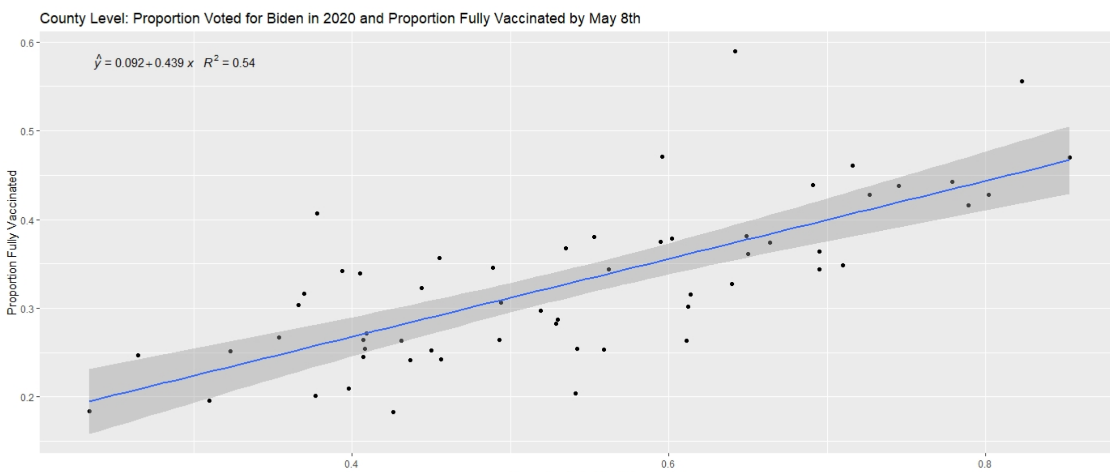

Figure 18: Proportion of a county vaccinated by May 8th and the proportion of a county that voted for Biden in 2020 election.

5 Vaccination Ratio amount Different Demography Groups

5.1 Methods

Vaccination data by demography and vaccination data by county and demography are used. Vaccination data by county and demography includes the estimated population of each demography group by county so it can be used to present vaccination rate in ratio. Vaccination data by demography includes gender variable while another one does not. These data were from December 15, 2020 to May 8, 2021, the 100th day of Biden administration.

For the study of race and age group, ratio of fully vaccinated population is used instead of using vaccinated population. Gender does not use this method because the gender ratio in California is approximately 1:1.

5.2 Results & Discussion

Three demography groups, gender, age, and race, were analyzed. The outcomes are the cumulative vaccinated population or cumulative vaccination rate.

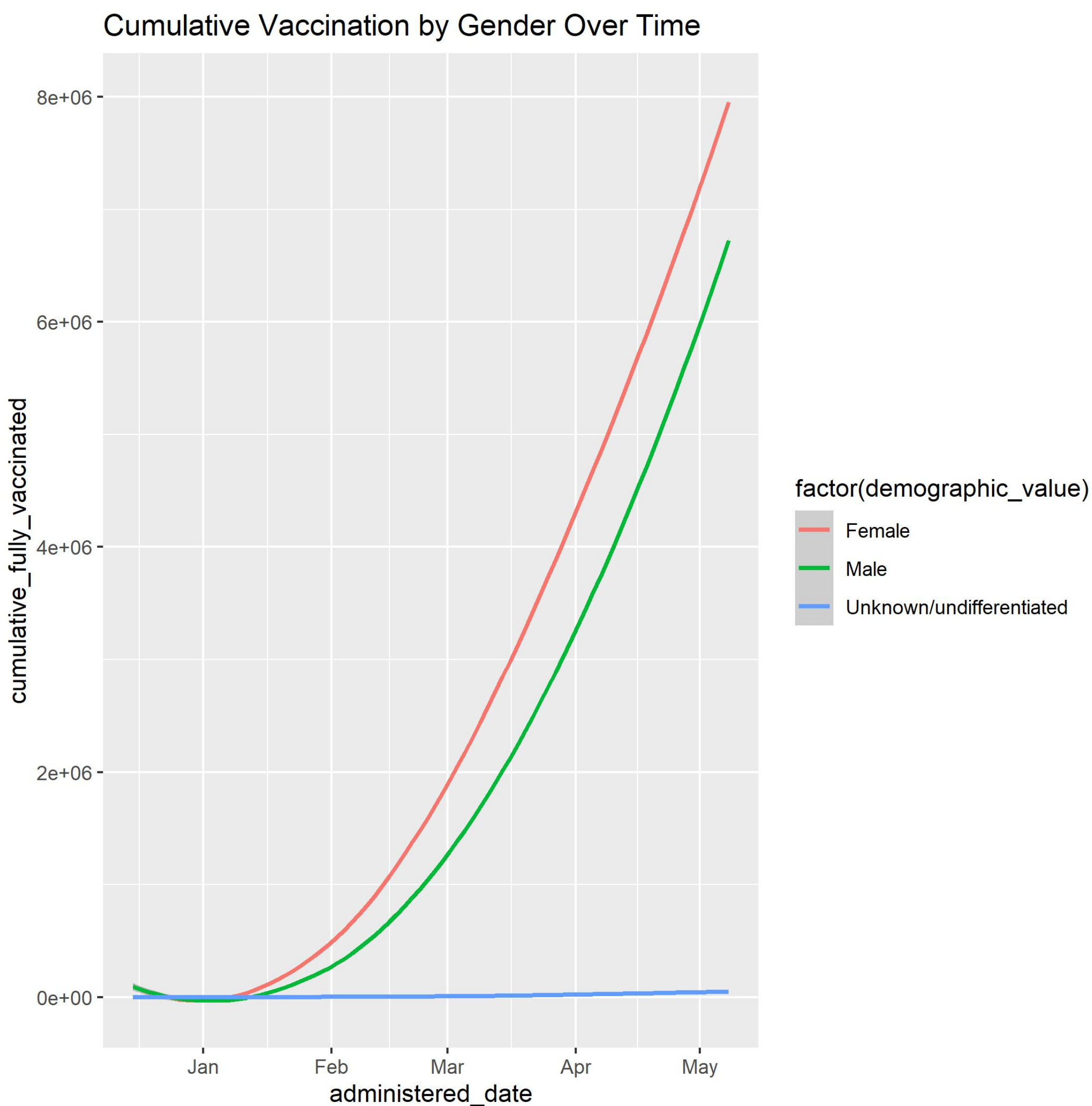

Figure 19: Fully Vaccination Population of Male, Female, and Unknown Gender in California by Date.

As Shown in Figure 19, the ratio of California is 99 men to 100 women, almost 1:1. It is interesting that in California, women are more willingly to get vaccine than men. In May 08, 2021, 7873689 women were fully vaccinated and 6681003 men were fully vaccinated, cumulatively; the difference was 1192686 people.

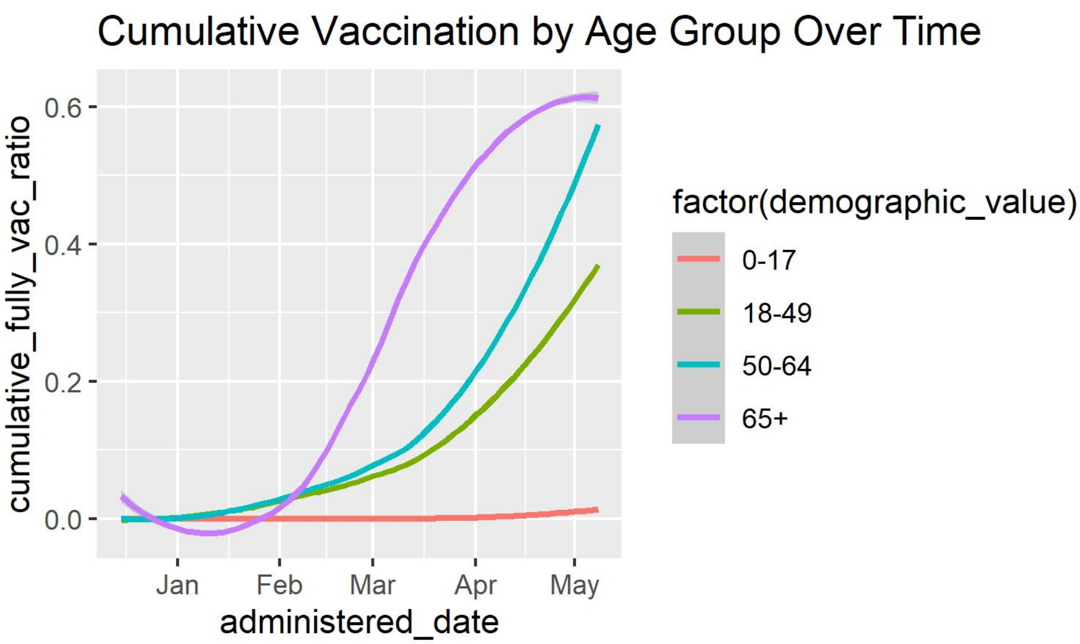

Figure 20: Fully Vaccination Population of Male, Female, and Unknown Gender in California by Date.

As Shown in Figure 20, the ratio of fully vaccinated 65+ people seems below 0; this is a consequence of using loess as geom smooth method, and the real data are above 0. The vaccination rate for elderly people increased rapidly after February 1, 2021.

The effect of California vaccination eligibility policies are not significant based on this plot. Beginning April 1, 2021, individuals age 50-64 years old would be eligible for vaccines; beginning April 15, 2021, every Californian age 16 and older would become eligible for vaccines. However, the slope of 50-64 does not increase after April 1, and the slope of 18-49 does not increase after April 15. One possible explanation is, many people between 18 to 64 got vaccines early as healthcare workers (Phase 1A) or food/agriculture, education/childcare, and emergency services workers (Phase 1B). The vaccination rate is very low for age 0-12 group because they did not have these opportunities to get vaccines early and the eligibility date for them is May 12, 2021.

As Shown in Figure 21, the vaccination rate of all races/ethnicity grew constantly with no significant change on the growth rate. Up to May 8, 2021, Native Hawaiian or other Pacific Islander has the highest vaccination rate, then it follows by Asian and White. Multiracial and Latino have the lowest vaccination rate.

There is no significant changes on the growth rate of vaccination, means there was no vaccination policy had significant racial discrimination.

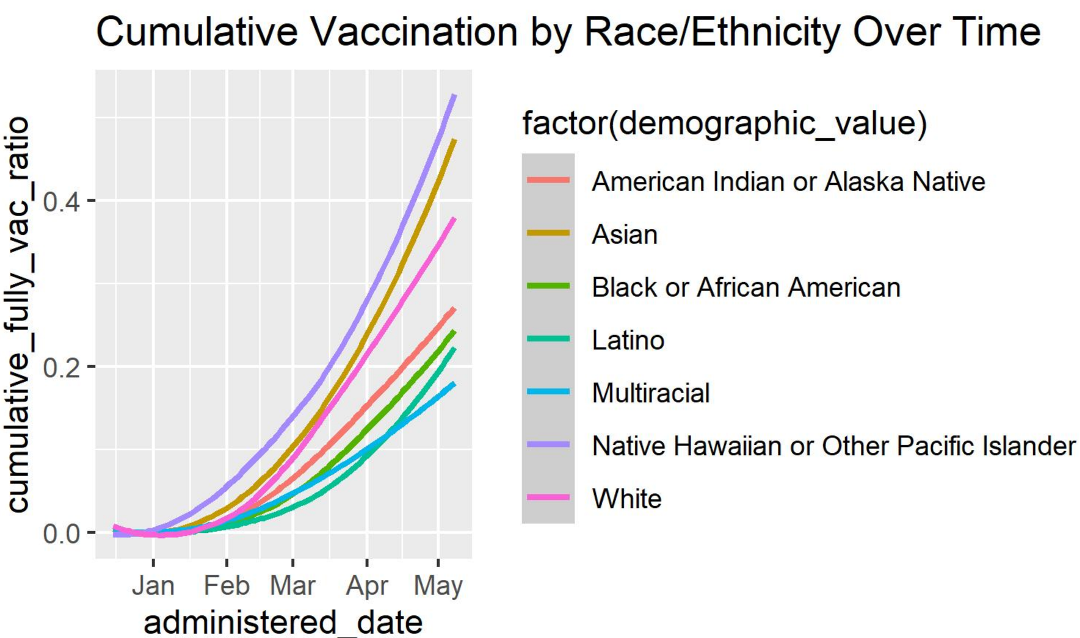

Figure 21: Fully Vaccination Population of Different Age Groups in California by Date.

6 Vaccine Equity

The Blueprint for a Safer Economy was instituted in California on August 30, 2020 in effort to assist counties with preventing the spread of COVID-19 and reduce the number of COVID-19 cases. California claims it views equitable vaccine administration as a mechanism for achieving safe and effective economic activity, prioritizing the hardest hit communities. We aim to investigate these claims by observing the impact of vaccine administration efforts on counties' Tier Assessment over time.

Tier Assignments are determined by a combination of variables but mainly Test Positivity (Excluding prison cases, 7 day average with 7 day lag) and Adjusted Case Rate for Tier Assignment(Rate per 100,000 population excluding prison cases, 7 day average with 7 day lag). There are four Tiers; Tier 1 is the most restrictive tier and Tier 4 is the least restrictive tier. Counties with populations below 106,000 were scaled differently.

As we gain more information about the spread and state of COVID-19 and as we reduce the risk of community transmission by vaccination, The Blueprint plans to adjust accordingly. The first change in Tier qualifications was implemented March 4th, 2021 with the notable difference being Case Rate must exceed 10 per 100,000 population to be Tier 1 where the previous requirement was greater than 7 per 100,000 population. The next two adjustments to Tier qualifications are to occur when two and four million vaccines are administered overall in California.

6.1 Methods

The data is compiled from Blueprint Data Charts from weeks '12-15-2020' to '3-15-2021', excluding week '2-23-2021' as it is not listed. The 'CA' observation was removed for weeks '4-06-2021' and '4-27-2021' since this observation was unique to these two weeks. For our initial interest in Tier movement, we subset our compiled data to include the variables County, Date of Tier Assessment, Final Tier Assessment, and Previous Tier Assessment. With this data and variables of interest, we were able to recover data for the week 2-23-2021. We then observe how many weeks each county spent in each Tier.

6.2 Results & Discussion

The Blueprint for a Safer Economy is a recognition that COVID-19 has not impacted all communities equally. It states "[a] heightened ability to protect individuals in these disproportionately impacted communities not only improves health in those communities but has critical impacts on the state as a whole". Although it is said that there is particular interest to vaccinate low vaccine equity quartile communities, specific efforts aren't listed.

Looking into Tier progression by county over the vaccine rollout period (Fig. 22), we see an overall trend into less restrictive tiers. On February 23rd, 2021 Trinity was the last county to be assigned a tier more restrictive than its' most recent tier. The only other incidence similar to this, in this time frame, occurred January 4th, 2021 in Humboldt County.

Recall that Tier 1 is the most restrictive Tier and Tier 4 is the least restrictive Tier. We may consider Tier Assignments as scores, the higher the overall score, the more a county was assigned less restrictive tiers. The lower the overall score, the more a county was assigned to more restrictive tiers. Table 2 specifies the ten counties with the five highest and lowest scores.

Rank	County	Score
1	Sierra	74
2	Alpine	69
3	Mariposa	54
4	Trinity	53
5	Lassen	46
...	...	...
54	Nevada	30
55	Inyo	29
56	Yuba	29
57	San Joaquin	28
58	Merced	26

Table 2: County Overall Tier Scores

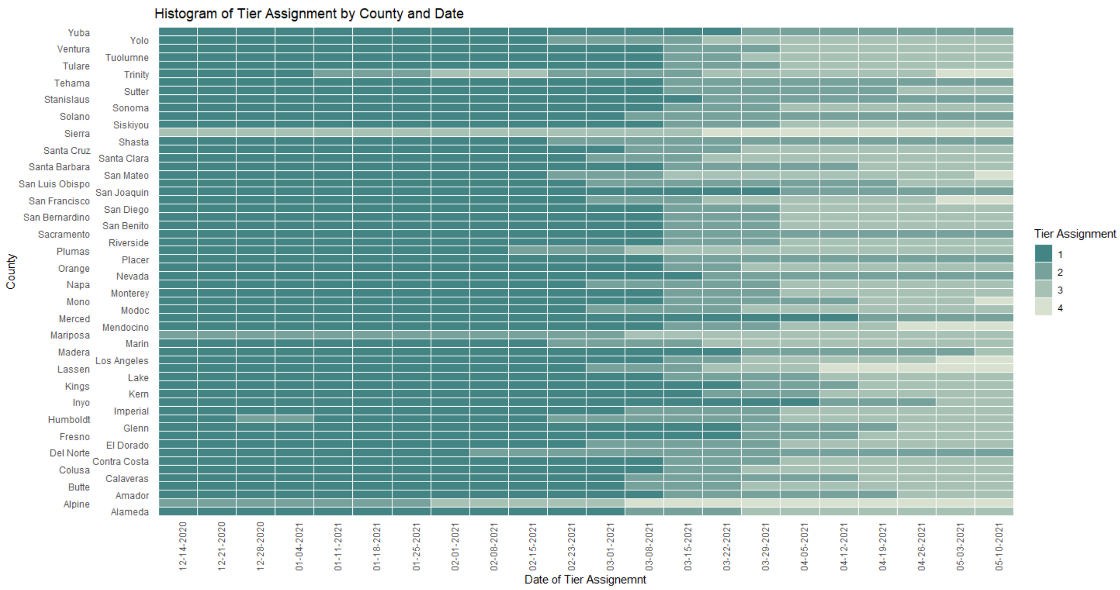

Figure 22: Tier Assignment over Time

The average score was 38 with a standard deviation of 8.323 meaning Sierra is 4.33 standard deviations above the mean and Merced is 1.44 standard deviations below the mean. These scores are heavily skewed with four outliers above the mean and zero below.

Next, we calculate how many weeks each county spent in each tier. From Figure 23, it appears that most counties spent more weeks in Tier 1, quickly transitioned through Tier 2, and are currently either in Tier 3 or Tier 4. Over this twenty-two week time span, California's fifty-eight counties spent 685 weeks in Tier 1, 283 weeks in Tier 2, 274 weeks in Tier 3, and 34 weeks in Tier 4.

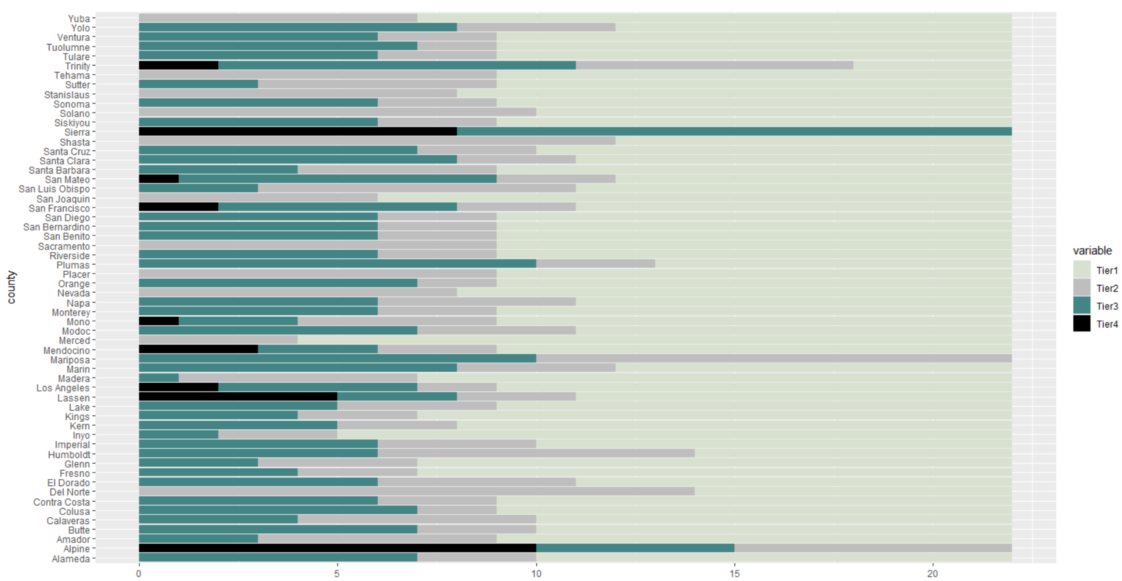

Figure 23: Number of Weeks in Each Tier

Although California's Blueprint includes efforts to make counties comparable, it should be noted that initial efforts to move out of Tier 1 could have been different for each county. We then may want to regard the progression of a county who began with a high Case Rate, say 16, and was able to achieved Tier 2 status in six weeks differently than a county who began with a lower Case Rate,say 11, and also achieved Tier 2 status after six weeks.

Whether or not the Blueprint was efficient in prioritizing harder hit communities is difficult to determine from what we've discussed in regards to Tier Assignments. However, Figure 16 suggests proportion of doses used is highly correlated with proportion of population, which seems to leave little room for considerations of disproportionately hit communities.

7 Conclusion

For most counties in California, there is an increasingly strong correlation between the rise in vaccination rates and decrease in infection rates. This observation is promising, but also sheds light on communities that have had a harder time recovering from the pandemic and likely need more assistance than their neighbors. This project looked at all 58 counties across the state with the objective of understanding the COVID-19 vaccines' impact and through our various methods we have identified which counties have stood out- for better or worse. By doing so, we have laid a path for decision-makers to further support communities in need and continue investigating where these disparities may originate from.

Hospitalization rates did decline as vaccination numbers increased in all counties. However, due to potential confounders, hospitalization decrease cannot be solely attributed to vaccinations, and further study is warranted. Additionally, Hospital bed counts did not effectively predict a county's vaccination rates. This is not a surprise because of the complex nature of healthcare and the specifics of vaccine distribution. As vaccine availability increased, regional patterns emerged across the state. By May 8th some counties achieved more than 50% vaccination while others failed to even reach 20%. The Tahoe region and Bay Area showed had some of the best vaccinated rates in the state but rural counties in the north and central parts of the state struggled to maintain a strong path to high vaccination levels. The most populous counties in the state all showed an expected slow but steady increase and these counties could likely reach high levels of vaccination within the next few months. Analysis comparing the population as proportion of the state and doses as proportion of all doses suggests that counties are using the appropriate amount of doses with respect to population size and differences in vaccination rates are likely due to other demographics.

Different demography groups have different vaccination rates. In California, Women are more willing to get vaccine than men. Native Hawaiian or Other Pacific Islander has the highest vaccination rate while Multiracial people has the lowest. The vaccination policies have no discrimination on gender or race/ethnicity, so these differences are probably caused by different cultures and social division of labor. People live in islands are less vulnerable to the pandemic because they are living in isolated region, but islanders have the highest vaccination rate. This situation makes them much safer than other races. Age group is affected by the vaccination policies. People of age above 65 are eligible very early, but their vaccination rate increased slowly at the beginning and increased fast after February. Even though there were two policies for the eligibility of people age 18-29 and 50-64, but the vaccination rates of these age groups are not significantly affected by these policies. Change of the slope at April 1 for 50-64 and April 15 for 18-49 were expected before this study, but these changes are not observed. One possible explanation is they got vaccine early for working reasons, such as healthcare worker and essential worker. The low vaccination rate of age 0-17 also support this hypothesis, because more of people in this age group are not eligible to work, and they had no chance to get vaccines for working reasons.

The movement of counties through Tier Assignments is a reflection of California's decreasing COVID rates. Due to qualifications needed to transition into less restrictive Tiers, our image of the state through the lens of Tiers may be a few weeks behind our reality. It is currently unclear if Blueprint efforts encouraged vaccine equity.

8 Acknowledgements

We sincerely thank Prof. Fushing Hsieh and Ben Roycraft for their feedback and assistance with this project. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1650042.

References

[1] Soumya Karlamangla and Rong-Gong Lin Il. How the ‘california miracle’ dissolved into a winter coronavirus nightmare. *Los Angeles Times*, Jan 2021.

[2] Ellina Abovian. California sees record-breaking covid-19 deaths, a lagging indicator of winter surge. *Los Angeles Times*, Jan 2021.

[3] Susie Neilson. The bay area is racing ahead of the country on vaccines. *San Francisco Chroncile*, May 2021.

[4] Centers for Disease Control and Prevention. Covid-19 vaccines work. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/work.html. Accessed: 2021-05-20.

[5] California COVID-19 Vaccination Program. Statewide covid-19 vaccines administered by county. https://covid19.ca.gov/vaccines/. Accessed: 2021-05-11.

[6] COVID-19 Response Data, Informatics, Surveillance, Clinical and Outbreaks (DISCO) Team. Statewide covid-19 cases deaths tests. https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/Immunization/ncov2019.aspx. Accessed: 2021-05-11.

[7] Elizabeth Chou, Yin-Chen Hsieh Catie McVey, Sabrina Enriquez, and Fushing Hsieh. Extreme-k categorical samples problem, July 2020.

[8] Nigel Bourne and Gregg N Milligan. Control and eradication of human and animal diseases by vaccination, chapter 3. *Vaccinology: An Essential Guide*, pages 43–58, 2015.

[9] California Hospital Association. Covid-19 hosptializations by county. https://data.chhs.ca.gov/dataset/2df3e19e-9ee4-42a6-a087-9761f82033f6/resource/15e2b847-d9c2-4523-bdc1-f22020da079e/download/covid-hospitalizations-data-dictionary.xlsx. Accessed: 2021-05-17.

[10] Alix Martichoux and Grace Manthey. California covid-19 vaccine tracker: How the state is doing, when you can get a coronavirus vaccine. *abc7 news*, 2021. https://abc7news.com/covid-vaccine-california-tracker-ca-newsom-update/9567632/ (visited 25May2021).

[11] Institute of Medicine (US) Committee on Monitoring Access to Personal Health Care Services and Millman M (editor). *Access to Healthcare in America*, chapter 2, A Model for Monitoring Access. National Academic Press (US), Washington (DC), 1993.

[12] Van Tieu. Lone hospital in yuba-sutter counties overwhelmed by coronavirus cases. *abc10*, 2020. https://www.abc10.com/article/news/health/coronavirus/lone-hospital-in-yuba-sutter-counties-overwhelmed-by-coronavirus-cases/103-e8825471-18b4-4afa-94cb-6c41a900ce97 (visited 08June2021).

[13] COVID-19 Response Data, Informatics, Surveillance, Clinical and Outbreaks (DISCO) Team. Updated covid-19 vaccine eligibility guidelines. https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/VaccineAllocationGuidelines.aspx. Accessed: 2021-06-08.

[14] Carl Latkin, Lauren A Dayton, Grace Yi, Arianna Konstantopoulos, Ju Park, Catherine Maulsby, and Xiangrong Kong. Covid-19 vaccine intentions in the united states, a social-ecological framework. *Vaccine*, 39(16):2288–2294, 2021.

[15] Matthew J Hornsey, Matthew Finlayson, Gabrielle Chatwood, and Christopher T Begeny. Donald trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy. *Journal of Experimental Social Psychology*, 88:103947, 2020.

[16] Allan James Vestal, Andrew Briz, Annette Choi, Beatrice Jin, Andrew McGill, and Lily Mihalik. Election 2020 california results. https://www.politico.com/2020-election/results/california, Nov 2020.