

## Group information

**Group members:** Jason Dinh, Kyle Huang, and William Fu

# PSTAT 100 Project Report

---

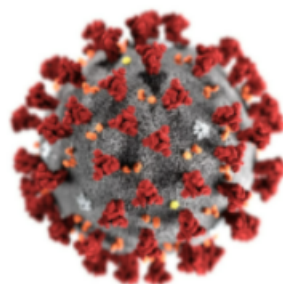
## 0. Abstract

Our report looks into COVID vaccinations, vaccine equity metric quartile, and political parties by each county or zip code. Our goal in this report is to see whether higher vaccine equity metric quartile means higher COVID vaccinations as well as whether higher vaccinations is tied to a Democratic majority county. We will look into the various plots to analyze the data and ultimately come up with a conclusion if there is any correlation.

---

## 1. Background

COVID-19 is caused by a coronavirus called SARS-CoV-2. Older adults and people who have severe underlying medical conditions like heart or lung disease or diabetes seem to be at higher risk for developing more serious complications from COVID-19 illness.



COVID-19 vaccines have been authorized for emergency use by the FDA due to the severity of COVID-19 in the United States.

This project seeks to see if different Vaccine Equity Metric Quartile shows different percentages of vaccinations in California. One of the data gives us doses from partial to full for each zip code in California and the other data tells us each counties' resident's political party in California. The motivation for collecting this data is to see how each community is doing on administering doses of the COVID-19 vaccine. We could potentially learn how different healthy community conditions have different accessibility of vaccines as well as learning if political parties of counties correlates with the amount of people vaccinated.

## 2. Data description

The data are Covid-19 vaccine doses of May 12, 2021 and it shows full, partial, and at least 1 dose coverage rates between different zip codes in California. There is also a dataset for registered people in California and their political party by county.

The data comes from the California Open Data Portal and the California Secretary of State, Dr. Shirley N. Weber. The COVID-19 sources are from California Immunization Registry 2015-2019 5-Year data. The political party sources are from the Report of Registration of February 10, 2021.

<https://data.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data-by-zip-code> & <https://www.sos.ca.gov/elections/report-registration/ror-odd-year-2021>

### Sample and measurement information

The data values were obtained through records of the California Immunization Registry and Report of Registration.

### Data Structure

For this study, the observational units of the vaccine and report of registration data are both persons. The variables are Zip Code Tabulation Area, Vaccine Equity Metric Quartile, the population of 16+, people partially or fully vaccinated, and the percent of partially, fully, or 1+ dose.

Table 1: variable description and units for each variable in the vaccination dataset

Name	Variable description	Type	Units of measurement
Zip Code Tabulation Area	Zip Code	Nominal	None
Vaccine Equity Metric Quartile	Health equity score quartile	Numeric	None
16+ Population	Total population aged 16+	Numeric	Persons
Persons Fully Vaccinated	Total number of people fully vaccinated	Numeric	Persons
Persons Partially Vaccinated	Total number of people partially vaccinated	Numeric	Persons
Percent of Population Fully Vaccinated	Percent of people fully vaccinated	Numeric	None
Percent of Population Partially Vaccinated	Percent of people partially vaccinated	Numeric	None
Percent of Population with 1+ Dose	Percent of people fully or partially vaccinated	Numeric	None

Table 2: example rows of vaccination data

	Zip Code	Tabulation Area	County	Vaccine Equity Metric Quartile	16+ Population	Persons Fully Vaccinated	Persons Partially Vaccinated	Percent of Population with 1+ Dose
0	94555	Alameda		4.0	29202	17056.0	5948.0	0.787754
1	94618	Alameda		4.0	14227	10004.0	1878.0	0.835173
2	94544	Alameda		2.0	63279	29493.0	10898.0	0.638300
3	94545	Alameda		3.0	27083	15052.0	4881.0	0.735997
4	94577	Alameda		3.0	39920	21952.0	6506.0	0.712876

Table 3: variable description and units for each variable in the report of registration dataset

Name	Variable description	Type	Units of measurement
County	County	Nominal	None
Eligible	Total amount of population eligible to register	Numeric	Persons
Total Registered	Total amount of population registered to vote	Numeric	Persons
Democratic	Total amount of population registered Democratic	Numeric	Persons
Republican	Total amount of population registered Republican	Numeric	Persons

Table 4: example of rows of report of registration data

	County	Eligible	Total Registered	Democratic	Republican
0	Alameda	1094829	961872	573201	105931
1	Alpine	925	896	390	219
2	Amador	27124	25989	7241	12114
3	Butte	148972	125269	44399	44754
4	Calaveras	35958	32017	8628	14799

---

## Aims

Let's take a look at our questions we have for these datasets.

1. *Do different VEM scores show different percentages of vaccinations?*
2. *Is there any correlation between political parties and vaccinations by county?*

---

## 4. Methods

Exploratory analysis aimed at what factors are correlated with vaccinations. Specifically, we will analyze the VEM scores and political parties and compare them to vaccinations in different zip codes and counties. Higher VEM scores suggest more accessibility to health care, however, we want to see if this will have a direct link to people getting vaccinations. If counties have a majority of a specific political party, we want to see how their vaccination rates compare to the vaccination rates in another county where the majority is the opposing political party. To analyze this, we will utilize visualization through point-and-line graphs and bar plots. In addition, we are going to use a scatter plot comparing VEM scores to percentage of population with at least 1 dose. We can also get a trend line from this which would tell us how VEM score affects vaccination rates. We will also be utilizing a scatter plot of proportion to compare democratic and republican majority counties. From this, we can make general conclusions on how factors like VEM and political parties correlate to vaccination rates.

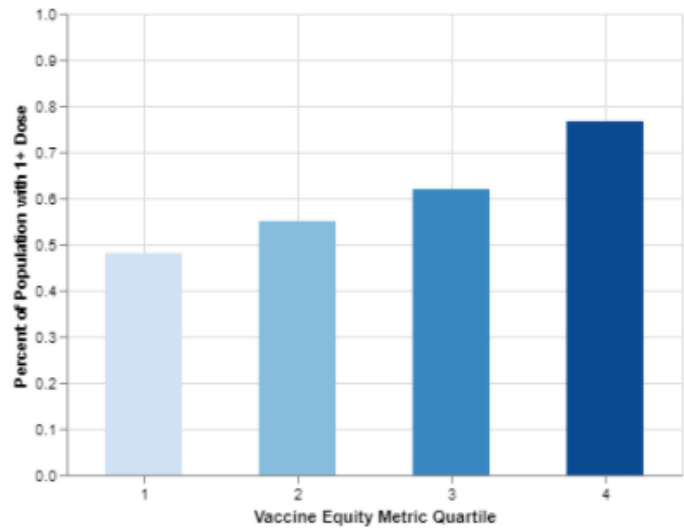
---

## 5. Results

### Average percentage of 1+ dose by the zip code's vaccine equity metric quartile

Exploratory analysis focused on the correlation between the average percentage of 1+ dose by the zip code's vaccine equity metric quartile. Vaccine Equity Metric Quartile is the health equity metric score quartile, with 1 indicating least healthy community conditions and 4 indicating healthiest community conditions. Figure 1 shows the average percentage of 1+ dose by their vaccine equity metric quartile of all zip codes.

Figure 1:

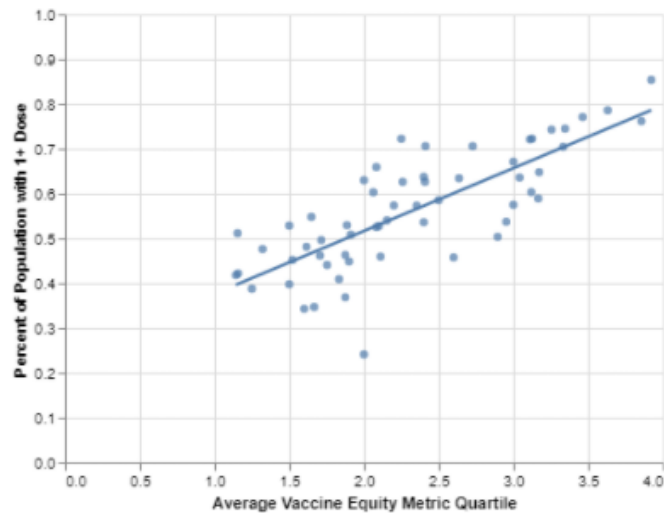


From this, there is an obvious correlation with higher vaccinations and higher vaccine equity metric quartile.

### Average percentage of 1+ dose by the county's average vaccine equity metric quartile

Instead of measuring by zip codes, we instead measure by county here. Since we are measuring by county, we find the average vaccine equity metric quartile of the county. Figure 2 shows that average percentage of 1+ dose by the average vaccine equity metric quartile of all the counties.

Figure 2:



Looking at the scatter plot and trend line, we can see that there is an upwards trend of vaccinations as the vaccine equity metric quartile increases.

## Finding the correlation between political parties and their vaccinations

From the second dataset, we can see which counties are dominated by what political parties with blue being Democrats and red being Republican. While there are many parties, we want to specifically focus on Democrats and Republicans. Figure 3 shows the percentage of each political party and their corresponding county.

Figure 3:

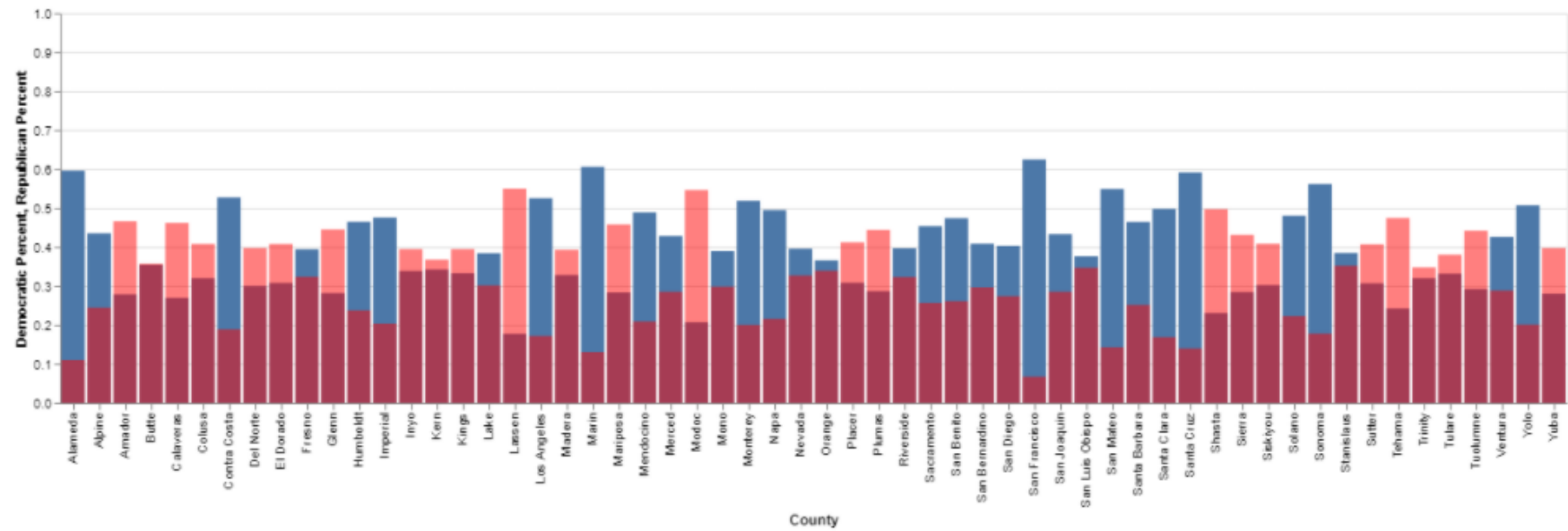
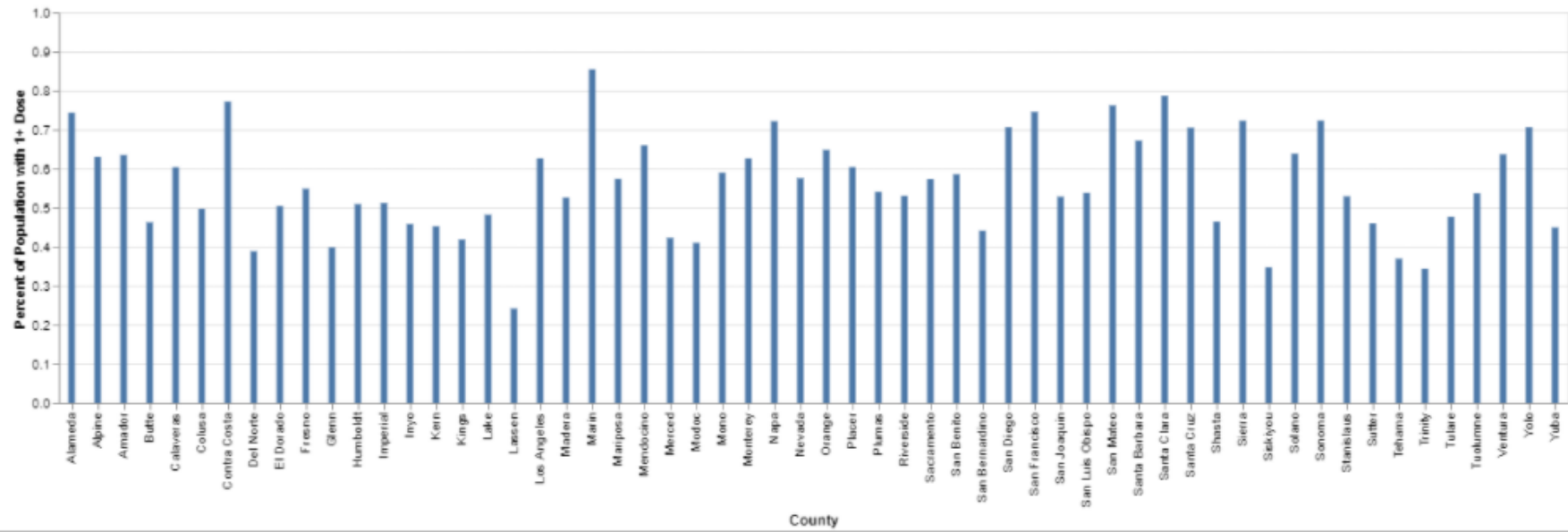


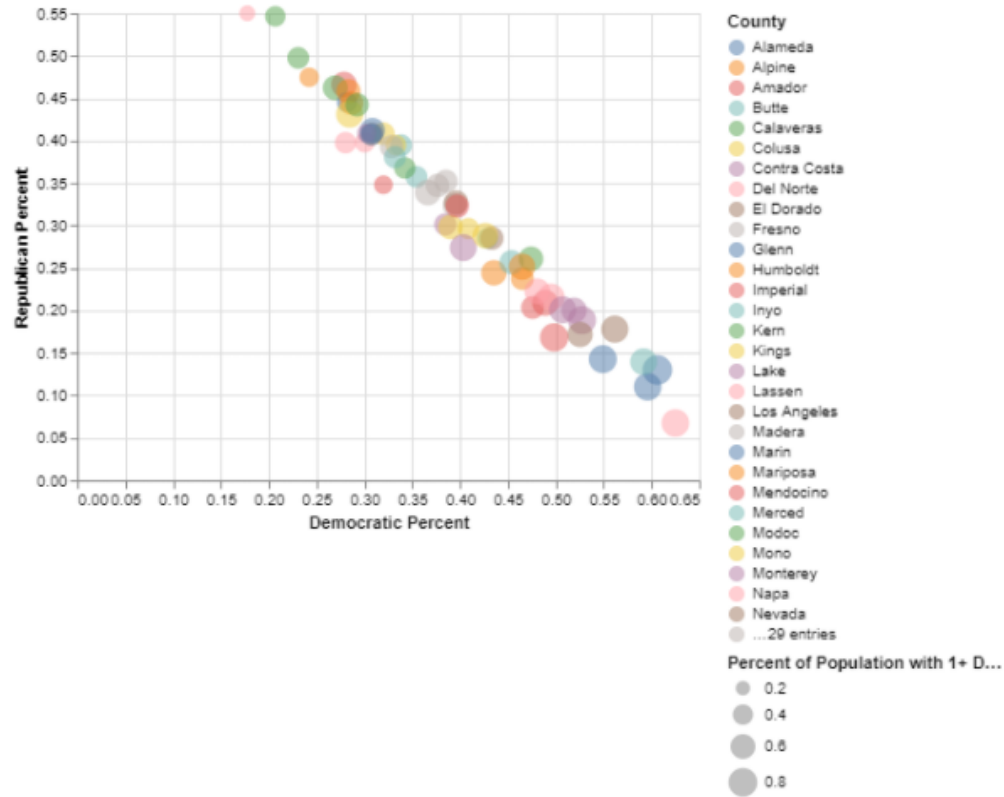
Figure 4 shows us the average percentage of 1+ dose of each county.

Figure 4:



We want to see if there is any correlation between political parties and their vaccinations. Figure 5 shows a scatter plot with the amount of vaccinations shown by how big the circle is.

Figure 5:



From figure 5, if we split the scatter plot in half with Republicans on the left side and Democrats on the right side, we can see that the smaller circles are on the Republican dominated side while the bigger circles are on the Democrat dominated side.



---

## 6. Discussion

This project analyzed the correlation between different California zip codes/counties and vaccine equity metric quartile as well as the correlation between political parties and the counties' vaccinations using data recorded from the California Immunization Registry 2015-2019 5-Year data and the Report of Registration of February 10, 2021. The analysis identified that there is a correlation of higher vaccination rates of counties with higher vaccine equity metric quartile and lower vaccination rates of counties with a lower vaccine equity metric quartile. The analysis also identified that there is a correlation of higher vaccination rates of counties that are Democrat dominated and lower vaccination rates of counties that are Republican dominated.