

Milestone #5

Saira Mayet, Jessica Pak, Andrew Nguyen

11/24/2021

Project Statement

We are monitoring COVID-19 vaccination rates among counties in California and in relation to age. Specifically, we are interested in whether there is any correlation between median age and vaccinated person prevalence at the county level. Utilizing two datasets, the California Census Data and the COVID-19 Vaccine Progress Dashboard, we intend on exploring, analyzing, and visualizing vaccination rates at the county level and the relationship between age and vaccination rate.

Methods

Of the two datasets of interest, the first describes COVID-19 vaccination administration across the state of California. This is sourced from the California Open Data Portal for the California Department of Public Health. Relevant fields include date, zip code, county, and raw counts of vaccination status, among other population information. The dataset starts from January 5th, 2021 and is continually updated to reflect new changes in the data, most recently updated on November 23rd, 2021.

The second dataset describes demographics (e.g. population, race/ethnicity, age, household size, etc.) for each California county, most recently updated in September 2021. This dataset was rehosted on Avery Richard's GitHub, and is sourced from Census data.

Partial subsetting of data was initially completed to keep the following variables in each dataset:

COVID-19 Vaccination Dashboard: Date, Zip Code, County, Population 5+, Number of Persons Fully Vaccinated, Number of Persons Partially Vaccinated

County Census Data: Name, Median Age

New Variables

After subsetting the data from both sources, the data were cleaned and new variables were created.

Data Cleaning

1. Mean imputation: county-level means of fully and partially vaccinated that will be used to replace NA values in dataset
2. Percent eligible population partially vaccinated = $\# \text{ of persons partially vaccinated} / \text{population } 5+$ (at county level)
3. Percent eligible population fully vaccination = $\# \text{ of persons fully vaccinated} / \text{population } 5+$ (at county level)
4. Merging relational data: county demographic dataset with vaccine administration dataset using key variable "county".

Visualizations

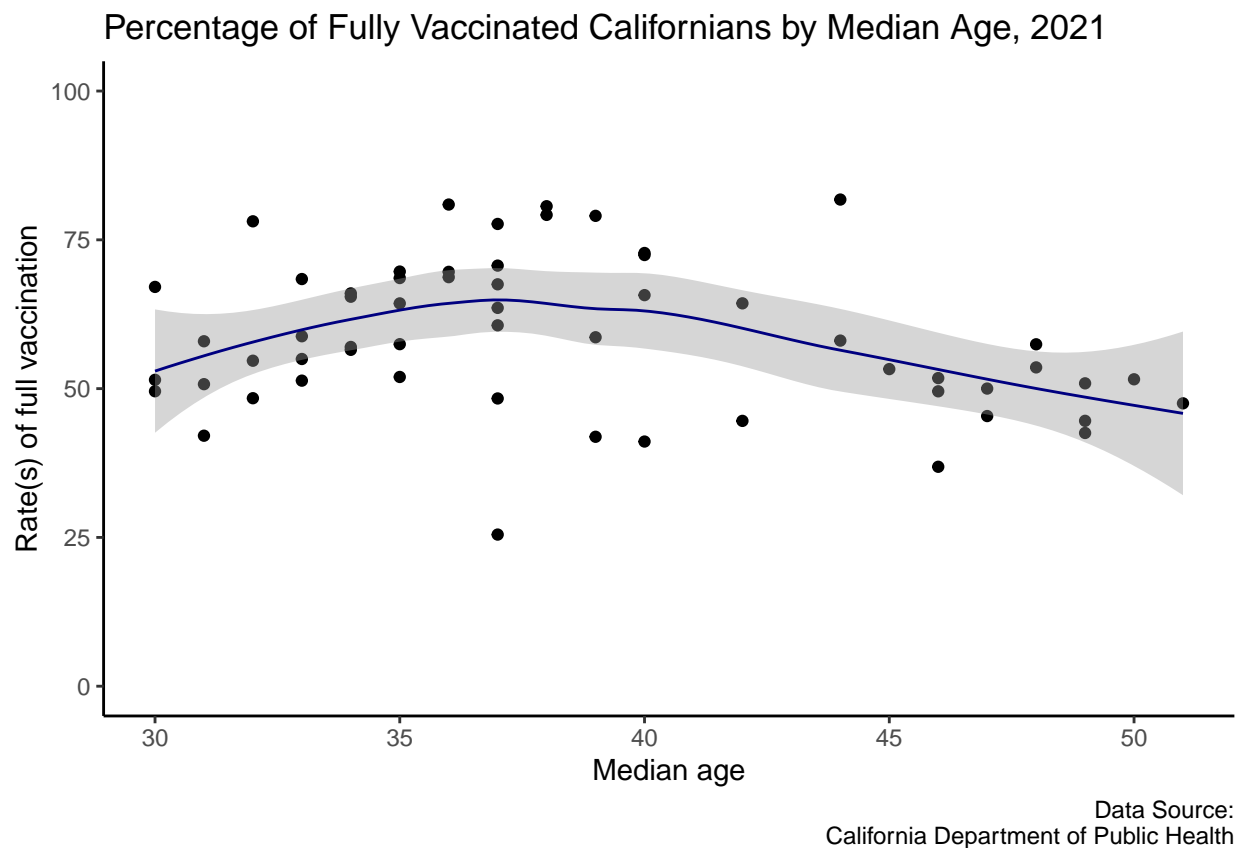
The Table, “Vaccination Rates for California Counties,” displays the partial vaccinated rate, fully vaccinated rate, total eligible population, and median age of each county in California. The summarized average rates, total eligible population, and median age for the entire state of California is calculated and presented at the top of the table.

Table 1: Vaccination Rates for California Counties

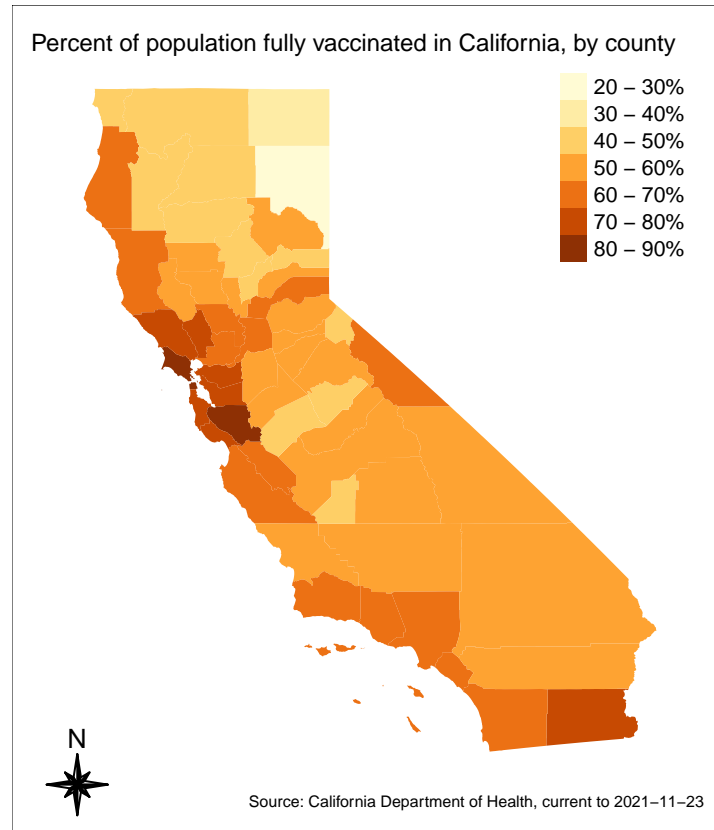
| California | 6.97 | 62.8 | 33,330,578 | 38 |
|-------------------|---------------------------|-----------------------|---------------------------|------------|
| County Name | Partially Vaccinated Rate | Fully Vaccinated Rate | Total Eligible Population | Median Age |
| Alameda | 7.67 | 77.67 | 1,565,553 | 37 |
| Alpine | 6.07 | 49.55 | 890 | 46 |
| Amador | 9.53 | 53.57 | 37,067 | 48 |
| Butte | 5.08 | 48.35 | 213,817 | 37 |
| Calaveras | 7.71 | 50.89 | 43,656 | 49 |
| Colusa | 6.52 | 56.52 | 20,153 | 34 |
| Contra Costa | 6.67 | 79.19 | 1,071,086 | 38 |
| Del Norte | 5.85 | 41.91 | 25,963 | 39 |
| El Dorado | 6.12 | 58.07 | 179,821 | 44 |
| Fresno | 7.60 | 57.96 | 911,080 | 31 |
| Glenn | 4.93 | 51.96 | 26,204 | 35 |
| Humboldt | 7.09 | 60.64 | 128,806 | 37 |
| Imperial | 18.50 | 78.11 | 161,453 | 32 |
| Inyo | 6.95 | 51.77 | 18,268 | 46 |
| Kern | 6.78 | 50.75 | 818,823 | 31 |
| Kings | 6.32 | 42.09 | 137,655 | 31 |
| Lake | 6.41 | 53.28 | 60,336 | 45 |
| Lassen | 3.06 | 25.48 | 26,919 | 37 |
| Los Angeles | 8.57 | 68.58 | 9,463,365 | 35 |
| Madera | 6.92 | 51.34 | 143,316 | 33 |
| Marin | 9.11 | 81.77 | 246,959 | 44 |
| Mariposa | 20.18 | 44.59 | 15,319 | 49 |
| Mendocino | 8.42 | 64.33 | 81,751 | 42 |
| Merced | 11.18 | 49.56 | 248,786 | 30 |
| Modoc | 3.36 | 36.87 | 9,384 | 46 |
| Mono | 7.68 | 67.53 | 12,259 | 37 |
| Monterey | 8.38 | 68.41 | 387,591 | 33 |
| Napa | 9.25 | 72.79 | 133,221 | 40 |
| Nevada | 7.43 | 57.45 | 92,519 | 48 |
| Orange | 6.94 | 69.62 | 2,986,910 | 36 |
| Placer | 6.24 | 65.71 | 367,860 | 40 |
| Plumas | 5.25 | 51.57 | 19,548 | 50 |
| Riverside | 6.85 | 57.00 | 2,255,664 | 34 |
| Sacramento | 6.85 | 64.32 | 1,427,122 | 35 |
| San Benito | 8.23 | 66.00 | 55,001 | 34 |
| San Bernardino | 6.14 | 54.69 | 1,991,511 | 32 |
| San Diego | 11.86 | 69.67 | 3,101,086 | 35 |
| San Francisco | 7.48 | 80.65 | 835,425 | 38 |
| San Joaquin | 9.36 | 58.80 | 688,728 | 33 |
| San Luis Obispo | 6.71 | 58.63 | 268,922 | 39 |
| San Mateo | 8.05 | 79.03 | 696,222 | 39 |
| Santa Barbara | 8.02 | 65.44 | 417,143 | 34 |
| Santa Clara | 7.22 | 80.93 | 1,833,854 | 36 |
| Santa Cruz | 6.57 | 70.67 | 278,046 | 37 |
| Shasta | 6.23 | 44.58 | 161,537 | 42 |
| Sierra | 3.15 | 47.53 | 2,632 | 51 |
| Siskiyou | 6.24 | 45.39 | 40,245 | 47 |
| Solano | 9.68 | 63.57 | 414,116 | 37 |
| Sonoma | 7.18 | 72.45 | 475,030 | 40 |
| Stanislaus | 8.95 | 54.98 | 505,820 | 33 |
| Sutter | 6.46 | 57.47 | 90,595 | 35 |
| Tehama | 4.67 | 41.11 | 66,959 | 40 |
| Trinity | 6.01 | 42.55 | 12,260 | 49 |
| Tulare | 7.06 | 51.48 | 420,906 | 30 |
| Tuolumne | 6.96 | 50.01 | 52,564 | 47 |

The scatterplot, “Percentage of Fully Vaccinated Californians by Median Age, 2021,” shows the rate of fully vaccinated individuals stratified by median age of California residents, as measured from California Department of Public Health’s latest online data repository for COVID-19. A trend line was fitted to follow the points on the plot, and a general trend of increasing fully vaccinated rates can be seen near the median age of 36-38.

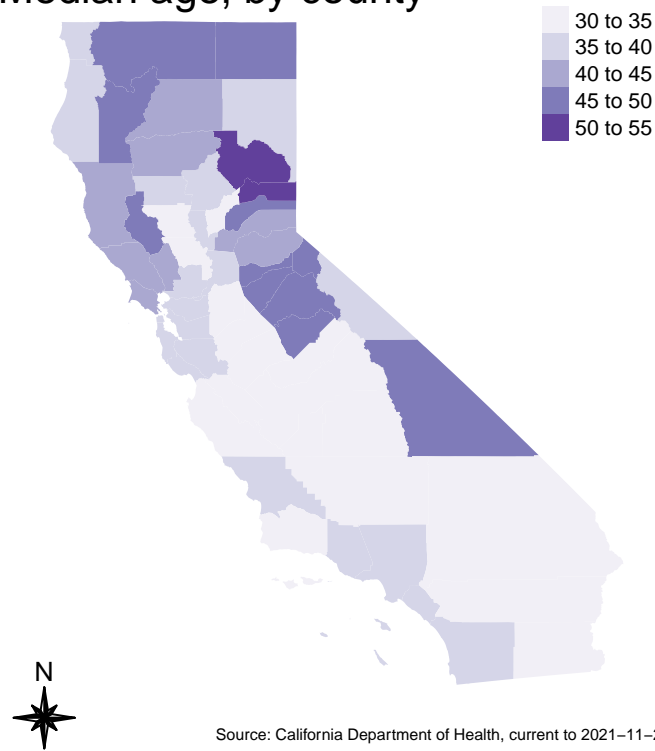
```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



The map shows the rates of full vaccination, by county, symbolized by a color ramp (from low to high rates of vaccination going from yellow to dark red). A color ramp was made with classes of 10 percent each. A second map shows the median age of each country, symbolized by a purple color ramp (younger to older median age going from light purple to darker purple). The five-step color ramp ranges from 30 to 55, with each step covering five years of age.



Median age, by county



This last visualization features points plotted by month and percent of full vaccination in aggregate ZIP codes falling under four categories of the Vaccine Equity Metric (VEM) quartiles, which is based on demographic indicators of socioeconomic disadvantage and health vulnerability (where 1 is least disadvantaged/vulnerable and 4 and most disadvantaged/vulnerable).

```
#get data per day at ZIP code level
vax_longitudinal <- read.csv(file_path_vax) %>%
  group_by(county) %>%
  mutate(county_partial =
    sum(persons_partially_vaccinated, na.rm = T)/
    sum(age5_plus_population, na.rm = T),
    county_fully =
    sum(persons_fully_vaccinated, na.rm = T)/
    sum(age5_plus_population, na.rm = T)) %>%
  ungroup() %>%
  mutate(persons_partially_vaccinated_2 =
    ifelse(is.na(persons_partially_vaccinated),
      age5_plus_population*county_partial,
      persons_partially_vaccinated),
    persons_fully_vaccinated_2 =
    ifelse(is.na(persons_fully_vaccinated),
      age5_plus_population*county_fully,
      persons_fully_vaccinated)) %>%
  group_by(county) %>%
  mutate(county_partial_2 =
    sum(persons_partially_vaccinated, na.rm = T)/
    sum(age5_plus_population, na.rm = T),
    county_fully_2 =
    sum(persons_fully_vaccinated, na.rm = T)/
    sum(age5_plus_population, na.rm = T),
    county_eligible_pop =
    sum(age5_plus_population, na.rm = T)) %>%
  ungroup()

#create aggregate dataset at VEM quartile level
vax_long_agg <- vax_longitudinal %>%
  group_by(as_of_date, vaccine_equity_metric_quartile) %>%
  summarize(percent_full_vem =
    sum(persons_fully_vaccinated, na.rm = T)/
    sum(age5_plus_population, na.rm = T)) %>%
  ungroup() %>%
  filter(is.na(vaccine_equity_metric_quartile) == FALSE) %>%
  mutate(date = as.Date.character(as_of_date)) %>%
  select(- as_of_date)
```

'summarise()' has grouped output by 'as_of_date'. You can override using the '.groups' argument.

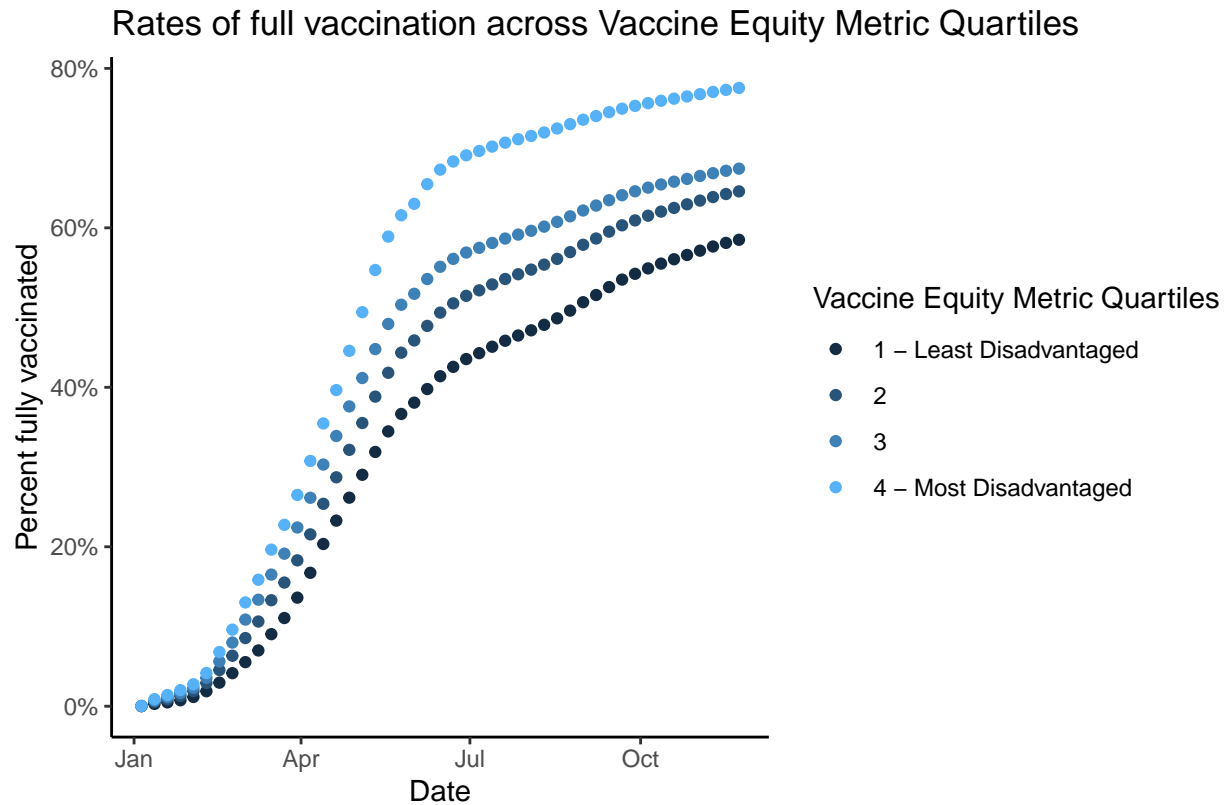
```
#Create longitudinal scatterplot visualization
ggplot(data = vax_long_agg, aes(x = date, y = percent_full_vem)) +
  geom_point(stat = "identity", aes(col = vaccine_equity_metric_quartile)) +
  theme_classic() +
  labs(x = "Date",
    y = "Percent fully vaccinated",
```



```

    title = "Rates of full vaccination across Vaccine Equity Metric Quartiles", caption = caption) +
  guides(color=guide_legend(title="Vaccine Equity Metric Quartiles")) +
  scale_color_continuous(labels =
    c("1 - Least Disadvantaged", "2", "3", "4 - Most Disadvantaged")) +
  scale_y_continuous(labels = scales::percent)

```



Source: California Department of Health, current to 2021-11-23

Results

Our analyses and data visualization shows that the highest vaccination rates were seen among the age group between 35-40. The general trend seen for this age group exhibited a full vaccination rate of approximately 70-73%, with rates seen above 75% as well.

Outliers were seen with higher rates of vaccination, approximately 85% for a median age of 44. Outliers were also found for lower rates of vaccination of approximately 25% for the median age of 37.

In terms of geography, higher vaccination rates were seen in counties such as San Francisco, Santa Clara, and Marin county with full vaccination rates at or above 80% and median ages of 38, 36, 44, respectively. Trailing behind are the counties of Alameda, Contra Costa, Imperial, and San Mateo with median ages of, 37, 38, 32, 39.

In terms of demographic indicators, it can be seen that rates of vaccination follow, from highest to lowest, the order of vulnerability derived from the Vaccine Equity Metric (VEM), from least to most – suggesting areas of improvement for equitable vaccine delivery. The aggregate of the least disadvantaged areas (VEM group 4) sees vaccination rates of 77%, whereas the most disadvantaged areas (VEM group 1) trails behind at 59%.

Discussion

According to our results, the minimum median age of persons receiving COVID-19 Vaccination in the state of California is 30, for Merced county. The maximum age was found to be 51 in Sierra county. The overall median for the state of California is 38. A wide range of ages was seen in the overall analysis.