Jaxon Abercrombie
PM-566: Introduction to Health Data Science

PM-566 Final Project Write-up

Investigating California's COVID-19 Vaccination Efforts by Region and County

Introduction (Background and Research Questions)

The ongoing COVID-19 pandemic has come under greater control with the roll-out of COVID-19 vaccines starting back in December 2020. California's governor, Gavin Newsom, speaks regularly about the success that our state has had with controlling the virus through immunization efforts. California is certainly diverse and varies in demographic composition and physical environment by county, so investigating how vaccine uptake has varied by county would be fascinating. Identifying which counties are behind on COVID-19 vaccination would be crucial for intervening and attempting to increase the percent of vaccinated individuals in a more targeted fashion. Furthermore, toward the beginning of the vaccine roll-out, discourse about which vaccine company people should receive their dose(s) from was frequent. Whether someone received Moderna, Pfizer, or Johnson & Johnson vaccines would dictate whether they received one or two doses and had greater immunity against variants. Because of this, investigating how California counties may differ by vaccine company distribution would also be exciting. Additionally, comparing vaccination rates alongside cases and deaths may reveal changes in vaccination attitudes based on the variation of COVID-19 morbidity and mortality. Lastly, especially because demographic characteristics heavily influence one's experience with COVID-19, analysis on how characteristics like age and race/ethnicity are related to vaccination will be carried out.

All that said, the primary question at hand is: *How have COVID-19 vaccination rates varied by county and/or region in California since their initial roll-out?* Furthermore, there are three secondary questions that dig deeper into the data sets used: (1) *How do trends in cases and deaths potentially affect immunization rates for California as a whole?*, (2) *How do vaccination efforts vary by vaccine company across the California regions (Pfizer, Moderna, Johnson & Johnson)?*, and (3) *How do vaccination efforts vary by age group and race/ethnicity by region?*

Methods (Data Sourcing, Wrangling, and Exploration)

Data for this project was found by searching through many California governmental databases for COVID-19 vaccination data to answer the primary and secondary questions at hand. Because the research questions were varied and required an abundance of variables, three different datasets were used. For a more detailed look at data for this project, check out this page of my website, which contains wrangling code and descriptions of the datasets used. Datasets were downloaded through R by URLs.

"Vaccines by California County" from the LA City Data website contains regularly updated variables including dose data for each vaccine company, population count per county, administered date of vaccine, and more. Its data spans from December 15, 2020 to early November 2021, which reflects how long vaccines have been available for. "Statewide COVID-19 Cases Deaths Tests" from the California Department of Health and Human Services is also regularly updated and contains data regarding COVID-19 cases, deaths, and tests that would help with the first secondary question. This set and the previously explained set were eventually merged for analysis and creation of visuals. In order to merge the two datasets, alterations to variable names and date ranges were made, as seen in the source code of the project. A region variable was created for the merged set based on the county variable, and percent variables were created by dividing cumulative vaccinated individuals by population.

The third dataset, "Vaccine demographic data," also from the LA City Data website, contains data about the vaccination of age groups and racial and ethnic groups across California and is updated regularly. This set was kept independent of the other two, since it contained sufficient information for its respective research questions. However, because the file size was so large, functions were created to extract a single day of information from the demographic datasets, create a region variable from county data, make a variable for percent individuals fully vaccinated was created, eliminate redundant rows for ease of plotting the data, and more (shown at this link).

Each dataset was explored by checking dimensions (especially those when merged to ensure correct merging), ensuring no NA values affected calculations and/or figures and tables, checking variable data types for consistency, and creating summary statistics that eventually became tables in the following Results section. Because the data came from governmental sources, they were reputable.

Results (Tables, Figures, and Analysis)

The following five pairs of figures and tables work to help answer the many questions from the introduction section of this write-up and are divided by horizontal lines. Interactive versions of the below figures are on the project website. Tables are in place to support the figures and provide a stationary visual for summary statistics that enables quantifiable differences for counties and/or regions.

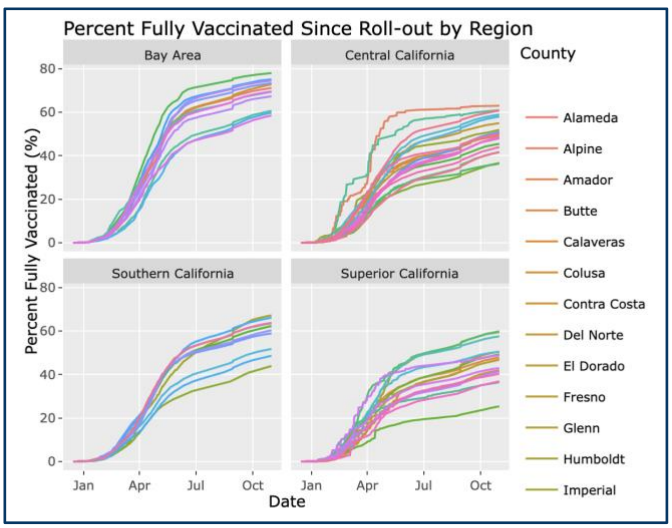

Figure 1: Vaccination Over Time by County and Region

The figure above depicts how vaccine efforts have been across four main regions of California, with country-specific data illustrated for each region's fully vaccinated population percentage. The data spans from the debut of the COVID-19 vaccine on December 15th, 2020 up until October 31st, 2021 (as with all following figures). Each region has a comparable number of counties, making these groupings representative and more reliable for drawing conclusions.

Table 1: Fully Vaccinated Population by County and Region

The above table supports Figure 1 and uses data from October 31st, 2021, the most recent day of full data for counties. For each region, descriptive statistics for the percent of fully vaccinated individuals are shown, with minimum, mean, maximum, and standard deviation values described at the county-level. For example, the the Bay Area minimum means that the county with the lowest percent of fully vaccinated individuals has a percent of 58.45%.

Region	Min %	Mean %	Max %	SD %
Bay Area	58.45	69.16	77.99	6.43
Central California	36.37	50.08	62.98	7.76
Southern California	43.93	58.64	67.21	7.90
Superior California	25.40	46.01	59.86	9.12

Based on Figure 1 and Table 1, each region experienced similar uptake in vaccination, as seen with relatively congruent plot shapes (Figure 1). It is evident that April 2021's expansion of vaccine eligibility influenced a rapid increase in vaccination, yet an eventual plateau came about, indicating that people who wanted to be vaccinated likely were. Certainly, there are regional differences, like the Bay Area having a consistently higher percent of fully vaccinated individuals and Superior California the least. In particular, the difference between the least and most fully vaccinated county in the Bay Area is less than 20%, while the same difference in Superior California is almost 35% (Table 1). Furthermore, the variability in vaccination among more rural regions like Central and Superior California is evident with lines deviating from each other more than the other two "urban" regions. Overall, it appears that urban areas (Bay Area and Southern California) have more vaccine uptake than rural areas (Central and Superior California) on average, which will be something to consider with subsequent figures on this website. Access comes to mind as an immediate hurdle for vaccination, as physical distance from urban centers may make rural living disadvantageous.

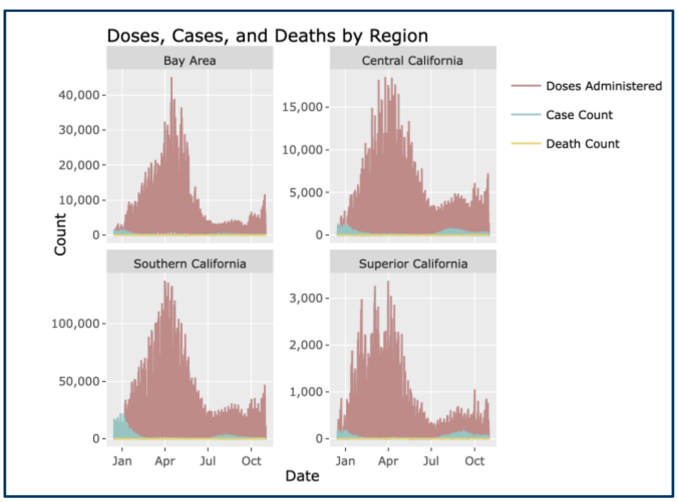

Figure 2: Doses, Cases, and Deaths by Region

The figure illustrates a potential relationship between vaccine uptake and cases and deaths. The y-axis is simply a count of cases, deaths, and doses administered in order to overlay the information. Ultimately, the purpose of the figure is to gauge potential changes in vaccination behaviors as a result of rising case and death counts, so numerical values are less important here.

Table 2: Doses, Cases, and Deaths by Region

The table supports Figure 2 uses data from December 15, 2020 to October 31, 2021. For each region, descriptive statistics for cases, deaths, and doses in order to depict daily records of cases and deaths, as well as show the vaccination capacity of each region through daily doses administered. This table shows data that spans almost an entire year of data, and it is important to view the table with Figure 2 in mind since the plot provides a lot of context to the ebb and flow of cases, deaths, and doses.

Region	Min Cases	Mean Cases	Max Cases	SD Cases	Min Deaths	Mean Deaths	Max Deaths	SD Deaths	Min Doses	Mean Doses	Max Doses	SD Doses
Bay Area	0	106.80	1752	174.45	0	1.34	32	2.92	0	3347.72	45081	4913.34
Central California	0	73.66	1369	141.74	0	1.14	23	2.49	0	1068.37	18514	2001.81
Southern California	1	589.08	22284	1667.61	0	10.83	304	31.54	0	9527.48	137449	16680.92
Superior California	0	14.48	192	23.39	0	0.21	7	0.60	0	207.27	3374	356.57

Based on Figure 2 and Table 2, the expansion of vaccine eligibility clearly created a surge in vaccination around April 2021. Around July of 2021, a rise in cases and deaths appear to be associated with an increase in vaccine uptake, likely because of the novel Delta variant. This period of increased vaccination happened with all regions except the Bay Area, which could be because of its already high vaccination rates. This relationship may also have been influenced by policy efforts to require vaccines in public spaces, eligibility for booster shots, and the realization of the effectiveness of vaccination. In the unfortunate situation that cases and deaths increase wildly in the upcoming winter, the vaccination capacity of each region are indicated in Table 2. For example, Southern California has a county with the capability of administering 127,449 doses in the case of emergency, while Superior has 3,374 (Table 2). Of course, these are relative to population size.

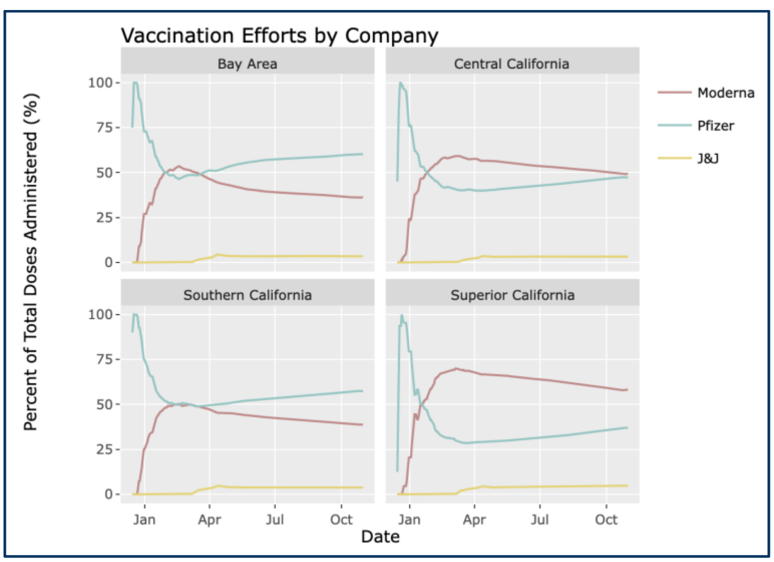

Figure 3: Share of Market by Company for Each region

This figure depicts the share of vaccine prevalence among the three main COVID-19 vaccine companies over the course of vaccine roll-out to October 31st. Share is defined as the percent a vaccine takes up in the market at a given time. Understanding how vaccination efforts differ by company may influence future roll-outs of vaccines, since timeliness of roll-out and targeting specific areas may make one company a preferable candidate compared to another.

Table 3: Vaccines Administered by Company for Each Region

This table uses data from December 15, 2020 to October 31, 2021. Data about dose counts for each company (Johnson & Johnson, Moderna, and Pfizer) rather than percent of the market like Figure 3 does is used, which helps to provide more context and numerical evidence.

Region	Min J&J	Mean J&J	Max J&J	SD J&J	Cum. J&J	Min Mod.	Mean Mod.	Max Mod.	SD Mod.	Cum. Mod.	Min Pfi.	Mean Pfi.	Max Pfi.	SD Pfi.	Cum. Pfi.
Bay Area	0	122.62	8037	472.84	472336.88	0	1096.26	15964	1707.77	4222801.5	0	2128.83	34818	3226.75	8200252.3
Central California	0	33.67	3004	123.99	216157.53	0	424.25	9727	892.37	2723702.9	0	610.44	10514	1120.72	3919017.7
Southern California	0	338.32	24711	1282.71	1085994.22	0	3668.22	62780	7427.10	11774981.4	0	5520.88	71114	8884.17	17722030.3
Superior California	0	8.83	861	29.72	45336.36	0	100.11	2664	203.51	514162.2	0	98.33	2729	197.24	505041.6

Based on Figure 3 and Table 3, there are clear differences in how vaccine companies share the COVID-19 vaccine market. When a vaccine came to market and whether it takes one or two doses to be fully vaccinated would certainly impact this data. Pfizer has dominance in more urban regions, and, despite Pfizer debuting on the COVID-19 vaccine market first, Moderna still dominates in rural areas (Figure 3). Johnson & Johnson vaccines, even if its single dose nature were accounted for, still share a limited amount of the vaccine efforts. In each region, Moderna became the dominant vaccine at similar points in time around February to March of 2021 and Pfizer regained dominance in urban regions thereafter (Figure 3). While Figure 3 shows percentages, Table 3 provides more context to numerical data for each vaccine, which demonstrates the magnitude of a vaccine's impact in a given area.

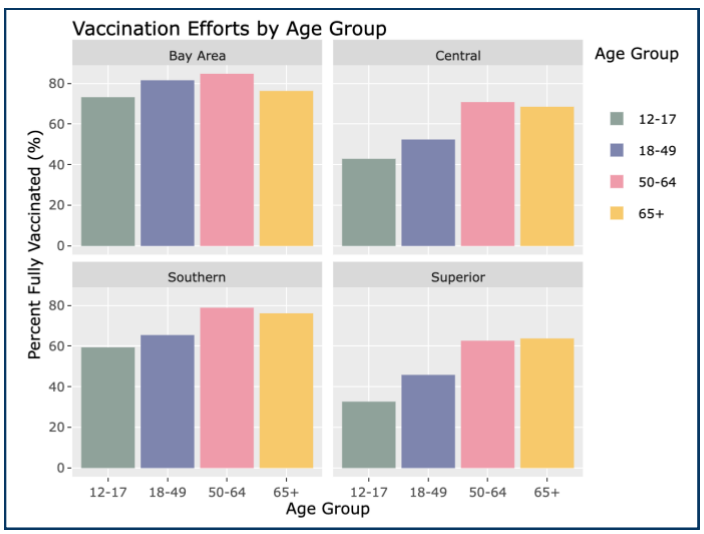

Figure 4: Fully Vaccinated Population for Regions, by Age Group

This figure depicts the current percentages (by October 31, 2021) of fully vaccinated individuals by age group for each California region. Using percentages makes comparisons between regions more tangible, since it is a standardized approach.

Table 4: Fully Vaccinated Population for Regions, by Age Group

This table supports Figure 4 and uses data from October 31, 2021. Data about each region, stratified by age group, is shown. Similar to Table 1, data is county-level, meaning the minimum for the Bay Area for the 12-17 group represents the county with the smallest vaccination percentage by October 31. Percentages like those in this table make data more comparable between region and demographic group, since they are standardized to be out of 100.

Region	Demographic	Min %	Mean %	Max %	SD %
Bay Area	12-17	55.42	73.21	93.63	9.33
Bay Area	18-49	63.95	81.54	96.10	10.45
Bay Area	50-64	76.79	84.75	90.60	4.72
Bay Area	65+	69.52	76.30	81.36	3.51
Central	12-17	22.12	42.89	60.37	9.91
Central	18-49	37.63	52.40	74.42	10.63
Central	50-64	47.01	70.80	96.02	11.74
Central	65+	44.17	68.51	81.84	9.00
Southern	12-17	41.92	59.38	96.34	14.72
Southern	18-49	48.45	65.44	81.14	10.81
Southern	50-64	64.31	78.88	90.20	8.10
Southern	65+	69.27	76.14	86.49	5.58
Superior	12-17	16.61	32.69	54.15	10.29
Superior	18-49	17.27	45.85	66.87	13.28
Superior	50-64	37.87	62.66	81.37	11.11
Superior	65+	44.90	63.76	78.20	8.32

Based on Figure 4 and Table 4, differences between age groups within regions are relatively small for urban regions and larger for rural regions. Similar to Figure 1, the Bay Area has the highest percentage of fully vaccinated individuals and Superior California the lowest. In almost every case, the 50-64 age group is vaccinated the most and the 12-17 group the least (Figure 4). These differences in vaccination percentages may be attributable to differences in eligibility phases; the elderly could be vaccinated first, with approval for youth vaccination only happening in more recent months. The eligibility differences may account for wider ranges between the highest and lowest groups within regions. For example, the mean percentage gap between the 12-17 and 65+ age group in Superior California is far larger than that of the Bay Area (~15 % vs. ~29%) (Table 4). More equivalent percentages across age groups within a region may demonstrate a steady availability and access to vaccines, since those who want the vaccine likely have received it already. Additionally, it is important to acknowledge how the age groups vary in size; 18-49 represents a far larger group than 12-17, for example. Regardless, even larger groups in number have the capability of outperforming others in percent.

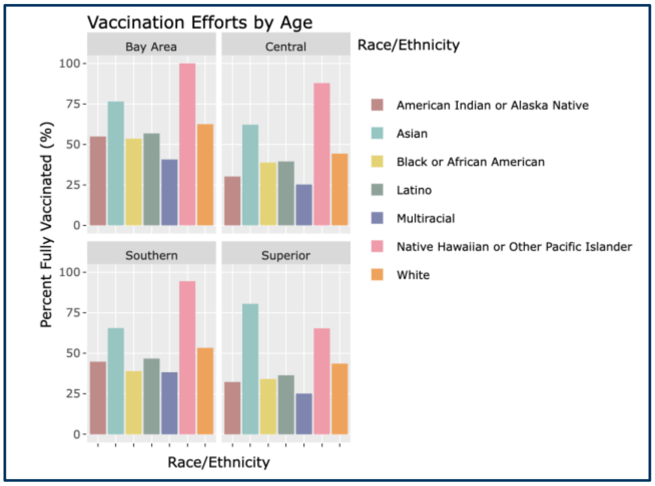

Figure 5: Fully Vaccinated Population for Regions, by Race/Ethnicity

This figure depicts the current percentages (by October 31, 2021) of fully vaccinated individuals by race and ethnicity for each California region. Using percentages makes comparisons between regions more tangible, since it is standardized.

Table 5: Fully Vaccinated Population for Regions, by Race/Ethnicity

The above table supports Figure 5 and uses data from October 31, 2021. Data about each region, stratified by race/ethnicity, is shown. Similar to Table 1 and 4, data is county-level, meaning the minimum for the Bay Area for the Asian group represents the county with the smallest vaccination percentage for that group by October 31.

Region	Demographic	Min %	Mean %	Max %	SD %
Bay Area	American Indian or Alaska Native	41.64	54.87	89.39	14.15
Bay Area	Asian	57.35	76.47	88.63	10.25
Bay Area	Black or African American	34.52	53.55	64.37	8.40
Bay Area	Latino	42.12	56.78	76.87	9.52
Bay Area	Multiracial	29.18	40.69	49.46	6.51
Bay Area	Native Hawaiian or Other Pacific Islander	50.60	74.44	87.67	14.31
Bay Area	White	46.70	62.47	75.68	7.02
Central	American Indian or Alaska Native	6.10	30.18	50.31	12.47
Central	Asian	38.48	62.13	86.54	12.87
Central	Black or African American	11.46	38.76	77.60	18.88
Central	Latino	26.82	39.47	60.71	6.97
Central	Multiracial	1.50	25.26	54.36	17.97
Central	Native Hawaiian or Other Pacific Islander	49.51	69.93	96.97	12.68
Central	White	19.72	44.29	60.41	10.32
Southern	American Indian or Alaska Native	23.68	44.82	71.44	13.12
Southern	Asian	38.25	65.57	79.60	11.93
Southern	Black or African American	12.83	39.00	56.95	13.25
Southern	Latino	38.18	46.75	62.25	7.96
Southern	Multiracial	1.03	38.34	54.57	14.92
Southern	Native Hawaiian or Other Pacific Islander	62.75	79.44	97.38	11.79
Southern	White	40.89	53.38	63.21	8.06
Superior	American Indian or Alaska Native	11.37	32.31	47.00	10.26
Superior	Asian	35.25	57.65	74.96	11.51
Superior	Black or African American	4.13	34.17	51.82	14.20
Superior	Latino	14.18	36.44	52.57	11.13

Based on Figure 5 and Table 5, Native Hawaiian and Asian populations significantly lead each vaccination percentage in each region, while multiracial, Latinx, and Black populations trail behind most other groups (Figure 5). Unlike age groups from Figure 4, the differences in rates between race and ethnicity are not attributable to expanding vaccination eligibility, since biological characteristics like age and conditions guided that expansion. Therefore, any discrepancy in vaccination rates are likely attributable to historical contexts, mistrust, or access because of structural inequity that non-white groups face most often. Regardless of region, there are visible gaps between racial and ethnic groups. The high percentages may be attributable to smaller population size and percentage us therefore easier to increase. Figure 5 allows for comparisons between regions, making conclusions like Native American populations faring better in the Bay Area compared to Superior California by about 22%, for example.

.

Conclusion/Summary (Overall Findings)

Overall findings for the primary question and multiple secondary questions of this project were deduced from the many figures and tables from the above section. There are clear needs for changes when looking at disparities across regions, like the need for rural areas to have greater assistance with administering vaccines and racial and ethnic disparities to be reduced through targeted efforts.

In regard to vaccination efforts by region overall, each region experiences a similar “S” curve, which demonstrates a natural surge in vaccination (after eligibility increased) followed by a plateau. However, looking at regions individually demonstrates a stark contrast in vaccination efforts when comparing more urban to more rural regions, with the former vaccinating more and latter not.

When looking at how cases and deaths may affect vaccine uptake, there appears to be an association in which a rise in cases and deaths is matched with a rise in administering vaccines. This could be attributable to imminent threats pressuring individuals to get vaccinated—or even simply be a factor of public policy enforcing vaccination.

Vaccine companies significantly differ in their share of a market by region, with Pfizer leading in more urban regions and Moderna leading in more rural regions. Furthermore, the Johnson & Johnson vaccines lack a substantial share of vaccines administered, even accounting for it being a single dose vaccine. It is clear that temporality in regard to vaccine release to market plays a role and that urban areas are easiest to reach through supply chain efforts.

By region, age groups slightly differ in percentage of fully vaccinated individuals. Overall, there are relatively small discrepancies between age groups within regions. Most often, older age groups are more likely to be vaccinated than younger age groups, which may be attributable to eligibility criteria that has expanded over time based on age.

Racial and ethnic groups have clear differences in percentage of fully vaccinated individuals. Compared to age, there are more obvious differences in racial and ethnic groups within regions, as observed with larger gaps in percentage of fully vaccinated individuals. Though these groups may vary by population count for a given region, Native Hawaiian and Asian populations are vaccinated most often and multiracial, Latinx, and Black populations the least.

Looking forward, a future version of this project with information about booster shots would be incredibly exciting, since uptake of booster shots may be vastly different than the initial doses of the COVID-19 vaccine. Now equipped with website and visualization skills, I feel eager to take on a new project!