Data Wrangling: COVID-19's Effect on the Population, Demographics, and the Economy by Counties in California

by

Jess Herbert
Jacob Mcgraw
Monica Orme
Diana Valderrama

Math 167PS

May 12, 2021

Abstract
In this data wrangling project, we have focused on visually summarizing county-level outcomes related to COVID-19 with respect to morbidity and mortality, unemployment rates and the labor force, and the distribution of vaccinations. To achieve a representative overview, we selected a county from each of four California regions. We chose Sacramento, Los Angeles, Colusa, and Santa Clara counties to represent the Central, Southern, Superior, and Bay Area regions, respectively, based on county size and availability of data. For each county, we present the distributions of monthly COVID-19 cases and deaths and the size of the labor force, compare the number of deaths by race and vaccination administration by race and age[1], and examine unemployment rates as a function of COVID-19 hospitalizations, both pre- and post-vaccine introduction. In this report, we explain the sets of data that were utilized to summarize COVID-19-related outcomes and present our findings.

Introduction
As most of our data covered a 15-month timespan, we decided to work with months as observational units for clearer analyses. However, as vaccination data spans only 5 months, visualizing this on a daily basis was feasible. We made use of several data sets in this project, including *CA_Unemployment*, *CA_covid_cases.deaths.tests*, *COVID-19 Vaccines Administered By Demographics*, *Local_Area_Unemployment_Statistics*, and *SC_County_COVID-19_cases_by_race_ethnicity_over_time*.

CA_Unemployment[I] originally included 11 columns denoting the Area Type (State, City, etc.) and Area Name (California, Colusa County, etc) for each observation; the date (mm/dd/yyyy), year, and month on which it was recorded; whether or not it is seasonally adjusted and its status as a preliminary or final finding; the number of people constituting the labor force on each day and how many of those were employed and unemployed people; and the unemployment rate. We also made use of *Local_Area_Unemployment_Statistics*[IV], which is a subset of this data set.

CA_covid_cases.deaths.tests[II] originally included 11 columns denoting the date (mm/dd/yyyy) of each observation; the area name, type, and population; and 7 COVID-19 related variables to report the number of cases, deaths, total tests, positive tests, reported cases, reported deaths, and reported tests.

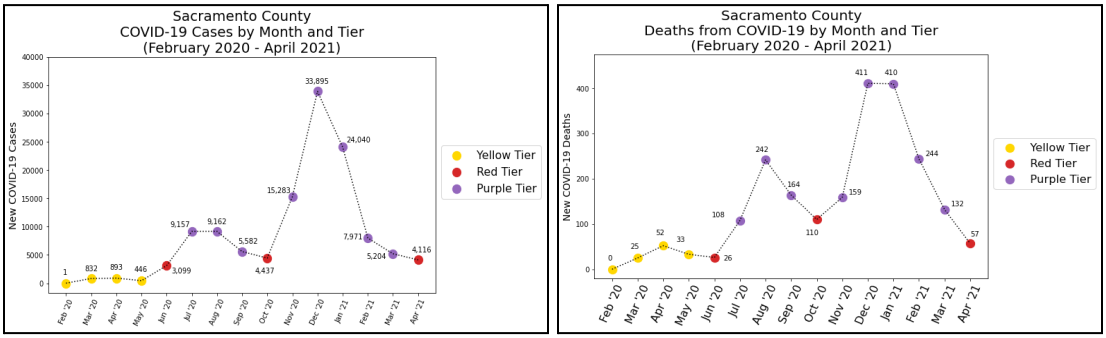
COVID-19 Vaccines Administered By Demographics[III] originally contained columns including the following information: county name; demographic categories such as age or race/ethnicity;

[1] Data were sparse for gender-related outcomes. Thus, gender was excluded from demographics and disparities analyses. Data including demographic categories for county-level economic outcomes was similarly unavailable. Hence, our project shifted from our original idea as we moved through the analysis based on the availability of good data.

demographic values such as age groups or races/ethnicities; an estimated population of each county along with estimations of different demographic groups; day of vaccination; the number of people who are partially and fully vaccinated measured daily and cumulatively; and the number of demographic groups that are not vaccinated.

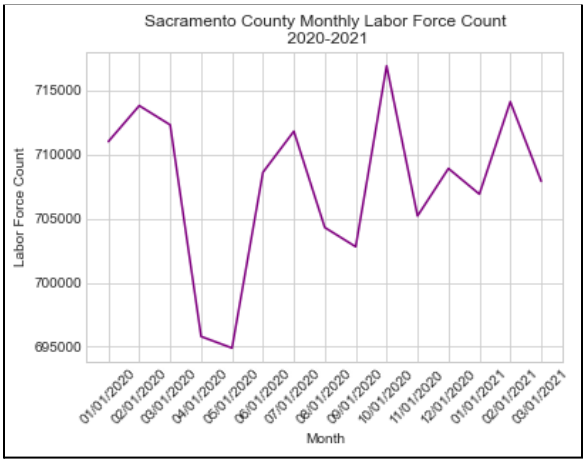
SC_County_COVID-19_cases_by_race_ethnicity_over_time[V] contained the following 4 columns: race/ethnicity, new COVID-19 cases on a weekly basis, the beginning date of that week, and the end date of that week.

(i): Sacramento County



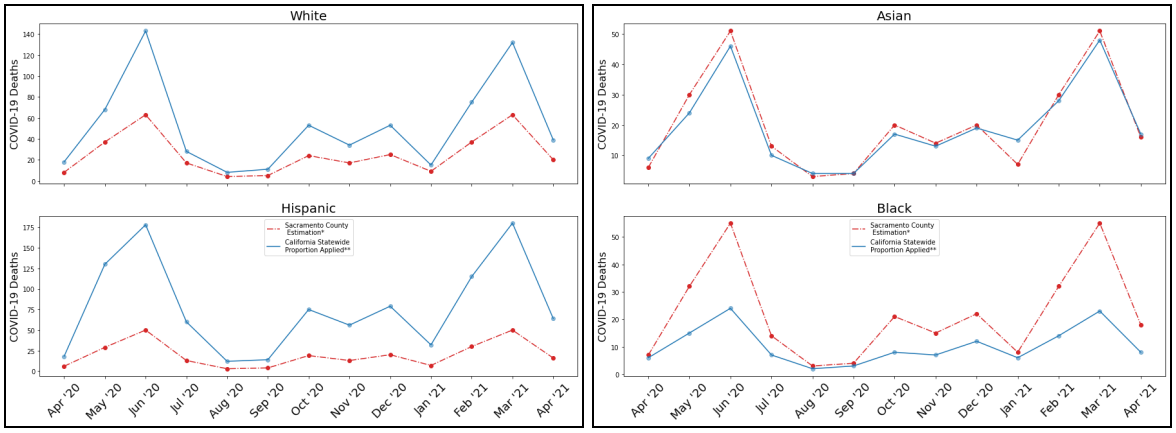
Prior to analyzing COVID-19-related outcomes with respect to race, we visualize the general morbidity and mortality experienced by Sacramento County overall, by assigning each month a risk level based on monthly new cases. To prepare the data for this analysis, we subset the data to include only counties and remove missing data. We clean the data using results from the data dictionary to eliminate contradictions between the 'deaths' and 'reported deaths' columns, for example. We use string formatting to replace the column containing dates as 'mm/dd/yyyy' by two columns containing 'Month' and 'Year' random variables. As we have chosen to use months as our observational units, and the risk level tiers defined by California's *Blueprint to a Safer Economy*[VI] are subject to change on a semi-monthly basis, we scale the quantitative indicators to monthly metrics to summarize a month's outcomes. To do this, we isolate the population column before grouping our data by month, then use the total number of monthly cases to calculate a random variable that denotes the number of monthly cases per 100,000 residents. We then scale the risk level tiers to be monthly benchmarks by multiplying the number of daily cases required for a tier by the number of days in a month. For example, the yellow tier is assigned to February 2020 because it saw less than 58 new cases through the month, or less than 2 daily new cases, per 100,000 residents. In Sacramento County, the number of COVID-19 cases in a month appears not to impact the number of deaths in that month but is greatly predictive of the number of deaths that will occur in the following month(s). During June 2020, for instance, cases

increased by approximately 700% from May, while deaths fell by 7%; July 2020 saw a 300% increase in cases, while deaths increased by 415%; and August 2020 saw cases flatten while deaths increased by 225%. A similar relationship between cases and deaths appears from November 2020 through January 2021, though in a decreasing fashion direction.

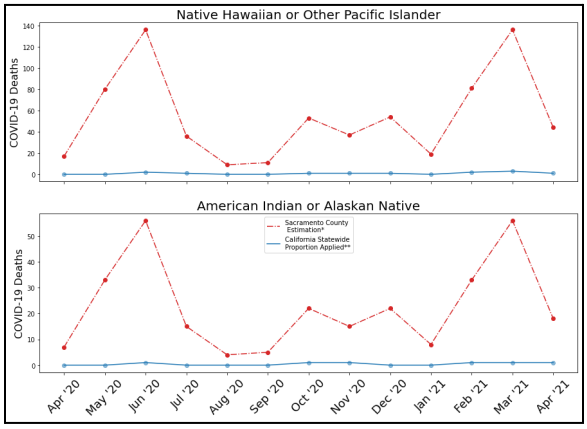


The data in this graph comes from the Labor Force and Unemployment data by county, which was provided by the California government Labor Market website. In order to display the trend of the labor force counts in Sacramento county on a monthly basis, the data needed to be cleaned. Prior to plotting the data, the data frame was subsetted to only the years 2020 and 2021, the data frame's columns were renamed, the dates were sorted first by year then by month, and string formatting was applied to remove commas and convert labor force counts to floats.

The plot of the monthly Labor Force counts in Sacramento county shows that the labor force count plummeted in March 2020, falling to about 696k participants from the 720k seen in February 2020. April 2020 saw the lowest count in the analysis year, with only 695k active participants in Sacramento county. However, the labor force then spiked in May 2020 and fluctuated by $\pm$10k through March 2021, which was the end date of the most recently available data. Comparing counts of the Sacramento County labor force to the county COVID-19 case and death counts, it is clear that COVID-19 had some impact on the size of the county's labor force. However, as compared to the other counties analyzed in this project, it appears Sacramento county's labor force was less impacted by the pandemic than other counties.



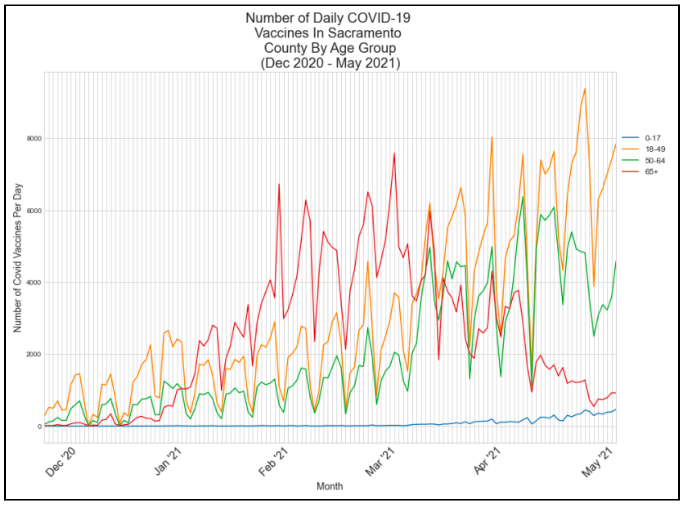
To analyze the Sacramento County COVID-19 mortality for different racial groups, we compare the estimated number of monthly deaths in Sacramento County[2] for a racial group to the monthly deaths per the California state-wide proportion[3] for the same time period. For all months included in the analysis, Sacramento County appears to have observed fewer deaths than California did for White and Hispanic residents. Asian county residents saw the most fluctuation, with some periods with county outcomes better than the state, and vice-versa. County residents identifying as Black, Native Hawaiian or Other Pacific Islander (NHPI) or American Indian or Alaskan Native (AIAN) appear to have faced much higher mortality than they did across the state, and to a nearly identical degree.



[2] Data were available for the total number of deaths in a county by month, and for the proportion of total deaths by race cumulatively. Thus, the *County Estimation* is generated by applying the proportion of total deaths represented by a racial group to the total number of deaths in a month, creating an estimated number of monthly deaths for each racial group.

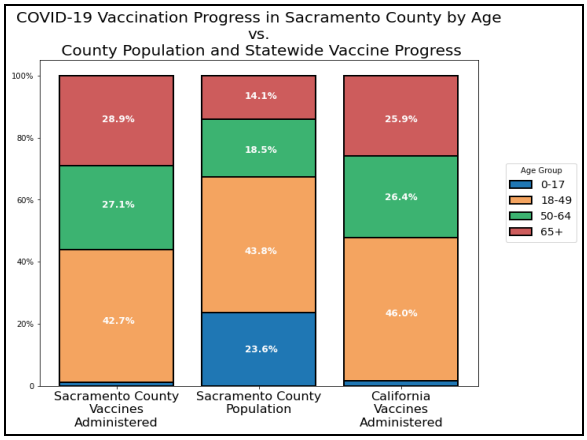
[3] Data were available for the statewide total number of monthly deaths by race, but no county-level observations. Thus, the *California Statewide Proportion Applied* data series is generated by taking the monthly proportion of statewide deaths represented by a racial group and applying it to the total number of deaths in a county for that month, creating an estimated number of monthly deaths for a racial group which a county should have seen if they saw outcomes equitable to statewide outcomes. This data was only available from April 2020 on, so we removed February and March 2020 from this portion of the analysis.

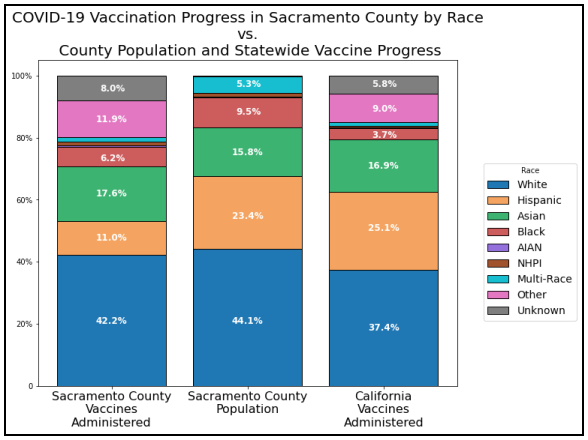
In order to look at daily vaccination data within our counties, we first had to subset the COVID-19 Vaccines Administered By Demographics data set to include the date, counties, and counts for daily partial and full vaccinations. A new column was created with a count of daily vaccinations, called "group_vaxed_day_total," which included both first and second dose counts per day, per county. After separating the data into each county, we were able to plot the daily vaccination counts for each age group. Since the COVID-19 vaccine was introduced after the beginning of the pandemic, this data is limited from December 2020 to May 2021.



For most of December and January, those aged 18-64 had the highest daily vaccination counts, presumably because those working in the healthcare system are likely to be in this age range. However, the age 65+ age group had many more daily vaccinations from early February - early April 2021. This is because, in California, those aged 65+ were eligible to be vaccinated before the rest of the general population. At the beginning of April 2021, vaccinations became available to those aged 16+, so we see that as senior vaccination counts decreased, both age groups 18-49 and 50-64 saw their vaccinations increase. Because, until recently, only those aged 16+ were eligible to receive the vaccine, the 0-17 age group has largely been ineligible, and thus vaccination counts for this age group were consistently very low. Interestingly, there are repeated peaks 7 days apart, indicating that one day of the week is the most popular for getting vaccinated. In Sacramento County, this day was Monday.

After looking at daily changes over time in vaccinations, we compare cumulative county vaccination rates to those for the entire state of California, and to the county's population demographics.

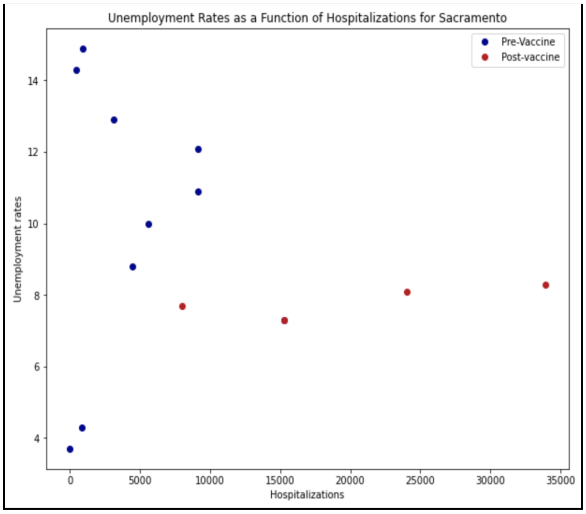


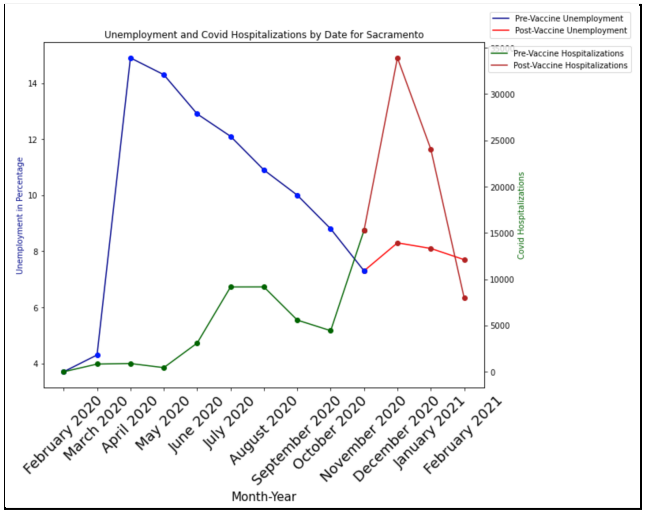


Cumulative vaccination rates by race and age were available on county and state COVID-19 dashboards[VII], so we first created a data frame of these values. We then import county census data with respect to race and age[VIII] and subset and group the observations to match the groups used in the cumulative vaccine data. Visualizing these outcomes in the stacked bar graph above, we see that Sacramento County's vaccines have been administered to older residents at a higher rate than the state's outcomes indicate and are represented by their population. This reveals that the county has outperformed the state with respect to vaccination of vulnerable age groups. However, with respect to racial groups, Sacramento County has underperformed. The rate at which vaccines are administered to Hispanic residents is less than half of the state's outcomes, and less than half of the Hispanic proportion of the county population. Black county residents have also been administered vaccines at a rate lower than their proportion of the population, but the county's outcome is significantly better than the state's, here.

In order to analyze hospitalization and unemployment rates, the data must be retrieved from the CSV file and manipulated in a number of ways so that it can be aggregated and analyzed. Both datasets track their observations by day, so our group needs to sum the values over their respective months and put that information into its respective data frame. Then, the data were subset to the county level. Subsequently, the data from the unemployment data frame was limited to the observations from 2020 to 2021. The observations were limited to February 2020 to February 2021 due to limitations of the data available in both datasets. Afterward, both datasets were subsetted to the months between February 2020 to February 2021 and given a column with a year and a month concatenated to serve as an indicator column. Furthermore, the data from the unemployment data frame was limited to the non-seasonally adjusted unemployment data to keep the estimations as accurate as possible. At this point, the datasets have 13 observations each, with each observation being indicated by the column that contains the month and the year in the specified time frame. The useful columns are the unemployment rates and the month/year indicator in the unemployment data frame, and the month/year indicator and the hospitalization rates in the COVID dataset. The data was plotted in the following ways: as a scatterplot with

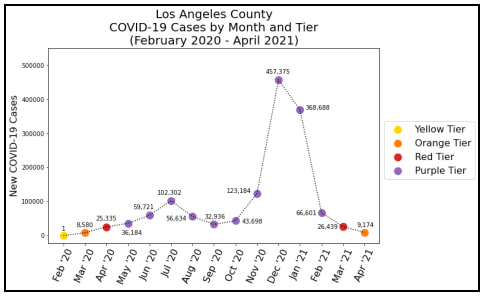
unemployment rates as a function of hospitalizations, and as a line graph of both unemployment rates and hospitalizations over time in terms of months.

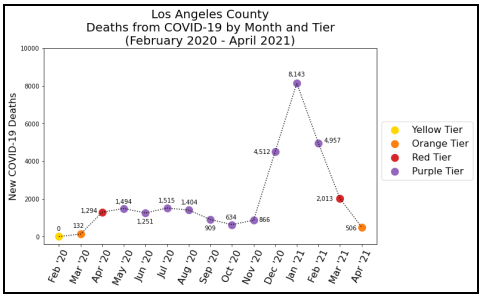




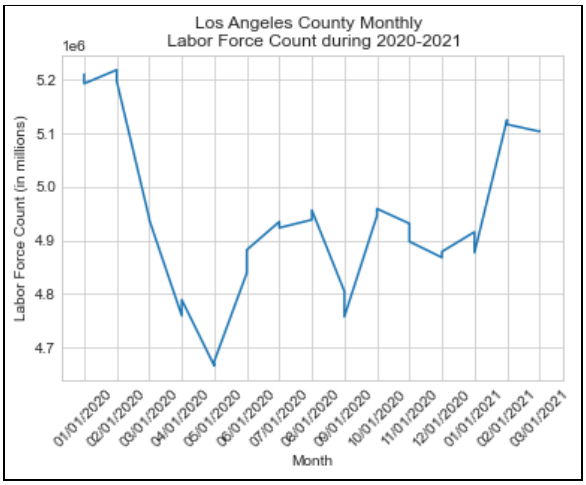
The scatterplot (above left) shows several interesting results. For one, the pre-vaccine hospitalizations seem to have a large amount of variance in terms of their unemployment rates, but seem to be relatively similar in terms of their hospitalizations, generally ranging from 0 to 10,000. For post-vaccine unemployment, the observations are generally uniform around the 7-8 percent unemployed mark and have a large amount of variance with respect to the number of hospitalizations. The line graph for Sacramento (above right) seems to explain some of the behavior displayed in the scatterplot. For one, although Sacramento's unemployment rate was relatively stable for the starting month of this analysis, in March 2020 the unemployment rate spiked dramatically and has been steadily returning to a more healthy unemployment rate as the economy gets closer to fully re-opening. The vaccines being administered seem to correlate in a slight initial boost in unemployment; however, the hospitalizations may explain this discrepancy. For one, before the vaccines were being administered, there was already a notable spike in COVID activity from September 2020 to November 2020, which only got more severe in the following month. This spike in COVID activity may be more explanatory than vaccines being administered, but more testing would be needed to draw any conclusions or correlations from these graphs. The hospitalizations for Sacramento also display an interesting pattern, they appear to be relatively not affected by COVID for several months. There was a minor spike in the following months, but the most significant spike in terms of hospitalizations was from September 2020 to November 2020.

(ii): Los Angeles County





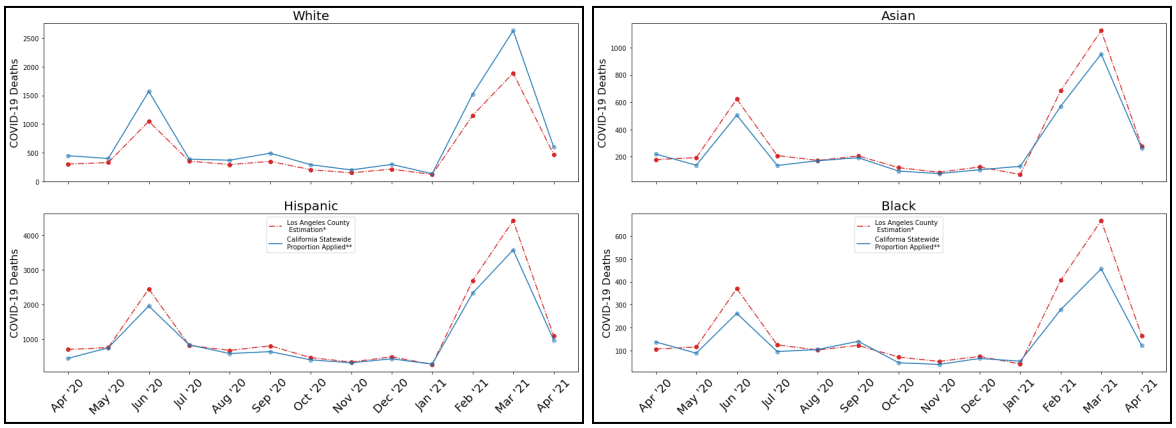
Analyzing the distribution of monthly morbidity and mortality in Los Angeles County, we see the pattern which occurred in Sacramento County is repeated - the number of new cases in a month, though not necessarily causative of the number of new deaths in that month, is greatly predictive of mortality in the coming month(s). For example, the number of cases in November 2020 increased by 180% from the previous month, while deaths only increased by 37%, but December 2020 then saw a 420% increase in deaths, while cases increased by 270%.



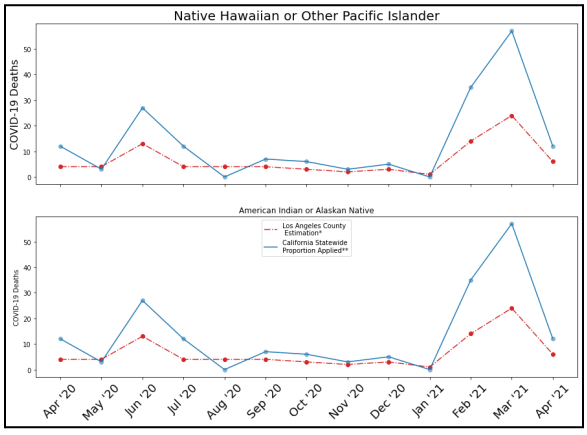
The data in this graph comes from the Labor Force and Unemployment data by county, which was provided by the California government Labor Market website. In order to display the trend of the labor force counts in Los Angeles County on a monthly basis, the data needed to be cleaned. Prior to plotting the data, the data frame was subsetted by only the years 2020 and 2021, the data frame's columns were renamed, the dates were sorted based on year first then the month, and the column for labor force counts had commas removed and were converted to float.

The plot of the monthly Labor Force counts in Los Angeles county shows that the labor force count plummeted in March 2020 from over 5.2 million people in the labor force in February to less than 4.8 million people in the labor force, which is approximately an 8% decrease in the labor force. The lowest peak in the labor count within Santa Clara county was about 4.65 million in May 2020. However, after that point, it can be seen that the labor force count spiked up slightly in June 2020 and fluctuated from then up until March 2021, which was the end date of

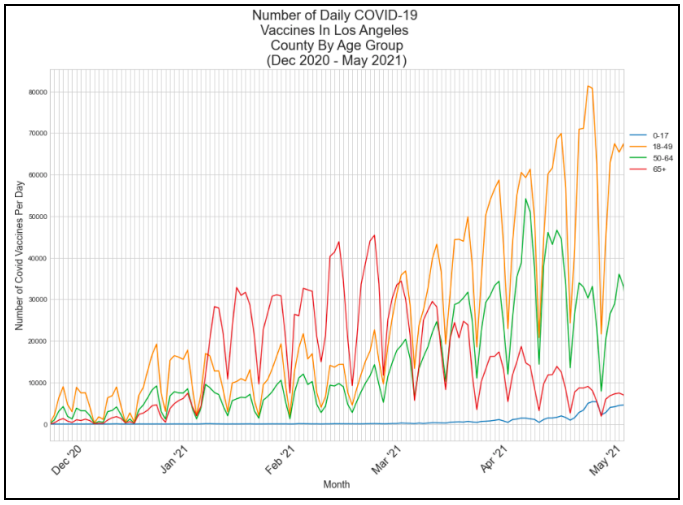
the most recently available data. Also, the labor force count never went up to the amount that it previously was prior to February 2020.



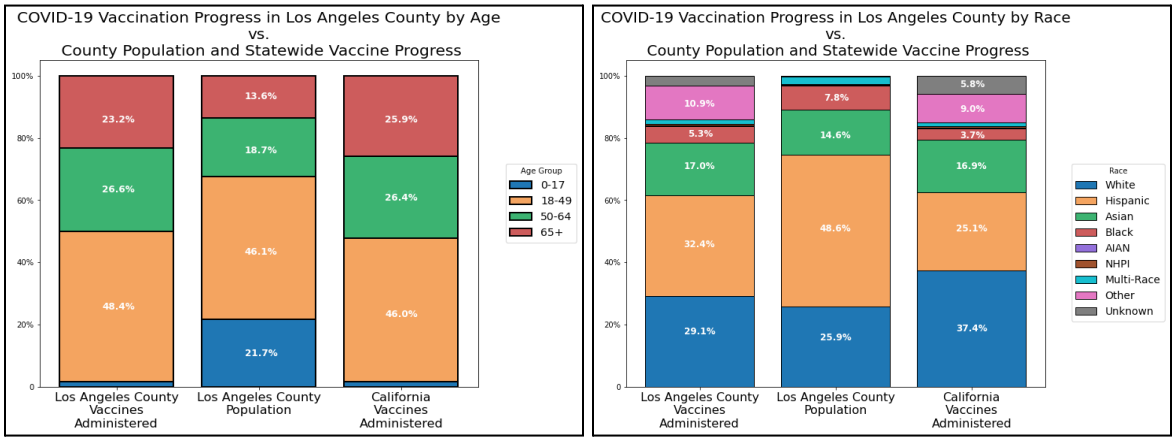
Analyzing the Los Angeles County COVID-19 mortality for different racial groups, we see Hispanic, Black, and Asian county residents saw similar outcomes, with mortality slightly higher through the analysis year, save peaks in June 2020 and March 2021, where county deaths were much higher than those in the state. The only racial group which saw consistently better county than state outcomes was White people, a disparity that was amplified during the two aforementioned peaks. NHPI and AIAN county residents saw much better outcomes than in Sacramento County, and during the peaks saw the smallest impact of all racial groups.



From our previously subsetted vaccination count data frame, we were able to focus on Los Angeles County.

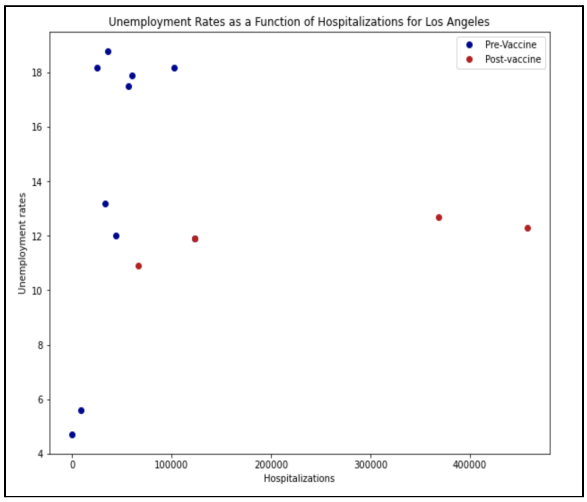


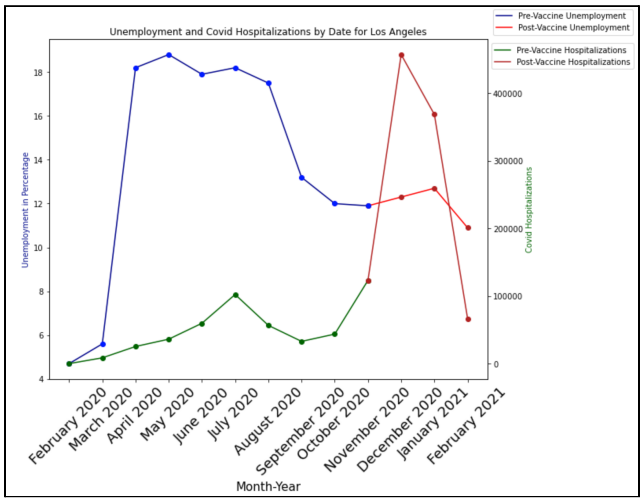
From the beginning of the vaccination period until the beginning of February, the 18-49 age group consistently had the highest daily vaccination counts, followed by the 50-64 age group. This is probably because those working in the healthcare system are usually between the ages of 18 and 64. From February to April, the 65+ age group had led the counts for daily vaccinations, and these counts consistently declined until the end of the data set. Again, this is due to the eligibility for 65+ beginning in February and opening to 50+ and eventually 16+ in April. Both age groups 18-49 and 50-64 increased their vaccination numbers as eligibility grew. Again, because the FDA had not approved vaccinations for those younger than 16 until recently, there has been a generally low number of vaccinations for those 0-17. However, there is an increase after eligibility opened up to 16+ in mid-April. For Los Angeles County, Tuesday is the most popular day of the week for getting vaccinated.



Visualizing cumulative county vaccination rates against those for the state, and against the county's population demographics, we see that Los Angeles County's vaccines have been administered to older residents at a higher rate than their respective population proportions, but the distribution skews younger than the state's. The 18-49 age group sees Los Angeles 2.5%

higher, 50-64 is 0.2% higher, and 65+ is 2.7% lower. This reveals that the county may not have done as well in reaching vulnerable age groups or vaccination. However, with respect to racial groups, Los Angeles County has performed well. Though the vaccination rates are lower than county population proportions in some cases, this is likely the result of 10.9% of county vaccinations being administered to 'Other' races. Yet, the county outperforms the state in all minority populations.

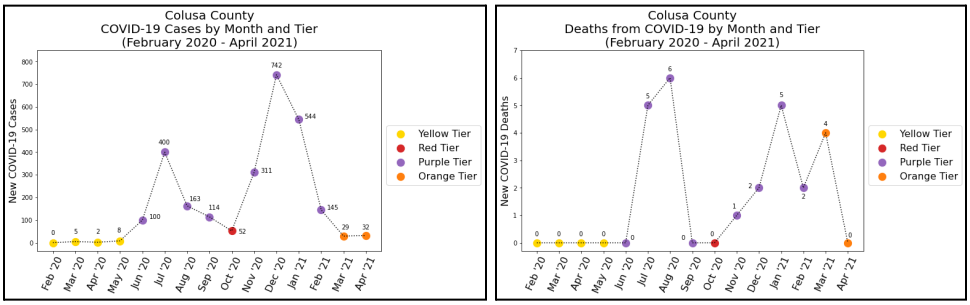




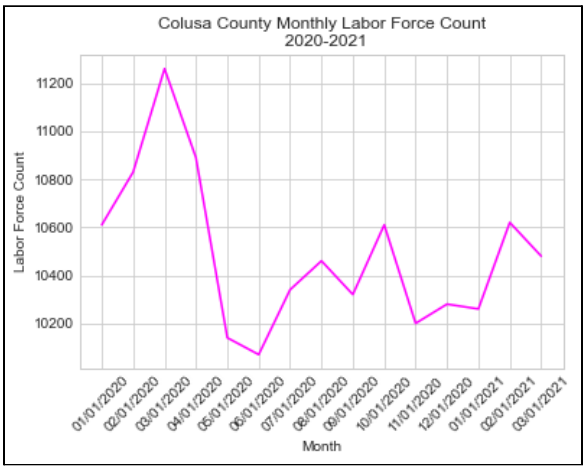
Los Angeles County displays several interesting behaviors with its scatterplot of hospitalizations versus unemployment rates. Like Sacramento, the pre- and post-vaccine observations have a particular pattern: the pre-vaccine observations tend to stay below the 10,000 mark in terms of hospitalizations and have far more variability in terms of their unemployment. In contrast, the post-vaccine observations tend to be in the 10 to 12 percent range and have a large variation in terms of their hospitalizations. However, the unemployment for the pre-vaccine seems to clump around the 18 percent unemployment rate, while Sacramento had much more variability in its pre-vaccine observations.

The line graph confirms several of the observations made in the scatterplot. In contrast to Sacramento, Los Angeles was not able to recover nearly as effectively. While Los Angeles initially performed well in February 2020, the county experienced a severe spike in unemployment and struggled to recover significantly for 5 months. While the unemployment did slowly recover, it again stagnated around the 11 to 12 percent mark for the rest of the time period. The hospitalizations follow the same pattern as Sacramento, with a small initial spike in hospitalizations and a very high spike in hospitalizations from September 2020 to November 2020.

(iii): Colusa County[4]



Analyzing the distribution of monthly morbidity and mortality in Colusa County, we encountered an issue as a result of Colusa's small overall population. We cannot accurately compare whether or not the Sacramento County trend is present - because the total number of deaths is so small, and there were zero deaths in eight of fifteen months, the percentage change between months is greatly inflated.



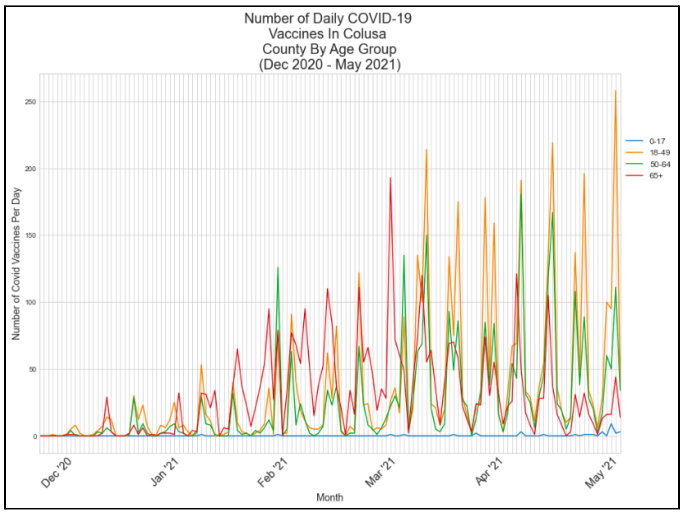
The data in this graph comes from the Labor Force and Unemployment data by county, which was provided by the California government Labor Market website. In order to display the trend of the labor force counts in Colusa county on a monthly basis, the data needed to be cleaned. Prior to plotting the data, the data frame was subsetted by only the years 2020 and 2021, the data frame's columns were renamed, the dates were sorted based on year first then the month, and the column for labor force counts had commas removed and were converted to float.

The plot of the monthly Labor Force counts in Colusa county shows that the labor force count plummeted in May 2020 from over 108k people in the labor force in April 2020 to less than 102k people in the labor force, which is approximately a 5.56% decrease in the labor force. That same

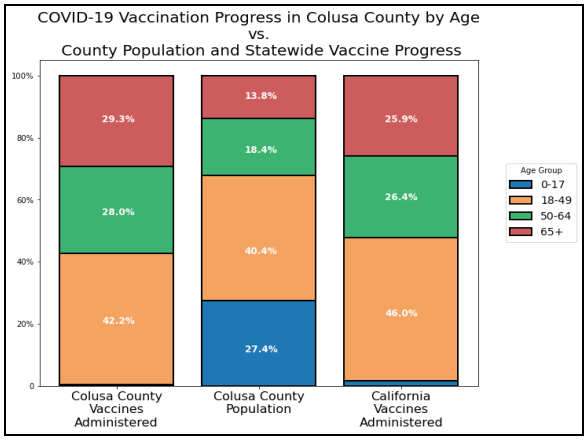
[4] Colusa County has not published a COVID-19 dashboard with cumulative figures. Thus, the protocol analyzing the estimated number of monthly deaths by race is excluded from this section.

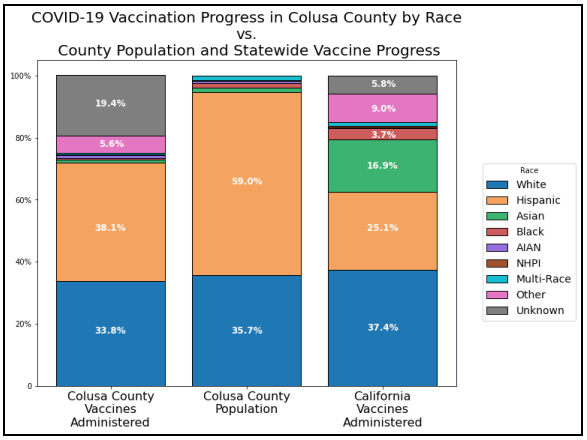
plummet was the lowest peak in the labor count. After that point, it can be seen that the labor force count went up slightly in July 2020 and fluctuated from then up until March 2021, which was the end date of the most recently available data. Also, the labor force count never went up to the amount that it previously was prior to February 2020. It's important to note that the fluctuation in labor force counts from July 2020 to March 2021 never reached its highest peak, which was over 112k people in the labor force in March 2020.

Additionally, if the labor force count is compared against the plots COVID case counts by tier for Colusa county as well as the other observed counties in the project and the labor force count trend across other counties, it can be noted that Colusa county entered the Purple Tier later than most of the counties observed. Right around the time that Colusa county first reached the Purple Tier in June 2020, the labor force count had plummeted around that time one month prior in May 2020.

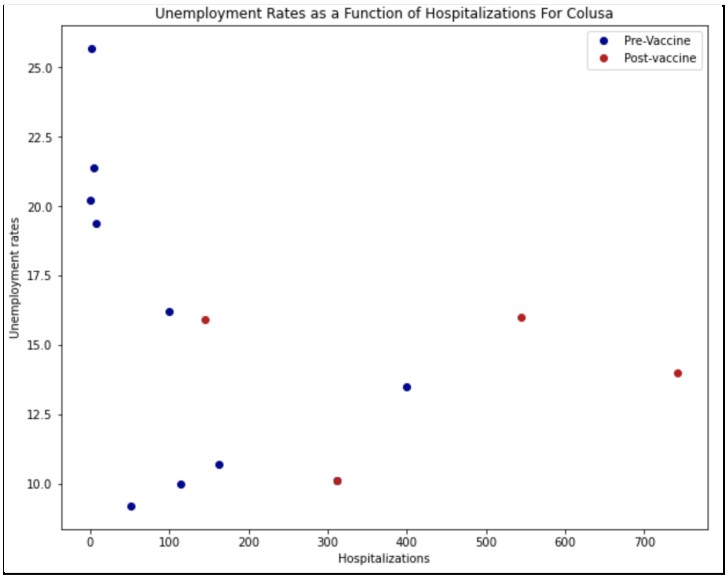


Unlike the two previously looked at counties, Colusa County did not have much difference between the three older age groups in regards to vaccination numbers. There was an overall increase in vaccinations as time went on, especially with the 18-49 age group. Despite eligibility for 16+ beginning in April, there was little increase in the 0-17 age group throughout the data set. Monday is the most popular vaccination day in Colusa County.



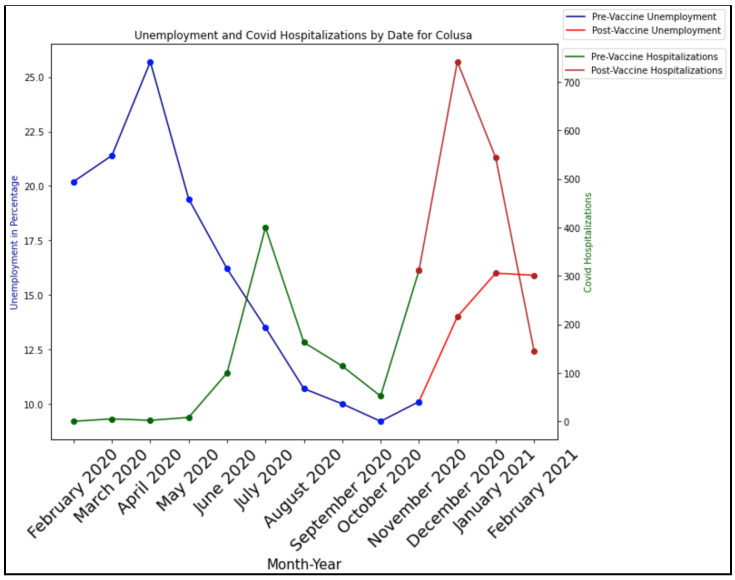


Comparing Colusa County's cumulative vaccination rates against those for the state, and against the county's population demographics, we see Colusa County has performed very well with respect to administering vaccines to vulnerable age groups. The 65+ and 50-64 age groups have received 29.3% and 28.0% of vaccines, respectively, while only 13.8% and 18.4% of the county population are represented by these age groups. However, Colusa County has only administered 38.1% of vaccines to Hispanic residents, who constitute 59% of their county population. However, we cannot draw conclusions from this observation, as the county's data quality with respect to vaccine administration was poor - nearly 20% of data regarding the race of vaccine recipients was 'Unknown'.



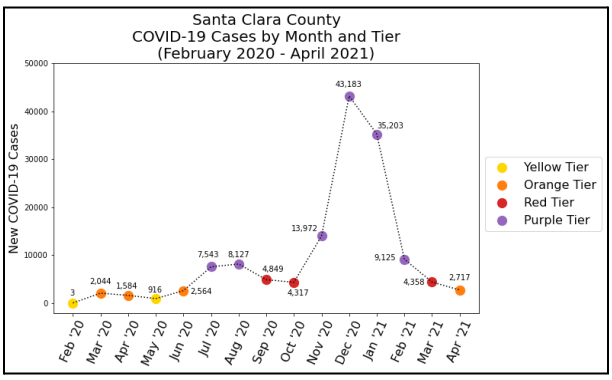
While this scatterplot does still vaguely follow the pattern from the previous two counties, this is the first observation that begins to Have more significant outliers than the experiment has produced. One of the pre-vaccine observations is at the 400 mark in terms of hospitalizations, where most of the observations are below the 200 mark. Furthermore, one of the post-vaccine

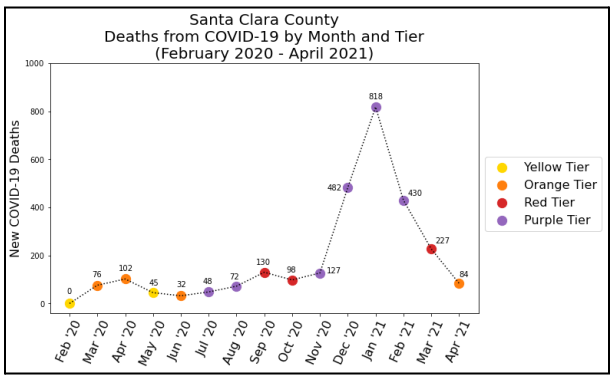
observations is in the ten percent region, where most of the observations are in the 15 percent region. It should be noted that the hospitalization rate is much lower and the unemployment rates are much higher in this observation in contrast to the counties in this experiment.



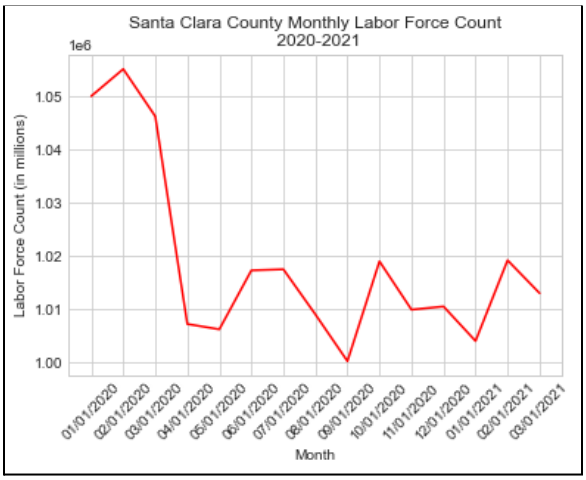
While this graph has the same general pattern as the other observed counties, there are some interesting discrepancies that should be noted. For example, in contrast to the other counties, the unemployment rate was already extremely high at the start of the experiment, being at 20 percent in February 2020 with a light spike in the next two months. Then, the unemployment in Colusa experienced a steep decline in the next 6 months, before experiencing a constant rise in employment for the rest of the observation period. The hospitalizations follow the same pattern as the previously observed counties, except with much smaller numbers, and a larger spike between the months of May and July.

(iv): Santa Clara County



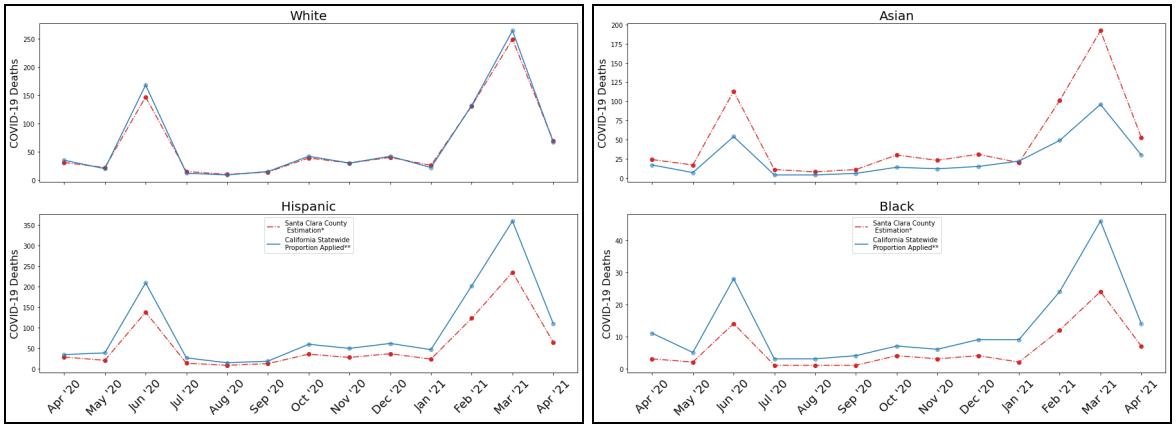


Analyzing the distribution of monthly morbidity and mortality in Santa Clara County, we see the pattern which occurred in Sacramento County is repeated - the number of new cases in a month, though not necessarily causative of the number of new deaths in that month, is greatly predictive of mortality in the coming month(s). For example, following a 225% increase in cases, and only a 30% increase in deaths, from October 2020 to November 2020, December 2020 saw an increase in cases and deaths of 210% and 280%, respectively, before January 2021 saw a 20% decrease in cases while deaths continued to rise, increasing by 70% from the previous month.

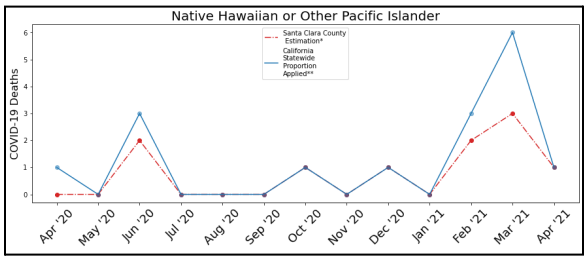


The data in this graph comes from the Labor Force and Unemployment data by county, which was provided by the California government Labor Market website. In order to display the trend of the labor force counts in Santa Clara county on a monthly basis, the data needed to be cleaned. Prior to plotting the data, the data frame was subsetted by only the years 2020 and 2021, the data frame's columns were renamed, the dates were sorted based on year first then the month, and the column for labor force counts had commas removed and were converted to float.

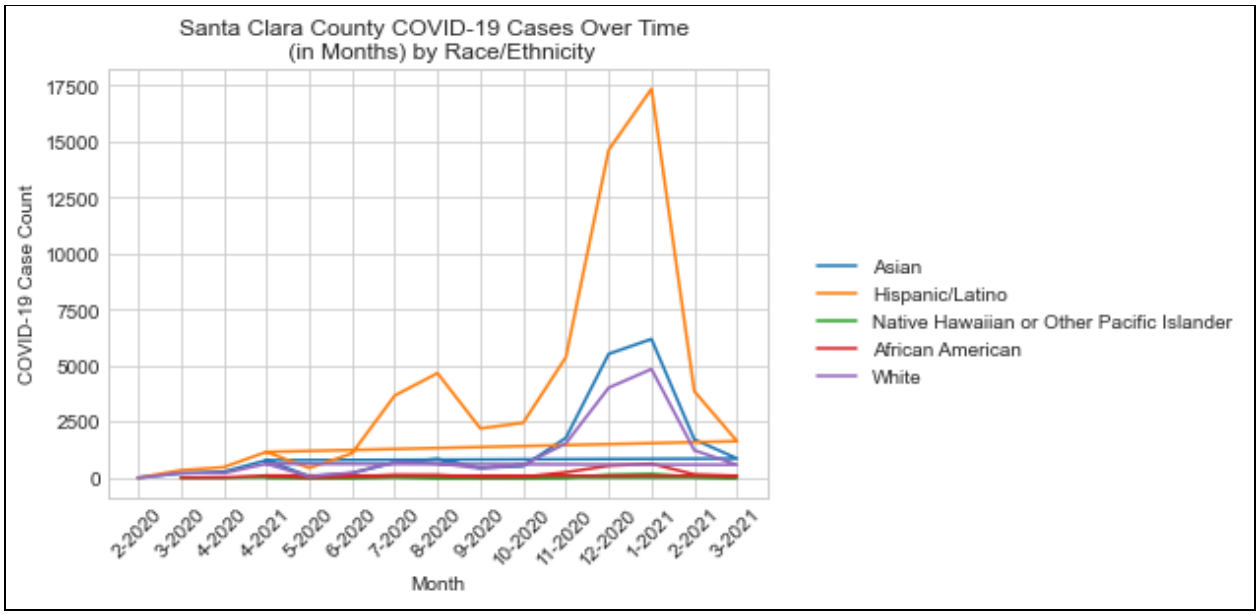
The plot of the monthly Labor Force counts in Santa Clara county shows that the labor force count plummeted in April 2020 from over 1.4 million people in the labor force in March 2020 to less than 1.1 million people in the labor force, which is approximately a 21.43% decrease in the labor force. The lowest peak in the labor count within Santa Clara county was about 1 million on September 1, 2020. However, after that point, it can be seen that the labor force count spiked up slightly on September 1, 2020, which likely attributed to the COVID-19 pandemic and the lockdowns that were in place as the county was during August 2021. The labor force count began to fluctuate from September 1, 2020 up until March 1, 2021, which was the end date of the most recently available data. Also, the labor force count never went up to the highest amount that it previously was prior to February 1, 2020.



Analyzing the Santa Clara County COVID-19 mortality for different racial groups[5], we see White residents of Santa Clara County saw outcomes nearly identical to those at the state level. NHPI county residents saw nearly identical outcomes to the state level as well, except for peaks in statewide deaths during April and June 2020 and February and March 2021. During these peaks, NHPI county residents saw increases in deaths, but not to the same degree as occurred statewide. Hispanic and Black county residents saw fewer deaths than the state through the entire year, while Asian residents saw precisely the opposite. save peaks in June 2020 and March 2021, where county deaths were much higher than those in the state.



[5] Santa Clara County's COVID-19 dashboard did not use "American Indian or Alaskan Native" as one of its racial categories, so this portion of the analysis does not include AIAN.



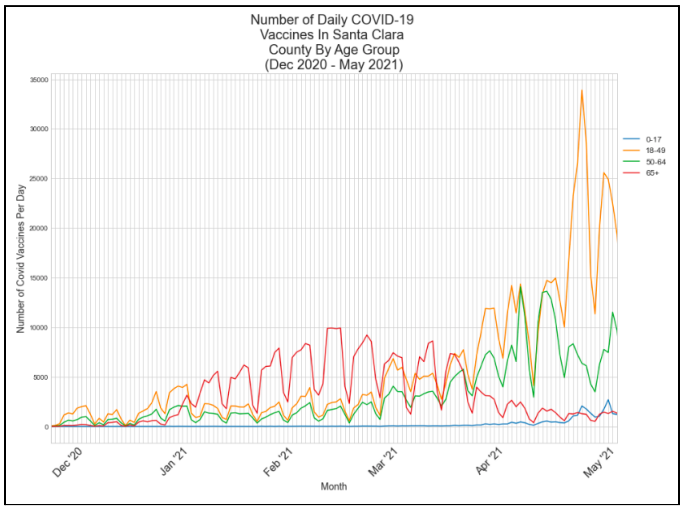
The data in this graph comes from Santa Clara's data on COVID-19 weekly cases by race and/or ethnicity. In order to display the trend of COVID-19 cases for each race/ethnicity on a monthly basis, the data needed to be aggregated and cleaned. Prior to plotting the data, the data frame's columns were renamed, a new column was created to label each of the case counts by month using the end date of the weekly case counts, the case counts had commas remove and were converted to float, groupby was applied to the data frame to group the cases for each race/ethnicity and month, and the data frame was finally melted, so that there would be one row for each month and race/ethnicity combination.

The plot of the estimated monthly COVID-19 in Santa Clara county shows that Hispanics often had cases at a higher frequency than most other races/ethnicities in the county since about April or June 2020, with the highest case count for Hispanics peaking at almost 17,354 cases in January 2021. The second highest case monthly COVID-19 trend was noted for Asians where the peaking at 6188 COVID-19 cases. It is clear based on the graph of estimated monthly COVID-19 in Santa Clara county that certain races/ethnicities had a higher rate of COVID-19 cases than others.

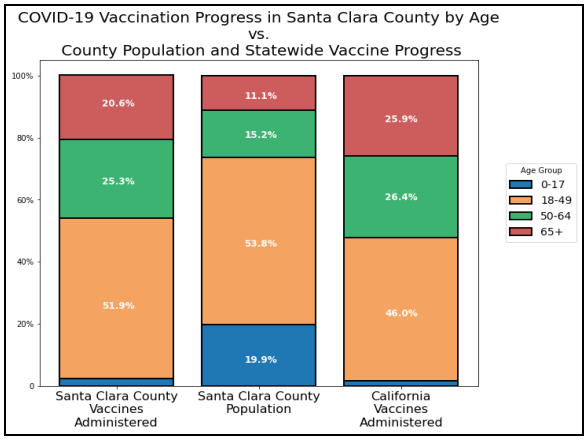
To discuss the summary statistics for the estimated monthly COVID-19 cases in Santa Clara County, on average for each race/ethnicity, African American people had 148 cases on a monthly basis, Asian people had about cases 1302 cases on a monthly basis, Hispanics/Latinos had about 3780 cases on a monthly basis, Native Hawaiian or Other Pacific Islanders had about 29 cases, and White people had about 1030 cases. These averages confirm the observations that were noticed by looking at trends alone. By observing these estimated monthly COVID-19 cases in Santa Clara County, it can be said that Hispanic/Latinos and Asians, on average, had higher COVID-19 cases on a monthly basis than most other groups.

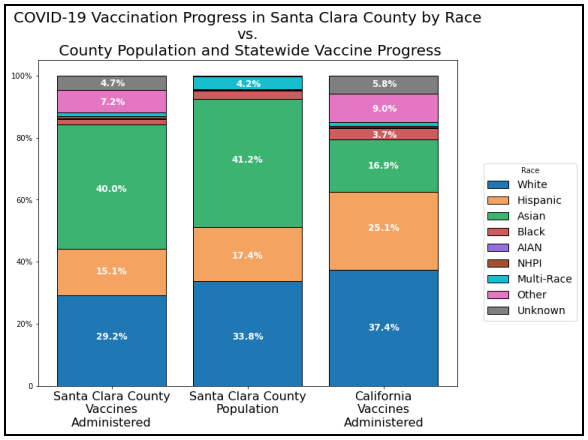
While the average case counts may be higher for certain groups than others, it is important to take into account that the estimated monthly case rates may not necessarily be the best way to show which races/ethnicities in Santa Clara county are most impacted as the plot did not account for the proportion of covid cases to the population counts of the various races/ethnicities. This is a plot that I would have made to the population counts by ethnicity in Santa Clara county on a monthly basis if the population data by race was available on a monthly basis. Also, the COVID case count graph was not made for the other counties in this project since the data was not available on a county level.

Here is the line graph comparing the vaccination numbers between the four age groups in Santa Clara County:

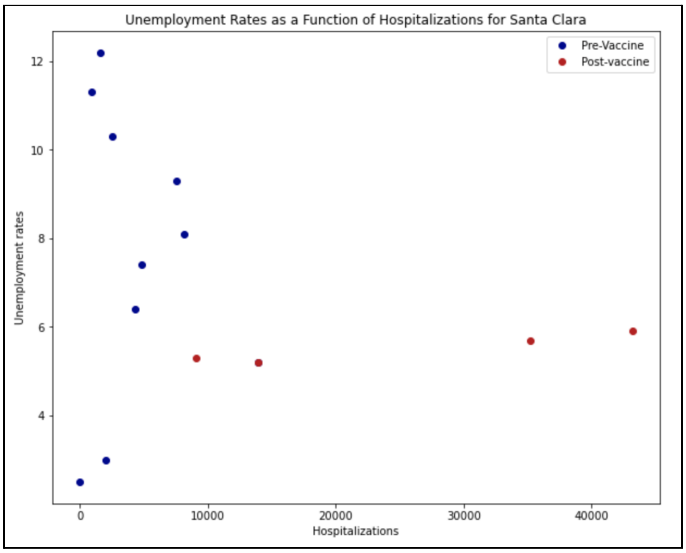


Vaccination numbers stayed low and each group had similar numbers until the beginning of February. As eligibility for 65+ began, that age group had dramatically higher numbers than the other age groups. Unlike the other three counties looked at in this report, there was not much growth in the 18-49 and 50-64 age groups before eligibility opened to all ages in April. After this point, there was a large increase in vaccinations among those aged 18-49 and 50-64. The biggest increase is in the 18-49 age group beginning mid-April. Throughout this time frame. The 0-17 age group remained low until the end of April, where a very small increase was seen. There is not a day that is more popular for vaccinations in Santa Clara County even though there is a similar pattern to the other counties.



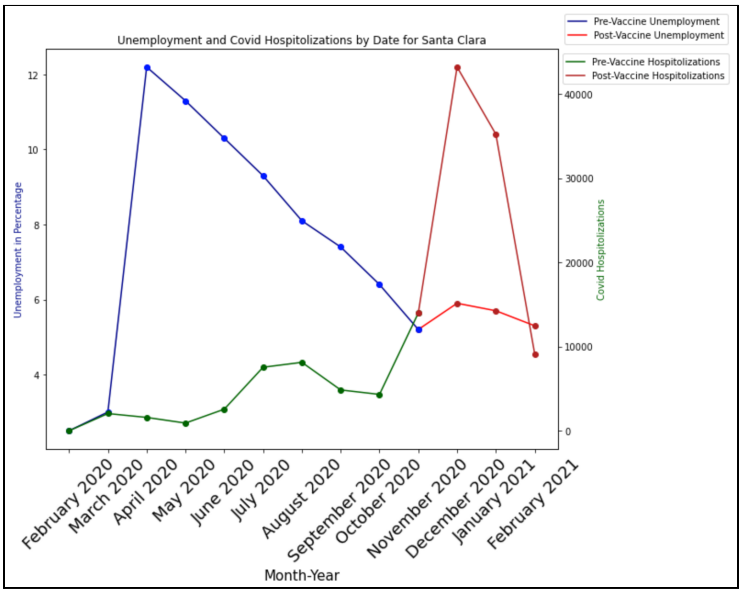


Comparing Santa Clara County's cumulative vaccination rates against those for the state, and against the county's population demographics, we see Santa Clara County performed poorly with respect to reaching vulnerable age and racial groups for vaccination. While older age groups received vaccinations at a higher rate than is represented in the population, the county had lower rates of vaccination for residents aged 50-64 and 65+ than the state did; residents aged 18-49 received the highest proportion of vaccines at 51.9%, with less than half of that going to the 50-64 age group, at 25.3%, and even fewer going to the 65+ residents, at 20.6%. With respect to racial groups, Santa Clara County administered vaccines to Hispanic, Asian, and Black residents at a rate lower than both the county population and the statewide vaccination rates. While it's possible these disparities would disappear if the 11.9% Other/Unknown data instead include a race, the disparities are uniform through racial categories, and the poor performance with respect to age indicates the likelihood of a strategic error regarding the distribution of vaccines is high.



Santa Clara returns to the pattern and unit size that we have observed in the previous counties. The pre-vaccine observations seem to appear under the 1000 mark for unemployment and have a

high variability for unemployment, while the post-vaccine observations have almost uniform unemployment as about 5 percent, and high variability in terms of their hospitalizations.



Santa Clara also shows a similar pattern to the other observations in our experiment. A very high spike in unemployment, followed by a decrease in unemployment over the course of the next 7 months where the unemployment level stabilizes. In terms of the hospitalizations, the line graph shows an initial low impact, followed by a low spike, followed by a drastic spike over the months of September to November.

References

I. California Local Area Unemployment Statistics

https://data.edd.ca.gov/Labor-Force-and-Unemployment-Rates/Local-Area-Unemployment-Statistics-LAUS-/e6gw-gvii

II. COVID-19 Time Series Metrics by County and State

https://data.ca.gov/dataset/covid-19-time-series-metrics-by-county-and-state1

III. COVID-19 Vaccination Progress in California:

https://data.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data/resource/e4aebae7-f004-420d-9737-259cfe7213c2

IV. Labor Market Information by County in California:

https://www.labormarketinfo.edd.ca.gov/geography/lmi-by-county.html

V. Santa Clara County COVID-19 Cases by Race/Ethnicity Over Time:

https://data.sccgov.org/COVID-19/COVID-19-cases-by-race-ethnicity-over-time/k7e9-hszp

VI. California Blueprint for a Safer Economy: https://covid19.ca.gov/safer-economy/

VII. Vaccination progress data: https://covid19.ca.gov/vaccination-progress-data/

VIII. Sacramento County Demographics Dashboard: https://datausa.io/profile/geo/sacramento-county-ca#demographics