UC Berkeley Data Analytics Bootcamp
Anna Chen, Melissa Acevedo, Nicholas Moore,
Ashlee Vo

Project Analysis

Project Topic and Goals

The purpose of our project is to analyze facets of the global pandemic COVID-19 within the scale of the United States. We aim to explore different social and socio-economic factors such as age, availability of health providers, employment, education, and urbanization on vaccination rates and death rates per state. The data we collected and used to conduct our analysis was sourced from the U.S Centers for Disease Control and Prevention as well as the U.S Census Bureau.

Death Rates and Vaccination Rates

We first wanted to explore one of the biggest known initiatives of COVID-19, which were vaccination rates. We began this analysis by creating a multi-bar chart to compare the two elements of death rate per state and raw deaths. This chart visualized that some states had a higher death rate relative to their total deaths, whereas other states had low death rates relative to their total deaths.

Raw Deaths vs Death Rate (R-value: 0.32):

Then, from what was visualized in the multi-bar chart we found the correlation between raw deaths and death rate per state. Utilizing the Pearson correlation coefficient, we were able to calculate the r-value as 0.32 which is considered a weak positive correlation. As expected, the difference between the count of raw deaths and death rates per state can have drastic effects on a population. For example, a state with a low raw death count but high rate will have more deaths relative to their population than a state with a high raw death count but low death rate.

Death vs Vaccination Rates (R-value: -0.75):

Lastly, we created a scatter plot to visualize the correlation between death rates and vaccination rates. The r-value was calculated to be -0.75 which is a strong negative correlation. This tells us that there is a strong correlation between death rates and vaccination rates. This supports the argument in support of vaccine effectiveness where when the percent of the total population vaccinated increases, the death rate decreases.

Vaccination Age Breakdown

Our project then conducted an analysis on age and vaccination rates. Individuals who were vaccinated were broken down into four different categories: ages 5+, 12+, 18+ and 65+ per state. We then created a multi-bar chart in order to visualize vaccination rates per state for each grouped age. The multi-bar chart allows us to see a breakdown of vaccination rates per state by

the categorized age groups and it appears that the older the age group, the vaccination rates tend to increase. To take this a step further, a pie chart was created to visualize vaccination rates on the national level. As expected, similarly to what we saw on the multi-bar, the older the age group, the higher the vaccination rate. The age group with the highest vaccination rates were seniors 65+ with a vaccination rate of 29.1%.

Available Health Providers and Vaccination Rates
Available Health Providers vs Death Rate (R-value: 0.18):
Another facet of COVID-19 our project explores is the amount of healthcare providers offering COVID-19 services and vaccination rates per state. We began this by creating a bar chart that visualizes the percent of available providers by state population. Some states were seen to have a higher ratio of providers relative to their population, however most were consistent. We then conducted a correlation analysis to see if there is a relationship between the number of total providers and death rates per state. The r-value was calculated to be 0.18 which is considered a very weak correlation. This conflicts with our hypothesis that states with more available vaccine providers will have a lower death rate.

Available Health Providers vs. Vaccination Rates (R-value: 0.18):
Another correlation analysis was conducted to explore if there is a correlation between the percent of fully vaccinated individuals and percent of providers per state population. Similarly, the r-value was calculated to be 0.18 which is also a generally weak correlation. This analysis tells us that the number or proportion of available health providers per state does not necessarily mean that the state will have lower death rates or higher vaccination rates.

Employment and Vaccination Rates
In this part of our research we look to explore if there is any relationship between the vaccination rate and the level of employment in the United States for COVID-19. We were curious about whether higher levels of unemployment would influence the amount of people getting vaccinated in the United States. Our hypothesis was that for states with higher levels of unemployment, a smaller percentage of the total population would be fully vaccinated compared to states with a lower percentage of unemployment. This is assumed because those unemployed are less likely to be required to be fully vaccinated (per working standard) or find the vaccination less appealing due to cost or social factors (such as uncertainty regarding its efficacy).

To start our analysis, we first created a bar graph plotting the percent total population vaccinated for each state versus the percent total population vaccinated for an overall picture of the relationship between the factors. The former is represented in blue, and the latter in red. We see that there is no obvious correlation between the two since we have states with a rather high vaccinated population having both high and low unemployment rates. This suggests that there is

little correlation between the two factors, which rejects our hypothesis that the two may be dependent on each other.

Furthering our research, we created scatterplots to support the results of our bar graph. The first scatterplot we created mirrors the bar graph, plotting vaccination rates on the y-axis against unemployment rates on the x-axis. We see that indeed there is no direct correlation between the two factors. The r-value is ~0.21, indicating a slight presence of a positive correlation between unemployment rate and vaccination rate, but with an r-value this low, it is too weak to make any conclusive statements. The results indicated that people are getting vaccinated regardless of employment status. For example, referencing the scatterplot, we see states with high unemployment rates between 3.0 - 3.5% with a percent of the total population vaccinated ranging from as low as ~55% to as high as ~81%.

If unemployment rates showed an insignificant trend between employment and vaccination, our last scatterplot strives to approach the topic from a different perspective by comparing percent of the total population in the labor force with vaccination rates per state. This scatterplot produced an r-value of 0.42, which is significantly greater than our scatterplot for unemployment rates. However, it is still a rather weak value which is not significant enough to make any conclusive statements. For example, referencing the scatterplot, at an approximately 65% population in the labor force, we have states with a percent total population vaccinated as low as ~53% to as high as ~88%. This means that we can officially reject the hypothesis, there is no significant correlation between the two coefficients as people in the United States are getting vaccinated regardless of employment status. Which makes sense as well since in the outbreak of a sudden pandemic, people want to be protected and safe regardless of working status. Also is a good sign for vaccine companies since this means that propaganda regarding the COVID-19 vaccine is effective.

Education and Vaccination Rates
The impact of educational levels on vaccination rates is a multifaceted and intriguing aspect of public health dynamics. Some research suggests that individuals with higher educational attainment may be more inclined to receive vaccinations due to a combination of factors. Higher education often correlates with increased health literacy, enabling individuals to better understand the importance of vaccines in preventing diseases. Moreover, individuals with advanced education levels may have better access to healthcare information and resources, facilitating informed decision-making regarding vaccinations. On the other hand, lower education levels may be associated with limited health information, potentially influencing vaccine hesitancy. Exploring this complex interaction between different educational levels and vaccination rates provides valuable insights for public health initiatives, allowing for more targeted and effective strategies to enhance overall community immunity.

In our investigation, we sought to uncover potential correlations between these variables across the U.S. states. We evaluated data for the highest educational level for people over 25 years old.

Our analysis started with the creation of scatter plots illustrating the distribution of vaccination rates relative to different education levels. While visual patterns were observed, we conducted a correlation analysis to quantify these relationships.

The findings revealed distinct patterns for each education category/level:

Less than High School Graduate (R-value: -0.25):
The correlation is negative and weak. As the percentage of individuals with less than a high school education increases, there is a tendency for the vaccination rate to slightly decrease.

High School Graduate, GED, or Alternative (R-value: -0.59):
The correlation is negative and moderately strong. As the percentage of individuals with a high school education, GED, or alternative increases, the vaccination rate tends to decrease. The relationship is more robust compared to the "Less Than High School Graduate" category. It is important to note that this category does not include people that went on to higher degrees.

Some College or Associate's Degree (R-value: -0.61):
A moderate to strong negative correlation suggests a significant association between an increase in the percentage of people with some college education or an associate's degree and a decrease in the vaccination rate. It is important to note that this category does not include people that went on to higher degrees.

Bachelor's Degree or Higher (R-value: 0.79):
The correlation is strong and positive. It highlights a substantial association between an increase in the percentage of people with a bachelor's degree or higher and a significant increase in the vaccination rate.

With these results, there is a consistent trend: higher education levels are associated and linked to higher vaccination rates. However, the strength of these correlations vary across education levels. In general, these findings support our hypothesis that the vaccination rate tends to be higher in states with a more educated population.

Urbanization and Vaccination Rates
Cities and urban landscapes have many theoretical disadvantages when faced with a virulent person to person transmitting disease; on average they have more people living and working within a geographic area, more avenues of transmission via increased public transit usage and

more people working in office or indoor jobs. There could also be theoretical advantages of being an urban resident in a pandemic such as increased resources (for example more vaccination sites, available doses and information) from local government or additional regulations enacted to prevent the spread of the disease.

While each of these individual factors could be a separate analysis, looking at the percent of the population in each US state which resides in an urban landscape would show, on the whole, if more urbanized states fared better in vaccinating their populations and reducing the amount of Covid-19 deaths.

Urban Percentage of Population vs Vaccinated Percentage of Population (R-Value: 0.404):
The data shows that there is a moderate positive correlation between the percent of a state's population that lives in urban areas and the percent of population which completed the primary vaccination series. Meaning that as the share of urban population increases, so does the percentage of population that was vaccinated.

Given that there was a negative correlation between the percent of population which was vaccinated and the amount of Covid deaths, this would suggest that the moderate positive correlation between urban populations and vaccinated populations might also have reduced deaths.

Urban Percentage of Population vs Covid-19 Death Rates per 100,000 People (R-Value: -0.159):
However when correlating the share of urban population to the rate of deaths from Covid-19, there was only a weak negative correlation. A number of states with comparable urban shares of population had significantly different rates of deaths for Covid-19. This suggests that even though more urban populations had more vaccinated people, the death rates may have been affected by many other factors.

Limitations of the Project
1. Granularity of the data
2. Massive number of possible factors
3. More in depth analysis can be conducted on populations at a smaller scale, considering population densities which may vary from state-wide data
4. Data availability and cleanliness

Conclusion
Conclusions that we drew from the study were that although some states had high raw deaths, other states had higher death rates relative to their population which can drastically change the effect of COVID-19 between states. We also found that there is a strong correlation where the higher the percentage of fully vaccinated, the lower the death rate which is a strong argument for

pushing to increase vaccination rate. In exploring differences in age groups and vaccination rates, we found that the older the age group, the percent of fully vaccinated within that group increases. Since vaccination effectiveness was established, initiatives could be taken to raise the vaccination rates for the age groups with lower percentages for preventative purposes. In respect to available providers relative to state population, our results showed that the amount and proportion of COVID-19 health centers relative to population has a minimal effect on vaccination and death rates. This could tell local governments and policy makers to focus more on preventative initiatives and community education over simply having more providers available. Our exploration of employment status showed that employment rates had little to do with vaccination rates. Regardless of employment status, people were still taking the initiative to get vaccinated. This could also tell local initiatives that there are other factors rather than employment to focus on in raising vaccination rate. Additionally, vaccination rates on levels of educational attainment varied. The most significant finding was that vaccination rates are the highest between populations who have earned a Bachelor's degree or higher. Therefore, education and initiatives for COVID-19 prevention could focus on groups with lower levels of educational attainment. Lastly, urban areas tend to have higher vaccination rates, however death rates between states varied, which may be attributed to other external factors that the study was limited in exploring.