DATA VISUALIZATION

PROJECT REPORT ON

Covid Vaccination Efficacy: A Comprehensive Analysis of COVID-19 Data of United States

Submitted By

Group 7

Sai Snehal Grandhi – SXG230071

Devang Makhdiya – DXM220063

Deepa Nalla – DXN230011

Chethan Rajendra – CXR220016

Tanya Shree Verma – TXV220020

Syed Tiham Zaki – TXS220016

Under The Guidance of:

Prof : Naser Islam

EXECUTIVE SUMMARY

The COVID-19 pandemic has had widespread impacts on individuals of all age groups in the United States during the year 2021. The rapid development and deployment of vaccines marked a significant achievement in the field of medicine and public health. The unprecedented speed of vaccine development was facilitated by advancements in vaccine technology, global collaboration, and robust regulatory processes.

The effectiveness of COVID-19 vaccinations in preventing symptomatic illness, severe disease, and death caused by the virus has been repeatedly shown by clinical trials and real-world data. Immunization has been essential in lowering the rates of virus transmission and preventing the virus's progress within communities. Ensuring equitable access to vaccines has been a significant challenge, with disparities observed in vaccine distribution and uptake across different demographic groups.

Our comprehensive dataset aims to delve into the incidence of COVID-19 cases and fatalities, categorized by age group and vaccination status, including the administration of booster doses. Through this analysis, we seek to elucidate the effectiveness of vaccination strategies and discern any disparities in insights and outcomes.

These detailed insights derived from our analysis are pivotal for informing evidence-based public health policies and interventions aimed at effectively combating the COVID-19 pandemic across diverse demographic segments. By understanding the complex interplay between age, vaccination status, and COVID-19 outcomes, policymakers and healthcare professionals can tailor targeted interventions to protect vulnerable populations, mitigate transmission risks, and ultimately save lives.

The deadly pandemic has highlighted the importance of global collaboration and solidarity in responding to public health emergencies. International collaborations in vaccine development, manufacture, and distribution have been crucial in expediting vaccine accessibility and guaranteeing just distribution of vaccines across nations with disparate resources and capacities.

TABLE OF CONTENT

INTRODUCTION

In the spring of 2020, the COVID-19 pandemic posed challenges in public health that the world had never seen before. In the United States, which went through the worst of the pandemic for several months, the need to mitigate the effects of the pandemic led to a vaccination campaign unlike any other seen in recent times. This campaign generated a wealth of data that allow for the study of vaccination rates by age and sex, and the distribution of those rates by state.

This information needs to be analyzed carefully, since early reports did suggest huge disparities in vaccination rates in some groups compared with others. Doing this will help us determine the success of public health strategies and suggest future directions in vaccination efforts and communication strategies. It will also explain variations in COVID-19 deaths from state to state. In March 2020, California, Texas, Florida, Pennsylvania, and New York were among the worst hit by the pandemic.

With almost daily doses of detailed data released through the vaccination rollout – whether it's the relative age and sex vaccination trends, the state-by-state COVID-19 mortality curves or the international comparisons – we're given a very rare opportunity to empirically investigate the behaviors of a complex population responding to dynamic public health policy and how this interaction impacts health outcomes. The general assumption is that as more people in the population are vaccinated, the mortality rate should decrease accordingly. However, there is complexity here that can be unpacked to improve tailored interventions.

This dataset is layered with meaning, and as we dig into it, we will describe the effects of vaccination uptake on how pandemics unfold directly and indirectly. By comparing the demographic composition of vaccinated populations in each state to their mortality, we can answer key questions about the efficacy of the vaccine campaign and the resilience of the country's healthcare systems. This introduction paves the way for in-depth discussion of the context, complexities, and consequences of COVID-19 vaccination worldwide, and the lessons we can gain from it to better prepare the public for future pandemics.

DATA DESCRIPTION

1. **Data Source:** https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Age-and-Sex-Trends-in-the-Uni/5i5k-6cmh/about_data

The dataset is a comprehensive record of COVID-19 vaccination progress across various states in the United States. Spanning from early 2021 to the latest entries, this dataset offers detailed daily insights into vaccination efforts, crucial for analyzing trends and the effectiveness of vaccination campaigns.

The dataset includes 17 columns with over 22,936 entries, representing individual records for different dates and locations. Each entry includes the following fields:

This dataset is instrumental for monitoring vaccination uptake and effectiveness across different demographics and geographic locations. It helps public health officials and policymakers evaluate the success of vaccination campaigns and strategize future initiatives. The data can be analyzed to determine vaccination rates over time, identify disparities in vaccine distribution or uptake, and assess the impact of vaccination on controlling the spread of the virus.

By offering both raw numbers and normalized data per hundred or per million, the dataset provides a robust foundation for comparative studies and temporal analysis. For instance, examining "Daily Vaccinations Per Million" alongside "People Fully Vaccinated Per Hundred" offers insights into how quickly different regions are approaching herd immunity. Moreover, metrics like "Share Doses Used" help evaluate the efficiency of vaccine deployment within states, crucial during times when vaccine supply might be limited.

Date: The specific date for the data entry.

Location: The U.S. state or territory to which the data pertains.

Daily Vaccinations: Number of vaccine doses administered on that day.

Daily Vaccinations Per Million: The rate of daily vaccinations per million people.

Daily Vaccinations Raw: Day-to-day changes in the total number of administered vaccine doses.

Distributed Per Hundred: Number of vaccine doses distributed per 100 people in the population.

People Fully Vaccinated: Total number of people who have received all doses prescribed by the vaccination protocol.

People Fully Vaccinated Per Hundred: Percentage of the population that has been fully vaccinated.

People Vaccinated: Total number of people who have received at least one dose of the vaccine.

People Vaccinated Per Hundred: Percentage of the population that has received at least one vaccine dose.

Share Doses Used: The proportion of distributed vaccines that have been administered.

Total Boosters: Number of booster doses administered.

Total Boosters Per Hundred: Number of booster doses administered per 100 people.

Total Distributed: Total number of vaccine doses distributed.

Total Vaccinations: Total number of vaccine doses administered.

Total Vaccinations Per Hundred: Total vaccinations administered per 100 people in the population.

Zero Line: A constant field used for plotting data in visualizations.

2. **Data Source:** https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/vaccinations/us_state_vaccinations.csv

The dataset is a detailed compilation of vaccination data across the United States, segmented by age and sex. This extensive dataset, with over 1,048,575 records, offers insights into the distribution and uptake of COVID-19 vaccines, highlighting demographic patterns.

This dataset is critical for understanding the vaccination coverage across different demographics and geographic locations. It allows public health officials and policymakers to identify areas with low vaccination rates and demographic groups that may require targeted outreach or additional resources. The inclusion of booster and second booster data is particularly valuable for assessing ongoing immunity and protection against COVID-19, especially in the face of emerging variants.

The data can be analyzed to evaluate the success of vaccination campaigns, the readiness of populations to face COVID-19 challenges, and the equity of vaccine distribution. Such analysis can inform decisions on resource allocation, booster shot campaigns, and public health strategies aimed at achieving widespread vaccine coverage and ultimately controlling the pandemic. Moreover, this dataset provides the foundation for detailed statistical analysis and modeling to predict future vaccination needs and assess the impact of vaccination on pandemic dynamics.

The dataset comprises the following 13 columns, each providing specific information related to COVID-19 vaccination trends:

Date: The date on which the vaccination data was recorded.

Demographic Category: Specific demographic details, specifying age groups and sex.

Demographic Category (group): A broader categorization of demographic data for analysis.

Location: The U.S. state where the data was collected.

Administered Dose1: The number of first doses of the COVID-19 vaccine administered.

Administered Dose1 pct age group: The percentage of the specified age group that has received at least the first dose of the vaccine.

Booster Doses: The number of booster doses administered.

Booster Doses Vax pct age group: The percentage of the specified age group that has received booster doses.

Census: The total population of the demographic category according to the most recent census data.

Second Booster: The number of second booster doses administered.

Second Booster Vax pct age group: The percentage of the specified age group that has received the second booster dose.

Series Complete Pop pct age group: The percentage of the specified age group that has completed the vaccination series.

Series Complete Yes: The total number of individuals in the specified demographic who have completed the vaccination series.

3. Data Source:

The dataset provides a comprehensive view of COVID-19 statistics across various states in the USA, with each record detailing the confirmed and probable cases and deaths, among other metrics. This dataset contains 21,900 entries and spans multiple states and dates, offering a broad perspective on the pandemic's impact.

The dataset serves as an essential tool for tracking the progression of the COVID-19 pandemic across different regions and over time. It provides a detailed breakdown of both confirmed and probable cases, which is vital for understanding the full scope of the virus's impact. The distinction between confirmed and probable cases helps in refining the accuracy of public health assessments and ensuring that measures are tailored to the actual situation.

By comparing these statistics with the national averages provided, policymakers, researchers, and public health officials can gauge the severity and trajectory of outbreaks in specific areas relative to broader national trends. The inclusion of data consent columns (for cases and deaths) also highlights the importance of data governance and privacy considerations in managing public health information during a crisis.

Overall, this dataset not only aids in immediate response efforts but also in long-term strategic planning and resource allocation to mitigate the effects of the pandemic. It supports detailed epidemiological studies and helps inform public communications about the ongoing public health responses.

Key columns in the dataset include:

Consent on Cases: Indicates whether there was consent to share data on confirmed cases for each state.

Consent on Deaths: Indicates whether there was consent to share data on confirmed deaths for each state.

Date: The date on which the data was recorded.

State: The name of the state where the data was collected.

USA: A general regional classification of the state (e.g., East Coast, West Coast, Central).

Confirmed Cases: The number of laboratory-confirmed COVID-19 cases.

Confirmed Deaths: The number of deaths attributed to laboratory-confirmed COVID-19 cases.

Mortality Rate: The percentage of confirmed cases that resulted in death.

National Average of Total Cases: The national average of total confirmed cases as of the date.

National Average of Total Deaths: The national average of total confirmed deaths as of the date.

New Cases: The number of new confirmed cases reported on the date.

New Deaths: The number of new confirmed deaths reported on the date.

Probable Cases: The number of probable COVID-19 cases.

Probable Deaths: The number of deaths attributed to probable COVID-19 cases.

Probable New Cases: The number of new probable cases reported on the date.

Probable New Deaths: The number of new probable deaths reported on the date.

Total Cases: The cumulative total of confirmed COVID-19 cases up to the date.

Total Deaths: The cumulative total of confirmed COVID-19 deaths up to the date.

DATA CLEANING

We used Microsoft Excel for this project to facilitate and simplify the work with the material collected. Sorting of the material is one of main available features in Excel and is especially useful when it comes to finding and deleting errors and duplication.

Conditional formatting within Excel can highlight any data points that meet certain criteria, useful for spotting potential inconsistencies. Through filtering features, for instance, sorting data by dates or alphabetical order, more detailed analysis of specific data points can be carried out.

To summarize, Excel's overall capabilities in terms of managing data become more effective with additional functional features such as sorting, filtering, and formatting. Importing and exporting data in various formats also helps make Excel a powerful tool that can manage a range of data activities.

Visualization Tool

Tableau - a business intelligence and data visualization tool were used for this project. It is the best fit for an enterprise that needs to integrate with multiple data sources. Thanks to Tableau you can connect to many different databases and spreadsheets of various formats and locations. It allows you to perform advanced analysis, such as running complex calculations and predict the future patterns and outcomes and detect patterns in your data so you can make better decisions based on your precise and actionable data insights.

Tableau’s eye-catching and easy-to-understand dashboards instantly communicate data in real time. A quick look not only shines a spotlight on contextual information but allows for drilling down to specific spot trends and outliers. Tableau’s advanced analytic tools are a vital resource for companies in identifying new opportunities and sharpening their operational decisions that will improve performance. The dashboards in Tableau are also customizable, which is important in helping to drive speed to decision for the enterprise.

In conclusion, Tableau is the most important tool which can help businesses and organization to deal with large sets of data from multiple sources. It also plays an important role by providing the best tools that can help to discover hidden trends and patterns within the data. Throughout this

report we use Tableau to analyze and visualize data that enabled us to make better decisions based on the findings.

DATA INSIGHTS AND FINDINGS

We used Tableau and Microsoft Excel to work with and visualize multiple data sets that can be used to identify trends and patterns important to the COVID-19 vaccination rate in multiple demographics and multiple states in the US. This visualization made it possible to spot key trends and patterns that can be used to understand the pandemic and the effect of vaccination towards curbing the pandemic.

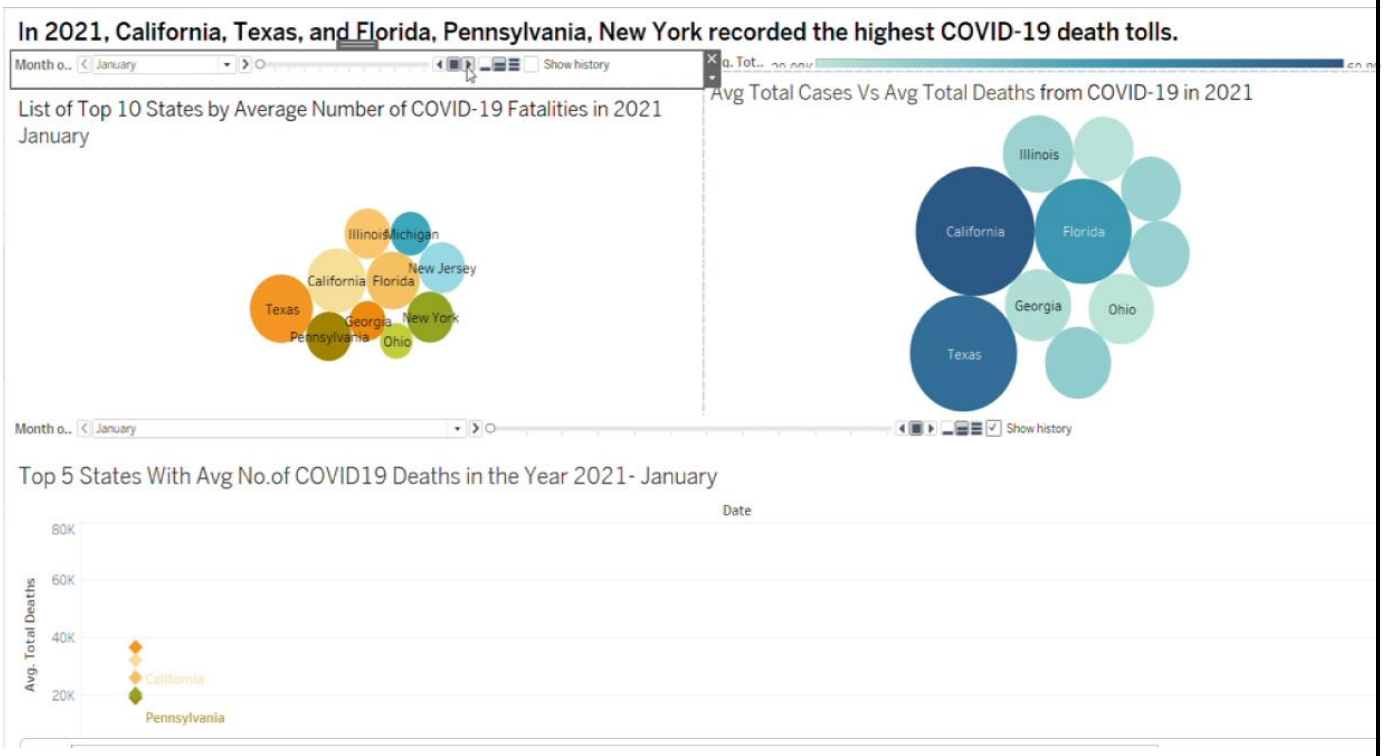
1. Vaccination Trends by Age and Sex: The age and sex trends in vaccination for COVID-19 in the United States were analyzed using the chart on the right. From the given chart, the older age groups were given priority in the initial phase of vaccination campaign and the vaccination rate of the old age people and the 16-19 age group were much higher than the other age groups. This might be an outcome of the targeted vaccination program by the US government. Moreover, all the states showed higher vaccination rate of females as compared to males. It should be noted that The UK government implemented a similar tax on sweetened beverages due to the similar trends in the US. Overall, we can conclude that the targeted vaccination program for specific groups of people has yielded positive results, and a public health campaign is required in the future to tackle any disparities.

2. State-Level Vaccination Coverage: From the ‘US State Vaccinations’ dataset, we can see that the states with the largest urban populations – New York, California and Texas – reached much higher numbers more quickly, but when we account for the population size, some of the smaller states – for example, Vermont and Maine – came out at the top, suggesting that state-level policy and healthcare infrastructure make a big difference.

3. Impact on COVID-19 Mortality Rates: Plotting vaccination data against COVID-19 mortality rates enabled the data to tell an important story. States that had early and successful vaccination rollouts, clearly showed a more dramatic decline in COVID-19 mortality.

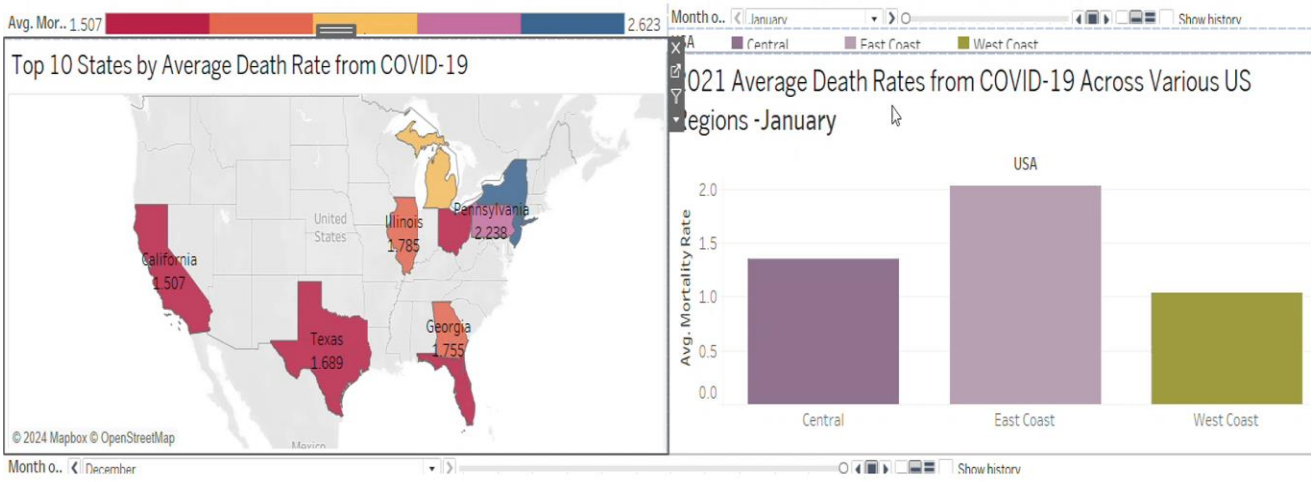
HYPOTHESIS 1

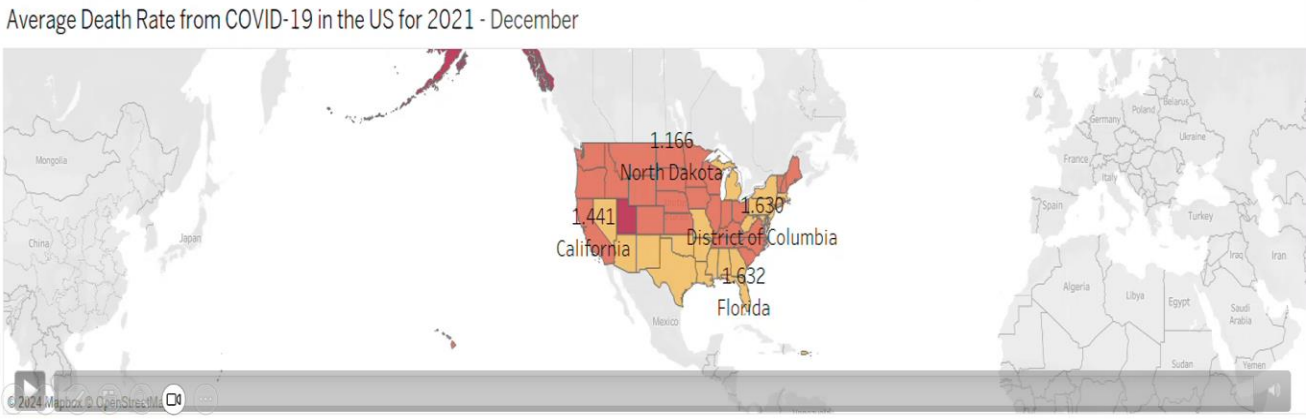
By the end of 2021, California, Texas, and Florida will continue to exhibit the highest COVID-19 death tolls among U.S. states, potentially comprising approximately 60% of the nation's total COVID-19 fatalities, based on historical trends and population density.



HYPOTHESIS 2

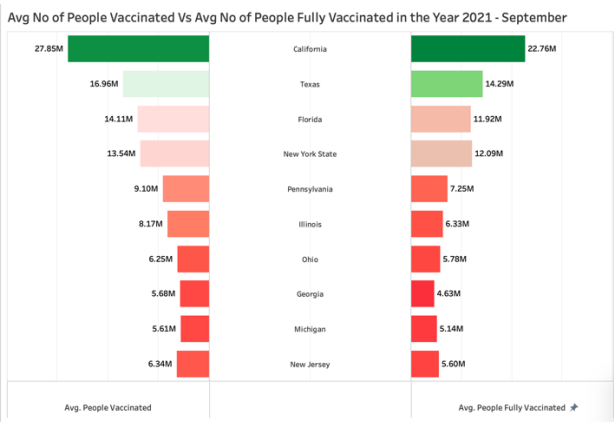
Throughout 2021, from January to December, California, Texas, Florida, Pennsylvania, and New York recorded the highest COVID-19 mortality rates, cumulatively accounting for over 50% of the United States' total mortality rate, as evidenced by trends in healthcare accessibility, population density, and the varying public health responses.

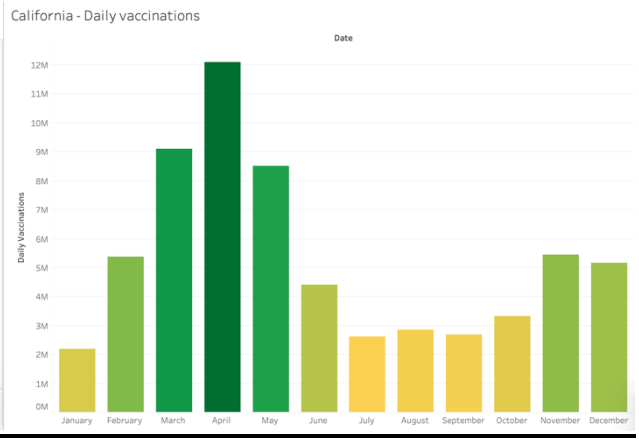


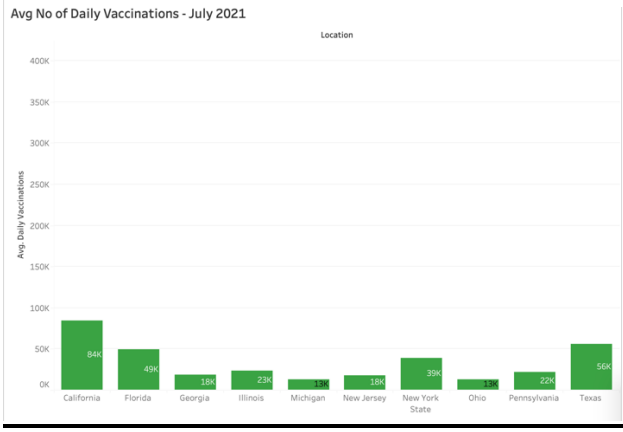


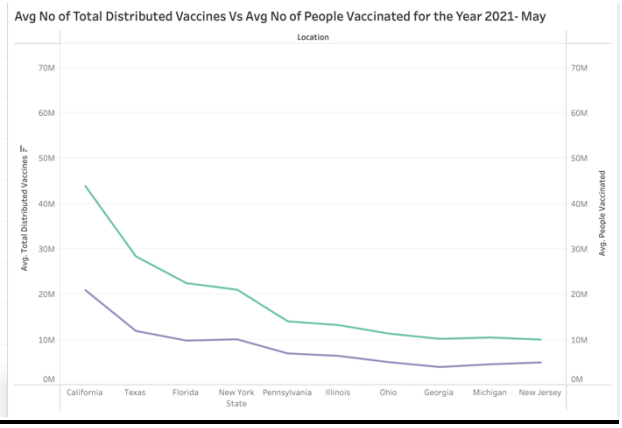
HYPOTHESIS 3

During the period of January 2021 to December 2021 in United States, Number of vaccinations distributed are the highest in California, Texas and Florida indicating heavy populated states have more transmission of virus. It indicates 35% of the daily vaccinations are comprised by the states California, Texas, and New York.



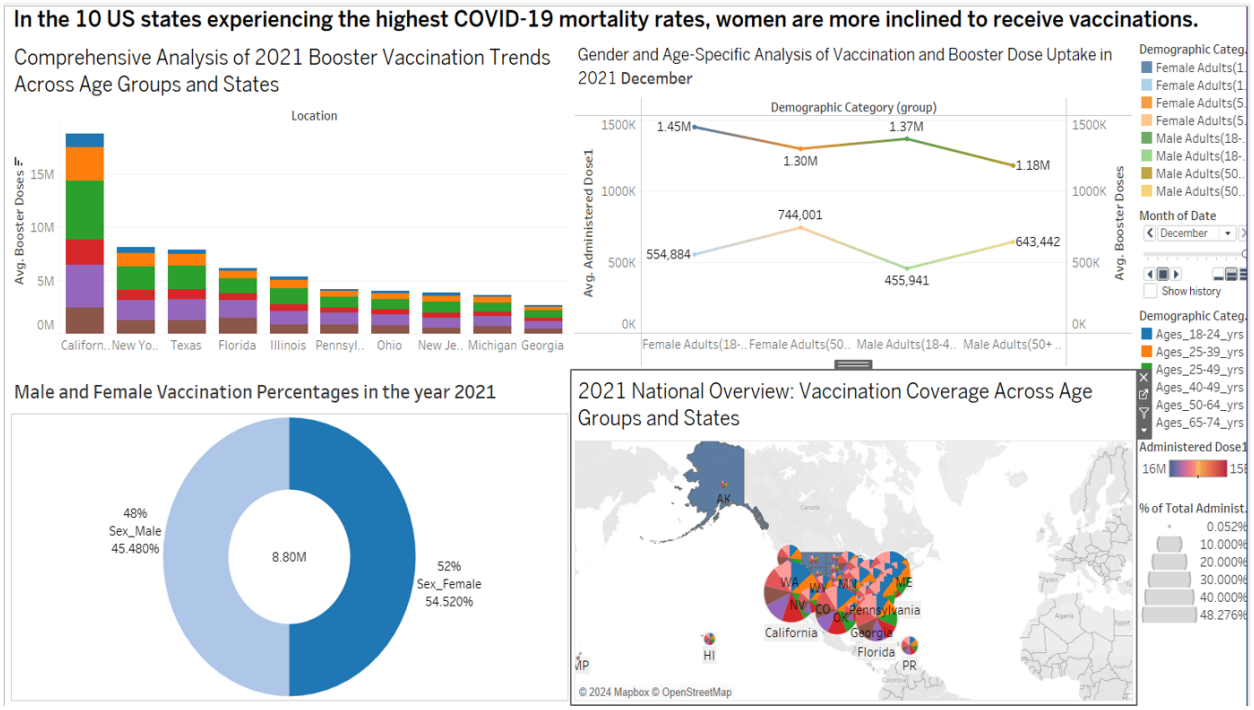






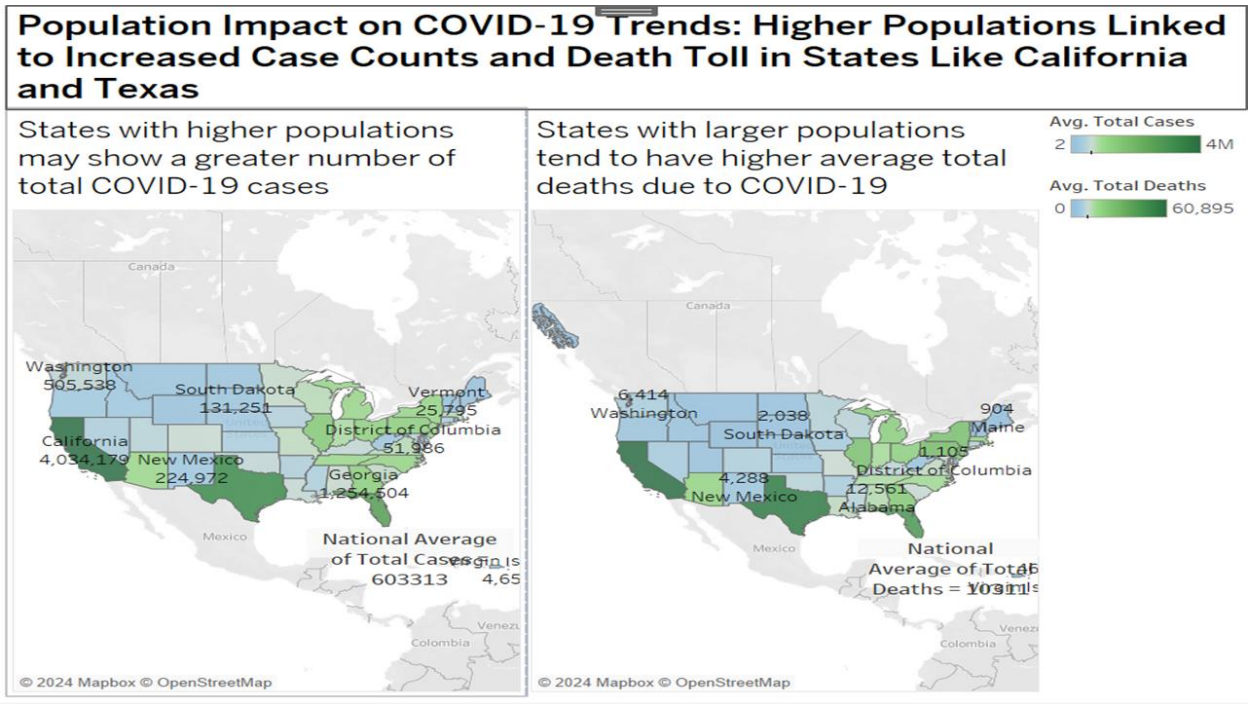
HYPOTHESIS 4

In the 10 US states experiencing the highest COVID-19 mortality rates, Women are more inclined to receive vaccinations.



HYPOTHESIS 5

By the end of 2021, heavily populated states like California, Texas, New Mexico, and Georgia are likely to have higher total COVID-19 case counts and death tolls compared to states with smaller populations. It is observed that densely populated areas may have a greater share of COVID-19 cases and fatalities, which could exceed 50% of the nation's overall figures. This projection is based on the historical trends that demonstrate a correlation between higher population densities and increased virus transmission rates.



CONCLUSION

HYPOTHESIS 1

By the end of 2024, California, Texas, and Florida will continue to exhibit the highest COVID-19 death tolls among U.S. states, potentially comprising approximately 60% of the nation's total COVID-19 fatalities, based on historical trends and population density.

- The graph showing deaths by COVID19 in 2021 indicates that, the five most populated states, California, Texas, Florida, Pennsylvania, and New York where most people died. This shows a relationship between the population density and pandemic impact. This means that the greatest population density in the world need the best health systems to cope with diseases prevention because clearly in these regions the death rate increases a lot from pandemic. It is fundamentally important that countries focused on public health policies have the best contingency plans to prevent some health crisis from happening again.

HYPOTHESIS 2

Throughout 2021, from January to December, California, Texas, Florida, Pennsylvania, and New York recorded the highest COVID-19 mortality rates, cumulatively accounting for over 50% of the United States' total mortality rate, as evidenced by trends in healthcare accessibility, population density, and the varying public health responses.

- The 2021 data visualizations confirm that California, Texas, Florida, Pennsylvania, and New York had the highest COVID-19 mortality rates in the nation, aligning with the hypothesized trends. The displayed shades of red across the maps directly correlate with the severity of mortality rates in these states, evidencing the vulnerability of populous regions. Over the year, a transition to lighter shades suggests a downward trend in mortality rates, reflecting the impact of improved medical treatments and the widespread adoption of vaccinations. This data not only substantiates the hypothesis but also captures a narrative of adaptation and resilience, marking a year of public health challenges and progress against the pandemic.

HYPOTHESIS 3

During the period of January 2021 to December 2021 in United States, Number of vaccinations distributed are the highest in California, Texas and Florida indicating heavy populated states have more transmission of virus comprising 35% of total vaccines distributed.

- Line chart demonstrates a surge in vaccine distribution, particularly in California and Texas, during the third and fourth quarters of 2021, coinciding with an increase in COVID-19 cases. Bar graph illustrates a trend towards full vaccination by December 2021, with California leading and New Jersey decreasing. The color change from red to green highlights the rise in vaccinations per state. Overall, these visualizations suggest a strong correlation between robust vaccination distribution efforts and a decrease in COVID-19 transmission, reinforcing the effectiveness of vaccination campaigns in curbing the spread of the virus.

HYPOTHESIS 4

In the 10 US states experiencing the highest COVID-19 mortality rates, Women are more inclined to receive vaccinations.

- As you can see in the visualizations done for this hypothesis. Whenever there is a direct comparison with males and females in the percentage of vaccinations received women are more. Also, in a comparison where it shows the number of booster doses and vaccinations received for male and female, Females are relatively more.

HYPOTHESIS 5

By the end of 2021, heavily populated states like California, Texas, New Mexico, and Georgia are likely to have higher total COVID-19 case counts and death tolls compared to states with smaller populations. It is observed that densely populated areas may have a greater share of COVID-19

cases and fatalities, which could exceed 50% of the nation's overall figures. This projection is based on the historical trends that demonstrate a correlation between higher population densities and increased virus transmission rates.

- The data visualization clearly presents the connection between the population size of a state and the severity of COVID-19's impact. States such as California, Texas, New Mexico, and Georgia, with their larger populations, have experienced significantly higher total case counts and death tolls from the virus compared to less populated states. This observation is consistent with historical trends that suggest areas with higher population densities are more vulnerable to virus transmission. As a result, by the end of 2021, these densely populated states collectively accounted for a significant portion of the nation's overall COVID-19 cases and fatalities, possibly exceeding 50% throughout the pandemic.

REFERENCES

- https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Age-and-Sex-Trends-in-the-Uni/5i5k-6cmh/about_data
- https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/vaccinations/us_state_vaccinations.csv