

Final Project

Claire Liu

Quarto

Abstract (150-200 words)

- **Purpose:** The abstract provides a concise summary of your project, including its objectives, key findings, and significance. Write this section last, after completing all other sections, to accurately reflect your project's focus and main results.
- **Guidelines:** Limit this section to 150-200 words. Briefly outline the purpose of your study, the approach you used, and the primary results and conclusions. The abstract should be clear, succinct, and give readers an immediate understanding of what your project entails.

The COVID-19 pandemic has had a profound impact globally, and its effects continue to be felt in present time. This study investigates the progression of COVID in the United States, focusing on state-level death rates and the impact of the pandemic's evolving nature from January 2020 to December 2024. The pandemic is divided into three distinct waves based on changes in population growths. Using data from the U.S. Census and CDC, the study evaluates population changes across states, COVID death rates, and case trends during this time. The findings indicate significant variations in death tolls, with larger states like California and Texas showing higher death rates, while smaller states like Vermont and Alaska reported fewer fatalities. Additionally, COVID cases steadily increased over time. Population growth trends suggest a gradual return to normality after 2021, though several outside factors could have impacted this. This study provides valuable insights into the effectiveness of public health measures and highlights the need for continued research to improve future pandemic preparedness in case something like this happens again and to prepare a more efficient response.

Introduction (500-600 words)

- **Purpose:** The introduction sets the stage for your project, presenting the background and rationale for your analysis. Explain why the topic is significant.

- **Guidelines:** Start with a broad overview of the topic, gradually narrowing down to your specific focus. Conclude with a clear statement of your research questions, hypotheses, or objectives. Use 2-3 paragraphs to establish a solid foundation for the rest of the paper.

The COVID-19 pandemic has been one of the most defining global health crises of the 21st century. Starting in late 2019, the virus spread rapidly across the globe, prompting unprecedented responses, including lockdowns, vaccination campaigns, and the unification of the world towards a common goal: to get rid of the virus. After half a decade since its inception, many of the issues stemming from covid can still be felt and linger within our society. We are able to gather a myriad of data concerning different facets of COVID to analyze and learn the evolution of COVID and its effect throughout the years from the beginning to the present time.

In this analysis, I aim to study the progression of COVID in the United States, focusing on statistics such as death rates and state-level performance across different phases of the pandemic. By dividing the pandemic timeline from January 2020 to December 2024 into distinct waves, we will assess how the virulence of the virus and its impact evolved over time based on changes driven by the emergence of new variants, changes in human behavior, and public health measures. Understanding these waves is important in evaluating the effectiveness of responses from the federal level, state level policies, and trends.

There are 3 main objectives to this study: to define the division of the pandemic into waves using data between 2020 and 2024, to compare and interpret state-level death rates for each period, and to evaluate the changing virulence of COVID-19 across these waves. This investigation is not only significant for its insights and reflections but also allow us to learn from the past and prepare for future pandemics. By understanding the successes and failures of different approaches during COVID-19, we can better equip ourselves for similar challenges in the future and be able to tackle them more efficiently.

Methods (600-700 words)

- **Purpose:** This section details the data sources, methods, and analytical techniques you used to conduct your analysis. It should be specific enough that someone else could replicate your study using the same resources and approach.
- **Guidelines:** Describe the dataset(s) you used, including information about data collection (e.g., sources, time frame). Outline your approach for cleaning and analyzing the data, including any statistical or computational methods applied. Clearly explain any assumptions or limitations in your approach.

In this study, I used 2 main datasets: population data from the census and also COVID data such as deaths and COVID cases from the CDC.

I was able to retrieve the population data from 2020 to 2023 by using a file found on the census.gov website: **Annual Estimates of the Resident Population for the United**

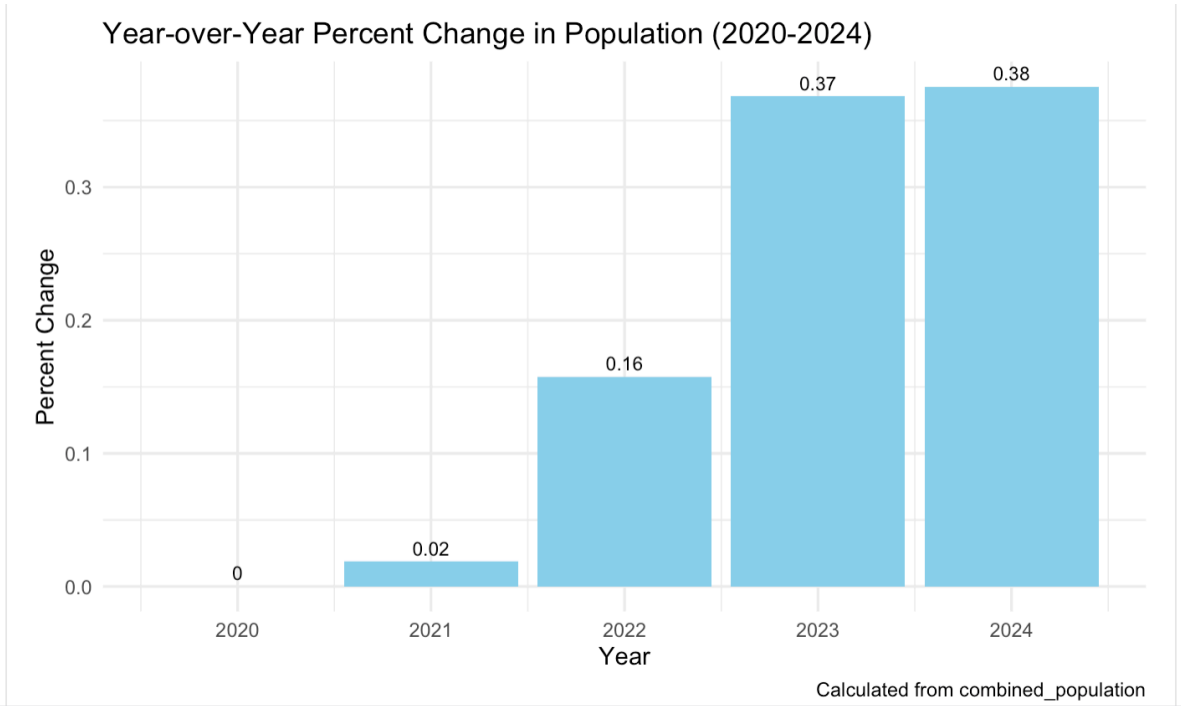
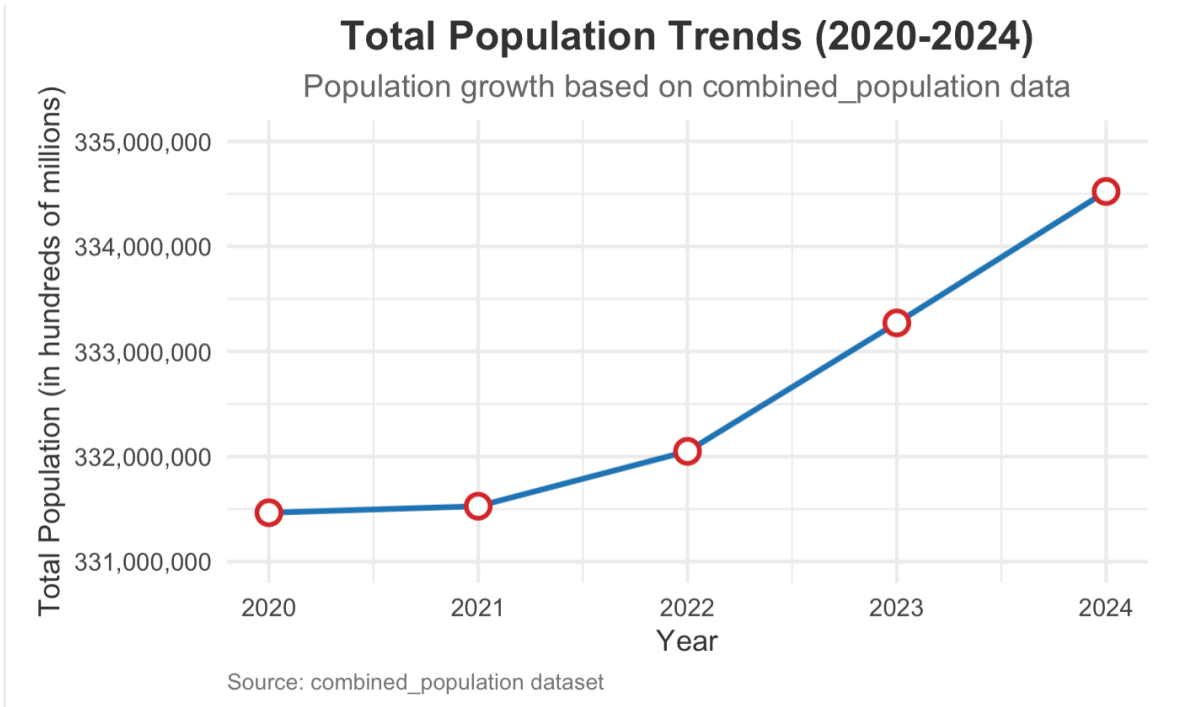
States, Regions, States, District of Columbia, and Puerto Rico: April 1, 2020 to July 1, 2023. Using this file, I was able to create a table which matched each state to their population in 2020, 2021, 2022, and 2023. Originally the data also included other metrics like total population for United States, the regions like West, East, etc., and Puerto Rico which I excluded from the data to simplify it. Since the file doesn't include population data for 2024, I had to estimate it. I figured that the best estimate would be based on the population growth rate from 2022 to 2023 and using this to estimate the growth rate from 2023 to 2024 since it would be consistent with recent data. To get the growth rate, I used this formula: $(2023 \text{ population} - 2022 \text{ population}) / 2022 \text{ population}$. Using this growth rate, I incorporated it to the 2023 population to get the population for 2024 by $\text{round}(2023 \text{ population} * (1 + \text{growth_rate_2023}))$. I combined the data into one table with the states and their populations for 2020 to 2024. With this table, I created 2 graphs: one depicting the total population trend and another showing the year over year percentage change in population. The one thing to make sure of is that the change for 2020 will be NA since there is no 2019 data to calculate it, so I just put a placeholder of 0 for 2020. Based on the percentage change graph, I noticed that there are deviations between 2021, 2022, while 2023 is the same as 2024. So I interpreted this as 3 waves.

To get the deaths and COVID cases from the CDC, I used an api to retrieve the data from the cdc.gov database. For deaths, I wanted to compare the states based on the 3 waves, so I made 3 separate tables for each wave. The death data is given by a table where each row has the state, date, and number of COVID deaths. For each table, I filtered by the year and grouped by states and summed the COVID deaths to get a table that matches each state to their COVID deaths. For COVID cases, I did something similar. The cases data is given by a table where each row has the state, date, and number of cases. I created a single table that groups the wave by the year and sums the total number of COVID cases. Using this, I can distinctly see the data for each wave and see how they have changed over time.

Results (500-600 words)

- **Purpose:** The results section presents the main findings of your analysis without interpretation. Organize the data logically to highlight key insights, using tables, figures, and charts to illustrate trends and comparisons.
- **Guidelines:** For each result, briefly describe it and refer to relevant visuals or tables where appropriate. Do not provide explanations or discuss implications in this section; focus only on presenting the findings clearly and accurately.

The analysis of COVID-19 deaths and cases across different U.S. states and years reveal important trends in the population and the impact of the pandemic over time. The pandemic period was divided into three waves based on the changes in the population growth rates.



Over the past four years, the plot shows a gradual increase, with significant changes occurring post-2021. From 2020 to 2021, the population grew by only 0.02%. However, the population

growth rate significantly increased from 2021 to 2022, rising by 0.16%. A more substantial rise was seen from 2022 to 2023, where the population grew by 0.37%, followed by a similar growth rate of 0.38% from 2023 to 2024. These year-over-year growth patterns suggest a noticeable shift in population trends, especially after 2021, which is why the pandemic period was divided into three distinct waves based on population growth.

state <chr>	total_deaths <dbl>
California	102123
Texas	101211
New York City	66930
Florida	65751
Pennsylvania	55603
New Jersey	54665
Illinois	50374
New York	48877
Ohio	45611
Michigan	37174
state <chr>	total_deaths <dbl>
North Dakota	4452
Montana	3760
Delaware	3204
District of Columbia	2935
New Hampshire	2427
Wyoming	1306
Maine	1282
Hawaii	1001
Alaska	678
Vermont	388

In 2020, California and Texas top the most COVID deaths with around 100,000 deaths, while Alaska and Vermont are the lowest with less than 1,000.

state <chr>	total_deaths <dbl>
Texas	145857
California	143703
Florida	116404
Ohio	61416
Pennsylvania	61390
Georgia	51755
North Carolina	45512
Michigan	44769
New York	44175
Arizona	41964

state <chr>	total_deaths <dbl>
Delaware	3717
Rhode Island	3537
New Hampshire	3496
South Dakota	2848
Wyoming	2775
North Dakota	2691
Alaska	2400
Hawaii	2066
District of Columbia	1994
Vermont	752

In 2021, the results are pretty similar with Texas and California and Florida close behind for the most deaths. Vermont still remains the lowest death.

state <chr>	total_deaths <dbl>
California	88402
Texas	70290
Florida	67514
Pennsylvania	46096
Ohio	45795
New York	36989
North Carolina	33819
Illinois	33153
Michigan	32052
Tennessee	28214

state <chr>	total_deaths <dbl>
Delaware	3161
Rhode Island	2673
Montana	2541
Hawaii	2361
South Dakota	2198
North Dakota	1777
District of Columbia	1400
Vermont	1356
Wyoming	1114
Alaska	953

In 2022 and 2023 combined, California goes back to the top with the most deaths. Alaska also goes back to the lowest.

period <chr>	total_cases <dbl>	avg_daily_cases <dbl>
2020	19802808	6600.936
2021	33816455	10838.607
2022-2023	51038265	11980.813

Looking at the cases table, the total number of COVID cases over the years showed a clear upward trend. In 2020, there were around 20 million reported cases, and this number increased significantly each year. In 2021, the number of cases rose to approximately 33 million, and in 2022-2023, the cases further increased to 51 million. The majority of these cases occurred in 2022, with 47 million cases reported.

Discussion (600-700 words)

- **Purpose:** In the discussion, interpret the significance of your findings, explore potential implications, and relate the results back to your initial research questions or hypotheses. This section allows you to discuss any patterns, unexpected findings, or limitations and suggest possible future research.
- **Guidelines:** Analyze your results in the context of your research question, linking them back to the background information from the introduction. Consider what your findings reveal, any limitations they may have, and how they might impact future work or policy. End with a brief conclusion summarizing your main insights.

The analysis of COVID deaths and cases over the course of the pandemic provides valuable insights into how different states in the U.S. have been impacted by the virus and shows a pattern reflecting the US. The effects of COVID are very nuanced. There are so many different factors and outside influences that contribute to our understanding of COVID. By examining the trends in deaths and cases, as well as changes in population growth during the pandemic, this study aims to understand differences across states, changes over time, and the potential factors contributing to these patterns.

The analysis of population growth tells an interesting picture. From 2020 to 2021, population growth was consistent, increasing by only 0.02%. However, from 2021 to 2022, the growth rate increased to 0.16%, and it continued to rise by 0.37% from 2022 to 2023. There can be a lot of different factors that influence this and the data doesn't necessarily tell us everything that can cause this. There can be behavioral, societal, political, or other reasons that contribute to the data which are not represented. The simplest explanation would be higher deaths and lower child births than normal but it doesn't account for something like migration. It would be more beneficial to have data surrounding these other factors and find how much of an impact they

had. But on the surface, it does indicate that the situation surrounding COVID got better gradually over time and things started to return to normal.

The analysis of COVID deaths per state don't really say too much besides that most of the deaths came from the start of the pandemic in 2020 and 2021. It isn't a surprise that the most deaths came from states with the most population and less deaths for smaller states. States with higher population also have more densely populated cities which make transmission of the virus much easier. It would be more helpful to know the proportion of the state population that died as a result from COVID. This would allow us to compare states and regions and see whether their policies are effective or not. We would also want to consider other factors such as healthcare systems available, demographic, socioeconomic factors, or others to discern the role that they played.

The analysis of COVID cases is an interesting one. It shows that the lowest total cases from the waves came from the start of the pandemic and then increased each year. This could be attributed to the evolving nature of the virus with new variants and more time allowing for more transmissions. This shows the challenges that came with trying to tackle COVID and how efforts might not have been enough to slow it down. With precautions like vaccinations and masks, we would expect the transmission to diminish and slow the growth of COVID. While these might not show, it could have stopped COVID from becoming an even bigger problem but is not reflected in this result. There could also be some cases where people with COVID never reported that they had it and it isn't shown in the data. This would be more prevalent for data that occurs near the beginning of the outbreak. People may not know that they have COVID and might attribute it to another disease or they might not have gotten tested to determine if it is COVID. Or maybe some people got COVID multiple times. Similarly to the last part, we want to incorporate more factors in our analysis to figure out what is accurate and contributes to the results of this data.

In conclusion, this analysis highlights the significant regional differences in the impact of COVID across the U.S. Larger states like California and Texas experienced higher death tolls, while smaller states such as Vermont and Alaska had fewer deaths. The steady increase in COVID-19 cases over the years shows the ongoing challenge of managing the pandemic, and the changes in population growth rates reflect the transition from an abnormal time to a more normal one. While the study provides valuable insights, it also highlights the need for more data and studies to fully understand the complex factors driving these trends. Future research should continue to explore the long-term effects of the pandemic and the lessons that can be applied to future public health responses so that we can do better.