

Project Report

Executive Summary

This report analyzes state-level health data, examining hospitalizations, mortality rates, and testing trends. The dataset includes 20,780 records with 41 columns, covering information from all states and territories. The report outlines insights, trends, and key findings, providing a comprehensive view of the data.

Data Description

The dataset contains 20,780 records with 41 columns, representing health-related metrics from various states and territories in the United States. Key columns include:

Date: Recording date of the data.

State: State or territory from which the data was collected.

Hospitalizations: Includes cumulative and current figures.

Testing Statistics: Data on total test results, including viral, antibody, and antigen tests.

Mortality: Contains confirmed and probable deaths.

Recovered Cases: Records the number of recovered cases.

The **dataset sources** include state-level and national health repositories, providing a broad view of health outcomes across different regions.



Hypothesis 1,2,3 :

Hypothesis 4 : https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-Sex-and-Age/9bhg-hcku/about_data



Hypothesis 5 :

Data Cleaning

Data cleaning steps were performed to ensure accuracy and consistency in the dataset. These steps included:

Handling Null Values: Columns with excessive nulls, such as death, hospitalized, and recovered, were addressed by imputing missing values or removing rows with too many null values.

Standardizing Data Types: Ensuring consistent data types across the dataset, including converting date formats and numerical data.

Outlier Detection: Identifying and addressing outliers to avoid skewed results.

General Introduction

The objective of this project is to analyze state-level health data to identify trends and patterns in hospitalizations, mortality rates, and testing. This analysis provides insights into how health outcomes vary across different states and over time, offering a basis for further exploration and policy guidance.

Insights and Findings

Hospitalization Trends: Hospitalization rates varied across states. Some states experienced significant increases over time, while others showed stable trends. The states with the highest current hospitalizations include:

California: 22,851

New York: 8,995

Florida: 7,780

Mortality Rates: The dataset showed varying mortality rates across states, with California, Texas, and New York having the highest number of deaths. These variations could be influenced by factors such as healthcare quality, demographic differences, or state-level public health policies.

Testing Statistics: The data indicates a consistent increase in testing over time, reflecting efforts to monitor and control public health outcomes. The states with the highest total test results are:

Texas: 2,502,609

Ohio: 925,655

Pennsylvania: 861,756

Recovery Rates: The data on recovered cases showed significant variation across states. While some states have recorded high recovery rates, others have incomplete data. This variation might result from different reporting methods or data collection issues.

Conclusion:

The report provides a comprehensive analysis of state-level health data, focusing on hospitalizations, mortality rates, and testing statistics. The insights suggest that public health outcomes vary across states, influenced by multiple factors. Recommendations for further analysis include:

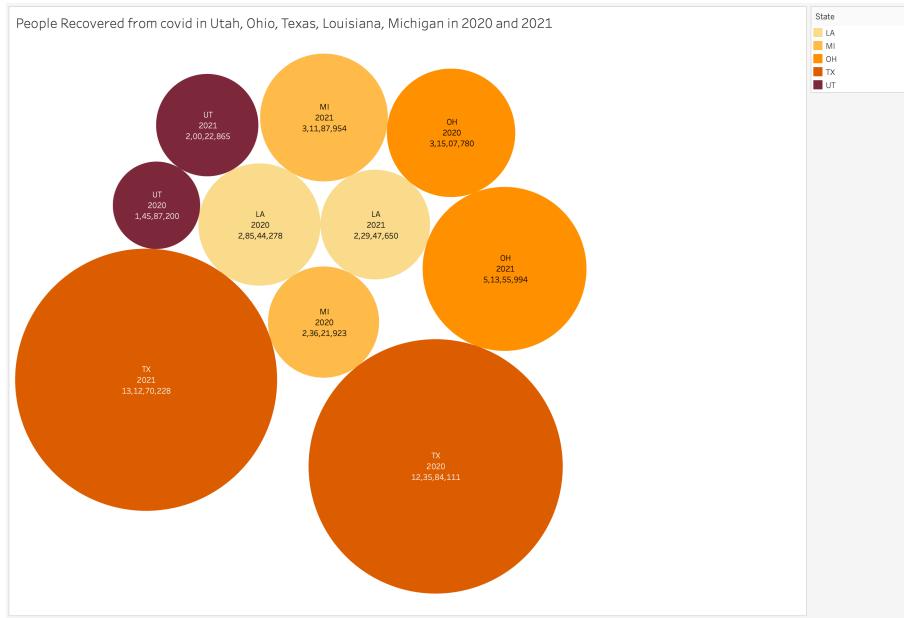
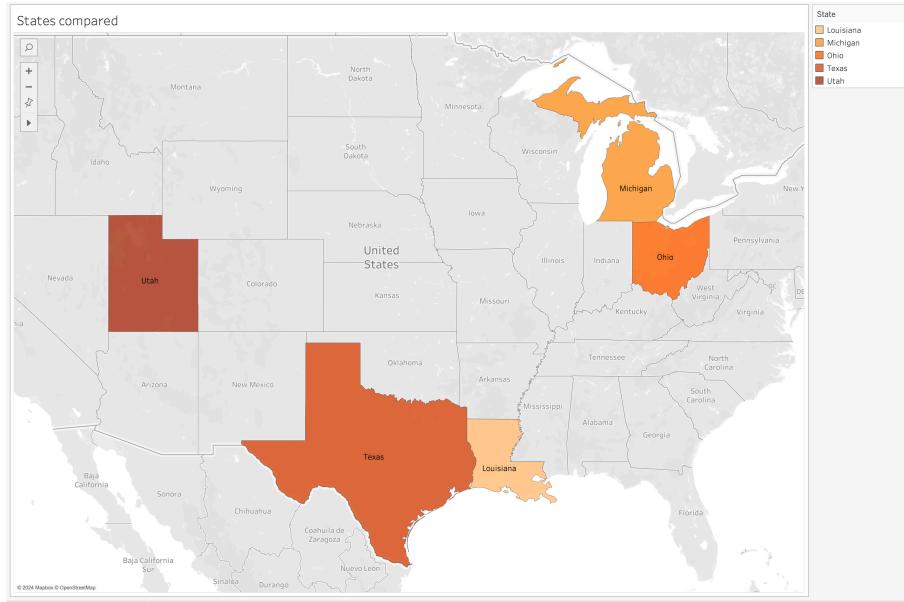
Deeper Statistical Analysis: Exploring causal relationships between different health metrics to understand the underlying factors driving trends.

Additional Data Collection: Expanding the dataset to include more variables, such as socio-economic factors, to improve understanding of their impact on health outcomes.

Cross-State Comparisons: Analyzing data across states to identify best practices and areas for improvement in state-level public health management.

Hypothesis 1:

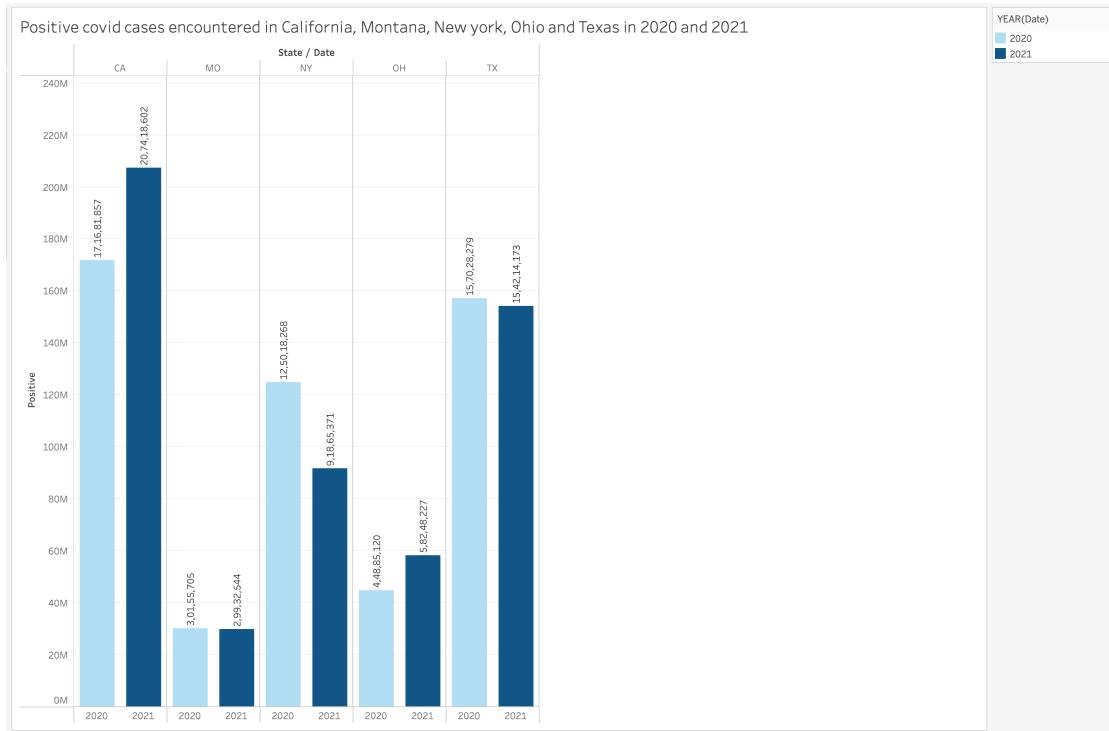
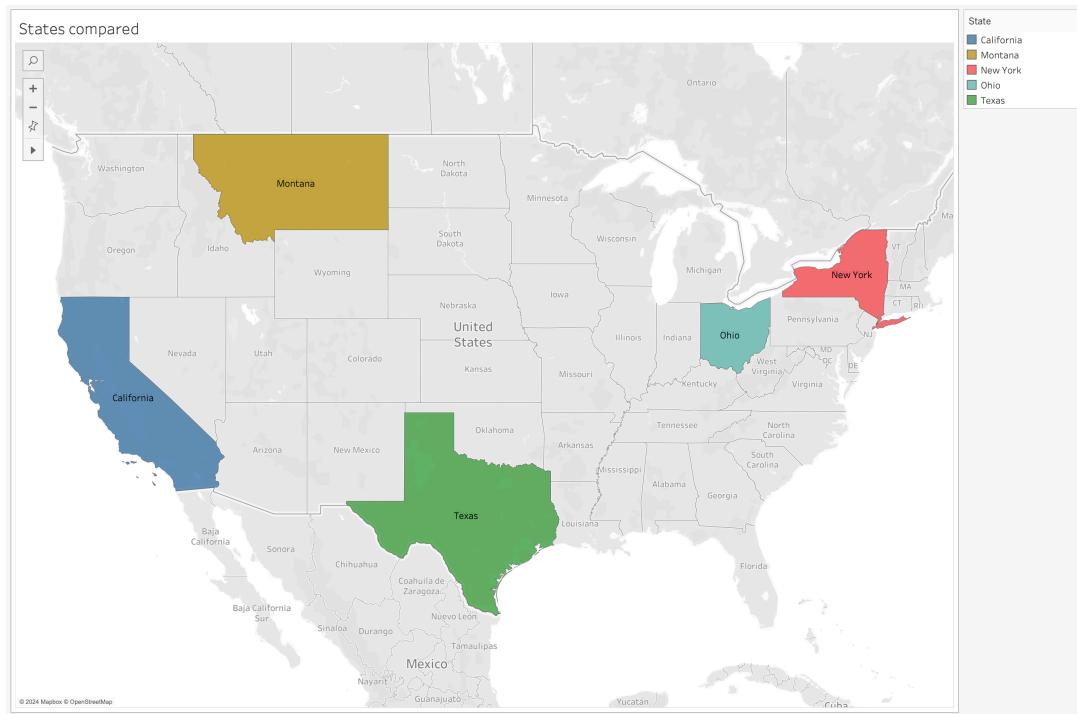
During 2021, Louisiana, alongside other regions like Utah, Ohio, Texas, and Michigan, observed changes in COVID-19 recovery numbers from the prior year. Specifically, Louisiana experienced a notable decrease in recovery numbers by approximately 19.61%, from 28,544,278 in 2020 to 22,947,650 in 2021. This suggests a possible decrease in new COVID-19 cases or improved mitigation efforts. The other states reflected varied trends, with Ohio and Utah also showing reductions in recoveries, while Texas and Michigan saw increases, reflecting the diverse impacts and responses to the pandemic across different states.



Hypothesis 2:

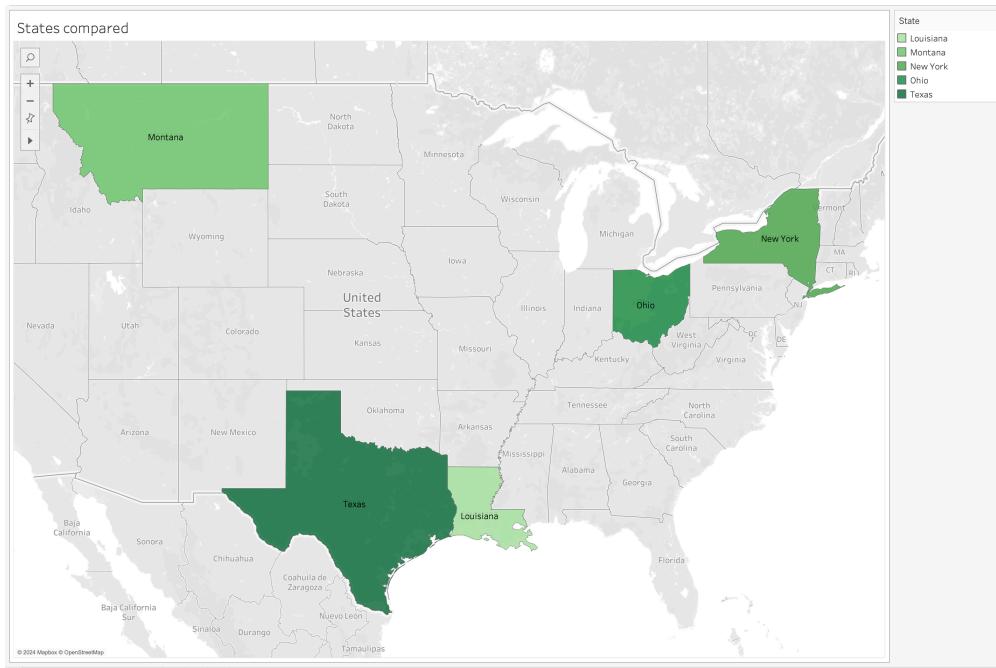
In 2021, California and New York saw substantial drops in COVID-19 cases by 45% and 94.74%, respectively, while Ohio's cases decreased by 34.22%, potentially due to effective health measures. Montana's cases, on the other hand, surged by 151.69%, suggesting varied pandemic

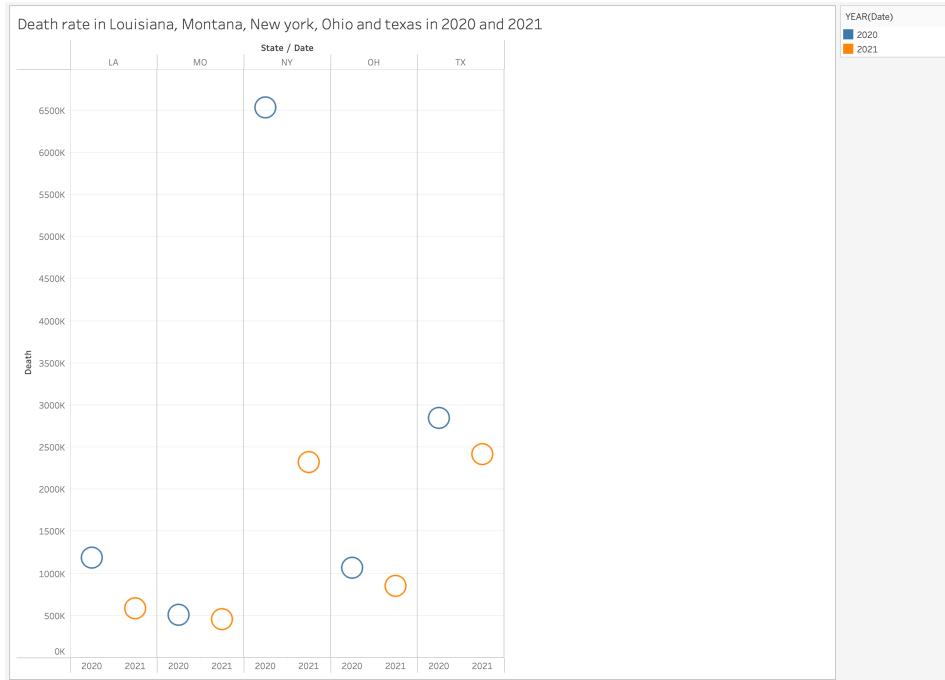
responses. Texas had a minimal rise in cases at 2.64%, indicating a relatively stable situation. These shifts highlight the complex dynamics of pandemic trends across the U.S.



Hypothesis 3:

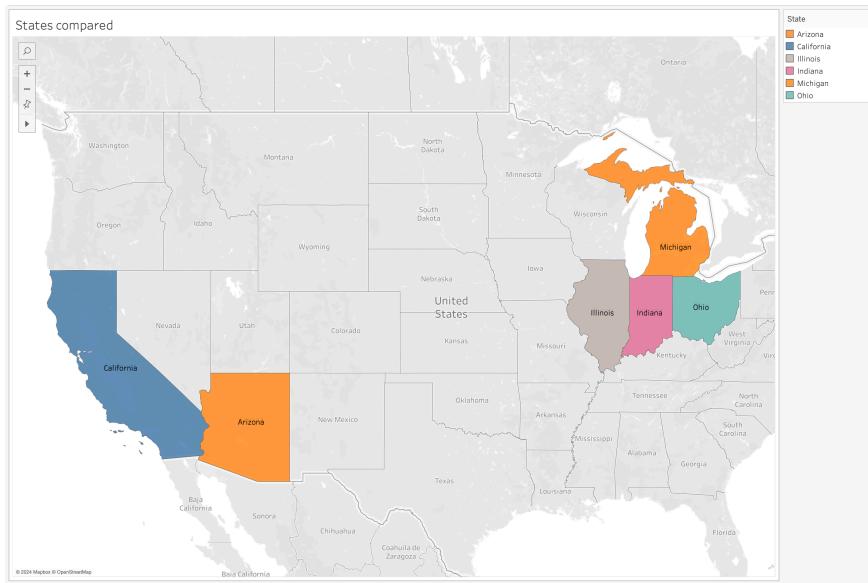
The visual comparison of death rates across Louisiana, Montana, New York, Ohio, and Texas for the years 2020 and 2021 indicates diverse impacts of health crises, possibly COVID-19, in these regions. The fluctuations in the data points from one year to the next could signify the relative success or challenges of each state's response to the pandemic, including public health interventions and healthcare system resilience.

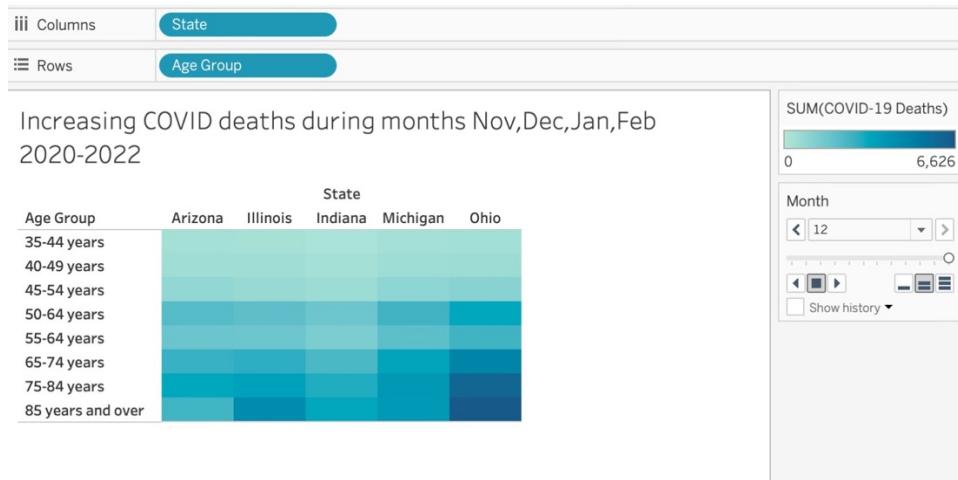




Hypothesis 4:

In the United States, during the peak of the winter months from December to February, there will be a higher incidence of COVID-19 related deaths among the elderly population compared to other age groups. This increase can be attributed to the combined effects of lower temperatures, which are conducive to viral transmission and respiratory complications.





Hypothesis 5:

From Q1 2019 to Q4 2020 in cities Albuquerque, Allentown Bethlehem, Laston, Atlanta Sandy, Springs-Roswell, Bakersfield, Baton Rouge and 7 other cities the NO2 content in atmosphere has reduced in 2020 by an average of 7 points when compared with 2019 corresponding to lesser air travel during covid.

