# Data Visualization: Assignment 3

Sathvik S Rao
*IMT2022082*
*IIIT - Bangalore*
Bangalore , India
Sathvik.Rao@iiitb.ac.in

Shreyas Arun Saggere
*IMT2022006*
*IIIT - Bangalore*
Bangalore , India
Shreyas.Saggere@iiitb.ac.in

Vishruth Vijay
*IMT2022507*
*IIIT - Bangalore*
Bangalore , India
Vishruth.Vijay@iiitb.ac.in

## I. Dataset

### A. Oxford OxCGRT Covid Response Tracker

In the first assignment, the OxCGRT dataset was extensively used to analyze various aspects of the pandemic, including confirmed cases, stringency index, government responses, economic interventions, and health-related policies.

It provided insights into how different governments implemented measures to contain the spread of COVID-19 and their corresponding impacts on the economy, society, and public health.

### B. Google COVID-19 Open Dataset

For the current assignment, the analysis was expanded by incorporating the Google COVID-19 Open Dataset. We felt that by including the Google COVID-19 Open Dataset, it would add both breadth and depth to our study. This dataset provided a wide range of data streams, including demographics, epidemiology, hospitalizations, and vaccinations.

*1) Demographics:* The Demographic data contained information related to the population of each region. It was used to draw correlations between the number of confirmed deaths of a region and the population of the region.

*2) Epidemiology:* The epidemiology data contained information related to the COVID-19 infections for each date-region pair. It was used to analyze confirmed cases over time, identify patterns, and perform clustering to group countries with similar case trends and response indices. This data was central to understanding temporal trends and refining the analysis for meaningful inferences.

*3) Hospitalizations:* The Hospitalizations data contained information related to patients of COVID-19 and hospitals.

*4) Vaccinations:* The Vaccinations data contained information related to deployment and administration of COVID-19 vaccines.

By incorporating data on demographics, hospitalizations, and emergency responses, deeper insights were drawn, validating the hypothesis and expanding the scope of inferences.

## II. Data Analytics Workflow

We have made use of the visual analytics workflow as depicted in the figure 1. Throughout the report, any instance of referring to workflow diagram shall mean the reference to this image. The process involves the following components:

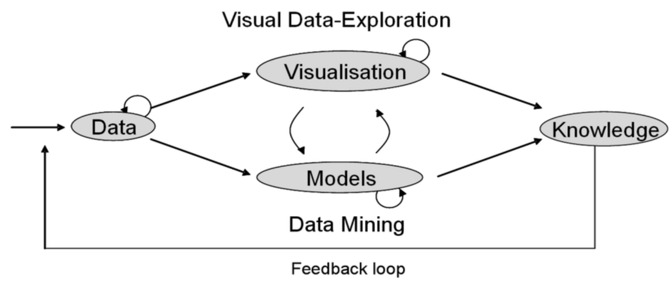


Fig. 1: Figure of Visual Analytics Workflow

1) Data: This step includes transformations made to understand the data better.
2) Visualization: This step involves creating interactive or static visual representations of the dataset to uncover patterns and trends.
3) Models: Machine learning models are applied to further obtain insights into the data. Example: Clustering models (K-means)
4) Knowledge: Observations and inferences drawn from the process including visualizations and models.
5) Feedback Loop: The workflow is iterative. Insights gained at any stage may lead to revisiting earlier steps, such as revising the visualization approach, refining the model, or collecting additional data.

How each stage of this workflow has been incorporated to understand the data and draw meaningful inferences have been elaborated on, in each individual workflow sections.

## III. Member Wise Contributions

This assignment consists of three distinct workflows, each comprising over two iterations. While the workflows were collectively planned, each team member took responsibility for exclusively working on one workflow for ease of splitting work.

- Vishruth worked on workflow 1, the analysis of confirmed COVID-19 cases and government response.
- Sathvik worked on workflow 2, the analysis of COVID-19 vaccinations in the United States.
- Shreyas worked on workflow 3, the analysis of COVID-19 deaths in the United States.

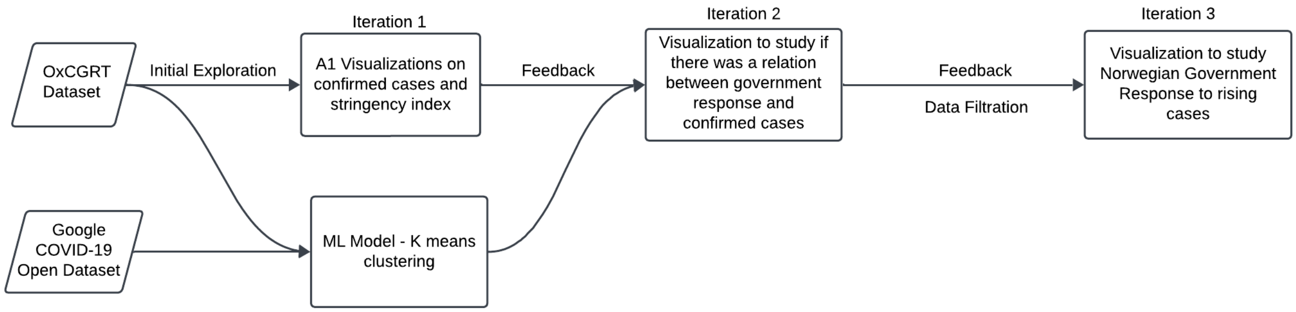


Fig. 2: The figure represents the workflow for studying the relation between government response index and COVID-19 confirmed cases

## IV. Analysis of Confirmed COVID19 Cases and Government Response

### A. Introduction

In the assignment A1, we analyzed the COVID-19 confirmed cases by continent vs time (please refer fig. 3) and the corresponding stringency index of different countries separately. This initial exploration led to understanding which countries were affected more adversely than others. A basic comparison between stringency index and confirmed cases did not lead to any major findings. Initial study lacked insight into how these government responses evolved as confirmed cases increased. Building on these findings, we hypothesized that governments tend to tighten their response measures as confirmed cases rise. To test this hypothesis, we adopted a data analytics workflow that included data transformation, clustering, and visualization, completing one feedback loop. Fig 2 provides an overview of the workflow.

### B. Iteration 2

*1) Formulating the Problem:* From the findings in Assignment 1, we observed that governments implemented varying levels of restrictions, represented by the stringency index. However, the relationship between confirmed cases and the stringency of government responses was unclear. The central question remained: *How do governments react to the growing number of confirmed cases in their countries?*

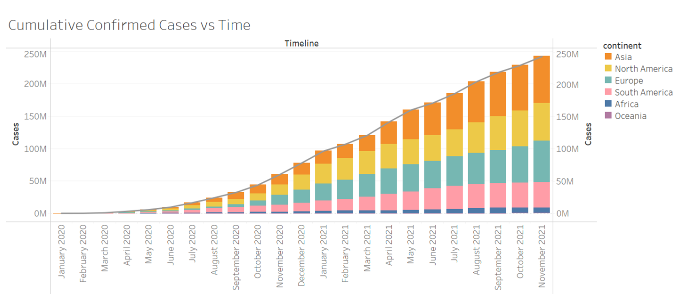


Fig. 3: Chart showing confirmed cases vs time across different continents (taken from A1).

*2) Data Transformation:* To investigate this, we integrated additional data from the Google COVID-19 Open Data repository, specifically the epidemiology dataset. This dataset included detailed records of confirmed cases. Data used to study the government response index was a subset of the OxCGRT COVID-19 dataset that we previously used for the A1 assignment. We combined country-level confirmed cases with their corresponding government response index for each month. A sample of this was studied, specifically June 2020 and December 2020.

*3) Model : Clustering Analysis:* Using the transformed data, I applied K-Means clustering to group countries based on their confirmed cases and government response index. The Elbow Method was employed to determine the optimal number of clusters, which was identified as 4 (please refer fig. 4).

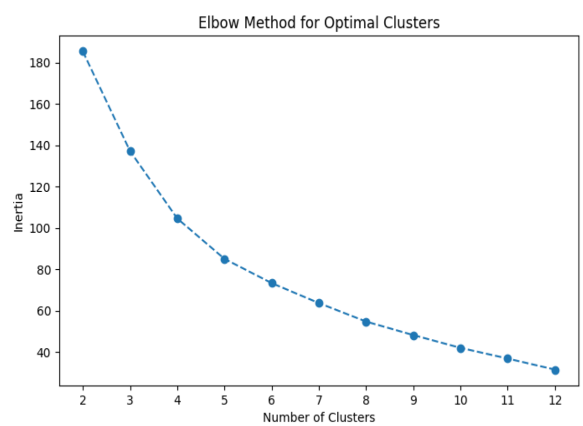


Fig. 4: Figure shows the elbow curve obtained for finding the optimal number of clusters

The scatter plots (fig. 5 and fig. 6) illustrate the clustering of countries based on their confirmed COVID-19 cases - Log Transformed (y-axis) and government response index (x-axis) for June 2020 and December 2020. The countries are divided into four clusters, each represented by a distinct color.

*a) Cluster 1: Low Cases, Low Government Response Index (Green):* These countries have a low number of confirmed

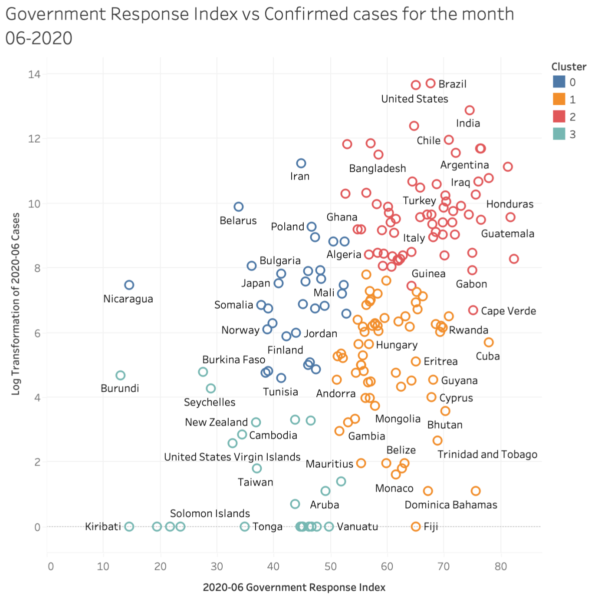


Fig. 5: Figure shows the scatter plot for government response index vs confirmed cases for the month of June, 2020.

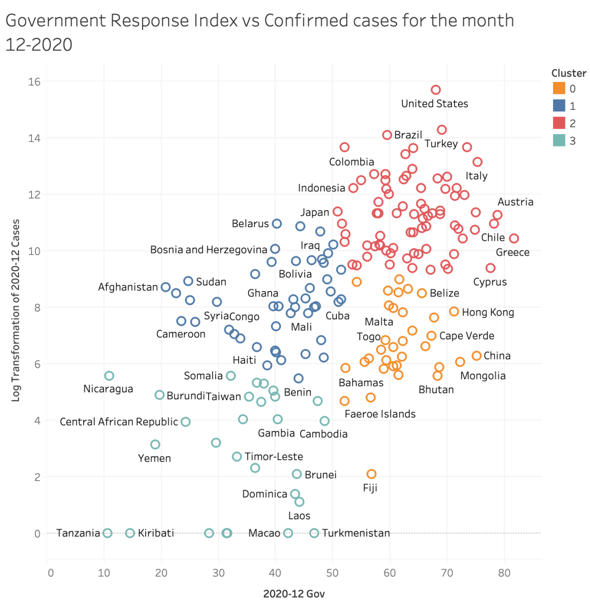


Fig. 6: Figure shows the scatter plot for government response index vs confirmed cases for the month of December, 2020.

cases. Correspondingly, their government response index is also low. Countries in this cluster may have experienced minimal impact from the pandemic during this period or implemented limited interventions.Their relaxed government measures reflect their relatively mild situation.

Examples: Burundi, Kiribati, Tonga, Solomon Islands.

*b) Cluster 2: Moderate Cases, Low Government Response Index (Blue):* This group has moderate confirmed cases. The government response index is in the range of 30 to 50. These nations likely adopted a balanced approach, where interventions were proportional to the rising case numbers.

Examples: Poland, Belarus, Norway, Somalia.

*c) Cluster 3: Moderate Cases, Moderate Government Response Index (Orange):* Governments in this group have adopted strict measures to control the rising number of cases. The cluster suggests a reactive pattern: higher cases prompted governments to tighten policies.

Examples: Hungary, Cyprus, Algeria, Rwanda.

*d) Cluster 4: Very High Cases, Very High Government Response Index (Red):* These countries exhibit the highest government response indices, and very high confirmed cases. Governments in this cluster implemented the strictest measures to combat the pandemic. These countries were among the hardest-hit during this period, necessitating aggressive government interventions. Despite stringent measures, their high case counts indicate the challenges of containing the virus in densely populated or highly connected regions, further investigated in the next iteration of the workflow.

Examples: Brazil, India, Chile, Bangladesh.

Please note that the clusters 1-4 as described in this section may not correspond to 0-3 in the images.

*4) Observations and Inferences:* The following are few of the observations made by studying the scatter plot:

- Across all clusters, a clear positive correlation is evident between confirmed cases and the government response index, supporting the hypothesis that stricter responses are triggered by a rise in cases.
- The clustering reveals that tailored government interventions aligned with case severity are a common approach. However, the effectiveness of such measures depends on factors like population density, healthcare infrastructure, and public compliance.

Norway's transition from 439 cases and a government response index of 38.85 in June 2020 to 13417 cases and an index of 53.42 in December 2020 demonstrates an escalation in both confirmed cases and governmental strictness over the latter half of the year. This can be sen evidently in the figures 7 and 8 respectively.

The rise in the response index aligns with the hypothesis that governments react proportionally to the severity of the pandemic to protect public health.

### *C. Iteration 3: Case Study - Norway*

*1) Introduction:* In the earlier phase of analysis, Norway's confirmed COVID-19 cases and government response index were studied. A pattern was observed where a rise in confirmed cases corresponded to an increase in the government response index, indicating stricter measures to mitigate the spread of the virus.

This raised a specific question for further exploration: *How do the peaks in confirmed cases correspond to changes in government response index across time?*

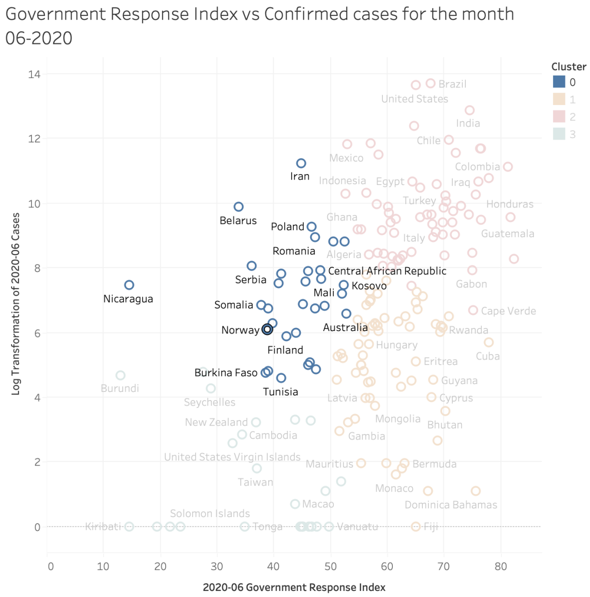


Fig. 7: Figure highlights Norway in the scatter plot for government response index vs confirmed cases for the month of June, 2020.

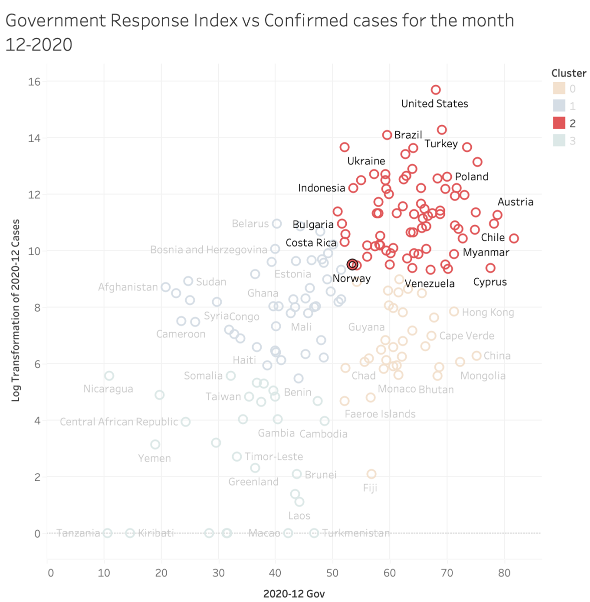


Fig. 8: Figure highlights Norway in the scatter plot for government response index vs confirmed cases for the month of December, 2020.

2) *Data Transformation:* Data was filtered to focus specifically on Norway, as it exhibited a clear trend of rising confirmed cases and an increasing stringency index over time. While not all countries necessarily follow this exact trend, the general pattern observed—where higher confirmed cases are associated with a stricter government response—holds for the majority of countries. This trend provides a basis for generalization, though exceptions may exist.

3) *Hypothesis:* It was hypothesized that:

- Peaks in COVID-19 cases would align with peaks in the government response index.
- This alignment supports the idea that governments implement stricter policies in response to rising cases, reflecting their attempt to control the spread.

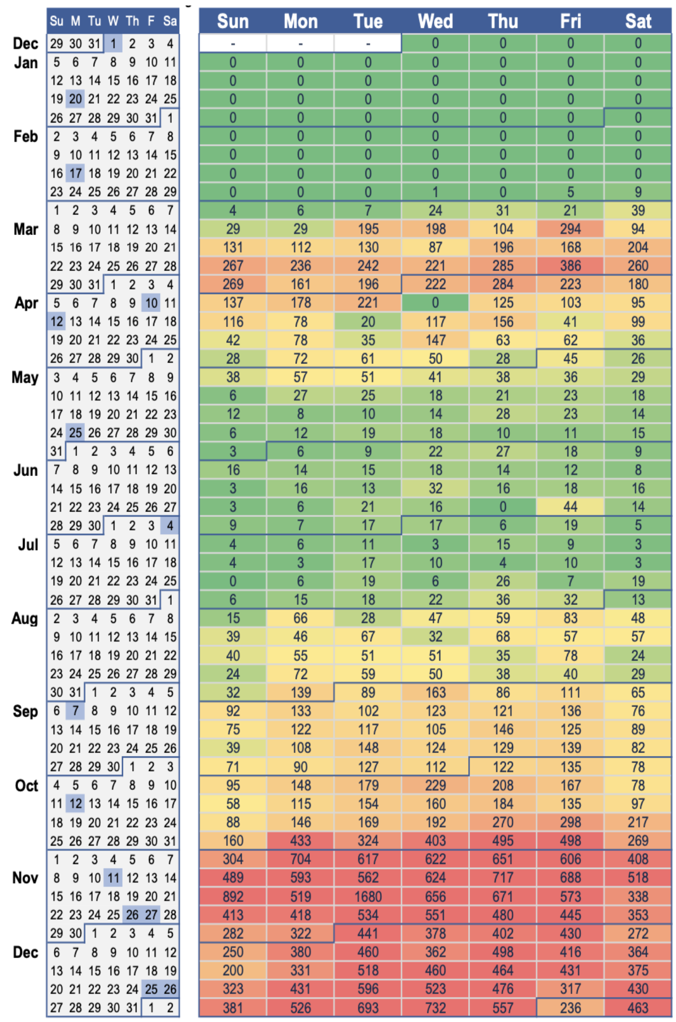


Fig. 9: Figure shows the calendar chart for the number of confirmed cases in Norway for the year 2020.

4) *Visualization:* The calendar chart (fig. 9) visualizes confirmed COVID-19 cases in Norway across the year, with colors indicating the intensity of cases. The green color in the chart signifies lower cases per day, yellow and orange signify moderate number of cases per day, and red signifies high number of cases per day. The calendar chart was plotted using a tool by Vertex42 [1], referenced in one of the lectures.

Notably, there are two distinct peaks observed in the chart: one during March/April and another in November/December. These red zones signify a significant rise in cases during these periods, highlighting two waves of infections.

Alongside the calendar chart, a line chart plotting weekly confirmed cases against months provides further clarity (fig.

10). The chart aligns with the calendar visualization, confirming the same two peaks in March/April and November/December.

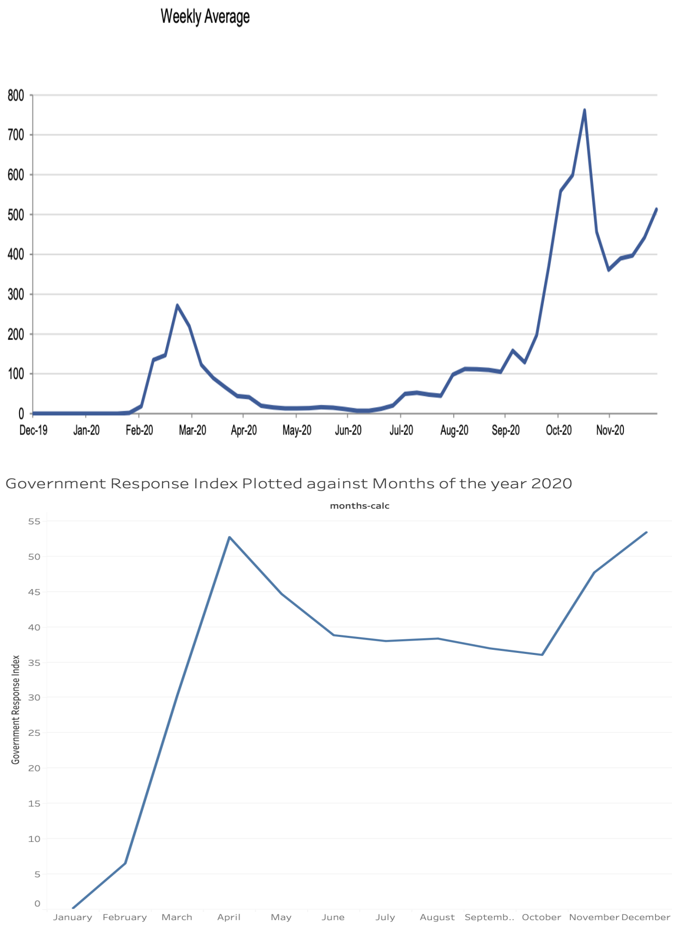


Fig. 10: The figure shows two images stacked vertically: the line chart for confirmed cases in Norway for the year 2020 and the government response index of Norway over the year 2020.

A separate line chart (fig. 10) depicts the monthly government response stringency index plotted against months. The chart reveals that the government response peaked concurrently with the surges in confirmed cases. During both March/April and November/December, the stringency index sharply increased, reflecting stricter measures implemented by the government to combat rising infections.

*5) Observations and Inferences:* The visualizations collectively support the hypothesis that as COVID-19 cases increased, the government responded with stricter measures. Both waves of infections, evident in the calendar chart and weekly cases line chart, correlate strongly with the peaks in the government response index. This indicates a reactive approach where government interventions were heightened in response to rising case counts.

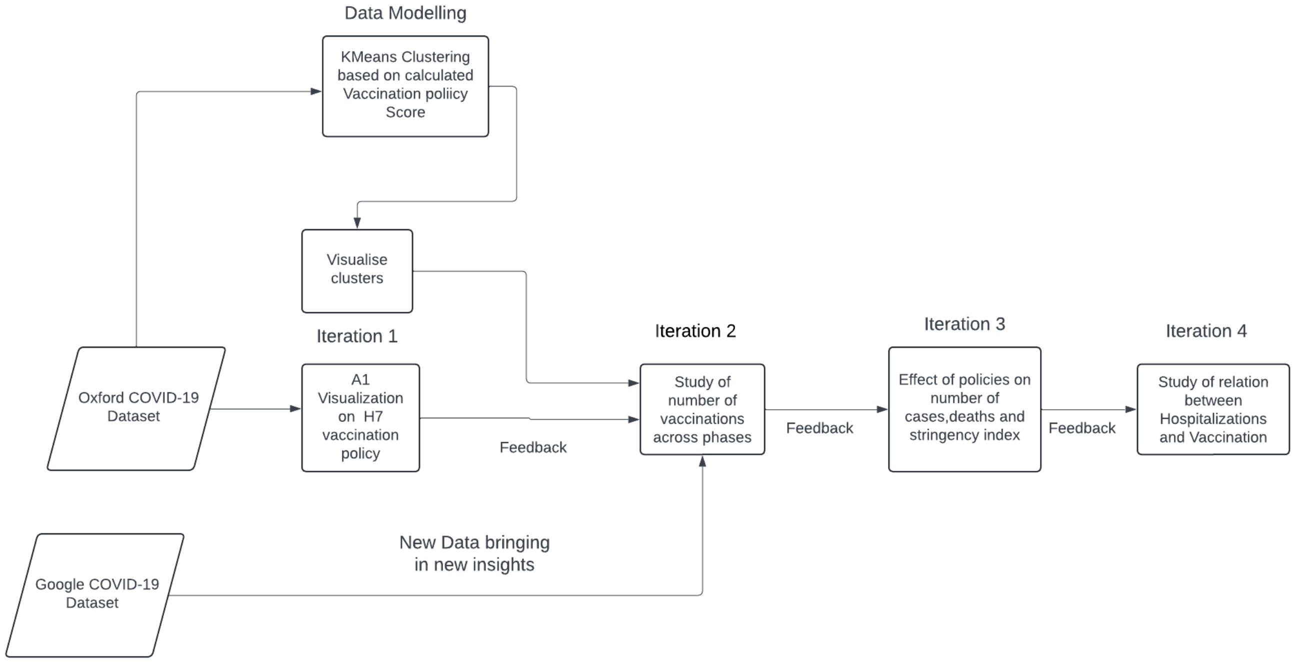


Fig. 11: The figure represents the visual analytics workflow for studying COVID-19 vaccinations in the United States

## V. Analysis of COVID-19 vaccinations in USA

### A. Introduction

The COVID-19 vaccination campaign in the United States began on December 14, 2020, which was one of the critical steps in fighting COVID-19. Vaccinations were initially given to the vulnerable groups and then were rolled out for the public. Pfizer, Moderna and Johnson & Johnson were the vaccines that were approved by the Food and Drug Administration (FDA). Our aim was to develop a visual analytics workflow in order to gain new insights into the strategy the US government and FDA undertook to distribute vaccines to the citizens.

### B. Data Handling

The datasets used for the above task were Oxford COVID-19 Government Response dataset (allotted in A1) [4], [5] and the Google COVID-19 Dataset [2]. The vaccination policy scores, Stringency index were calculated and obtained from the former and the data for number of vaccinations, population, number of hospitalizations, number of ICU patients were obtained from the latter. The datasets were joined based on Region codes and dates. The tools used for visualizing and data cleaning are Python [3] and Tableau [7].

### C. Workflow

*1) Iteration 1:* Figure 12 is the visualization borrowed from Assignment 1. It compares how different countries staged their vaccination drives. More specifically, we focus on how the United States conducted its vaccination drive.

**Inferences**:

Figure 12 illustrates the evolution of vaccination policies in India, China, and the United States from December 2019 to December 2021.The H7 vaccination policies (0-5) are mentioned in Table III. The plot reveals that all three countries started with no vaccination policy (level 0) until late 2020, which aligns with the development and approval timelines of COVID-19 vaccines. China appears to have implemented its vaccination policy earliest, around November-December 2020, followed by India and the United States. The United States was one of the quickest to jump to phase 5 of the vaccination policy. Unlike India, the United States did not follow the vaccine distribution in a staged, prioritized manner.

*2) Data Modelling:* Figure 13 was derived from K means clustering of countries based on the calculated mean vaccination score from January 2021 to December 2021. The K value was chosen to be 5, by using the elbow method as seen in figure The vaccination score for each country is calculated based on the following groups, with each group assigned a weight. The weights reflect the priority given to different demographic and occupational groups. The weights chosen were higher for higher risk groups and lower for lower risk groups.

The groups and their corresponding weights are given in Table I and Table II. The vaccination score for each country is calculated using V1_Score, V2_Score, V3_Score

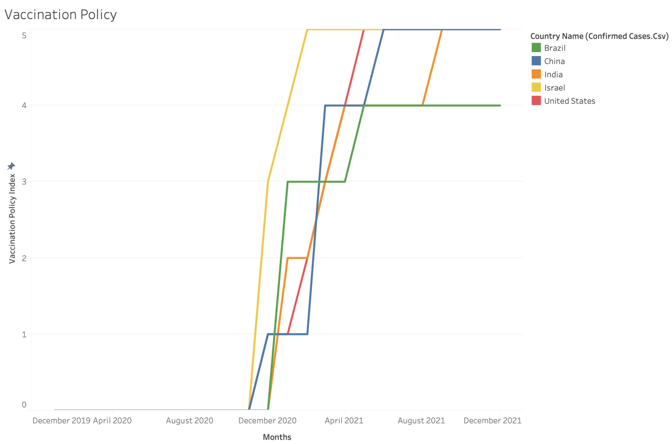


Fig. 12: H7 Vaccination Policy index of the USA, China, India, Brazil and Israel over time. Borrowed from A1

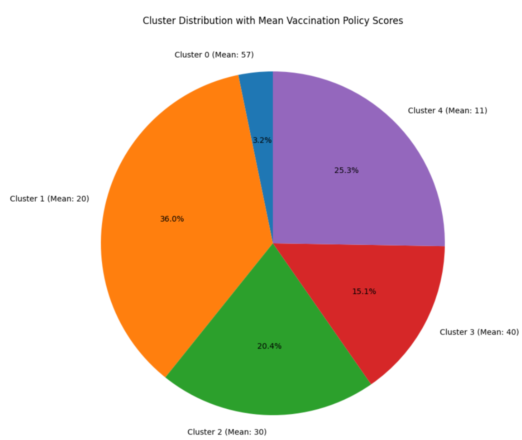


Fig. 13: Clustering of countries based on mean vaccination score

$$\text{V1\_Score} = \sum_i (\text{V1}_i \times \text{Weight}_i)$$

$$\text{V2\_Score} = \sum_i (\text{V2}_i \times \text{Weight}_i)$$

$$\text{V3\_Score} = \sum_i (\text{V3}_i \times \text{Weight}_i)$$

$$\text{Vaccination\_Score} = \frac{(\text{V1\_Score} + \text{V2\_Score} + \text{V3\_Score})}{3}$$

Here $V1_i$ represents the V1 index for the $i^{th}$ category as mentioned in Table I.

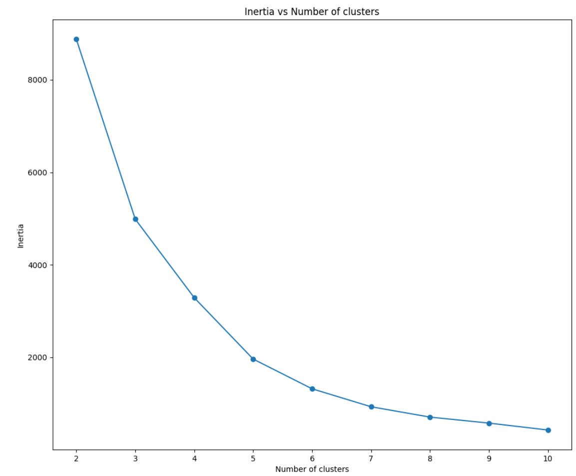


Fig. 14: Inertia(Sum Squared Error) vs Number of clusters plotted for determining the suitable number of clusters

**Inferences**

Given below are some of the countries in each cluster.

- Cluster 0: Consists of 6 countries. Belgium, Togo and Malta belong to this cluster.
- Cluster 1: Consists of 67 countries. India, Japan and Russia are some of the countries in this cluster.
- Cluster 2: Consists of 38 countries. UK, Turkey, Sweden made it to this cluster.
- Cluster 3: Consists of 28 countries. New Zealand, China, Slovenia made it to this cluster.
- Cluster 4: Consists of 47 countries. USA,UAE and Nigeria are some of the countries in this cluster.

The United States falls within a cluster of countries with an average vaccination score of 11, which is on the lower end of the spectrum. This could be attributed to the country's decentralized approach, stemming from its federal system of government. The variability in vaccine prioritization and distribution across different states led to inconsistencies, significantly lowering the U.S.'s V1 Prioritization score and ultimately contributing to a lower overall vaccination score.

In contrast, nations like the UK and India adopted a more centralized vaccine distribution system, which was more effective. Additionally, the swift move across H7 Vaccination policy is also one of the reasons for a low prioritization score.

*3) Iteration 2:* In Iteration 2 we visualize, the number of vaccinated people across different phases of the H7 Vaccination policy in the United States. The time period chosen was from January 2021 to December 2021. The knowledge from the previous iteration and the data modelling gives us insights into the global performance of the United States in regard to

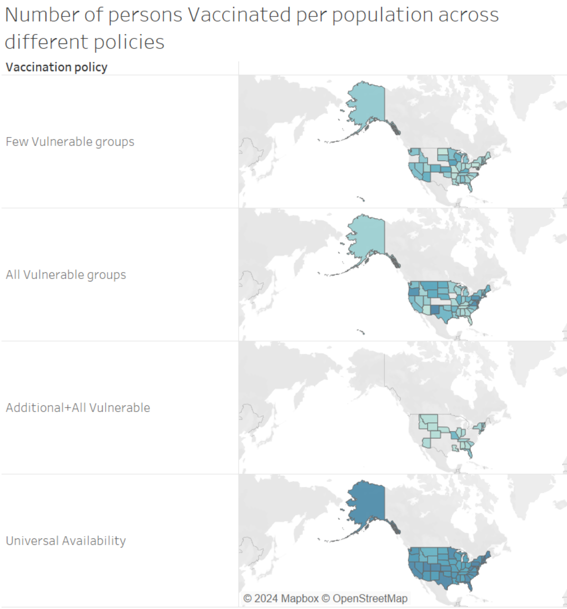


Fig. 15: Vaccinations per 100 population shown across different phases in the United States

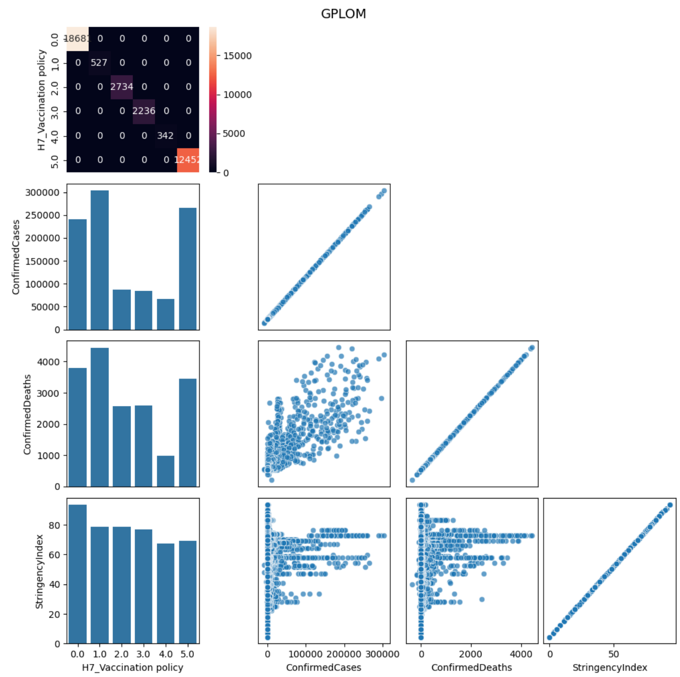


Fig. 16: GPLOM using the variables, H7 Vaccination Policy, Confirmed Cases, Confirmed Deaths and Stringency Index in the United States

the vaccination policies. Figure 15 is the visualization used for this iteration.

**Inferences**

Although the U.S. had a low vaccination score because of its lower V1 Prioritization score, the government did a good job of distributing vaccines throughout the country. As shown in the choropleth in Figure 15, which plots state-by-state vaccination rates per 100 people over various H7 Vaccination Phases, most vaccinations took place during Phase 5, when vaccines became widely available. In fact, as shown in Figure 12, it is clear that almost no states adopted a phased approach. Phase 2, which targeted the few vulnerable groups, was not adopted by any state. This means that while the U.S. was great at making vaccines widely available, the phased prioritization scheme was not fully followed by the states.

*4) Iteration 3:* An important question now is how the number of vaccinations affected the trends in COVID-19 deaths and cases, and whether the government relaxed strict COVID-19 protocols as vaccination rates rose. This analysis examines that relationship by considering the interaction between the number of vaccinations and the H7 vaccination policy established in the previous iteration. The iteration consists of Figure 16 as its visualization. The period reviewed in this visualization is from January 2021 to December 2021.

**Inferences**

From the Generalized Plot Matrix in figure 16, it is easy to see the relation between the number of cases and number of deaths, i.e. as the number of cases rises so does the number of deaths. For the same reason, the number of cases and deaths have a similar chart when plotted against the H7 vaccination policy. The general trend observed is the reduction in number of deaths and number of cases as H7 Vaccination policy increments. But surprisingly, the maximum number of deaths and cases were observed during phase 5,i.e. universal availability. This is likely due to the reduced efficacy on the Delta variant of SARS-COV2 virus [6].

Another observation that can be made is that the stringency index reduced as the H7 Policy incremented. The US government relaxed some of the restrictions after rolling out the vaccines to the public.

*5) Iteration 4:* From the previous iteration, we obtained the knowledge of the relation between the vaccination policy with number of cases and deaths. A further improvement on this study is to introduce the number of hospitalizations across different phases. This would help us understand if the vaccines had any effect on the severity of cases. The iteration consists of Figure 17 as the visualization.

**Inferences**

Figure 17 is a violin plot which shows that the number of hospitalized cases dropped as the H7 Policy increments. We see a similar pattern to the confirmed cases and deaths when plotted against the H7 Policy. Likely due to Delta variant, there was a rise in the number of hospitalization in phase 5.

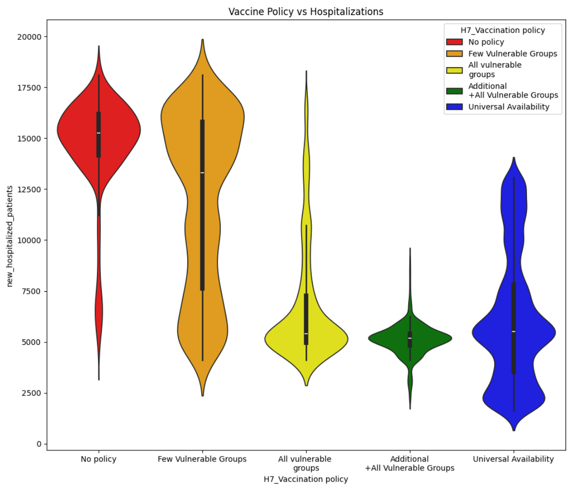


Fig. 17: A violin plot showing the number of hospitalization across different H7 Policies in the United States

*6) Iteration 4 Contd:* The first part of iteration 4, gave us insights into the number of hospitalizations in the US across different phases of H7 policy. Figure 18 shows the normalized daily average of the number of COVID-19 Hospitalizations per 100 population. Using the choropleth, we can identify Kentucky and Oklahoma to have higher hospitalizations, and Alaska, Washington states to have the lowest hospitalizations. The selected states can be seen in figure 19. We use these states and explore how vaccinations per population (fractional) in these states affected the number of deaths per 100 population. This is visualized in figure 20.

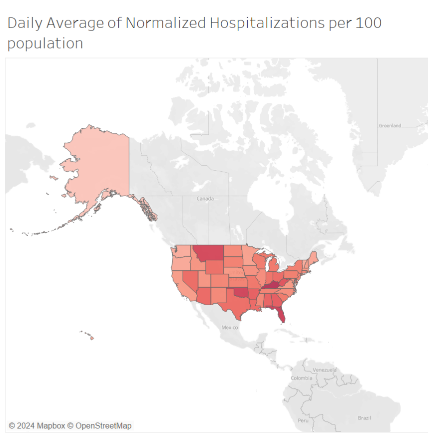


Fig. 18: A choropleth showing the Normalized Number of Hospitalizations per 100 population in different states of the United States

**Inferences**

Figure 20 represents a time series braided chart of 2 variables: Cumulative vaccinations per population(fractional) and Normalized Number of Hospitalizations per 100. The time period studied was from January 2021 to September 2022. We can observe that the number of vaccinations per population across all these states reached a stagnant value of about 0.7-0.8 at the same rate. Although they had the same vaccination rate, Kentucky and Oklahoma had more hospitalizations due to the fact that their population densities are much greater than that of Alaska and Washington.

*D. Conclusion*

The four iterations of the workflow were designed to analyze and compare the global performance of the USA's vaccination drive and its impact on the country.

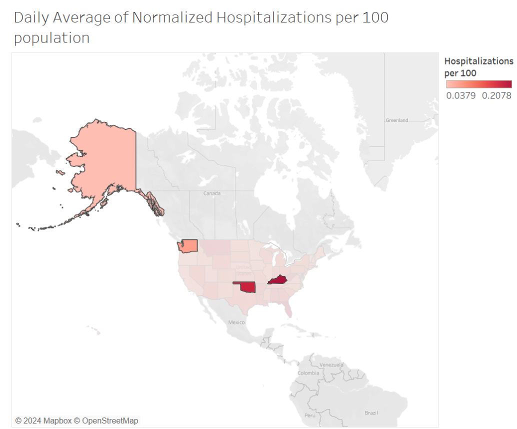


Fig. 19: A choropleth which is similar to figure 17, but focuses on the states: Kentucky, Oklahoma, Washington and Alaska

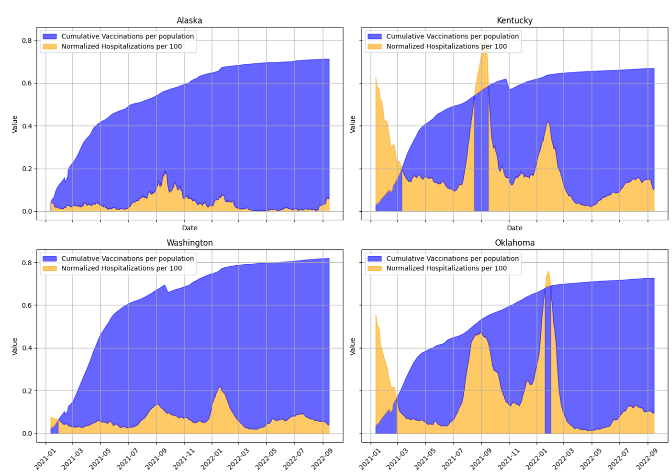


Fig. 20: A Braided chart which represents the time series data of the Normalized number of Hospitalization per 100 population and Vaccination per population in Alaska,Washington,Kentucky and Oklahoma

| Category | Weight |
|---|---|
| Healthcare workers/carers (excluding care home staff) | 5 |
| Frontline/essential workers (when subcategories not specified) | 4 |
| Police/first responders | 4 |
| At Risk 80+ yrs | 5 |
| Clinically vulnerable/chronic illness/significant underlying health condition (excluding elderly and disabled) | 4 |
| Pregnant people | 3 |
| General 80+ yrs | 3 |
| General 70-74 yrs | 2 |
| General 60-64 yrs | 2 |
| General 50-54 yrs | 1 |
| General 40-44 yrs | 1 |
| General 30-34 yrs | 1 |
| General 20-24 yrs | 1 |
| General 16-19 yrs | 1 |

TABLE I: Category and Weight Table

| Indicator | Description |
|---|---|
| V1 – Vaccine prioritization | 1, 2, 3, 4... – category has been selected for prioritization; number represents the rank of prioritization; equal-ranked categories will share the same number |
| V2 – Vaccine eligibility/availability | 0 - vaccines are not being made available to this category<br>1 - vaccines are being made available to this category. |
| V3 – Vaccine financial support | 1 - full cost borne by the individual (or through private health insurance) or no policy<br>2 - partially funded by government and individual pays nominal fee<br>3 - fully covered by government funding, FREE. |

TABLE II: Vaccine Indicators Table

| H7 Vaccination Policy | Description |
|---|---|
| 0 | No availability |
| 1 | Availability for ONE vulnerable group (key workers/clinically vulnerable/elderly) |
| 2 | Availability for few vulnerable groups |
| 3 | Availability for ALL vulnerable groups |
| 4 | Partial additional availability |
| 5 | Universal availability |

TABLE III: H7 Vaccination Policy

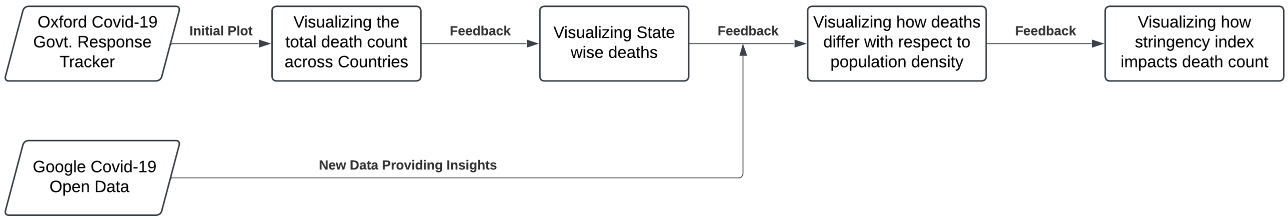


Fig. 21: The figure represents the workflow for studying COVID-19 Deaths focused on the United States

# VI. Analysis of COVID-19 Deaths in the United States

## A. Introduction

The COVID-19 pandemic is associated with a huge global impact that is characterized by the death toll in various regions. This has been associated with the complex interaction of demographic, policy, and socio-economic variables. This workflow tries to explain the factors influencing COVID-19 deaths across the U.S. states with a focus on the month of December, which happens to be a critical period that characterizes widespread transmission with differences in government response. It seeks patterns and correlations in such indicators as population density, per capita mortality rates, and the stringency of imposed government restrictions. The findings aim to provide insights into how these factors interact and offer data-driven perspectives on effective pandemic management strategies.

## B. Data and Methodology

1) **Data Sources:** The workflow utilizes data from two primary sources: The Oxford COVID-19 Government Response Tracker [5] and the Google COVID-19 Open Dataset [2].
2) **Data Processing:** Data transformations were applied differently for each step in the workflow, tailored to the specific requirements of the analysis. Detailed descriptions of these transformations are provided within their respective sections.

## C. Workflow

*1) Iteration 1:* In this iteration, the raw data from the Oxford COVID-19 Government Response Tracker [5] is used to generate a heatmap depicting confirmed deaths across countries. This visualization is adapted from Assignment-1 **??**.

**Inferences**

The heatmap highlights that India, the United States, and Brazil have reported the highest death counts compared to other countries. Subsequent iterations focus on a more detailed analysis of deaths within the United States.

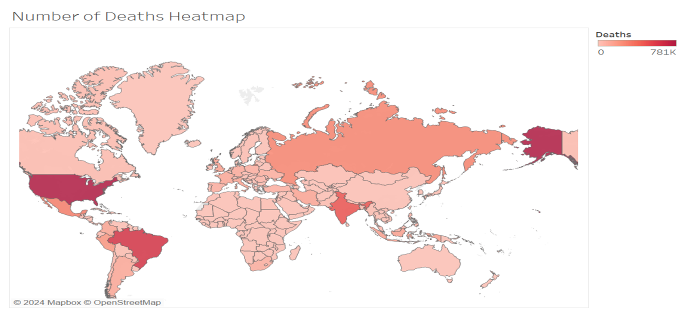


Fig. 22: The figure is a heatmap that shows the number of total deaths across countries.

*2) Iteration 2:* This iteration utilizes state-wise time series data for the United States to create a treemap visualization of total death counts across various states(See Figure 23). The data is aggregated monthly, with the death counts for the last day of each month used to construct the treemap. Darker colors in the treemap indicate higher death counts for a specific state and month.

**Inferences**

The hierarchical treemap visualization provided a detailed perspective on COVID-19 deaths across different regions in the United States, segmented by month. This enabled a deeper understanding of regional variations and seasonal trends. Larger sections in the treemap, representing higher death counts, were notably observed in states with greater population density. This is exemplified by states like New York and California, which recorded some of the highest death counts. The month-wise breakdown further revealed significant spikes corresponding to the pandemic waves observed during 2020-2021.

By combining spatial and temporal insights, a hypothesis can be formulated that the number of death counts were high in regions where the population density was high.

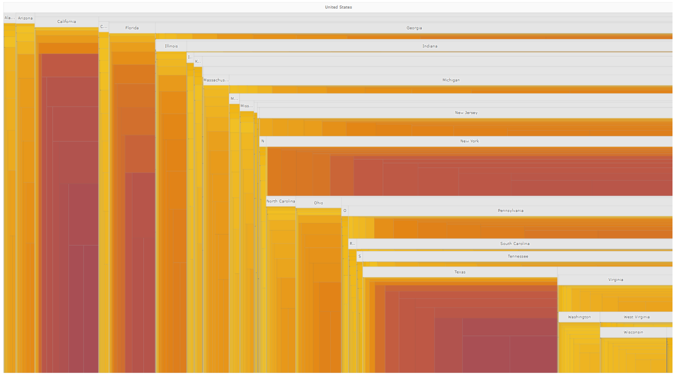


Fig. 23: Treemap that has region wise and month wise deaths for each state in the United States

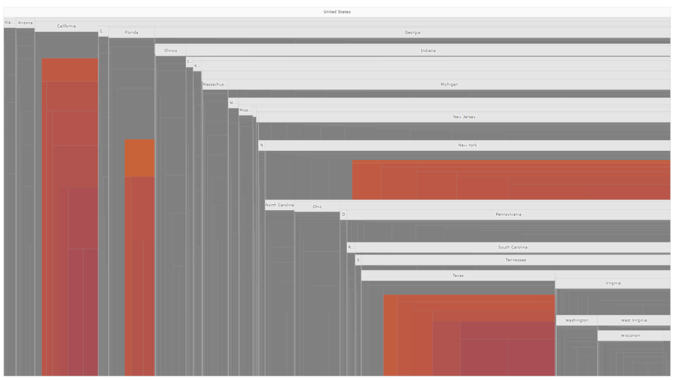


Fig. 24: Treemap that shows region wise and month wise data for each state in the United States filtered for deaths > 50,000

*3) Iteration 3:* This iteration builds upon the hypothesis formulated in VI-C2. To investigate this further, the Oxford COVID-19 Government Response Tracker and the Google COVID-19 Open Data (Demographics) were filtered for the United States and merged based on state names. This integration provided both confirmed death counts and population data, focusing on the month of December.

A new field, population density, was calculated for each state by dividing the population by the state's area in square kilometers.

The resulting visualization(See Figure 25) is a choropleth, where circles are used as glyphs. The size of each circle represents the population density, while the color reflects the death count, using a linear orange color scale.

**Inferences**

At first glance, the hypothesis formulated in VI-C2 appears to hold true, with states such as New York, New Jersey, and California—areas with high population density—showing higher death counts(See Figure 27).

However, a closer examination reveals that states like Maryland, Delaware, and Massachusetts, despite having high population densities, exhibit relatively lower death counts(See Figure26). This observation suggests that the death count is not solely dependent on population density. Instead, there may be a more complex relationship, potentially linking population density with deaths per capita rather than raw death counts.

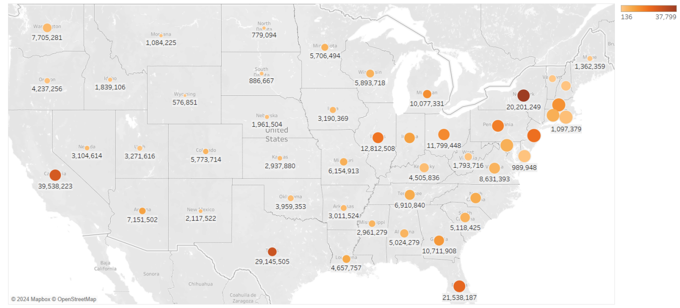


Fig. 25: Choropleth of The United States, visualizing the number of deaths and the population density

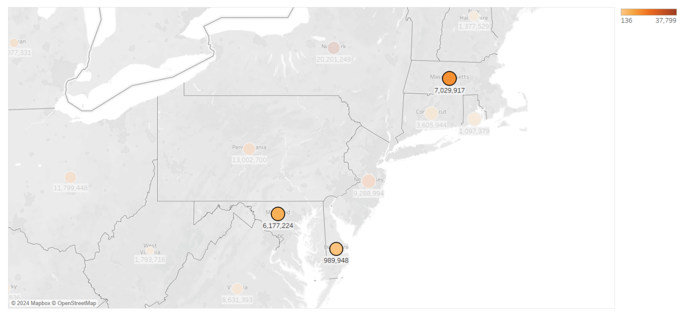


Fig. 26: Choropleth of The United States visualizing the number of deaths and the population density, focussing on the states that have a high population density but a low death count

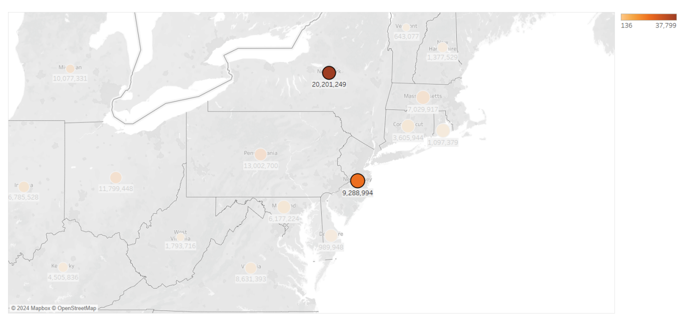


Fig. 27: Choropleth of The United States visualizing the number of deaths and the population density, focussing on the states that have a high population density and a high death count

*4) Iteration 4:* In this iteration, the analysis delves into a more complex relationship between population density and deaths.

The goal is to explore the potential connection between deaths per capita and population density, hypothesizing that regions with higher population densities are likely to experience more deaths per capita compared to less densely populated regions.

Additionally, this iteration seeks to understand the role of the stringency index. It is hypothesized that two states with similar population densities may exhibit differing death rates per capita depending on how stringent their government regulations were. If a state had stricter regulations, it was expected that the death per capita would be lower; conversely, less strict states might experience higher death rates.

To visualize these relationships, a scatter plot(See Figure 28) was created for each state, with the color of the circles representing the stringency index, using a linear color scale.

**Inference**

From the regression curve plotted, we observe a weak correlation between population density and deaths per capita. Furthermore, states that are positioned above and below the regression line show varying stringency indices, suggesting that the stringency of regulations did not have a significant impact on deaths per capita for many states.

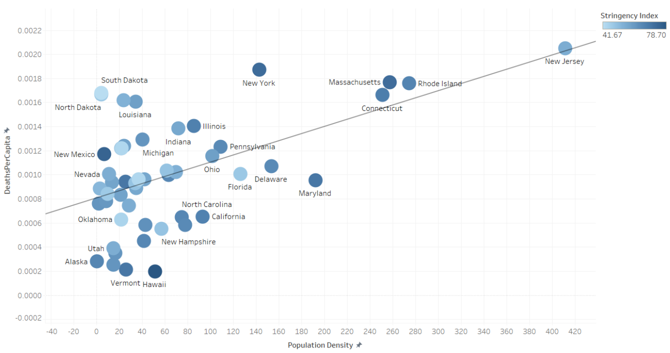


Fig. 28: The figure is a scatter-plot visualizing the relation between Deaths-Per-Capita and Population Density the color of the circle indicates how strict the regulations were, with darker color indicating more strict regulations. The line is a regression curve that best fits the data

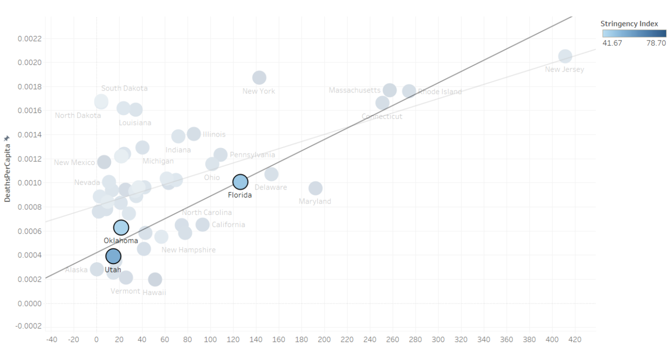


Fig. 29: The figure shows the states that have a lower Death per Capita value when compared to the regression curve in 28 despite having a low Stringency Index

## D. Conclusion

States with similar outcomes in terms of deaths per capita are clustered by shared characteristics, such as high population density and moderate stringency levels, but the presence of outliers highlights the importance of local factors, including healthcare infrastructure and policy compliance. This analysis underscores that a one-size-fits-all approach to pandemic response may not be effective, emphasizing the need for tailored strategies.

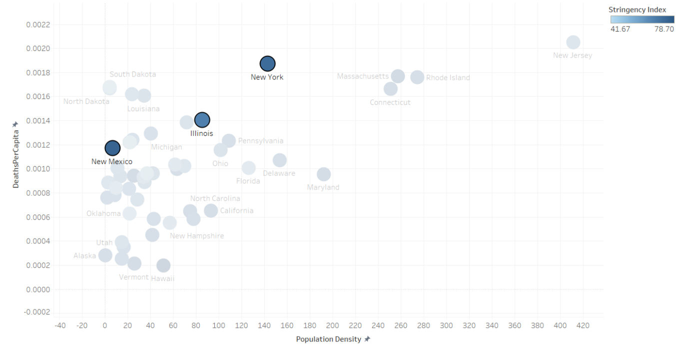


Fig. 30: The figure shows the states that have a higher Death per Capita value when compared to the regression curve in 28 despite having a high Stringency Index

## References

[1] Calendar chart visualization - vertex42.
[2] Google covid-19 data.
[3] Matplotlib for visualizations.
[4] Oxcgrt github documentation.
[5] Oxford covid-19 government response tracker.
[6] Reduced efficacy of vaccines on the delta variant.
[7] Tableau for visualizations.