COVID-19 Individual Project Report

For Data Visualization (INF552)

Yuelin HU[a], Alex Quach[a], Yujia CHENG[a], Yuxin LIU[a]

[a]*École Polytechnique, Palaiseau, France*

December, 2023

Abstract

This report details the development of an interactive data visualization tool using D3.js to represent COVID-19 case trends in the United States over time. The project focuses on rendering dynamic graphs that allow users to view trends for cases, deaths, vaccinations, hospitalizations, etc., of all the different counties in the US. Our work is presented here.

1. Introduction

The COVID-19 pandemic has generated vast amounts of data, necessitating effective visualization tools for comprehension and analysis. This project aims to create an interactive visualization to track the progression of COVID-19 cases in the United States, employing JavaScript and D3.js to render dynamic, responsive graphs.

The datasets used for this visualization project come from the following repositories: COVID-19 Data in the United States, HealthData.gov, Centers for Disease Controls and Prevention, and The Covid Tracking Project.

1.1. Problem Description

The COVID-19 pandemic has presented an unprecedented global health crisis, with far-reaching impacts on various aspects of human life, including healthcare systems, economies, and day-to-day activities. One of the primary challenges in managing such a pandemic is the effective understanding and interpretation of vast amounts of data related to the virus's spread, its effects on different demographics, and the efficacy of measures taken to control it.

Graphs and tools are crucial in translating raw data into actionable insights, aiding policymakers, healthcare professionals, and the general public in understanding the pandemic's dynamics.

1.2. Motivation

The motivation behind this project stems from the desire to understand the COVID-19 pandemic, mainly focusing on how it has affected the various states and counties in the United States, and how it has evolved from 2019 to 2023. Data visualization serves as a powerful tool, offering clarity and insight into the complexity of pandemic-related data.

2. Datasets & Graphs

2.1. Total Reported Cases & Total Reported Deaths

The datasets utilized for the Total Reported Cases Graph and the Total Reported Deaths Graph comprises daily updates of COVID-19 cases and deaths reported in the United States. The purpose of this visualization is to provide a clear and concise overview of the pandemic's progression over time, highlighting trends and patterns in the spread of the virus in the United States.

2.1.1. Technical Development

The graph was developed using D3, since it allows for interactive and dynamic graph creation. The process involved importing the dataset using D3, selecting appropriate graph types, and configuring data fields for the x-axis (time) and y-axis (case counts). D3's built-in functionalities were utilized to refine the visualization, such as filtering options and tooltips for detailed information on hover.

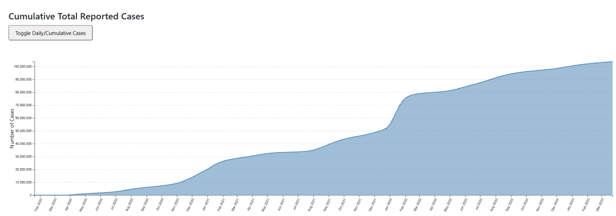


Figure 1: Cumulative Cases Graph.

2.1.2. Visualization Design and Features

The designs of the Total Reported Cases Graph and the Total Reported Deaths Graph are time-series line graphs, chosen for their effectiveness in displaying trends over time. Key features include:

- **Visualization Toggle**: Users can toggle between viewing the total reported cases cumulatively, or view the daily reported cases.

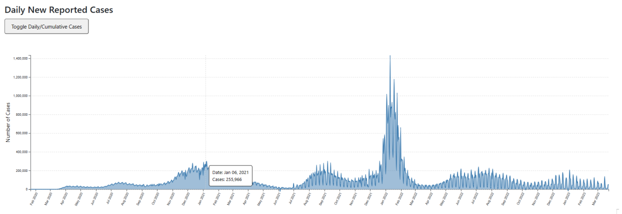


Figure 2: Daily Cases Graph.

- **Hover Tooltips**: Hovering over any point on the graph displays a tooltip with exact numbers and dates, providing detailed information without cluttering the visual.
- **Horizontal Reference Lines**: The graphs provides dotted horizontal reference lines at fixed intervals.

2.1.3. Rationale for Design Choices

- **Line Graph Selection**: A line graph was chosen as it is ideal for showing changes over time. Its continuous nature allows for easy tracking of the pandemic's trajectory and identification of trends.
- **Visualization Toggle**: This allows the users to see two difference sides of the same dataset. The cumulative graph helps see how quickly the number of cases grew, and the daily graph helps see which dates saw the most case growth.
- **Hover Tooltips**: This allows users to quickly see how many new cases a particular date had.
- **Horizontal Reference Lines**: The reference lines allows users to quickly identify around what number of cases a particular date had.

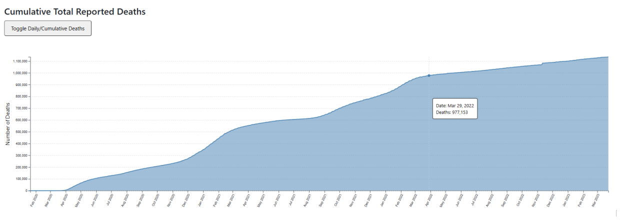


Figure 3: Cumulative deaths graph.

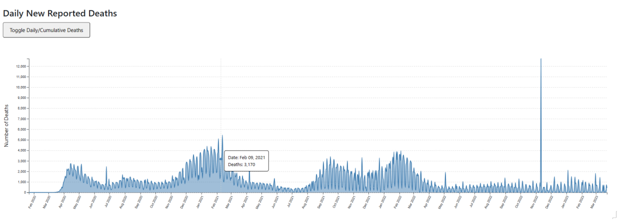


Figure 4: Daily Deaths Graph.

2.2. New Hospital Admissions by Age

2.2.1. Data Processing

We utilized the dataset titled "COVID-19 Reported Patient Impact and Hospital Capacity by State Timeseries" from HealthData.gov. This dataset provides state-aggregated data for hospital utilization in a timeseries format, dating back to January 1, 2020. Then we employed Python for data integration and filtering to calculate the new hospital admissions across seven age groups.

2.2.2. Statistical Significance

- Understanding changes in health status and potential medical needs across different age groups, particularly during pandemics or other health crises.
- Assisting policymakers and hospital administrators in predicting and preparing necessary medical resources such as beds, equipment, and staff.
- Providing researchers with data to analyze health challenges faced by specific age groups at certain times.
- Offering the public critical information regarding health risks and preventive measures, especially for high-risk groups.

2.2.3. Visualization

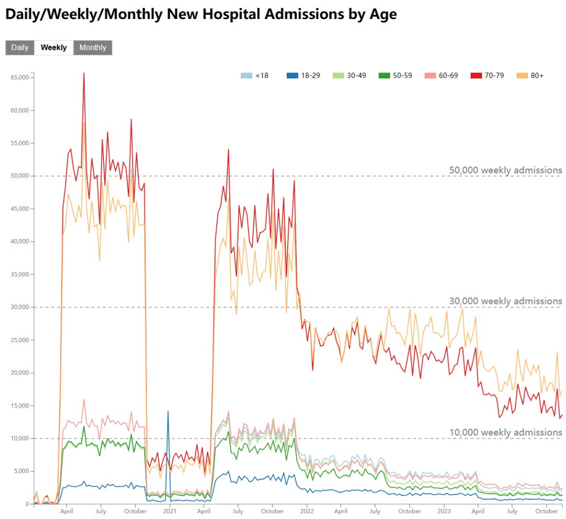


Figure 5: Weekly New Hospital Admissions by Age

Figure 5 presents an interactive line chart that illustrates the number of new hospital admissions by age group, displayed on a daily, weekly, or monthly basis. At the top of the chart, three buttons — "Daily," "Weekly," and "Monthly" — enable users to switch the data display's granularity. Each line correlates with an age group, with color coding matched to the legend, allowing users to identify which line corresponds to which age group.

1. **Interactive Switching:** Users can switch between daily, weekly, or monthly data views by clicking buttons above the chart. This dynamic update is accomplished by binding event listeners to the button clicks. And there are effects that change color when hovered and selected.

Figure 6: 3 Buttons

2. **Axis:** The horizontal axis signifies time, with ‘d3.scaleTime‘ mapping time data to appropriate pixel positions. The timeline extends from approximately April 2021 to October 2023, from left to right. The vertical axis indicates the number of hospital admissions, with ‘d3.scaleLinear‘ mapping admissions numbers to pixel locations. The colored lines represent the fluctuations in hospital admissions by age group over time.
3. **Reference Lines:** Dashed lines serve as benchmarks for weekly admission figures, such as ”50,000 weekly admissions” and ”30,000 weekly admissions,” aiding users in quickly assessing data against these benchmarks.

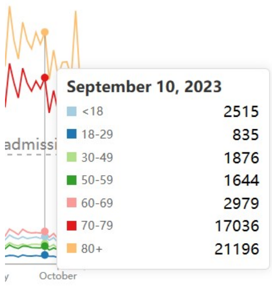


Figure 7: Tooltip

4. **Mouseover Interactions:** Hovering over any data point triggers a tooltip, providing specific data values like the date and number of admissions by age group. Each tooltip is accompanied by a corresponding vertical marker line, connecting data points of a specific date across all age group lines.

2.3. Impact of COVID-19 on State-Level

This interactive visualization offers a comprehensive view of COVID-19’s impact in the United States by examining confirmed cases, related deaths, and hospitalizations. The dataset is from The COVID Tracking Project. We have designed and developed three types of visualizations using D3.js: **bubble chart**, **line graph** and **bar chart** to show how states in the US were impacted by COVID-19.

Hovering over the chart elements engages a linkage effect across the three charts, simultaneously revealing detailed information for the state in focus.

2.3.1. Bubble Chart

Figure 8 shows the cumulative COVID-19 hospitalizations for each state.

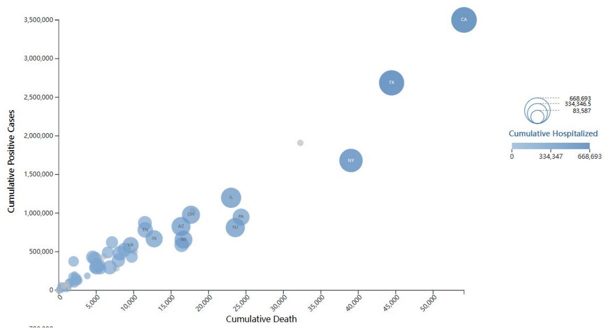


Figure 8: Bubble Chart

1. **Axes**: The x-axis is labeled “Cumulative Death” and indicates the total number of COVID-19 related deaths per state.
   The y-axis is labeled “Cumulative Positive Cases” and shows the total number of confirmed COVID-19 cases per state.
2. **Bubbles**:
   - Each bubble corresponds to a state, labeled with the state’s abbreviation.
   - The position of each bubble is based on the state’s data for deaths (x-axis) and cases (y-axis).
   - The size and color of each bubble is proportional to the total number of COVID-19-related hospitalizations in the state.
3. **Legend**:
   - **Circle Legend** The circle legend visualizes the relative size of the data by using circles of different sizes. The user can estimate the relative number of hospitalizations by the size of the circles.
   - **Rectangle Legend** The rectangular legend indicates the range of the data by shades of color, from the lowest value (light blue) to the highest (dark blue). This color coding allows the user to quickly understand the distribution of the data.
4. **Data Points**:
   - States with a larger number of cases and deaths appear toward the top-right of the chart.
   - States with fewer cases and deaths are positioned toward the bottom-left.

5. **Interactivity**: When the user hovers over the corresponding bubble, it will be highlighted, with a black border, and linked to the line graph and bar chart, displaying the full name of the corresponding state in the middle of the entire data visualization layout.

2.3.2. Line Graph

Figure 9 shows the cumulative positive cases and cumulative deaths in each state over time.

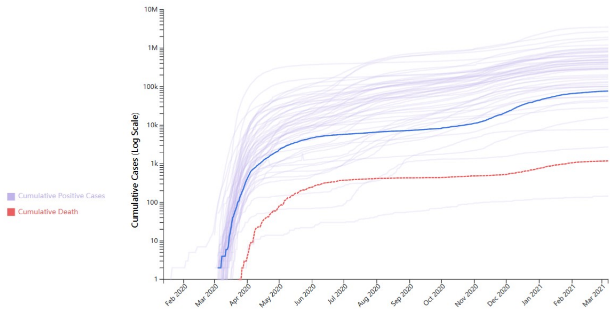


Figure 9: Line Graph

1. **Logarithmic Scale**: The y-axis is logarithmic, as indicated by the labels 1, 10, 100k, and so on, up to 10 million. A logarithmic scale is used to represent data that covers a wide range of values in a condensed manner. Each tick mark on the axis represents an exponential increase in the number of cases and deaths, which allows for easier visualization of the growth patterns over time, especially when the growth is exponential.
2. **Default State Lines**: When the mouse is not hovering over the chart, the purple lines represent the cumulative number of positive COVID-19 cases for various states.
3. **Hover Interaction**: Upon hovering over a particular line, it transforms into a bold blue solid line. This emphasizes the data for that state.
4. **Corresponding Data Visualization**: Simultaneously with the line becoming bold, a corresponding dotted red line appears, representing the cumulative number of deaths in that state.
5. **Linked Visualizations**: Similar to the bubble chart, this graph is interactive and linked with two other charts. This means that when interacting with this chart, the other two charts will also display corresponding effects or data visualizations reflective of the selected state's data.

2.3.3. Bar Chart

Figure 10 shows cumulative ventilator use and cumulative ICU cases per state. The x-axis is labeled with state abbreviations. Some is unavailable for certain states in the data set.

1. **Categories**: The light orange portion of the bar represents individuals in the Intensive Care Unit (ICU).
   The blue portion represents individuals who require ventilators.
   The dashed line running across the chart likely represents the cumulative number of deaths. This is a separate scale on the right side of the chart, which correlates with the primary y-axis for hospitalization data.
2. **Interactivity**: When hovering over a bar, a black border appears around the bar, enhancing its visual prominence against the other bars, simultaneously linking the data displays of the corresponding states of the other two graphs.

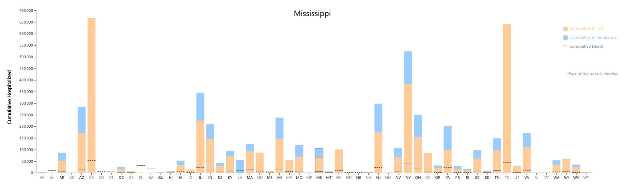


Figure 10: Bar Chart

2.4. Cases, Deaths per Capita, Hospitalized and Vaccinations Statewide

In this subsection, we embark on a crucial journey to effectively communicate complex data in the contemporary digital age. Our aim is to develop an interactive choropleth map that provides a vivid portrayal of the COVID-19 pandemic's impact across the United States. This visualization focuses on key metrics such as cases, deaths per capita, hospitalizations, and vaccinations. Through this, we aspire to enhance data accessibility and understanding, serving as a valuable tool for policymakers, healthcare professionals, and the general public in their decision-making processes.

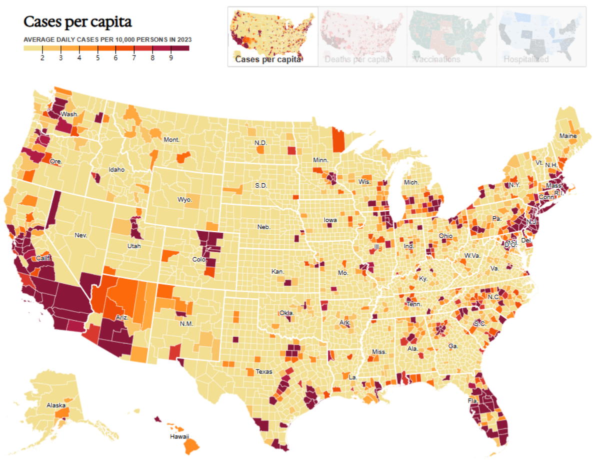
2.4.1. Dataset Description

The data for this map is sourced from reliable and up-to-date public health databases. The primary datasets include:

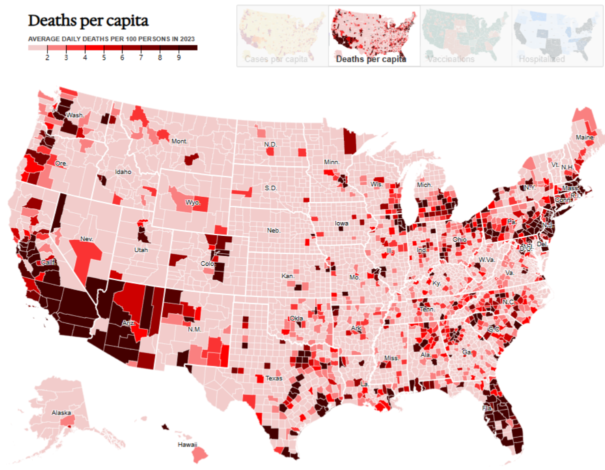
Case and Death Data: Sourced from a 2023 dataset detailing cases and deaths county-wise across the U.S. from public database, providing a comprehensive view of the pandemic's spread and severity.

Vaccination Data: This dataset is sourced from Centers for Disease Control and Prevention, which encapsulates the vaccination rates at the county level.

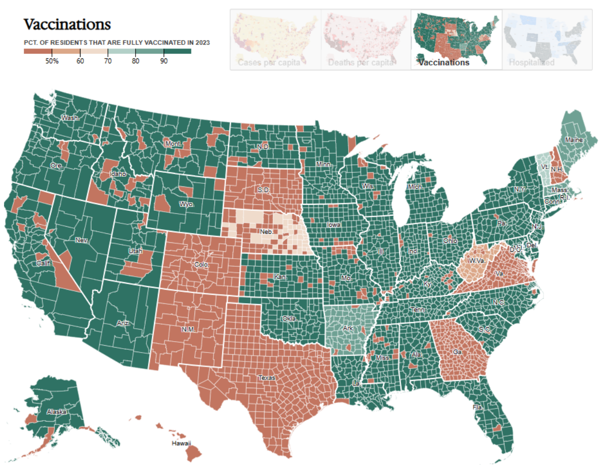
Hospitalization Data: Sourced from the Reported Patient Impact and Hospital Capacity by State dataset in HealthData.gov, it provides state-level data on hospitalizations due to COVID-19, crucial for understanding healthcare system impacts.



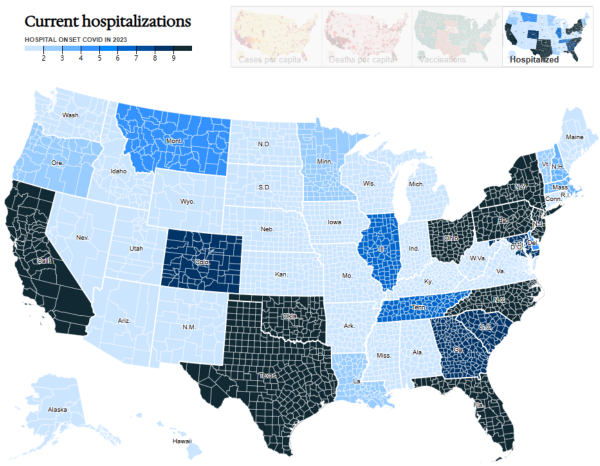
(a) Cases per capita.



(b) Deaths per capita.



(c) Vaccinations in counties.



(d) Current hospitalizations.

Figure 11: Choropleth Maps

2.4.2. Data Processing

The data processing stage of this data visualization is crucial for accurate and effective presentation of COVID-19 metrics. Our approach involves several key steps:

Normalization of Data: We adjusted the units of COVID-19 cases and deaths to improve the visualization of data ranges. This includes scaling cases per 10,000 people and deaths per 100 people to ensure the data is comparable and easy to interpret on a choropleth map.

Geographical Mapping: We meticulously matched each dataset's cases, deaths, vaccinations, and hospitalization data with corresponding state and county information. This involved the use of FIPS codes to accurately associate data with specific geographical locations.

Handling Missing Data: For areas with null or missing data, we implemented a system to display a "No data available" message, ensuring clarity and transparency in our visualizations.

Data Source Integration: Recognizing limitations in available datasets, we sought additional sources for vaccination and hospitalization data that were more aligned with our project's needs. Specifically, hospitalization data was represented at the state level due to the lack of county-specific information.

Temporal Filtering: We focused exclusively on data from the year 2023, filtering out all historical data to provide a current and relevant snapshot of the pandemic's impact.

2.4.3. Visualization settings

The visualization is implemented using D3.js for manipulating documents based on data. The main features of the implementation are as follows:

Interactive Map: A geo-mapped visualization representing the U.S., with the ability to display data at the county level. The map allows users to switch between different data views (cases, deaths, vaccinations, and hospitalizations).

Color Coding: Each data category is represented using a distinct color scale shown in Fig. 12, enhancing readability and differentiation. For instance, cases are represented in shades of orange, deaths in red, vaccinations in blue, and hospitalizations in light yellow.

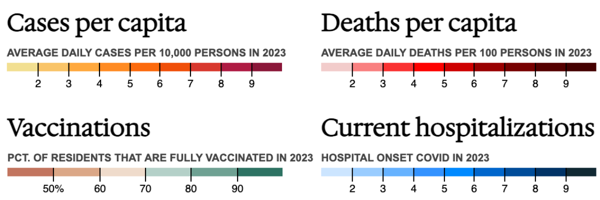


Figure 12: Scaled Color Legends of Maps

Thumbnail Interaction: Positioned over the map, this bar houses clickable thumbnails shown in Fig. 13 representing different data types. Each thumbnail is an image button that switches the map to display the corresponding data, with designed transitions to the new data representation, providing a seamless user experience. Using CSS, they exhibit changes in opacity on hover and become fully opaque when active, providing clear feedback to the user. Each thumbnail is accompanied by a label for better understanding, and the layout is designed to be intuitive and accessible.

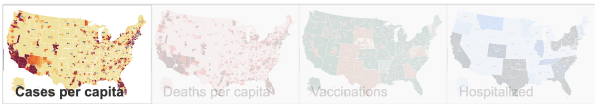


Figure 13: Button of Thumbnail

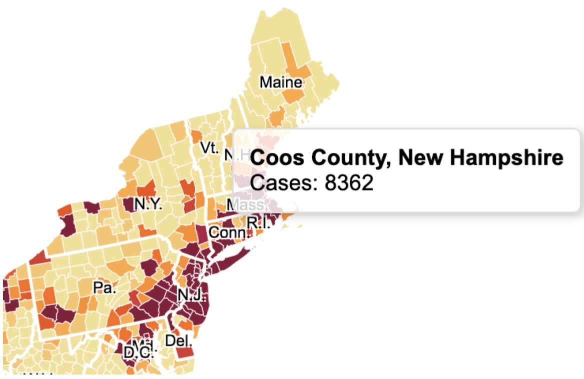


Figure 14: Tooltip of Maps

Tooltip Functionality: Hovering over a specific county on the map displays a tooltip with detailed data, such as the number of cases, deaths, or vaccination rates, as well as corresponding county and state names, which we can observe in Fig. 14.

Aesthetic and Functional Design: Each data type is represented with its own distinct and visually appealing color scale, chosen for clarity and user engagement. These color scales are designed to transition smoothly as users switch between different data types, enhancing the interactive experience. Typography is another key element; state and county names are displayed in carefully selected fonts like 'Kalnia' and 'Hedvig Letters Serif', chosen for their legibility and aesthetic appeal, with font sizes adjusted for readability against the map's spatial constraints. The outlines of states and counties are finely tuned, featuring a white stroke color and width that clearly delineate each region. Moreover, the map's interactive elements, such as hover effects and clickable regions, are not only functional but also contribute to the aesthetic appeal, ensuring a seamless and engaging user interface. Together, these design elements coalesce to present complex COVID-19 data in a format that is both accessible and visually compelling for a diverse audience.

2.5. HTML Illustration

In the following section, we present a comprehensive illustration of all the graphs and visualizations developed for this project, seamlessly integrated into a single HTML framework. Each visualization, whether it be line graphs, bubble charts, or interactive maps, is an embodiment of our commitment to delivering clear, informative, and engaging data on COVID-19's impact. Our work as well as the detail codes are presented here.

3. Conclusion

In conclusion, this project represents a significant endeavor in visualizing the multifaceted impact of the COVID-19 pandemic in the United States. Through the utilization of diverse datasets and the application of advanced data visualization techniques using D3.js and other tools, we have successfully created a series of interactive and informative graphs and maps. These visualizations not only provide a clear and comprehensive overview of the pandemic's progression, including cases, deaths, hospitalizations, and vaccinations, but also offer valuable insights into the temporal and geographical spread of the virus. The integration of aesthetic and functional design elements ensures that the complex data is accessible and engaging to a wide audience, including policymakers, healthcare professionals, and the general public. This project stands as a testament to the power of data visualization in understanding and responding to global health crises.