

Project Metadata

Title: *Pandemic Trends: A Visual Exploration of COVID-19 Data*

Team Members:

Name	Email	Spire ID
Hemangani Nagarajan	hemanganinag@umass.edu	34036775
Kavisha Parikh	kavishaprana@umass.edu	34733769
Varshini Venkataraman	vvenkatarama@umass.edu	34829219

GitHub: <https://github.com/HEMANGANI/571-Project>

Background & Motivation

The COVID-19 pandemic has been one of the most significant global health crises, affecting millions worldwide. Understanding how the virus spreads across different regions, its impact on mortality rates, and recovery patterns is crucial for gaining insights into pandemic dynamics and informing future public health responses.

Our motivation for this project stems from the opportunity to work with an extraordinarily vast and multimodal dataset that captures numerous dimensions of the pandemic. The COVID-19 data includes temporal, geographic, demographic, and epidemiological information that, when viewed as raw numbers alone, cannot reveal the complex interrelationships and patterns that defined the pandemic's progression. By transforming these complex, multimodal datasets into accessible, interactive visualizations, we aim to uncover hidden patterns, unexpected correlations, and subtle, invisible trends in tabular data formats.

Working with this rich dataset provides an exceptional learning opportunity to develop our data visualization skills while potentially revealing insights that could contribute to pandemic understanding. The visual exploration of this data can illuminate spatial-temporal patterns of virus spread, identify unexpected hotspots, and demonstrate the effectiveness of different regional responses.

The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) has maintained comprehensive datasets tracking COVID-19 cases, deaths, and recoveries worldwide. By leveraging these datasets, we can create meaningful visualizations beyond simple counts and ratios to reveal the complex story of how the pandemic unfolded across different geographical regions and timeframes.

Project Objectives

The primary goal of our project is to develop interactive visualizations that provide insights into the spread and impact of COVID-19 across different regions and timeframes. Through these visualizations, we aim to answer the following key questions:

1. **How did COVID-19 spread across different regions over time?**
 - By visualizing the progression of confirmed cases, we can identify when and where major outbreaks occurred and how quickly the virus spread in different areas.
2. **What were the mortality and recovery patterns across different regions?**
 - By comparing death and recovery rates across different countries and states, we can identify severely affected regions and those managed the pandemic more effectively.
3. **When did the spread of COVID-19 slow down in different regions?**
 - By analyzing trends in case numbers over time, we can identify when various regions reached their peak infection rates and when the spread began to slow.
4. **How did different waves of infection compare across regions?**
 - By examining time-series data, we can identify and compare different waves of infection across regions.
5. **What are the relationships between confirmed cases, deaths, and recoveries?**
 - By analyzing correlations between these metrics, we can gain insights into the overall impact of the pandemic.

Data

JHU CSSE COVID-19 Dataset:

GitHub Repository: <https://github.com/CSSEGISandData/COVID-19.git>

- **Important Note:** The dataset was not updated after March 10, 2023 (3/10/2023), as JHU CSSE concluded their data collection efforts. Our analysis will cover the pandemic from its beginning until this cutoff date.
- Key datasets include:
 - time_series_covid19_confirmed_global.csv:
 - Contains daily cumulative confirmed COVID-19 cases for all countries outside the US.
 - time_series_covid19_deaths_global.csv:
 - Records cumulative death counts for each country/region over time, following the same structure as the confirmed cases file.
 - time_series_covid19_recovered_global.csv:
 - Contains recovery data by country/region over time, though it's worth noting that recovery reporting has been less consistent.
 - time_series_covid19_confirmed_US.csv:
 - Provides county-level data for the United States, with significantly more geographic granularity.
 - time_series_covid19_deaths_US.csv:
 - Contains US county-level death data and includes population data for each county, which enables per capita calculations.

Data Processing

Our data processing approach will involve several key steps to ensure that the data is clean, consistent, and ready for visualization:

1. Data Extraction and Integration:

- We will extract relevant data from the JHU CSSE time-series files.
- For global data, we will combine information from the confirmed cases, deaths, and recoveries datasets.
- For US data, we will integrate state-level information from the confirmed cases and deaths datasets.

2. Data Cleaning and Standardization:

- Handling missing values by imputing them based on surrounding data points or excluding them from the analysis.
- Standardizing country and region names to ensure consistency across datasets.
- Converting date formats to a standardized format for easier time-series analysis.

3. Compute Quantities:

- **Case Rates:** Calculating per capita cases by normalizing confirmed cases by population.
- **Death Rates:** $(\text{Total Deaths} / \text{Total Confirmed Cases}) \times 100$.
- **Recovery Rates:** $(\text{Total Recoveries} / \text{Total Confirmed Cases}) \times 100$.
- **Growth Rates:** day-over-day and week-over-week growth rates in confirmed cases.
- **Doubling Times:** Computing how quickly cases doubled in different regions.
- **Peak Identification:** Identifying when different regions reached peak infection rates.
- **First Onset Dates:** Determining when COVID-19 first appeared in different regions.

4. Time-Series Processing:

- Converting cumulative data to daily new cases, deaths, and recoveries.
- Creating rolling averages (7-day, 14-day) to smooth out reporting inconsistencies.
- Identifying trends and patterns in the time-series data.

5. Categorical Aggregation:

- Aggregating data by geographical regions (countries, states, continents).
- Creating time-based aggregations (monthly quarterly summaries).
- Classifying regions based on severity metrics (case rates, death rates).

We will implement this using JavaScript and D3.js for the web-based visualization, with preprocessing done in Python to generate clean, structured datasets ready for visualization.

Visualization Design

Our visualization design will focus on creating interactive, informative, and user-friendly representations of COVID-19 data. Based on our brainstorming and design exploration, we have developed the following visualization components:

Brainstorming Ideas

more useful for visualizations

COVID19 Data

General dataset

Daily reports

Time Series Data

Sub datasets

Confirmed

deaths

recovered

global & US

Filter

Data points

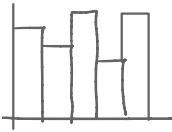
- Date
- Location
- Cases
- Deaths
- Recovery
- Population
- Vaccine
- Source



- heatmap
- cases doubling
- regions
- Bit complex to understand X
- We can see in map more clearly



- area maps
- growing areas
- how covid progresses
- add more features in next iterations
- confirmed cases



- top N regions
- it can be cases, deaths or recovery
- all sub-datasets are included
- changes according to different time periods
- location + other data used

1. —□
2. —○
3. —△
4. —■
5. —□



- definitely include mapcharts
- easier to perceive the location
- combine location and other datasets information
- refined view



- Region wise recovery/deaths percentage
- compare across multiple regions
- normalize to 100%
- link it to bar graph

Categorize

include trend lines

area represents cases

trend lines for deaths, recovery etc ...

Trends in time-series data

Analytical Visualization

Split into 2



Combine

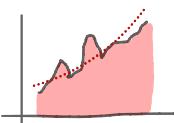
clicking country → changes piechart



Geographic Visualization

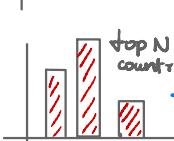
Colour coded highlight hotspots gradient to show spread and intensity concentration

Combine & Refine



Trends Graph

Different waves can be monitored how the trend changes over time



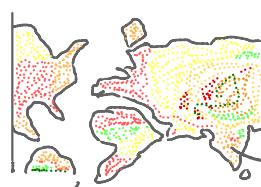
top N countries



Deaths

vs Recovery

hot spots highlighted red



Global Covid Spread

tool tip shows detailed info



Cases vs Deaths
vs Recovery

Linked charts

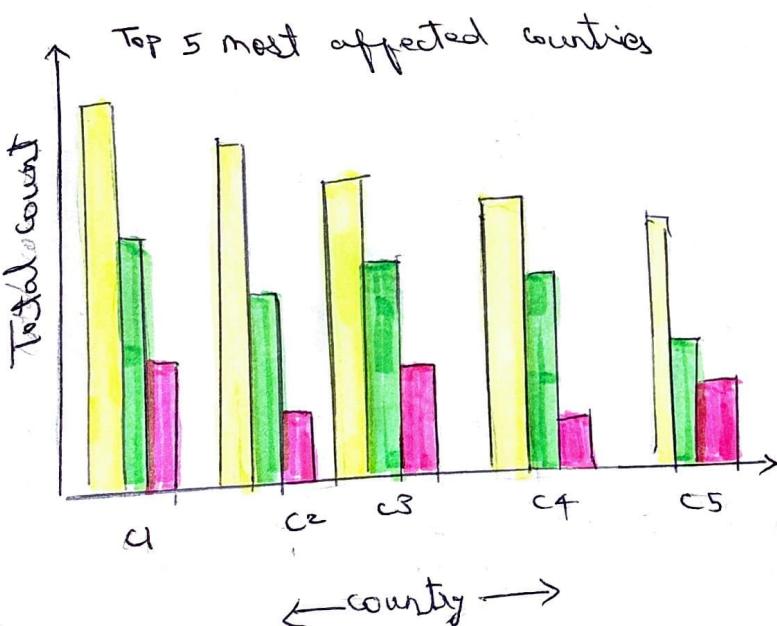
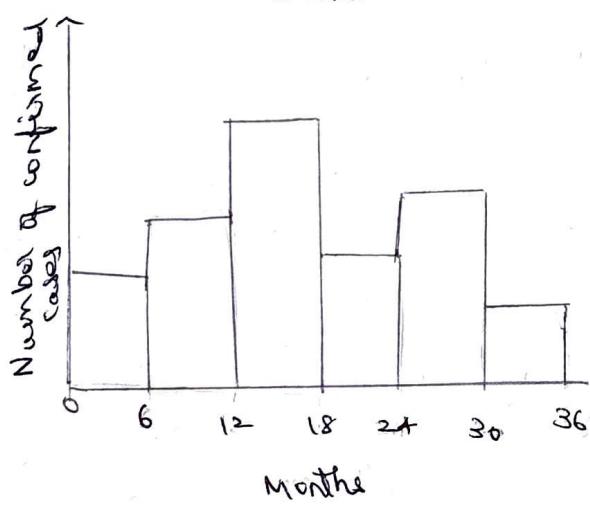
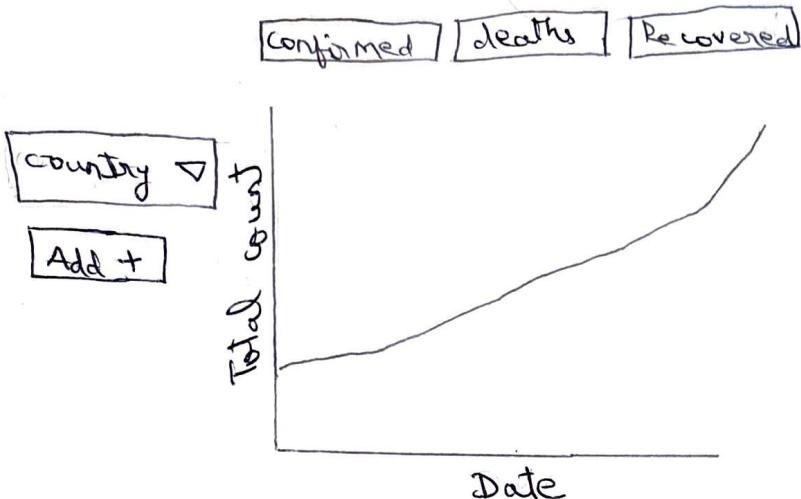
which region is more effective

Region cruise

relationships in these ratios

Identification of highly affected areas.

Time series analysis over different periods of time.



Metadata:

Title: Pandemic Trends: A visual exploration of COVID-19 Data

Sheet: 2

Task: Trend over different time periods

Operations

- * The user can select country from dropdown for which he wants to see the graph.
- * Add button: Adds a new country dropdown for comparisons (max 5).
- * Confirmed, Deaths and recovered buttons add a new trend line for the country selected.
- * Hover over a bar or line to get the exact value of that point.
- * Hover over a country in the cluster column chart to get exact value of point.

Discussions

- 1) What are the advantages of having these graphs?
- 2) Do these accurately convey the trend?

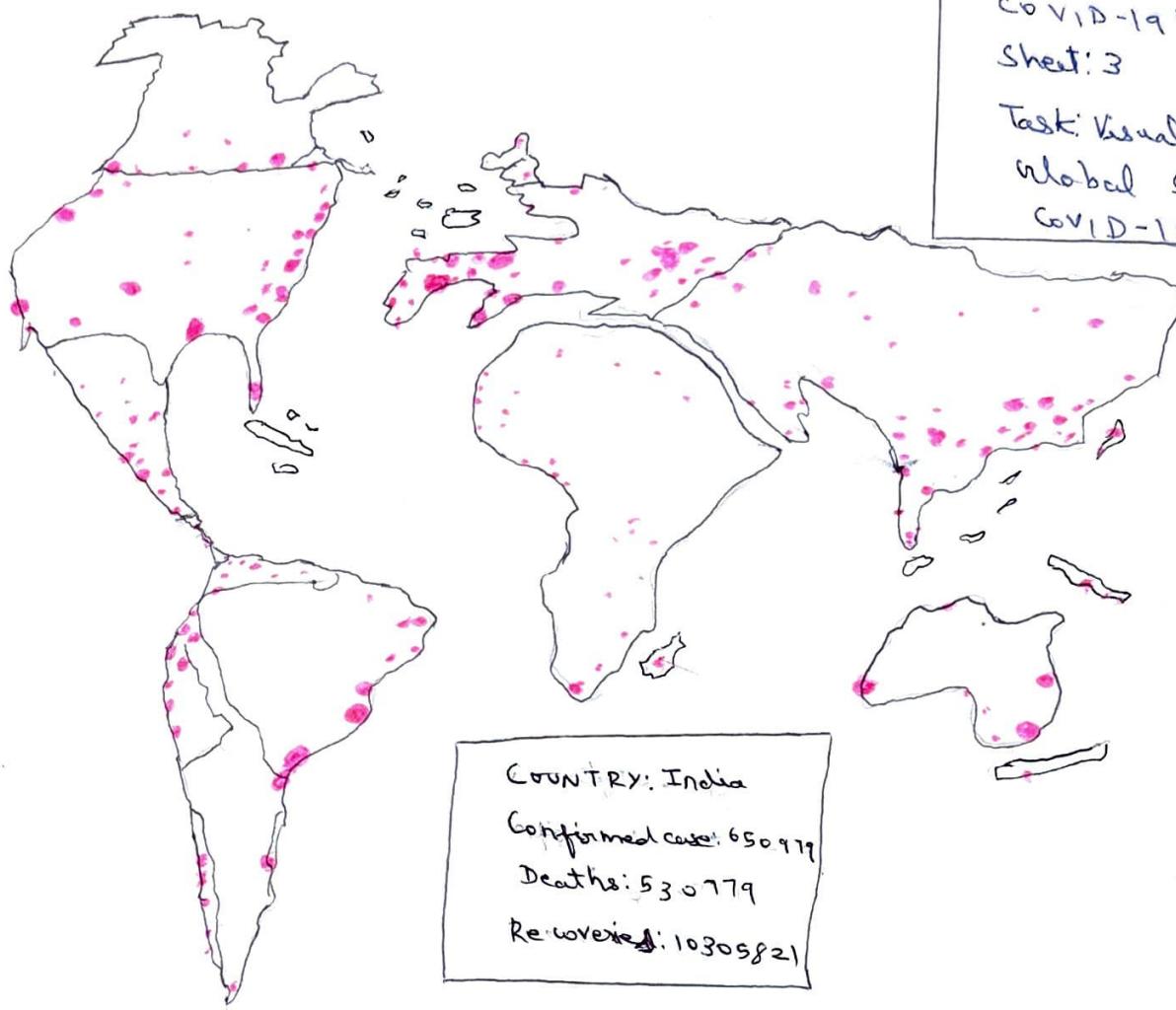
Global COVID-19 Spread: Total Confirmed Cases, Deaths and Recoveries

Metadata:
Title:

Page - 3

Pandemic Trends: A Visual Exploration of COVID-19 Data
Sheet: 3

Task: Visualizing the Global Spread of COVID-19



Operations:

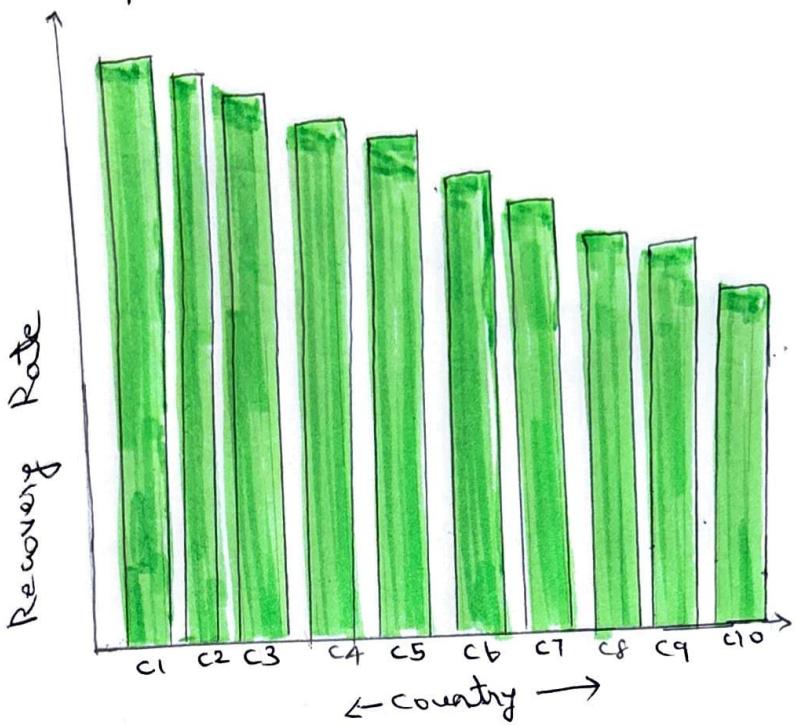
* When a user hovers over a country, the country's name, the number of confirmed cases, deaths and recoveries are displayed.

* Additionally, the intensity of COVID-19 can be indicated using color coding:
red for severe, yellow for moderate and green for safe.

Discussions:

- 1) What type of graph, other than line and bar graphs, would better illustrate the spread of COVID-19 across the globe?
- 2) What kind of information should be included when the user hovers over the graph?
- 3) Would using color coding to represent confirmed cases be effective?

Top 10 countries with best recovery rate



Country ▼



Recoveries And Deaths

Metadata:

Title:

Pandemic Trends: A Visual Exploration of COVID-19 Data

Sheet: 4

Task: Displaying countries with best recovery rates.

Operation:

* For the bar chart: When a user hovers over the bars, the value will be displayed.

* For the pie chart: The user can select the country they want to view the number of deaths or recoveries for.

Discussion:

- 1) How can we visualize the deaths and confirmed cases for all countries and will this be useful?
- 2) Should we display data about the top 10 countries with best recovery rates?
- 3) What are the advantages of these two charts?

Focus

Page - 5

Dashboard Design, Chart selection, Extracting requirement and relevant Data processing steps.

Meta Data

Title: Pandemic Trends: A visual exploration of COVID-19 Data.
Names: Kavisha; Varshini, Hemangani.
Tasks: Dashboard Design, chart placement.

PANDEMIC TRENDS: A visual Exploration of COVID-19 Data.

Rise of COVID Cases over 3 yrs

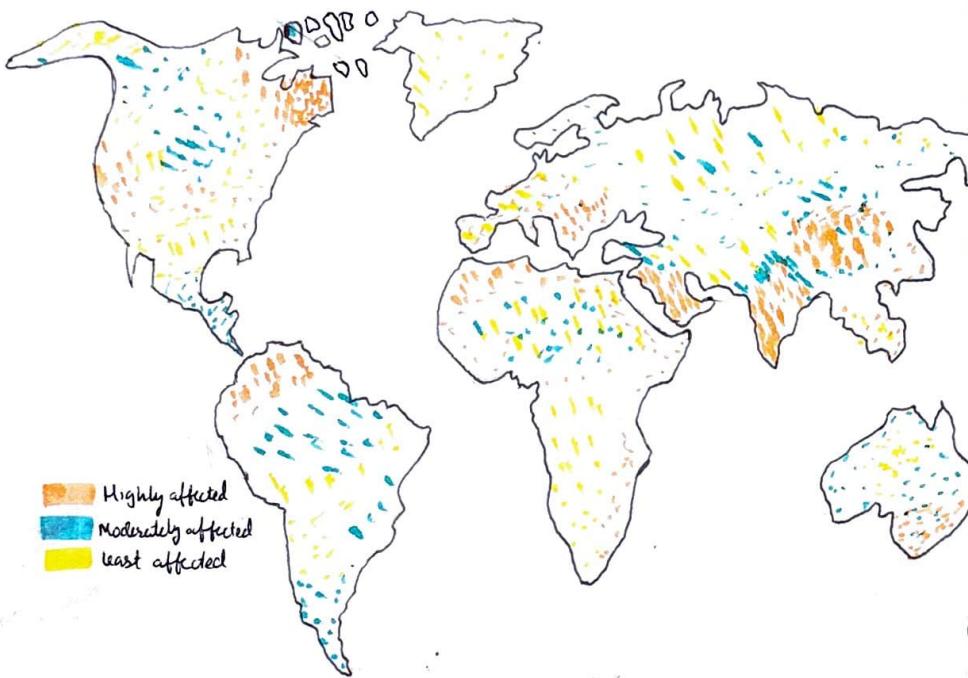
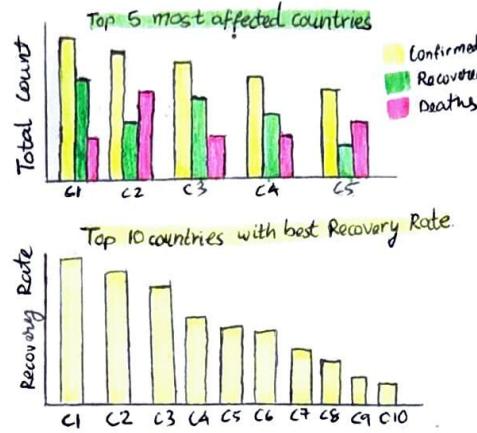
Confirmed Deaths Recovered

Country ▾

Add +

Total Count

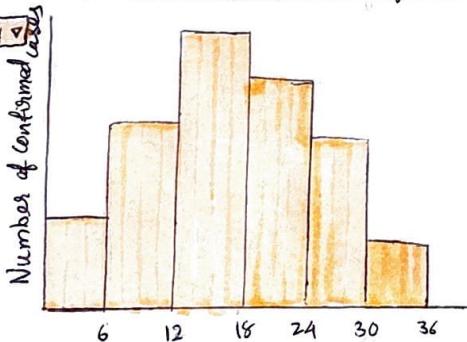
Date (22/3/2020 - 30/3/2023)



Global COVID-19 Spread: Total Confirmed Cases

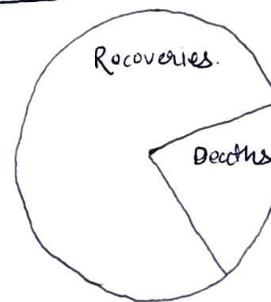
Number of confirmed cases every 6 Months

Country ▾



Recoveries and Deaths

Country ▾



Operations

- Country Dropdown: User selects country from dropdown for which he wants to see the graph.
- Add Button: Adds a new country dropdown for comparison (Max 5)
- Confirmed, Deaths and recovered Buttons: add a new trendline for the country selected.
- Hover over a bar or line to get the exact value of that point.
- Hover over a country in the choropleth map to get exact number of confirmed cases, deaths and recoveries.

Detail.

Software Dependencies.

Frontend: D3.js, CSS, Bootstrap, HTML.

Backend: Excel, Python, Pandas, Numpy, Flask.

Calculations

Recovery Rate graph: Filtering countries with cases above a threshold and computing $(\text{total recoveries} / \text{total confirmed}) * 100$

6-month graph: Data is cumulative, so in order to get values for this graph, subtract row every 6-month date column, e.g., for bar chart, first bar value will be 22 July 2020 - Jan 22 2020

Pie chart calculation: Death and recovery ratio calculation relative to total outcomes.

Must-Have Features

The following features are essential for our COVID-19 data visualization project:

1. **Interactive Time Series Visualization**
 - Line chart showing progression of COVID-19 metrics over time
 - X-axis: dates
 - Y-axis: confirmed cases, deaths, or recoveries (user-selectable)
 - Users can select specific countries/regions to display
 - Visualization supports data from all five key datasets (global confirmed, global deaths, US confirmed, US deaths, global recovered)
2. **Global Choropleth Map**
 - Color-coded world map displaying either confirmed cases, deaths, or recoveries based on user selection
 - Toggle to view either total cumulative numbers or per capita rates
3. **Comparative Bar Chart by Time Period**
 - Bar chart showing progression over 6-month periods
 - X-axis: 6-month periods
 - Y-axis: selected metric (deaths, recoveries, or confirmed cases)
 - Users can select specific countries or view data for all countries
 - Clear visual distinction between time periods
4. **Top N Countries Clustered Column Chart**
 - Clustered bar chart showing the top N most affected countries
 - Three bars for each country representing total cases, deaths, and recoveries
 - Countries selected based on the highest total confirmed cases
 - Clear labeling and a consistent color scheme for each metric
5. **Recovery Rate Analysis Dashboard**
 - Visualization of top N countries based on recovery percentage
 - Adjustable threshold for minimum number of cases to be included
 - Clear presentation of recovery rate calculation methodology
 - Sorting functionality to arrange countries by the recovery rate
6. **Country-Specific Pie Chart**
 - Pie chart for the proportion of deaths and recoveries for a selected country
 - The total of deaths and recoveries represented as %
 - Clear labeling of each segment with percentages
 - Updates dynamically when the user selects different countries
7. **Basic Filtering and Selection Controls**
 - Country/region selector with search functionality
 - Metric selector (confirmed cases, deaths, recoveries)
 - Clear user interface elements for interacting with visualizations
 - Consistent layout and design across all visualizations

Optional Features

Our primary focus will be on implementing the must-haves, the following features would enhance the project if time permits:

1. Enhanced Interactive Elements

- Hover functionality showing detailed statistics when hovering data points
- Tooltips with comprehensive information about specific data points
- Click interactions that update multiple visualizations simultaneously
- Interactive legends that allow filtering by clicking on legend items

2. Wave Analysis Visualization

- Identification and highlighting of pandemic waves across different regions
- Comparison of wave intensity, duration, and timing between countries
- Annotations indicating potential causal factors for different waves

3. First Onset Visualization

- Visualization showing when COVID-19 first appeared in different regions
- For US data: visualization of first onset dates by state
- For global data: visualization of spread patterns from initial outbreaks

4. Doubling Time Heatmap

- Color-coded grid showing how quickly cases doubled in different regions
- Ability to track changes in doubling time as the pandemic progressed
- Helps identify where the spread was accelerating or decelerating

5. Mobile-Optimized Interface

- Responsive design tailored explicitly for smartphone and tablet viewing
- Touch-friendly controls for all interactive elements
- Simplified views for smaller screens

6. Data Export Functionality

- Option to download visualized data in CSV or JSON format
- Ability to save or share specific visualization states
- Export high-resolution images of visualizations

7. Informational Overlays

- Optional educational content explaining epidemiological concepts
- Contextual information about major pandemic events
- Methodology explanations for calculations like recovery rates

Project Schedule

Our project timeline spans 10 weeks with key milestones as follows:

Week 1 (March 3-9)

- Setup GitHub repository and project structure
- Explore and understand the COVID-19 datasets
- Initial data exploration and preprocessing

Week 2 (March 10-16)

- Implement core data processing functions
- Set up the D3.js framework
- Begin initial visualization components

Week 3 (March 17-23)

- Implement time series, choropleth map, and bar chart visualizations
- Create a basic user interface for visualization selection

Week 4 (March 24-30)

- Develop remaining visualizations (top countries, recovery analysis, pie chart)
- Integrate visualizations into a cohesive interface

Week 5 (March 31-April 6)

- Project milestone preparation and presentation
- Implement feedback from milestone review

Week 6 (April 7-13)

- Refine visualizations based on feedback
- Performance optimization and testing
- Begin implementing optional features

Week 7 (April 14-20)

- Continue optional feature implementation
- User testing and refinement

Week 8 (April 21-27)

- Final UI polish and optimization
- Complete documentation

Week 9 (April 28-May 4)

- Create project screencast
- Final testing and quality assurance

Week 10 (May 5-11)

- Final submission
- Group feedback and retrospective

Task Distribution:

- Hemangani Nagarajan
 - Data Processing and Time Series Visualization
- Kavisha Parikh
 - Geographic Visualizations and Integration
- Varshini Venkataraman
 - Analytics Visualizations and Testing

Related Work

Our project was inspired by several well-known COVID-19 visualizations and research studies that helped shape our design and goals:

1. Johns Hopkins University COVID-19 Dashboard

Link: <https://coronavirus.jhu.edu/map.html>

This was one of the first and most widely used dashboards during the pandemic. It provided real-time data on COVID-19 cases across the world. We followed its approach to show data over time and across regions, but added more interactive features and comparison tools in our visualizations.

2. The New York Times COVID-19 Tracker

Link: <https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>

This tracker showed detailed data at the state and county levels in the U.S. It influenced our U.S. time series chart, especially in how we presented trends for individual states and made comparisons easier to understand.

3. Our World in Data: COVID-19 Visualizations

Link: <https://ourworldindata.org/coronavirus>

This site offers interactive charts with a global focus, including time series and per-capita views. It helped guide the design of our global line chart (by Kavisha Parikh), especially for comparing trends between countries over time.

4. COVID-19 Data Visualization Using D3.js (Kim et al., ACM CHI 2021)

This paper showed how D3.js can be used to build powerful and interactive pandemic visualizations. It gave us ideas for how to structure our charts, make them interactive, and keep them scalable even with large datasets.

5. The COVID-19 Mobility Network Model (Chang et al., Nature 2021)

Link: <https://www.nature.com/articles/s41586-020-2923-3>

Although this study focused on mobility patterns, it showed how detailed location and time data can reveal hidden patterns in virus transmission. This inspired us to include geographic context through a choropleth map (planned for future work).

6. Class Resources and CICS 571 Visual Examples

We also used examples discussed in class, like Hans Rosling's Gapminder visualizations and the CDC COVID-19 Data Tracker. These helped us understand the importance of clear, interactive, and meaningful visuals in public health communication.

Design Evolution

Visualizations Considered and Implemented

1. Time Series Line Chart for total confirmed cases

- Use of position along a common scale for clear trend comparison across time. This adheres to Cleveland & McGill's principle of using the most accurate visual encoding for continuous data.

2. Bar Chart of Confirmed Cases by Country

- Sorted bars provide easy ranking, and vertical layout optimizes space for long country names.
- Equivalent to the "Top N Countries Clustered Column Chart"

3. Global Choropleth Map

- Use of hue and intensity to indicate death count; spatial encoding makes geographic patterns intuitive. Color choice aligns with standard choropleth color ramps.

4. Recovery vs. Death Pie Chart

- Captures part-to-whole relationships using area, though less precise than bar charts—still effective for high-level understanding.

Visualizations changed or not Implemented

1. Clustered Column Chart with All Three Metrics (not visualized)

- We focused on single-metric bar charts rather than multi-metric clusters.
- This simplification improves readability and avoids visual clutter, aligning with Tufte's principle of maximizing data-ink ratio.

2. Death v/s recovery rate Stacked bar chart

- To know the reason for this change, read the Note mentioned below.
- Vertical bars and sorting allow accurate comparison of the death rate.

Implementation

Our interactive web-based visualizations were developed using **D3.js**, with data preprocessing handled in **MS Excel**. Each visualization was designed to explore different dimensions of the COVID-19 pandemic, including spread, severity, recovery trends, and regional comparisons. Below, we describe the purpose and functionality of each implemented visualization, along with screenshots illustrating their key features.

1. Time Series Line Chart (Country Comparison)

Intent:

To track the evolution of COVID-19 confirmed cases in selected countries over time and compare how different nations experienced waves of infection.

Functionality:

- Users can select one or more countries using a world map.
- Clicking a country in the world map updates the chart, and "reset" resets the view.
- Lines are colored distinctly to differentiate each country.
- The chart provides precise temporal tracking and trend comparison.

Design Considerations:

- Position along a common scale (y-axis: cases) ensures accurate trend comparison.
- Color coding improves readability when multiple countries are selected.

2. Global Bar Chart (Confirmed Cases by Country)

Intent:

To highlight the top countries most impacted by COVID-19 based on confirmed cases.

Functionality:

- Users can choose the data type (Confirmed Cases, Deaths).
- Option to display top N countries (e.g., Top 10, Top 20).
- “Normalize by Population” allows per capita comparison.
- The interactive date slider lets users filter by custom time ranges.

Design Considerations:

- Sorted vertical bars help viewers quickly identify relative magnitudes.
- The date above the slider mentions the exact date range used by dragging the interactive date slider.

3. Global Choropleth Map (Time Series)

Intent:

To present a spatial view of COVID-19 severity across the world for any given date, enabling geographical pattern recognition.

Functionality:

- Users can select a metric (Confirmed Cases, Deaths).
- A date slider animates progression over time.
- Darker shades indicate higher counts for the selected metric.

Design Considerations:

- Color intensity encodes severity, leveraging intuitive visual associations.
- The temporal slider enables exploration of pandemic evolution over time.

4. Death Rate Dashboard (Bar Chart)

Intent:

To identify the countries where the most number of deaths relative to total cases occurred.

Functionality:

- Displays the **top N countries** based on death rate in a selected year.
- Users can set a **minimum confirmed case threshold** to filter unreliable data.
- Sorting options enable ascending or descending views.

Design Considerations:

- Bar length is used for recovery rate comparison—a highly accurate encoding.
- Hover tooltips display exact values of percentage, total cases and recoveries for each country.

5. Recovery vs. Death Pie Chart by Country:

Intent:

Offers a quick proportional comparison of outcomes (deaths v/s recoveries)

Functionality:

- Users can select one country using a dropdown menu.
- User can select the year for which he wants to view the information.

Design Considerations:

- Hover tooltips display exact values of percentage for the sector of pie chart over which it is hovering

6. 6-Month Comparative Bar Chart

Intent:

Enables understanding of how the pandemic evolved over major phases.

Functionality:

- Users can select one country using a dropdown menu.
- Users can select a metric (Confirmed Cases, Deaths).

Design Considerations:

- Hover tooltips display exact values of number of cases for the bar over which it is hovering

Evaluation

Insights Gained from the Visualizations

Through our interactive visualizations, we uncovered several valuable insights about the COVID-19 pandemic:

- **Temporal Spread:** The time series charts clearly revealed how different countries experienced infection peaks at different times. For instance, Afghanistan had a more prolonged and erratic pattern, while Ireland showed distinct wave-like behavior—highlighting how geographic, policy, and healthcare differences impacted the pandemic's timeline.
- **Geographic Impact:** The choropleth map effectively highlighted global hotspots at various points in time. We observed that countries like the US, India, and Brazil consistently remained high in cumulative deaths and cases, while some smaller nations showed spikes relative to population size.
- **Death Trends:** The death rate dashboard exposed interesting disparities. Some countries with high rates in one year would not be present in list of top 20 in the subsequent years. This indicates robust healthcare responses by these countries.
- **Comparative Patterns:** The global bar chart and 6-month period chart helped track how the pandemic evolved over phases. For instance, we noticed that the second half of 2020 saw sharp increases in many regions, correlating with the emergence of new variants and fluctuating lockdown policies.

Answering the Research Questions

1. How did COVID-19 spread across different regions over time?

The time series and choropleth map visualizations made the spread and timing of infection waves clearly visible across countries and continents.

2. What were the mortality and recovery patterns across regions?

The pie charts and death-recovery dashboard allowed us to compare deaths and recoveries in both relative and absolute terms.

3. When did the spread slow down in various regions?

Using time series curves, we identified plateau phases and declining trends in countries like Ireland and Japan.

4. How did different waves of infection compare?

Multiple peaks in the line charts corresponded to different pandemic waves, confirming their timing and intensity in specific countries.

5. What are the relationships between confirmed cases and deaths?

Our multi-chart setup, especially the death-recovery stacked bar dashboard, helped establish correlations and disparities between metrics.

Effectiveness and Limitations

Our visualizations successfully provided an engaging, interactive way to explore the pandemic's trends. Users could slice the data by time, region, and metric, which offered flexibility and deeper insight. The clear layout and consistent design helped users focus on the data rather than navigating complexity.

However, we also encountered several **limitations**:

- **Inconsistent Recovery Data:** As noted in the JHU dataset, recovery data was not uniformly reported across countries. We had to change several of our visualizations to accommodate this limitation.
- **Performance Lag:** As the dataset was large and included global time-series data, some visualizations (especially the choropleth map and top N bar charts) had minor lags during updates, especially when filtering large date ranges.
- **Limited Multi-Metric Charts:** We chose to simplify many visualizations by focusing on one metric at a time. While this improved clarity, it also meant losing out on more nuanced, side-by-side comparisons of cases, deaths, and recoveries in the same view. Especially since we had to remove the global recovery dataset.
- **Mobile Responsiveness:** While we prioritized desktop usability, we did not fully optimize the interface for mobile or tablet viewing—an area of potential improvement for future iterations.

Future Improvements

- Incorporate **annotations** on major pandemic events (e.g., vaccine rollouts, variant emergence) for context.
- Add **wave detection** and visual overlays to make wave identification automatic and clearer.
- Implement **export and share functionality**, allowing users to save filtered views or download insights.
- Explore **animation** or storytelling features to guide users through significant patterns interactively.