Process Book

Project Title: COVIDStat Visualizer - Interactive COVID-19 Insights for

EU/EEA

Names, UIDs, and e-mails:

• Hima Mynampaty, u1528521, <u>u1528521@utah.edu</u>

Praneeth Chavva: u1465506, u1465506@utah.edu

• Hemasundar Tatipudi: u1465343, <u>u1465343@utah.edu</u>

GitHub Repository Link

Background and Motivation

The COVID-19 pandemic has significantly impacted all regions of the world, and understanding the differences in outcomes across countries can provide valuable insights for policy decisions and future health crises. We chose to concentrate on the European Union (EU) and the European Economic Area (EEA) due to the region's diverse pandemic responses, which includes a variety of lockdown measures, vaccination strategies, and testing protocols. This project combines our interests in public health, data visualization, and interactive technologies to create a valuable tool for analyzing COVID-19 trends in these regions. Furthermore, we are motivated by the tool's ability to inform policymakers and health professionals about the efficacy of interventions in mitigating the virus's impact.

Project Objectives

Our primary goal is to develop an interactive dashboard that provides insightful and easily accessible COVID-19 data visualizations for the EU/EEA region.

The main questions we want to answer with this visualization are:

 How have COVID-19 cases and death rates varied over time in different EU/EEA countries?

- What effect did vaccination campaigns have on reducing cases and deaths in the region?
- How do testing and positivity rates correlate with case and death trends in various countries?
- Which countries responded most effectively, and what factors contributed to their success (e.g., vaccination rates, testing strategies, and government responses)?
- How did the pandemic's waves (e.g., first wave, post-vaccine era) differ regarding case surges and control strategies?

What We Aim to Learn and Accomplish

We want to get a better understanding of the dynamics of the COVID-19 pandemic in the EU/EEA region by

- Graphing trends in cases, deaths, vaccinations, and testing rates across countries.
- We are comparing the results of different public health interventions and vaccination campaigns.
- Identifying patterns in pandemic progression and highlighting effective mitigation strategies.

Benefits:

- Inform Policy and Research: Provide clear, data-driven insights that will help policymakers and researchers understand the effectiveness of various response measures.
- 2. **Public Awareness:** Help the general public visualize the pandemic's impact in their own and neighboring countries, fostering a better understanding of global health dynamics.
- 3. **Decision-Making:** Provide actionable insights for future pandemic preparedness by emphasizing the links between vaccination efforts, testing rates, and case reductions.
- 4. **Comparative Insights:** Allow users to compare countries and identify best practices that resulted in lower case and death rates, potentially informing future strategies.

By answering these questions and meeting our goals, we will create a tool that improves understanding of the pandemic's impact in the EU/EEA and facilitates better decision-making for public health responses.

Data

We will collect data from a publicly available source called European Centre for Disease Prevention and Control (ECDC) for detailed EU/EEA statistics.

Data Sources

- Data on 14-day notification rate of new COVID-19 cases and deaths
- Data on COVID-19 vaccination in the EU/EEA
- Data on testing for COVID-19 by week and country
- Data on country response measures to COVID-19

Data Processing

We anticipate extensive data cleanup to ensure the datasets are accurate, complete, and ready for visualization.

The main tasks for data processing are:

- 1. <u>Filter by region (EU/EEA):</u> Because the project is specifically focused on the EU/EEA region, we will filter the data to include only countries from that region. Non-EU/EEA countries will be removed from the datasets to ensure that we are working with accurate geographic data.
- 2. <u>Handling Missing Data</u>: Several columns, including "tests_done," "positivity_rate," and "vaccination rates," contain missing or incomplete entries. We will close these gaps by either imputing missing values using statistical methods (for example, forward or backward filling based on trends) or filtering out incomplete records.
- 3. **Standardizing Date Formats:** The datasets include time-based fields (e.g., "year_week," "date_start") that must be standardized for consistent

- time-series analysis. We will convert these fields to a standard date format to facilitate filtering and visualization.
- 4. <u>Aggregating Data:</u> To calculate summary statistics, we will aggregate data by country and over specific time periods (e.g., weekly or monthly). Total cases and fatalities by country. Vaccination rates vary by population segment and country. Testing and positivity rates vary over time. Cumulative data (e.g., total cases, deaths, and vaccinations per capita).
- 5. <u>Merging Datasets:</u> To ensure that all relevant data is available in a single view, we will combine datasets using common fields such as country, country code, and time period (e.g., year_week). This will enable us to visualize multiple metrics (cases, deaths, vaccinations, and testing) simultaneously.

Quantities to Derive:

- 14-day case and death rates per 100,000 population for easier country-tocountry comparisons.
- Vaccination coverage rates by age group and risk category.
- **Testing and positivity rates** for evaluating testing effectiveness.
- Case-to-vaccination correlation metrics, showing how vaccination efforts correlate with the reduction in case counts and deaths.

Implementation:

We will use Python for data processing, utilizing libraries like Pandas for data manipulation, merging, and cleaning. The workflow will include:

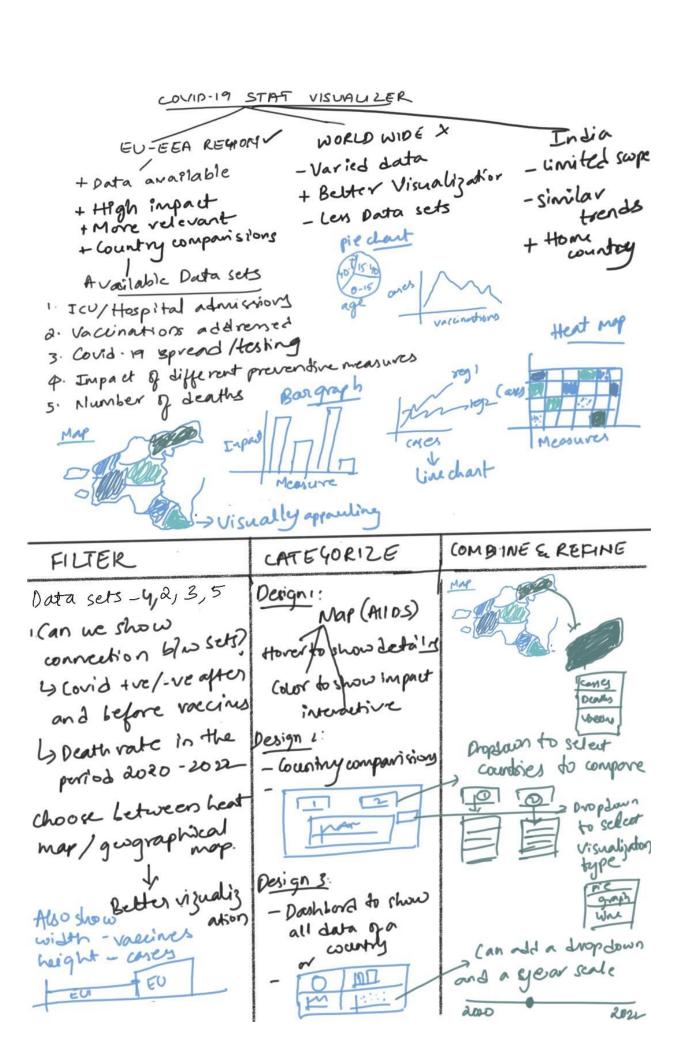
- Loading datasets into dataframes.
- Cleaning and imputing missing values using methods such as forward fill, backward fill, or mean imputation.
- Aggregating and transforming data to calculate derived metrics like case rates, death rates, and vaccination percentages.
- Merging the datasets on common keys to create a unified dataset for visualization.

Visualization Design:

We used the Five Design Sheet Methodology to discuss, design, weigh advantages and disadvantages to come up with a final realization.

Brainstorm: Our aim from the beginning is to visualize COVID-19 data in an interactive and useful way. The main conflict was to decide on which region to focus on. We considered doing it for the whole world, India and the EU/EEA region. We concluded on EU/EEA due to different positives mentioned in our Ideation sheet below.

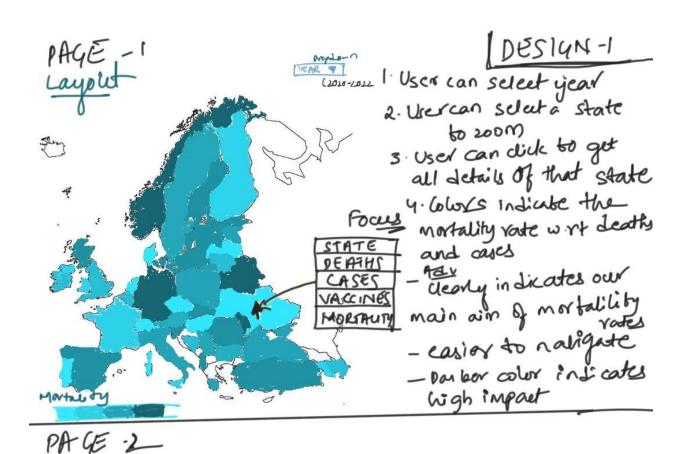
The discussion continued to decide on which data sets were available and which data sets to use. Later we came up with very rough sketches on how we could represent the data from the data sets. As there were many ideas, we focussed on filtering the data sets, categorizing the data, coming up with designs to represent the data chosen and finally combining them all in a refined way.

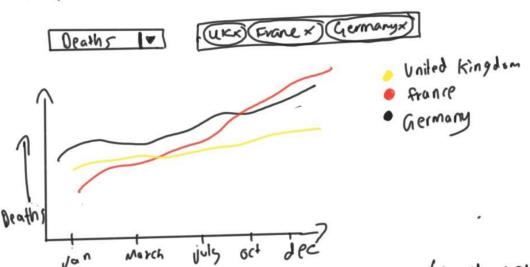


Design 1: As we had multiple ideas and things to represent in the web application, we have decided to do 3 screens of visualizations.

Page1 is our main screen with a EU/EEA region map with sequential color pallet, where colors represent the mortality rate (number of deaths in relation to number of positive cases after normalizing). Page2 is a screen used to compare derived data from the data sets available between different states. Page 3 is a dashboard with all data at a single place of a particular state.

We have discussed how our screens would look, what are our major focus areas, what are the features available for users, advantages and disadvantages of all three screens.



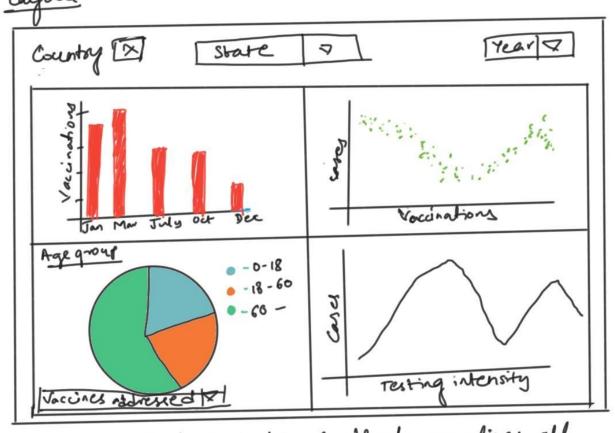


1. User can select the parameter of type (death, cases etc) 2. User can choose the states he wants to compare

3. Each country data is represented by a different color in a line graph

- Comparing data (time series) from different lines is easier and convinient

- Drop to way provide a way to dynamically choose options



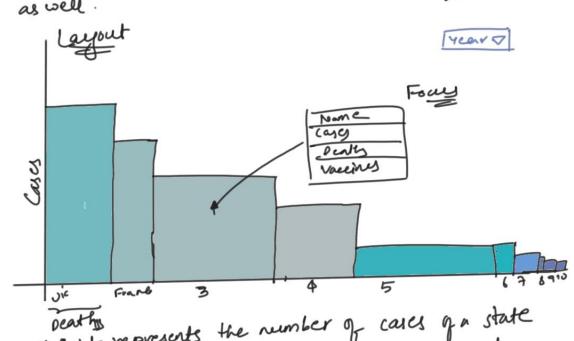
- 1. A consolodated dashbord, that visualizes all trends with available Lata sets in a particular state in a year
- 2 Bar graph represents the vaccination intensity in
- 3. Scatter plot represents vaccinations given in relation
- 4. he dark represents the affect of different factors
- 5 The bone graph represents the cases count in ansociation with testing intensity
- tre gires à single dashbord à analyse various trends in a state to be prepared

Design2: We thought of different ways we could represent the main screen. One of the main competitors to the map was the bar graph below as it would allow us to show three features of the data(cases, mortality and vaccinations) as the main focus, without having to hover/click on the bars.

But the design had its own disadvantages which are discussed in the sheet.

Plot 1 canbe represented as boar grouphs as well.

DES14N-2



- Height represents the number of cases of a state - with represents the number of deaths sue to

- Color represents the intensity of vaccination done

- Wicking on a bar gives more details

the more values can be represented at once

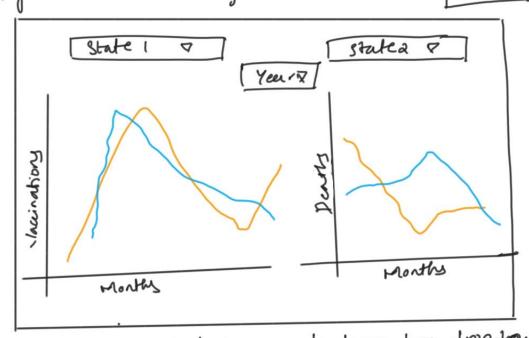
-ve states with lessen impact are not very visible

- ve Navigating a map is easier for end users

- ve Visually not as good as a map

Design3: Instead of representing in a single line graph, initially we had an idea of comparing only two states using two dropdowns to select the state and two different graphs representing two different metrics.

But the main flaw is the constraint of having to select only two states. Other points are weighed as well.



- One of our initial design was to have two dropdowns to select two states of the country

- Two graphs / scatter plots representing two different

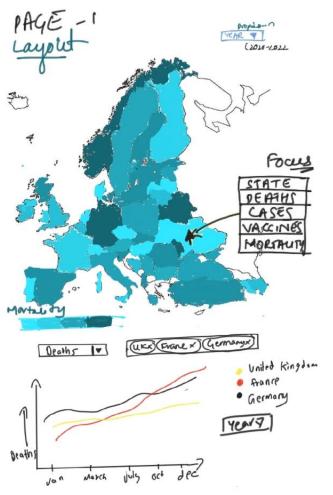
the Basier to understand trends between two states tre less complexity during implementation

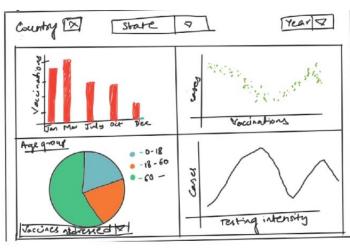
- ve No option to compare multiple states at once

-re Dashboard can be used to see multiple graphs/ trends at the same time

Realization: We debated on which designs to incorporate and concluded on using the ones that best fit a users requirement and ease of navigating the visualization.

The key data are being used, navigations and selections that a user can perform, advantages of this visualization, time range of our visualization are discussed and we came to a conclusion!





REALISATION

- -All data is represented in the map
- selecting multiple state to compare data
- scleding which Lata to compare
- Dahbord to represent wrielation between netric data
- Fearibility to relect different years

operations/

- -Years range 2020 fo
- Data sets acquired from EU/EEA regions
- Scale for a year-12 months
- -Mortality rate in relation to normalize cases and deaths
- Colors used in fovor
- Graphs chosen based on readability

Must Have Features:

Interactive Dashboard

- EU/EEA Overview Map: An interactive map highlighting COVID-19 statistics by country with choropleth visualization.
- Country Filters: Options to select and filter country for specific visualizations.
- Time Frame Filters: Controls to filter data by specific time periods (e.g., prevaccination, post-vaccination or based on a specific wave) and/or years.
- Key Metrics Visualization: Use of charts and plots to display trends in cases, deaths, testing, and vaccination rates.
- Tooltips and Pop-ups: Hover-over tooltips and pop-up windows providing detailed information for selected data points.

Comparative Analysis

- Country Comparison Charts: Line and bar charts comparing COVID-19 trends and responses between two countries.
- Case-to-Vaccination Correlation Metrics: Visualization of the correlation between vaccination rates and case reductions.

Visualization for Vaccination

- Vaccination Coverage Charts: Pie charts and bar graphs displaying vaccination coverage rates by age group and risk category.
- Vaccination Impact Analysis: Scatter plots showing the relationship between vaccination coverage and case/death rates, with trendlines for correlation analysis.

Optional Features

Advanced Filtering

- Custom Date Ranges: Users can define custom date ranges for more specific time frame analyses.
- Data Download Option: Ability for users to download raw data or visualizations for offline analysis.

Comparing Past Data

• Pandemic Waves Visualization: Detailed comparisons of different pandemic waves and their impact on cases, deaths, and responses.

Predictive Analytics

- Trend Forecasting: Implementation of predictive models to forecast future trends based on historical data.
- Scenario Analysis: Visualization of potential future scenarios based on different vaccination and testing strategies.

User Customization

- Customizable Dashboards: Allow users to customize their dashboard view by selecting which metrics to display and how to arrange them.
- Saved Views: Feature for users to save and revisit their preferred dashboard configurations.

Project Schedule

Week 1 (Sept 13)

- Announce the Project: Decide and announce the project.\
- Project Proposal: Begin the process book by outlining the background, goals, and schedule.
- Finalize project requirements and objectives.

Week 2 (Sept 20)

- Verify data from the European Centre for Disease Prevention and Control (ECDC).
- Set up the project repository and initial environment.
- Start data cleaning and preprocessing.
- **Project Review**: Discuss the project with the instructor, get feedback, and explore potential improvements or ideas.

Week 3 (Sept 27)\

- Handle missing data and standardize date formats.
- Filter non-EU/EEA countries from the datasets.
- Handle inconsistent data and perform necessary data aggregations.

Week 4 (Oct 4)

- Merge datasets by country and time period.
- Complete feature engineering for derived features.
- Validate the prepared data and check for any inconsistencies.
- Perform exploratory analysis to summarize statistical data and generate simple plots.

Week 5 (Oct 11)

- Build prototypes for charts and plots to be used in the visualizations.
- Develop a choropleth map with EU/EEA country-level stats.

Week 6 (Oct 18)

- Add additional functionalities and filters to enhance comparisons and visualizations.
- Optimize data handling to ensure smooth dashboard performance.

Week 7 (Oct 25)

- Complete the visualization dashboards, including all required charts and plots.
- Consider adding any optional features to enhance the dashboards.
- Project Milestone: Submit the code, process book, and prototype of the working visualizations.

Week 8 (Nov 1)

- Test and debug visualizations on different screen sizes and with various data types.
- Improve interactivity with tooltips and hover transitions.
- Peer Feedback: Gather feedback from peers to identify areas for improvement.

Week 9 (Nov 8)

- Refine the visualizations and user interface based on feedback.
- Perform thorough testing for bugs and performance issues.

Week 10 (Nov 15)

Develop optional features to make the dashboard more informative.

Week 11 (Nov 22)

- Finalize documentation and create user guides for navigating the project.
- Project Screencast Submission: Submit a screencast video showcasing the project.

Week 12 (Nov 29)

- Prepare the dashboard for deployment.
- Complete the final process book.
- Review the project internally to ensure all areas of project are thorough.

Week 13 (Dec 6)

- Validate the final code, data, process book, and README file for the GitHub repository.
- Create a final release on GitHub repository and upload the screencast video.
- Final Project Submission: Hand-in the final project.

Team responsibilities

- Hima Mynampaty: Line charts, comparative analysis, and scatter plots.
- Praneeth Chavva: Data processing, map-based visualization development.
- Hemasundar Tatipudi: Dashboard integration, filters, and interactivity.

FEEDBACK ON INITIAL PROPOSAL FROM MEETING WITH TA PROMA

(review meeting on 09.19.2024, entry by Praneeth Chavva)

On 09.19.2024, we met with TA Proma regarding our initial project proposal for the "COVIDStat Visualizer." Below are the key points and takeaways from the discussion:

- Project Focus: Proma advised us to refine the project's primary focus. Instead of addressing several aspects (case rates, death rates, vaccination campaigns, and positivity rates), we should focus on one key aspect. For example, we might concentrate on how vaccination efforts influenced case and death rates across EU/EEA countries.
- Simplicity and Storytelling: Proma emphasized that the visualizations should tell a story. Rather than overwhelming users with multiple data points, the project should guide users through a logical flow—starting with foundational metrics and progressing toward key insights. This approach will help users better understand the data and its implications.
- 3. Visualization Refinement: It was suggested that we reduce the number of visualizations and focus on those most critical to telling the story. An interactive map showing vaccination rates and a line chart illustrating case trends over time could suffice to present the data effectively.
- 4. Next Steps: We must revise the proposal to reflect this simplified approach and create a prototype that aligns with the storytelling method. Our next meeting will involve reviewing these adjustments.
- 5. Upcoming Deadlines: Prioritize data cleaning and developing the initial visualizations, ensuring they follow a narrative structure. Revised visualizations and mockups should be ready for the following review.

Updated Project Plan

Following TA Proma's feedback, we've refined the project to emphasize a clear, story-driven approach. This updated plan narrows the focus to key aspects of COVID-19 waves and government interventions across EU/EEA countries, using four main visualizations to guide users through a cohesive narrative with a polished, user-friendly design.

Project Scope & Direction

Our goal remains the same: to explore the impact of government interventions on COVID-19 cases throughout the pandemic. To ensure clarity, we'll focus on distinct pandemic phases and the effectiveness of responses, structuring visualizations to show the progression through each wave. Users will gain insights into how interventions influenced the spread and severity of COVID-19.

<u>Visualization Design and Ideas</u>

The design concepts from the original proposal remain intact. We continue to use interactive elements like timelines, bar charts, and line graphs to show trends and intervention impacts, but these are now organized with a clearer focus to enhance narrative flow.

- 1. First Wave vs. Post-Vaccine Era (Line Graph)
 - Objective: Show contrasting trends in cases and deaths between the prevaccination and the post-vaccine era.
 - Design:
 - Separate line graphs for cases and deaths with each graph containing two lines, color-coded to differentiate pandemic phases.
 - o An X-axis timeline and a Y-axis for case/death counts.
 - Interaction: Users can select countries and view tooltips on hover, providing context for case/death counts by date.

2. Timeline of Government Interventions and Case Peaks

- Objective: Illustrate the timing of interventions and how they align with case peaks.
- Design:
 - An X-axis for the timeline throughout the covid era
 - Y-axis displaying COVID-19 cases, allowing peaks to stand out visually.
 - Visual distinctions for intervention types, such as shapes or colors (e.g., "BanOnAllEvents" vs. "ClosDaycare").
- Interaction: Country selection and hover tooltips to show start and end dates, intervention type, and cases during intervention periods.

3. Comparing Case Reductions Between Waves (Bar Chart)

- Objective: Highlight the impact of the intervention by comparing case reductions across pandemic waves.
- Design:
 - Each country's bars are grouped by pandemic wave (e.g., First Wave, Second Wave), showing reductions visually.
 - o A consistent color scheme across waves is used to enhance comparability.
- Interaction: Country filter and clickable bars to reveal detailed breakdowns of case reductions and interventions.

4. Government Response Stringency vs. Case/Death Trends (Line and Bar Graph)

- Objective: Correlate government response stringency with COVID-19 case/death trends.
- Design
 - Dual Y-axes: one for response stringency (as a line) and another for case/death counts (as bars), sharing a time-based X-axis.
 - Overlay design to make comparisons visually straightforward.
- Interaction: Users can hover to view stringency scores, cases, and death counts by time point, with country selection to see national trends.

Alignment with the Original Proposal

This updated plan preserves our original objectives while following TA Proma's guidance for a focused, user-friendly approach. Although the overall focus is now more streamlined, the core design concepts—interactive maps, timelines, bar charts, and line graphs—remain consistent with our initial proposal. We maintain the intended depth of analysis by keeping these original design ideas, now presented with a clearer, narrative-driven structure. This will allow us to deliver a meaningful, cohesive perspective on COVID-19 waves and responses across EU/EEA countries while enhancing accessibility and insight for users.

Data Preprocessing:

To support this refined narrative, we've conducted extensive preprocessing to ensure data consistency, usability, and relevance for the visualizations.

Cases and Deaths Dataset

- Handling Missing Values: Forward-filled missing values in key columns (`rate_14_day`, `weekly_count`, `cumulative_count`) to maintain continuity.
- Date Conversion: Converted `year_week` to a proper datetime format (assuming Monday as the start of the week).
- Data Integrity Check: Added checks on cumulative cases and corrected any inconsistencies to ensure data accuracy.
- Pandemic Wave Categorization: Segmented data into different pandemic waves for a more structured analysis.
- Population Normalization: Created a `cases_per_million` metric to normalize case numbers by each country's population.
- Time Aggregation: Included monthly and quarterly aggregations to support both granular and broader analyses.
- Data Cleaning: Removed unnecessary columns, retaining only those relevant to our visualizations.
- Sorting: Sorted the data by country and date to ensure consistent chronological order.

Response Graph Dataset

- Country Filtering: Filtered the dataset to include only EU/EEA countries, based on a predefined list.
- Date Conversion: Converted `date_start` and `date_end` into a proper datetime format for time-based analysis.
- Handling Missing `date_end` Values: Filled any missing `date_end` values with the latest available date for continuity.
- Stringency Score Calculation: Added a `stringency_score` to quantify intervention severity, assigning weights between 0 and 1 based on intervention type, with a default weight of 0.2 for unspecified interventions.
- Stringency Score Normalization: Scaled `stringency_score` to a range of 0-100 for consistency across interventions.
- Intervention Duration Calculation: Calculated `intervention_duration` as the number of days between `date_start` and `date_end`.
- Intervention Category Assignment: Classified interventions into broad categories (e.g., `Lockdown`, `Masks`, `Gathering`), assigning "Other" for unspecified interventions.
- Aggregated Stringency Data: Created an aggregated dataset with daily stringency scores by summing scores for each country and date.
- Categorizing Intervention Severity: Divided the stringency scores into severity levels (e.g., Low, Medium, High) for a simplified overview.

Testing Dataset

- Handling Missing Values: Applied forward fill and interpolation for `tests_done` and `positivity_rate`.
- Date Conversion: Converted `year_week` into a date format (assuming Monday as the start of the week).
- Testing Effectiveness Metric: Calculated `testing_effectiveness` as the percentage of `new_cases` relative to `tests_done`.
- Testing Effectiveness Categorization: Classified `testing_effectiveness` into 'High',
 'Medium', and 'Low' using predefined thresholds.
- Rolling Averages: Added 7-day rolling averages for `tests_done` and `positivity_rate` to support smoother time series visualization.
- Data Cleaning: Removed unnecessary columns (e.g., `country_code`, `level`, `region`).
- Sorting: Sorted by `country` and `date` to ensure chronological consistency.

Vaccination Dataset

- Date Conversion: Converted `YearWeekISO` to a date format for chronological analysis.
- Sorting: Sorted by `ReportingCountry` and date to support cumulative calculations.
- Cumulative Vaccination Calculation**: Calculated cumulative vaccinations and created a `CumulativeVaccinationRate` as a population percentage.
- Vaccination Rate (Weekly): Added `VaccinationRateWeekly` to show weekly dose administration.
- Rolling Averages: Added a 7-day rolling average (`VaccinationRate_7d_avg`) for smoother trend visualization.
- Vaccination Milestones: Categorized countries based on vaccination milestones: <25%, 25%+, 50%+, and 75%+ vaccinated.
- Cumulative Validation: Recalculated cumulative doses for consistency, adding a `CumulativeCheck` column for validation.
- Data Cleaning: Dropped unused columns, focusing only on those relevant for visualization.

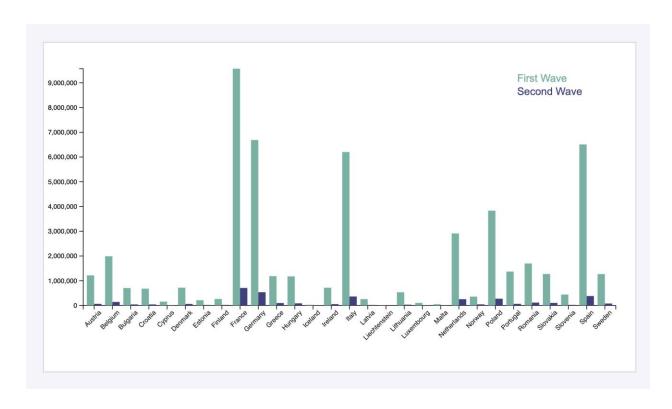
Additionally, we may need to perform some data processing to enhance the clarity and precision of the visualizations.

<u>Progress on Visualization Prototypes for "Pandemic Waves and Government Responses</u>

With our data preprocessing completed, we have begun developing and refining visualization prototypes that form the core of our analysis on COVID-19 trends and government responses. These prototypes use sample subsets of our data, providing initial layouts and basic interactivity. This approach ensures we are on the right track while allowing us to gather feedback and refine before adding advanced features.

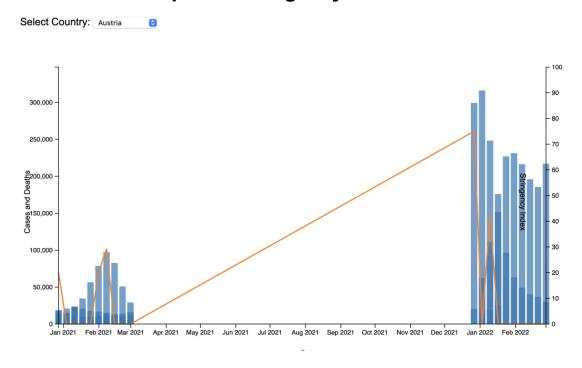
Completed Visualizations

1. COVID-19 Case Reductions Comparison Between Waves



- Overview: This bar chart compares COVID-19 case reductions across different pandemic waves for EU/EEA countries, emphasizing intervention impact over time.
- Features: It includes a basic layout with countries on the X-axis and case reductions on the Y-axis. Each wave data is color-coded to represent specific waves, allowing for easy comparison.
- 2. Government Response Stringency Index vs. Cases and Deaths

Government Response Stringency Index vs. Cases and Deaths

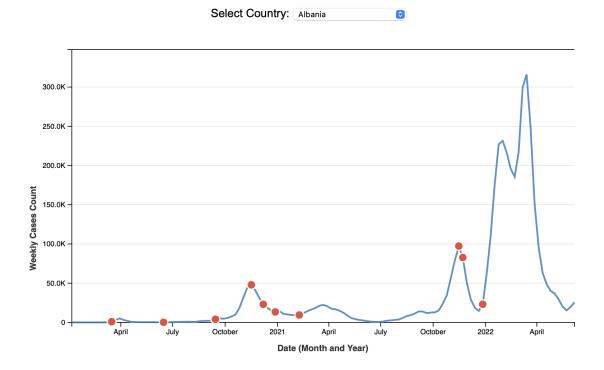


Overview: This visualization overlays government response stringency with COVID-19 cases and deaths over time, providing insights into how varying response levels influenced case and death trends.

Features: A drop-down menu allows users to select specific countries. The chart displays stringency as a line graph, with bars representing cases and deaths. Color-coded elements provide visual cues to differentiate data types.

3. Weekly Cases Line Chart with Response Measures

Weekly Cases and Response Measures by Country

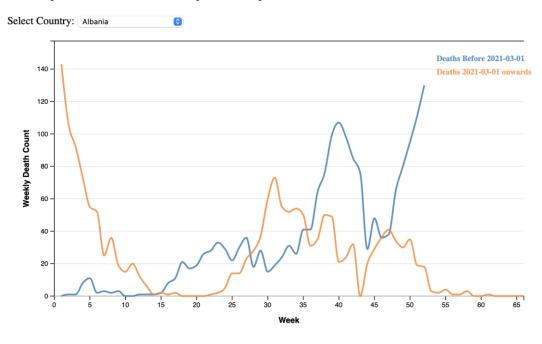


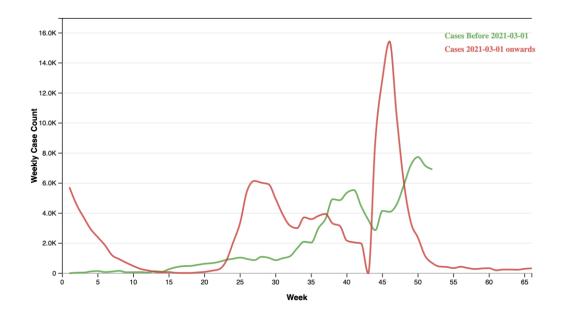
Overview: A line chart showing weekly COVID-19 cases with government intervention markers. This allows users to see how specific responses impacted case trends in each country.

Feature: Users can select a country from a drop-down menu, and intervention measures are highlighted on the timeline to show the correlation with case counts

4. Weekly Deaths and Cases Line Charts

Weekly Deaths and Cases by Country





Overview: This visualization compares weekly COVID-19 cases and deaths across selected countries.

Features: Drop-down options enable country selection, allowing users to focus on specific regions. The separate lines for cases and deaths make it easy to see the effect of pandemic phases on these metrics over time.

Next Steps for Enhanced Visualization

Building on these prototypes, we will implement additional features to improve usability and visual appeal:

Expanded Interactivity: We plan to incorporate hover tooltips, clickable elements, and more filtering options to enable a deeper dive into specific data points and trends.

Animations: Adding animations will help guide users through trends, highlighting changes over time and drawing attention to significant shifts in the data.

Drill-Down Features: For detailed exploration, we'll add drill-down options, allowing users to delve into specific interventions, responses, or time periods.

Stylistic Enhancements: We will standardize colors, fonts, and labels to create a polished, cohesive appearance, enhancing readability and user experience.

This iterative approach, starting with prototypes, ensures that each visualization aligns with our narrative and effectively conveys insights on COVID-19 waves and government responses. By developing these foundations, we are well-prepared to expand on interactivity and presentation, aiming for a final dashboard with precise, insightful analysis.