Project Proposal

Project Title: COVIDStat Visualizer - Interactive COVID-19 Insights for EU/EEA

Names, UIDs, and e-mails:

- Hima Mynampaty, u1528521, u1528521@utah.edu
- Praneeth Chavva: u1465506, u1465506@utah.edu
- Hemasundar Tatipudi: u1465343, u1465343@utah.edu

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:

1. **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 the daily number of new reported COVID-19 cases and deaths by EU/EEA country
- 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. **Filter by region (EU/EEA):** 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. **Handling Missing Data**: 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. **Aggregating Data:** 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. **Merging Datasets:** 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.

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

Quantities to Derive:

- **14-day case and death rates per 100,000 population** for easier country-to-country 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.

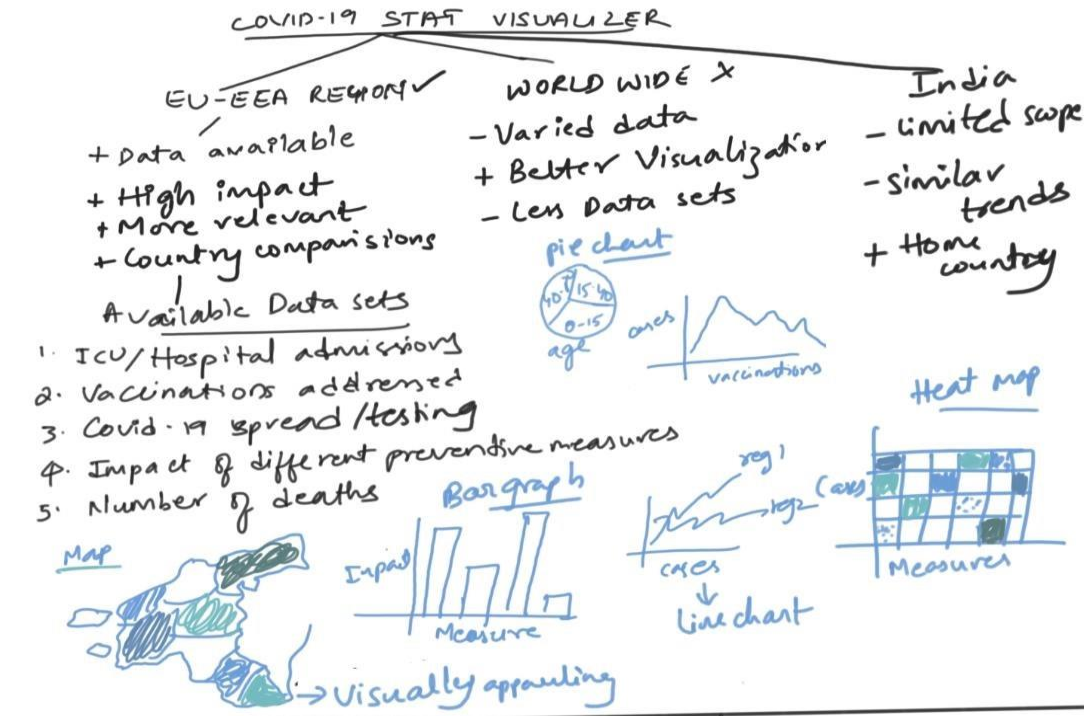
Visualization Design:

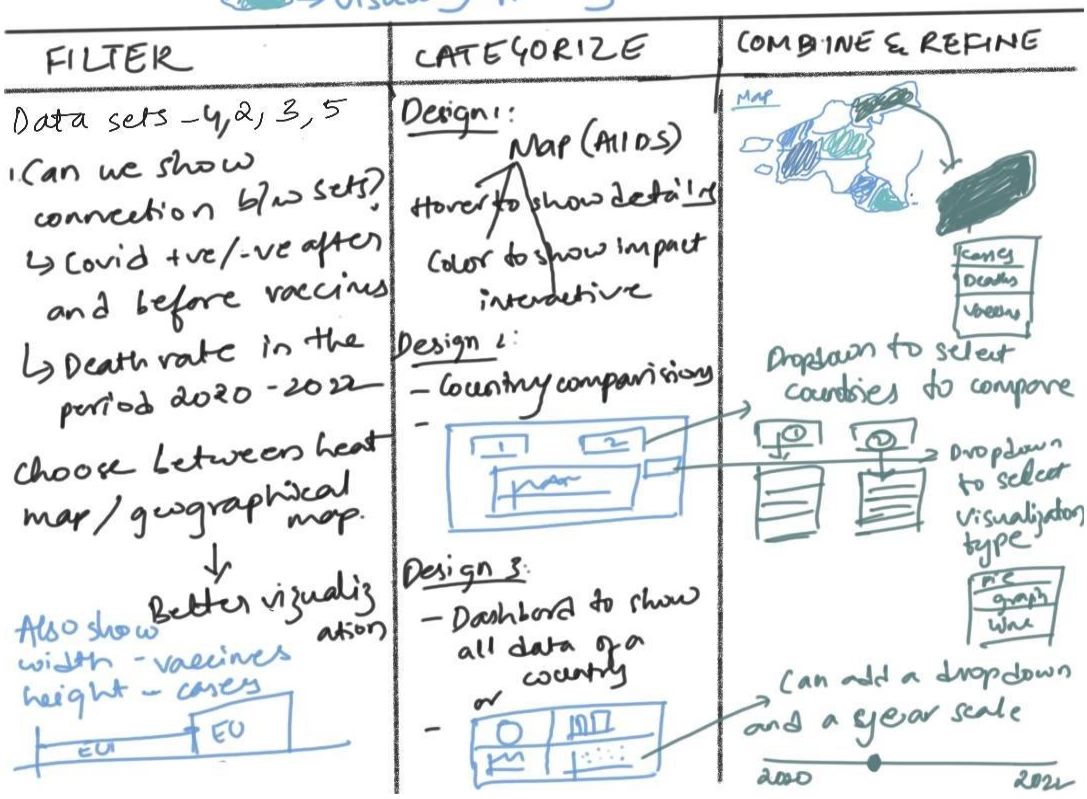
We used the Five Design Sheet Methodology to discuss, design, weigh advantages and disadvantages to come up with a final realization.

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.



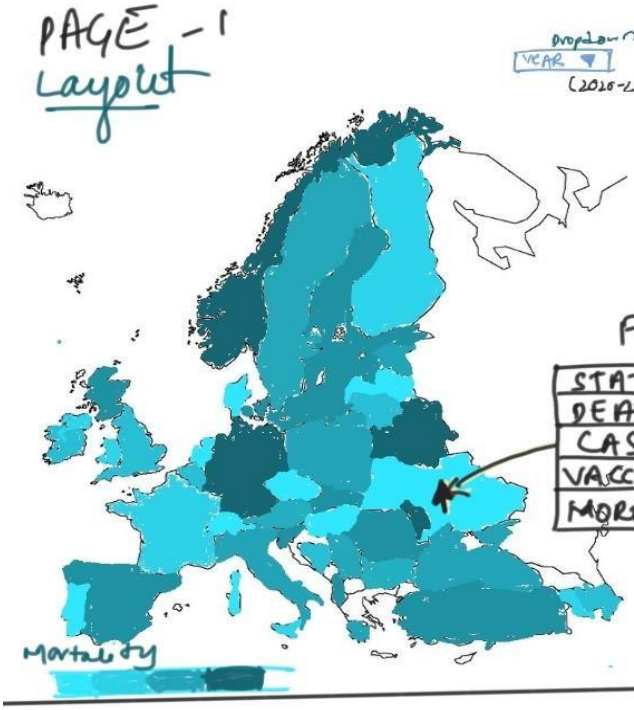


Design 1: As we had multiple ideas and things to represent in the web application, we have decided to do 3 screens of visualizations.

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.

PAGE -1
Layout



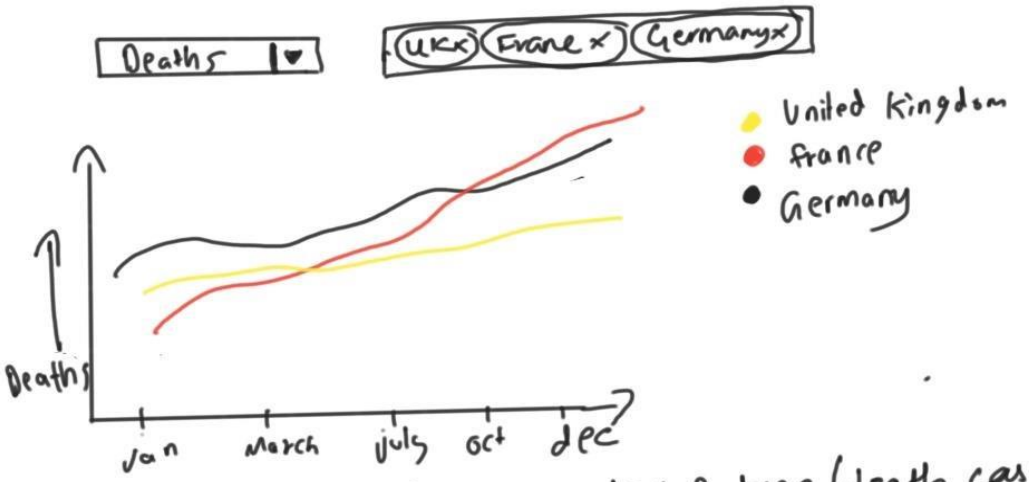
DESIGN -1

1. User can select year
2. User can select a state to zoom
3. User can click to get all details of that state
4. Colors indicate the mortality rate w.r.t deaths and cases

Focus

- Adv
  - clearly indicates our main aim of mortality rates
  - easier to navigate
  - Darker color indicates high impact

PAGE -2



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

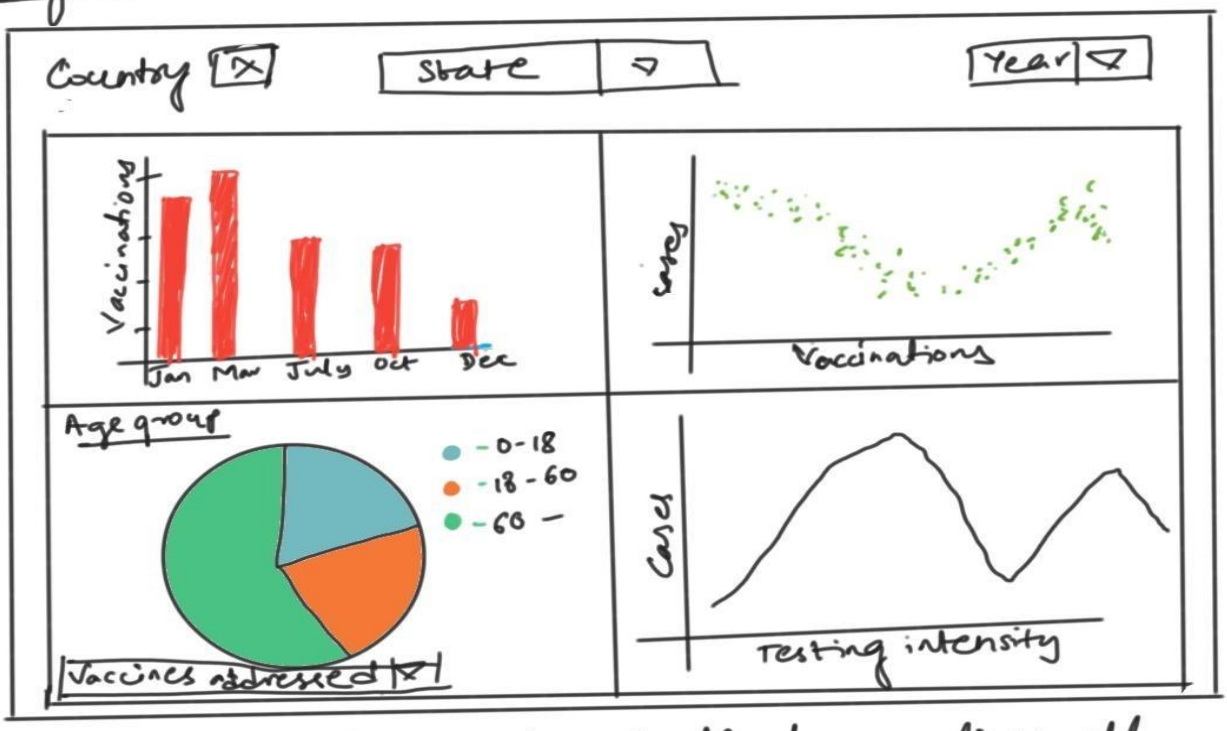
Adv

- Comparing data (time series) from different lines is easier and convinient
- Dropdowns provide a way to dynamically choose options

PAGE-3

DESIGN-I

Layout



1. A consolodated dashbord, that visualizes all trends with available data sets in a particular state in a year
2. Bar graph represents the vaccination intensity in all months
3. Scatter plot represents vaccinations given in relation to total cases in the state
4. Pie chart represents the affect of different factors on age groups
5. The line graph represents the cases count in association with testing intensity

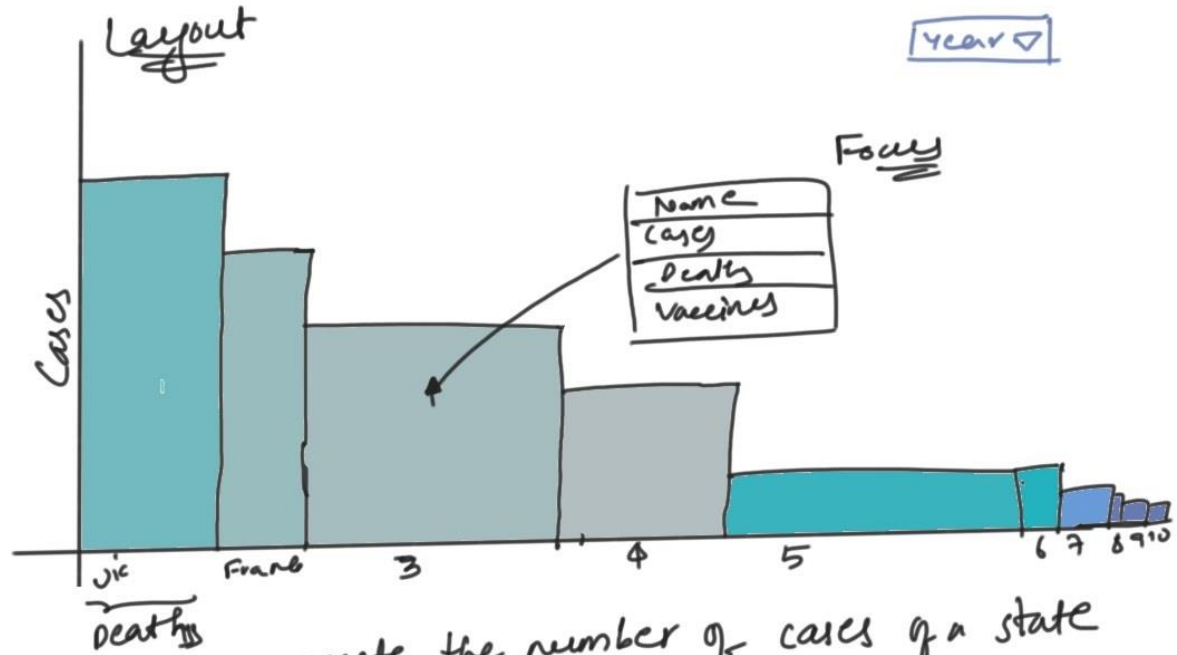
+ve Gives a single dashbord to analyse various trends in a state to be prepared

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 can be represented as bar graphs as well.

DESIGN - 2



- Height represents the number of cases of a state
- Width represents the number of deaths due to covid-19
- Color represents the intensity of vaccination done
- Clicking on a bar gives more details

+ve more values can be represented at once

-ve states with lesser impact are not very visible

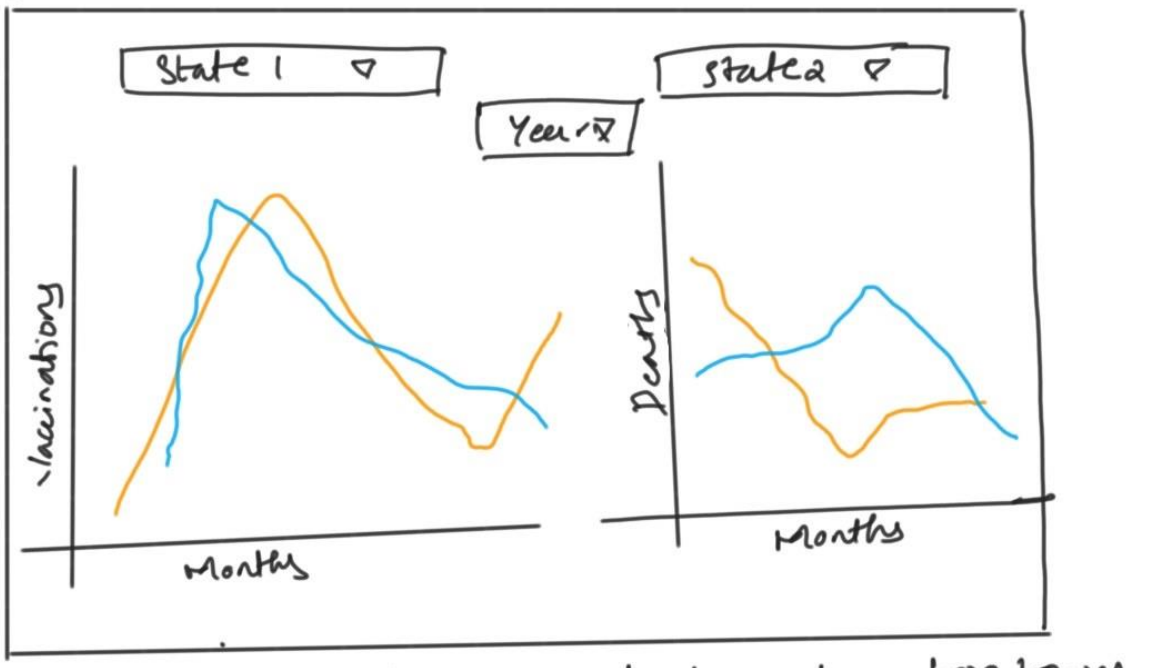
-ve Navigating a map is easier for end users

-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.

Page 2 can also be designed as below [Design-3



- One of our initial design was to have two dropdowns to select two states of the country
- Two graphs / scatter plots representing two different metrics

+ve Easier to understand trends between two states

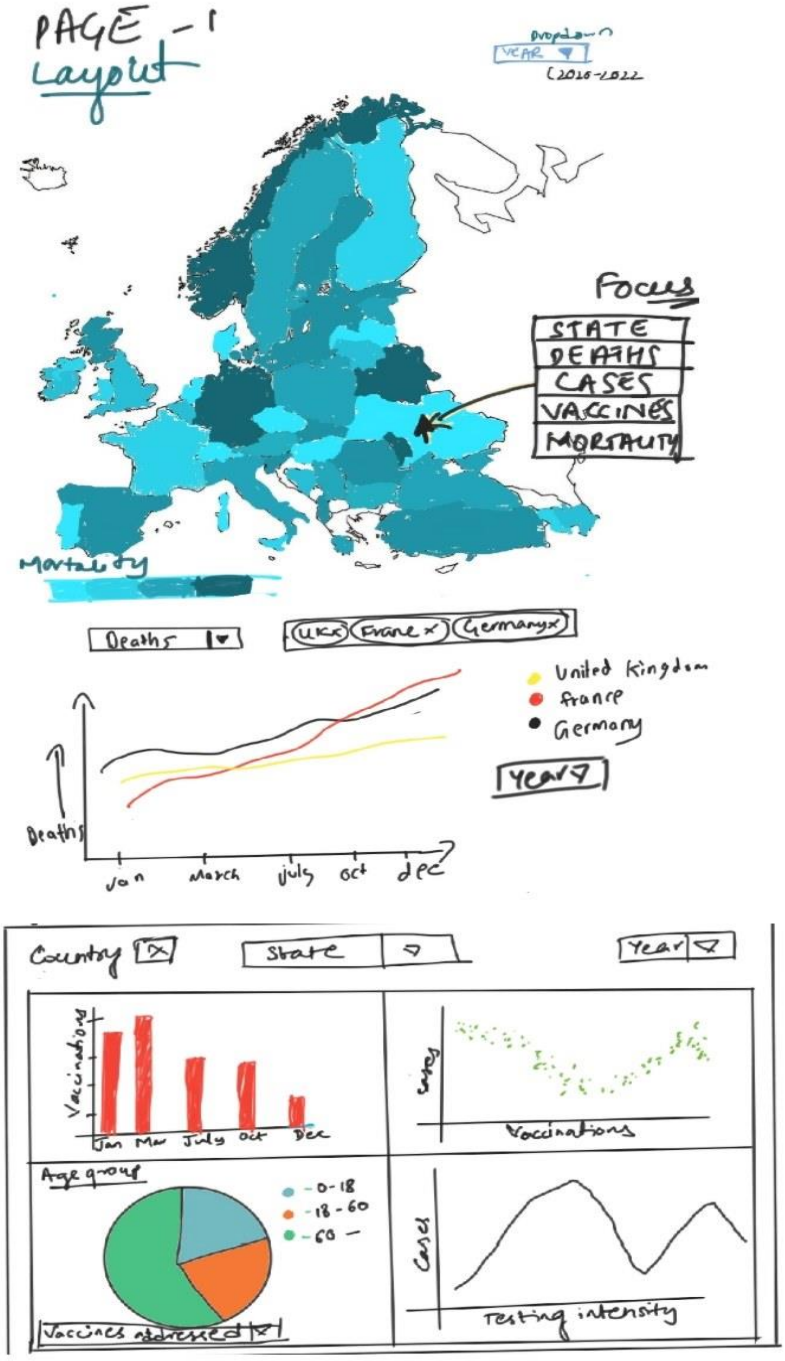
+ve Less complexity during implementation

-ve No option to compare multiple states at once

-ve 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
- selecting which data to compare
- Dashboard to represent correlation between metric data
- Feasibility to select different years

operations

Details

- Years range - 2020 to 2022
- Data sets acquired from EU/EEA regions
- Scale for a year - 12 months
- Mortality rate in relation to normalized cases and deaths
- Colors used in favor of color blindness
- 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., pre-vaccination, 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.