

Project Plan

Step 1: Data Collection

Objectives: - Assemble all necessary datasets from reliable, publicly available sources. - Ensure all datasets are updated, consistent, and cover the target timeframe (2020–2023).

Tasks:

- 1.1. Gather daily confirmed COVID-19 case data from: - Johns Hopkins CSSE COVID-19 Data Repository (via CSV or API). - COVID-19 Data Hub R package if needed.
 - 1.2. Gather vaccination data including: - Total doses administered. - Number of people partially and fully vaccinated. - Vaccination rates per hundred people.
 - 1.3. Assemble ICU healthcare stress indicators: - Number of ICU patients. - ICU bed capacity (if available). - Hospitalization rates.
 - 1.4. Focus countries: - United States - United Kingdom - Canada - China(To be decided) - Singapore(To be decided)
-

Step 2: Data Cleaning and Preprocessing

Objectives: - Prepare datasets for analysis by ensuring consistency, completeness, and usability.

Tasks:

- 2.1. Clean the datasets by: - Removing duplicate rows. - Handling missing values (either by imputation or removal depending on extent). - Standardizing country names and date formats.
 - 2.2. Perform necessary joins/merges: - Combine case counts, vaccination rates, and ICU occupancy into a single unified dataset by **Country** and **Date**.
 - 2.3. Create meaningful derived variables: - Vaccination coverage rate: **% fully vaccinated population**. - Case fatality rate: **(deaths / confirmed cases) * 100**. - ICU usage rate: **(icu_patients / estimated ICU capacity)** if data available.
 - 2.4. Filter datasets: - Focus only on data from 2020 to 2023. - Include only the five countries selected.
-

Step 3: Research Questions and Visualizations

Objectives: - Answer core research questions through carefully designed plots and summaries.

We can use a **pair chart** to show the difference between vaccinated and fully-vaccinated and see how that plays a role in the death rate

Research Questions & Associated Plots:

3.1. How did daily COVID-19 case counts evolve across countries over time? - Line plots of new daily cases by country from 2020 to 2023. - Facet wrap by country to show comparative timelines. - Highlight key surge periods (e.g., Delta and Omicron waves).

3.2. What is the relationship between vaccination rates and death rates in the US? - Scatter plot: - X-axis: % fully vaccinated population. - Y-axis: death rate per 100,000 people. - Color points by country or year. - Add trendline (linear regression or LOESS) to detect correlation.

3.3. How did ICU occupancy change during pandemic surges? - Stacked area chart or grouped bar chart: - ICU patients per country over months. - Compare against surges in cases.

3.4. (Optional) Additional plot: - Map visualization (small world map) showing vaccination rates by country.

Step 4: Final Analysis and Reporting

Objectives: - Summarize results, draw conclusions, and prepare final deliverables.

Tasks:

4.1. Write an introduction summarizing background context and goals.

4.2. For each research question: - Present associated plots. - Provide short written explanations interpreting the observed trends.

4.3. Discussion Section: - Compare findings across countries. - Highlight surprises, expected patterns, and possible causes (e.g., vaccine hesitancy, variant emergence, healthcare infrastructure).

4.4. Conclusion Section: - Summarize the main lessons learned. - Suggest future research directions.

4.5. Cite all datasets properly using `.cs1` files.

4.6. Compile and render final report in `.pdf` format using Quarto.

Notes

- Large datasets (e.g., OWID full data) will not be pushed to GitHub; a local download script will be provided.
- Ensure consistent citation style using APA format (.csl file applied in Quarto).