COVER PAGE

1. ​​Table of Contents

​NOTE: MAX 2500 words

1. Introduction

The World Health Organization's ambitious goal of curbing COVID-19 transmission through widespread vaccination has encountered significant hurdles along the way. From the outset, the effectiveness and duration of available vaccines have been limited, impacting the ability to achieve sustained immunity. Compounding this challenge is the emergence of highly transmissible variants, evading immune responses and leading to breakthrough infections among the vaccinated. Furthermore, incomplete vaccination rates and unequal distribution have left populations vulnerable, impeding efforts to reach immunity. Amid these complexities, the imperative for enhanced vaccines becomes evident—a critical need underscored by the ongoing struggle to contain transmission. Equally important is the continued adherence to public health measures, such as mask-wearing and physical distancing, which remain essential safeguards in the face of evolving virus strains.

To navigate this multifaceted landscape, a comprehensive and holistic approach is indispensable. This approach demands the integration of robust vaccination programs, bolstered by equitable distribution strategies to ensure global coverage. Concurrently, investment in vaccine innovation is pivotal to enhance efficacy and durability against new variants. Moreover, fostering international collaboration is imperative, transcending borders to address the pandemic as a collective global challenge.

Against this backdrop, our project endeavours to explore the intricate interplay between vaccination rates, mortality outcomes, and socio-economic factors during this unprecedented crisis. Through data visualisation, we seek to identify patterns and insights that can inform targeted interventions and policy responses, contributing to the broader mission of combatting COVID-19 and advancing global public health resilience.

1. Tools Used

The project team has a diverse range of skills in various data visualisation tools, notably Python and Tableau. We have leveraged these proficiencies to analyse our datasets and extract insights for our report. Our visualisations, crafted using both Python and Tableau, are sophisticated and impactful, addressing our predefined research questions effectively.

The decision to use tableau, a powerful data visualisation tool was based on the following reasons:

1. **Wide range of visualisation options**: users can choose from an array of data visualisation types to communicate their chosen insights.
2. **Data connectivity**: Tableau enables users to integrate and join multiple data sources together into a single workbook. This is especially relevant for this project which utilises three separate but relevant datasets.
3. **Advanced analytics/calculations:** advanced analytics capabilities are built into the software enabling calculations to be performed (i.e., using calculated fields) without having to manipulate the original datasets.

The reason for using Python are the following:

1. **Data Analysis Libraries**: Python language has data analysis libraries such as pandas and numpy that are powerful to manipulate, prepare and analyse datasets, facilitating a smooth workflow from data cleaning to visualisation.
2. **Visualisation Libraries:** Python also has libraries for data visualisation such as Matplotlib, Seaborn and Plotly. These libraries offer a wide range of customizable plotting functions and styles to create diverse and insightful visualisations.
3. **Simplicity of the language**: Python is known for its readability and simplicity, making it accessible to both beginners and experienced programmers. The syntax for generating visualisations is straightforward, allowing analysts to focus more on data exploration and less on coding complexities.
4. Datasets

The project used open-source datasets that were available in Kaggle. These datasets include information pertaining to the case numbers and mortality rates associated with COVID-19 (‘COVID-19 Global Statistics’), the vaccination efforts of different countries (‘COVID-19 World Vaccination Progress’) and statistics regarding the socioeconomic status of different countries that were collated by World Bank prior to the onset of the pandemic (‘Life Expectancy and Socio-Economic’).

COVID-19 Global Statistics

The COVID-19 Global Statistics Dataset includes 231 records and 14 columns. Variables include the total cases, new cases, total deaths, new deaths, total recovered, new recovered, active cases, serious/critical cases, total tests, tests (per million), deaths (per million), total cases (per million), and population figures for each country.

*Summary Statistics*

KEVIN, NAHID

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable Name** | **Type (str, num)** | **Mean** | **SD** | **Min** | **Max** | **Count** | **Number of Missing Values** |
| Country | Str |  |  |  |  | 239 | 0 |
| Total Cases | Str |  |  |  |  | 239 | 0 |
| New Cases | Str |  |  |  |  | 12 | 227 |
| Total Deaths | Str |  |  |  |  | 234 | 5 |
| New Deaths | Float | 15.857143 | 16.647608 | 1.000000 | 37.000000 | 7 | 232 |
| Total Recovered | Str |  |  |  |  | 190 | 49 |
| New Recovered | Str |  |  |  |  | 17 | 222 |
| Active Cases | Str |  |  |  |  | 191 | 48 |
| Serious, Critical | Str |  |  |  |  | 60 | 179 |
| Tot Cases/1M | Str |  |  |  |  | 230 | 9 |
| Deaths/1M pop | Str |  |  |  |  | 225 | 14 |
| Total Tests | Str |  |  |  |  | 213 | 26 |
| Tests/1M pop | Str |  |  |  |  | 213 | 26 |
| Population | Str |  |  |  |  | 229 | 10 |

COVID-19 World Vaccination Progress

Data is collected daily from ‘Our World in Data GitHub’ repository for Covid-19, merged and uploaded. The dataset contains 15 variables: Country, iso code, date, total vaccinations, people vaccinated, people fully vaccinated, daily vaccinations raw, daily vaccinations, total vaccinations per hundred, people vaccinated per hundred, people fully vaccinated per hundred, daily vaccinations per million, vaccines, source name, source website and contains 86,512 records. KEVIN, NAHID

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable Name** | **Type (str, num)** | **Mean** | **SD** | **Min** | **Max** | **Count** | **Number of Missing Values** |
| country | str |  |  |  |  | 86512 | 0 |
| iso\_code | str |  |  |  |  | 86512 | 0 |
| date | str |  |  |  |  | 86512 | 0 |
| total\_vaccinations | Float | 4.592964e+07 | 2.246004e+08 | 0.000000e+00 | 3.263129e+09 | 43607 | 42905 |
| people\_vaccinated | Float | 1.770508e+07 | 7.078731e+07 | 0.000000e+00 | 1.275541e+09 | 41294 | 45218 |
| people\_fully\_vaccinated | Float | 1.413830e+07 | 5.713920e+07 | 1.000000e+00 | 1.240777e+09 | 38802 | 47710 |
| daily\_vaccinations\_raw | Float | 2.705996e+05 | 1.212427e+06 | 0.000000e+00 | 2.474100e+07 | 35362 | 51150 |
| daily\_vaccinations | Float | 1.313055e+05 | 7.682388e+05 | 0.000000e+00 | 2.242429e+07 | 86213 | 299 |
| total\_vaccinations\_per\_hundred | Float | 80.188543 | 67.913577 | 0.000000 | 345.370000 | 43607 | 42905 |
| people\_vaccinated\_per\_hundred | Float | 40.927317 | 29.290759 | 0.000000 | 124.760000 | 41294 | 45218 |
| people\_fully\_vaccinated\_per\_hundred | Float | 35.523243 | 28.376252 | 0.000000 | 122.37000 | 38802 | 47710 |
| daily\_vaccinations\_per\_million | Float | 3257.049157 | 3934.312440 | 0.000000 | 117497.000000 | 86213 | 299 |
| vaccines | str |  |  |  |  | 86512 | 0 |
| source\_name | str |  |  |  |  | 86512 | 0 |
| source\_website | str |  |  |  |  | 86512 | 0 |

Life Expectancy and Socio-Economic

This dataset contains 16 variables: Country name, country code, region, income group, year, life expectancy world bank, prevalence of undernourishment, CO2, health expenditure %, education expenditure %, unemployment, corruption, sanitation, injuries, communicable, noncommunicable and contains 3,306 records.

*Summary Statistics*

KEVIN, NAHID

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable Name** | **Type (str, num)** | **Mean** | **SD** | **Min** | **Max** | **Count** | **Number of Missing Values** |
| Country Name | str |  |  |  |  |  | 0 |
| Country Code | str |  |  |  |  |  | 0 |
| Region | str |  |  |  |  |  | 0 |
| IncomeGroup | str |  |  |  |  |  | 0 |
| Year | integer | 2010.00000 | 5.478054 | 2001.000000 | 2019.000000 | 3306.000000 | 0 |
| Life Expectancy World Bank | float | 69.748362 | 9.408154 | 40.369000 | 84.356341 | 3118.000000 | 188 |
| Prevalence of Undernourishment | float | 10.663654 | 11.285897 | 2.500000 | 70.900000 | 2622.000000 | 684 |
| CO2 | float | 1.574924e+05 | 772641.5 | 1.000000e+01 | 1.070722e+07 | 3.154000e+03 | 152 |
| Health Expenditure % | float | 6.364059 | 2.842844 | 1.263576 | 24.230680 | 3126.000000 | 180 |
| Education Expenditure % | float | 4.589014 | 2.119165 | 0.850320 | 23.270000 | 2216.000000 | 1090 |
| Unemployment | float | 7.890760 | 6.270832 | 0.100000 | 37.250000 | 3002.000000 | 304 |
| Corruption | float | 2.860513 | 0.621343 | 1.000000 | 4.500000 | 975.000000 | 2331 |
| Sanitation | float | 52.738785 | 30.126762 | 2.377647 | 100.000004 | 2059.000000 | 1247 |
| Injuries | float | 1.318219e+06 | 5214068 | 4.304900e+02 | 5.563676e+07 | 3.306000e+03 | 0 |
| Communicable | float | 4.686289e+06 | 18437270 | 3.301600e+02 | 2.685646e+08 | 3.306000e+03 | 0 |
| NonCommunicable | float | 7.392488e+06 | 29326880 | 2.481820e+03 | 3.246378e+08 | 3.306000e+03 | 0 |

1. Data Integration

Describe the process of addressing the research question from the three selected datasets. Ex: Data Integration, Joining, etc.

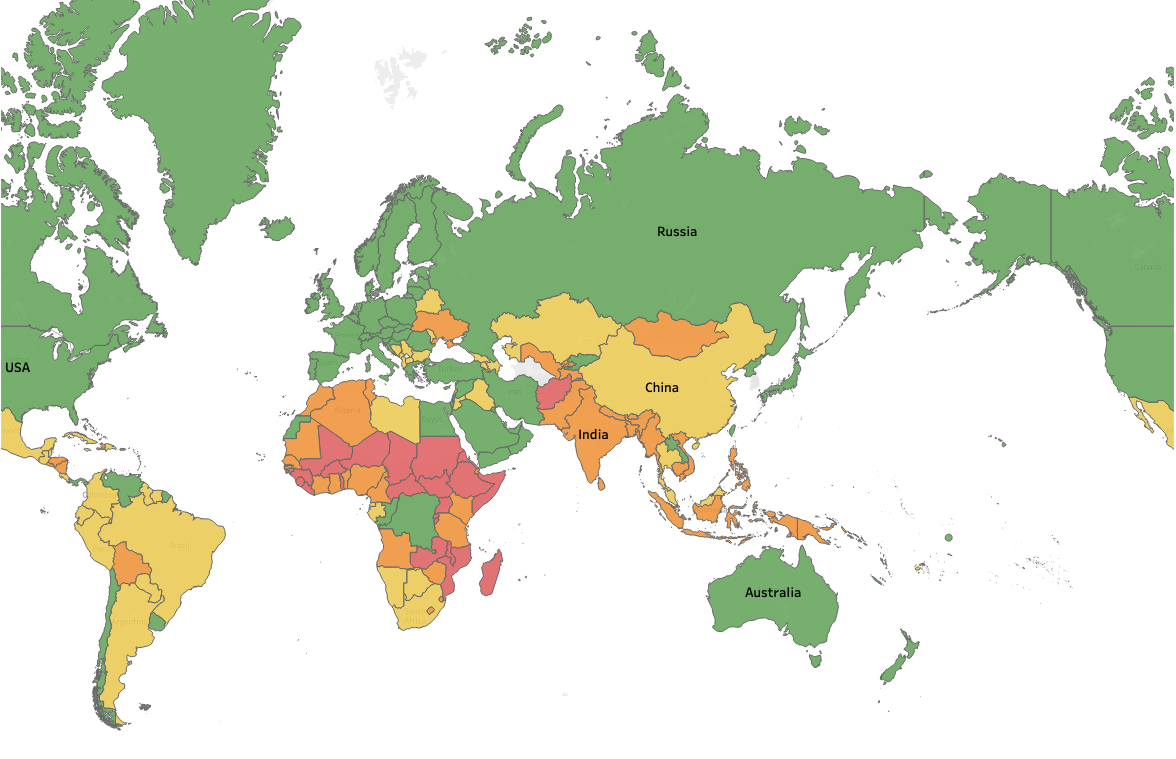
FAITH, NAHID, TYSON - Discuss how data was integrated/joined for your vizualisations to answer your question assigned.

Nahid:

Faith: Data Integration: The COVID-19 data and country vaccination data was merged using the pd.merge() function from the pandas library IN PYTHON. The datasets were joined on the 'Country' and 'Iso Code' columns from the COVID-19 data and the 'country' and 'iso\_code' columns from the country vaccination data, respectively.

Tyson: Datasets were joined in Tableau by the linking field ‘Iso Code’ which was a common variable between the ‘COVID-19 World Vaccination Progress’ and ‘Life Expectancy & Socio-Economic (world bank)’ datasets. This variable was manually created in the ‘COVID-19 Global Statistics Dataset’ by assigning each ‘Country’ with their respective ISO code. This enabled…

1. Tableau Plots

Complexity, number of plots (24-30 plots) from different datasets, Quality of Plots, Description of Plots.

A screenshot of a phone

Description automatically generated

**Figure 1:** Countries colour-coded according to income bracket determined in the World Bank ‘Life Expectancy and Socio-Economic’ dataset.

1. Python Plots

Complexity, number of plots (24-30 plots) from different datasets, Quality of Plots, Description of Plots.

ADD FROM FOLDER

1. Discussion

Discussion of Findings (The meaning of the plots). How the plots can help to analyse and answer the described problem (research Questions)

Among the findings how to utilize different plots to narrate a single story.

FAITH, NAHID - create initial discussion points based on visualisations

KEVIN and JO to synthesise and create the discussion flow.

*Q1. Describe the* ***interaction*** *between COVID-19 testing, vaccination rates and mortality rates over time. FAITH*

*Q2. What* ***countries*** *have experienced the greatest* ***disparities*** *of COVID-19? NAHID*

Q3. What are the associations between COVID-19 vaccination rates and mortality cases considering socio-economic factors? TYSON (sorry I went a bit overboard!)

Socioeconomic status (SES) considers the overall economic and social position of the individuals within a nation. Key indicators of a country's SES include: gross domestic product (GDP), accessibility of education and the quality of a nation's healthcare system among others (Fotso & Kuate-Defo., 2005). According to the World Health Organisation, countries from a lower socioeconomic position will suffer from lower levels of health (WHO, 2022). Therefore, it can be inferred that the adverse effects of COVID-19 will have a greater impact on those countries with relatively lower SES. In order to determine if this relationship existed, different components of SES were examined in relation to the case numbers, mortality and vaccination rates of different countries over the period spanning from from Q1 2020 to Q4 2022.

As income is proven as the greatest determinant of SES (Darin-Mattsson et al., 2017), the income bracket associated with each country is visualised as a reference in Figure 1. It was found that countries considered ‘high income’ (as determined by the ‘Life Expectancy & Socio-Economic’ dataset) observed a greater number of cases and deaths (per one-million people) compared to countries from lower income groups (Figure 2). However, this may be attributed to a large number of unreported cases and lack of testing in lower-income countries (Oxfam International., 2022). This relationship was examined and found to be true in Figure 3, highlighting that the total number of cases is directly correlated to the number of COVID-19 tests performed across countries from different income brackets.

Figure 4 demonstrates a clear positive correlation between COVID-19-related deaths (per 100 people) and each country’s health expenditure as a proportion of Gross Domestic Product (GDP). It is unexpected that the trendline is taking this trajectory as conventional wisdom would suggest that countries investing a greater proportion of their GDP in public health would have a greater capacity to reduce the number of deaths related to COVID-19. Although, the effectiveness of public health measures can be influenced by factors beyond financial investment, including the suitability of healthcare infrastructure to handle a viral pandemic among others. The relationship between COVID-19-related deaths (per 100 people) and a country’s education expenditure as a proportion of GDP is not so pronounced (Figure 5). Despite the expectation that countries with a higher proportion of GDP expenditure on education would experience fewer COVID-19 related deaths due to increased awareness of hygiene practices and adherence to safety measures (i.e, increased vaccination uptake), this correlation is not observed.

The vaccination uptake response for different components of SES was also observed. By grouping countries according to income, it is observed that countries with higher-income status demonstrated a faster rate of vaccination uptake relative to lower-income countries (Figure 4). As studies have proven that lower SES is associated with more access barriers and subsequently worse health outcomes (McMaughan et al., 2020), it can be assumed that a lack of access (i.e., physical and financial) may have been attributed to lower vaccination uptake among relatively low SES groups. Interestingly, when evaluating the relationship between people fully vaccinated and rates of mortality (Figure 7), higher-income countries with a greater proportion of people fully vaccinated also observed the greatest number of COVID-19-related deaths (per 100 people). However, this may be attributed to a greater proportion of highly vulnerable older-aged individuals being concentrated among high-income countries, leading to more COVID-19-related deaths (Bayati., 2021). Additionally, lower-income countries reported a lower-number of cases (Figure 3) and may have attributed COVID-19-related deaths to alternative causes. In order to explore alternative causes for a higher rate of vaccination uptake among high-SES countries, education expenditure (% GDP) was plotted against people fully vaccinated (per 100 people). From the trendline, it is obvious that relative expenditure on education increases the proportion of individuals that are fully vaccinated which aligns with intuition.

Tyson Graphs Limitations (KEVIN, JO):

* Real versus absolute expenditure values for the income dataset.
* In Figure 3, case numbers are higher than test numbers (potentially due to errors in data or different testing methods).

*Q4. What implications do these findings have for public health policies and interventions aimed at mitigating the impact of the COVID-19 pandemic? JO- In this question no plots are required, wait for Q1 to Q3 to be completed and write paragraph drawing conclusions.*

1. Conclusion

Summarize the work.

KEVIN and JO

Link to Python code

https://colab.research.google.com/drive/1R4syd36yAaRcAaRxGsSlp7QMibeeYiUh?authuser=1#scrollTo=o\_h3qBP0wzl-

References

Bayati M. (2021). Why Is COVID-19 More Concentrated in Countries with High Economic Status?. *Iranian journal of public health*, *50*(9), 1926–1929. https://doi.org/10.18502/ijph.v50i9.7081

*Covid-19 death toll four times higher in lower-income countries than Rich Ones*. Oxfam International. (2022, March 3). https://www.oxfam.org/en/press-releases/covid-19-death-toll-four-times-higher-lower-income-countries-rich-ones

Darin-Mattsson, A., Fors, S., & Kåreholt, I. (2017). Different indicators of socioeconomic status and their relative importance as determinants of health in old age. *International journal for equity in health*, *16*, 1-11.

Fotso, J. C., & Kuate-Defo, B. (2005). Measuring socioeconomic status in health research in developing countries: should we be focusing on households, communities or both?. *Social Indicators Research*, *72*, 189-237.

McMaughan, D. J., Oloruntoba, O., & Smith, M. L. (2020). Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. *Frontiers in public health*, *8*, 231. https://doi.org/10.3389/fpubh.2020.00231

*Social Determinants of Health* (no date) *World Health Organization*. Available at: https://www.who.int/health-topics/social-determinants-of-health#tab=tab\_1 (Accessed: 29 April 2024).

Global Covid-19 Vaccination Strategy in a Changing World. (2022, July). Global Covid-19 Vaccination Strategy in a Changing World - July 2022 update. Retrieved from [WHO](https://www.who.int/publications/m/item/global-covid-19-vaccination-strategy-in-a-changing-world--july-2022-update)