Exploring Socioeconomic factors and Testing Availability in COVID-19 Infection Rates: A Global Analysis

Introduction:

COVID-19, a highly infectious strain of the coronavirus family, was first reported at the end of December 2019 in Wuhan, China. It was declared a public health emergency in January 2020 by the World Health Organisation (WHO) before it was escalated to a global pandemic just months after in March of 2020 (Coronavirus disease (COVID-19) pandemic, n.d.).The pandemic had incomparable effects to societies worldwide, propelling the need to understand variables and factors influencing global infection rates. In this report, we explore the relationships between the different socioeconomic factors, testing availability and infections rates, to uncover patterns and insights that might help us prepare for similar incidents in the future.

The socioeconomic factors we'll be exploring are extreme poverty rates and the Human Development Index (HDI). While poverty rates are subjective worldwide, the World back has set an international poverty line of $2.15 per day to maintain global consistency for monitoring extreme poverty (Poverty - Our World in Data, n.d.). The human development index (HDI), assesses a countries growth & development. The HDI takes into account 3 major parameters. Health, which is measured by life expectancy at birth. Education, which is measured by the mean years of schooling for adults above 25 years of age and expected years of schooling at birth. Lastly, Standard of living, which is measured by the Gross national income per capita (GNI) (Human Development Index | Human Development Reports, n.d.). Understanding these relationships is essential in developing and fostering public health interventions, policy decisions, travel restrictions and resource allocations for future contingencies.

Aims:

- Identifying any association between HDI and COVID-19 infection rates
- Analysing Covid-19 testing and Infection rates
- Investigating the relationship between extreme poverty and infection rate

Methodology:

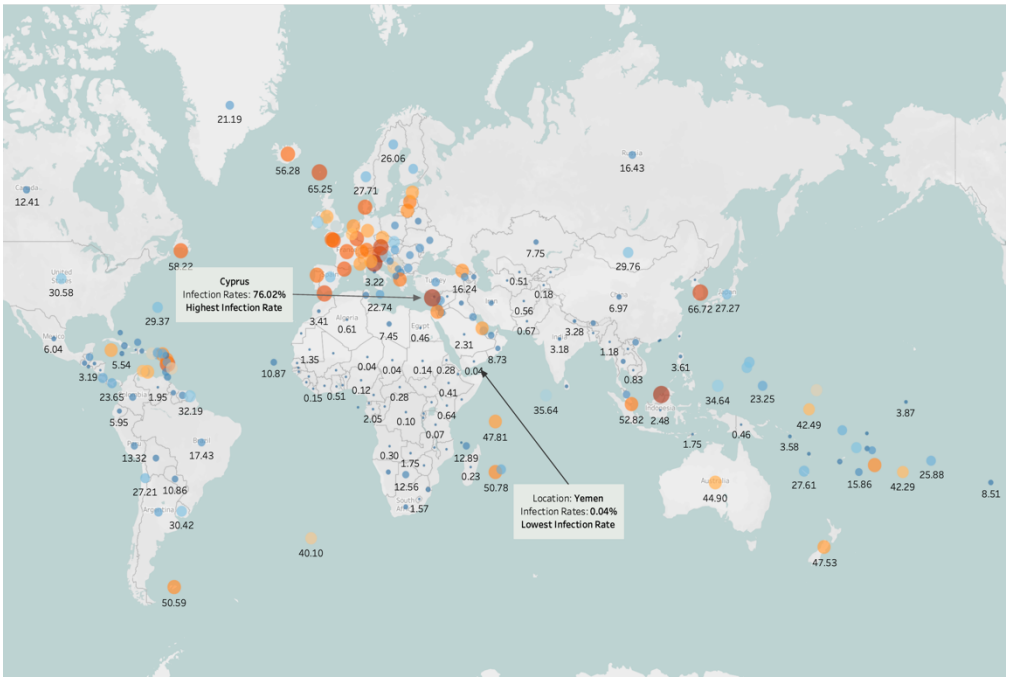
Data was sourced from 'https://ourworldindata.org/covid-deaths', includes data from dates 01-01-2020 to 02-02-2024. During the preparation of the data, I encountered a few discrepancies involving the date format, altering the date format from string to date.
Inconsistencies were found in the TestingPer1000 attribute; hence I used aggregations to calculate this attribute anew with the provided data. It is best to keep in mind that the population value does not increase or decrease during this time period, so some discrepancies will be present
Also, while parsing through the data, I discovered, 243 locations were listed, this included all countries and territories worldwide. For consistent and reliable data, we only required UN (united nations) recognised countries, hence I joined to another table that contained internationally recognised countries.

Excel and SQL queries such as aggregation, CTE windows, joins were used for analysis.
Tableau visualisation for easy accessibly, comprehension and identifying patterns and trends.

Analysis results:

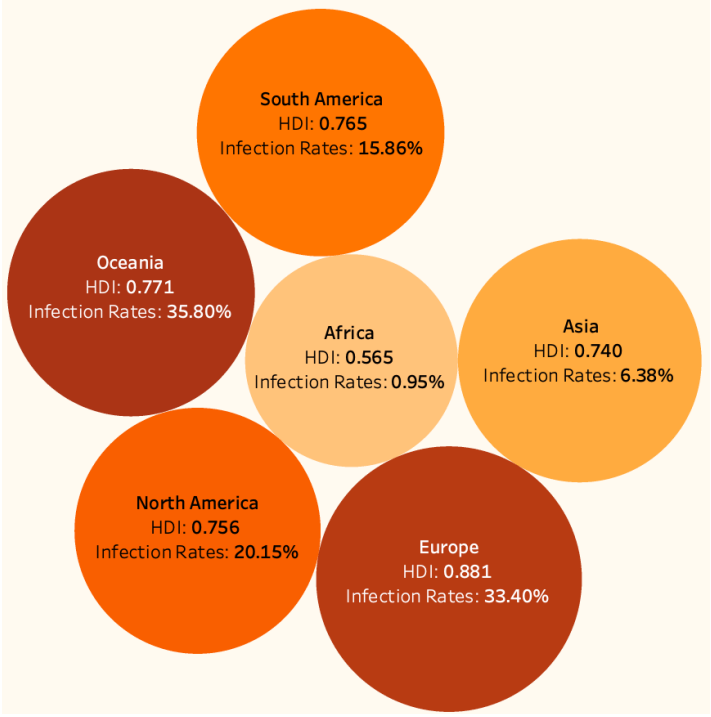
The primary aim of this analysis was to investigate the relationship between socioeconomic factors and infection rates globally.

Graph 1: Map of global infection rates



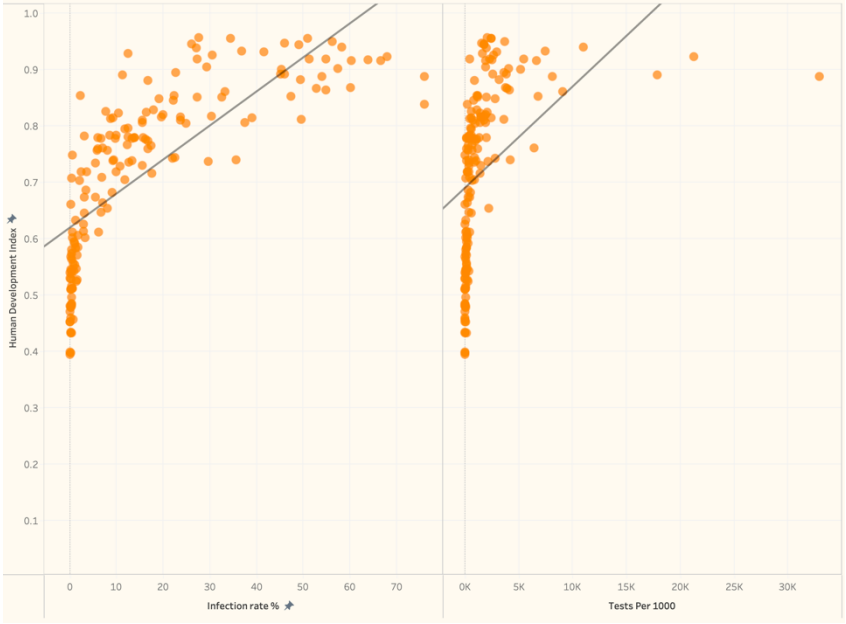
Global infection rates were calculated relative to the population of each country. On the map, it is clear that we have certain hotspots in Europe, with Cyprus having the highest infection rates worldwide (76.02%). Yemen, Chad and Niger all equally had the lowest infection rates globally (0.04%).

Graph 2: Average Continental HDI and Infection Rates



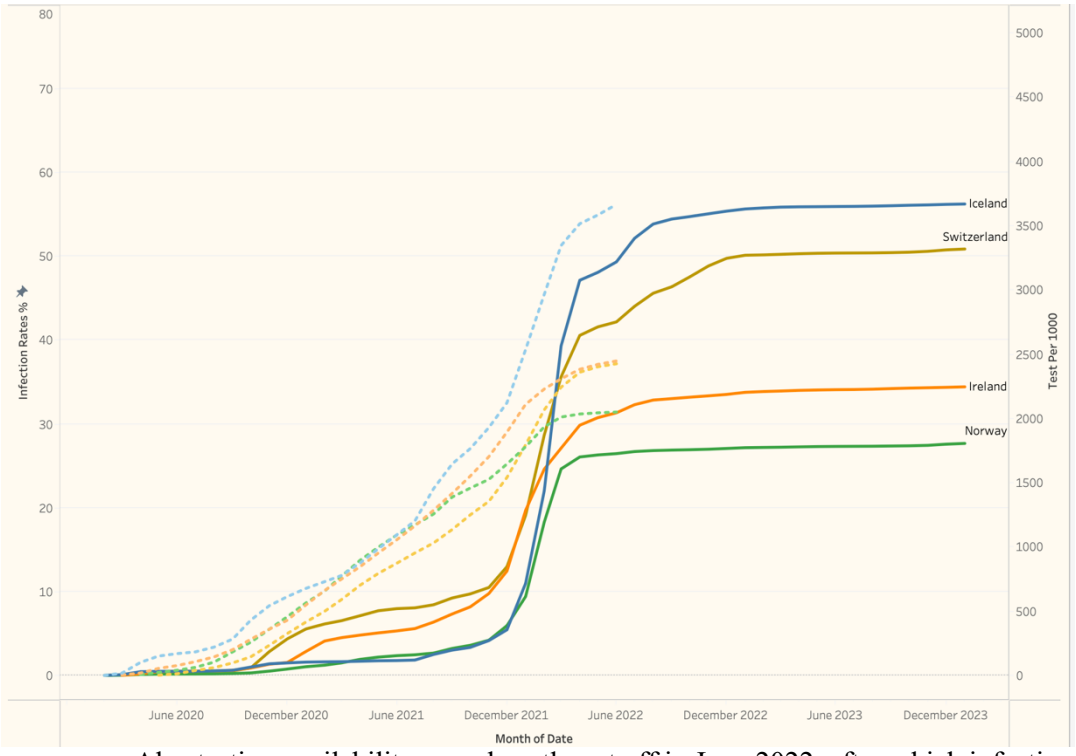
Scaling back and observing the infection rates from a continental perspective. In contrast with the above observation, we discover that Oceania had the highest infection rates relative to its population (35.08%), followed by Europe (33.08%). Africa had the lowest average human development index (HDI) as well had the lowest infection rates.

Graph 3: Relationship Between HDI, Testing Availability and Infection rates.



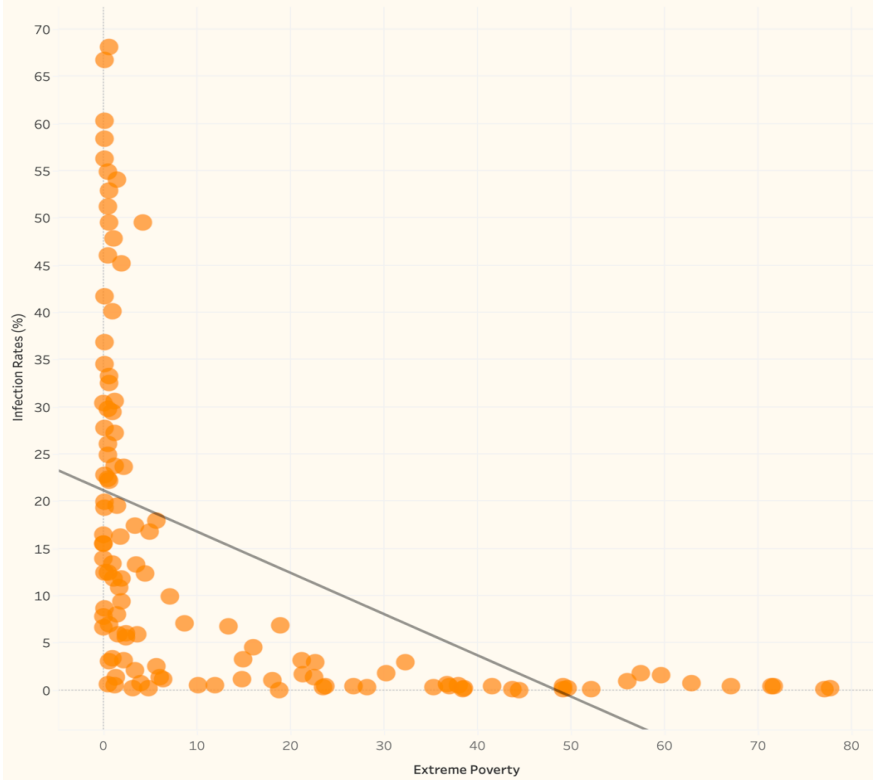
This graph demonstrates a positive correlation between HDI and Infection rates with a correlation coefficient of 0.754 and an R-squared values of 0.570. This means approx. 57% of the variability seen in infection rates can be attributed to fluctuations in HDI.
It is also important to point out that the correlation between HDI and test availability is positive although it has a lower coefficient of 0.434. However, with a lower R-squared value of 0.182, it suggests that HDI and test availability alone do not influence infection rates. Other factors such as healthcare access, infection control policies and the documentation process might have impacted it, nevertheless further analysis is required to explore these factors and its impact on infection rates.

Graph 4: Trend analysis: Infection Rates and Testing Availability in high HDI countries



This graph illustrates the top 4 countries with the highest HDI. It shows that although there's a steady increase in testing availability, it isn't until the December of 2021, that there is a surge in infection rates for all 4 countries. This could be attributed to the winter season, due to many spending more time indoors, resulting in less ventilation and can easily enable virus transmission.
Also testing availability was abruptly cut off in June 2022, after which infection rates in all but Switzerland reaches a plateau. It isn't until January of 2023, that Switzerland reaches a plateau. This might be due to lack of documentation or the cessation of testing altogether. The plateau might also be attributed to the halt of reporting confirmed cases. To examine this more closely, we'll need to look in to the reporting practices and public health policy changes during this time period.

Graph 5: Relationship between Infection Rate and Extreme Poverty



This graph illustrates a negative correlation with a correlation coefficient of -0.502. When extreme poverty increases, infection rates tend to decrease. However, with an r squared value of 0.252. It is hard to say if extreme poverty directly affects infection rates as only 25.2% of the variability seen in infection rates can be attributed to extreme poverty and even then, other factors such as (testing practices, health access, correct documentation) can potentially cause this effect, as stated above.

Conclusion:

In conclusion, this report aimed to examine the relationships between socioeconomic factors (such as HDI and extreme poverty), Testing Availability and Infection rates. We found that HDI and Infection rates had a positive correlation, indicating that high levels of HDI is associated with higher infection rates. Additionally, we found that HDI and Testing availability also had a positive correlation. It is important to note that although this association is present, it is incorrect to assume high infection rates, is strictly owed to high HDI levels and testing availability. There are many other facets that potentially play significant roles in infection rate reports such as public health policies and emergencies, housing policies, healthcare access and reporting practices.

Further analysis is required to between these variables and how they link into HDI and viral transmission. Providing insight into policies that require to be revised, in order to be better equipped for potential future health emergencies.

References:

1) *Coronavirus disease (COVID-19) pandemic* (n.d.). Available at: https://www.who.int/europe/emergencies/situations/covid-19 (Accessed: 18 March 2024).

2) *Human Development Index | Human Development Reports* (n.d.). Available at: https://hdr.undp.org/data-center/human-development-index#/indicies/HDI (Accessed: 18 March 2024).
3) *Poverty - Our World in Data* (n.d.). Available at: https://ourworldindata.org/poverty (Accessed: 18 March 2024).