Contributions

Name	***Report Contribution***	***Code Contribution***
Zohd Khan	1. Introduction for first approach 2. Data Acquisition Section 3. Data Processing Section 4. General Life Expectancy Analysis 5. Adult Mortality Rates 6. Impact of Communicable Diseases on Life Expectancy: 7. Relating Government Health Expenditure to Life Expectancy: 8. Relating GDP Per Capita Over Time to Life Expectancy: 9. Conclusion for first approach	Section 1 1. WHO Data Fetching 2. Data Preparation 3. Visualization 1: Life Expectancy over Time 4. Visualization 2: Life Expectancy by Gender 5. Visualization 3: Government Health Expenditure per Capita 6. Visualization 4: Mortality by Communicable Diseases Heatmap Section 2 1. World Bank Data Fetch & Preparation 2. Visual 1: GDP Per Capita Visualization ***3.*** Visual 2: Adult Mortality Rate Visualization
Asuka Iwamoto	1. Introduction for second approach 2. Comparing Life Expectancy in Developed vs Developing Countries 3. Connecting Life Expectancy to Infant/Maternal Mortality 4. Exploring Differences in Child Vulnerable Diseases 5. Conclusion for second approach	1. Section 3

Comparison of Life Expectancy in Different Countries

Zohd Khan, Asuka Iwamoto

Dec 15th, 2023

Introduction:

In this report, we incorporate a multifaceted approach to compare life expectancy between countries. We first examine variations in life expectancy across three major countries—America, China, and India—and aim to identify the socioeconomic and health factors that are most closely associated with these differences. We particularly explore the following questions:

- How has Life Expectancy changed over the past few decades between the countries of India, China, and America? How does gender play into this?
- Do mortality rates, which should be negatively associated with life expectancy, corroborate our findings on general life expectancy?
- Which major diseases have contributed to differences in life expectancy, specifically by increasing mortality rates?
- How does health expenditure compare between countries, and does this comparison explain differences in life expectancy?
- How does GDP per Capita compare between countries, and does this comparison also explain differences in life expectancy?

We will then compare variation in life expectancies among developed countries and developing countries, and the socioeconomic and health factors that may play a role in these differences. Some questions we explored were:

- How has life expectancy evolved differently between developed countries and developing countries around the world?
- How does disease prevalence and differences in healthcare availability differ in developing countries vs developed countries? Do they contribute to differences in life expectancy?

To execute both these approaches, we utilize Python to query the **GHO OData API** from the World Health Organization **(WHO),** as well as **World Bank API.** We then produce several visualizations to better understand the data and draw meaningful conclusions. Given the importance of data acquisition, we will first offer background to the data as well detail our data acquisition process before delving into the focus of our analysis.

Data Acquisition:

GHO API -

To query from the GHO OData API, we first had to develop a solid understanding of the query interface as well as the different parameters available. The GHO Portal utilizes an Open Data Protocol (OData) which allows anyone to access its data by forming requests that use the appropriate parameters. The two key parameter types for this specific API are **dimensions** and **indicators.** In this API, a dimension is a category by which the data can be filtered, such as "SpatialDim" for country and "TimeDim" for year. An indicator, in this context, refers to a specific dataset offered within the API, and must be identified with unique codes such as "WHOSIS_00007" to fetch the data. We were able to access information about the different indicators and dimensions available in the data by referring to URLS provided in the API's documentation.

```
indicator_codes = ["WHOSIS_000007", "WHOSIS_000001","ORALHEALTH_UHC_GOVSCHEME","NCD_BMI_PLUS1C"]
#List of country codes
countries = ["IND","USA","CHN"]
def fetch_data_for_country(indicator_codes, country_code):
    base_url = "https://ghoapi.azureedge.net/api/"
    dataframes = {}

    for indicator_code in indicator_codes:
        # Constructing the URL for fetching data for a specific country and indicator
        url = f"{base_url}{indicator_code}?$filter=SpatialDim eq '{country_code}' and TimeDim ge 2000 and TimeDim le 2019"
```

Although the GHO API's documentation claimed to offer users a simple querying experience, we had to put extensive effort into fetching the data, particularly because we initially struggled to find the best way to acquire and store several indicators (datasets) in our code. To acquire several datasets, we created a list object named "indicator_codes" along with another list designating our three countries of interest, and we then created a "for loop" to iterate through the "indicator_codes" list and complete the appropriate steps to acquire the data. Among these steps, we created a formula for the URL we used for querying, which included the base URL and our two lists, as well as the appropriate filtering for our
desired dimensions based on the syntax requested by the API. After completing the API request, we stored the data in a local dictionary, and assigned a different key to the DataFrames (values) associated with each indicator by using python code **key = f"{country_code}_{indicator_code}".** Once we

returned the local dictionary, we created a global dictionary and called our **fetch_data_for_country** function to store all dataframes in the global dictionary to enable us to work outside of the main function.

World Bank API -

Halfway through our report, we realized that the GHO OData API did not offer data of the highest quality and flexibility for certain variables. Therefore, we chose to also acquire data from the World Bank API, which, similar to our initial data source, operates on an open data protocol. However, querying this API in Python was much easier using the **wbgapi module**, which is a module built for interfacing with the World Bank API. Within the wbgapi module, we simply used the function **wb.data.Dataframe()** function, and entered the appropriate indicator code, country list, and year range as arguments to the function. Since we used this function from the module, we performed all our queries from this API in separate code cells. Despite the simplicity of querying the WordBank API, we do not regret also querying the GHO API as it served as helpful practice in extracting data from more difficult datasources.

```
df = wb.data.DataFrame('NY.GDP.PCAP.CD', ['IND', 'CHN', 'USA'], range(2010, 2021))
```

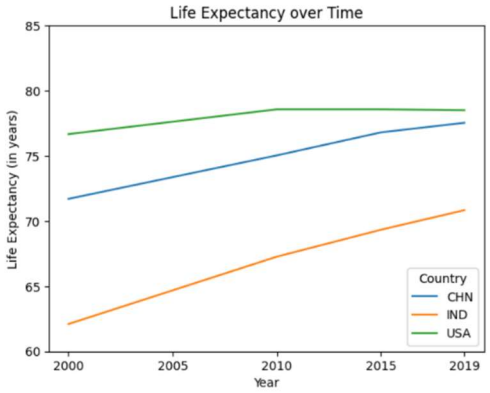
Data Processing:

We performed our data processing in a typical fashion, specifically with tasks such as using the concat function from Pandas to join the different dataframes into one single dataframe. In our respective codes, we performed our own unique steps, with one example being to use the “map” method to transform the data when standardizing gender values. After that, we removed both columns that contained little data or offered little value, as well as all other NA values in the data using the “.dropna” method. A crucial tool we used to accomplish these steps was **Microsoft Excel,** which allowed us to better understand the data due to its visual interface and the many data manipulation features it offers. I (Zohd) personally had a hard time understanding how to clean initially, mainly because I thought I needed to have different cleaning sections for each visualization. However, after spending considerable time looking into the data for my visualizations , I was able to get it all into one code cell for my part as I knew which exact columns needed to be dealt with.

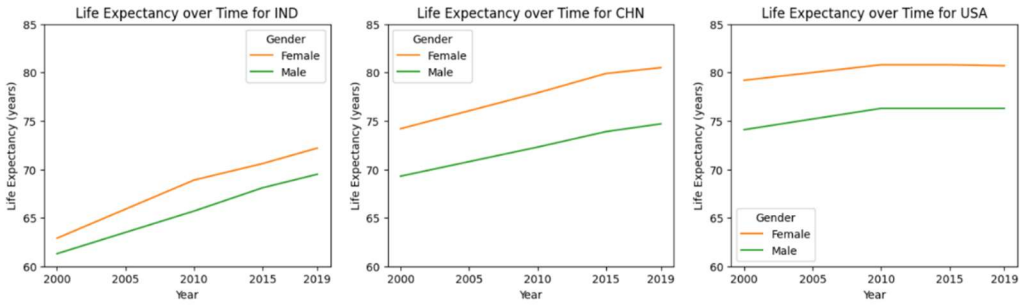
General Life Expectancy Analysis:

To begin the analysis aspect of our report, we use data from the GHO API on Life Expectancy at Birth to examine the general relationship of life expectancy over time between our countries of interest as well as between Males and Females. It is important to note that some of the plots in the report, including these

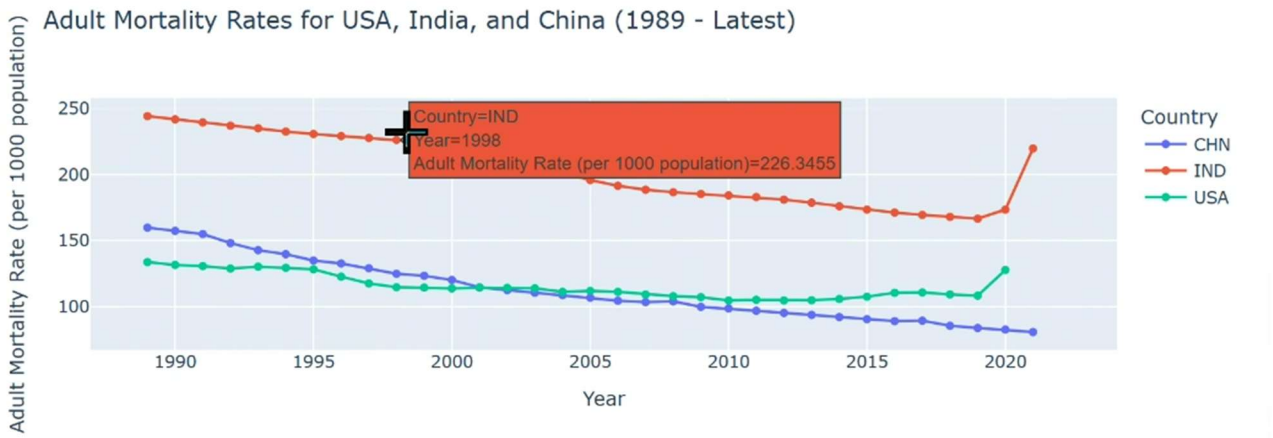
first two ones, data was only available for the 21st century, so we specifically had to query for this time window. However, this time window is still suitable for our research purposes, and hence we were not concerned enough to completely change our data source.



From the first graph (depicted above), it is evident that although India has historically had the lowest life expectancy averages among the three countries, its rate of increase in life expectancy at the beginning of this decade has surpassed the others. Furthermore, China also showed growth, whereas America's life expectancy has largely plateaued over the last ten years. This overall trend by country, as depicted in our initial graph, is corroborated by the second graph. The latter graph (depicted below) reveals a consistent pattern across all three countries: on average, men have a shorter life expectancy than women. While a simple Google search can shed light on the reasons for this gender disparity, our report will not delve into these causes. Instead, we will focus on pinpointing the specific factors that contribute to India's lower overall life expectancy through further visualization, as these may be due to more nuanced reasons.



Adult Morality Rates:



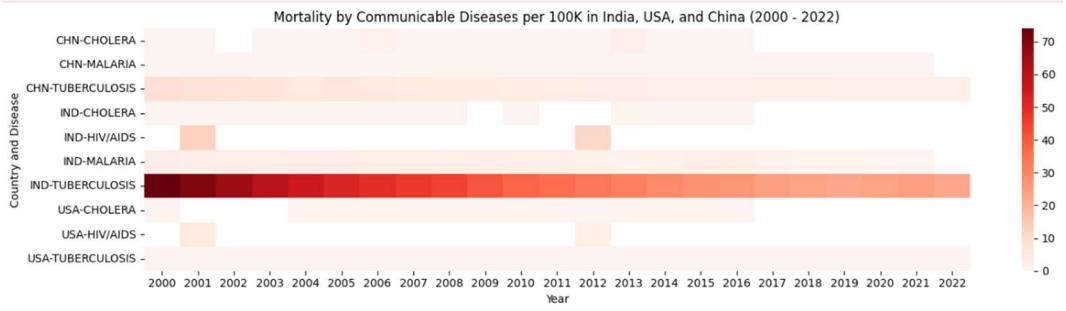
Considering that adult mortality rates—a measure of the risk of dying between the ages of 15 and 60—are crucial for calculating life expectancy, we turned to this statistic to contextualize our earlier findings on life expectancy. Our interactive plot (depicted above) computes the average of male and female data from the World Bank API to illustrate adult mortality rates per 1,000 people over recent years in the countries of interest. The trend from this plot **corroborates** our findings with the life expectancy plot as there is somewhat of an inverse relationship in the lines from the two plots for each respective country. This is expected as when more people die, life expectancy will decrease, and vice versa. The lines in the plot for India and America had a slight but steady **decrease** in mortality rates up until the **COVID-19 Pandemic** hit in 2020. Notably, India, despite having **the highest mortality** rate for the past 35 years, saw its rates decrease most sharply throughout the 21st century up until the COVID-19 pandemic hit in 2020.

In contrast, China did not experience a significant spike in mortality during the pandemic, despite the first reported case originating in Wuhan. This could suggest that data from 2020 onwards might not fully reflect reality, given the politically charged discussions surrounding the pandemic, as well as potential misreporting of true death counts. While such inaccuracies could be part of the data for any disease, they are particularly pertinent to COVID-19 since it is arguably the most impactful pandemic of the century.

Due to the unnatural skew of the data from 2020 onwards, we may put more focus on the data prior to this year, specifically when trying to identify how very specific factors contribute to changes in life expectancy in upcoming sections. To summarize the importance of this section, we have proven the validity of our initial life expectancy plots, partially due to the inverse relationship that is to be expected between life expectancy and mortality rates, and because we have used two different data sources for each plot, which ensures that our results are more credible due to a variety of data sources.

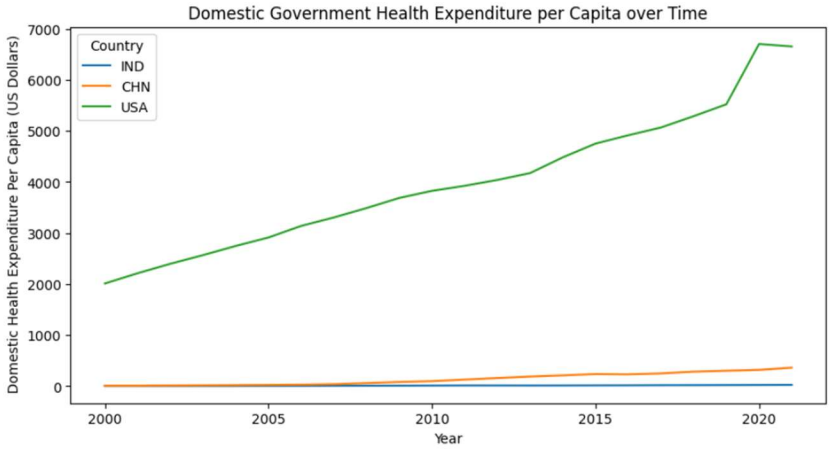
Impact of Communicable Diseases on Life Expectancy:

We now explore which common diseases have contributed to the differences in life expectancy. Unfortunately, the GHO API did not have data available on mortality for what we would expect to be among the deadliest diseases, such as influenza, which resulted in our selection of diseases having relatively low mortality rates. However, the heatmap below still reveals that relative to all other countries, India had a significant struggle combatting mortality from Tuberculosis in the early 2000s. This reveals that one reason India may have lower life expectancy averages compared to the other countries is the poor ability it does in combatting disease, as it did not only have a relatively higher mortality rate due to Tuberculosis, but also due to covid as shown indirectly in the previous chart. We next look at certain factors that may explain why India does such a poor job of disease control and prevention when compared to China and America, which in turns results in the disparities in life expectancy we have observed.



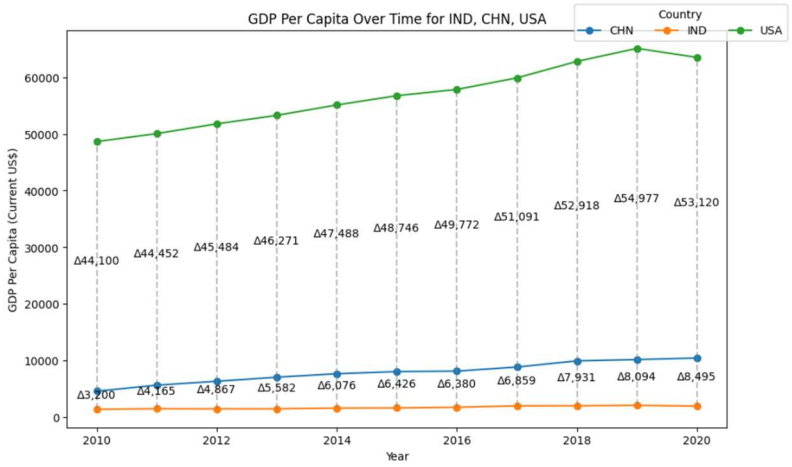
Relating Government Health Expenditure to Life Expectancy:

Another potential factor to affect the life expectancy of these different countries is Healthcare expenditure per Capita (per person) in U.S dollars, which we illustrate in the plot below. This plot clearly reveals that India and China spend significantly less compared to the United States as far as health infrastructure is concerned. The implication of this is self-explanatory, as we can deduce that India low government health expenditure may be a primary reason that it has trouble combatting diseases, and thereby has a lower life expectancy on average. For this broader topic question, we will lastly explore a standard economic measure between countries to see if perhaps one reason India struggles to deal with its health expenditure is its economic capabilities.



Relating GDP Per Capita Over Time to Life Expectancy:

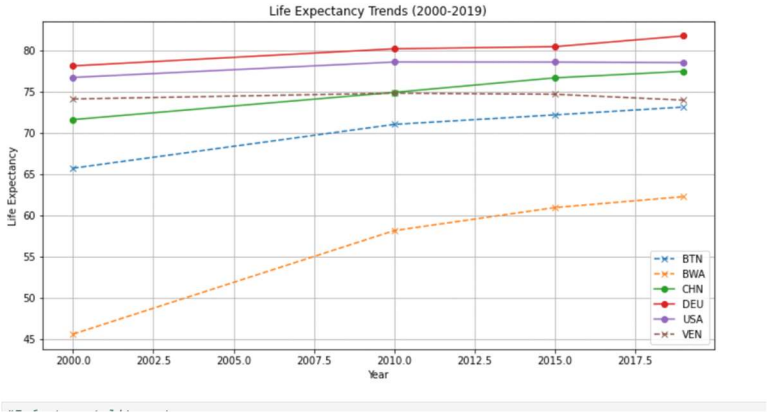
Below we plot the GDP Per Capita over the past 10 years for our 3 countries of interest, and we see countries in the same order, with India having the lowest GDP per capita compared to other countries. One potential may be that India's line is nearly constant over time, which is extremely unlikely and may indicate another issue with the data provided by the GHO API. However, the general idea that India struggles in every facet possible to the extent that it has a lower life expectancy than the other countries is supported in this plot as well. We can say this because India would have a very long way to go before reaching the United States, even if its GDP per Capita is rising. Therefore, we can also attribute India's lower life expectancy averages it performing worse in economic metrics relative to America and China.



Comparing Life Expectancy in Developed vs Developing Countries:

We now begin the next approach in our report. Looking at the differences between the life expectancies as well as disease prevalence in the 3 major countries in the world of USA, China, and India caused me to question the variation between countries with more distinct differences, specifically in development. I chose the countries of USA, China, and Germany, to represent developed countries, and chose the 3 countries of Bhutan, Venezuela, and Botswana to represent developing countries.

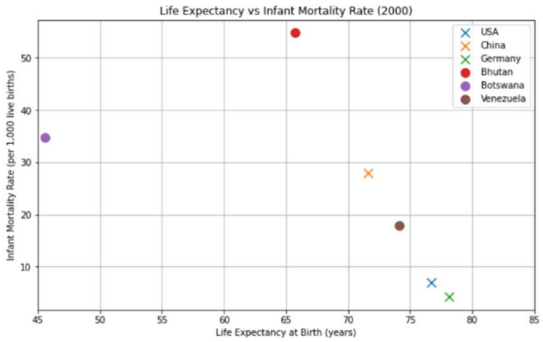
First, we compare the difference in life expectancy for these countries. The developing countries are labeled with a dotted line, while the developed countries are labeled with a straight line.

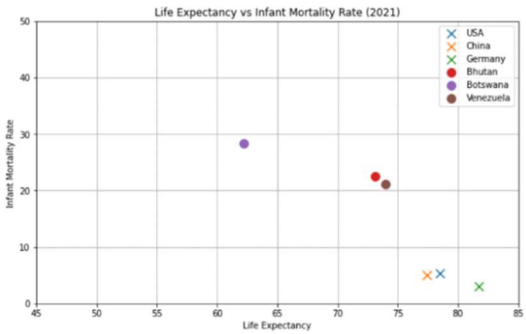


As one can see, Botswana has a distinctly low life expectancy compared to the other 5 countries. We can also observe that all 3 of the developing countries have a very stable life expectancy trend, as their life expectancy has not varied relatively for the 20 years graphed. Botswana and Bhutan on the other hand show a clear positive trend in their life expectancy throughout the 20 years, conveying a noticeable increase in their life expectancy. We will be exploring what changes in health factors occurred or did not occur in these countries to possibly connect to this life expectancy trend.

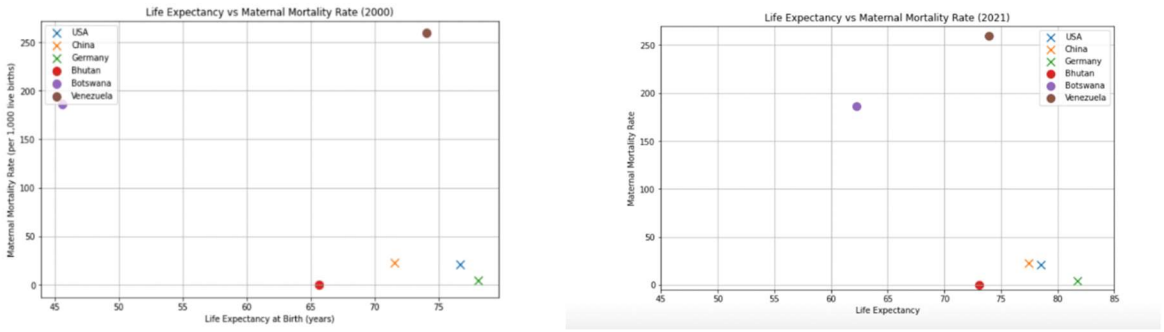
Connecting Life Expectancy to Infant/Maternal Mortality:

Inspired by the variation in trends observed in Neonatal Mortality in our analysis of USA, China, and India, we decide to look at the variation in infant mortality in these countries.





Just like our life expectancy data, there is no drastic change in the developed countries of USA and Germany, marked with an x (developing countries are plotted with a o). China experienced a drop in infant mortality rate of over 20 percent, but did not drastically shift their life expectancy, as the increase in life expectancy was less than 5 years. In contrary, Botswana's life expectancy experienced a large increase of over 15 years in the past 2 decades, but their infant mortality rate remains high at around 30 deaths. The most noticeable growth in both infant mortality rate and life expectancy is seem in Bhutan, as their infant mortality dropped by nearly 30 percent, as well as increasing their life expectancy by around 8 years. Now we will look at the maternal mortality ratio, or the number of deaths of mothers per 100,000 births.



Once again, there is no drastic change in the maternal mortality ratio for developed countries of USA, China, and Germany, observing a similar trend as the lack of change in life expectancy. Overall, the maternal mortality ratio did not drastically change in variation among the 6 countries in the past 20 years. The developed countries continued to maintain a ratio of close to 0, while Botswana has maintained a distinctly high ratio of around 200 deaths compared to the other 5 countries for both 2000 and 2020.

Exploring Differences in Child Vulnerable Diseases:

We decided to look at the two major diseases that are particularly vulnerable to infants under 1 year old, particularly with the ability to be prevented through vaccines within 1 year of an infant's birth. We seek the association between a country's infant mortality that we observed previously with the disease prevalence of Tuberculosis and Measles.

From this data, we can observe a measles outbreak in Botswana in 2020. From this, we can infer that Botswana does not offer a widespread distribution of the measles vaccine. We can also see a noticeable prevalence of both Tuberculosis and Measles for Bhutan and Venezuela, and only the Measles for Botswana in 2014. The generally stable prevalence of both Measles and Tuberculosis for Bhutan and Venezuela throughout the decade indicates that vaccines for both diseases are likely not distributed widely in both countries. However, the case value of Tuberculosis in Botswana is consistently below 5 through the decade. The vaccination distribution of Tuberculosis in the past 10 years may have been a factor contributing to Botswana's increased trend in life expectancy.

From this plot, we can see that the case report values for Tuberculosis are below 5 for both USA and Germany throughout the decade, allowing us to assume vaccinations for Tuberculosis have been distributed widely for all 10 years in the visualization. We also observe a significant prevalence of Measles in Germany throughout the decade, letting us suggest that Measles is not a required or widely distributed infant vaccine in Germany. This is a surprising statistic counteracting the high life expectancy trend of Germany. However, all of these

trends in prevalence are not comparable to the scale of the large Measles prevalence occurring in China throughout the 10 years, with an especially large breakout of over 50,000 reported cases in 2014.

Conclusion

Following the execution of our first approach, it is evident that life expectancy had a similar trend in India, America, and China, specifically in that it held a slight but steady increase for all 3 countries from the start of the 21st century to the start of the COVID-19 pandemic. Despite following a similar trend of growth, the countries clearly differed in the average life expectancy estimates, with India being significantly lower than the other two countries. Not only did we verify these findings by exploring mortality rates from a separate API, but we also identified several reasons for why India's life expectancy averages were significantly lower, with these reasons being the India's poor public health infrastructure as well as its inability to support its health infrastructure due to relatively worse economic conditions. Although we arrived at these conclusions, it was clear throughout the process of completing this report that the GHO Odata API did not have best data available, which indicates that some of our results could be slightly inaccurate. This was surprising to us as we expected a renowned organization like the WHO to provide complete data that is suitable for a robust analysis, so we did not realize the poor data quality until midway through the report. Despite this limitation, we still worked to acquire, process, and visualize the data given to the best of our ability, and we believe that our first approach did an adequate job of answering our primary questions of interest, with the most glaring issue being out of our control.

Through comparing the life expectancies of developed and developing countries, this allowed us to observe a change in life expectancy seen specifically in developing countries. In general, we can conclude that developing countries experience less change in health factors and mortality rates throughout the years, likely from stable healthcare and socioeconomics. Meanwhile, developing countries saw a noticeable increase in not only life expectancy, but mortality rates for both the mother and child, allowing us to connect the increasing trend in life expectancy to be connected to better services surrounding families and giving birth. Finally, an analysis in disease prevalence of child vulnerable disease allowed us to suggest which diseases were likely not widely vaccinated in both developed and developing countries, at times seeing a connection to the increase in life expectancy with diminished prevalence of certain diseases. However, the disease prevalence in even developed countries like Germany let us know that there are still many ways a country can improve to make their citizens healthier and live longer, happier lives.