Development in the Global Healthcare Industry

Arjun Naik

April 29, 2022

The GitHub repository for this project can be accessed at:
github.com/Aleph-Null-123/Development-in-the-Global-Healthcare-Industry
The code used throughout this project, along with the raw data sources can be found here.

Contents

1 Introduction

1.1 Background

As said in Arjun Naik's "Global Access to Quality Healthcare,"[1] many people in highly developed countries take good healthcare for granted. Unfortunately, access to healthcare is distributed unevenly throughout the world. According to the World Health Organization, over 50 percent of the population does not have access to essential services such as healthcare. However, in a developed country such as the US, over 90 percent of the population has access to quality healthcare, whereas in a country such as India, only 37 percent of the population has access to healthcare.[2]

1.2 Significance

This is significant, because the healthcare industry is one of the most vital structures to the workings of the world. If such a significant portion of the world is unable to access quality healthcare, what would this say about global development in general?

1.3 Question

Access to quality healthcare introduces a new concept, the concept of progress in the healthcare industry, in that improved access to quality healthcare is a subset of progress in the healthcare industry. Since progress in the healthcare industry and global development are both positive changes, it is reasonable to expect that these changes will reflect each other. This paper adds to this idea by investigating the question,

Does global *progress* in the healthcare industry reflect global *development* in general? Does global *development* in general reflect global *progress* in the healthcare industry?

2 General Approach

In order to investigate this question most effectivley, this paper will be taking real world data that represents aspects of the healthcare industry. These data sets will be used to represent the "*progress* in the healthcare industry" that the question asks about.

Next, we will be taking real world data representative of the "global *development* in general" that the questions ask about. This data set will be representative of global development because it will be a "World Development Indicator."[3]

After acquiring the data, we pick a set of countries that we reason to be representative of the whole data set, in terms of overall development.

Using this new set of countries, we compare each of the selected indicator datasets with the chosen indicator for overall development per country. This comparison will give us information about the relationship, or namely, the similarities and differences between progression in the healthcare industry and overall global development.

Next, using our conclusions about the relationships between healthcare progression and global development, we connect to qualitiative information about advancements in order to make an attempt to learn more about this relationship.

[1] Naik, Arjun. *Global Access to Quality Healthcare.* 2022. 28 Apr. 2022. *GitHub*, github.com/Aleph-Null-123/Global-Access-to-Quality-Healthcare

[2] "World Bank and WHO: Half the World Lacks Access to Essential Health Services, 100 Million Still Pushed into Extreme Poverty Because of Health Expenses." World Health Organization, World Health Organization, www.who.int/news/item/13-12-2017-world

[3] Our data set will be taken from World Bank's list of World Development Indicators.

3 Data

3.1 Deciding upon the Datasets to Use

In terms of the data that we must gather, we need to fulfill two categories:

1. Data that represents global development in general.
2. Data that represents progress in the healthcare industry.

The first requirement will be a little easier to satisfy—global development, despite being general, can be measured by a spectrum of things, from economic standing to political situations within a country.

However, progress in the healthcare industry will be a little more difficult to satisfy. Using any single dataset that is within this concept will likely make our data hyperspecific to one subset of progress in the healthcare industry, and may misrepresent the question we are asking which would make our analysis and conclusion biased.

Taking these things into account, the most reasonable approach on collecting the data is the following:

> We take **one single dataset** that is representative of global development. Since it can be measured by a spectrum of things, which can be used independently, this is viable. For progress in the healthcare industry, we take **multiple datasets** in order to represent this. This will allow us to ensure that our data covers a wider range of possibilites, and to prevent us from making our data hyperspecific on one facet of the thing we are trying to represent, progress in the healthcare industry. In addition, we find that many countries have data missing for one phenomenon we pick, so choosing many phenomena would allow us to carry out the analysis even with this situations.

Now that we've made clear the manner in which we plan on choosing datasets, we must now actually choose the datasets. All the datasets used in this project are from World Bank.[4] Full versions of the data in ".CSV" file format can be viewed on the GitHub repository[5] for this project.

For the dataset that will be representative of global development, we will be taking data from the dataset, GDP Per Capita.[6] This is representative of global development from an economic standpoint. In addition, World Bank recongizes this dataset as a "World Development Indicator," which means it indicates global development, which perfectly fits our motives for using this data.

For the datasets that will be representative of overall progress in the healthcare industry, we pick the following datasets:

- Total Population[7]
- Crude Birth Rate for every 1,000 people[8]
- Crude Death Rate for every 1,000 people[9]
- Incidence of Tuberculosis for every 100,000 people[10]
- Life Expectancy at Birth [11]

These datasets are fairly spread out throughout the concept of progression in the healthcare industry, which makes it viable.

[4] data.worldbank.org/indicator
[5] github.com/Aleph-Null-123/Development-in-the-Global-Healthcare-Industry
[6] data.worldbank.org/indicator/NY.GDP.PCAP.CD
[7] data.worldbank.org/indicator/SP.POP.TOTL
[8] data.worldbank.org/indicator/SP.DYN.CBRT.IN
[9] data.worldbank.org/indicator/SP.DYN.CDRT.IN
[10] data.worldbank.org/indicator/SH.TBS.INCD
[11] data.worldbank.org/indicator/SP.DYN.LE00.IN

3.2 Describing the Data

Below is the “head,” or first five rows of our GDP dataset (taken as an example). Currently, it is hard to use for our purposes, so something must be done to the data before it is used in the analysis.

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...	2012	2013	2014	2015	2016	2017	2018	2019	2020	Unnamed: 65
0	Aruba	ABW	GDP per capita (current US$)	NY.GDP.PCAP.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	24712.493263	26441.619936	26893.011506	28396.908423	28452.170615	29350.805019	30253.279358	NaN	NaN	NaN
1	Africa Eastern and Southern	AFE	GDP per capita (current US$)	NY.GDP.PCAP.CD	147.612227	147.014904	156.189192	182.243917	162.347592	180.214908	...	1736.166560	1713.899299	1703.596298	1549.037940	1431.778723	1573.063386	1574.978648	1530.059177	1359.618224	NaN
2	Afghanistan	AFG	GDP per capita (current US$)	NY.GDP.PCAP.CD	59.773234	59.860900	58.458009	78.706429	82.095307	101.108325	...	638.845852	624.315455	614.223342	556.007221	512.012778	516.679862	485.668419	494.179350	516.747871	NaN
3	Africa Western and Central	AFW	GDP per capita (current US$)	NY.GDP.PCAP.CD	107.932233	113.081647	118.831107	123.442888	131.854402	138.526332	...	1965.118485	2157.481149	2212.853135	1894.310195	1673.835527	1613.473553	1704.139603	1777.918672	1710.073363	NaN
4	Angola	AGO	GDP per capita (current US$)	NY.GDP.PCAP.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	5100.097027	5254.881126	5408.411700	4166.979833	3506.073128	4095.810057	3289.643995	2809.626088	1776.166868	NaN

[12]

Every dataset used in this project follows the same format: the columns (after appropriate cleaning) are the years, spanning from 1960 to 2020. The rows (after appropriate cleaning and reformatting) are the countries.

3.3 Cleaning the Data

Now that we’ve acquired that data from the relevant source, we must “clean” the data such that it fits the criteria we will be using in order to make the data most suitable for the purposes of our analysis. The purpose of this section is to explain these criteria and how we will clean the data in order to fit the critera.

3.3.1 Criteria

The main criteria that our data needed to meet was that the “Country Name” was part of a list of countries that:

- are publicly recognized as countries
- would have a sufficient amount of informaion if research on the country was needed for further analyses

Now that our criteria is defined, we clean our data by doing the following things:

1. Remove unnecessary columns that clutter the data, (i.e. [‘Indicator Name’, ‘Indicator Code’, ‘Country Code’])
2. Remove all rows such that the Country Name in that row doesn’t exist in the list of countries that meet our criteria

To see the actual algorithm used, view the GitHub repository for this project.

Now, our data looks like this and is more usable for the analysis:

Country Name	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Aruba	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	24985.013919	24712.493263	26441.619936	26893.011506	28396.908423	28452.170615	29350.805019	30253.279358	NaN	NaN
Afghanistan	59.773234	59.8609	58.458009	78.706429	82.095307	101.108325	137.594298	160.898434	129.108311	129.32976	...	591.190030	638.845852	624.315455	614.223342	556.007221	512.012778	516.679862	485.668419	494.179350	516.747871
Angola	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	4615.468219	5100.097027	5254.881126	5408.411700	4166.979833	3506.073128	4095.810057	3289.643995	2809.626088	1776.166868
Albania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	4437.142612	4247.630047	4413.062005	4578.633208	3952.802538	4124.055390	4531.019374	5287.663694	5395.659532	5246.292306
Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	43334.811413	38686.461264	39540.724814	41303.929371	35770.776704	37475.635059	38964.904478	41791.969837	40897.330873	NaN

[13]

With this one, we only have the columns we need and the countries we want. We perform this cleaning procedure on all the datasets we decided on from section 3.1. The only notable deviation is that we divide all data entries in the population dataset by 1,000,000. This way, population will be measured in the millions, which might make it easier for us to compare data.

Now that we have the data that will be used for the analysis, after the cleaning procedures, we can now proceed to the next steps.

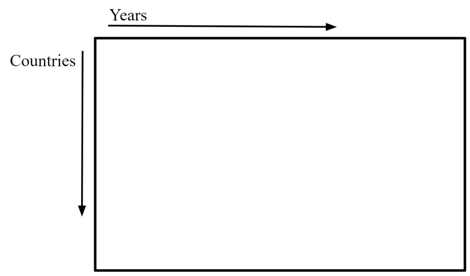
[12] Pandas and JupyterLab

[13] Pandas and JupyterLab

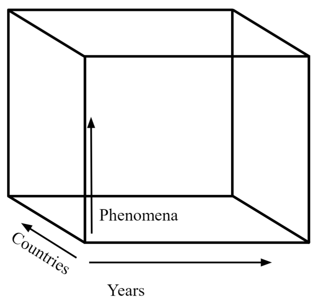
4 The Three-Dimensional DataFrame

4.1 Concept

As detailed in section 3, this paper uses multiple datasets. Every dataframe follows the same format, with the columns being the years and the rows being the countries. These can be visualized like the following:



We have many phenomena that we are trying to model in this paper, and they are all in separate dataframes. However, we can combine them into one single three-dimensional dataframe by "stacking" the dataframes on top of each other. Here, the columns would be the years, the rows would be the countries, and the "layers" would be the measured phenomena. This can be visualized like the following:

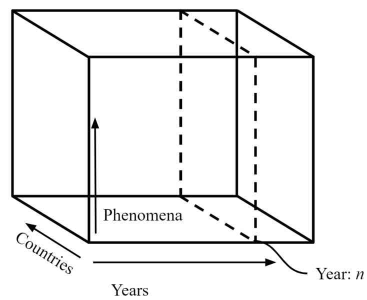


4.2 Usage

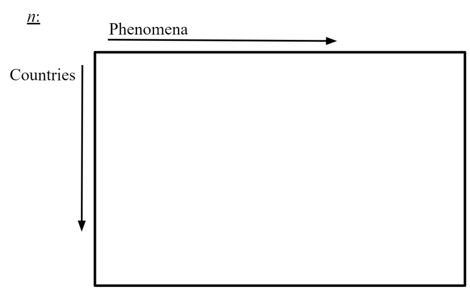
In this paper, the three-dimensional dataframe will be used to create two-dimensional dataframes that involve only two out of the three axes. With the multiple two-dimensional dataframes themselves, we were only able to model countries and years together, for each phenomenon individually. However, the three-dimensional dataframes will allow us to do this with any pair of axes.

4.3 "Slicing"

The way that this will be done involves "slicing" the three-dimensional dataframe, across one of the axes for a given value on the axis. For example, if we needed to model a dataframe with the measured phenomena and countries for a given year, we would slice the three-dimensional dataframe across the years' axis for that given year, like the following:

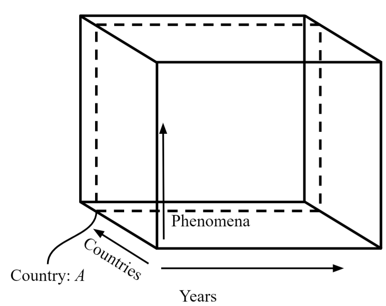


This would produce the following two-dimensional dataframe:



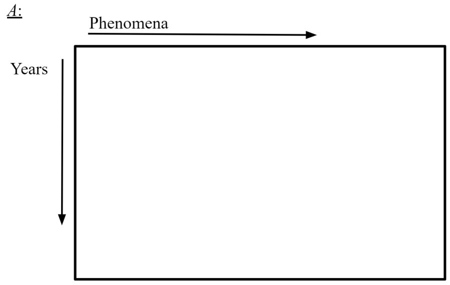
To view the algorithm used to slice the dataframe and produce the 2D dataframe, view the GitHub repository for this project.

Similarly, if we needed to model a dataframe with the measured phenomena and years for a given country, we would slice the three-dimensional dataframe across the countries' axis for that given country, like the following:



This would produce the following two-dimensional dataframe:

A:



5 Set of Countries

5.1 Strategy for Choosing

Part of our general approach in answering the question is picking a set of countries that we reason to be representative of all the data in terms of overall development. What would this look like, though? The approach we take in creating this set of countries is to take countries in the top, middle, and bottom 1 percentile of the dataset that represents overall development.

The dataset that we chose to represent overall development is the GDP per capita dataset, so we will use this dataset to find representative countries using the top, middle, and bottom 1 percentiles. However, our dataset is two-dimensional, so there is more than one column. This would make it impossible to produce a single set of countries with percentiles, since every year, the countries' GDP Per Capita's will change, so each year would have different countries in the top, middle, and bottom 1 percentiles.

5.2 Viable Options

This deterrent means we have to take a different approach. We have two main viable options:

1. Pick a designated year between 1960–2020 from the dataset, which would result in a single number for every country, which would make it possible to calculate the countries in the top, middle, and bottom 1 percentiles.
2. Or, we could produce a number for each country based on the data that is representative of the GDP per capita over all the years (1960–2020) for every country. Since it produces a single number, it would be possible to calculate the countries in the top, middle, and bottom 1 percentiles.

The first option would produce a single number, but might not be representative of the whole data. For example, if we take the year 2020, and calculate the countries in the top, middle and bottom 1 percentiles based on that, this number will be heavily biased towards more recent developments and recent data. Central to this project is the *development* happening in the countries and their *progress* in the healthcare industry. This makes the representation of all years, not only countries, crucial to our analysis.

Thus, we will go with the second option, producing a number for each country that is representative of *all the years* as well.

We will use the **mean** of the GDP per capitas for the years 1960–2020, doing this for every country. For a given country, this number will be representative for that country's GDP per capita and its development throughout the years, since it takes into account every data point.

5.3 Choosing

We now choose the countries based on our strategy. First, for each country we take the mean of the country's GDP per capita. This produces a one-dimensional dataset, which means each country now only has one number value that is representative of all years (the mean).

Next, we calculate the countries in top, middle, and bottom 1 percentiles based on their GDP per capita's.

To see the actual algorithm used to calculate these countries, view the GitHub repository for this project.

We get the following output:

	Top	Middle	Bottom
0	Liechtenstein	Botswana	Burundi
1	Monaco	Turkmenistan	Somalia

[14]

Thus, our chosen countries are the following:

1. **Top 1 Percentile**
   - Liechtenstein
   - Monaco
2. **Middle 1 Percentile**
   - Botswana
   - Turkmenistan
3. **Bottom 1 Percentile**
   - Burundi
   - Somalia

6 Analysis Strategy

Now that we have the set of countries, we proceed to the analysis. To reiterate, the first step of the analysis is to see if there is a *correlation* between the countries' overall development (measured by GDP per capita) and the countries' progression in the healthcare industry.

In this paper, everytime correlation is computed, we will use the Pearson correlation coefficient, which is determined by the following equation:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}}$$

The proceeding sections do this first step on the groups of countries that were chosen in section 4.3.

6.1 Disclaimer

Please note that the images produced in the following sections show only the "head" of the produced dataframe, and the resulting dataframe actually contains all indexes. In addition, although it may appear that there are many missing values, depicted by "NaN", this is likely only because the head shows the data with the earliest years, where data is less likely to be present. The missing data doesn't affect any of the computations performed on the dataset, and the computations will exclude the missing values.

[14] Pandas and JupyterLab

7 Legend

The following sections will primarily use the following acronyms for the indicators, for the sake of simplicity:

1. GDP—GDP per capita
2. POP—total population
3. CBR—crude birth rate
4. CDR—crude death rate
5. TBC—incidence of tuberculosis
6. LFE—life expectancy at birth

8 Top 1 Percentile of Countries

We start with the countries in the top 1 percentile: Liechtenstein and Monaco.

8.1 Liechtenstein

We start with Liechtenstein. We begin by slicing the dataframe along the countries' axis on Liechtenstein, which produces a two-dimensional dataframe that looks like the following:

	GDP	POP	CBR	CDR	TBC	LFE
1960	NaN	0.022461	NaN	NaN	NaN	NaN
1961	NaN	0.022813	NaN	NaN	NaN	NaN
1962	NaN	0.023043	NaN	NaN	NaN	NaN
1963	NaN	0.023165	NaN	NaN	NaN	NaN
1964	NaN	0.023236	NaN	NaN	NaN	NaN

[15]

To reiterate the disclaimer from section 6.2, the visual hint that there is a lot of missing date is likely to the dates previewd in the head of the dataframe, and does not affec our analysis.

Now that we have the slice for Liechtenstein, we can compare the values that indicate progress in the healthcare industry with the GDP per capita, or overall development.[16]

To do this, we comnpute their correlations, using the Pearson correlation coefficient:

Computing the correlation between GDP per capita and overall population gives us that their correlation coefficient is approximately **0.955717**.

Computing the correlation between GDP per capita and crude birth rate gives us that their correlation coefficient is approximately **-0.855147**.

Computing the correlation between GDP per capita and crude death rate gives us that their correlation coefficient is approximately **-0.128496**.

Computing the correlation between GDP per capita and life expectancy gives us that their correlation coefficient is approximately **0.885644**.

[15] Pandas and JupyterLab
[16] We omit TBC due to missing data

8.2 Monaco

We do a similar process for Monaco, by first taking a slice at Monaco, which produces the following two-dimensional dataframe:

	GDP	POP	CBR	CDR	TBC	LFE
1960	NaN	0.022461	NaN	NaN	NaN	NaN
1961	NaN	0.022813	NaN	NaN	NaN	NaN
1962	NaN	0.023043	NaN	NaN	NaN	NaN
1963	NaN	0.023165	NaN	NaN	NaN	NaN
1964	NaN	0.023236	NaN	NaN	NaN	NaN

[17]

Now that we have obtained the slive, we use the same process that we used on Liechtenstein to compare the indicators: using the correlation.[18]

Computing the correlation between GDP per capita and overall population gives us that their correlation coefficient is approximately **0.965727**.

Computing the correlation between GDP per capita and crude birth rate gives us that their correlation coefficient is approximately **-0.975807**.

Computing the correlation between GDP per capita and crude death rate gives us that their correlation coefficient is approximately **-0.976538**.

Computing the correlation between GDP per capita and incidence of tuberculosis gives us that their correlation coefficient is approximately **0.269116**.

9 Middle 1 Percentile of Countries

We proceed with the countries in the middle 1 percentile: Botswana and Turkmenistan.

9.1 Botswana

We start with Botswana. Like with the previous countries, our first step is to slice the three-dimensional dataframe for Botswana, producing the following two-dimensional dataframe:

	GDP	POP	CBR	CDR	TBC	LFE
1960	60.493958	0.502733	47.281	17.687	NaN	49.179
1961	64.176140	0.512688	47.059	17.275	NaN	49.684
1962	68.050349	0.523777	46.824	16.867	NaN	50.171
1963	71.106439	0.535692	46.589	16.464	NaN	50.641
1964	75.955918	0.547870	46.362	16.066	NaN	51.099

[19]

Again, we compute the correlations:

Computing the correlation between GDP per capita and overall population gives us that their correlation coefficient is approximately **0.973384**.

[17] Pandas and JupyterLab
[18] We omit LFE due to missing data
[19] Pandas and JupyterLab

Computing the correlation between GDP per capita and crude birth rate gives us that their correlation coefficient is approximately **-0.939469**.

Computing the correlation between GDP per capita and crude death rate gives us that their correlation coefficient is approximately **-0.614902**.

Computing the correlation between GDP per capita and incidence of tuberculosis gives us that their correlation coefficient is approximately **-0.896620**.

Computing the correlation between GDP per capita and life expectancy gives us that their correlation coefficient is approximately **0.568411**.

9.2 Turkmenistan

We proceed to Turkmenistan. Again, our first step is to slice the three-dimensional dataframe for Turkmenistan, producing the following two-dimensional dataframe:

	GDP	POP	CBR	CDR	TBC	LFE
1960	NaN	1.603254	45.710	15.946	NaN	54.471
1961	NaN	1.658364	45.552	15.578	NaN	54.897
1962	NaN	1.715408	45.121	15.173	NaN	55.326
1963	NaN	1.773854	44.433	14.736	NaN	55.757
1964	NaN	1.833065	43.534	14.279	NaN	56.186

[20]

Again, we compute the correlations:

Computing the correlation between GDP per capita and overall population gives us that their correlation coefficient is approximately **0.875476**.

Computing the correlation between GDP per capita and crude birth rate gives us that their correlation coefficient is approximately **-0.336663**.

Computing the correlation between GDP per capita and crude death rate gives us that their correlation coefficient is approximately **-0.771347**.

Computing the correlation between GDP per capita and incidence of tuberculosis gives us that their correlation coefficient is approximately **-0.890081**.

Computing the correlation between GDP per capita and life expectancy gives us that their correlation coefficient is approximately **0.958766**.

10 Bottom 1 Percentile of Countries

Finally, we move on to the countries in the bottom 1 percentile: Burundi and Somalia.

[20] Pandas and JupyterLab

10.1 Burundi

We start with Burundi. As always, we start by taking a slice at the three-dimensional dataframe for Burundi, producing the following two-dimensional dataframe:

	GDP	POP	CBR	CDR	TBC	LFE
1960	70.051910	2.797925	48.510	23.226	NaN	41.281
1961	71.167188	2.852438	48.446	22.930	NaN	41.592
1962	73.435331	2.907320	48.391	22.629	NaN	41.907
1963	78.514621	2.964416	48.338	22.328	NaN	42.225
1964	86.161550	3.026292	48.283	22.031	NaN	42.540

[21]

Now, we compute the correlations:
Computing the correlation between GDP per capita and overall population gives us that their correlation coefficient is approximately **0.698225**.

Computing the correlation between GDP per capita and crude birth rate gives us that their correlation coefficient is approximately **-0.257508**.

Computing the correlation between GDP per capita and crude death rate gives us that their correlation coefficient is approximately **-0.715310**.

Computing the correlation between GDP per capita and incidence of tuberculosis gives us that their correlation coefficient is approximately **-0.880945**.

Computing the correlation between GDP per capita and life expectancy gives us that their correlation coefficient is approximately **0.748257**.

10.2 Somalia

The final country we do the operations for is Somalia. We start by slicing the dataframe at Somalia, producing the following dataframe:

	GDP	POP	CBR	CDR	TBC	LFE
1960	65.479716	2.755967	47.630	26.748	NaN	36.976
1961	68.106397	2.814125	47.520	26.367	NaN	37.374
1962	70.813049	2.874215	47.424	25.990	NaN	37.773
1963	73.607204	2.936478	47.339	25.616	NaN	38.175
1964	76.480398	3.001160	47.261	25.246	NaN	38.578

[22]

Again, we compute the correlations:

Computing the correlation between GDP per capita and overall population gives us that their correlation coefficient is approximately **0.951175**.

Computing the correlation between GDP per capita and crude birth rate gives us that their correlation coefficient is approximately **-0.826070**.

Computing the correlation between GDP per capita and crude death rate gives us that their correlation coefficient is approximately **-0.928148**.

[21] Pandas and JupyterLab
[22] Pandas and JupyterLab

Computing the correlation between GDP per capita and incidence of tuberculosis gives us that their correlation coefficient is approximately **-0.882634**.

Computing the correlation between GDP per capita and life expectancy gives us that their correlation coefficient is approximately **0.947697**.

11 Analysis of Results

Combining all of the correlations that were computed in the previous sections, we can produce the following dataframe:

	POP	CBR	CDR	TBC	LFE
Liechtenstein	0.955717	-0.855147	-0.128496	NaN	0.885644
Monaco	0.965727	-0.975807	-0.976538	0.269116	NaN
Botswana	0.973384	-0.939469	-0.614902	-0.896620	0.568411
Turkmenistan	0.875476	-0.336663	-0.771347	-0.890081	0.958766
Burundi	0.698225	-0.257508	-0.715310	-0.880945	0.748257
Somalia	0.951175	-0.826070	-0.928148	-0.882634	0.947697

[23]

The columns represent the indicators that were compared with GDP per capita to get the correlation.

Computing the means of the correlations for each given phenomena gives us the following:

POP	0.903284
CBR	-0.698444
CDR	-0.689123
TBC	-0.656233
LFE	0.821755

[24]

We now analyze these correlations for each indicator:

11.1 Population

Here, the average correlation was approximately 0.9, which is a very strong positive correlation. This is expected, since the population of countries would grow over time, just as they develop over time. More insight about how this relates to the healthcare industry can be taken from the crude birth rate There aren't any correlations present in the dataset, before the average was taken, that deviate from this conclusion.

11.2 Crude Birth Rate

Here, the average correlation was approximatley -0.7, which is a negative correlation. This means that as a country's GDP per capita goes up, the crude birth rate goes down. This shows that they reflect each other. This is expected, since as a country becomes more developed, people would be able to get better access to things such as contraceptives, which would prevent the crude birth rate from increasing too much. In this way, low population growth is a sign of development. Turkmenistan and Burundi both seem to differ a little from this strong negative correlation, both displaying weak negative correlations. In "Global Access to Quality Healthcare," Arjun Naik argues that "healthcare is distributed unevenly throughout the globe

[23] Pandas and JupyterLab
[24] Pandas and JupyterLab

because of poverty, lack of education and political instability in developing countries."[25] High crude birth rate is another implication of poor access to healthcare. Thus, it is likely that the causes to unequal access to healthcare that Naik details cause high crude birth rate. Since the CBR deviates from the average for Turkmenistan and Burundi, it is likely that poverty, lack of education and political instability don't affect these countries as much as other developed and developing countries, or that it affects it differently, which would cause this difference.

11.3 Crude Death Rate

Here, the average correlation is approximately -0.7 again, which is a negative correlation. This means that as a country's GDP per capita goes up, the crude death rate goes down, which shows that they reflect each other. This is expected, since fewer people would die on average in a country that is more developed, because the developed country would have better access to healthcare, as detailed in Naik's work. An anomaly to this trend appears in Liechtenstein's correlation, which is a weak negative correlation. In "Global Access to Quality Healtcare," Naik argues that the healthcare disparities in large affect only developing countries.[26] In this study, Liechtenstein was used as a developed country, since it was found in the top 1 percentile of our GDP per capita data, which is what we used to indicate development. Thus, since the country that differs is so greatly developed, it is likely that a change in the crude death rate of that country isn't because of its development, which is why the correlation found — -0.12 — is so weak.

11.4 Incidence of Tuberculosis

Here, the average correlation is approximately -0.7, which, again, is a negative correlation. This means that as a country's GDP per capita goes up, tuberculosis becomes less common. This means that they reflect each other. This is expected, because a developed country is less likely to be affected by disease. This is also shown in Naik's work, where he argues that the unequal access to healthcare implies higher burden of disease in developing countries.[27] Naik's claim also can be used to explain the only anomaly in this set of correlations: that of Monaco. In this study, since Monaco was found in the top 1 percentile of countries in terms of GDP, and thus development, the burden of disease would be much less than that of a developing country. Because of this, the slight development of Monaco would not affect the incidence of tuberculosis as much as it would for a country that is developing more from an undeveloped state, since the latter type of country would have already been affected by burden of disease, as detailed in Naik's work. This explains why the correlation for Monaco doesn't match the average.

11.5 Life Expectancy

Here, the average correlation is approximately 0.8, which is a strong positive correlation. This means as a country's GDP per capita goes up, so does the average life expectancy at birth of that country. This means that they reflect each other. This is expected, since a more developed country would have better medical support, which would lead to higher life expectancy. There are no anomalies to this trend in the set of correlations for this indicator.

12 Conclusions

In section 11, we discussed how every indicator for progress in the healthcare industry reflected overall development and vice versa, using the average correlations of the chosen countries. When there were exceptions to this trend, we were able to clearly explain these excpetions using phenomena present in Naik's "Global Access to Quality Healthcare." Since each indicator for progress in healthcare relfected overall development

[25] Naik, Arjun. *Global Access to Quality Healthcare.*
[26] Naik, Arjun. *Global Access to Quality Healthcare.*
[27] Naik, Arjun. *Global Access to Quality Healthcare.*

in general, we can make the argument that $\boxed{\textbf{yes,}}$ **global progress in the healthcare industry reflects global development in general, and global development in general reflects global progress in the healthcare industry.**

13 Further Inquiry

The investigation that occurred in this paper was largely centered around ideas presented by Arjun Naik, in his work, "Global Access to Quality Healthcare." However, in this paper, we took a statistical approach to answer our question, whereas in his paper, Naik engaged with other primary and secondary sources in order to answer his question of, what causes the unequal access to quality healthcare globally, and what does this imply? He answered this without taking a statistical approach. His main conclusions were that poverty, lack of education, and political stability caused the unequal access, and this implied infant mortality, burden of disease, and stalling of development, primarily in developing countries. A direction of further inquiry is to test the validity of Naik's claim by answering his question by taking a statistical approach. To do this, one would need to collect or gather data about the specifc things that Naik argues cause the unequal access, and the phenomena this inequality causes.