**Child Mortality Analysis**



**Prepared by**

**Brian De Villa, Matthew Jankowski,**

**Katherine Misyutina, Abhi Shah**

**for use in CS 418**

**at the**

**University of Illinois at Chicago**

**Spring 2020**

**Table of Contents**

[**I.** **Introduction** 2](#_Toc39499839)

[**II.** **Data** 3](#_Toc39499840)

[**III.** **Methods and Analysis** 4](#_Toc39499841)

[**IV.** **Case Studies** 8](#_Toc39499842)

[**Nigeria** 8](#_Toc39499843)

[**Haiti & Dominican Republic** 8](#_Toc39499844)

[**India** 8](#_Toc39499845)

[**V.** **Results** 9](#_Toc39499846)

[● How does the child mortality rate relate to the development in countries worldwide? 9](#_Toc39499847)

[● How does child mortality of measles compare to other diseases? 9](#_Toc39499848)

[● What kind of diseases are highly linked to underdeveloped and developing countries compared to developed ones? 9](#_Toc39499849)

[**VI.** **Conclusion** 9](#_Toc39499850)

[**Appendix A: Works Cited** 10](#_Toc39499851)

[**Appendix B: Code** 12](#_Toc39499852)

[Introduction 12](#_Toc39499853)

[Load in Data 13](#_Toc39499854)

[Calculations 13](#_Toc39499855)

[Functions to help with plot 14](#_Toc39499856)

[Create Graphs 15](#_Toc39499857)

[Population 15](#_Toc39499858)

[GDP 15](#_Toc39499859)

[Acute Lower Respiratory 15](#_Toc39499860)

[Injuries 16](#_Toc39499861)

[Measles 16](#_Toc39499862)

[Prematurity 16](#_Toc39499863)

[Sepsis 17](#_Toc39499864)

[GDP Increase from 2000-2015 17](#_Toc39499865)

[Merge Dataset 17](#_Toc39499866)

[GDP vs Disease 19](#_Toc39499867)

[Correlation Matrix 20](#_Toc39499868)

[**Appendix C: References** 21](#_Toc39499869)

# [**Introduction**](#l9kb77ogp9rw)

Even as medicine has evolved throughout the years, we still have many countries that lack the proper health care to assist everyone. More than medicine, people themselves have increased substantially throughout the years. In this study, we try to look at death mortality for each country, specifically in children. Children under the age of five are dying due to diseases and injuries that are either curable or prevented by a better standard of living. In this study, we have chosen 32 countries from each continent to get a good sample size. The diseases and injuries we have chosen are: Acute Lower Respiratory Infection, Measles, Sepsis, Prematurity, and Injuries. These were used with the population and GDP of each of the chosen countries. We wanted to see if we can see a relationship between GDP and the number of diseases for each country. Also, we wanted to find out whether there was a decrease in cases over the years.

There are three questions that we are trying to answer with our data:

1. How does the child mortality rate relate to development in countries worldwide?
2. How does child mortality of measles compare to other diseases?
3. What kind of diseases are highly linked to underdeveloped and developing countries compared to developed ones?

The three hypotheses/predictions that were predicted based on the questions:

1. Countries with lower child mortality rates have a higher Gross Domestic Product.
2. Countries that have had a large increase in Gross Domestic Product over the past 2 decades had a large decrease in child mortality.
3. Poor countries such as African nations, India, and Haiti, have the highest amounts of cases.

In our data, it can be seen that countries that are already developed have the smallest percentage of child mortality. Measles has a lower child mortality rate compared to other conditions. Diseases such as Acute, sepsis, and measles are more prevalent in developing countries compared to developed ones.

# [**Data**](#l9kb77ogp9rw)

The data we used had to be credible and detailed. Sources included:

* [Our World in Data, “GDP per capita”](#y4a88iz19adx)
* [Wikipedia, “List of countries by past and estimated future population”](#xb80j46eu91e)
* [Global Health Observatory data repository](#g20tz82yeixp)

In the uncleaned dataset for GDP per Capita from 1990 to 2017, we can see each country's GDP per capita. We also noticed that for some countries the GDP was not listed or was zero because recording data for those countries didn’t begin until later.

In the uncleaned world population dataset, we can see every country’s population in the thousands from 1985 to 2015.

In the uncleaned diseases dataset for child mortality, we also can see that the child’s mortality rate is from 0-4 years, 0-27 days, and 1-59 months.

We filtered the datasets to the year 2000, 2005, 2010, and 2015 to get a brief overview by having a 5-year interval. We also selected the following countries:

**Asia:**

India

Philippines

Bangladesh

China

Saudi Arabia

**Africa:**

Uganda

Kenya

Ethiopia

Morocco

South Africa

Nigeria

Burundi

**Europe:**

Poland

Russia

Germany

Ukraine

Serbia

Albania

**North America:**

United States

Canada

Mexico

Dominican Republic

Guatemala

Haiti

**South America:**

Ecuador

Colombia

Brazil

Chile

**Oceania:**

Australia

New Zealand

Solomon Islands

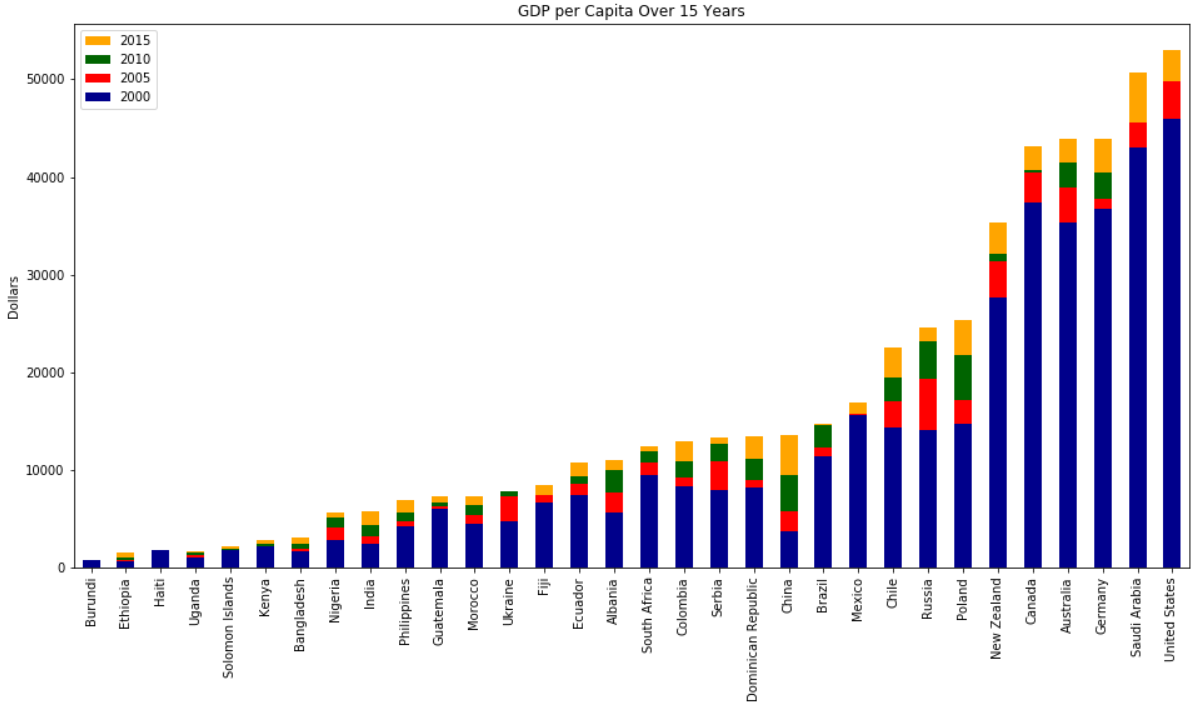
Fiji

We then restricted each disease dataset for child mortality to primarily focus on the age group where the children are less than 5 years old. The diseases we analyzed were Acute Lower Respiratory Infections, prematurity, sepsis, measles, and injuries.

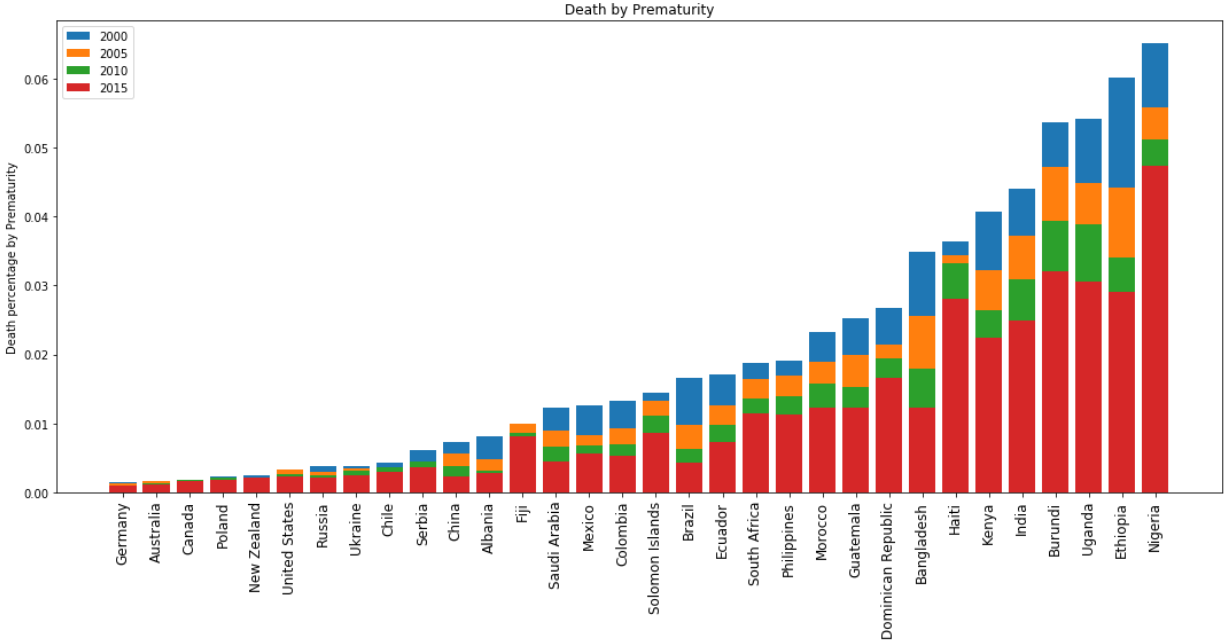
# [**Methods and Analysis**](#l9kb77ogp9rw)

In order to reproduce our data, we first looked at the dataset for the world’s GDP per Capita from the years 1990 to 2017. We then chose sets of countries with diverse GDP from different continents. Next, we decided to observe five diseases or conditions within these countries. We also web-scraped the population for each country from Wikipedia. We then specified the data to have a 5-year interval starting from the year 2000 and going to 2015.

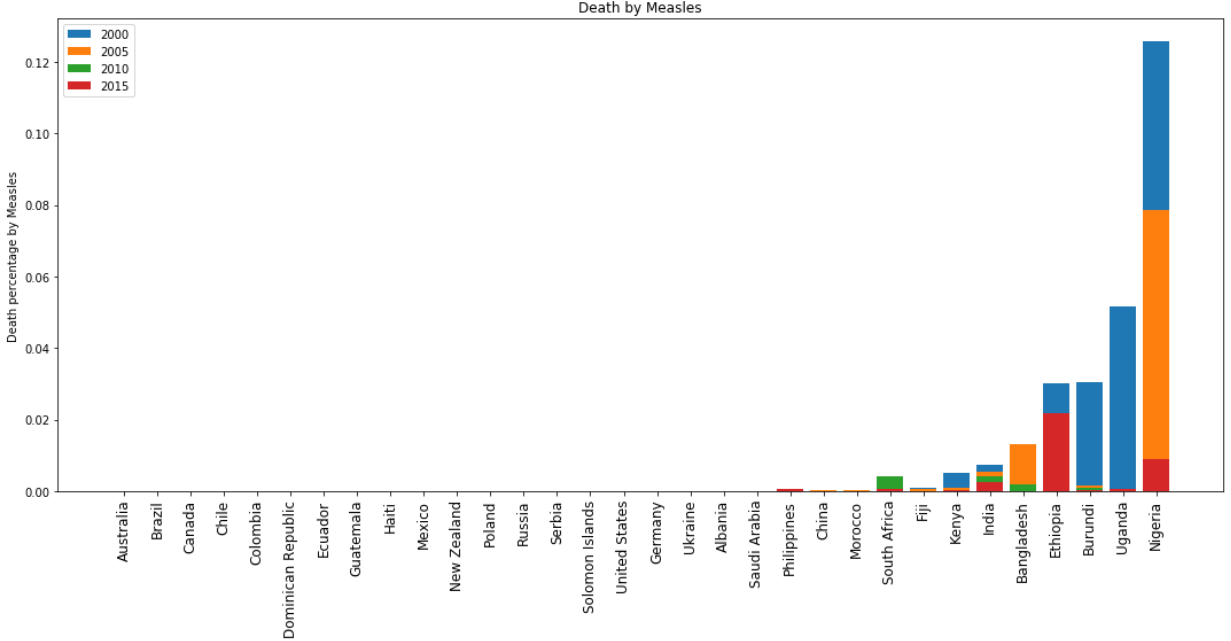
After cleaning the data, we used Python’s pandas and matplotlib libraries to plot and analyze what the data contains. Each group member from the team took some of the data and created various graphs and plots that displayed the data in visual form. For example, in order to understand the GDP, we created a vertical stacked bar graph, showing the change of GDP in each country over time:



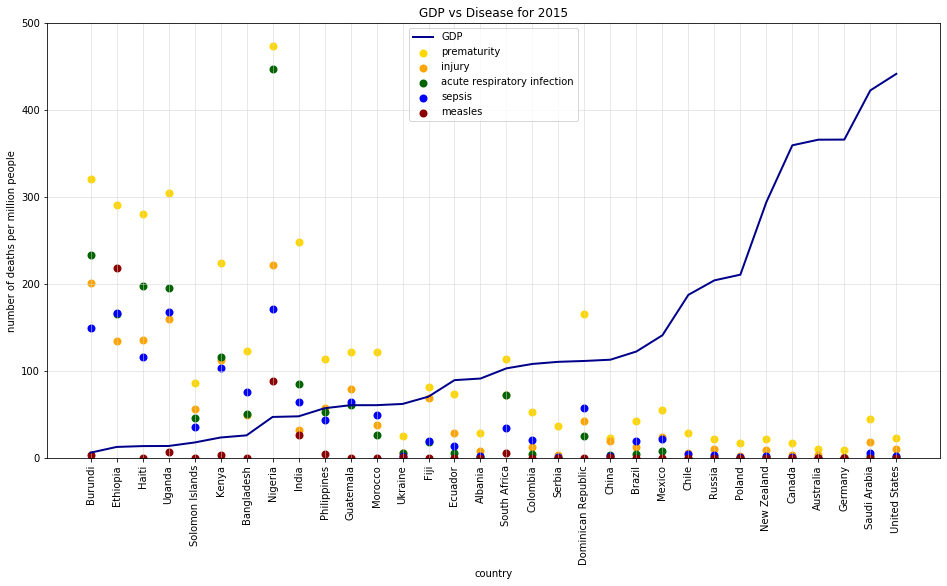
By creating similar graphs for each particular disease, we obtained graphs with the y-axis somewhat reversed. For example, the chart displaying the mortality of children dying from prematurity is the following:

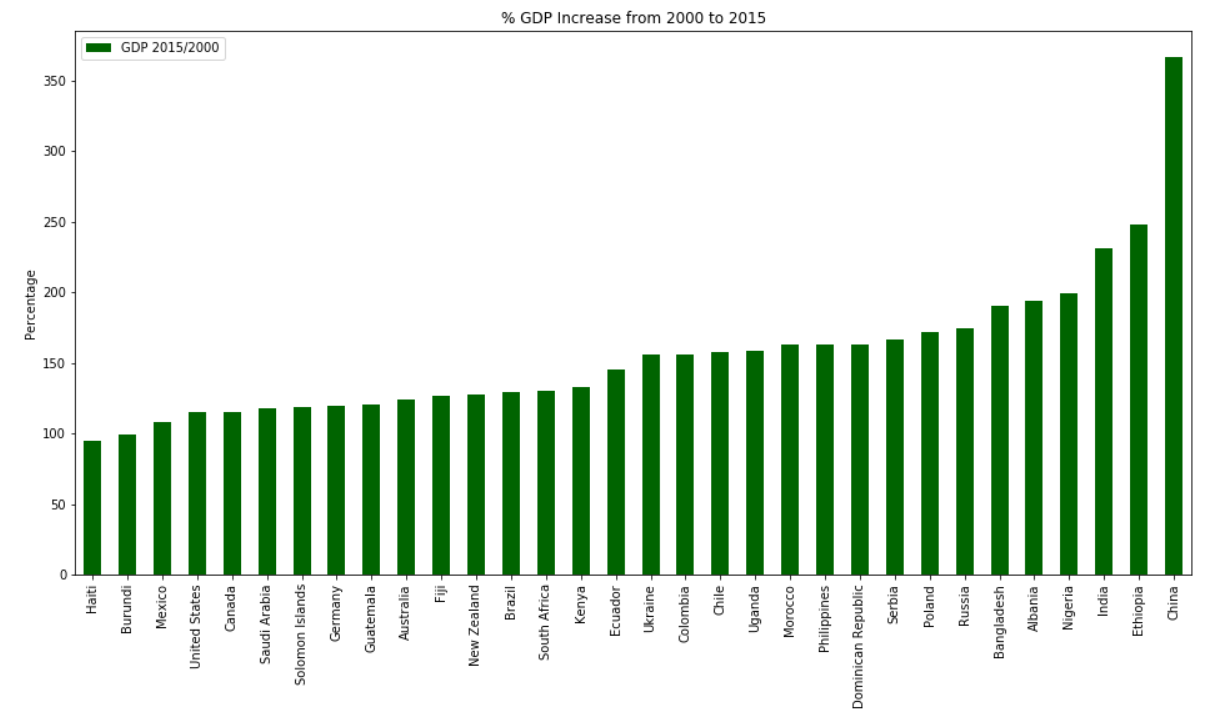


While most diseases showed considerable amounts of cases, we found that there are few cases of measles in most countries. However, some African countries still struggle with this disease. The following chart reveals the rarity of measles in most countries worldwide as well as the large drop from 2000 to 2015 in Nigeria, Uganda, and Burundi:

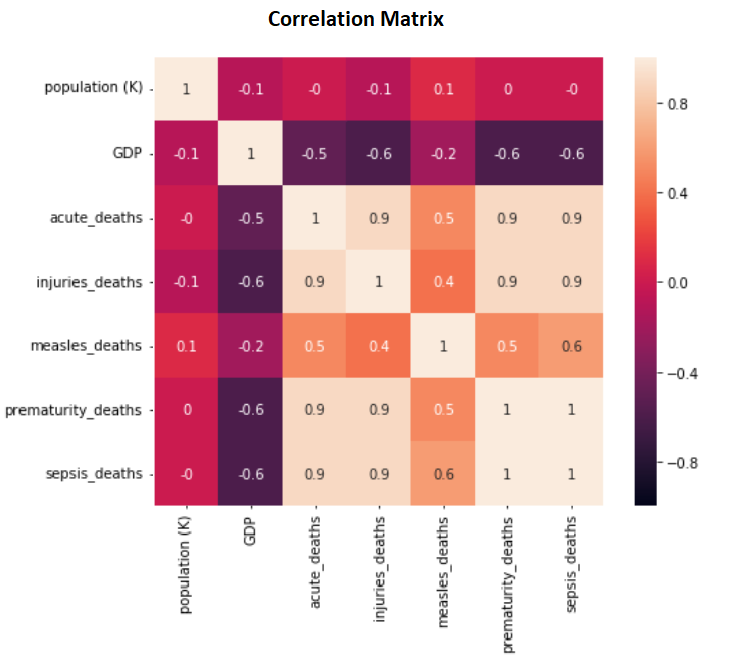


After doing similar and other things to other portions of the datasets, we were able to compare how GDP and diseases compare for various countries. Our hypotheses were validated when our graphs showed more cases for less developed nations and fewer cases for more developed nations:



We focused not only on GDP by country but also on how it changed over time. We found that overall the GDP increased for most countries from 2000 to 2015. We can also observe that Burundi’s and Haiti's GDP has stayed the same over the years. Most countries' GDP has increased by 1.5 times or more from 2000 to 2015. China's GDP increased drastically, by 3.5 times.

We also created a correlation matrix to get a more numerical sense of the relationship between various attributes. To do that, we used the seaborn library in python. Then we used matplotlib to create the following visual matrix:



The correlation matrix tells us several stories. First, the dark purple color we see on the matrix stands for a strong negative correlation between the GDP and the various illnesses. This means that high GDP usually signifies low disease, supporting our predictions. The peach color stands for a positive correlation, which exists between the diseases themselves. This means that if one disease is prevalent, others are too. Though this is not always the case, it often is. Finally, the population didn’t seem to have a correlation with the other attributes.

To see more resources or see the distribution for the matrix and scatter plot for different years, refer to Appendix C.

# [**Case Studies**](#l9kb77ogp9rw)

### **Nigeria**

Nigeria’s GDP per capita has doubled since 2000. It has been growing steadily at 7% a year due to its oil industry. From 2000 to 2015, the child mortality rate has fallen by 50%, but is still very high due to:

* Government corruption
* Terrorism caused by Boko Haram and ISIS, which cause displacement and famine in many regions.
* Poverty, with half the population earning less than $2 a day.

### **Haiti & Dominican Republic**

Both Haiti and the Dominican Republic have higher child mortality from injury values due to natural disasters devastating the islands in the Caribbean islands. In September 2004, Hurricane Jeanne killed 3,000 in Haiti. From August to September 2008, Hurricanes Fey, Gustav, Hanna, and Ike killed 800. On January 10th, 2010, a 7.0 magnitude earthquake hit Haiti, killing approximately 300,000.

### **India**

Across 184 countries, the rate of preterm birth ranges from 5% to 18% of babies born. Almost 1 in 10 babies are born prematurely. In India, out of 27 million babies born every year, 3.5 million babies born are premature. In low-income settings, half of the babies born at or below 32 weeks (2 months early) die due to a lack of feasible, cost-effective care.

# [**Results**](#l9kb77ogp9rw)

We revisit the three questions that we were trying to solve before performing our findings on child mortality:

### How does the child mortality rate relate to the development in countries worldwide?

### How does child mortality of measles compare to other diseases?

### What kind of diseases are highly linked to underdeveloped and developing countries compared to developed ones?

We found that looking at the graphs and correlation matrix, we can note that as GDP increases, child mortality caused by respiratory infection, injuries, prematurity, and sepsis decreases. Countries with lower GDP lack medical care and resources that developed countries have. Due to this, many infants get sick and die. Countries with higher GDP have the resources to prevent many mortality cases. For example, Germany has the 3rd highest GDP per capita and has little to no child deaths from disease.

Based on the measles disease graph, one can note that 12 out of the 32 countries were affected by measles while most countries overall have been affected by acute lower respiratory infections, sepsis, prematurity, and various injuries. Based on the correlation matrix, there is only a slight correlation between the mortality rate of measles and that of other diseases. Measles seems to have much fewer cases than other diseases. Because of this, the correlation is harder to spot.

According to our results, all the diseases turned out more or less to be more prevalent in under-developed and developing countries. Child mortality caused by injury and prematurity are more prominent in developing countries, due to the fact that the other causes are more easily prevented in countries with more money to fund medical care, better access to sanitation across the population, and better access to clean water across the population.

# [**Conclusion**](#l9kb77ogp9rw)

Based on our results, we were able to successfully answer our three questions. We also confirmed that our hypotheses were correct based on the supporting data. Additionally, we were able to notice trends and see patterns. We noted that some continents, such as Africa, have similar demographics throughout, while others, such as Asia and North America, vary largely from country to country. Finally, we provided an overview that shows the data in an understandable way to most readers. The analysis findings can be used by data scientists, decision-makers, and anyone interested in this topic.

# **A**[**ppendix A: Works Cit**](#l9kb77ogp9rw)**ed**

“GDP per Capita.” *Our World in Data*, ourworldindata.org/grapher/gdp-per-capita-worldbank?tab=chart&year=earliest&time=latest&country=USA.

“List of Countries by Past and Estimated Future Population.” *Wikipedia*, Wikimedia Foundation, 4 Apr. 2020, en.wikipedia.org/wiki/List\_of\_countries\_by\_past\_and\_estimated\_future\_population.

“GHO | By Category | Number of Deaths by Country - Acute Lower Respiratory Infections.” *World Health Organization*, World Health Organization, apps.who.int/gho/data/view.main.ghe1002015-CH9?lang=en.

“GHO | By Category | Number of Deaths by Country - Injuries.” *World Health Organization*, World Health Organization, apps.who.int/gho/data/view.main.ghe1002015-CH17?lang=en.

“GHO | By Category | Number of Deaths by Country - Measles.” *World Health Organization*, World Health Organization, apps.who.int/gho/data/view.main.ghe1002015-CH6?lang=en.

“GHO | By Category | Number of Deaths by Country - Prematurity.” *World Health Organization*, World Health Organization, apps.who.int/gho/data/view.main.ghe1002015-CH10?lang=en.

“GHO | By Category | Number of Deaths by Country - Sepsis and Other Infectious Conditions of the Newborn.” *World Health Organization*, World Health Organization, apps.who.int/gho/data/view.main.ghe1002015-CH12?lang=en.

“The History of Natural Disasters in Haiti - UFondwa-USA, Inc.” *UFondwa*, 23 Feb. 2018, ufondwa.org/history-natural-disasters-haiti/.

“Hurricane Relief.” *USAID: Caribbean Hurricane Update*, 15 Oct. 2004, web.archive.org/web/20041015085858/hurricane.info.usaid.gov/.

“Is India's Healthcare System Failing Premature Babies Even Before They Are Born?” *Healthy Newborn Network*, 17 Nov. 2018, www.healthynewbornnetwork.org/news-item/is-indias-healthcare-system-failing-premature-babies-even-before-they-are-born/.

Pallardy, Richard. “2010 Haiti Earthquake.” *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., 15 Jan. 2010, www.britannica.com/event/2010-Haiti-earthquake.

Rastogi, Aruna. “Preterm Birth.” *Preterm Birth | National Health Portal Of India*, 23 May 2016, www.nhp.gov.in/disease/reproductive-system/female-gynaecological-diseases-/preterm-birth.

Vank, Mason. *World Flag Map*. 31 Jan. 2015, commons.wikimedia.org/wiki/File:World\_Flag\_map.png.

“The World Factbook: Nigeria.” *Central Intelligence Agency*, Central Intelligence Agency, 1 Feb. 2018, www.cia.gov/library/publications/the-world-factbook/geos/ni.html.

# [**Appendix B: Code**](#l9kb77ogp9rw)

The code for cleaning and creating plots can be found in our GitHub repo at <https://github.com/ashah244/CS-418-Project>. The latter is here for reference:

## [Introduction](#_u3k3maz7cgqh)

from IPython.core.interactiveshell import InteractiveShell

InteractiveShell.ast\_node\_interactivity = "all"

import numpy as np

import pandas as pd

from matplotlib import pyplot as plt

import seaborn as sn

CS418 Data Science Final Project

Names: Abhi Shah, Brian De Villa, Katherine Misyutina, Matthew Jankowski

Dataset Topic:

Child Mortality for children under the age of 5. Compare different countries and diseases. Show differences between developed and developing countries.

Criteria:

In early April (exact date to be announced) you must submit your progress report. Your progress report must contain the following:

1)An introduction part to your data:

\* Data spec: describe your data. Include the format and any assumptions about your data, size of the dataset

\* A link to your full data in downloadable form, you can keep your data on Google Drive, Box, DropBox, GitHub, or personal website

\* A sample of your data ( n = 10 - 50)

\* A report of your data collection process

\* How did you collect your data

\* How did you clean your data

\* Mention any difficulties you faced in the beginning steps

2) A summary of challenges and observations you have made so far. A brief mention of your next steps and what you plan to do with your data as you move into the analysis (If you are already in the analysis phase you can mention that as well) Group member duties

## [Load in Data](#_7nrampm0rsbg)

This section loads in all the cleaned data that we did.

dfPopulation = pd.read\_csv('../datasets/CleanedPopulation/Clean\_population.csv')

dfPopulation2 = pd.read\_csv('../datasets/CleanedPopulation/Clean\_population.csv', index\_col=0)

dfAcute\_Lower\_Respiratory\_infections = pd.read\_csv('../datasets/CleanedDiseases/CleanedAcute\_Lower\_Respiratory\_infections.csv')

dfInjuries = pd.read\_csv('../datasets/CleanedDiseases/CleanedInjuries.csv')

dfMeasles\_mortality = pd.read\_csv('../datasets/CleanedDiseases/Cleanedmeasles\_mortality.csv')

dfPrematurity = pd.read\_csv('../datasets/CleanedDiseases/CleanedPrematurity.csv')

dfSepsis = pd.read\_csv('../datasets/CleanedDiseases/CleanedSepsis.csv')

dfGDP = pd.read\_csv('../datasets/CleanedGDP/Clean\_gdp2.csv', index\_col=0)

dfGDP2 = pd.read\_csv('../datasets/CleanedGDP/Clean\_gdp2.csv')

## [Calculations](#_kifaprgbffkg)

This section gets the percent of diseased children over the total population of the country during that year. It also gets the difference in GDP from 2000 to 2015.

#calculates the percent of diseases over the total population for each country in each year

acute\_lower\_respiratory\_infection\_percentage = (dfAcute\_Lower\_Respiratory\_infections.iloc[:, 1:5]/(dfPopulation.iloc[:,1:5] \* 1000)) \* 100

injuries\_percentage = (dfInjuries.iloc[:, 1:5]/(dfPopulation.iloc[:,1:5] \* 1000)) \* 100

measles\_mortality\_percentage = (dfMeasles\_mortality.iloc[:, 1:5]/(dfPopulation.iloc[:,1:5] \* 1000)) \* 100

prematurity\_percentage = (dfPrematurity.iloc[:, 1:5]/(dfPopulation.iloc[:,1:5] \* 1000)) \* 100

sepsis\_percentage = (dfSepsis.iloc[:, 1:5]/(dfPopulation.iloc[:, 1:5] \* 1000)) \* 100

gdp\_difference\_percentage = (dfGDP['2015']/dfGDP['2000']) \* 100

## [Functions to help with plot](#_cb0oljq4553j)

Here there will be 2 functions. One to create the percentages dataframe and another to sort the percentages. These functions will help later on to plot in the way we want to plot.

#makes some empty dataframes to be used

dfAcute\_percentage = pd.DataFrame(columns =['Countries', '2000', '2005', '2010', '2015'])

dfInjuries\_percentage = pd.DataFrame(columns =['Countries', '2000', '2005', '2010', '2015'])

dfMeasles\_percentage = pd.DataFrame(columns =['Countries', '2000', '2005', '2010', '2015'])

dfPrematurity\_percentage = pd.DataFrame(columns =['Countries', '2000', '2005', '2010', '2015'])

dfSepsis\_percentage = pd.DataFrame(columns =['Countries', '2000', '2005', '2010', '2015'])

dfGDP\_difference\_percentage = pd.DataFrame(columns = ['GDP 2015/2000'])

#creates a dataset that combines both countries name and newly made percentages

def createPercentageDataframe(tempDF, percentageDF, countriesDataFrame):

tempDF['Countries'] = countriesDataFrame

for year in percentageDF:

tempDF[year] = percentageDF[str(year)]

createPercentageDataframe(dfAcute\_percentage, acute\_lower\_respiratory\_infection\_percentage, dfAcute\_Lower\_Respiratory\_infections.iloc[:, 0])

createPercentageDataframe(dfInjuries\_percentage, injuries\_percentage, dfAcute\_Lower\_Respiratory\_infections.iloc[:, 0])

createPercentageDataframe(dfMeasles\_percentage, measles\_mortality\_percentage, dfAcute\_Lower\_Respiratory\_infections.iloc[:, 0])

createPercentageDataframe(dfPrematurity\_percentage, prematurity\_percentage, dfAcute\_Lower\_Respiratory\_infections.iloc[:, 0])

createPercentageDataframe(dfSepsis\_percentage, sepsis\_percentage, dfAcute\_Lower\_Respiratory\_infections.iloc[:, 0])

dfGDP\_difference\_percentage['GDP 2015/2000'] = gdp\_difference\_percentage

#sorts the created diseases percentage countries names and population dataset countries names in the same way

#so when we plot all indexes line up.

def sortingByPercentage(percentageDataF, dfPopulation):

percentageDataF = percentageDataF.sort\_values(by=['2000','2005','2010','2015'])

dfPopulation['country'] = pd.Categorical(dfPopulation['country'], percentageDataF['Countries'])

dfPopulation = dfPopulation.sort\_values('country')

return percentageDataF, dfPopulation

## [Create Graphs](#_e0wy2fx37z22)

Plots the population, GDP, and the five diseases

### [Population](#_xazh97ooc8z2)

dfPopulation2 = dfPopulation2.sort\_values(by=['2000', '2005', '2010', '2015'])

\_, ax = plt.subplots()

dfPopulation2['2015'].plot(kind='bar', ax=ax, color='Orange',figsize=(16,8),legend=True)

dfPopulation2['2010'].plot(kind='bar', ax=ax, color='DarkGreen',legend=True)

dfPopulation2['2005'].plot(kind='bar', ax=ax, color='Red',legend=True)

dfPopulation2['2000'].plot(kind='bar', ax=ax, color='DarkBlue',legend=True)

plt.title('Countries by Population')

plt.xlabel('country')

plt.ylabel('population')

plt.show();

### [GDP](#_l07le32zomty)

dfGDP = dfGDP.sort\_values(by=['2015','2010','2005','2000'])

\_, ax = plt.subplots()

dfGDP['2015'].plot(kind='bar', ax=ax, color='Orange',figsize=(16,8),legend=True)

dfGDP['2010'].plot(kind='bar', ax=ax, color='DarkGreen',legend=True)

dfGDP['2005'].plot(kind='bar', ax=ax, color='Red',legend=True)

dfGDP['2000'].plot(kind='bar', ax=ax, color='DarkBlue',legend=True)

plt.title("GDP per Capita Over 15 Years")

plt.xlabel("")

plt.ylabel("Dollars")

plt.show();

### [Acute Lower Respiratory](#_l2u65fzcbh6y)

plt.figure(1, figsize = (15,8))

dfAcute\_percentage, dfPopulation = sortingByPercentage(dfAcute\_percentage, dfPopulation)

plt.bar(dfPopulation.iloc[:, 0], dfAcute\_percentage.iloc[:, 1])

plt.bar(dfPopulation.iloc[:, 0], dfAcute\_percentage.iloc[:, 2])

plt.bar(dfPopulation.iloc[:, 0], dfAcute\_percentage.iloc[:, 3])

plt.bar(dfPopulation.iloc[:, 0], dfAcute\_percentage.iloc[:, 4])

plt.title("Acute Lower Respiratory Infection")

plt.xticks(rotation = 90)

plt.ylabel("Acute Lower Respiratory Infection Percentage")

plt.legend(["2000", "2005", "2010","2015"])

plt.tick\_params(axis = 'x', labelsize =12)

plt.tight\_layout()

plt.show();

### [Injuries](#_59c04vugog47)

plt.figure(2, figsize = (15,8))

dfInjuries\_percentage, dfPopulation = sortingByPercentage(dfInjuries\_percentage, dfPopulation)

plt.bar(dfPopulation.iloc[:, 0], dfInjuries\_percentage.iloc[:, 1])

plt.bar(dfPopulation.iloc[:, 0], dfInjuries\_percentage.iloc[:, 2])

plt.bar(dfPopulation.iloc[:, 0], dfInjuries\_percentage.iloc[:, 3])

plt.bar(dfPopulation.iloc[:, 0], dfInjuries\_percentage.iloc[:, 4])

plt.title("Death by Injuries")

plt.xticks(rotation = 90)

plt.ylabel("Death percentage by Injuries")

plt.ylim(0, .05)

plt.legend(["2000", "2005", "2010","2015"])

plt.tick\_params(axis = 'x', labelsize =12)

plt.tight\_layout()

plt.text("Haiti", .04, "Haiti actually has a .37", horizontalalignment='left')

plt.show();

### [Measles](#_794pl05wf4ze)

plt.figure(3, figsize = (15,8))

dfMeasles\_percentage, dfPopulation = sortingByPercentage(dfMeasles\_percentage, dfPopulation)

plt.bar(dfPopulation.iloc[:, 0], dfMeasles\_percentage.iloc[:, 1])

plt.bar(dfPopulation.iloc[:, 0], dfMeasles\_percentage.iloc[:, 2])

plt.bar(dfPopulation.iloc[:, 0], dfMeasles\_percentage.iloc[:, 3])

plt.bar(dfPopulation.iloc[:, 0], dfMeasles\_percentage.iloc[:, 4])

plt.title("Death by Measles")

plt.xticks(rotation = 90)

plt.ylabel("Death percentage by Measles")

plt.legend(["2000", "2005", "2010","2015"])

plt.tick\_params(axis = 'x', labelsize =12)

plt.tight\_layout()

plt.show();

### [Prematurity](#_o0zocdhzqa0u)

plt.figure(4, figsize = (15,8))

dfPrematurity\_percentage, dfPopulation = sortingByPercentage(dfPrematurity\_percentage, dfPopulation)

plt.bar(dfPopulation.iloc[:, 0], dfPrematurity\_percentage.iloc[:, 1])

plt.bar(dfPopulation.iloc[:, 0], dfPrematurity\_percentage.iloc[:, 2])

plt.bar(dfPopulation.iloc[:, 0], dfPrematurity\_percentage.iloc[:, 3])

plt.bar(dfPopulation.iloc[:, 0], dfPrematurity\_percentage.iloc[:, 4])

plt.title("Death by Prematurity")

plt.xticks(rotation = 90)

plt.ylabel("Death percentage by Prematurity")

plt.legend(["2000", "2005", "2010","2015"])

plt.tick\_params(axis = 'x', labelsize =12)

plt.tight\_layout()

plt.show();

### [Sepsis](#_bhe78wty1oao)

plt.figure(5, figsize = (15,8))

dfSepsis\_percentage, dfPopulation = sortingByPercentage(dfSepsis\_percentage, dfPopulation)

plt.bar(dfPopulation.iloc[:, 0], dfSepsis\_percentage.iloc[:, 1])

plt.bar(dfPopulation.iloc[:, 0], dfSepsis\_percentage.iloc[:, 2])

plt.bar(dfPopulation.iloc[:, 0], dfSepsis\_percentage.iloc[:, 3])

plt.bar(dfPopulation.iloc[:, 0], dfSepsis\_percentage.iloc[:, 4])

plt.title("Death by Sepsis")

plt.xticks(rotation = 90)

plt.ylabel("Death percentage by Sepsis")

plt.legend(["2000", "2005", "2010","2015"])

plt.tick\_params(axis = 'x', labelsize =12)

plt.tight\_layout()

plt.show();

### [GDP Increase from 2000-2015](#_4hpofxaaoc01)

This part shows all the gdp from 2000-2015 and shows the percent change

dfGDP\_difference\_percentage = dfGDP\_difference\_percentage.sort\_values(by=['GDP 2015/2000'])

\_, ax = plt.subplots()

dfGDP\_difference\_percentage['GDP 2015/2000'].plot(kind='bar', ax=ax, color='darkgreen',figsize=(16,8),legend=True)

plt.title("% GDP Increase from 2000 to 2015")

plt.xlabel("")

plt.ylabel("Percentage")

plt.show();

## [Merge Dataset](#_vqrevi57n1kf)

This section merges to create new datasets that combine the gdp and disease death per million.

merged2015 = pd.DataFrame()

merged2010 = pd.DataFrame()

merged2005 = pd.DataFrame()

merged2000 = pd.DataFrame()

#If year == 2015, indexOfYear = 4

#If year == 2010, indexOfYear = 3

#If year == 2005, indexOfyear = 2

#If year == 2000, indexOfyear = 1

#function that merges the year wanted and gdp into one dataframe

def dfGdpAndDisease(year, popYear, gdpYear, acuteYear, injuriesYear, measlesYear, prematurityYear, sepsisYear, indexOfYear, merged):

#reads in only the year column we want

popYear = popYear.iloc[:,[0,indexOfYear]]

gdpYear = gdpYear.iloc[:,[0,indexOfYear]]

acuteYear = acuteYear.iloc[:, [0,indexOfYear]]

injuriesYear = injuriesYear.iloc[:,[0,indexOfYear]]

measlesYear = measlesYear.iloc[:,[0, indexOfYear]]

prematurityYear = prematurityYear.iloc[:,[0, indexOfYear]]

sepsisYear = sepsisYear.iloc[:,[0, indexOfYear]]

#then renames the columns inside of it

popYear = popYear.rename(columns={year : 'population (K)'}) #, inplace=True)

gdpYear = gdpYear.rename(columns={'Country' : 'country', year : 'GDP'}) # , inplace=True)

gdpYear.GDP = gdpYear.GDP.astype(int)

acuteYear = acuteYear.rename(columns={'Countries' : 'country', year : 'acute\_deaths'}) # , inplace=True)

injuriesYear = injuriesYear.rename(columns={'Countries' : 'country', year : 'injuries\_deaths'}) #, inplace=True)

measlesYear = measlesYear.rename(columns={'Countries' : 'country', year : 'measles\_deaths'}) #, inplace=True)

prematurityYear = prematurityYear.rename(columns={'Countries' : 'country', year :'prematurity\_deaths'}) #, inplace=True)

sepsisYear = sepsisYear.rename(columns={'Countries' : 'country', year : 'sepsis\_deaths'}) #, inplace=True)

#merge them into one dataset

merged = pd.merge(popYear, gdpYear, on='country', how='inner')

merged = pd.merge(merged, acuteYear, on='country', how='inner')

merged = pd.merge(merged, injuriesYear, on='country', how='inner')

merged = pd.merge(merged, measlesYear, on='country', how='inner')

merged = pd.merge(merged, prematurityYear, on='country', how='inner')

merged = pd.merge(merged, sepsisYear, on='country', how='inner')

return merged

merged2015 = dfGdpAndDisease('2015', dfPopulation, dfGDP2, dfAcute\_Lower\_Respiratory\_infections, dfInjuries, dfMeasles\_mortality, dfPrematurity, dfSepsis, 4, merged2015)

merged2010 = dfGdpAndDisease('2010', dfPopulation, dfGDP2, dfAcute\_Lower\_Respiratory\_infections, dfInjuries, dfMeasles\_mortality, dfPrematurity, dfSepsis, 3, merged2010)

merged2005 = dfGdpAndDisease('2005', dfPopulation, dfGDP2, dfAcute\_Lower\_Respiratory\_infections, dfInjuries, dfMeasles\_mortality, dfPrematurity, dfSepsis, 2, merged2010)

merged2000 = dfGdpAndDisease('2000', dfPopulation, dfGDP2, dfAcute\_Lower\_Respiratory\_infections, dfInjuries, dfMeasles\_mortality, dfPrematurity, dfSepsis, 1, merged2000)

## [GDP vs Disease](#_3b2w9l5v70b0)

This plots the GDP vs Disease for each year. Comparing these graphs we can answer our questions better than the other graphs.

#plot to show GDPvsDisease

def plotGDPvsDisease(mergedYear,counter, year):

plt.figure(counter, figsize=(16,8))

plt.plot('country', 'GDP', data=mergedYear, color='darkblue', linewidth=2, label="GDP")

plt.scatter('country', 'prematurity\_deaths', data=mergedYear, color='gold', linewidth=2, label="prematurity")

plt.scatter('country', 'injuries\_deaths', data=mergedYear, color='orange', linewidth=2, label="injury")

plt.scatter('country', 'acute\_deaths', data=mergedYear, color='darkgreen', linewidth=2, label="acute respiratory infection")

plt.scatter('country', 'sepsis\_deaths', data=mergedYear, color='blue', linewidth=2, label="sepsis")

plt.scatter('country', 'measles\_deaths', data=mergedYear, color='darkred', linewidth=2, label="measles")

plt.title("GDP vs Disease for " + year)

plt.xlabel('country')

plt.ylabel('number of deaths per million people')

plt.legend(loc='center')

plt.grid(linestyle='-', linewidth='0.5', color='lightgray')

plt.xticks(rotation = 90)

plt.legend(loc = "upper center")

plt.ylim(0,500)

plt.show();

plotGDPvsDisease(merged2015, 1, '2015');

plotGDPvsDisease(merged2010, 2, '2010');

plotGDPvsDisease(merged2005, 3, '2005');

plotGDPvsDisease(merged2000, 4, '2000');

## [Correlation Matrix](#_clqaluk3pojk)

A correlation matrix for each year comparing all the graphs and seeing all they relate to each other.

def adjustGDP(mergedDataset):

corrMerged = mergedDataset

corrMerged['GDP'] = mergedDataset['GDP'] \* 0.1

return corrMerged

corrMerged2015 = adjustGDP(merged2015)

corrMerged2010 = adjustGDP(merged2010)

corrMerged2005 = adjustGDP(merged2005)

corrMerged2000 = adjustGDP(merged2000)

def createCorrelationMatrix(adjustedDataset):

corrMatrix = adjustedDataset.corr()

corrMatrix = corrMatrix.round(1)

return corrMatrix

corrMatrix2015 = createCorrelationMatrix(corrMerged2015)

corrMatrix2010 = createCorrelationMatrix(corrMerged2010)

corrMatrix2005 = createCorrelationMatrix(corrMerged2005)

corrMatrix2000 = createCorrelationMatrix(corrMerged2000)

def plotCorrelationMatrix(corrleationDataset, year, counter):

plt.figure(counter, figsize=(8,6))

plt.title(year)

sn.heatmap(corrleationDataset, annot=True, vmin=-1, vmax=1, xticklabels=1, yticklabels=False, square=True)

plt.show();

plotCorrelationMatrix(corrMatrix2015, '2015', 1)

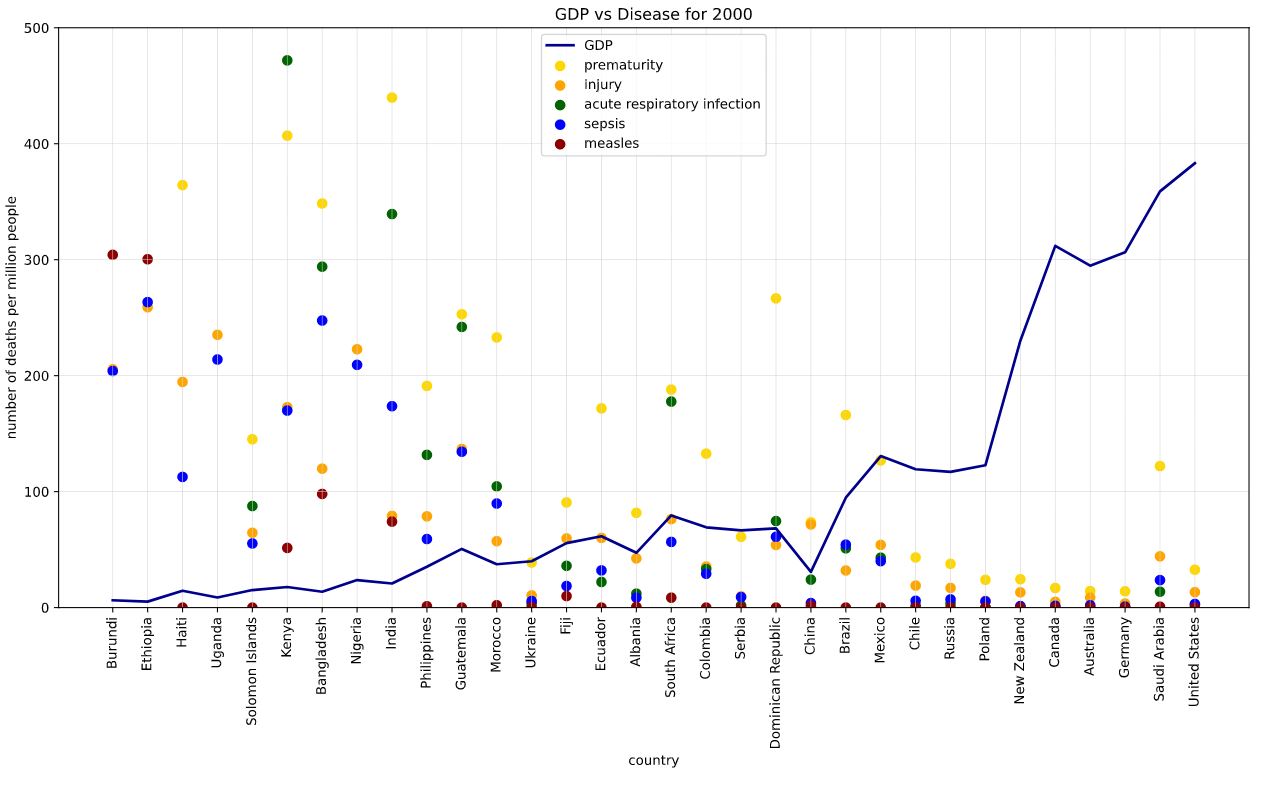
plotCorrelationMatrix(corrMatrix2010, '2010', 2)

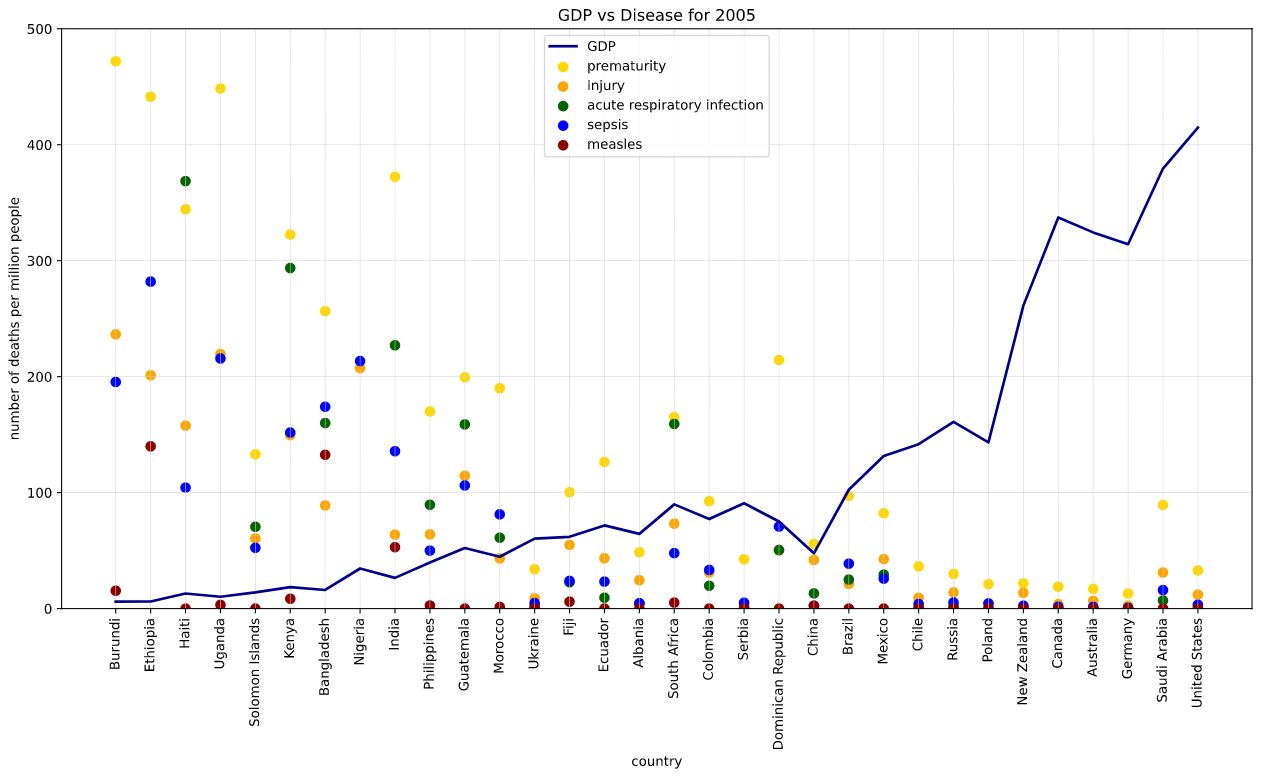
plotCorrelationMatrix(corrMatrix2005, '2005', 3)

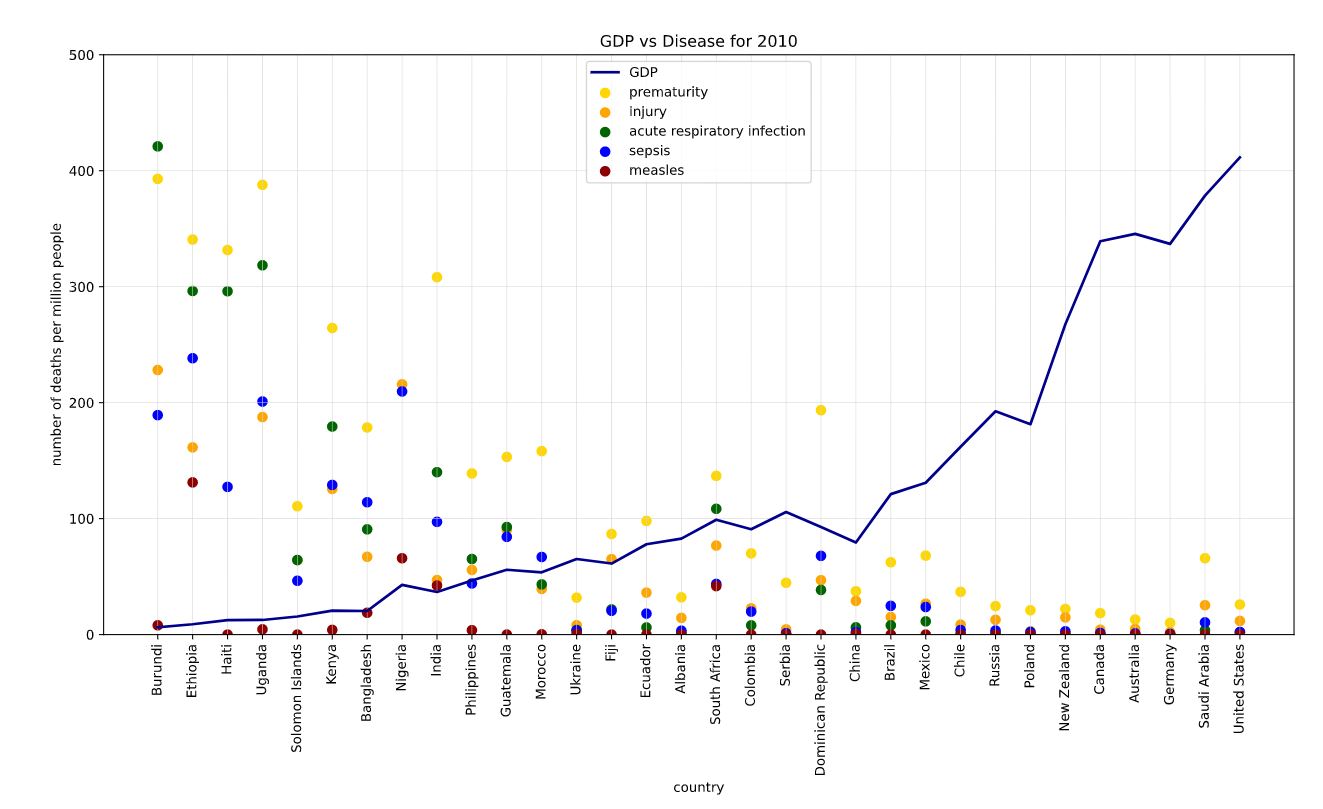
plotCorrelationMatrix(corrMatrix2000, '2000', 4)

# [**Appendix C: References**](#l9kb77ogp9rw)

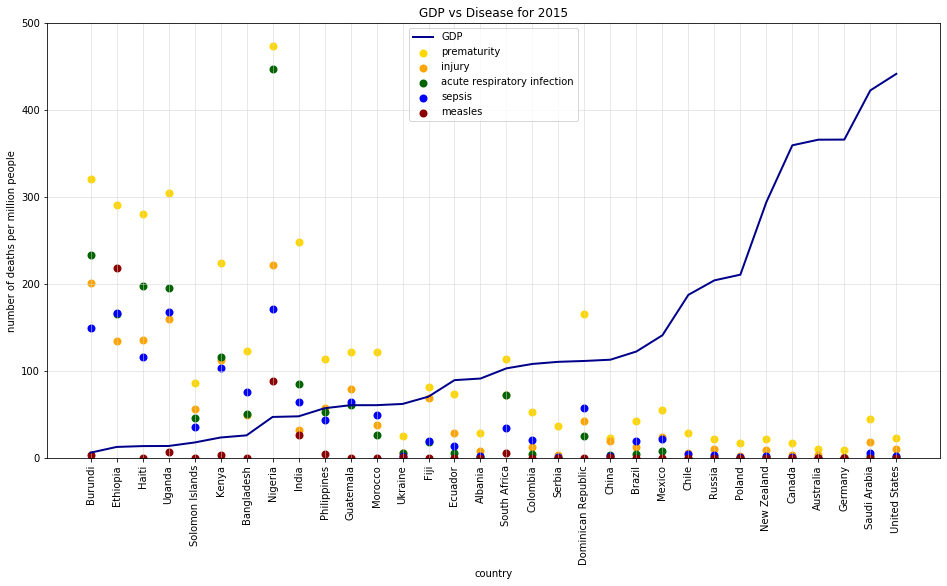
Here one can find graphs and charts not shown in the main part of the report.

2000 GDP vs Disease:

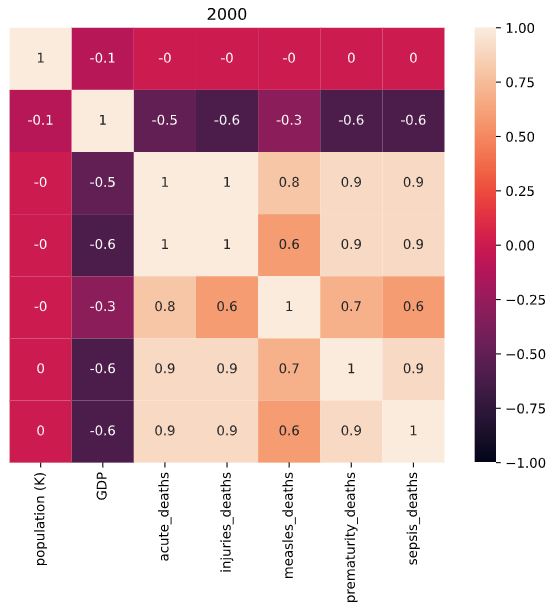
2005 GDP vs Disease:

2010 GDP vs Disease:

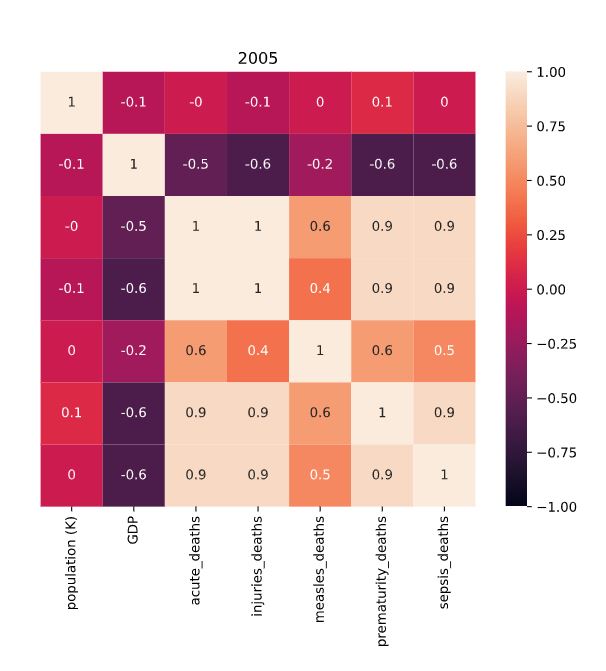
2015 GDP vs Disease:



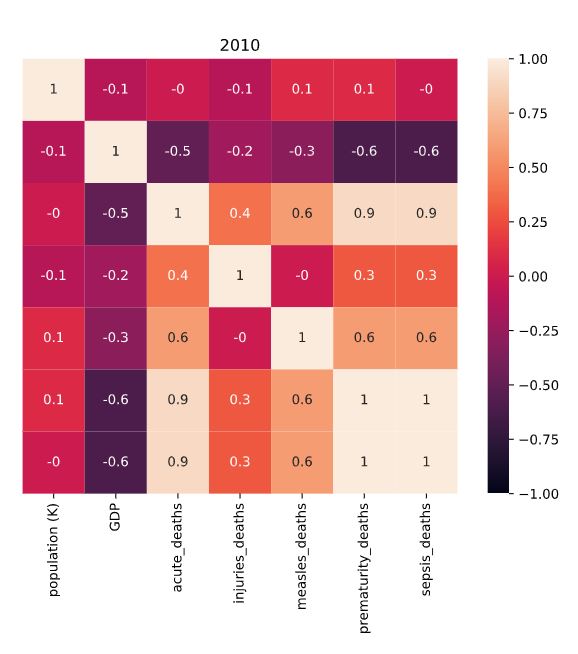
2000 Correlation Matrix:



2005 Correlation Matrix:



2010 Correlation Matrix:



2015 Correlation Matrix:

