**Data Analysis and Visualization - Boot Camp**

**Group Project**

**TEAM-THIS**

Tom Harmon, Ruth Ineh, Zohaib Syed, and Michael LeFrancois

**Background**

Cholera is transmitted through impacted waters, or contaminated food causing infection of the small intestine by some strains of the bacterium Vibrio Cholerae and can result in dehydration (Wikipedia contributors “Cholera”). Cholera is especially endemic in areas lacking adequate clean water infrastructure.

This project compares Cholera cases to water resource indicators with data available from the [Food and Agricultural Organization of the United Nations](http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en) (FAO) and the [World Health Organization](http://www.who.int/gho/database/en/) (WHO). For this analysis two groups of cholera-endemic border countries are investigated: Niger and Nigeria in West Africa and Myanmar and Thailand in Southeast Asia. Comparisons also consider standards of living indicated via the Gross Domestic Product (GDP). Finally, beyond the 4 target countries examined, a 2016 dataset of all countries was also used to understand possible correlations between water resource indicators and Cholera incidence rate.

**Research Questions**

To understand a potential connection between Cholera outbreaks and water resource indicators the following questions were investigated:

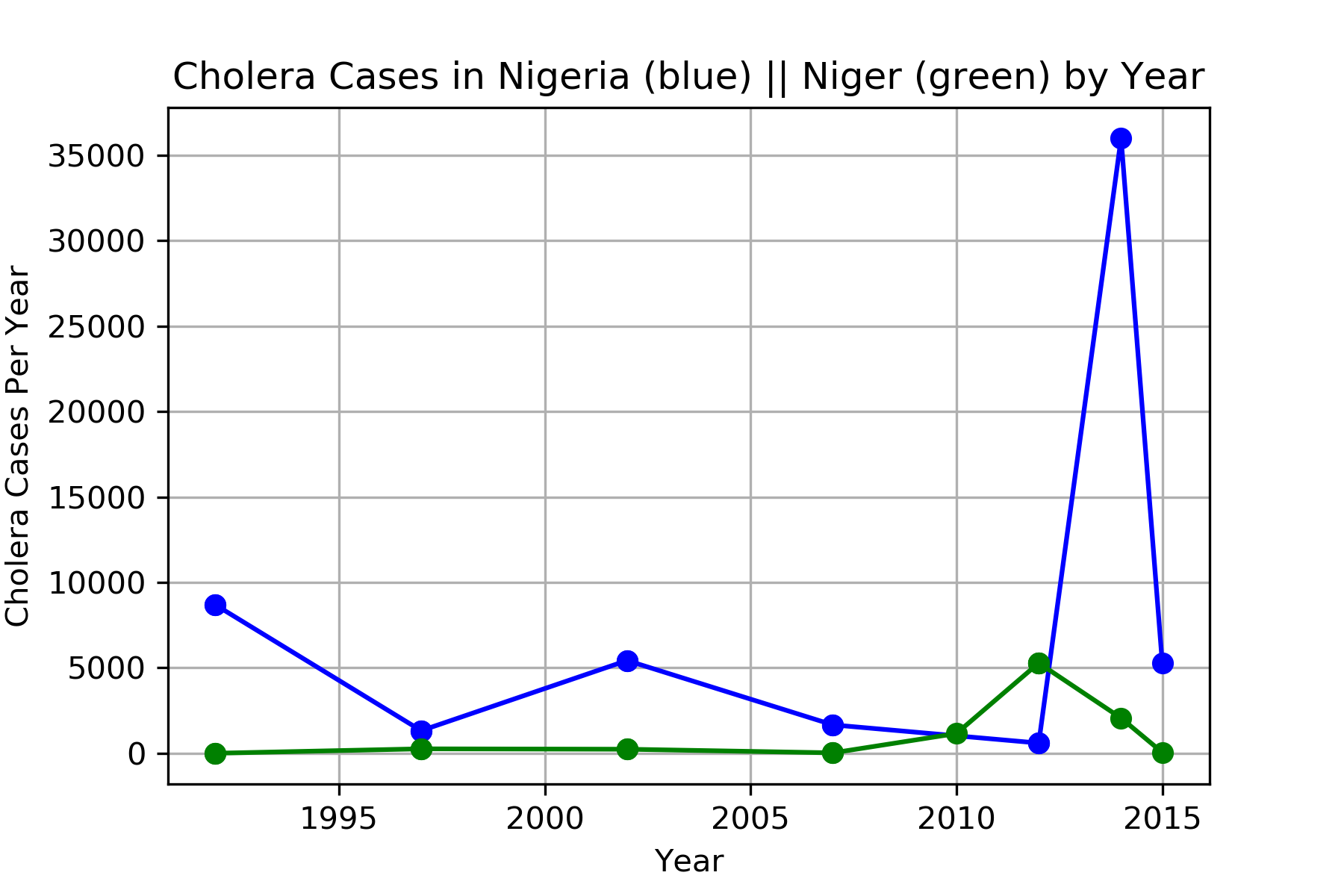
* Is there a correlation between Cholera rates between bordering countries?
* Is there a correlation between water resource indicators and Cholera rates?

**Analysis**

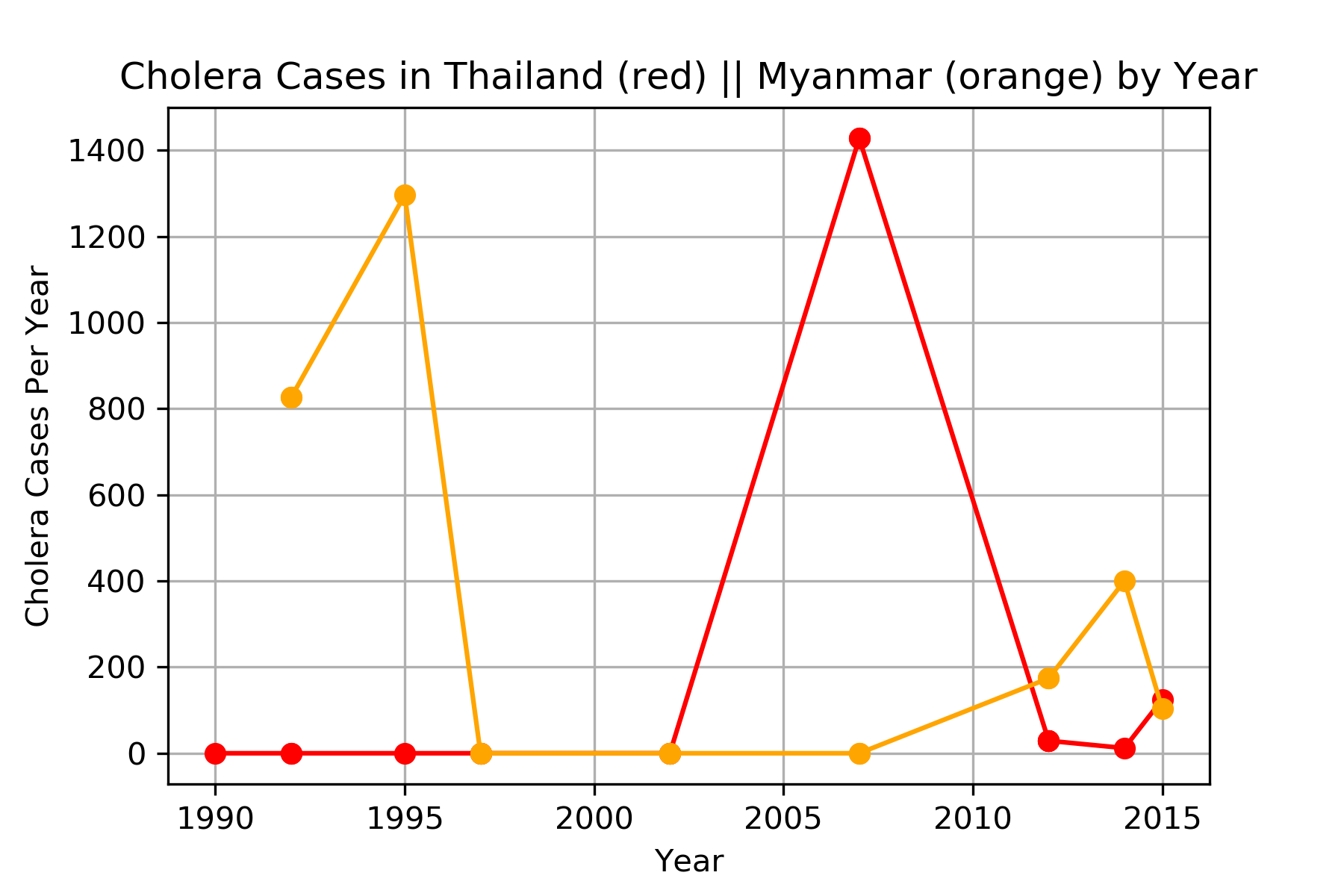
A small sampling of Cholera data in the four target countries over 1990 to 2015 provided the basis for an analysis of Cholera outbreaks. Population and GPD data are summarized in   
**Table** 1; notably, in each pairing one country (Nigeria and Thailand, respectively) has a much higher population and Gross Domestic Product (GDP). Cholera incidences over time between bordering countries are shown in **Figures 1** and **2**. Incidences in the West African countries tended to rise in tandem, with higher cases in Nigeria for all but one year, and an especially high outlier data point correlating with a regional outbreak in 2014. Southeast Asian countries also seem to track together for periods of time and also express similar peaks in number of cases, but in different years.

**Table 1. Country Profiles**

|  |  |  |
| --- | --- | --- |
| **Country** | **Population (Million)** | **GDP (Million USD, 2016)** |
| Nigeria | 186 | 405.1 |
| Niger | 20.6 | 7.5 |
| Thailand | 69 | 406.8 |
| Myanmar | 53 | 67.4 |

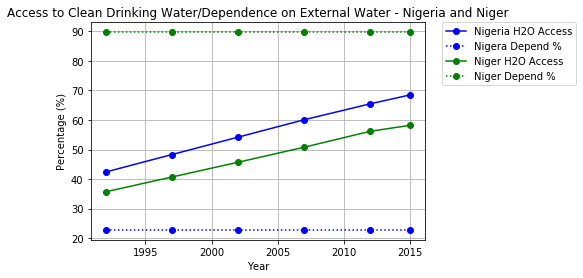


**Figure 1**: Cholera Cases in West African Border Countries.

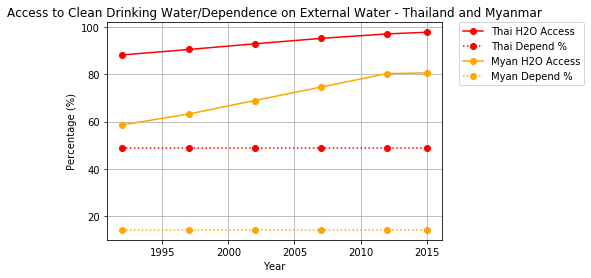


**Figure 2**: Cholera Cases in Southeast Asian Border Countries.

Water resource indicators were also plotted for each border country to help understand potential correlations between cholera incidence, the quality of the nation’s water infrastructure, and shared water. Both access to clean drinking water and dependence are shown for West African (**Figure 3**), and Southeast Asian countries (**Figure 4**). Dependency expresses the percent of total renewable water resources originating outside the country and can vary between 0 to 100 percent. A dependency ratio of zero equates to no water from neighboring countries, whereas 100 percent equates to full dependency from upstream countries.

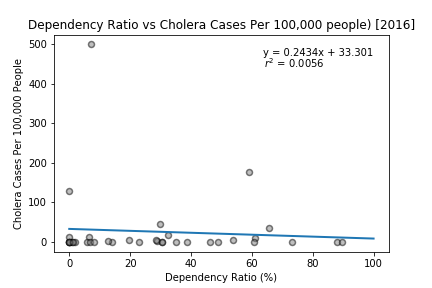


**Figure 3**: Access to Clean Water and Water Dependence in Africa.

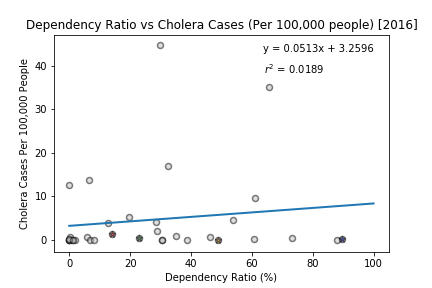


**Figure 4**: Access to Clean Water and Water Dependence in Asia.

For the purposes of discussion thus far, only a selection of bordering countries where analyzed, over a period of 14 years. To support the analysis of potential correlation between water resource indicators and Cholera rates over a larger sample set, all countries with available 2016 data are analyzed in **Figures 5** and **6**.



**Figure 5**: Water Resources Indicator Versus Cholera Cases. Dataset includes outliers



**Figure 6**: Water Resources Indicator Versus Cholera Cases. Outliers removed.

**Conclusions**

This analysis was based on a limited dataset, only using bordering countries within two different parts of the world, for a total of 4 countries. In addition, other categories beyond water resource indicators could also be available to help understand potential correlations to Cholera incidence rates. The data supports the following conclusions:

* Bordering countries show Cholera rates similar to each other and see similar peaks in number of cases, but not necessarily in the same year.
* Access to clean water tends to increase for each country over time, while dependence on water remains constant throughout.
* Rates of Cholera tend to be independent from water resource indicators (within each country).

A direct correlation between water resource indicators and Cholera incidences was not observed from this data. As noted, the entirety of available information was not incorporated into this analysis, which limits a more complete view. A regression analysis of a global dataset, with and without outliers, does not suggest a strong correlation with a low R2 in both cases. (Notably, major outliers in the global analysis are experiencing civil unrest.) However, it can be inferred, with the overall increase in regression statistics without outliers, a higher water dependency ratio increases the potential risk to Cholera infection.

**Source Data for Analysis**

FAO:

Aquastat database of water resource and water quality information.

[**http://www.fao.org/nr/water/aquastat/data/query/results.html**](http://www.fao.org/nr/water/aquastat/data/query/results.html)

WHO:

Database of Cholera results.

[**http://apps.who.int/gho/data/node.country**](http://apps.who.int/gho/data/node.country)