**COVID-19 Numbers in relation to both latitude and GDP**

Analyzing the effects: infections, infections per population, and deaths of COVID-19 on a global scale seeing what regions/countries have been affected more and less severely, see if the countries’ latitude and GDP has a relationship to the COVID infection numbers, and make conclusions to why or why not these correlations exist.

**Hypothesis:** Latitude and COVID have a positive relationship due to environmental factors such as weather, as a location gets further away from the equator, there will be more infections and deaths due to people being inside more often, similar to the relationship with influenza. GDP of a location and COVID effects have a negative relationship: as a country is more wealthy, they will be able to better handle the overall effects of the pandemic, particularly when it comes to deaths and deaths per cases.

**Does latitude positively or negatively affect the COVID-19 infection rate/deaths?**

**Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated**

We have hypothesized that latitude could affect COVID effects due to countries having different environmental, cultural, and weather considerations. As you can see from the above boxplots above the infections and deaths from COVD-19 are influenced by significant outliers that will affect any relationship overall, this will have to be taken into consideration for any analysis. Since the primary factor we are investigating is distance from the equator, we have taken the absolute value of the countries latitude in order to better represent its location in visual form. First we will look at Cases vs. Degrees from the equator:

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

As you can see in the visualizations above this number is significantly influenced by 3 significant outliers and dropping them from the dataset significantly affects the overall correlation. There is a slight positive relationship between number of overall cases and the distance from the equator however the correlation is weak to none overall. Next we will look at the deaths per 100,000 people vs. distance from the equator:

Chart, scatter chart

Description automatically generated

Again there is a slight positive relationship between deaths per 100,000 people in the countries population and the distance from the equator, this time with a stronger correlation of .02269 but still weak overall but analyzing on a global scale, this is the strongest correlation that was found. The possible reasons that there was a stronger correlation could be due to colder temperatures and potentially higher population densities more located in cities as the latitudes increase in absolute value. Eliminating extreme outliers in this case did not have the same effect as noted on the previous example, this is primarily due to having less significant outliers overall. Since, the highest correlation found between COVID effect numbers and distance from the equator is this weak, it appears that COVID-19 infections/deaths are not significantly related to a countries location on the earth and that other factors are to blame for higher and lower impacts.

**Does latitude play a more significant role when we look deeper into countries of similar population size?**

On the larger scale, we found that there was almost no correlation between COVID numbers and distance from the equator and surmised that there must be other factors at play. Controlling for one of those factors, population size, we took a deeper look to see if the correlations changed when looking only at countries with similar population sizes. We did this by breaking down our countries into approximately equal groups by population, and the results were quite interesting! Some of the ranges had negative correlation, some had minimal correlation, and others had low but statistically significant correlation. The highest correlations for total cases came from the population range of 110k to 799k and 11.5m-18.9m:

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

However, there were also ranges with negative or minimal correlation:

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

This pattern continued when comparing with the other statistics for death:

And for mortality rate (death per 1000 cases):