Causes of Death Around the World (1990-2019)

A straightforward way to assess the health status of a population is to focus on mortality (death). By combining it with morbidity (suffering from diseases), we get a more encompassing view on health outcomes.

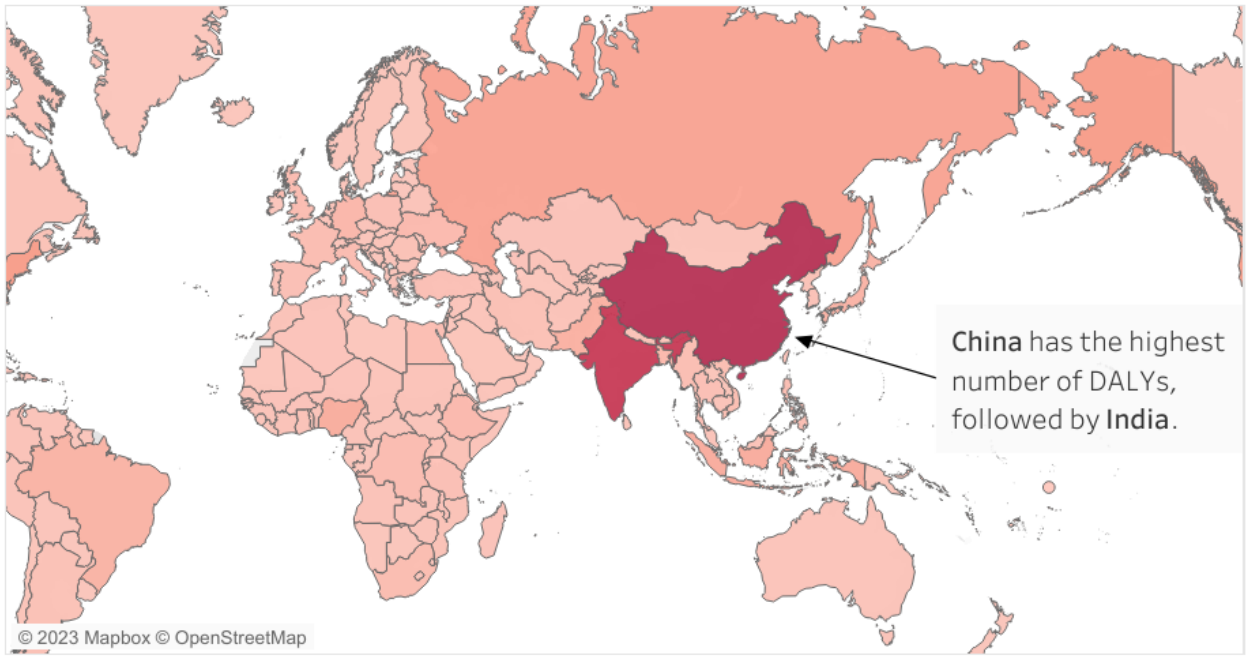
The sum of mortality and morbidity is referred to as the "**burden of disease**" and can be measured by a metric called "**Disability Adjusted Life Years**" (DALYs).

Conceptually, one DALY is the equivalent of losing one year in good health because of either premature death, disease, or disability.

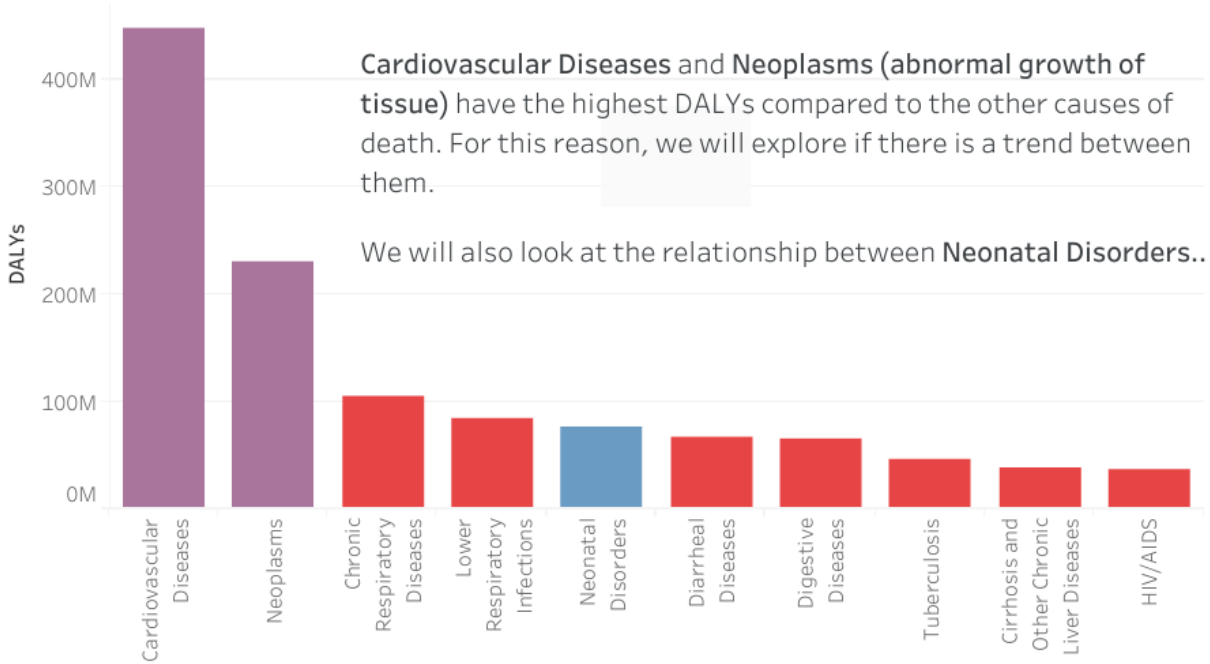
Disability Adjusted Life Year is a measure of overall disease burden, expressed as the cumulative number of years lost due to ill-health, disability or early death

= YLD
Years Lived with Disability

\+ YLL
Years of Life Lost



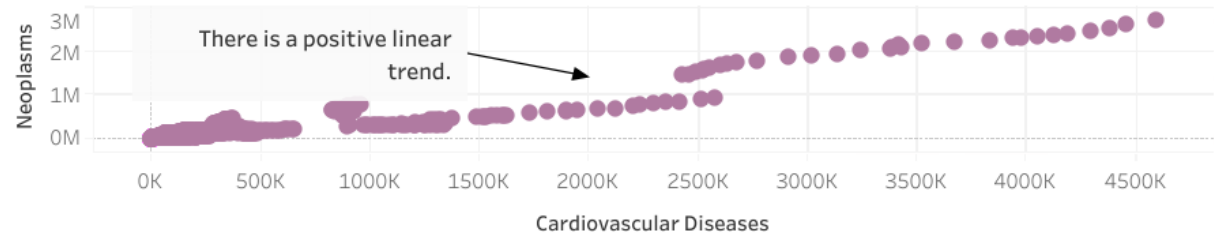
Top 10 Causes of Death Worldwide



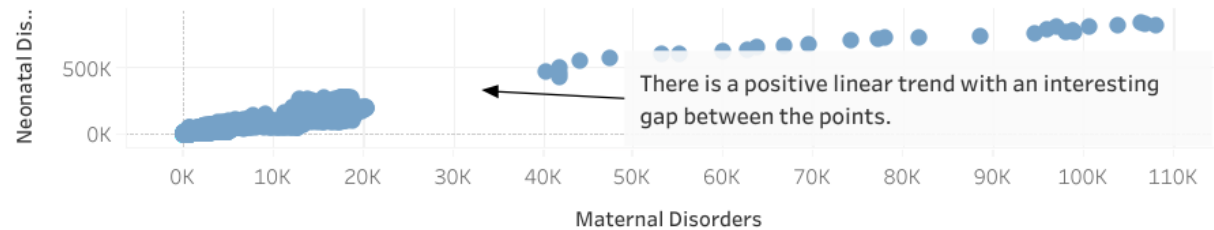
Cardiovascular Diseases and **Neoplasms (abnormal growth of tissue)** have the highest DALYs compared to the other causes of death. For this reason, we will explore if there is a trend between them.

We will also look at the relationship between **Neonatal Disorders..**

Cardiovascular Diseases vs Neoplasms

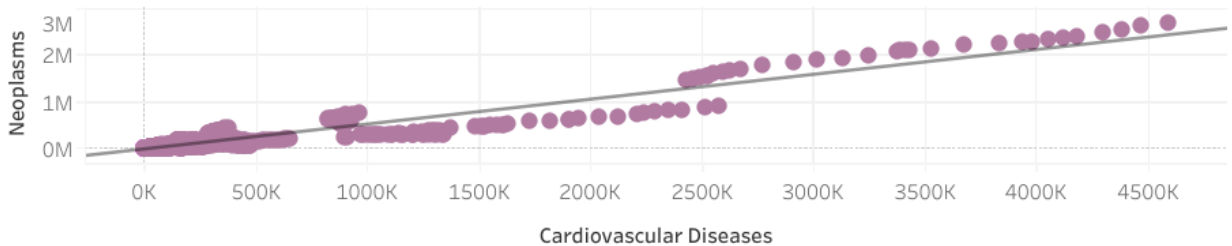


Maternal Disorders vs Neonatal Disorders



Linear Regression

Cardiovascular Diseases vs Neoplasms

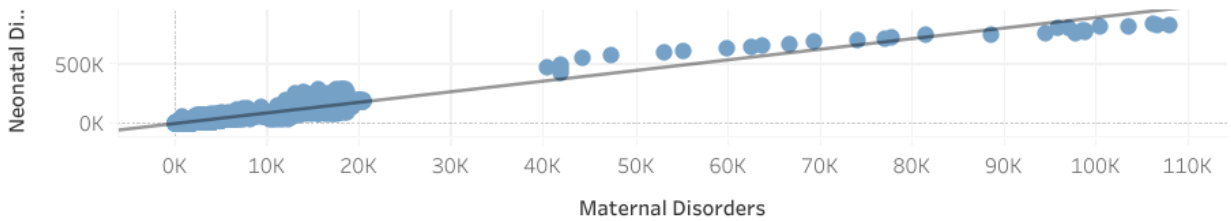


The positive linear trend means as the number of DALYs for cardiovascular diseases increases, the number of DALYs for Neoplasms also increase.

The R-squared value is 0.921 which is very close to 1. This indicates that the linear model is a good fit for the data.

The regression line seems to fit most of the data, but there is a group of data points that are further from the line. Also, the majority of points are concentrated in the bottom left corne..

Maternal Disorders vs Neonatal Disorders



The positive linear trend means as the number of DALYs for maternal disorders increases, the number of DALYs for neonatal disorders also increase.

The R-squared value is 0.942 which is very close to 1. This indicates that the linear model is a good fit for the data.

The majority of points are concentrated in the bottom left corner of the plot. There is also an interesting gap between the bottom left points and the top right points. Conducting a cl..

Cluster Analysis

The data is divided into two groups based on similar and/or closely related values from the entire dataset.

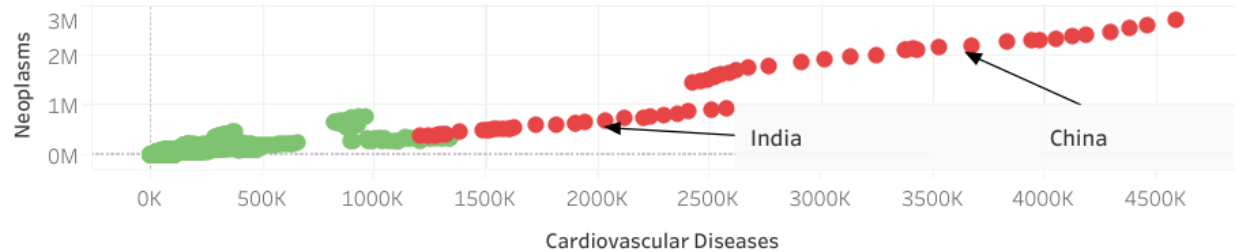
The red cluster has higher DALY's (Disability Adjusted Life Years) on average than the green cluster for cardiovascular diseases and neoplasms. This makes sense since the red cluster consists of China and India which have the largest populations.

However, China has significantly more DALY's in both categories than India, even though their population size is similar. This could be due to another factor such as China having a h..

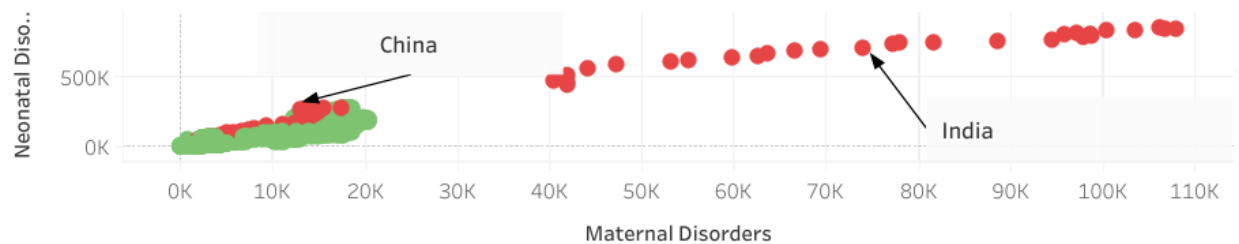
The red cluster has higher DALY's on average than the green cluster for maternal disorders and neonatal disorders.

It is interesting that the number of DALY's are not as high for China which has a similar population size as India. This could be due to another factor such as India having a higher fertility rate than China, especially because China had a one-child policy during most of the years the data was collected from.

Cardiovascular Diseases vs Neoplasms



Maternal Disorders vs Neonatal Disorders



Final Thoughts

Trends do exist between countries for the leading causes of death from 1990 to 2019.

Linear regression showed a strong positive linear relationship between Cardiovascular Diseases and Neoplasms, as well as between Maternal Disorders and Neonatal Disorders.

One limitation of this study is timeliness since the most recent data is from 2019.

Next steps include further investigation of how factors such as life expectancy, population size, sanitation, and access to medical help might have impacted the data. The results could then be used to decrease the numb..