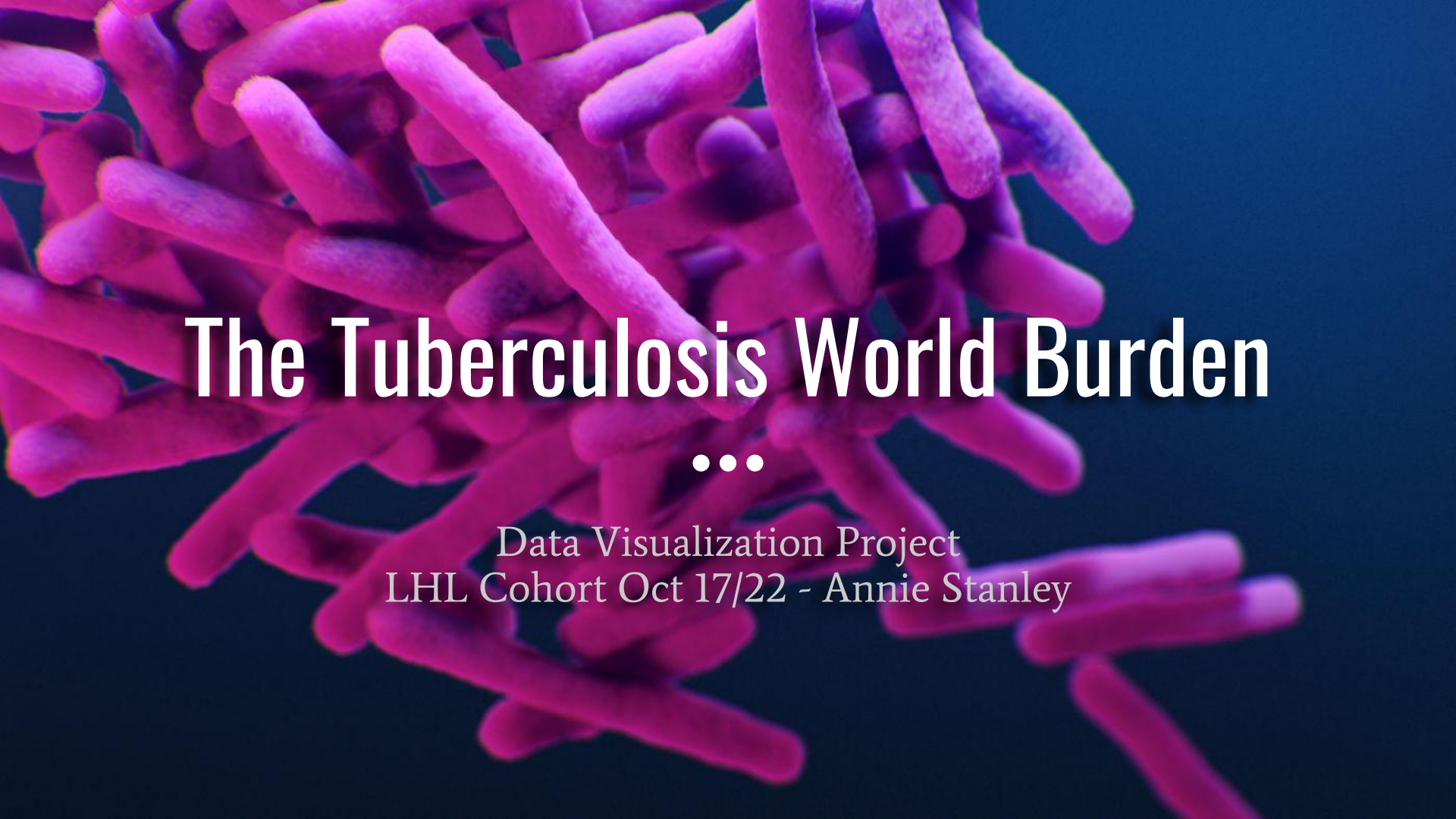


The Tuberculosis World Burden

•••

Data Visualization Project
LHL Cohort Oct 17/22 - Annie Stanley

Project Flow

Basic Steps:

1. Connected data
2. Inspected factors and data types
3. Adjusted data types where necessary
4. Explored the data through various visualizations
5. Identified important features
6. Looked for trends within those features
7. Came up with meaningful questions based on key features
8. Made dashboards based on each question

Basic Methods:

- Lots of data to explore meant lots of time exploring various combinations of data
- Had some ideas of what would correlate, but also spent some time playing around to see if I could find anything meaningful
- Combined some readily available data in order to expand my results

Results:

Important Factors:

- Population
- Geographic location
- Year
- Estimated mortalities (both including and excluding HIV)
- Estimated prevalence
- The variations of those numbers per 100 000 people
- Outside income data

Possible Biases:

- Countries provide their own data, and anything further is extrapolated or predicted
- Countries may not feel comfortable giving accurate numbers
- Predictions may not always be accurate (as we will see later)

Interesting Patterns/Trends/Outliers:

- China and India both have much larger populations and overall TB numbers than many other countries
    - Weren't excluded completely, but they were left out of some visualizations/some data was manipulated in order to show an even spread of data
- Rates of TB have decreased over time, although there was an uptick of deaths in 2002, and because of Covid-19, our forecast may not be completely accurate...

Main Visualization Types:

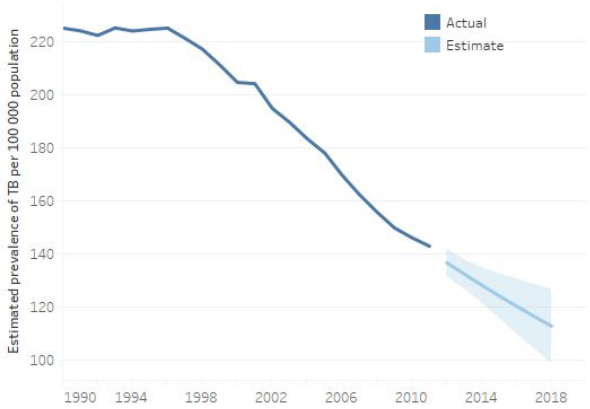
- Maps - Lots of this data was geographical, and maps are the quickest and easiest way to make that information easy to parse
- Line Graphs - Less exciting, but effective for showing changes over time

Q1 - How Has TB Prevalence Changed Over Time

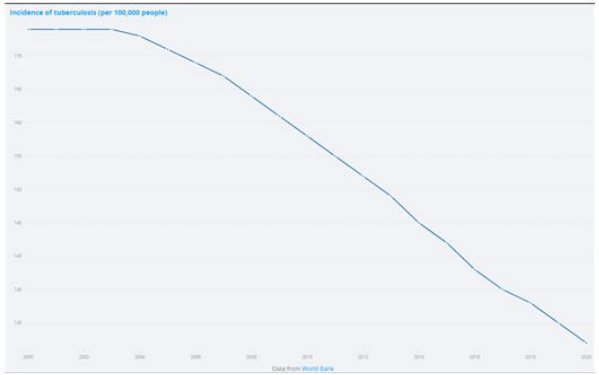
- Simple line graphs with a forecast
- Compared this with current and up to date data as a comparison
  - Model is not actually accurate, as Covid-19 caused TB rates to increase again
- Case detection rate refers to how many cases are actually reported
- We can see that the TB prevalence has decreased as case detection rates have increased
- Chose two separate graphs as opposed to dual, as they have such different y-axis

How Has TB Prevalence Changed Over Time?

Estimated TB Prevalence Over Time (W/ Forecast)



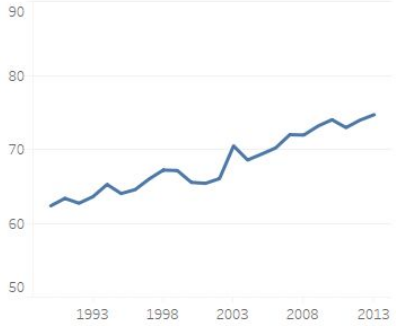
World Bank Tuberculosis Data (Actual)



An Increase In Detection Rates

An increase in case detection rates leads to more accurate data. Higher case detection rates are also assumed to correlate with national funding towards tuberculosis funding. Thus, higher case detection rates correlate with lower cases of tuberculosis over time.

Case Detection Rate



Forecast Accuracy

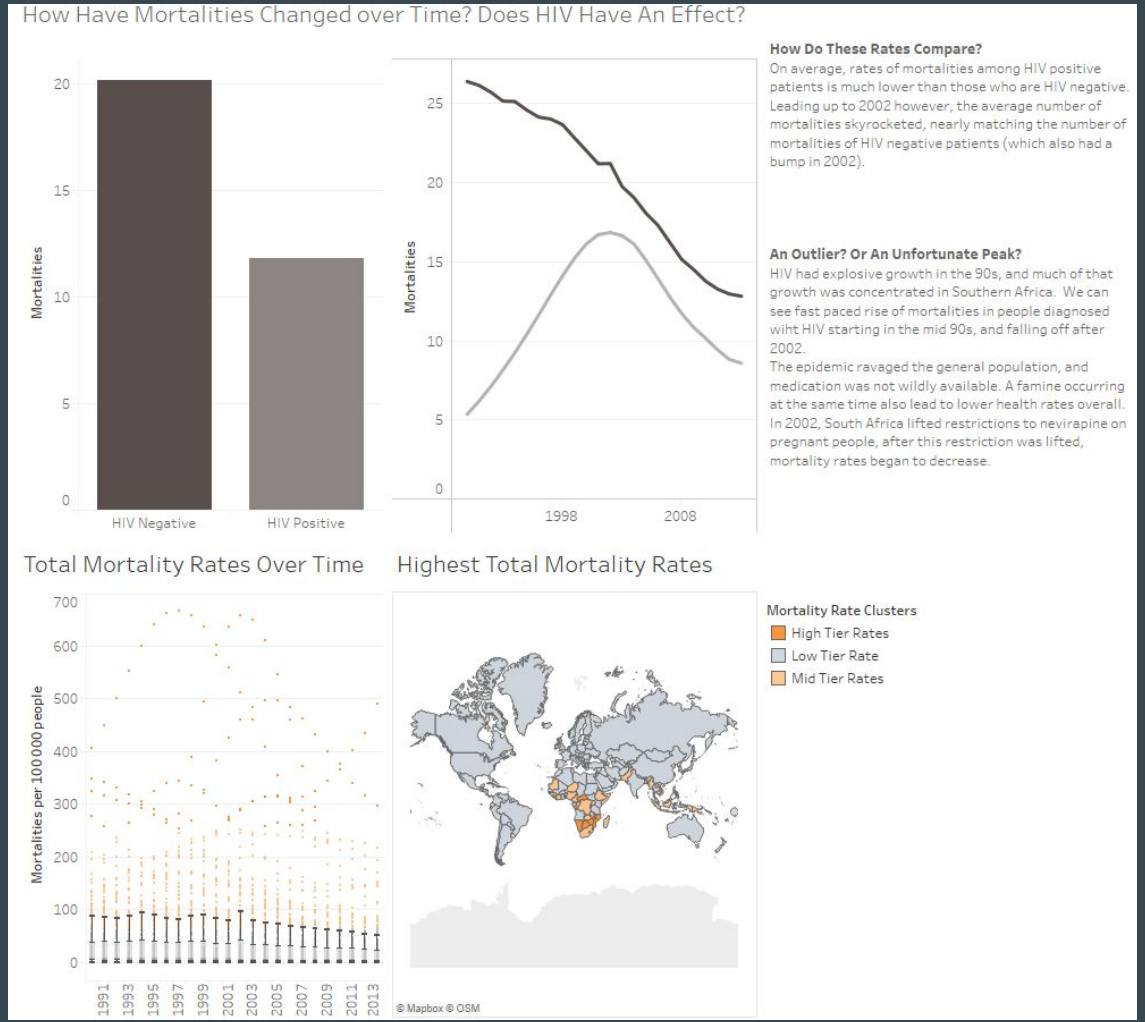
The forecast in our data seems to follow the trend shown in the data collected the World Bank. Based only on this data, we would feel safe to assume that this trend would continue to decline.

However, as per WHO's 2022 Global Tuberculosis Report, the incidence rate of TB rose by 3.6% between 2020 and 2021. WHO suggests that this is likely from fewer resources, due to Covid-19.

World Health Organization

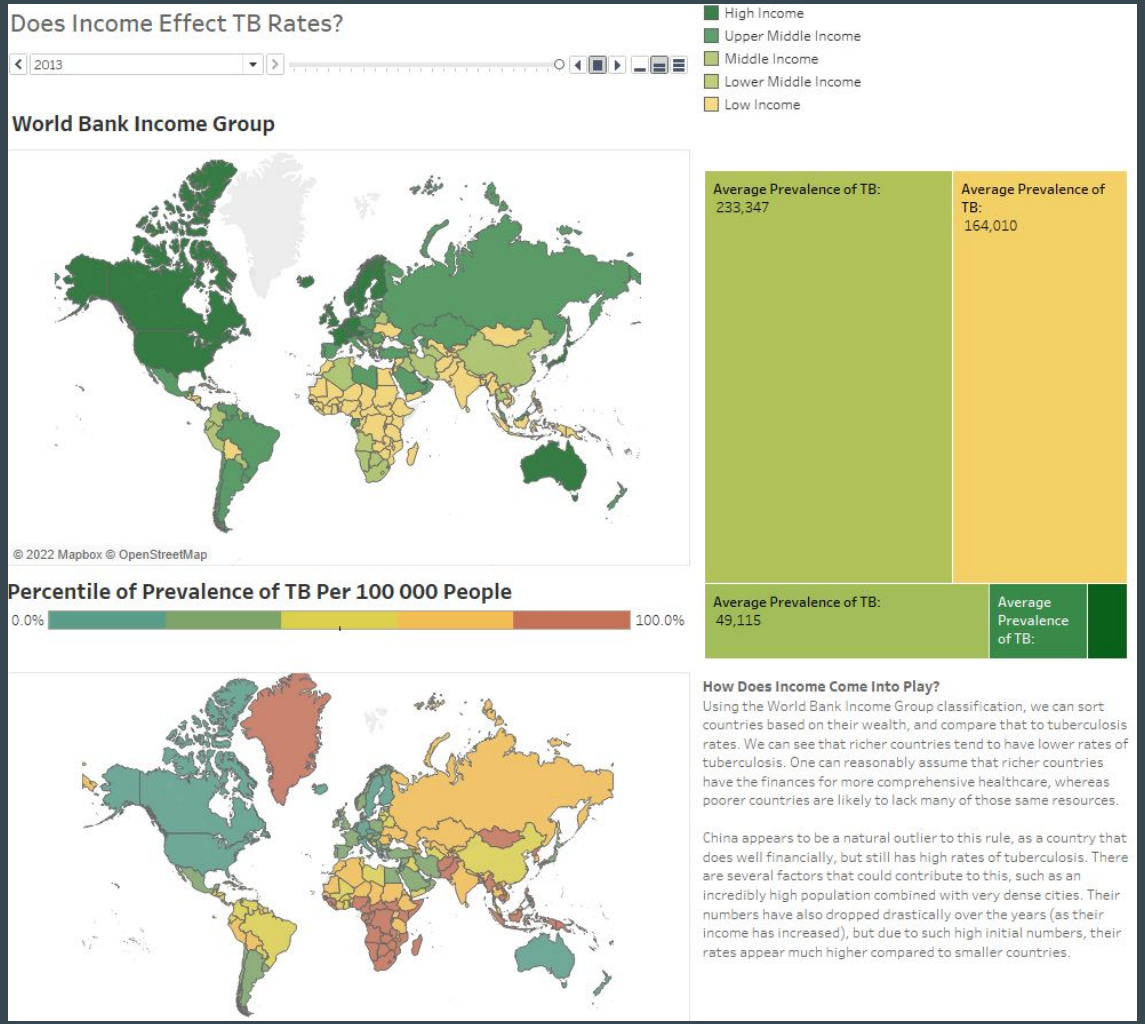
Q2 - How Has the Mortality Rate Changed?

- Chose some bar and line graphs to represent the difference between the mortality rate of HIV negative and HIV positive people
  - Specifically chose muted colours so as not to any biases
- Chose a box plot to show the outliers over time
  - Wanted to show that these numbers are still significant, even if they are 'outliers'
- Took these outliers and added their representative countries to a map, to show that these were mostly concentrated to a single region



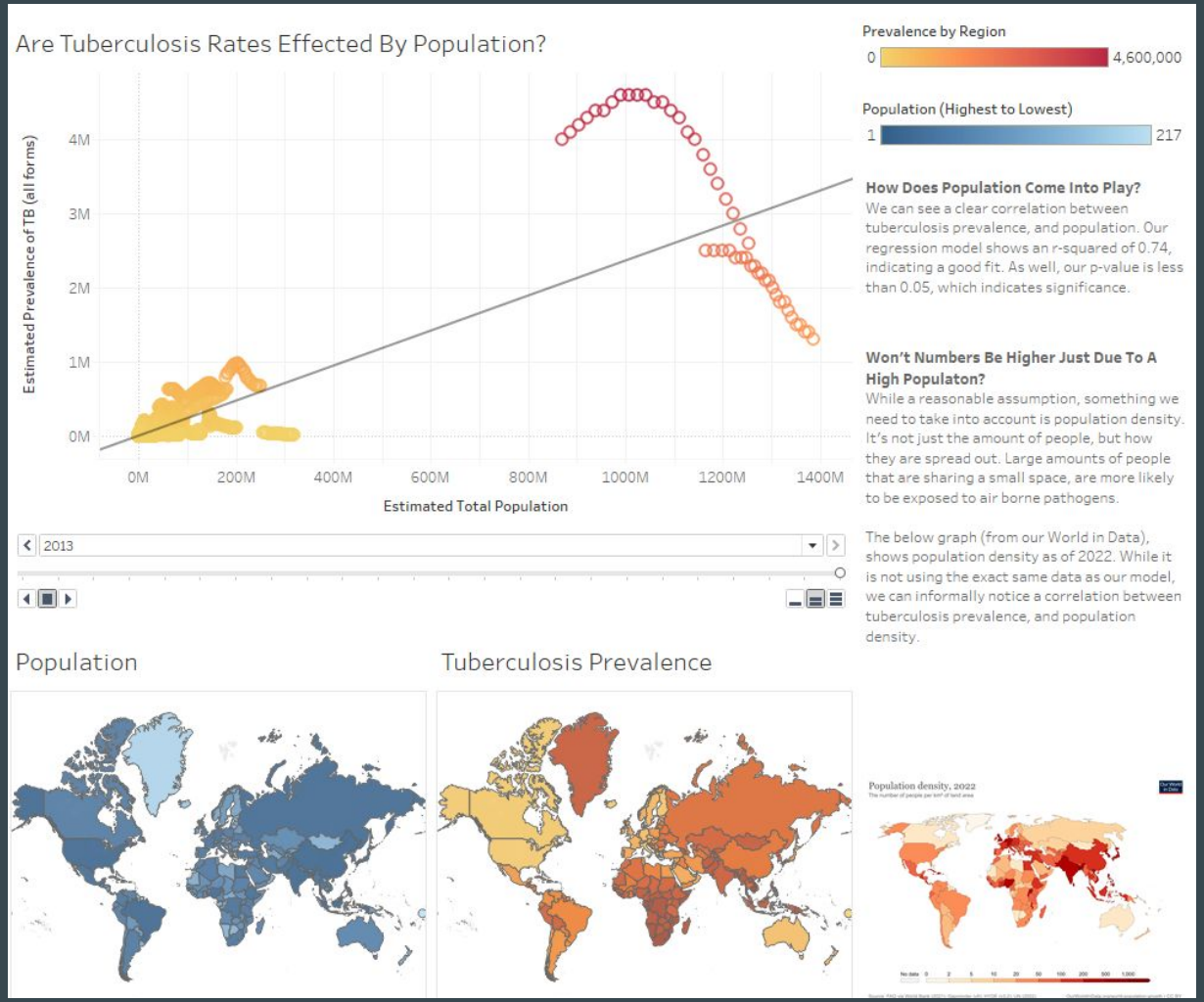
Q3 - Does Income Effect TB Rates?

- Created maps with income groups as well as TB prevalence
  - Results can be filtered by year, country, and income group
  - Having the maps side by side allow for an instant visual confirmation of any correlations in the data
- The heatmap combines these results, and highlights the stark differences between the TB rates of lower income countries and richer countries



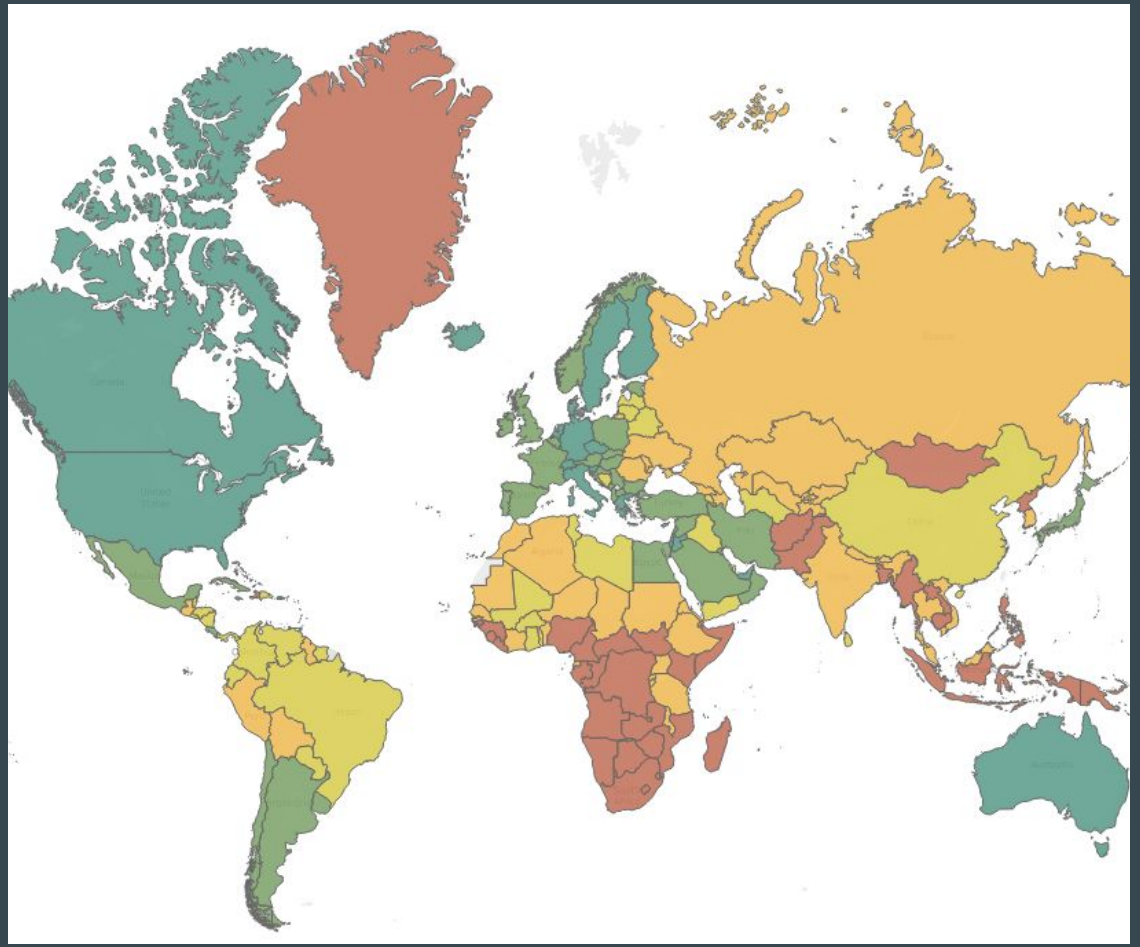
Q4 - How Does Population Come Into Play

- This line graph shows total population, and TB prevalence
- The trend line shows significant correlation
- While this may seem obvious, we have to inspect closer
- It's not just the higher numbers, it's the density
- People closer together are more likely to be exposed to airborne pathogens passed on by others
- If we compare the Tuberculosis Prevalence map vs the smaller map of Population Density (courtesy of Our World in Data), we can see how much they line up



Challenges/More Time

- It was a challenge to choose what data is ‘most relevant’ when it involves people’s lives
- I would like more time to explore this data fully, and bring in other datasets to make more meaningful comparisons and draw out conclusions on a more micro level
- This kind of data deserves to be treated with care and respect, and I would like to take time out of the context of rushing through a project, in order to give it that respect



Thank You For Your Time

•••

Data Visualization Project
LHL Cohort Oct 17/22 - Annie Stanley