# Final Project

Minchen Wang

## 1. Introduction

According to Centers for Disease Control and Prevention, "One fourth of the world's population, is infected with TB". World Health Organization (WHO) has been committed to accelerate efforts in ending TB and reach all affected people with prevention and care. The dataset I chose is the worldwide Tuberculosis Burden data from 1990 to 2013, which was provided by countries to WHO for the Global Tuberculosis Report. I'd like to study on what's the distribution of TB prevalence and how the worldwide TB prevalence changes from 1990 to 2013.
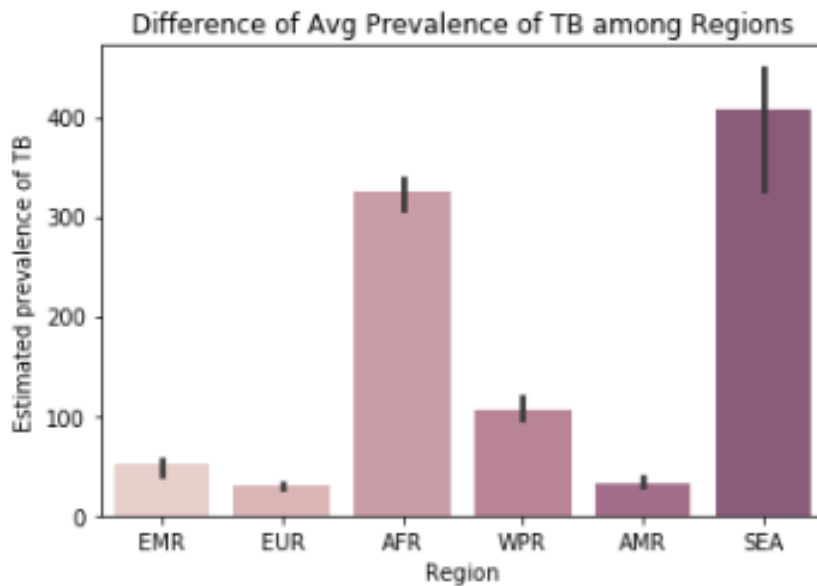
## 2. Summary of Data

- Histogram
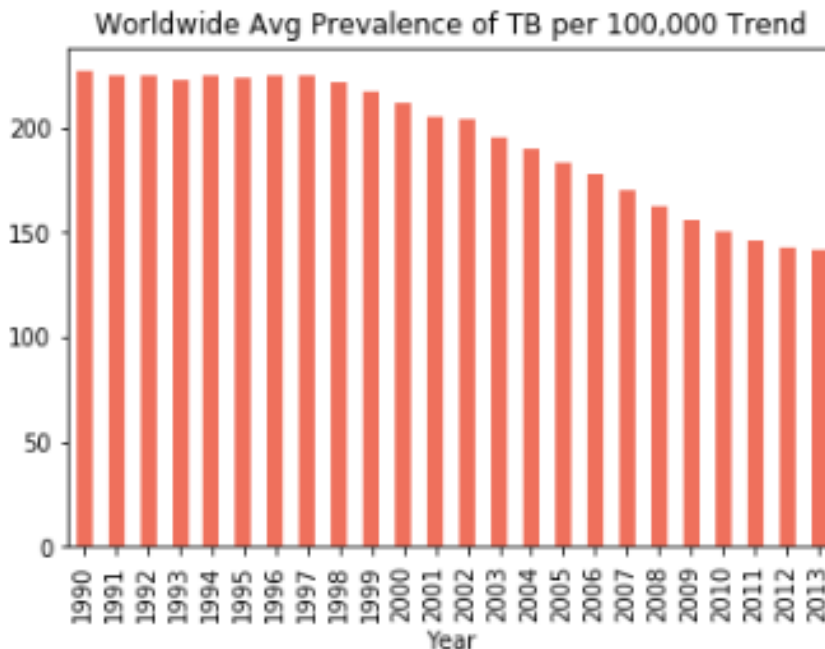


I'd like use a histogram to see the trend of average prevalence of TB per 100,000 population worldwide. The plot shows that regardless of the region, the prevalence of TB got almost one thirds decrease worldwide during 1990 to 2013.
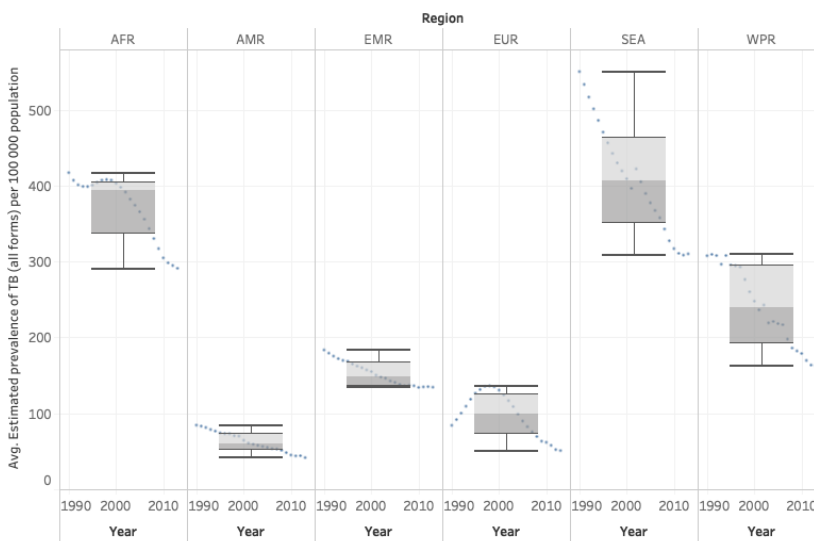
- Barplot



Difference of Avg Prevalence of TB among Regions

After looking at the time trend, I use the Barplot to have a look at the Region difference of the average prevalence of TB. It could be found that Africa and southeast Asia have much higher value than other regions in the world.
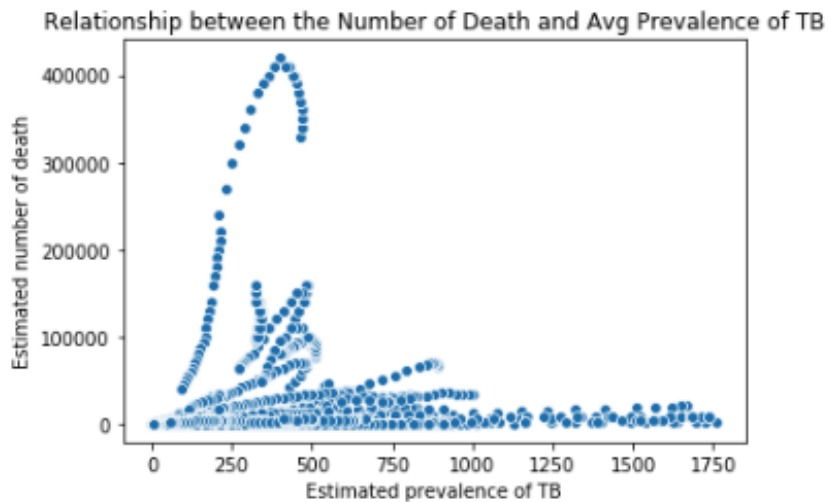
- Boxplot



The Range of avg prevalence of TB per Region

From the boxplot when can see the range of the prevalence change during the recorded years for each region. For the regions with relatively higher prevalence, it changed more than those with lower prevalence, which shows TB got effective control in the places with severe epidemics.

- Scatterplot



The scatterplot shows the relationship between the number of death and average prevalence of TB. For most points, these two values have positive relationship and the slopes are close. However, there are some points that have much higher deaths rate than others, which we could digger more.

- Bubble Map



The bubble plot shows the distribution of the average prevalence of TB per 100,000 people per year during 1990 to 2013 by country. The larger the bubble is, the severe the problem was in that country. It could be found that Africa and Southeast Asia have two clusters as for the prevalence of TB.

- Chloropleth Map



Avg Prevalence of TB per 100,000 *by Country*

Avg. Estimated prevale..
2        1,360
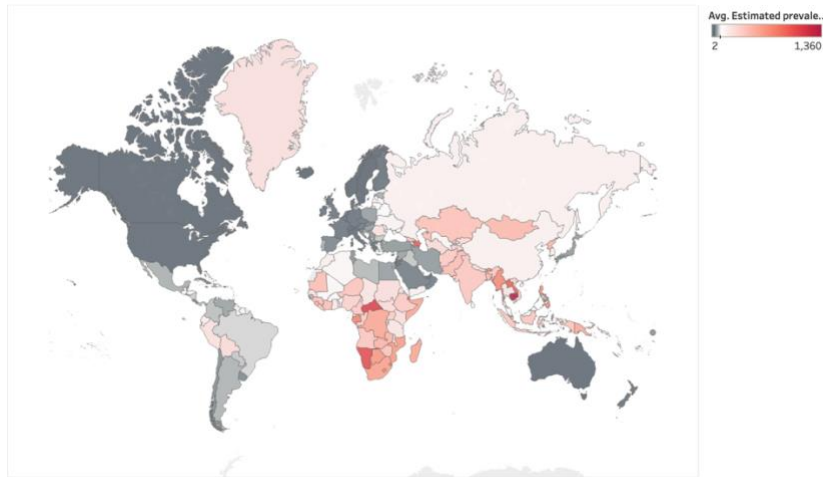
Map based on Longitude (generated) and Latitude (generated). Color shows average of Estimated prevalence of TB (all forms) per 100 000 population. Details are shown for Country or territory name. The data is filtered on Region, which keeps 6 of 6 members.

The chloropleth map discloses the same information as the bubble map, while in this map, the color of the country indicates the number of average TB prevalence per 100,000.

- Connection Map



The Change of The Center of Avg Prevalence of TB

SUM(Estimated prevale...
890        9,985
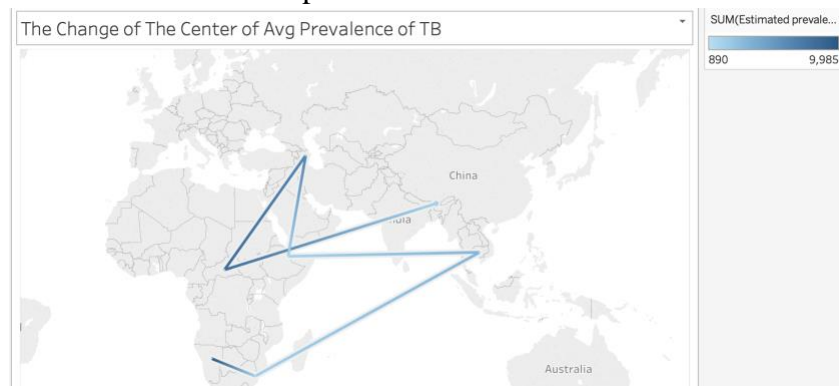
The connection map shows the change of the center of the prevalence of TB every year during 1990 to 2013.

- Heat Map
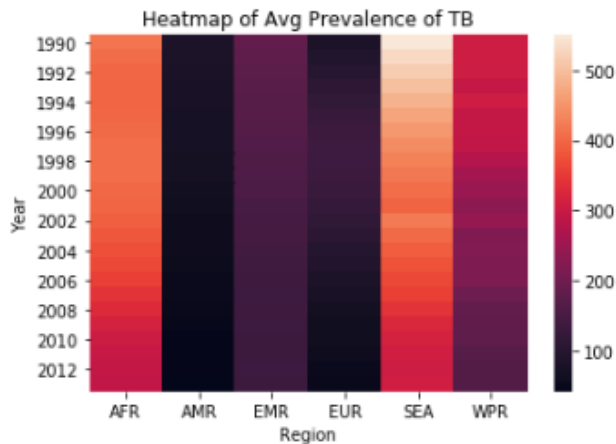
Heatmap shows the relationship between three variables: Year, Region and Average Prevalence of TB. It seems that region is a more important variable that decides of the Average Prevalence of TB, compared with time.

- Stacked Area Graph
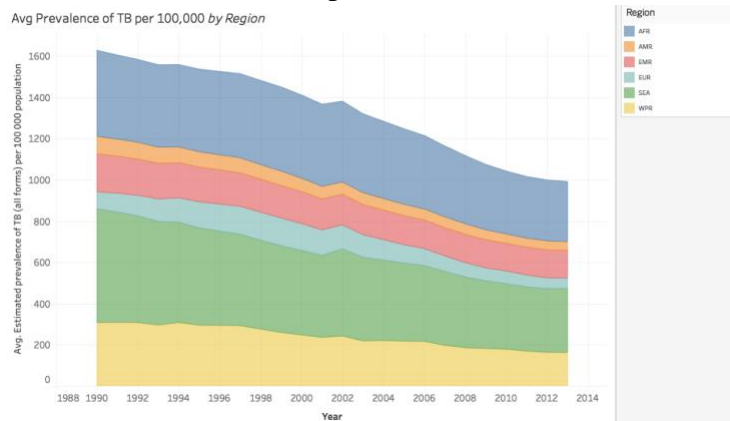


The stacked area graph displays the change of average prevalence of TB per 100,000 by region. As a whole, we can see that all the regions' TB prevalence has decreased a lot in the recorded time period, and the decrease slopes other than WPR look similar. However, there are still big gaps between different regions in terms of TB Prevalence.
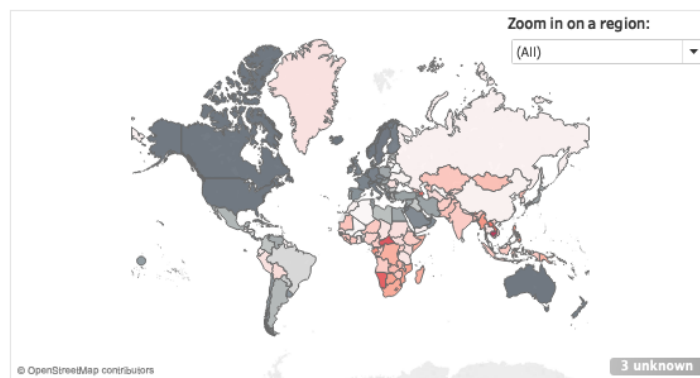
- Treemapping
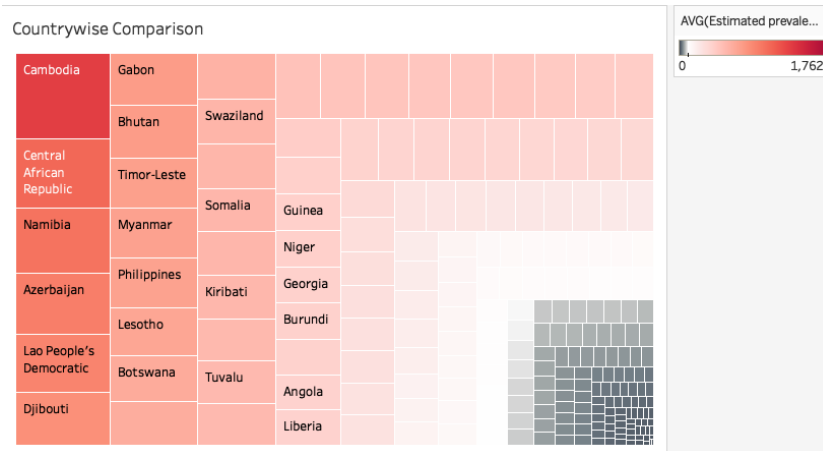
From the treemapping, we can get a base sense of how much percentage each country accounts for for the prevalence of the TB worldwide. You can easily find that the top countries are some African and Southeastern Asian countries.
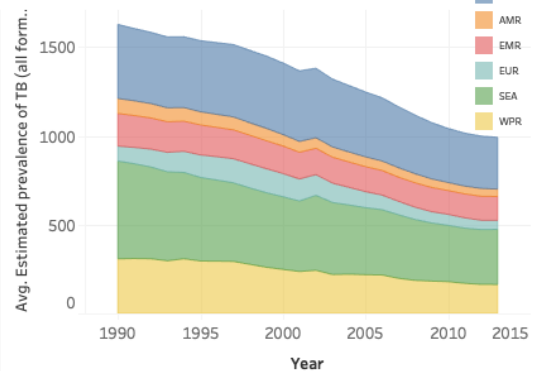
### 3. Storyline

The first plot at the upper left answers the first question. It is a map which reveals the distribution of the average prevalence of TB per 100,000 people per year during 1990 to 2013 by country. In the map, the color of the country indicates the number of average TB prevalence per 100,000. The gray means TB is very rare in these countries, which can find out that most are developed countries. While redder the color, higher the number and severer the situation. Based on the map, it is clear that Africa, South-East Asia and Western Pacific have much higher average TB prevalence compared with other places in the world.

The other two plots are aiming to answer how the worldwide TB prevalence changes from 1990 to 2013.The one at upper right is an stacked area chart which displays the change of average prevalence of TB per 100,000 by region (AFR=Africa; AMR=Americas; EMR=Eastern Mediterranean; EUR=Europe; SEA=South-East Asia; WPR=Western Pacific). As a whole, we can see that all the regions' TB prevalence has decreased a lot in the recorded time period, and the decrease slopes other than WPR look similar. However, there are still big gaps between different regions in terms of TB Prevalence. The other one at bottom is a line chart which shows annual TB prevalence per 100,000 population for each country from 1990 to 2013. In the plot, the redder the color is, the severe TB prevalence is in that country. We could find that the value of TB prevalence of each country has decreased to certain level ( less than 1000 per 100,000 population) before 2010.

## 4. Conclusion

All in all, we can conclude that from 1990 to 2013, TB prevalence has a great decrease worldwide, while there still are exists geographical difference.

## 5. Link to Github

https://github.com/minchenw/Data_Viz