OBJECTIVE OF STUDY:

Presented by: Sahay, Anupam

CIS:5270

Business Intelligence

Analytical Study on Tuberculosis by Country (Who) using Tableau

SUbmited to: DR. Shilpa Balan

March 30, 2018

TB is an infectious disease caused by the bacillus Mycobacterium tuberculosis. It typically affects the lungs but can also affect other sites. The disease is spread when people who are sick with pulmonary TB expel bacteria into the air, for example by coughing. Overall, a relatively small proportion (5–15%) of the estimated 1.7 billion people infected with M. tuberculosis will develop TB disease during their lifetime. However, the probability of developing TB disease is much higher among people infected with HIV, and also higher among people affected by risk factors such as under-nutrition, diabetes, smoking and alcohol consumption. Tuberculosis is the ninth leading cause of death worldwide and leading cause from a single infectious agent above HIV/AIDS. Tuberculosis death rates are among HIV – negative people are falling every year. Globally, the TB mortality rate is falling every year. TB incidence is falling at about 2% per year; this needs to improve to 4–5% per year by 2020 to reach the first milestones of the End TB Strategy. By doing this project I would be able to analyze the tuberculosis burden in every country. The TB death rates among HIV positive and negative people and the necessary action taken by the WHO in reducing of the TB deaths. The analysis can also be done on the geographical locations of the world and how has it changed over the time.

ABOUT TUBERCULOSIS DATASET:

One -third of the world’s population is thought to be infected with TB. Ending TB is not just public health problem, but development challenge and opportunity. In order to end TB epidemic, countries will need to strengthen their health and social sector. To eradicate TB, we can focus on particular areas. Tuberculosis dataset includes number of deaths due to tuberculosis (HIV+ / HIV-), morality rate, incidences caused by TB, year, locations, estimated world population.

**A.) DATASETS LINKS :** [**http://apps.who.int/gho/data/node.main.1315?lang=en**](http://apps.who.int/gho/data/node.main.1315?lang=en)

INSIGHTS:

Insights 1: Estimated incidence of TB cases who are HIV + (high bound and lower bound) every

year?

Insight 2: Estimated number of deaths from TB (all forms).

Insights 3: Estimated prevalence of TB (high bound and lower bound) in all forms

Insights 4: Estimated mortality of TB cases (all forms)

Insights 5: Most TB affected regions in the world.

Insights 6: Case detected rate of TB every year

**B.) DATA CLEANING:**

In order to get correct results for analysis it is very crucial that data should be clean enough otherwise results can be manipulated. In Tuberculosis dataset following cleaning is been done.

1.) Repeated Values

There were repeated values for Bulgaria

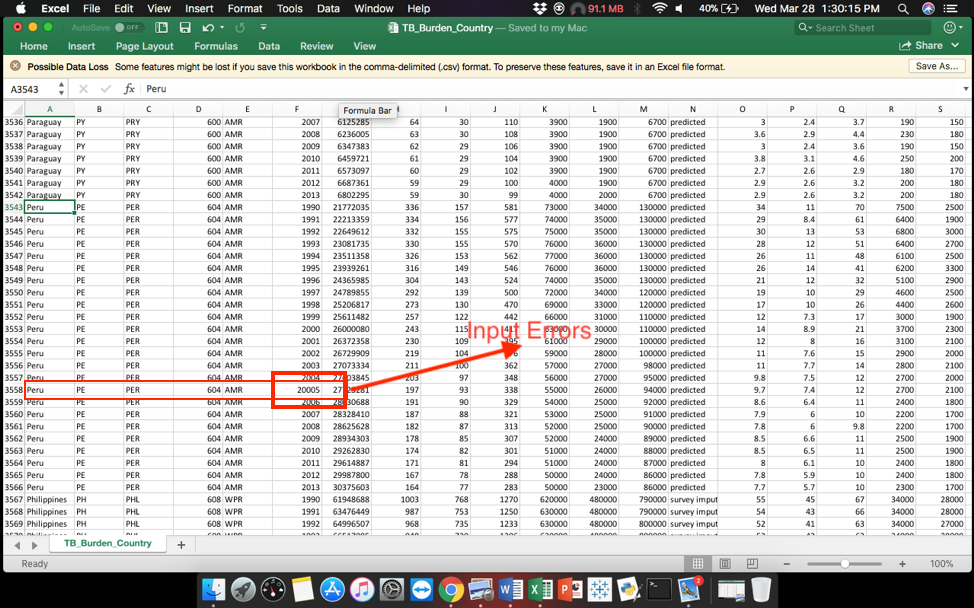


Repeated values were cleaned.

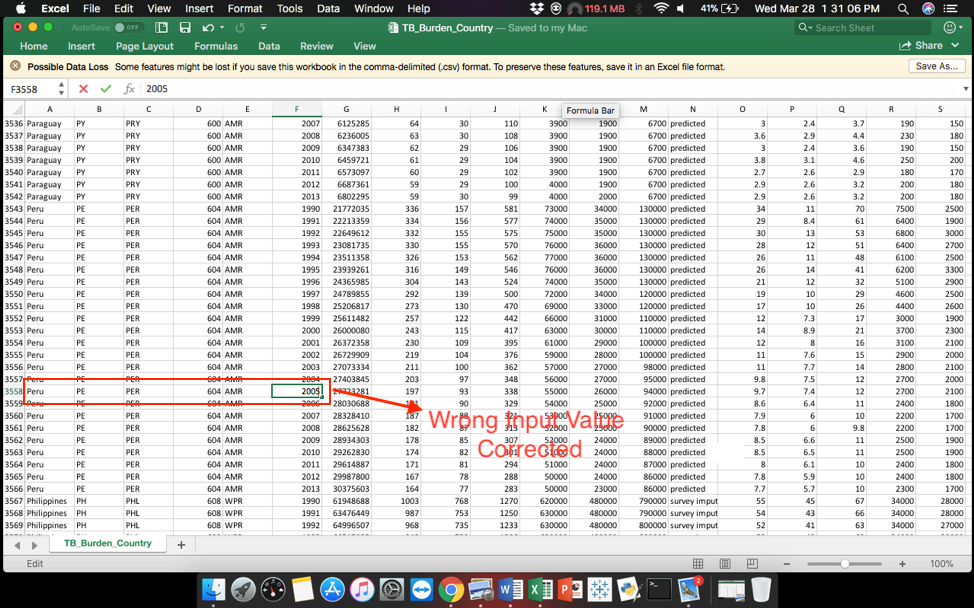


2.) Invalid Values

The year columns have few input errors.

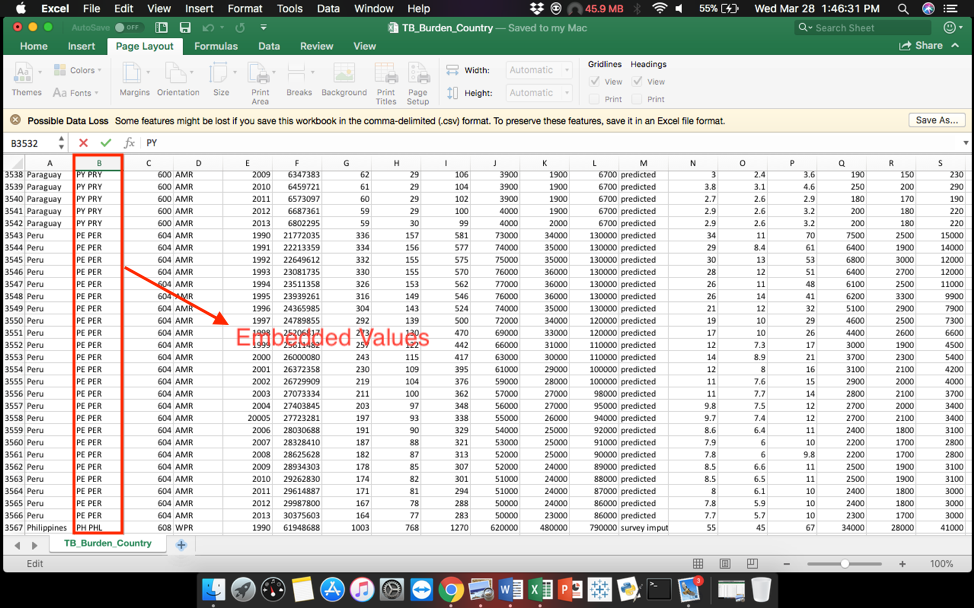


The format of year was cleaned



3.) Embedded Values

Two types of character country/territory codes were available in the same column

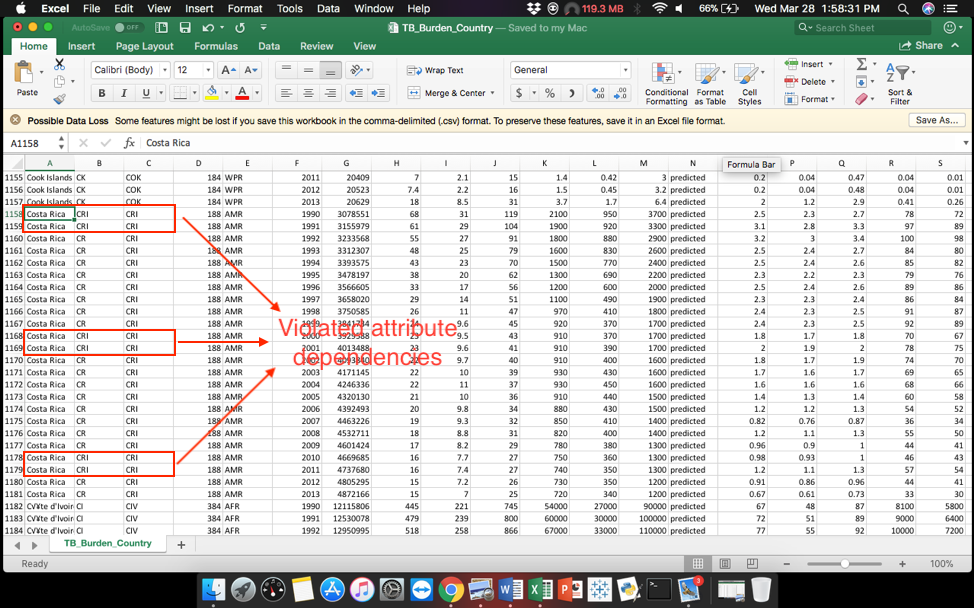


Now there are two different columns containing the values for country/territory code

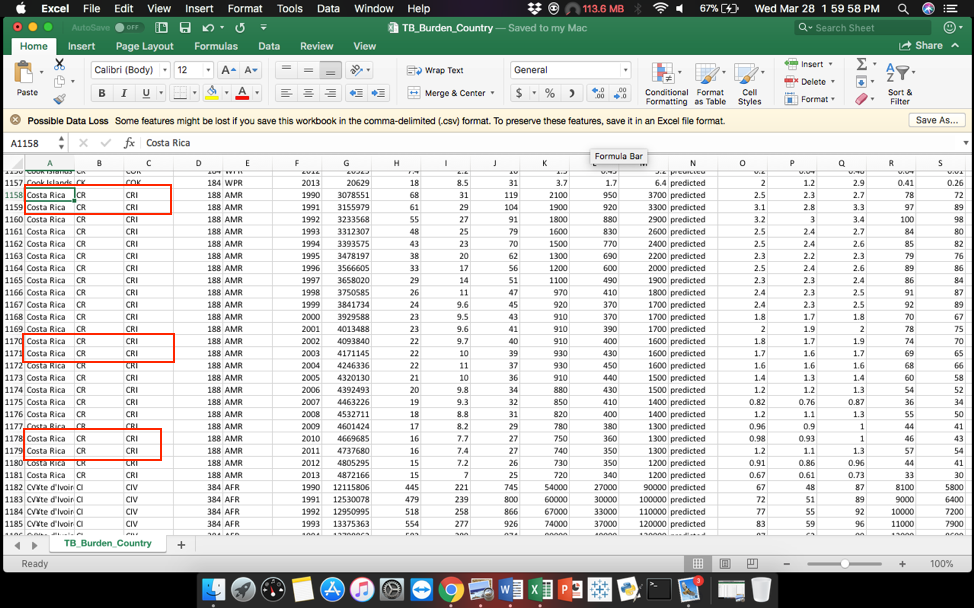


4.) Violated Attributes dependencies

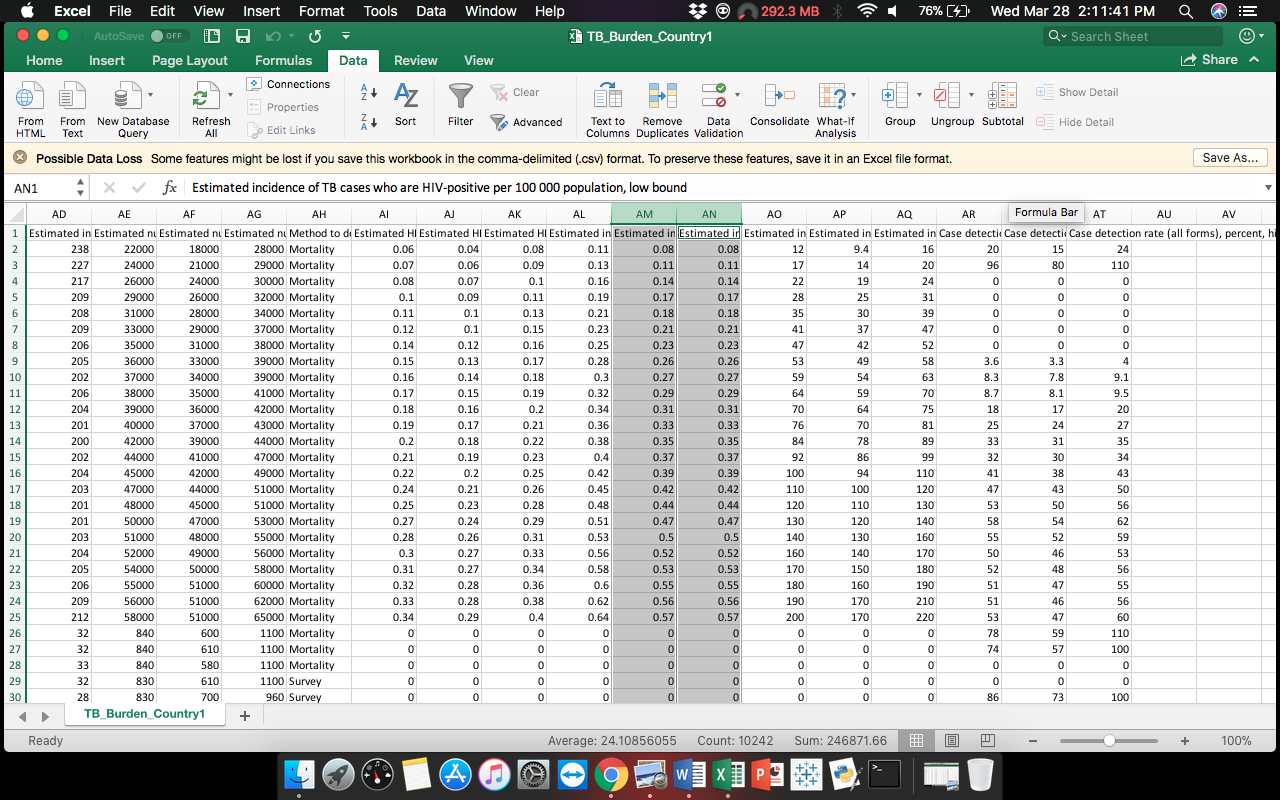
The type 1 and type 2 of the country/territory codes for are few entries were same



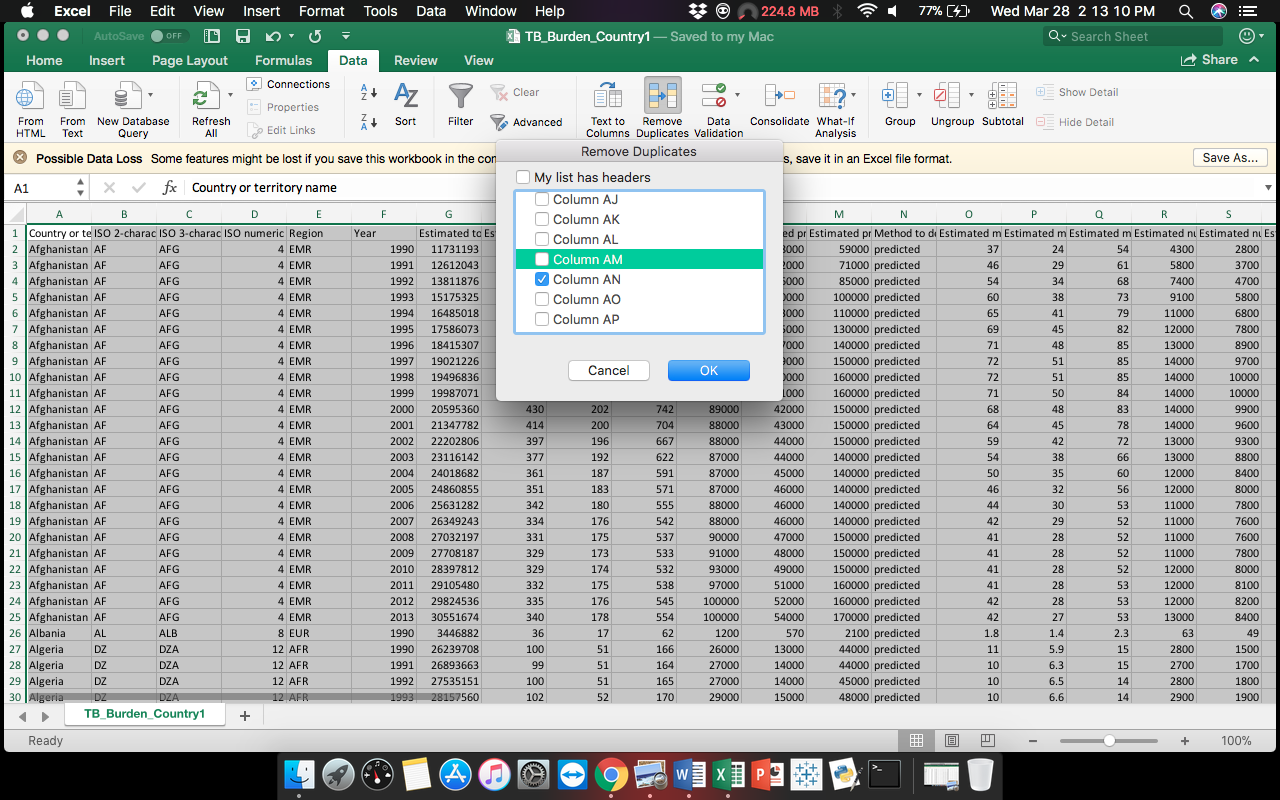
The values were cleaned



5.) Duplicate Records



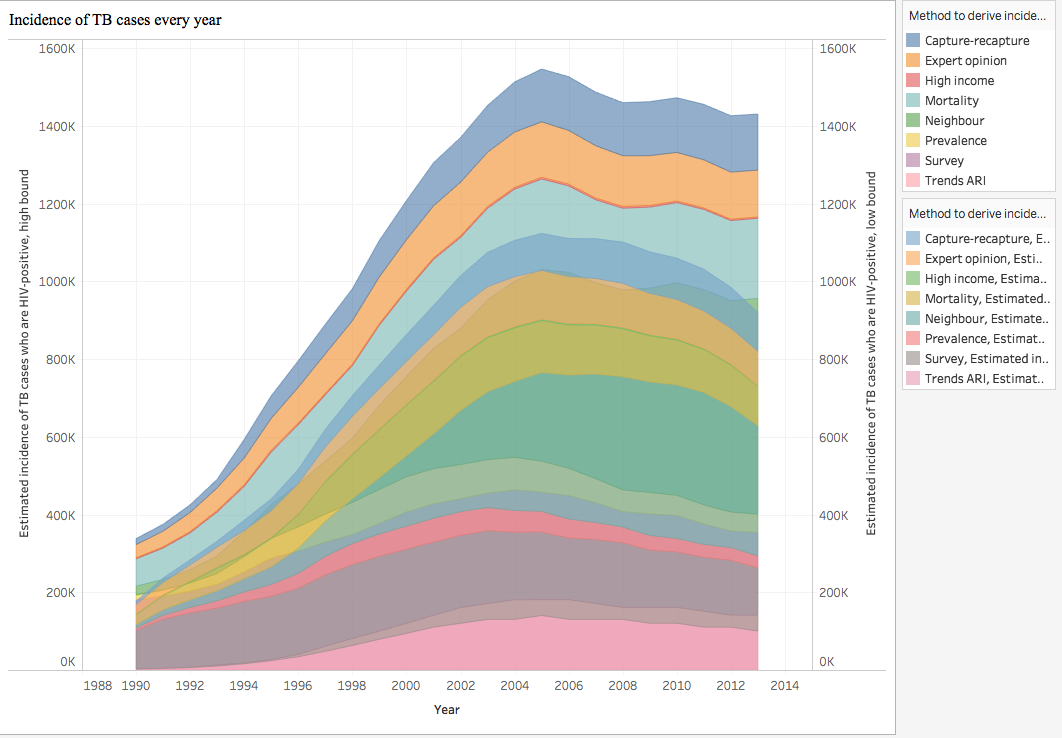
Column AM and AN were the same records of “Estimated incidence of TB cases who are HIV positive per 100000 population, low bound”, column AN was removed





**C.) DATA VISUALIZATIONS:**

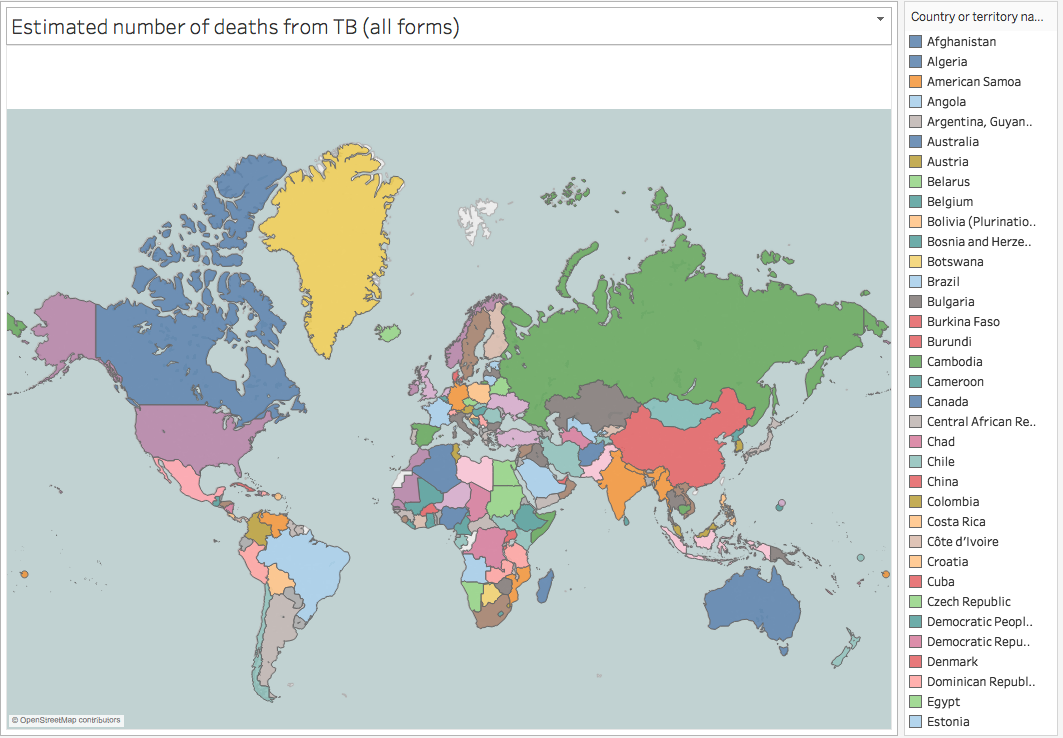
**Insight 1:** Incidence of TB cases (HIV +) every year.



[Used: Area Charts]

The area charts show the estimated incidences of TB cases who are HIV positive (high bound and low bound). It also tells us the methods to derive the TB incidence estimates every year. In the analysis it is clear that estimated incidence of TB cases who are HIV positive, high bound in year 2005 is at its peak and the method to derive the estimated incidences is capture – recapture method. These analyses help the WHO’s authorities to take careful take precautions to reduce the TB incidences in the subsequent years which will help in them to eradicate TB.

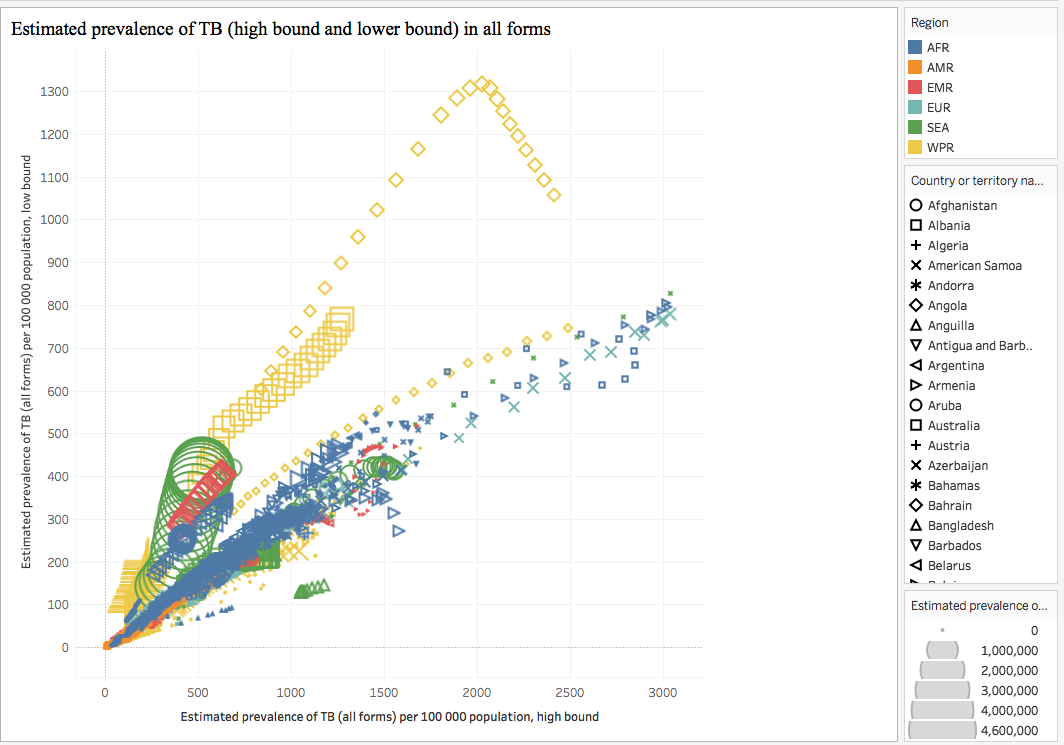
**Insight 2:** Estimated number of deaths from TB (all forms).



[Used: Geographical Maps]

The geographical map shows us all the countries in the world and tells us the estimated number of deaths from TB (all forms). It also shows us the estimated population number of each country. The countries are grouped according to color. The orange color countries show the highest estimated death from TB. According to analysis India has recorded the highest number of TB deaths (taking into account the estimated population of the country). These records can be helpful to common people and to the health authorities of the country to reduce the TB death rates.

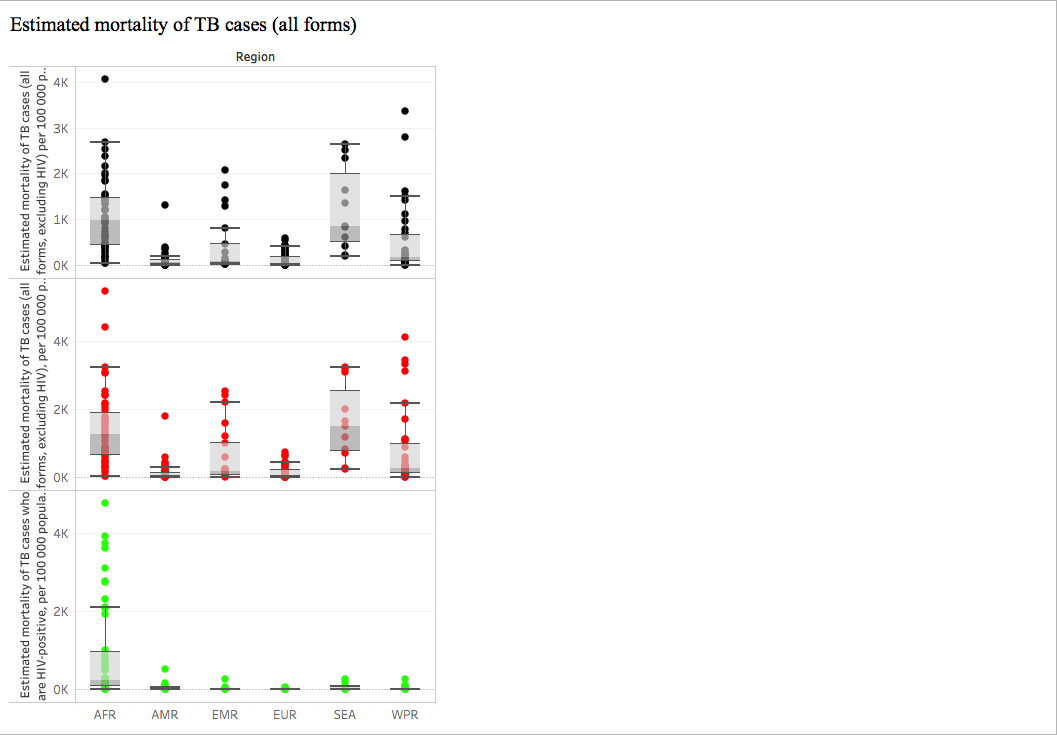
**Insights 3:** Estimated prevalence of TB (high bound and lower bound) in all forms



[Used: Scatter Plots]

In this chart we analyze the prevalence of TB (in HIV+/HIV-) in each region / country. It also describes the method by which the prevalence of TB is derived in each country. After the analysis we know that Cambodia was the highest prevalence of TB in year 1999, and the method to drive the prevalence estimates was survey imputed. This report of the prevalence of TB (all forms) can help the government to take necessary actions to reduce the TB prevalence in their country and help the WHO to eradicate TB.

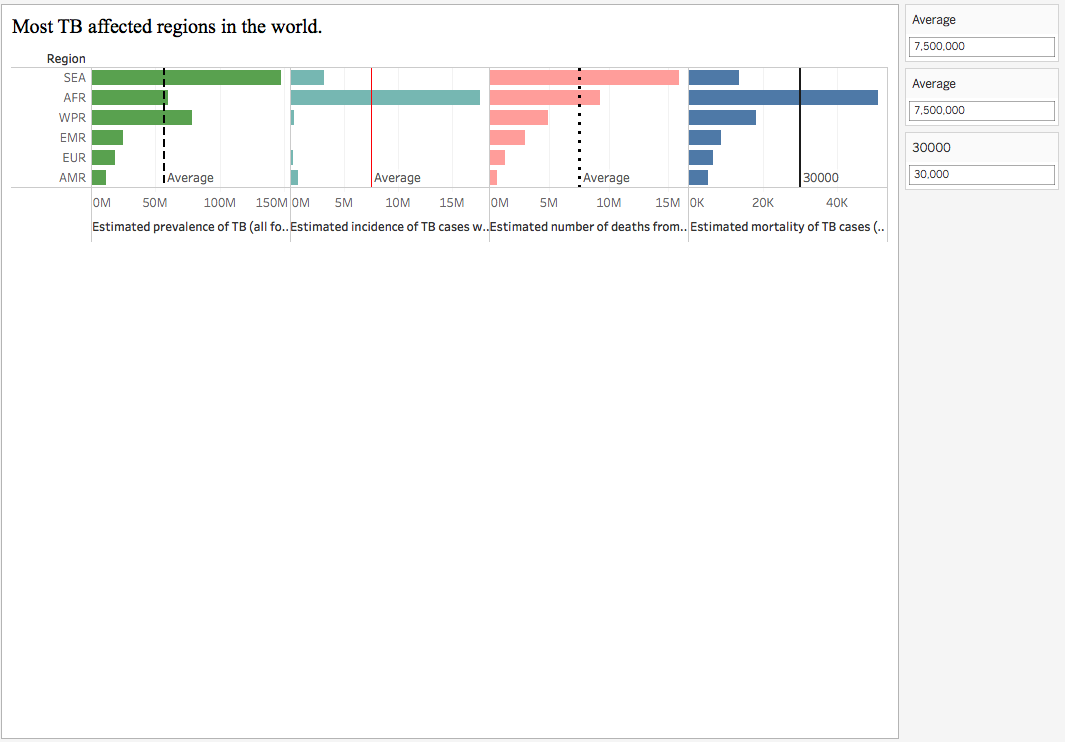
**Insights 4:** Estimated mortality of TB cases (all forms)

****

[Used: Box-and-Whisker Plots]

The box and whisker plot are used to analyze the mortality of TB cases in each region. The Central African Republican have the highest in the whole analysis. This also brings to our notice that each region has its own outliers. The upper whisker, lower whisker and the median gives an indication that the morality rate of TB is decreasing which will eradicated TB if all the health authorities of every county take suitable steps to prevent TB given by World Health Organization.

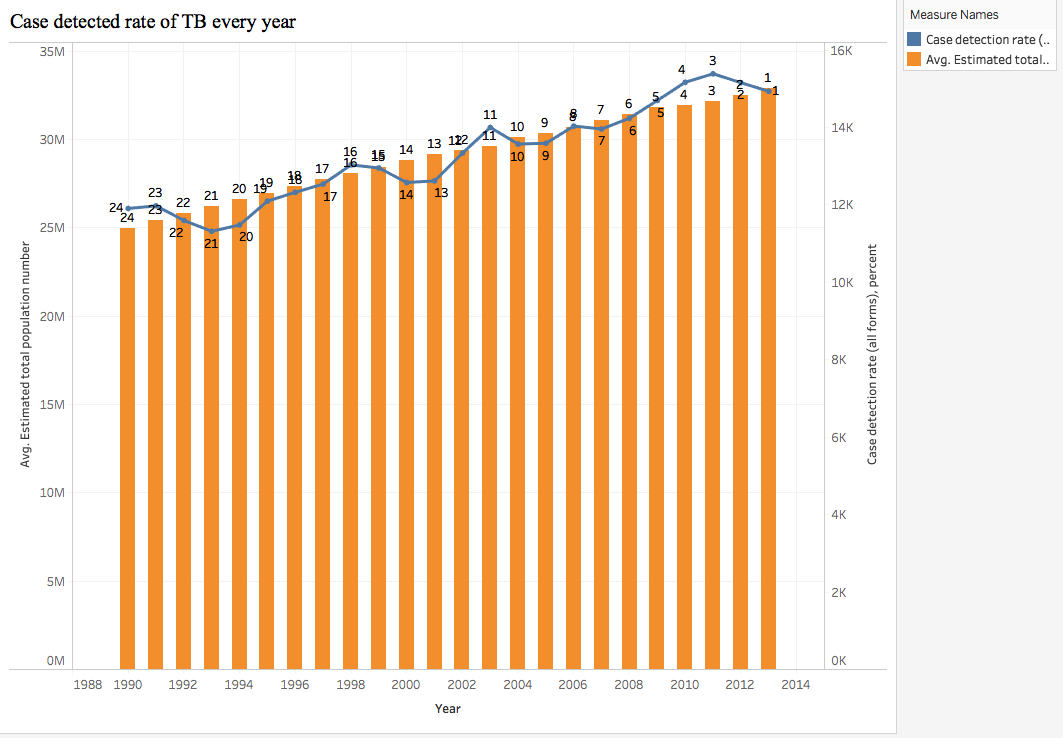
**Insights 5:**  Most TB affected regions in the world.

****

[Used: Reference Lines]

This insight gives us the highlight for estimated prevalence of TB, estimated incidences of TB cases who are HIV positive, estimated number of deaths from TB (all forms), estimated morality of TB cases vs regions. Reference line give the average in each different column. This indicates that when we take proper care of the people affected with TB we can reduce it. The proper detection of TB on time is very important to eradicate TB from the world, for which we need to educate the people about TB in most affected regions of the world.

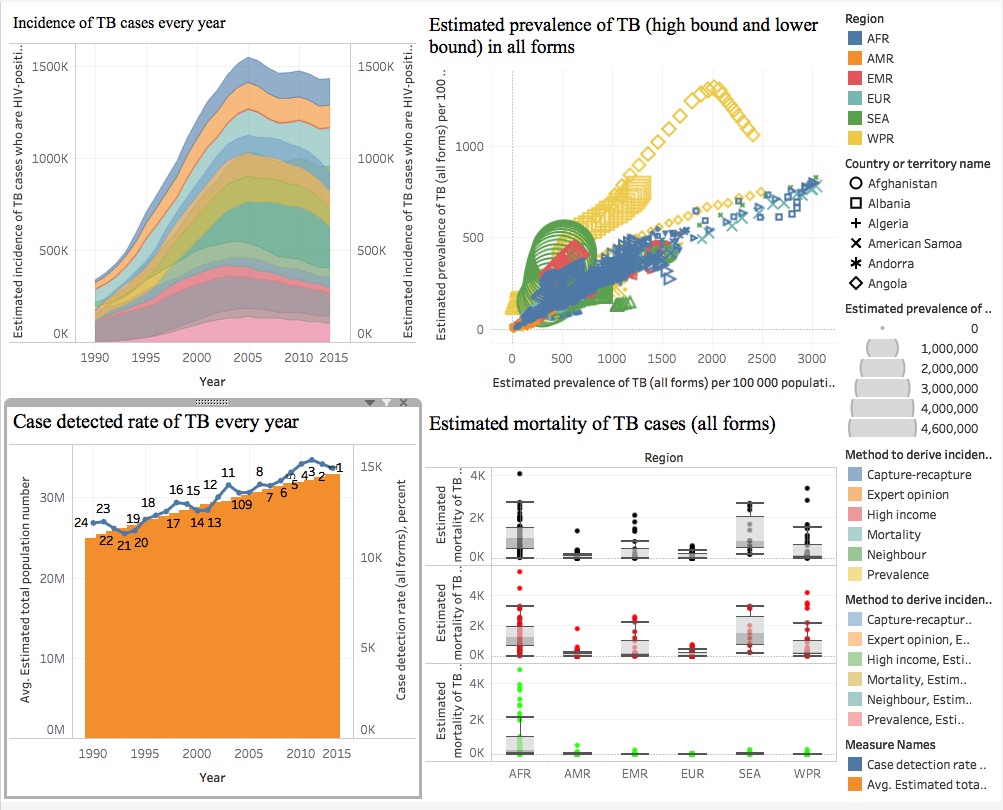
**Insights 6:** Case detected rate of TB every year



[Used: Rank, Filters]

This analysis provides the case deduction rate of tuberculosis and the average estimated population number of the countries / regions every year. We rank the case detection rate (all forms) and we can see that it is increasing every year which is good indication to reduce the TB.

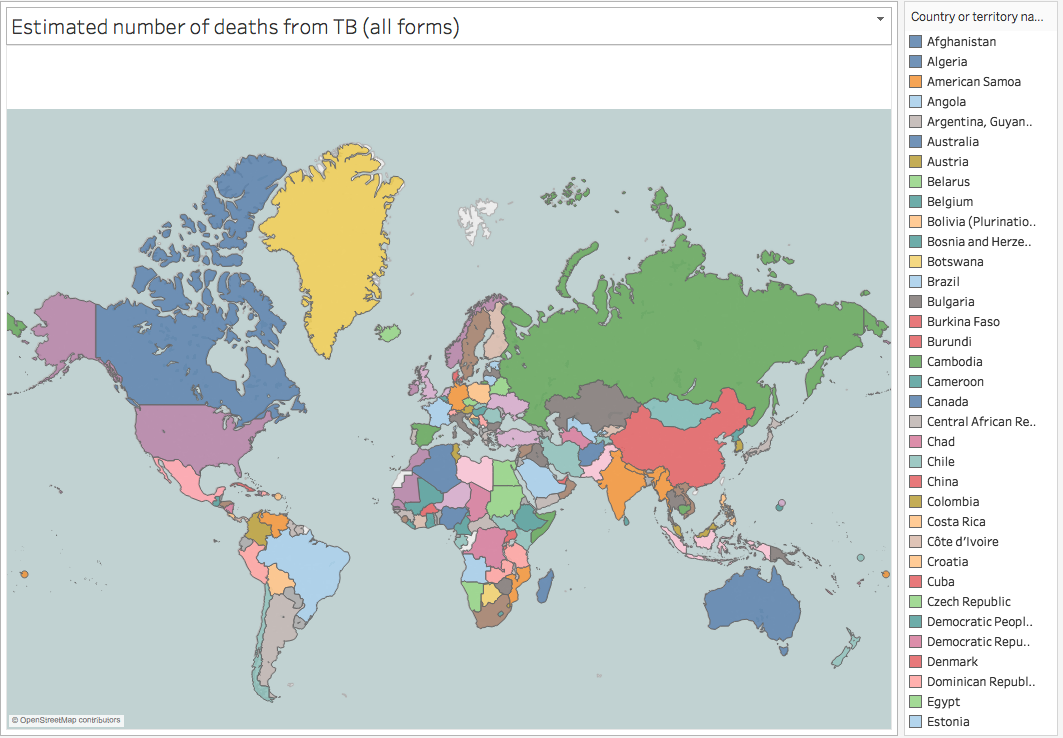
**D.) Dashboards**



**E.) STORY TELLING:**

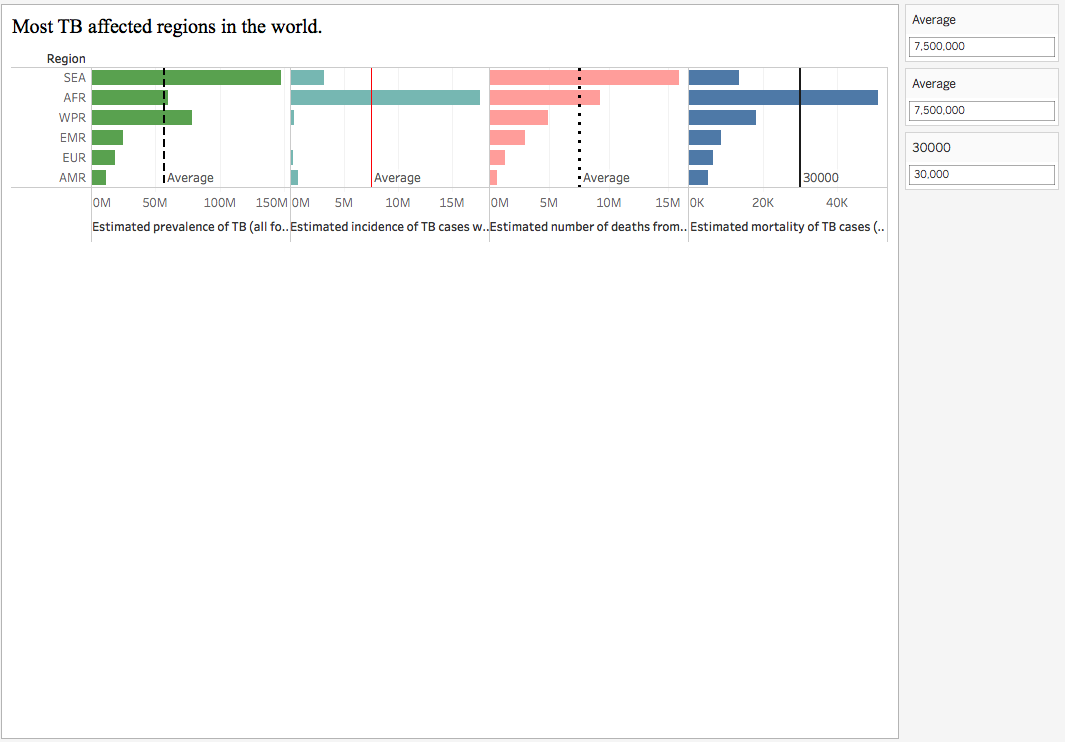
Tuberculosis (TB) is a potentially serious infectious disease that mainly affects your lungs. The bacteria that cause tuberculosis are spread from one person to another through tiny droplets released into the air via coughs and sneezes. Once rare in developed countries, tuberculosis infections began increasing in 1985, partly because of the emergence of HIV, the virus that causes AIDS. HIV weakens a person's immune system so it can't fight the TB germs. In the United States, because of stronger control programs, tuberculosis began to decrease again in 1993, but remains a concern. Many strains of tuberculosis resist the drugs most used to treat the disease. People with active tuberculosis must take several types of medications for many months to eradicate the infection and prevent development of antibiotic resistance. Tuberculosis can also affect other parts of your body, including your kidneys, spine or brain. When TB occurs outside your lungs, signs and symptoms vary according to the organs involved. Since the 1980s, the number of cases of tuberculosis has increased dramatically because of the spread of HIV, the virus that causes AIDS. Infection with HIV suppresses the immune system, making it difficult for the body to control TB bacteria. As a result, people with HIV are many times more likely to get TB and to progress from latent to active disease than are people who aren't HIV positive. From 2000 to 2015, global and national efforts to reduce the burden of tuberculosis (TB) disease were focused on achieving targets set within the context of the Millennium Development Goals (MDGs).

Estimated: 1.3 million(range: 1.16–1.44 million) HIV-negative people died from TB and 0.37million (range: 0.33–0.43 million) HIV-positive people died from TB. Global treatment success has increased slightly in recent years. A total of 138 countries and territories reported treatment outcomes for people started on treatment in 2014. In contrast, treatment failure was highest in the WHO European Region (13%), and the death rate was highest in the WHO African and South-East Asia regions (20%).



Treatment success was less than 50% in China, India, Peru, the Philippines and Ukraine, due to high death rates in India (21%) and Ukraine (17%), high rates of treatment failure in Ukraine (18%) and loss to follow-up or missing data in all five countries (19–60%) (no treatment outcome data were reported by Uzbekistan).

Funding for and implementation of high-quality TB-specific interventions should result in detection of people with TB and curative treatment; in turn, this should have a direct impact on TB mortality. Shortening the duration of disease through detection and treatment of cases will also reduce the prevalence of TB disease, and therefore, transmission. There will be an impact on TB incidence if transmission can be reduced sufficiently and if preventive treatment of people with latent TB infection is effectively implemented on a large scale. At the same time, a range of factors besides TB-specific interventions influence levels of TB disease burden, by affecting population susceptibility to both TB infection and the risk of developing TB disease once infected. These include overall levels of wealth and the distribution of wealth (measured e.g. as GNI per capita, the proportion of people living in poverty), the overall coverage and quality of health services and the prevalence of HIV and other risk factors for TB.



If we need to eradicate the tuberculosis we have to ensure that health-care systems are capable of delivering the right care. And the right care requires rapid development and dissemination of new tools, including quick point-of-care diagnostic tests, safe and fast-acting drugs, and an effective TB vaccine.

**References:**

1. Global tuberculosis report 2017

<http://www.worldlifeexpectancy.com/cause-of-death/tuberculosis/by-country/>

2. Mayo Clinic

<https://www.mayoclinic.org/diseases-conditions/tuberculosis/symptoms-causes/syc-20351250>

3. National Center of Biotechnological Information

<https://www.ncbi.nlm.nih.gov/books/NBK258651/>

4. <https://www.weforum.org/agenda/2015/09/how-can-we-eradicate-tuberculosis/>