BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION

ALY6070- Third Quarter, Term A, Dr. Venkata Duvvuri

Module 4, Week Six

Bringing it All Together: Storytelling and Application

by

Group 1 – Anubhav Rastogi, Manvinder Kaur, and Anastasiya Krymskaya

Assignment Completion Date: 05/18/19

Silicon Valley, Northeastern University, CA.

BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION

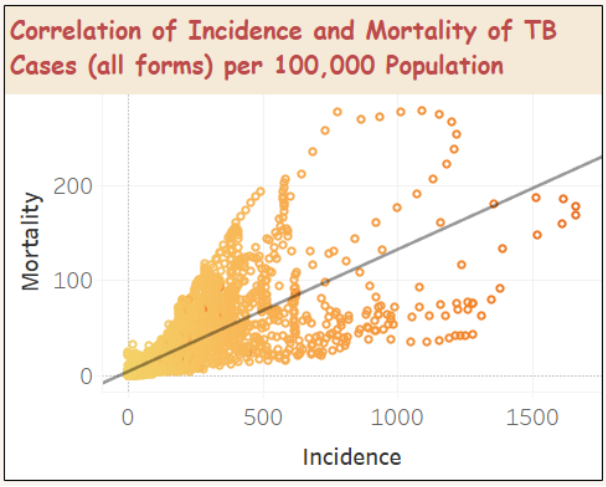
INTRODUCTION

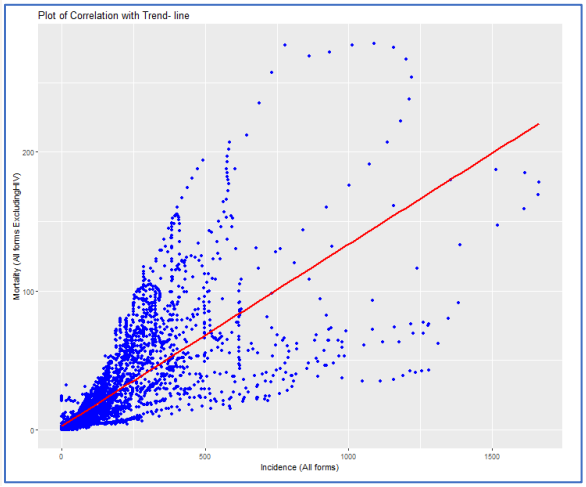
The report sights our findings on the Tuberculosis Burden dataset. Tuberculosis (TB) is a disease caused by bacteria called Mycobacterium tuberculosis. The bacteria usually attack the lungs, but they can also damage other parts of the body. TB spreads through the air when a person with TB of the lungs or throat coughs, sneezes, or talks.

The Dataset we used includes 5120 rows and 47 columns. The variables we used for analysis were country, population, year, estimated mortality, death, prevalence and incidence. We used both R-studio and Tableau to generate graphs and reactive applications. Further, we have used certain R libraries such as shiny, readxl and ggplot to prepare graphical representation of our analysis and findings.

QUESTIONS ANSWERED

1. Is there a correlation between Incidence and Mortality of TB cases (all forms) per 100,000 population?



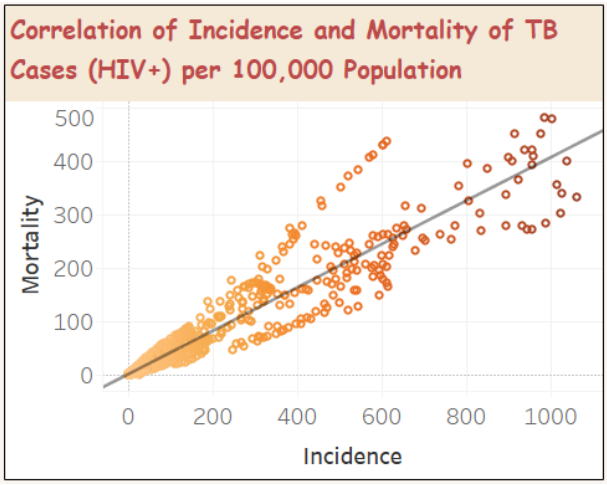


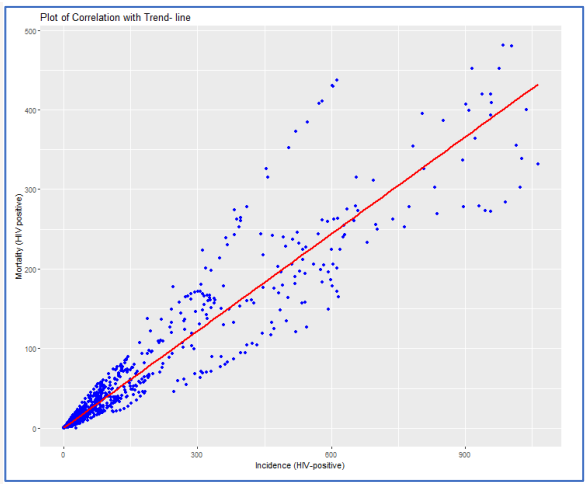
Yes, there is a positive correlation between these two variables. Higher the incidence, higher is the mortality. The trendline is also showing the positive correlation in both the visualizations

using Tableau and R-shiny. Below are few calculated fields P-value, Intercept, Standard error and R-Squared (0.55509) which further supports our conclusion.

Term	Value	StdErr	t-value	p-value
Estimated incidence (all forms) per 100 000 population	0.128825	0.0017749	72.5833	< 0.0001
intercept	3.28506	0.448903	7.31796	< 0.0001

2. Is there a correlation between Incidence and Mortality of TB cases (HIV+) per 100,000 population?



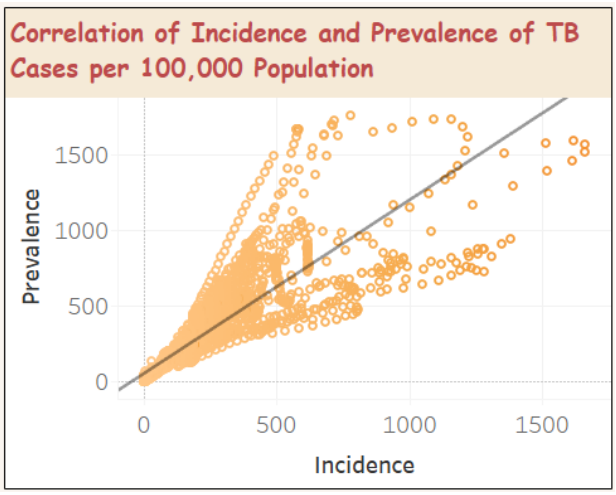


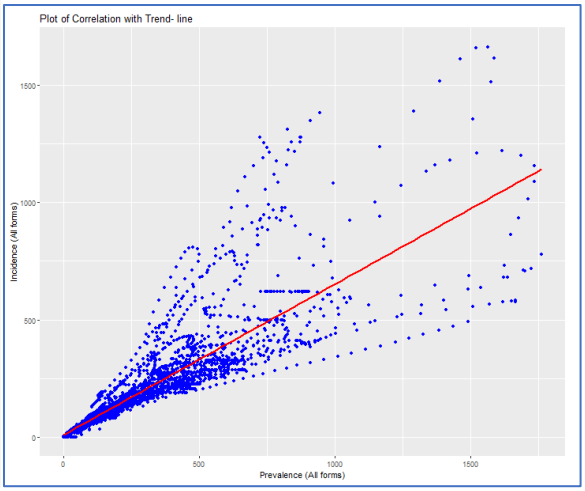
Yes, there is a strong positive correlation between these two variables. It means that HIV + patients have much less chances to recover from TB. The trendline is showing the very positive correlation as the R-Squared is 0.91 which is close to 1 and the standard error is also very low in both the visualizations using Tableau and R-shiny. Below are few calculated fields P-value, Intercept, Standard error and R-Squared (0.91330) which further supports our conclusion.

Term	Value	StdErr	t-value	p-value
Estimated incidence of TB cases who are HIV-positive per 100 000 population	0.405872	0.0025678	158.062	< 0.0001
intercept	0.442306	0.418086	1.05793	0.290194

3. Is there a correlation between Incidence and Prevalence of TB cases (all forms) per 100,000

population?



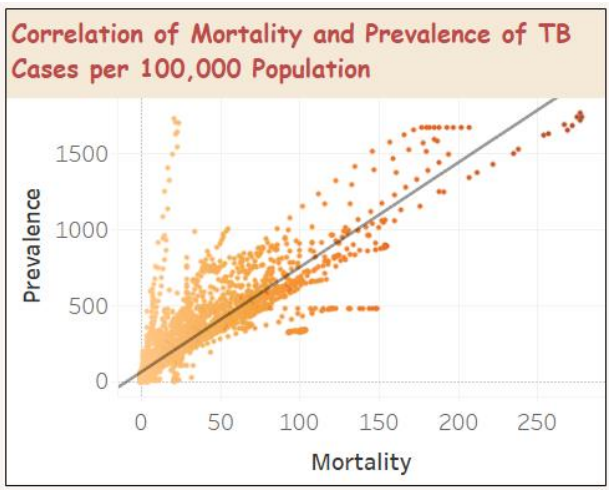


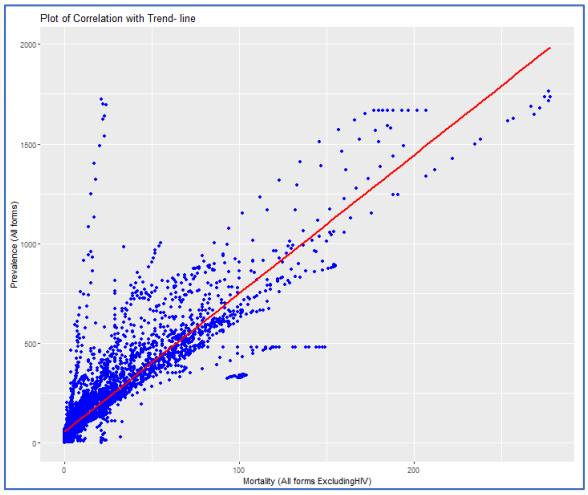
First, let's learn the difference between prevalence and incidence. Prevalence refers to the number of cases of TB that are present in a population at a given time. Prevalence includes newly diagnosed people, plus people who were diagnosed in the past, and people who haven't even been diagnosed. Incidence refers to the number of new cases of TB that develop in a given period. Incidence means the number of people who are newly diagnosed with TB in a given period (usually a year). There is a positive correlation between these two variables. The trendline is also showing the positive correlation in both the visualizations using Tableau and R-shiny. Below are few calculated fields P-value, Intercept, Standard error and R-Squared (0.73643) which further supports our conclusion.

Term	**Value**	**StdErr**	**t-value**	**p-value**
Estimated incidence (all forms) per 100 000 population	1.14746	0.011302	101.527	< 0.0001
intercept	51.5753	3.06419	16.8316	< 0.0001

4. Is there a correlation between Prevalence and Mortality of TB cases per 100,000 population?

BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION

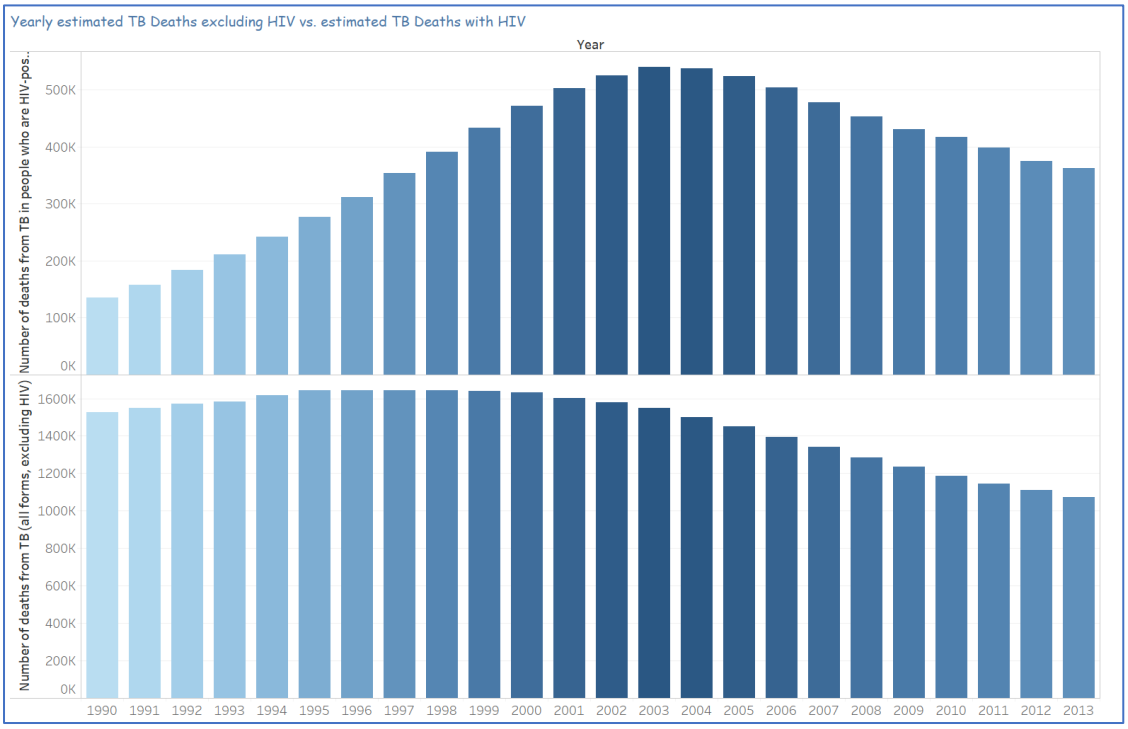




There is a strong correlation between prevalence and mortality. The trendline is also showing the positive correlation in both the visualizations using Tableau and R-shiny. Below are few calculated fields P-value, Intercept, Standard error and R-Squared (0.76862) which further supports our conclusion.

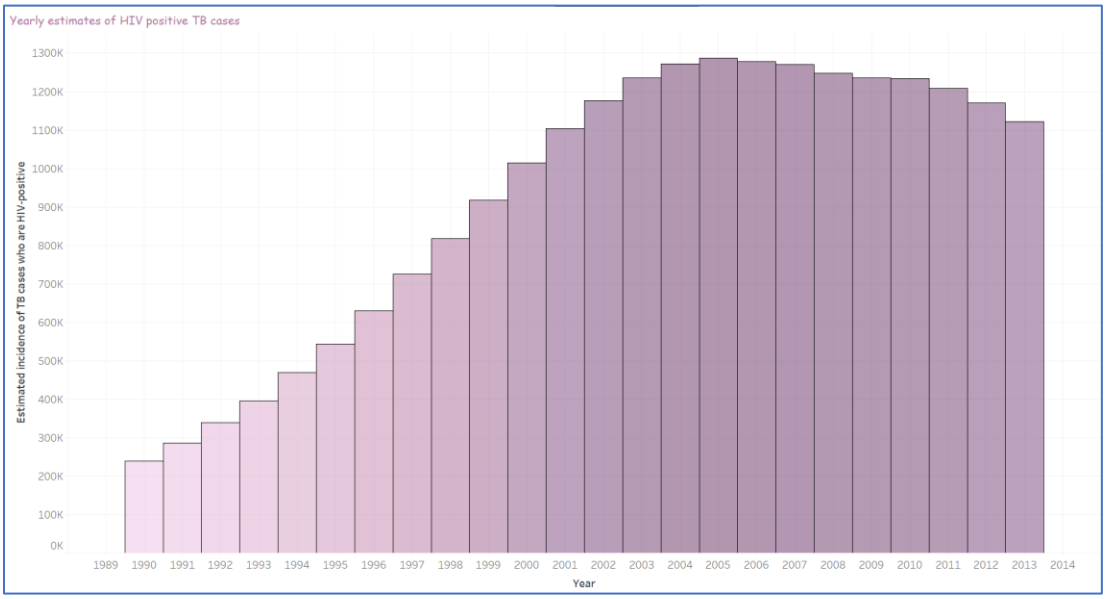
Term	**Value**	**StdErr**	**t-value**	**p-value**
Estimated mortality of TB cases (all forms, excluding HIV) per 100 000 population	6.87646	0.0571373	120.35	< 0.0001
intercept	60.7093	2.35907	25.7344	< 0.0001

5. Is there any pattern associated with estimated TB deaths excluding HIV and estimated deaths with HIV?



A peak year by number of deaths with HIV was 2003, after that number of deaths started decreasing each year. The number of estimated deaths of TB excluding HIV was increasing during the years 1990-1998 and decreasing during the years 1998-2013.

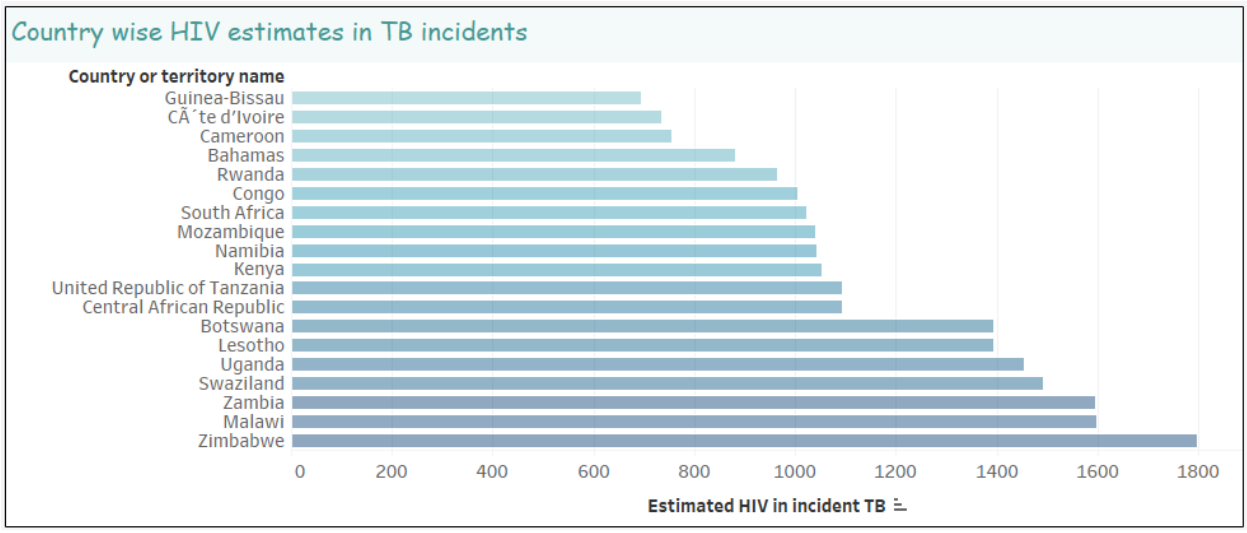
6. Is there any pattern associated with estimates of HIV positive TB cases?



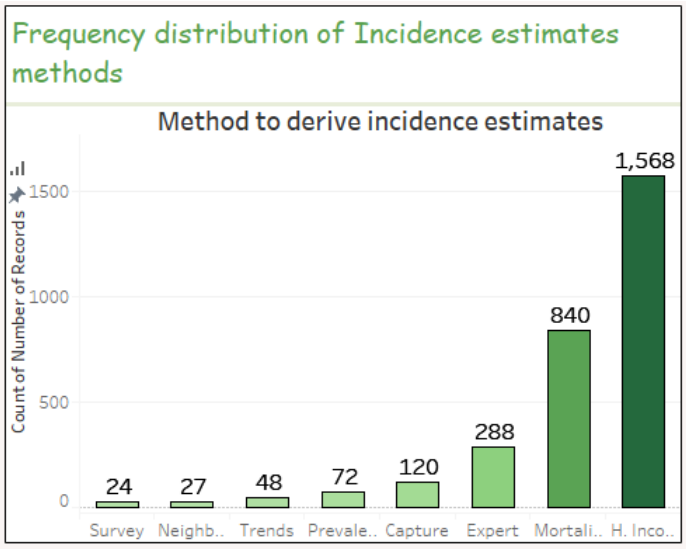
BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION

The number of HIV positive cases increased dramatically during the years 1900-2005 and decreased slightly during the years 2005-2013.

7. What are the countries with the highest percent of HIV in incident?

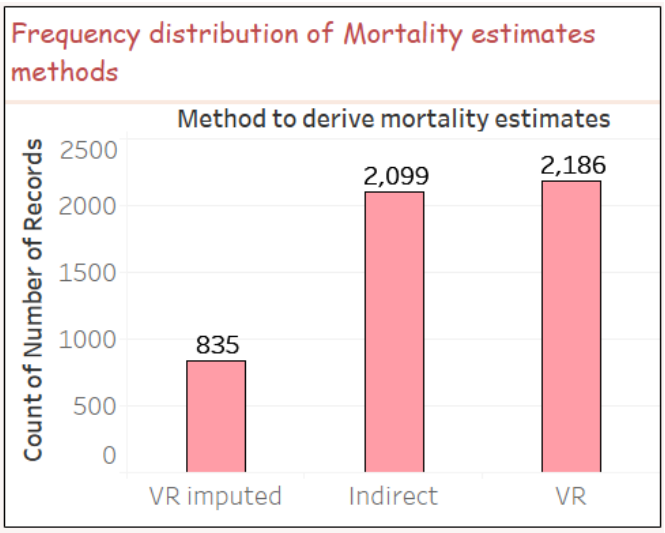


8. What are the prevalent methods to derive incidence estimates?



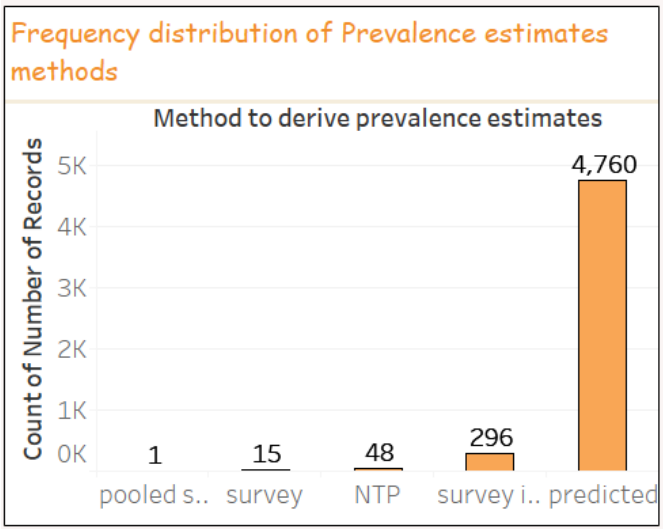
The prevalent methods are high income, mortality and expert.

9. What is the prevalent method to derive mortality estimates?



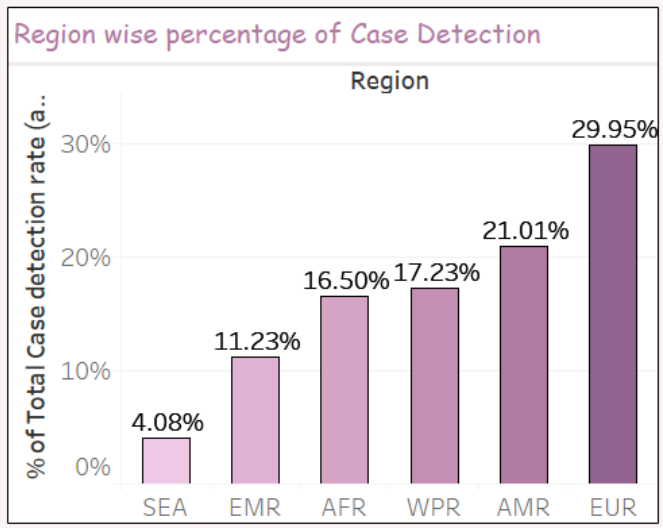
Vital registration (VR) is the prevalent method to derive mortality estimates and the second one is indirect. It is considered that both methods give the same results.

10. What is the prevalent method to derive prevalence estimates?



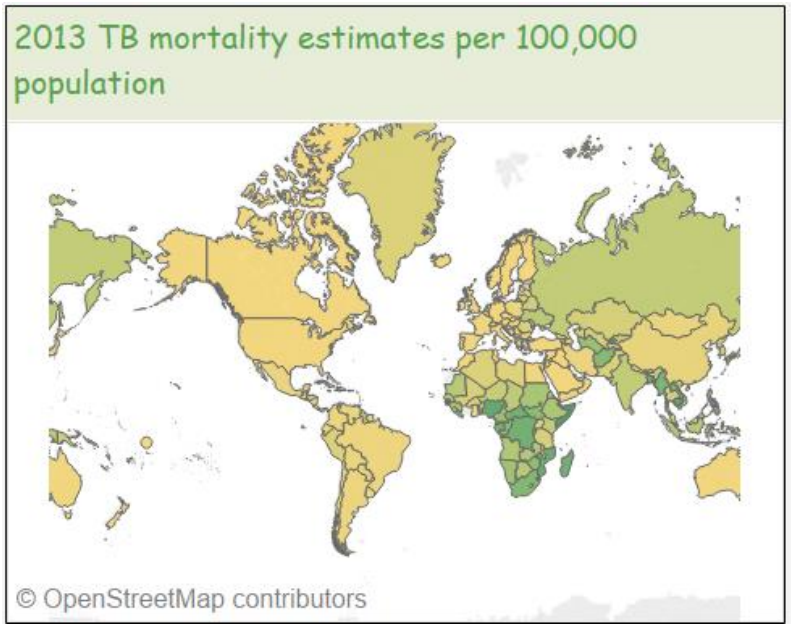
Predictive method is used in 92% of cases for prevalence estimates.

11. What are the regions with the highest percent of case detection?



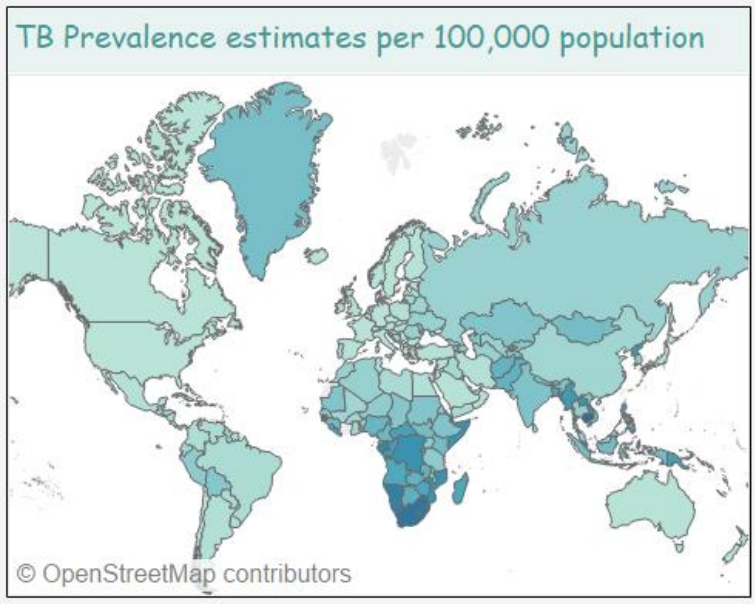
Europe (30%) and America (21%) are regions with the highest percentage of case detection.

12. What are the mortality estimates country wise in the year 2013?



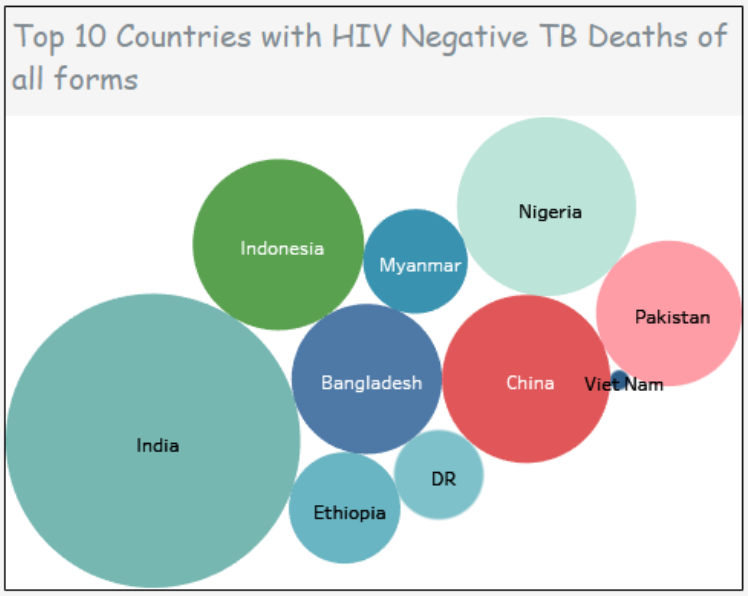
The highest mortality is detected in Somalia, Cambodia, Nigeria and Bangladesh. The lowest mortality is in United States, Canada, Switzerland, Netherlands, Norway and Sweden.

13. What are the prevalence estimates per 100,000 population country wise in the years 1990-2013?



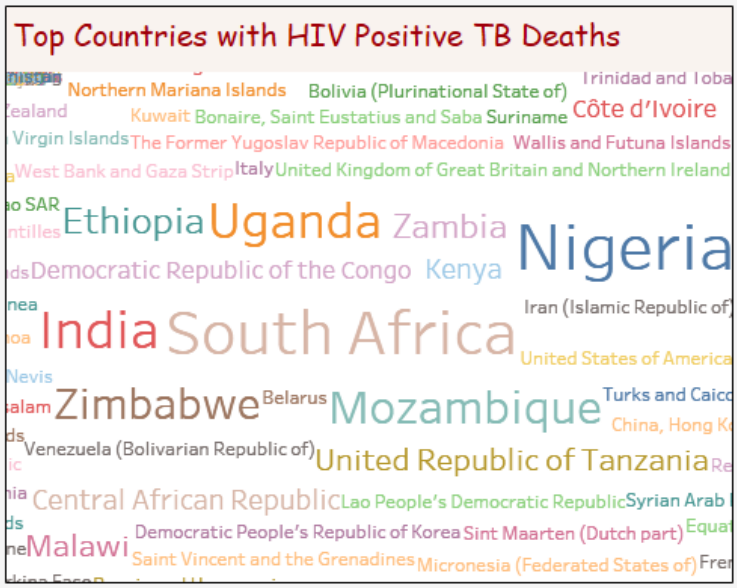
The highest prevalence is in South Africa, Namibia, Gabon, Cambodia. The lowest is in United States, Canada, Germany and Finland.

14. What are the top 10 countries with Deaths of TB (all forms) excluding HIV?



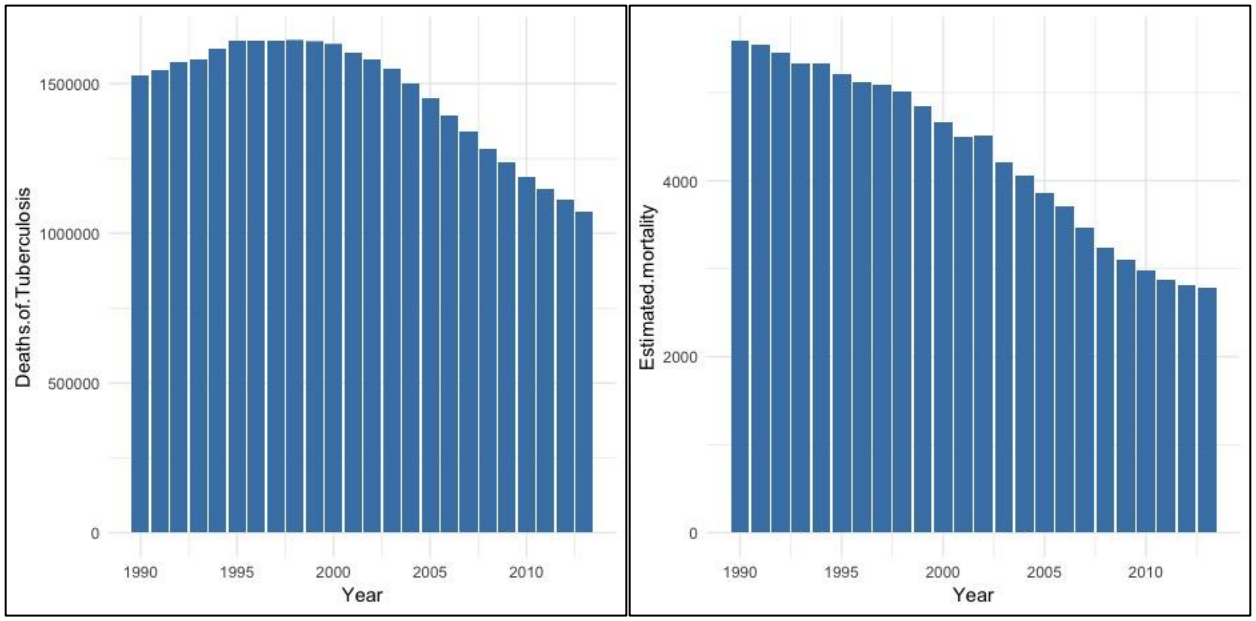
Those countries are India, Indonesia, China, Bangladesh, Pakistan, Nigeria, Ethiopia, Myanmar, Congo, Vietnam.

15. What are the top countries by deaths of TB (HIV +)?



Those countries are Nigeria, South Africa, Mozambique, India, Uganda, Zimbabwe, Ethiopia, Malawi.

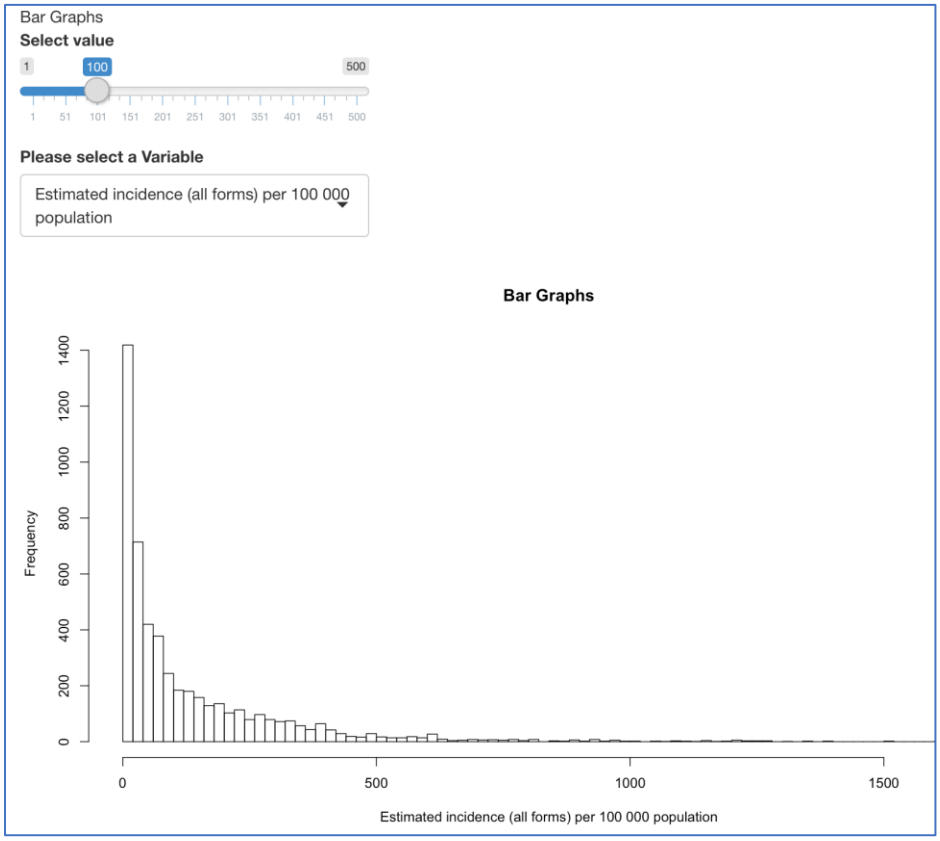
16. Is there any pattern between deaths and estimated mortality per 100,000 population of TB (all forms)?



As we can see from the graphs, even though there was an increasing tendency for number of deaths, the mortality goes down all the time during the years 1990-2013. It means that there is a positive

tendency and percent of people who die because of TB was decreasing.

17. Is there any pattern associated with estimated incidence (all forms) per 100,000 population?

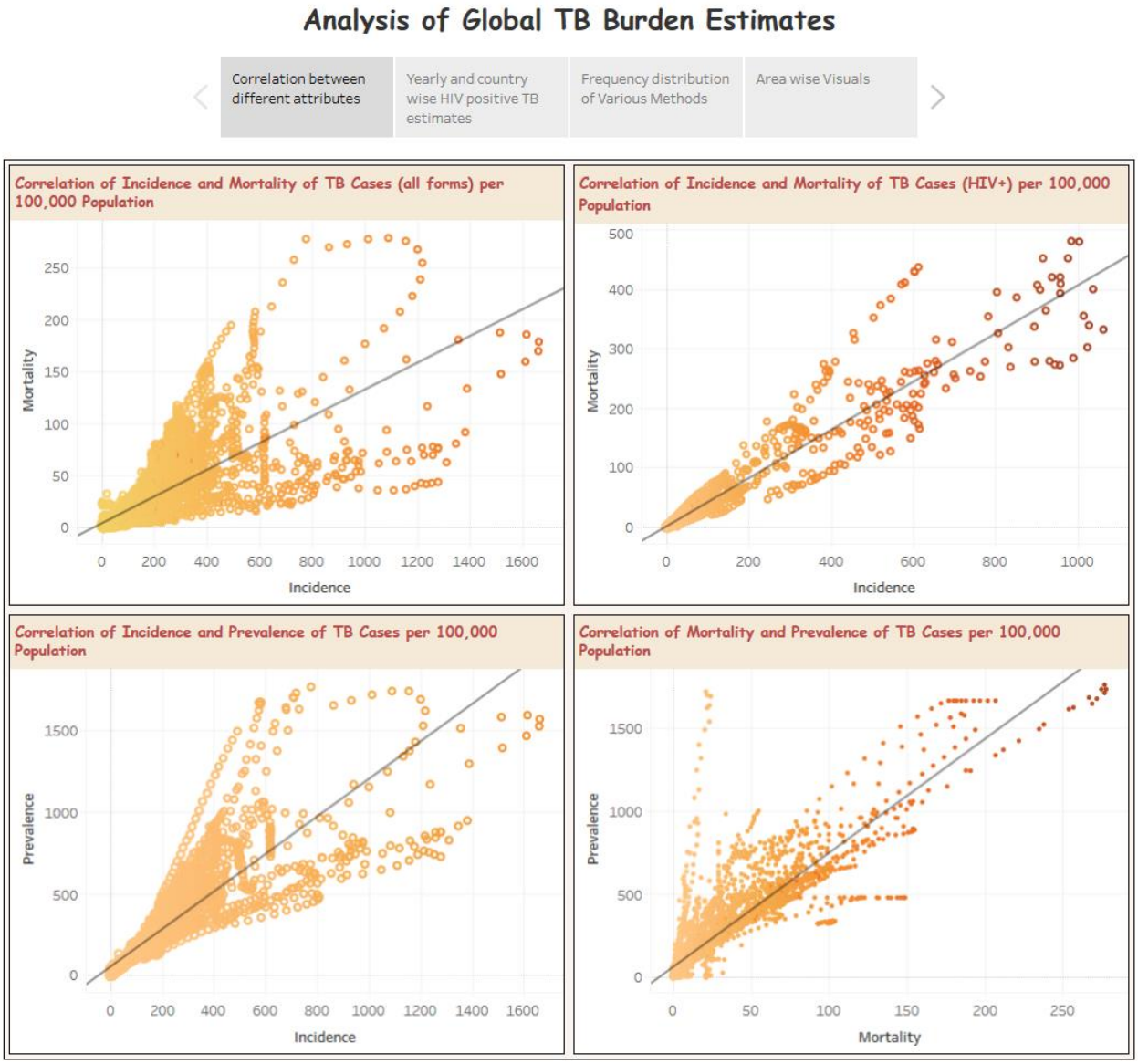


The median value of estimated incidence is 61, for low bound 57 and for high bound 69. Depending on the year and country, the range of estimated incidence differs from 0 to 1662. Estimated incidence higher than 700 can be observed extremely rarely, in less than 0,03% cases. Thus, there exists a declining trend which can be observed from the above chart.

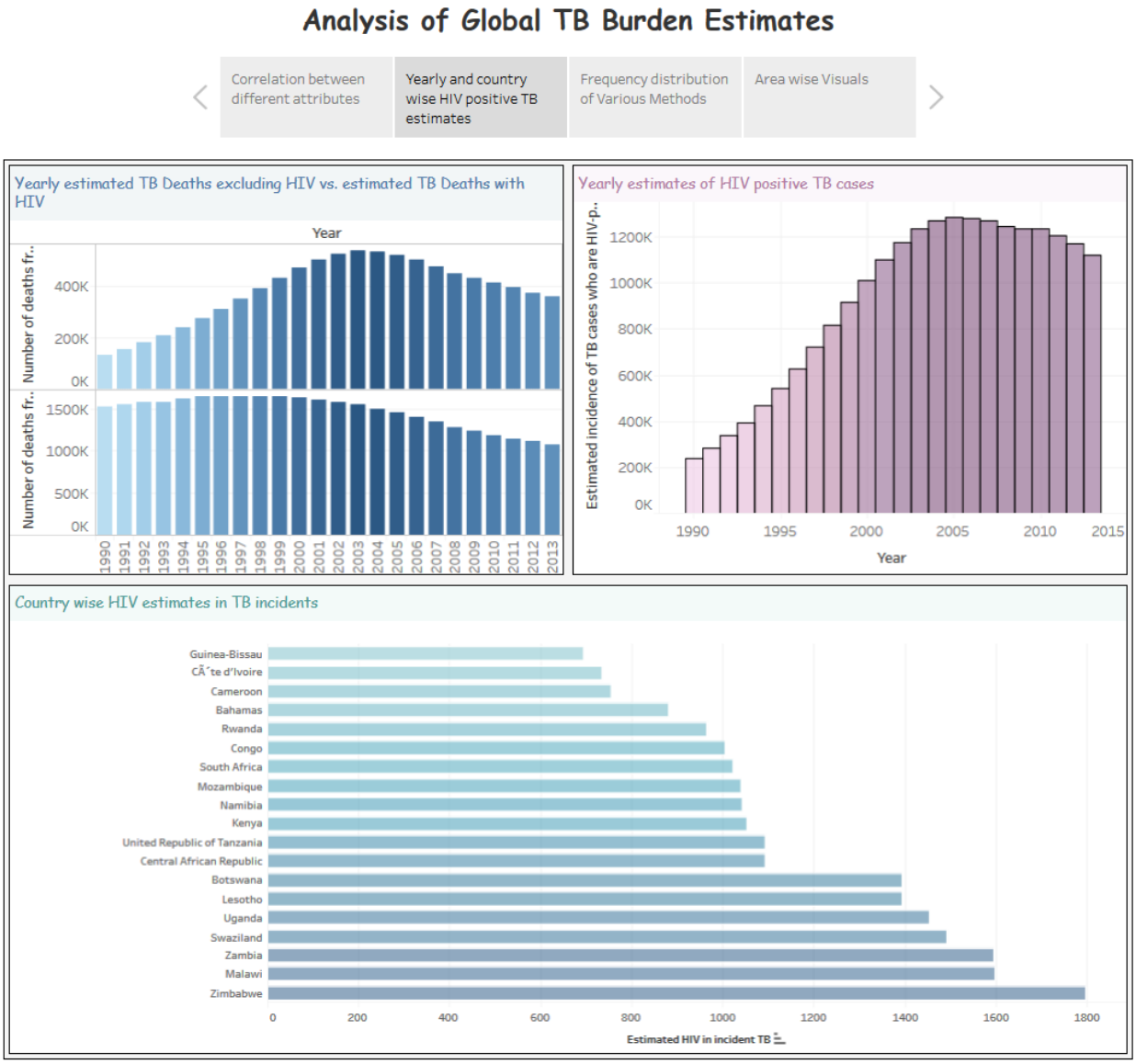
The dashboard story can be accessed from the Link - Link

Following are the screenshots from Tableau

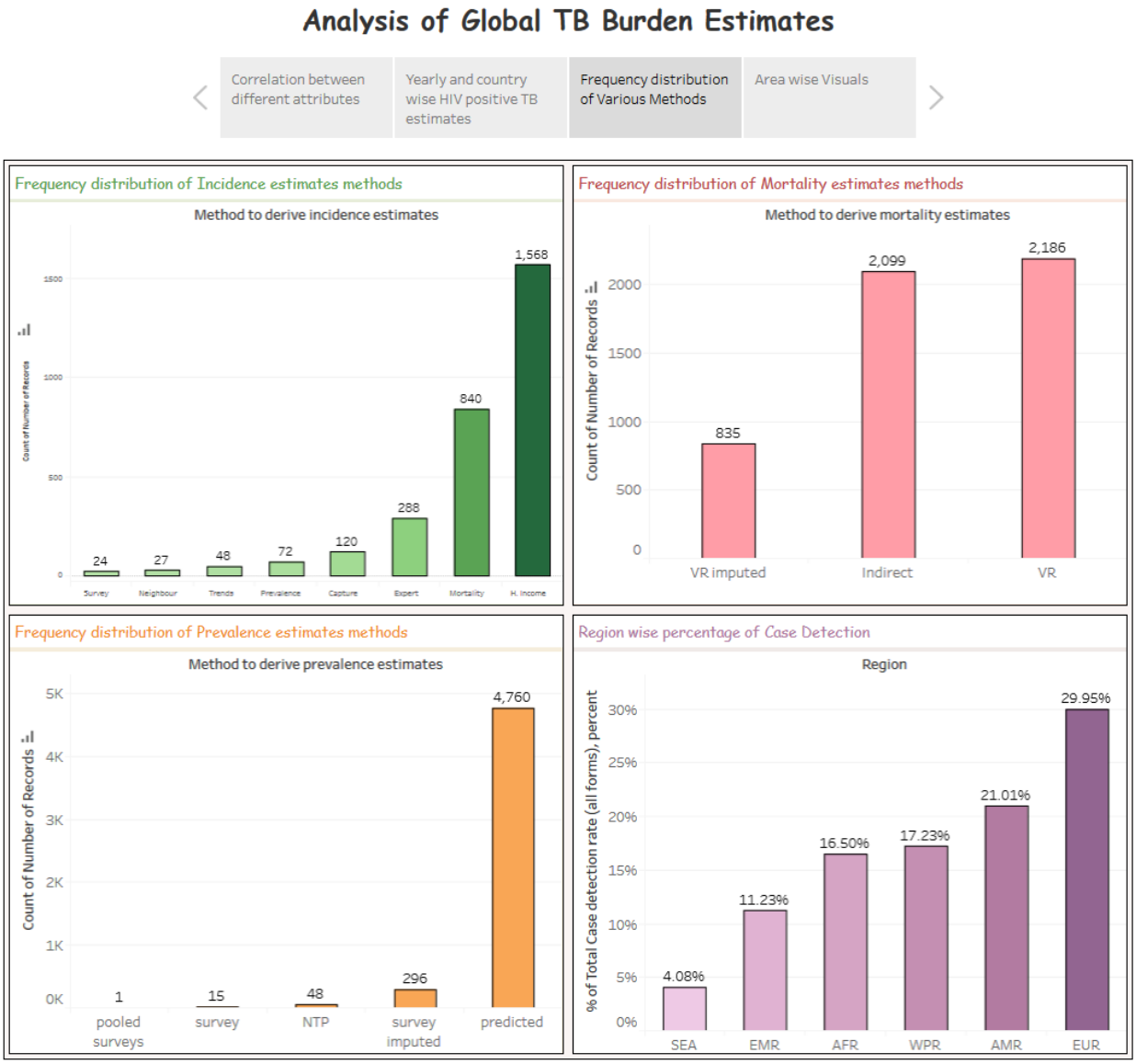
BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION

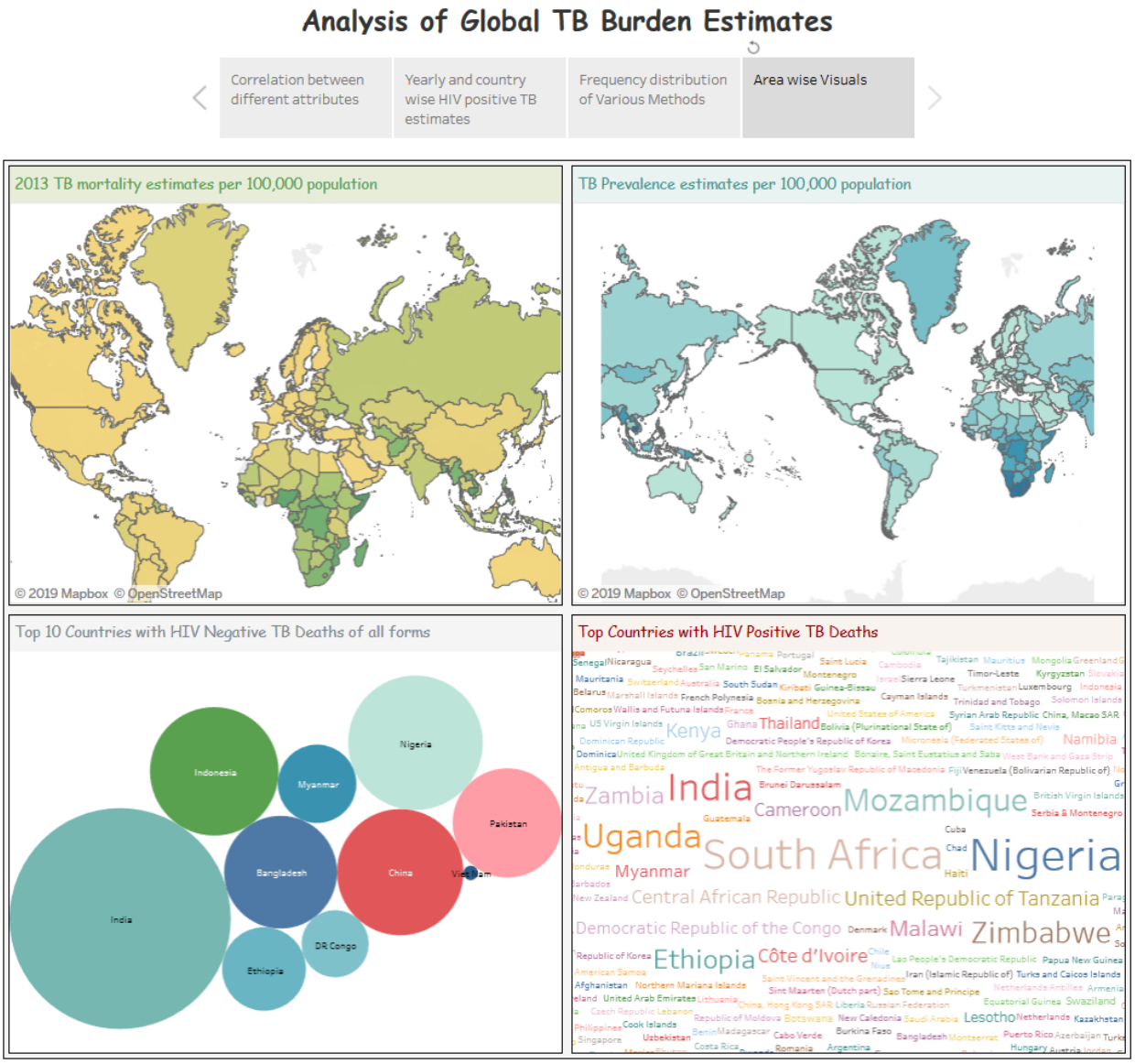


BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION



BRINGING IT ALL TOGETHER: STORYTELLING AND APPLICATION





FINDINGS:

After performing analysis of the provided dataset on tuberculosis in years 1990-2013, we came to important conclusions. First, the mortality has decreased, that is a positive sign. Another important conclusion is that TB mortality depends on financial situation in the country, the higher GDP is, the lower TB mortality is. The lowest mortality is in Norway, United States and Canada. The highest mortality of TB is in Somalia, Nigeria and Cambodia.

Throughout the world, the median value of estimated incidence per 100,000 population is 61.

Deaths of Tuberculosis are positively correlated with HIV+ results. The countries with the highest percentage of HIV+ incidents are Zimbabwe, Malawi and Zambia.

There is a positive correlation between incidence and mortality and prevalence and mortality. For the second one, the correlation is stronger, as incidence is newly detected cases and prevalence includes people who were diagnosed in the past.

The highest percentage of case detection belongs to Europe (78%) and the lowest is in Africa (51%).

REFERENCES

1. Cole Nussbaumer Knaflic, (2015). *Storytelling with data.* Retrieved from http://www.bdbanalytics.ir/media/1123/storytelling-with-data-cole-nussbaumer-knaflic.pdf
2. Annabel Kanabus, (2018). "*Information about Tuberculosis*". Retrieved from https://www.tbfacts.org/tb-statistics/