Hello everyone, so for this project, I created a tableau story on analyzing the burden of tuberculosis throughout the world.

title slide:

A bit of background information:

- TB is an infectious disease, primarily affecting the lungs

- Transmitted through droplets from coughs and sneezes

- A leading causes of death from infectious diseases worldwide

- Control Efforts focus on early detection, vaccination, and effective treatment

NEXT SLIDE

Overview slide:

* To start off we will be comparing totals and the normalized data by population.
* Next, I will cover the mortality and TB-HIV co-infection rates.
* I’ll then review some more positive trends that can be seen.
* And finally, I’ll summarise the key points.

NEXT SLIDE

Data normalization – incidences per 100k map

* The graph in the bottom left shows the average *total* yearly cases.
  + You can that China and India are far ahead of any other country when it comes to total case count.
* The map and bar graph on the right, however, show the number of cases per 100k by country.
  + Looking at this map, we see that China and India are no longer some of the top countries once their populations are normalized.
* You can also see that in the top 15 countries there is less variance between the top 2 countries and the rest of the list.

NEXT SLIDE

100k cont:

* Here we can start to identify which regions are the most heavily affected.
  + The bar graph on the left is the same as from the previous page but has now been filtered by region.
  + We can see the overall average (of all countries) and the average for each region as we change the filter.
  + As you can see, the two regions that are furthest above the overall average are Africa and SE Asia
* Other important statistics to look at are the prevalence and prevalence rate.
  + Prevalence is the number incident cases during a specific time and within a specific region.
    - The prevalence rate is that value divided by the population for the same region and time.
    - A lower prevalence indicates fewer active cases.
    - Since prevalence rate is divided by the population it is also a normalized value
  + In the bottom left we can see the prevalence rate for our 6 regions over the course time.
  + On the bottom right, we can see a scatter plot of the population plotted against the prevalence.
    - As you can see, India and China can be considered outliers due to their large populations. We can remove them for now to get a closer look at all the other data points.
    - You can see that as time goes on, the prevalence fluctuates between individual countries, but overall decreases.
      * As the dots move closer to the Y axis, their prevalence rate, and thus cases are going down.

NEXT SLIDE

Mortality and HIV

* To determine the burden of TB, we need to look at the mortality rates and TB-HIV co-infection rates.
* A lower mortality rate could mean that certain control efforts are working.
* As we can see from these graphs, the 2 most affected regions are SE Asia and Africa. They outpace the 3rd highest mortality rate by a significant margin.
  + Overtime, we see that the overall deaths however does decrease.
* Looking at our prevalence vs population graph, we can see that the clusters tend to be grouped based on prevalence, rather than population.
* One very convincing trend we see in our data is the relationship between TB-HIV infections and the mortality rate.
  + In this graph in the bottom right, we can see a trend line with its confidence bands.
  + Visually inspecting the data seems to be an obvious trend.
  + The R-squared value for the trend line is 0.94 implying a very good fit for the data.
  + This means that the mortality rate for HIV positive TB patients is highly correlated.
    - This may feel somewhat obvious, as HIV is an immunodeficiency virus, however if the goal is to minimize deaths, then it’s important to make note of that.

Progress

* Even though TB is still responsible for many deaths annually, there have also been many positive trends identified in the fight against the disease.
* As we saw with the prevalence, all 6 of the regions currently have a negative trend for their prevalence rates up to 2013.
* We can also see that co-infection deaths (in the top right) have been trending downward since the height of the HIV epidemic in the late 90’s/early 2000’s.
* Watching the graph progress in the bottom right we can also see that the trendline’s slope is becoming less steep.
  + For this graph, that shows that there are less deaths per case as time goes on.
  + This trendline has an R-squared value of 0.78 and can be a reliable fit.

Recap

* So just to recap everything that was covered,
* When looking at the burden of a disease, it’s important to normalize the data by population.
* Data shows that since 1990 the most burdened regions are Africa and Southeast Asia
  + Africa suffered from TB significantly in the late 90s during the HIV epidemic.
* Despite the large burden TB places on the different countries, there have been positive trends,
  + This may indicate that the current control methods are effective
* With more time, it would be interesting to take a deeper dive into the data, to identify where some of the more heavily burdened countries are struggling.
  + Additional data would also aid this research.
    - This data set was limited to incident cases, and mortality, but other socioeconomic factors were not included.