Analysis of the US weekly Nationally Notifiable Disease Surveillance Data

PROBLEM STATEMENT

- The aim of this assignment is to analyze the data provided by the “National Notifiable Disease Surveillance System” hosted on Project Tycho’s website for the effects in each state from the years 1888-2013. We Look into the diseases with highest number of cases, deaths and mortality. Further we do statewise and year wise analysis of Tuberculosis and analyse the patterns regarding number of cases, deaths and mortality for Tuberculosis. We also take into account the population for statewise analysis of the Tuberculosis. The population data is created by us by retrieving the individual year population of each state in US from “Macro Trends Website”.

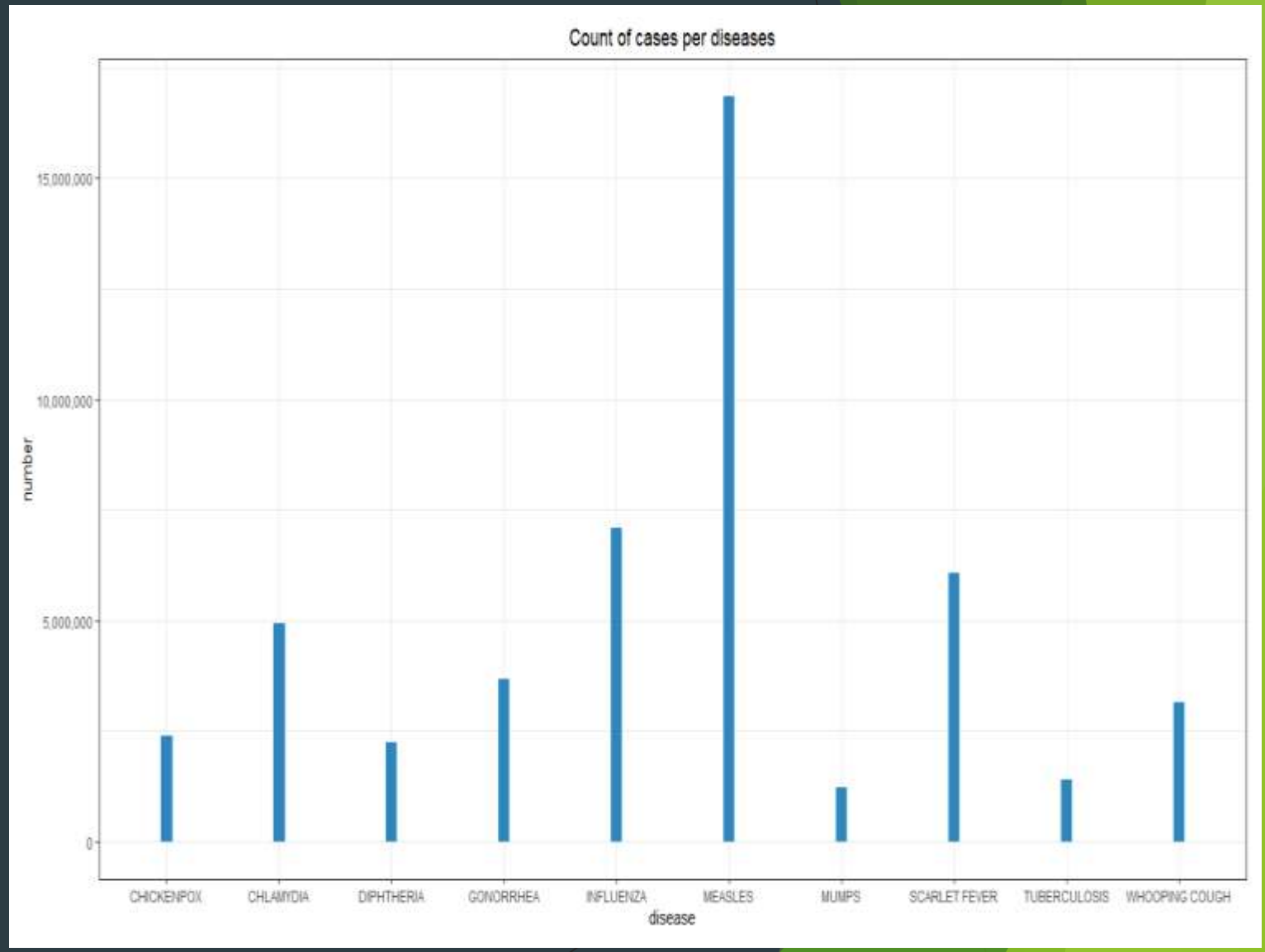
ABBRIVATIONS

AL : Alabama
AK : Alaska
AZ : Arizona
AR : Arkansas
CA : California
CO : Colorado
CT : Connecticut
DE : Delaware
FL : Florida
GA : Georgia
HI : Hawaii
ID : Idaho
IL : Illinois
IN : Indiana
IA : Iowa
KS : Kansas
KY : Kentucky
LA : Louisiana
ME : Maine
MD : Maryland
MA : Massachusetts
MI : Michigan
MN : Minnesota
MS : Mississippi
MO : Missouri
MT : Montana
NE : Nebraska
NV : Nevada
NH : New Hampshire
NJ : New Jersey
NM : New Mexico
NY : New York
NC : North Carolina
ND : North Dakota
OH : Ohio
OK : Oklahoma
OR : Oregon
PA : Pennsylvania
RI : Rhod Island
SC : South Carolina
SD : South Dakota
TN : Tennessee
TX : Texas
UT : Utah
VT : Vermont
VA : Virginia
WA : Washington
WV : West Virginia
WI : Wisconsin
WY : Wyoming

TB : Tuberculosis
USA : United States of America
EDA : Exploratory Data Analysis

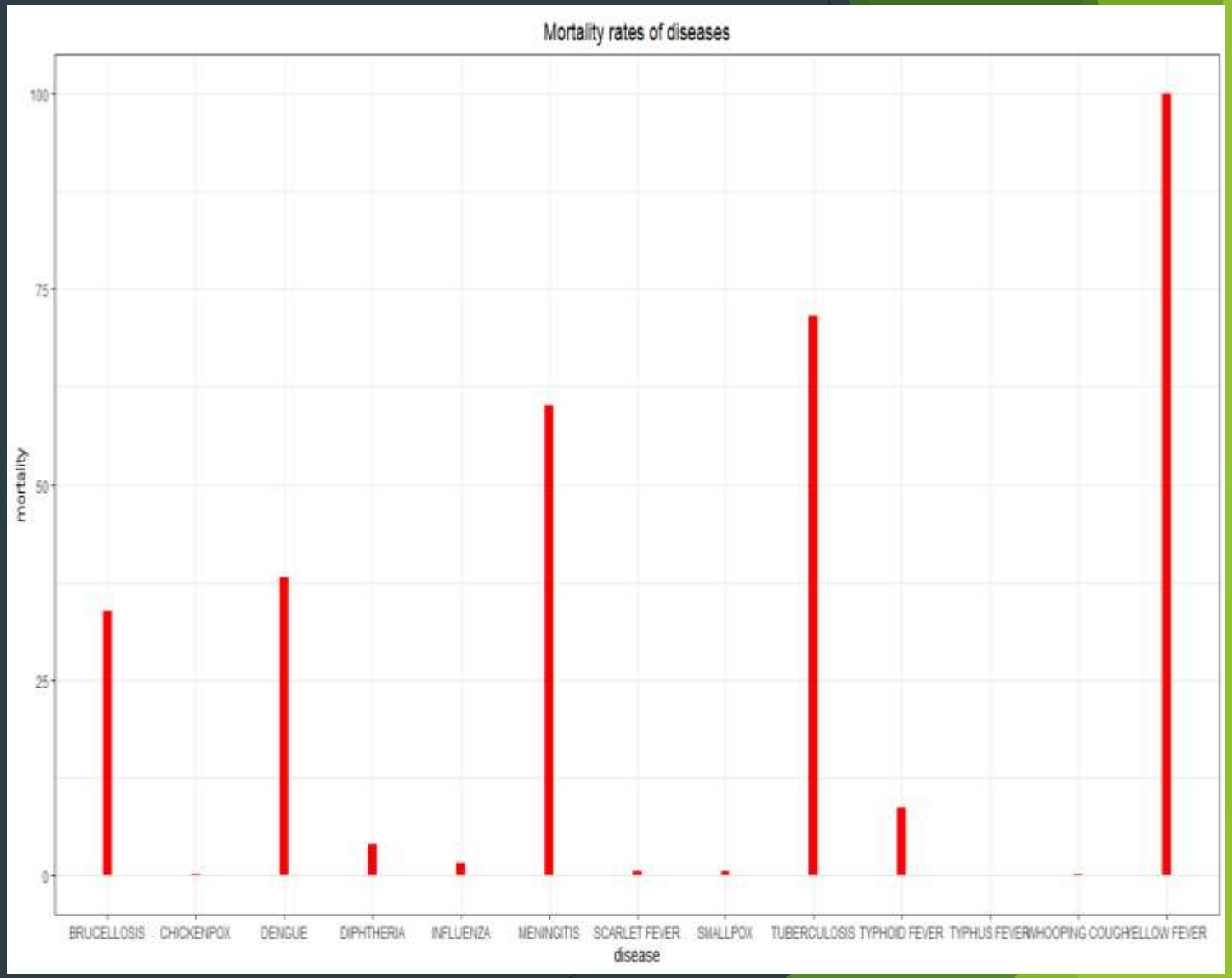
PLOT: COUNT OF CASES PER DISEASE

Analyzing top 10 most contagious diseases. As we can see, Measeles had highest number of cases followed by Influenza, Scarlet Fever, Chlamydia, Gonorrhea, Whooping Cough, Chickenpox, Diphtheria, Tuberculosis, Mumps.



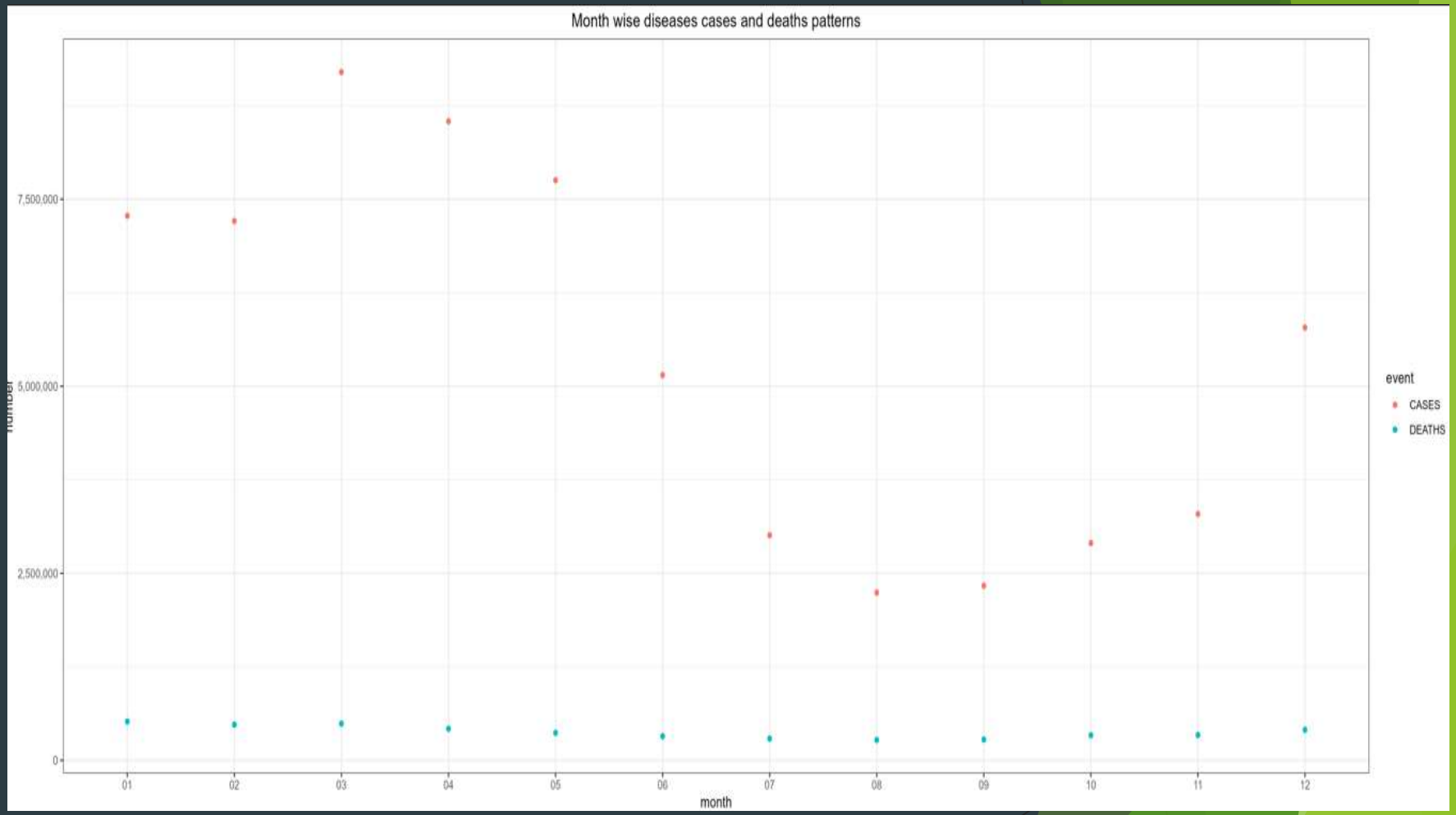
PLOT: DISEASES WITH THE HIGHEST MORTALITY

Yellow fever has unusually high mortality rate of 100%(Yellow fever generally has a mortality between 20% and 50) followed by Tuberculosis, Meningitis and Dengue.



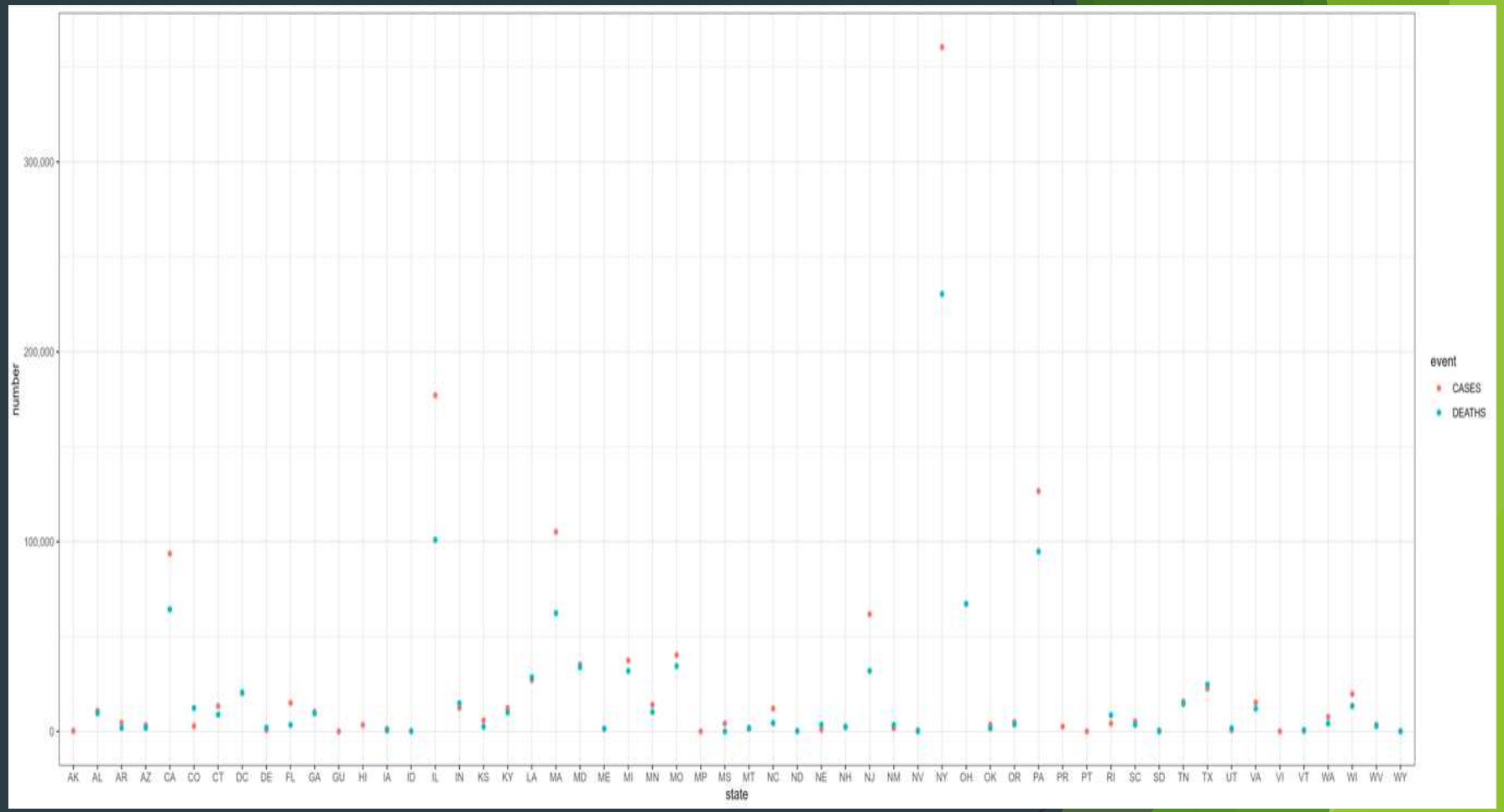
PLOT: SEASONAL TRENDS ANALYSIS OF EVENTS

The graph shows a high spread of the diseases from the month of September till March. The increase in number of cases is accompanied by increase in number of deaths. From March, we witness a decline in number of cases and deaths which goes in till August. The higher number of cases may be attributed to winter as the period from Sep-Oct to Feb-Mar have lower temperatures. In summers diseases tend to be less contagious.



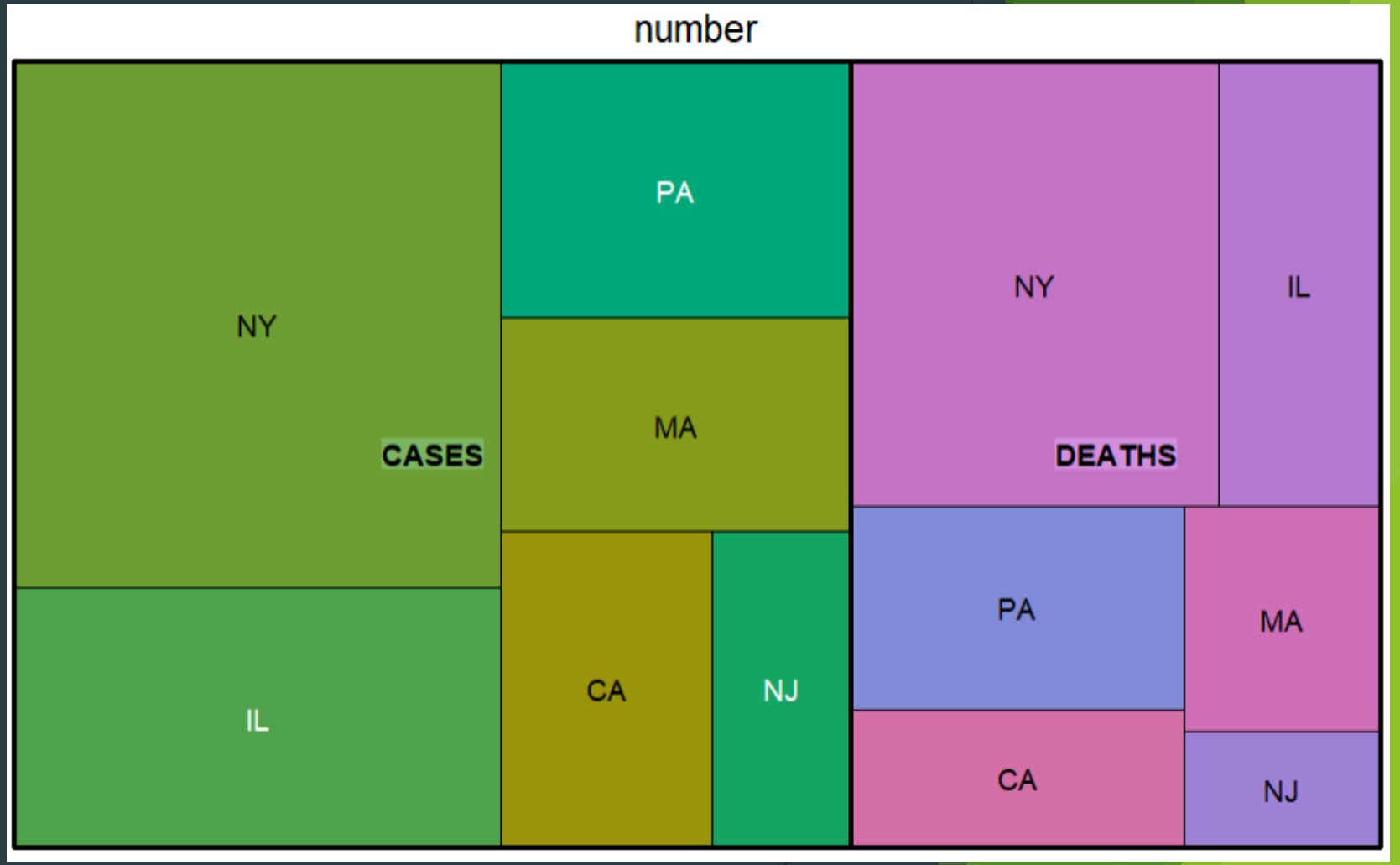
PLOT: IMPACT OF TUBERCULOSIS STATE WISE

CO has more deaths then cases which is not possible. Also few states don't have any data either on deaths or on cases for Tuberculosis. So we pick NY, IL,MA, PA, CA, OH for state wise comparison.



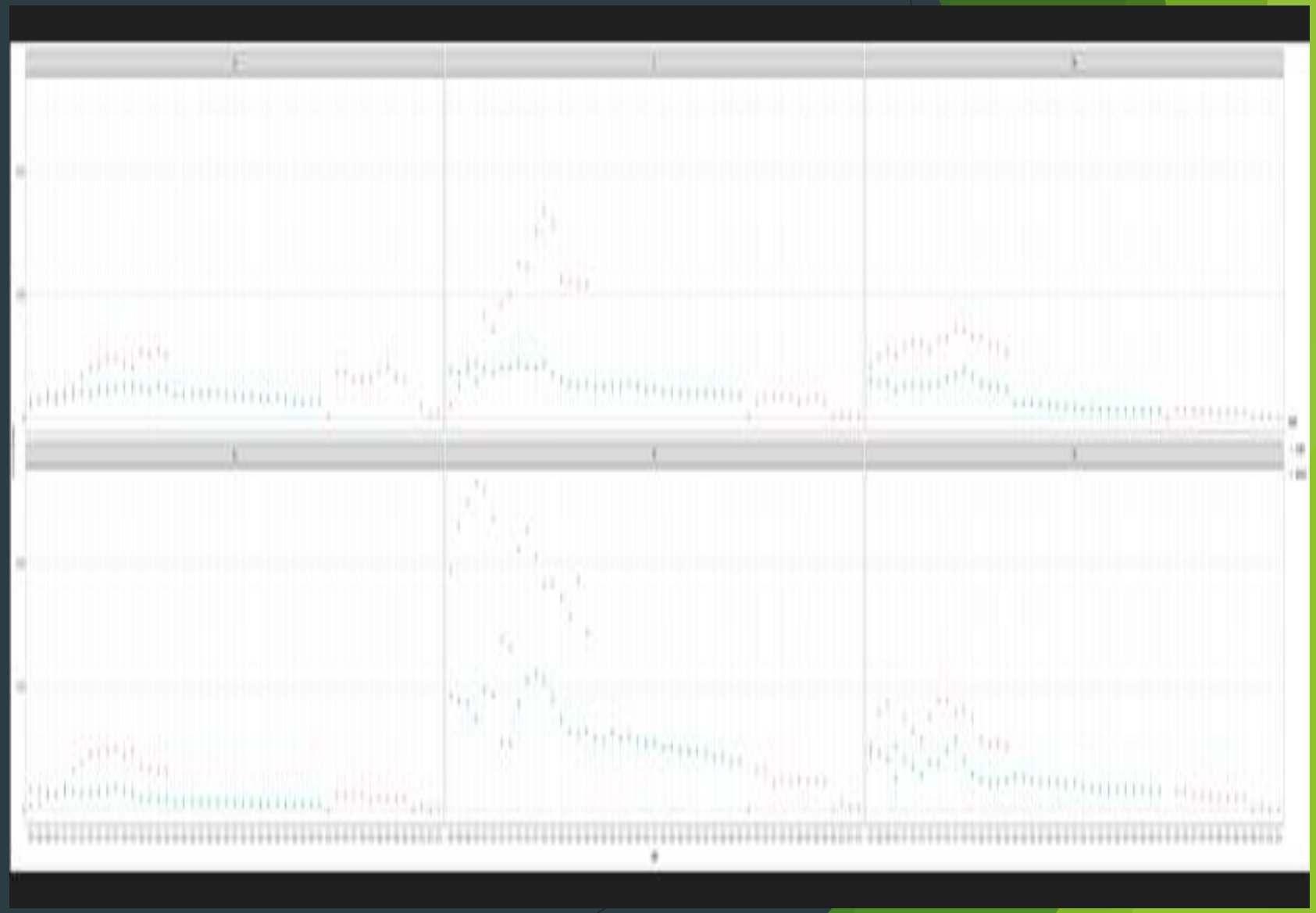
PLOT: TREEMAP FOR COMPARISION OF CASES AND DEATHS FOR SELECTED STATES

We group the TB data by state event and year.As We can see, New York has highest number of Cases followed by Illinois, Pennsylvania and Massachusetts. New York had most deaths followed by Illinois, Pennsylvania and California. Further we analyse year wise patterns for these states.



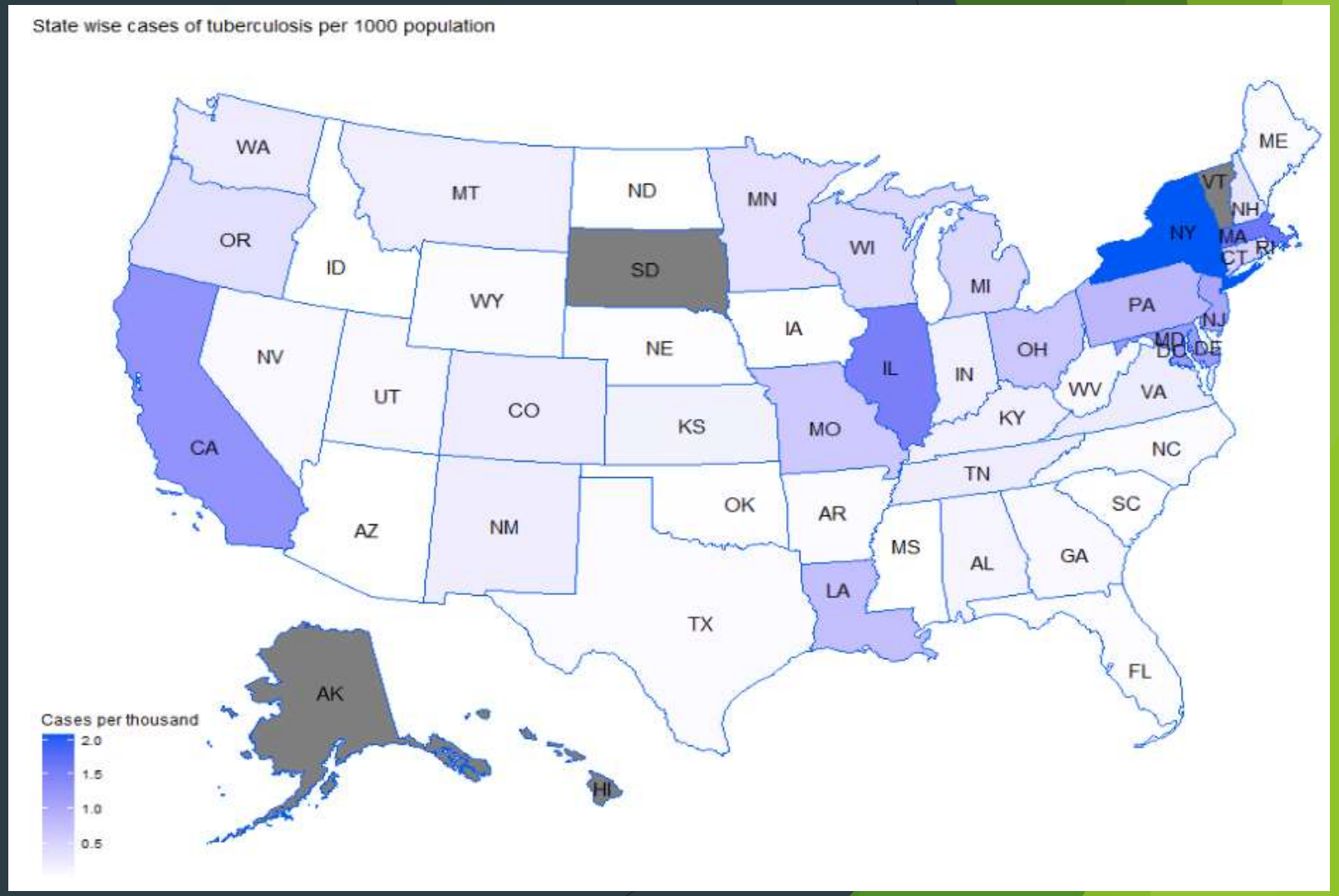
PLOT: Comparison for Tuberculosis cases between states by year

Now we do the analysis year wise. To get a better picture. Higher number of cases in followed by higher number of deaths in all states with exception in case of California, New Jersey and Illinois in the years 1907-1911. In these years the number of deaths are unusually high compared to number of cases. Number of cases and deaths tend to be higher for period 1915-1920 with New York showing very high cases and deaths in the years 1907-1911.Post 1920 we see decline in cases. This can be attributed to public clinics and better prevention education campaigns taken during those times. As we can see for each state, data for both cases and deaths is only available till 1923. So for further analysis, we will only consider years 1907-1923.



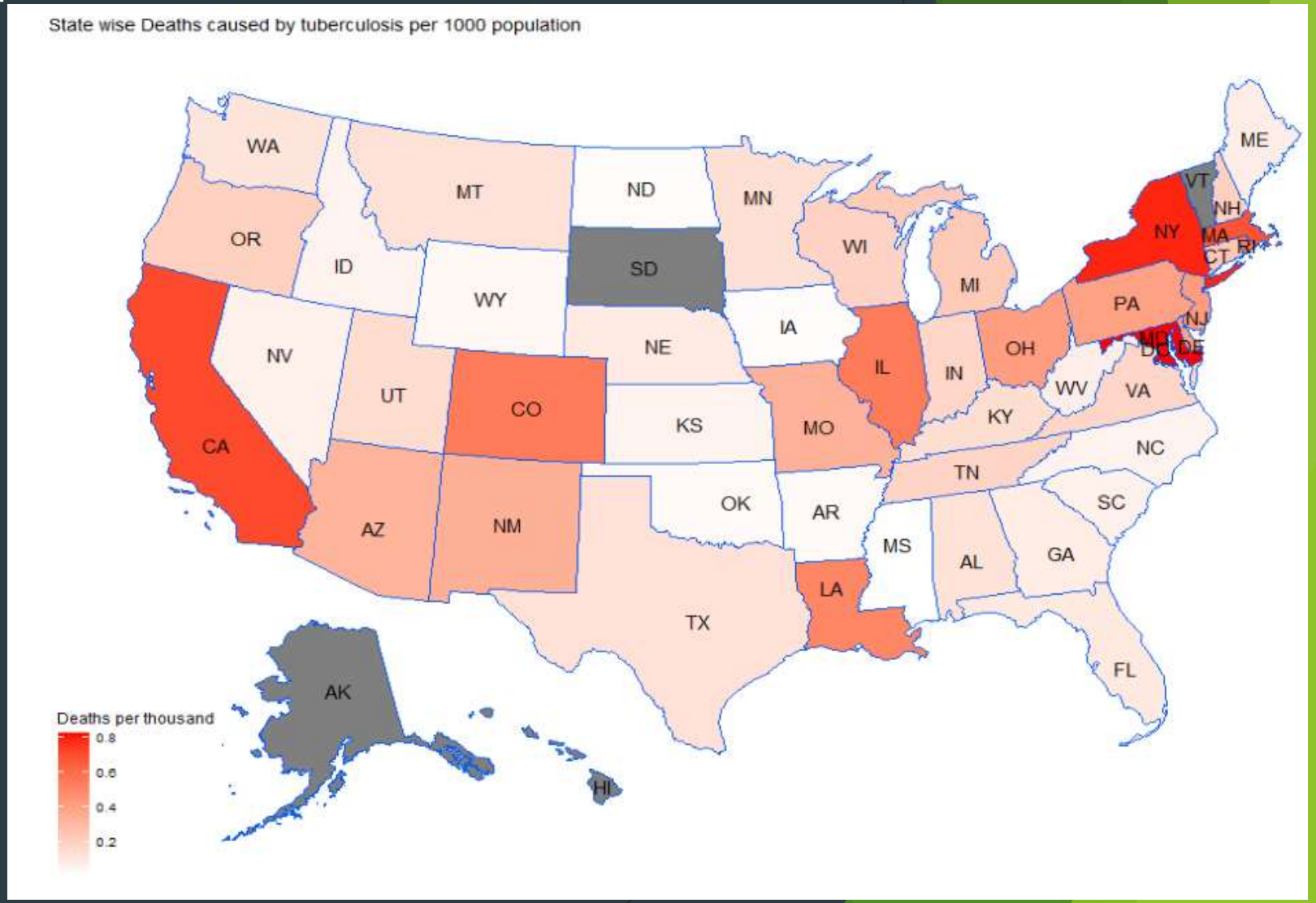
PLOT: COMPARISON OF STATES ON NUMBER OF TUBERCULOSIS CASES PER 1000 PEOPLE

For per 1000 population cases, New York had highest cases of tuberculosis, followed by Massachusetts, Illinois, California and Maryland.



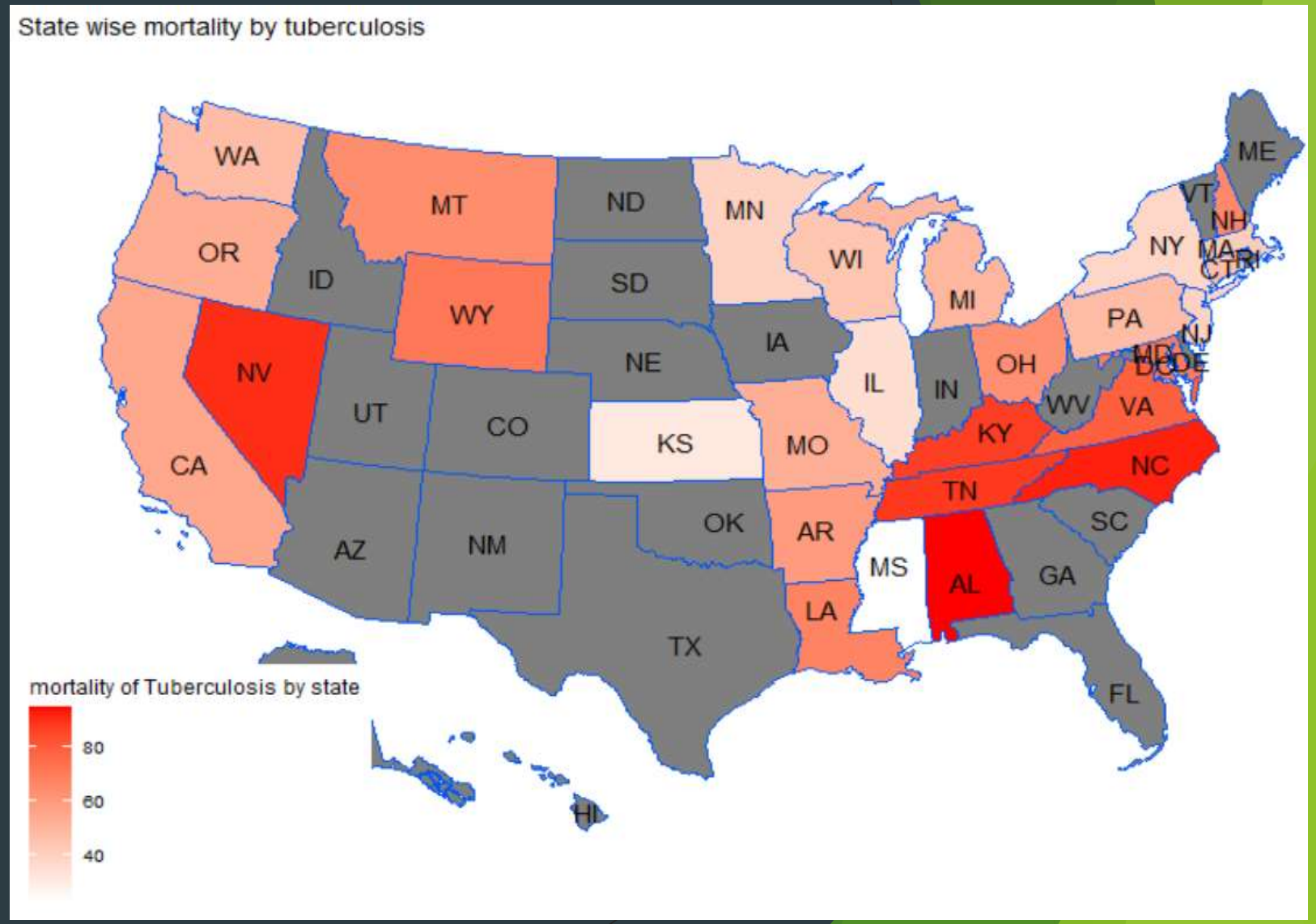
PLOT: Comparison of states on number of Tuberculosis deaths per 1000 people

Here, as we can see Maryland has highest number of deaths per 1000 population, followed by New York, California, Massachusetts and Rhode Island. Deaths per thousand in Rhode Island is considerably high given that it does not have a lot of cases per 1000 population.



PLOT: Comparison of states by mortality for Tuberculosis

As we can see a lot of states had mortality above 100 which is not possible. One of the states with mortality higher than 100 is Rhode Island. This combined with higher number of deaths per 1000 population could mean higher reported deaths in Rhode Island or then actual deaths or lower number of total reported deaths.



CONCLUSION

- Overall analysis of the data shows that NY, IL and PA are amongst the states with highest events for cases and deaths.
- Seasonal trend suggest that the diseases tend to spread more in winter season.
- We observed that Measles, Influenza, and Scarlet Fever are amongst the most contagious diseases. And Yellow Fever, Tuberculosis and Meningitis are amongst the diseases with highest mortality.
- State wise analysis of Tuberculosis, NY, CA, IL, MA are among the states with highest number of cases and deaths.
- Population wise analysis we saw NY, CA, MA, MD amongst the states with highest number of cases and deaths with MD registering comparatively higher deaths per 1000 population.

THANK YOU