

# CSC605 Data Science

Fall 2022

## Opioid Epidemic Data Analysis Stage II

### **Instructor:**

Dr Somya Mohanty

### **Members:**

Aditi Darandale

Manish Shah

Priyanka Budavi

Raveena Rakesh

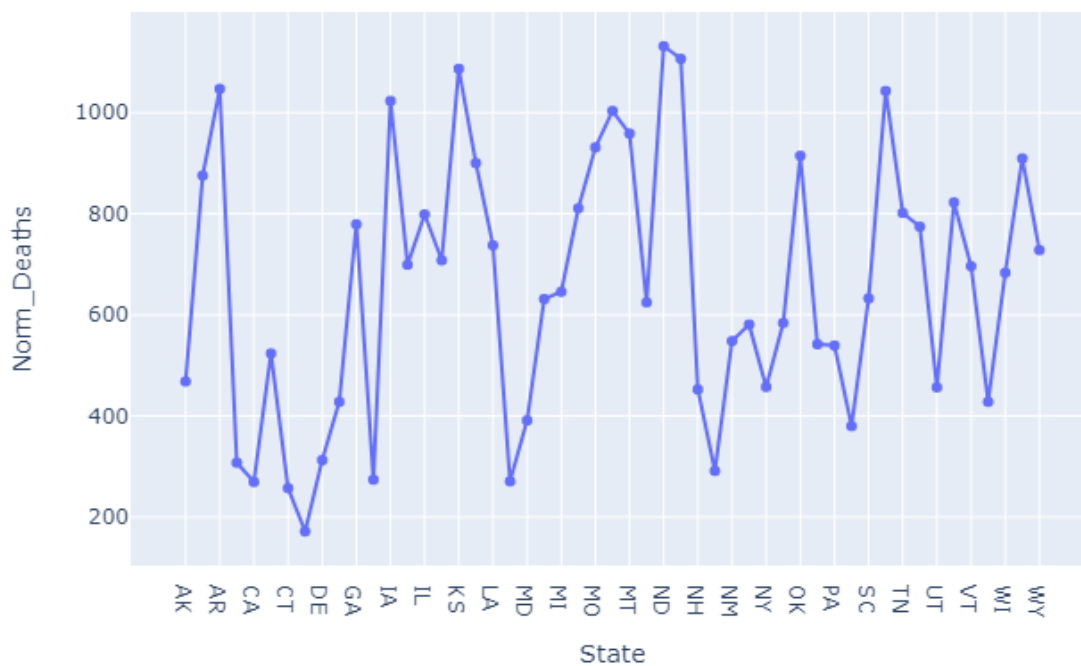
## Task 1: Importing historical data and identifying peaks in Opioid mortality.

The main idea behind this task is to analyze the peaks generated by the mortality rates, when we analyze them for each state and year.

To analyze the mortality rate for each state,

- First we read the 1999-2020 dataset, clean the dataset by filling the nan values.
- Normalize death rate by population per 100,000. So that the scale for measuring the rates across various states is the same.
- Groupby the states and then calculate the mean on the normalized deaths. Then we plot the values for each state, as shown in the graph below.

Death rate across different states

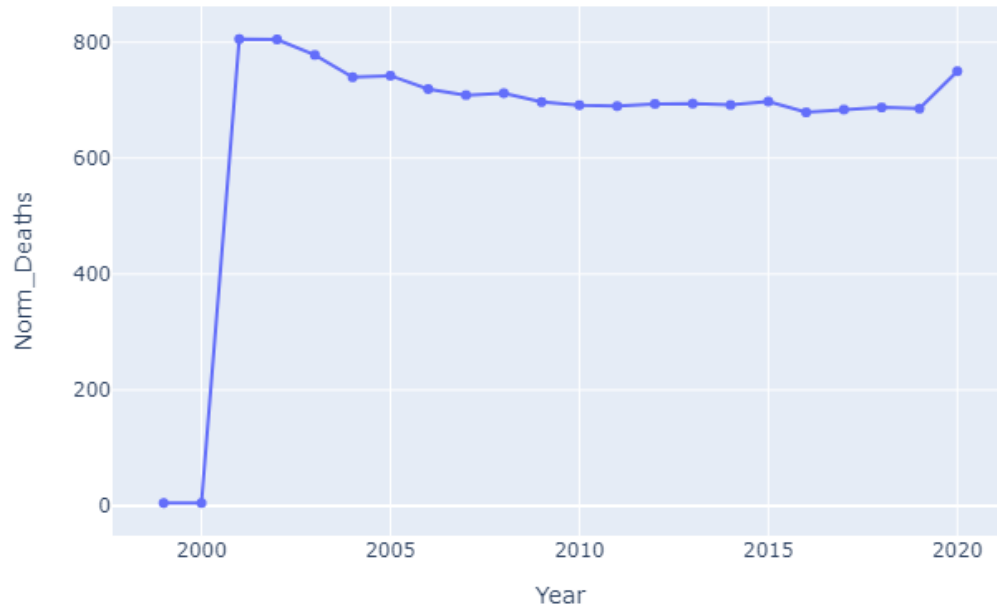


- From the above generated plot we can see that the highest mortality rate is across state ND and the mean death rate is 1131.521.
- Here, we hypothesize that the mortality rates across each of the states maybe influenced by the population and hence as the population increases so does the mortality rate increases.
- The peaks could have also been due to the opioid dispensing rate.

To analyze the mortality rate for each year,

- Since all the pre-processing is already done on the data, we only need to groupby the years and then calculate the mean on the mortality rate.

Opioid Mortality rate across different years across US



- In plot-2 the highest mortality rate was in the year 2001 with a mean death rate of 734.8614.
- Also, we can observe from the graph that the opioid mortality rate is significantly fluctuating, there was a sudden spike in the early 2000's and then the rates have been constant since then, this maybe because there is a lot of missing data for the years 1999 and 2000.
- The peaks could have also been due to opioid dispensing rate.

## Task 2: Identify trends in states

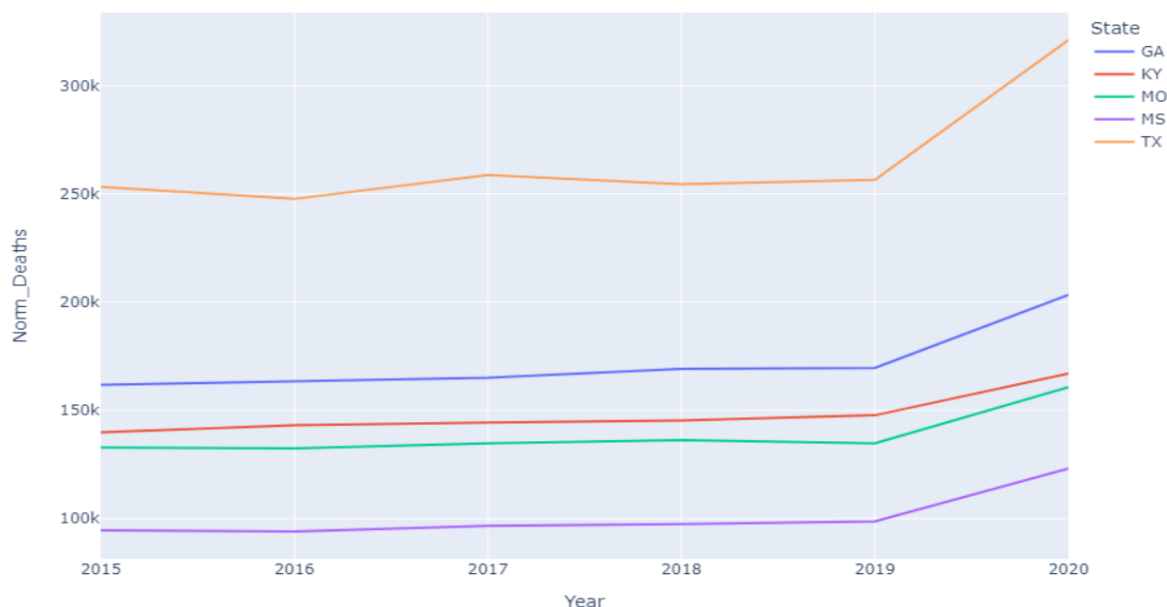
For this task we used '1999-2020\_Drug\_Overdose\_By\_Category.cs' to find the trends in the state. The dataset had only county information along with state abbreviation, so I split the County into state and county to get separate state information as this will be helpful for performing group by operation.

To find the increasing trend I used last 5 years information to see if there was an increase in the mortality rates. So, to compare each state with previous years we need to get the difference in the years for deaths column. Once the difference is obtained, we will group state wise and calculate the sum of the differences in death values. This will definitely give an idea of total deaths in each state per year. Thus, by using this information we can sort the deaths value and find the top five states with increasing values and top five states with decreasing values.

I also checked the plots to see if the trends show an increasing or decreasing trend for the last five years. The challenge involved in this task was to make sure that when we calculate the `diff()` the new state should have sum as zero because the values should be compared within the state. So in order to achieve this I filled the value with zero at the start of the year for each state.

### Task 3: Plot a graph of top 5 increasing and decreasing states in a line graph

#### Increasing Trend:

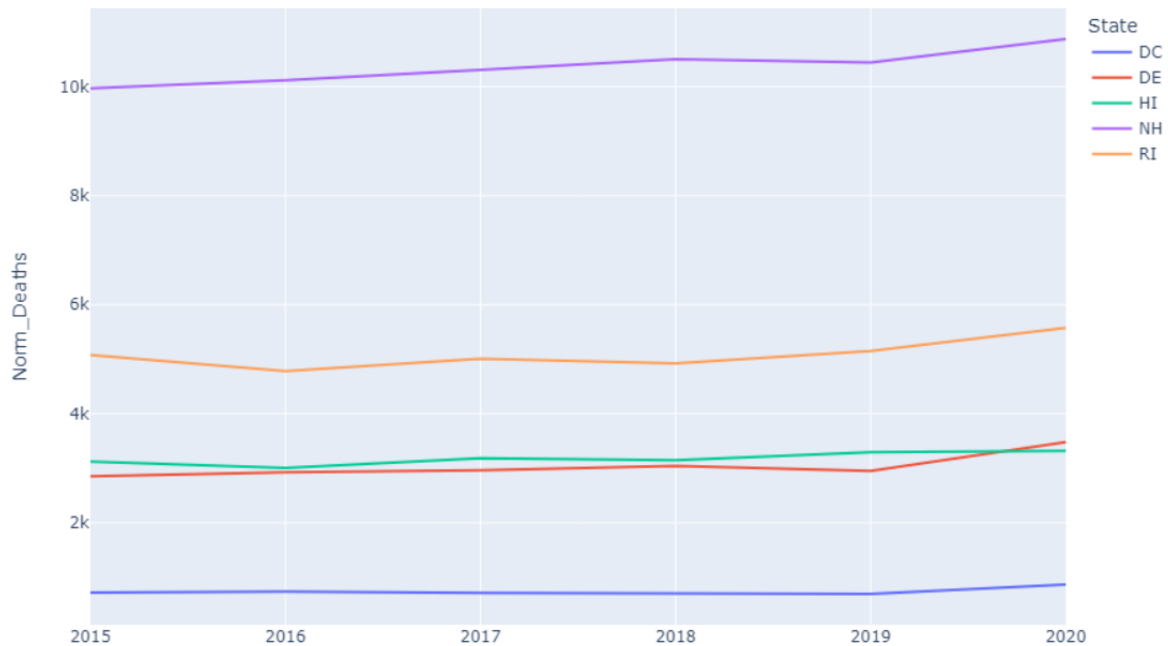


#### Inferences :

- Texas has the highest population whereas MS has the lowest population so the death rates are high in Texas as compared to other states.
- Since 2000 the use of opioid has increased in the US , but the reports say that in Texas, from 2015 there was an increase in female deaths compared to male deaths.

- So from 2015 to 2019 the deaths were not increasing they were almost consistent but in 2019 there was a slight increase in the deaths.

## Decreasing Trend:



## Inferences:

- Washington DC had the highest decrease in death rate over the period of 5 years and then the states i.e DE, HI , RI and then NH were the states in the decreasing number of Norm\_Deaths.
- The above states had special programs granted by the US government such as Prescription Drug Monitoring Programs (PDMPs) which helps in reducing the rate of opioid prescription.
- DC has the lowest population compared to the other 4 states which results in the lowest number of opioid deaths.

## Task 4: Plot a scatter plot graph of normalized mortality by state with respect to the log of the population.

- For this task, we have taken the opioid data from 1990's to 2020.

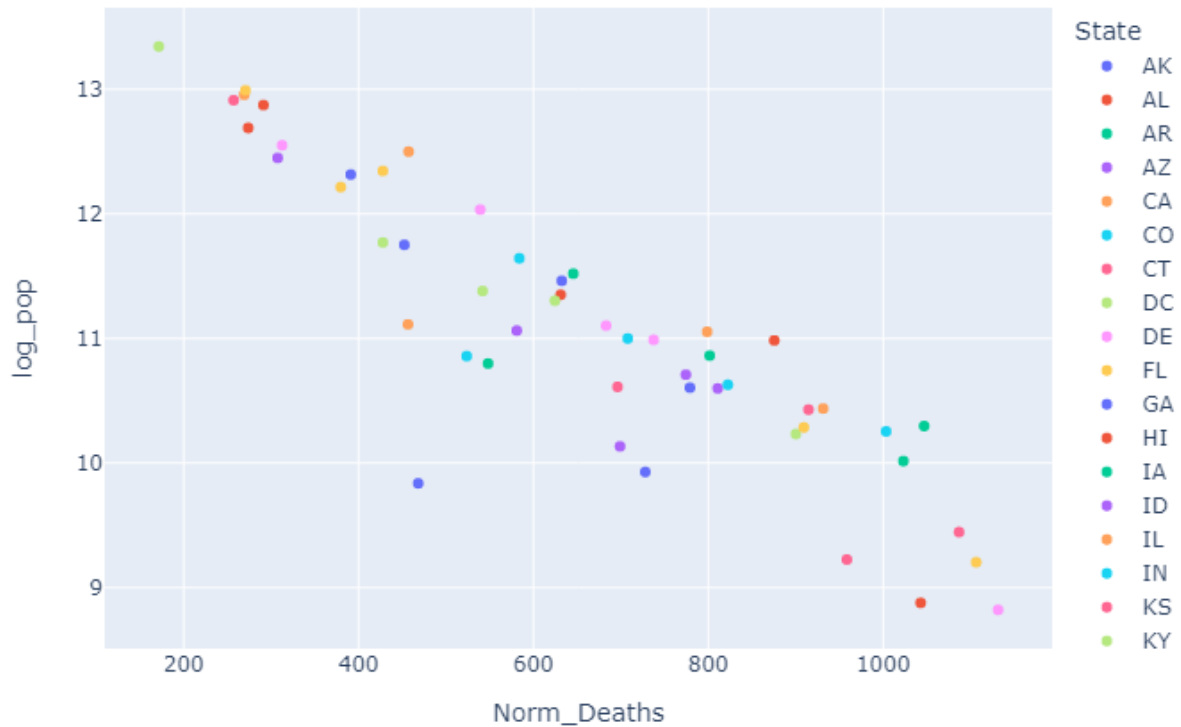
- Initially, determining the mortality rate by dividing the death value by population per 100000. Taking the natural log of population value reduces the large values of population typically in millions to 100,000.
- By grouping the data by county and plotting the total sum of Norm\_Deaths and Log of population in a scatter plot we get:

Scatter plot of Norm\_Deaths vs. Log of Population



Here, we have color coded the dots by state.

- The plot shows an inverse kind of relationship between log\_pop and Norm\_Deaths. It shows a kind of exponentially decreasing relation for all counties as norm death value increases.
- The dots are color coded by States and most of the cases are in the range of norm death value between 700 to 1400 with log population value between 8 to 10. Similarly grouping the data by state and plotting the scatter plot we get:



- The plot shows the values of norm\_deaths values of 50 states given the log of population.
- From the plot, we can see that there exists a negative relationship between the state mortality rate and log of population. It is in sync with the above county based plot.
- The plot also shows that there exists a linear relationship between norm\_deaths and log\_pop.