



Clustering and Data Fitting

Name: Sania Rashid

Student ID: 22030629

Github Repository Link: <https://github.com/srashid95/Clustering-Assignment>

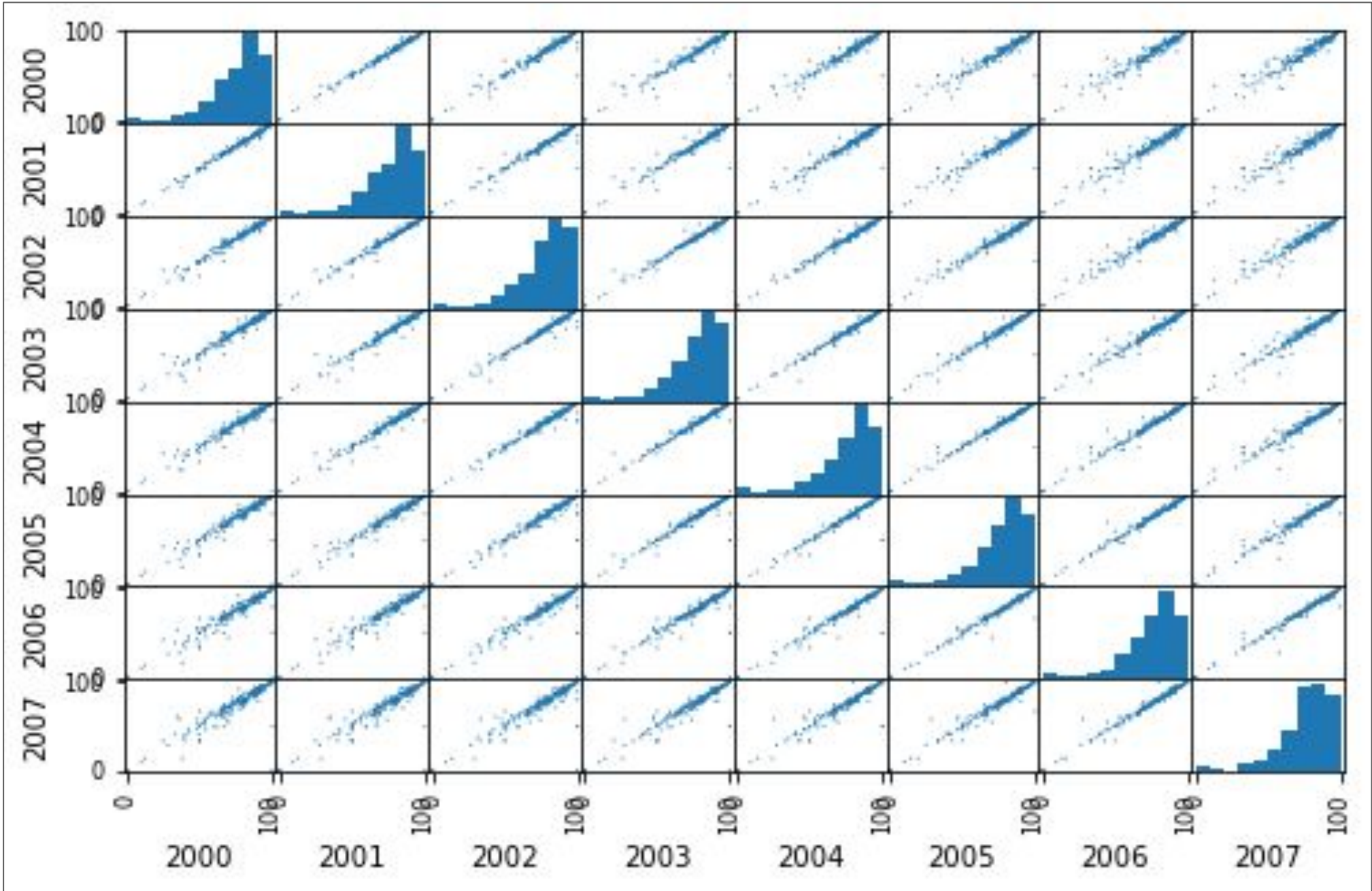
Abstract

Since last 40 to 50 years, the death rate has been decreased especially in the youth of ages between 20 to 24 years. I decided to research on factors and indicators that can show some positive results in this ratio. I found the death rate data from world bank website and two other factors named agricultural nitrous oxide emission and the access of clean fuel in the population (by percentage). After working on it, I observed that both of the factors affect the probability of death of youth from age 20 to 24 positively. The death rate is inversely proportional to these both indicators.

Youth death rate associated with agricultural nitrous oxide and clean fuel access

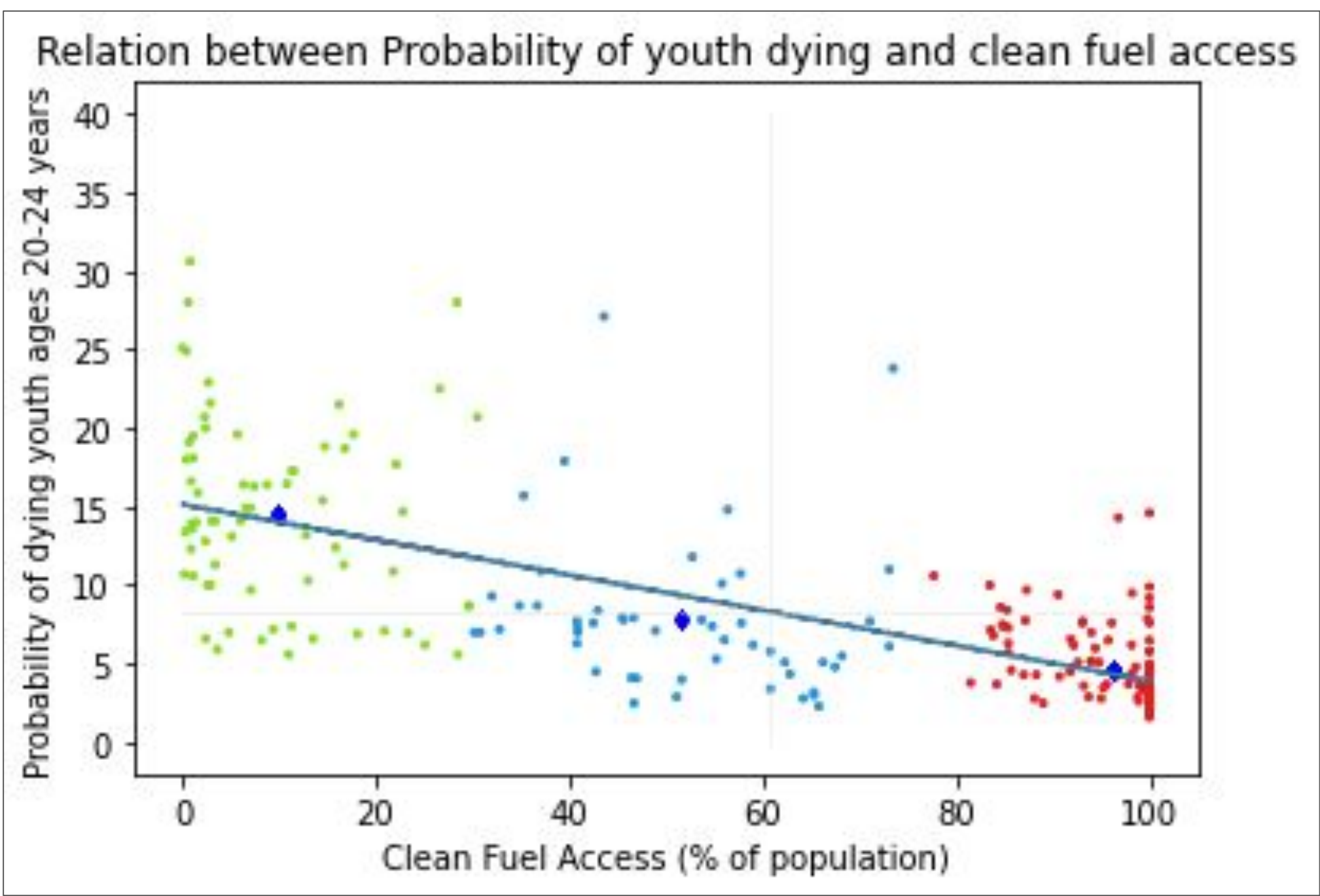
Summary

- During last few years, death rate of the youth has incredibly decreased and it shows a downward trend in current time series.
- I was curious to study the reasons behind this downward trend for death rate in youth.
- I got the data from world bank website about youth death rate and some other indicators that I think can help.

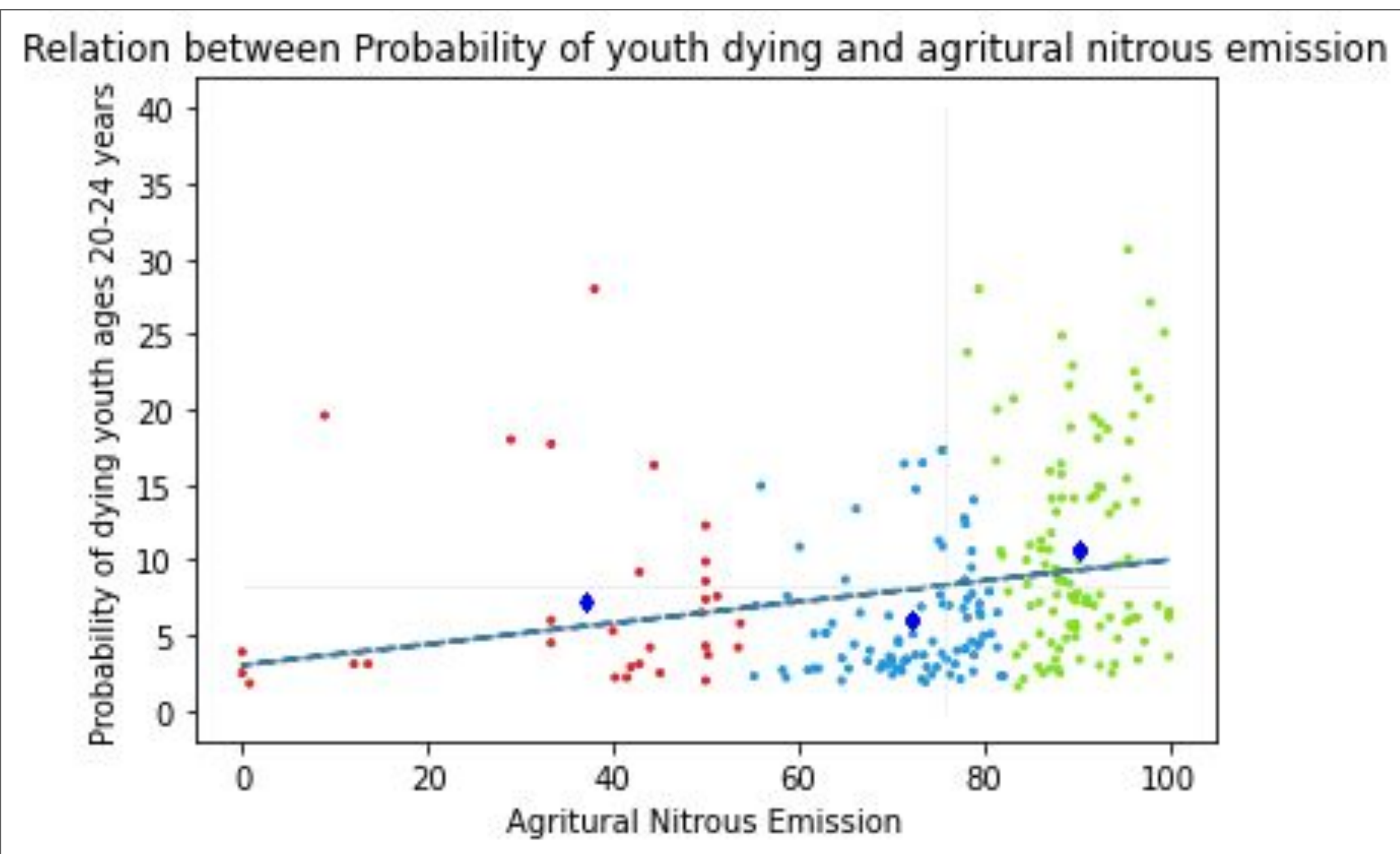


- The preceding scatter matrix shows the data for agricultural nitrous gas emission and its relation between the years from 2000 to 2007.

Introduction

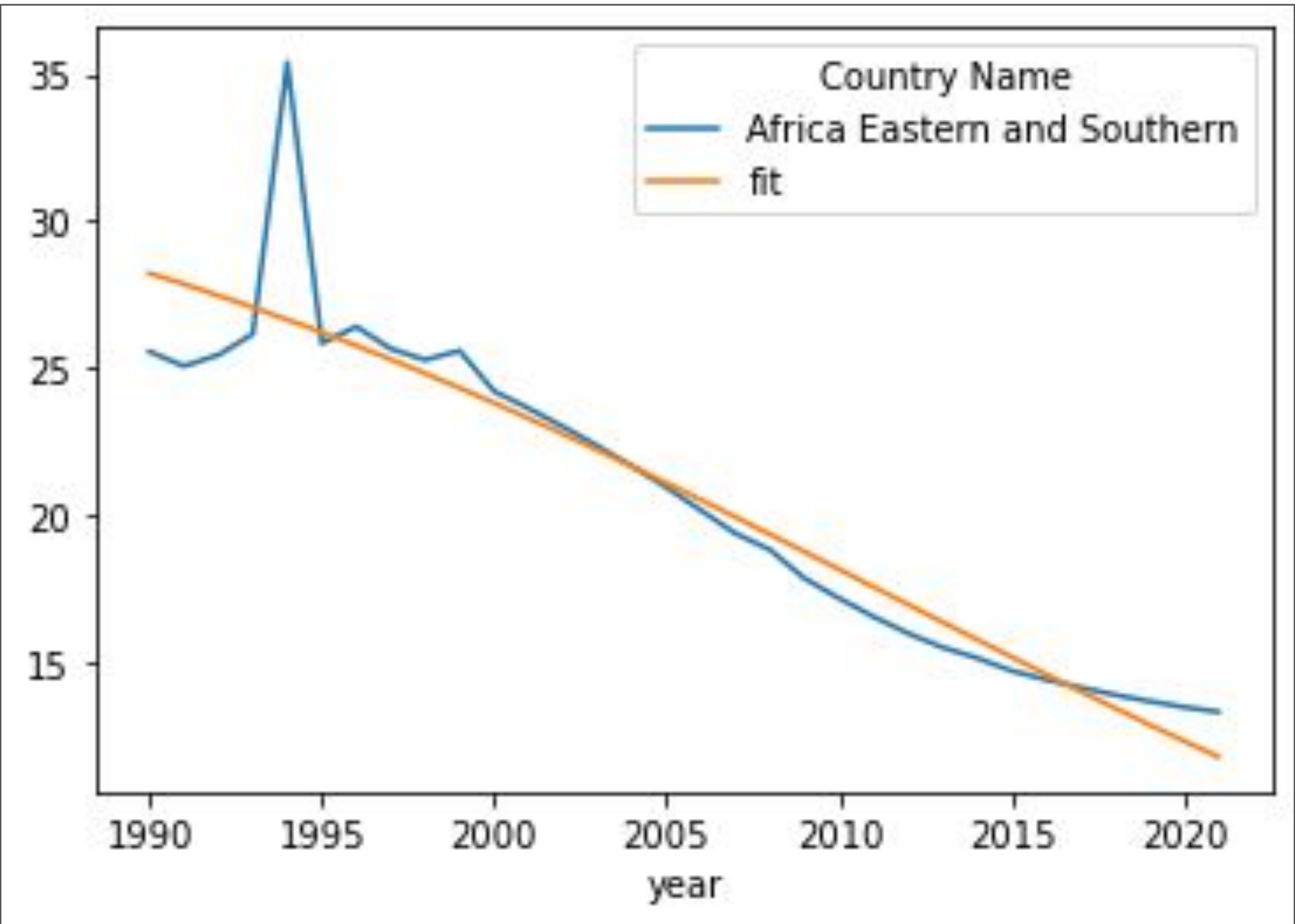


- I have created some clusters using scatter plot.
- In the preceding plot, we can observe that there is a good relation between clean fuel access and probability of dying youth between the age 20 to 24 years.
- These both indicators are inversely proportional to each other. More the access to clean fuel, less the probability of dying youth between the age 20 to 24 years.



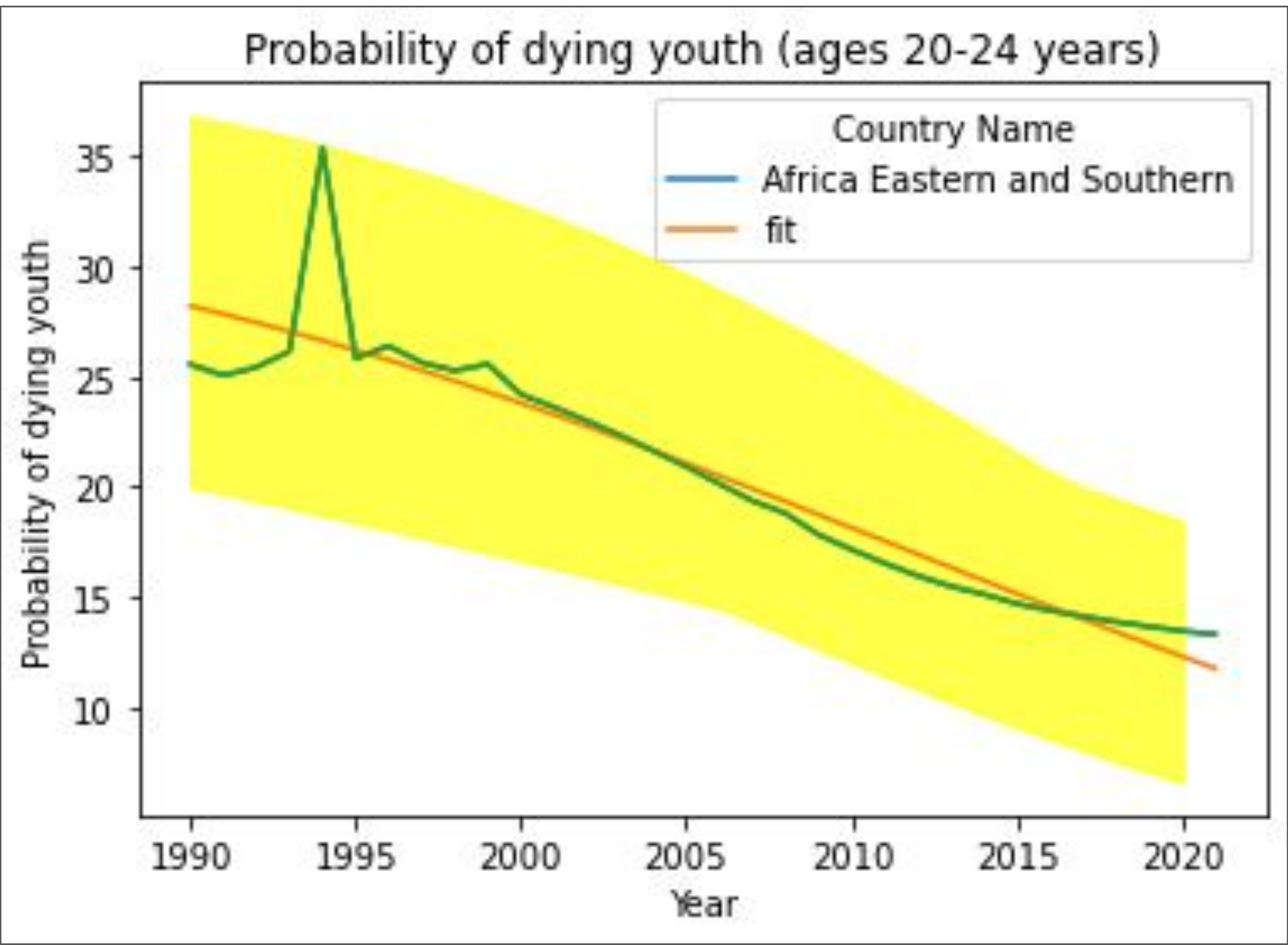
Data Fitting

- In the following image, I tried to fit the data for probability of dying youth from ages 20 to 24 years.



Error Boundries

- The preceding plot displays the fitting of data for probability of death rate of youth.
- That explains good, but to make it better, we need to draw the error boundaries against the fitted line.



Conclusion

- After observing the output of the above plots, we can conclude that probability of dying youth is inversely proportional to clean fuel access and agricultural nitrous gas emission.