

STATISTICAL ANALYSIS OF CRIMES IN MALAYSIA BASED ON ARTICLE AND EXTENDED WORK

Anirudh Aman Arun Praveen

Brief introduction to Malaysia and India data

Initially we have crime data of Malaysia in excel format and the sheet in excel where divided by state(include federal territory) wise and first page was Malaysia data. Column in excel where represent data for particular type of data, example murder,rape & etc And the row where represent data year wise from 2006 to 2017

So we collected similar type of data for India from 2001 to 2016 and state wise, from <https://ncrb.gov.in/> Collected population data of Malaysia (from <https://www.dosm.gov.my/> i.e department of statistics Malaysia)and India(from Wikipedia) (population data not collected for the states)

(all the files we collected are in the drive)

R code

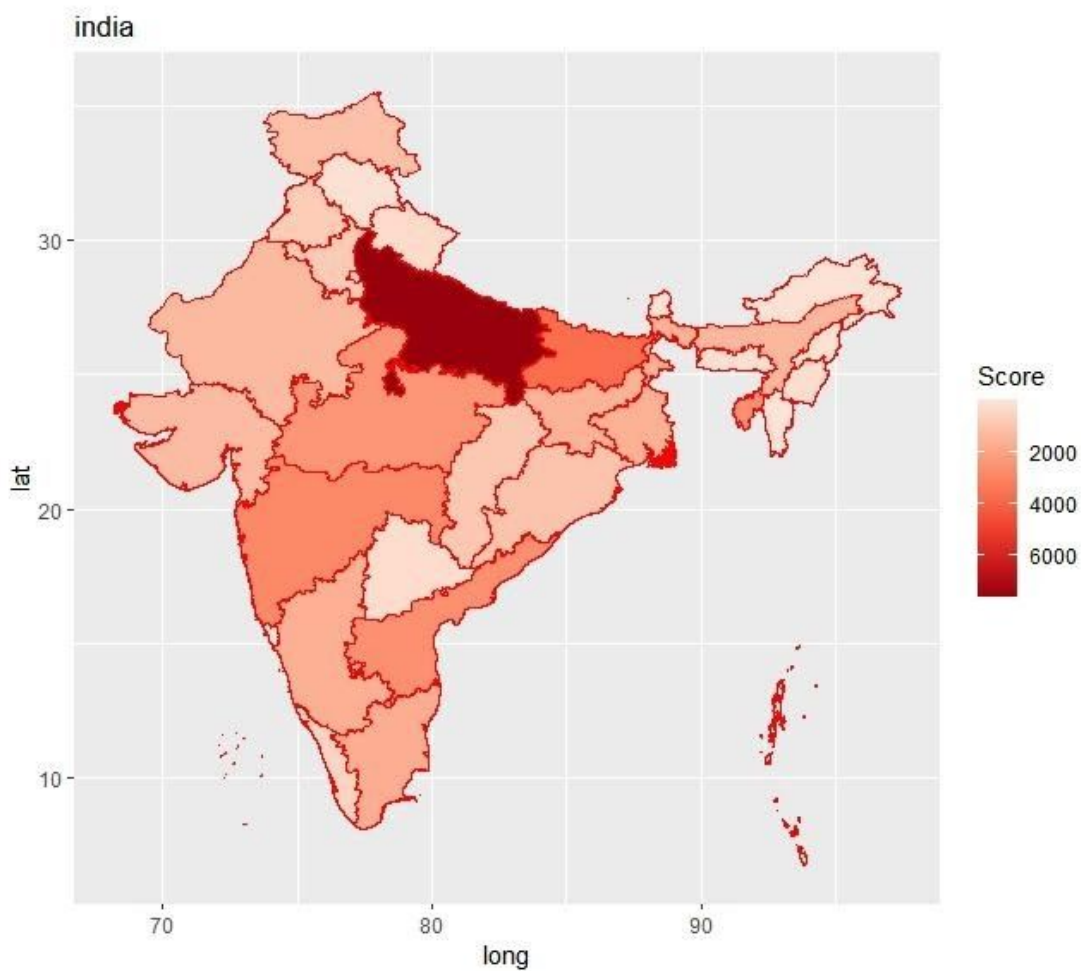
Since we had state wise data, so decided to plot map of India and Malaysia, colour the state by red such a way that dark red represent high number of data and light red represent low number of data. Sample of map is shown below.

So how to plot map in R?

We took help from stack overflow and studied how to plot map and colour them according to number and to plot map we need shape file of India and Malaysia. We collected shape file from www.gadm.org both Malaysia and India

And we made video of these plot and presented in PPT

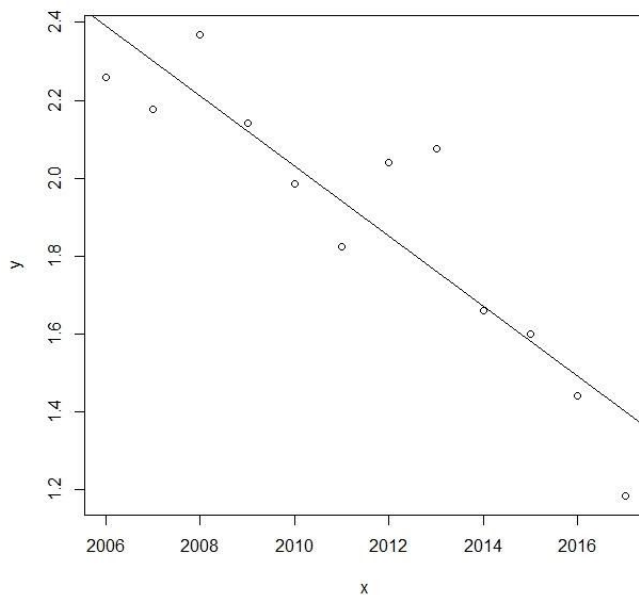
Packages used are -> maptools, rgeos, ggplot2 and rgdal



Here is brief note on what we did in R code

We read the excel file and converted into 3 dimensional array for Malaysia array is "mat" and for India array is "mati". $mat[i,j,k]$ i represents the year j represent the particular type of crime and k represent the year, similarly for India

We re plotted type of crime v/s year for Malaysia that are already present in pdf. Then we created function **linearreg** that take two vector of same length and plot them with best fitted line (slope and intercept are calculated by method that discussed in class) and studied all possible plot with x axis represent year. The best fitted one are presented in PPT (also one showed below)



The function also return coefficient of determination etc
 Then we studied linear regression for crime rate (i.e number of crime per one lakh people) here we used population data,best one are in PPT.

Statistical out put and inference made in PPT

(Only some parts are included)

As observed from map plots of crimes in Malaysia it's obvious that Selangor (State) has most crime records over the categories of the data taken from 2006-17. After looking over this data we searched for the recent works of Selangor police and we observed that a total of 19,810 cases were recorded in 2019 compared to 21,338 cases in 2018.

And least Crime recorded State/Federal Territory -

As observed from the plots, Pahang(State) remains one of the safest states in Malaysia (2006-2017). Even in recent years it recorded a 8.7 per cent drop in its crime index rate from 3,582 cases in 2018 to 3,269 cases in 2019 according to data provided by Pahang Police Chief.

Inference from Map plots of India

Most Crime recorded State/Union Territory

Under the categories considered for crimes in India we clearly observe that Uttar Pradesh has most recorded crimes. On a further look to recent data we found that there was a 74.50% fall in cases of robbery in 2020 as compared to 2016.

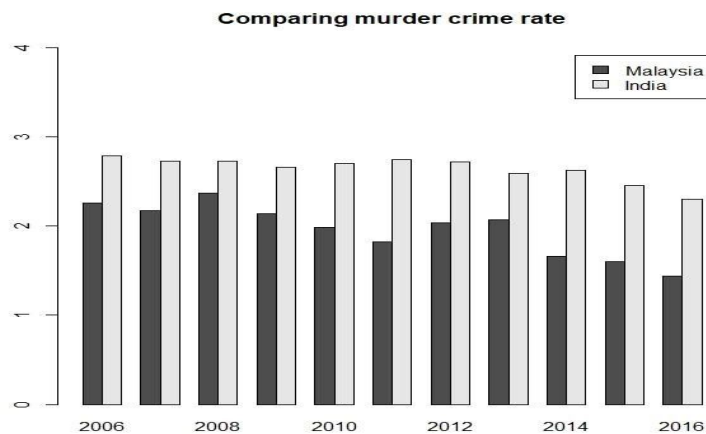
Some bias in data obtained for India:-

1. Bias on crime records in villages
2. Report expose bias in minds of cops against Muslims

(SEVERAL STUDIES HAVE, OVER THE YEARS, SHOWED THAT A HIGHER PROPORTION OF MUSLIMS—RELATIVE TO THEIR POPULATION-SHARE—TEND TO GET ARRESTED THAN MOST OTHER RELIGIOUS/CASTE GROUPS. FOR EXAMPLE IN NOVEMBER 2016, THE INDIAN EXPRESS REPORTED THAT, AT AN ALL-INDIA LEVEL, 15.8% OF ALL CONVICTS WERE MUSLIMS AS COMPARED TO THEIR POPULATION SHARE OF 14.2%.

WE CAN INCLUDE MANY MORE BIAS TO HOW THE CRIMES ARE RECORDED SUCH AS THE SOCIAL STATUS OF THE OFFENDER, HATE CRIME ETC.)

Comparison of murder crime rate

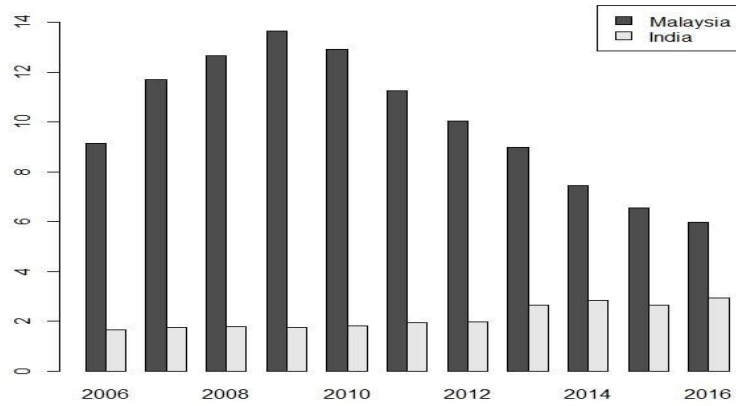


After looking at the double bar plot we observe:-

1. India has the higher murder rate in every year under consideration.
2. Malaysia seems to have its rate reduced more in comparison to India.

Comparison of rape crime rate

1. Malaysia has the higher crime rate in every year under consideration.
2. Malaysia's plot shows right skewness and hence rate decreases but in India the rates seems to increase over the years considered.



Now lets see some questions:

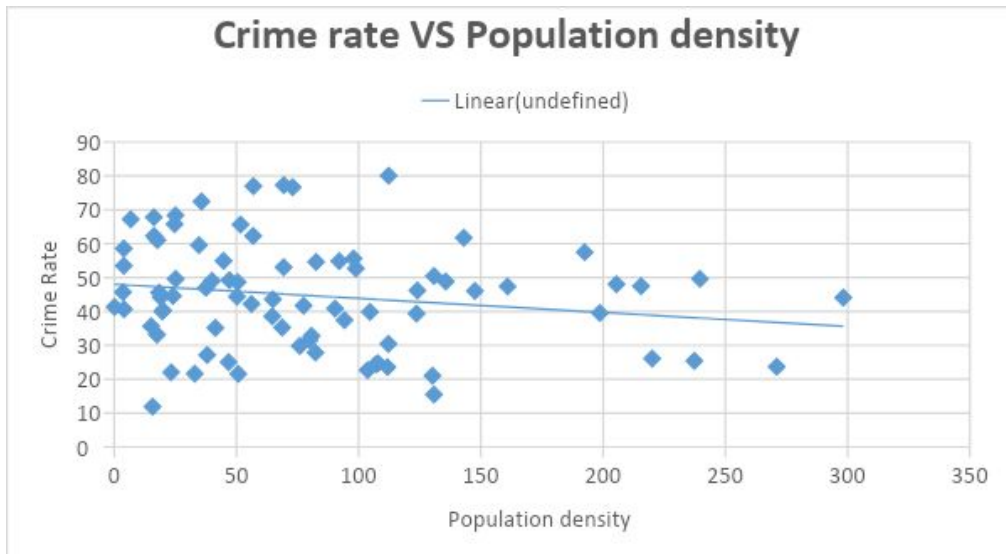
Will Improving **education index** reduce **Crime rates**?

Does **crime rate** has something to do with **population density**?

How **unemployment** rate effects **crime rate**?

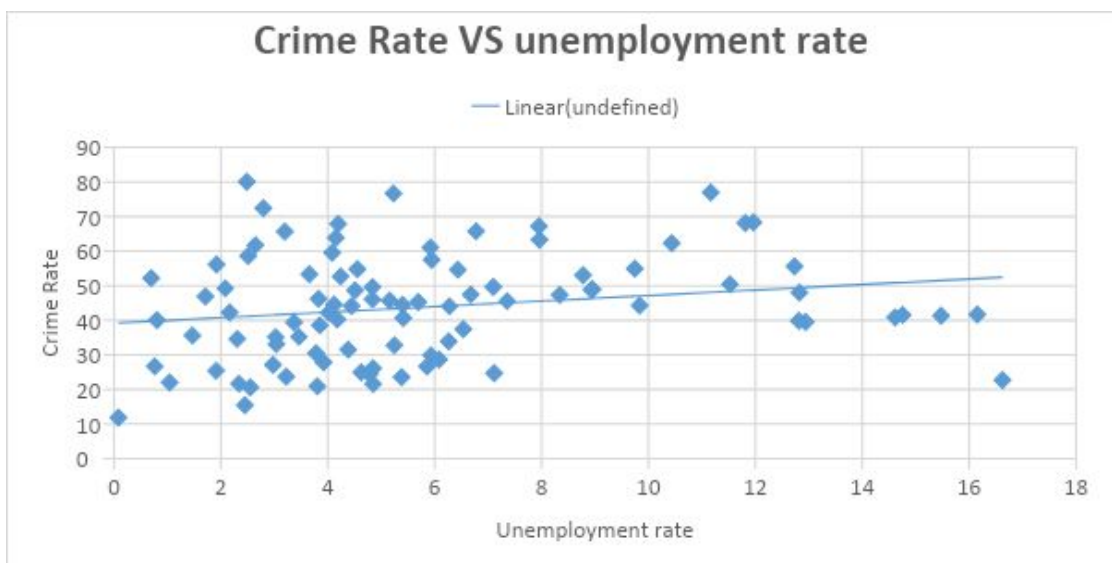
How **GDP per capita** and **life expectancy** effect **Crime rate**?

To answer these , we collected data of countries of world
 We removed some outliers, which were making the plots
 skewed



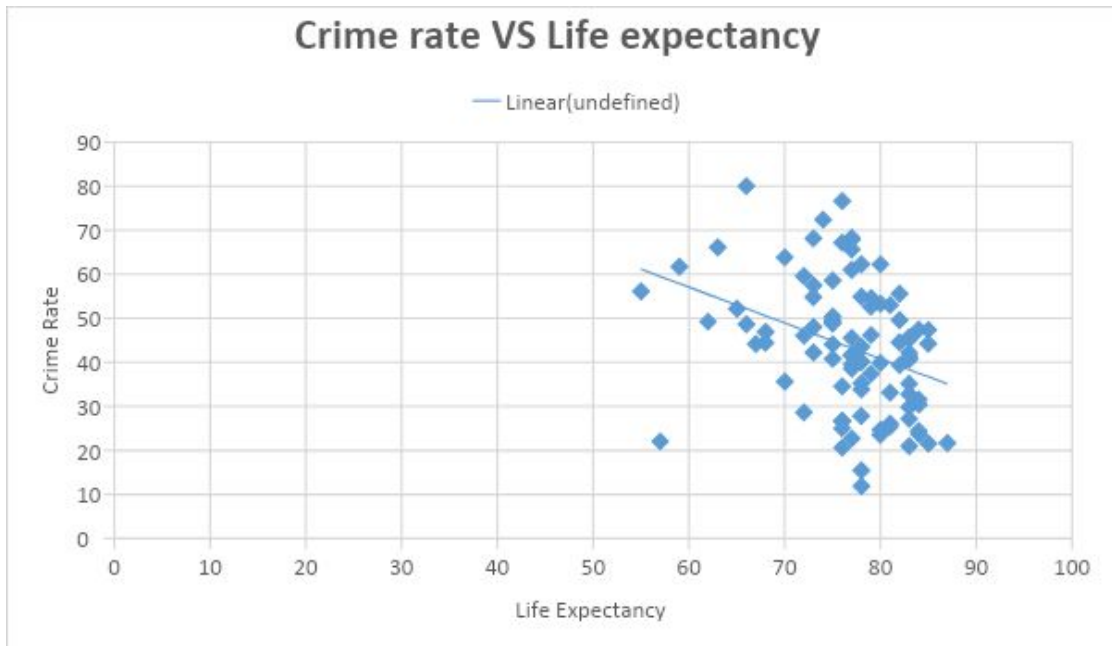
The regression line is a bad estimate of the data points as Coefficient of correlation is very small (0.0319)

The slope is slightly negative, so it kind of looks that with increase in population density crime rate is slowly decreasing

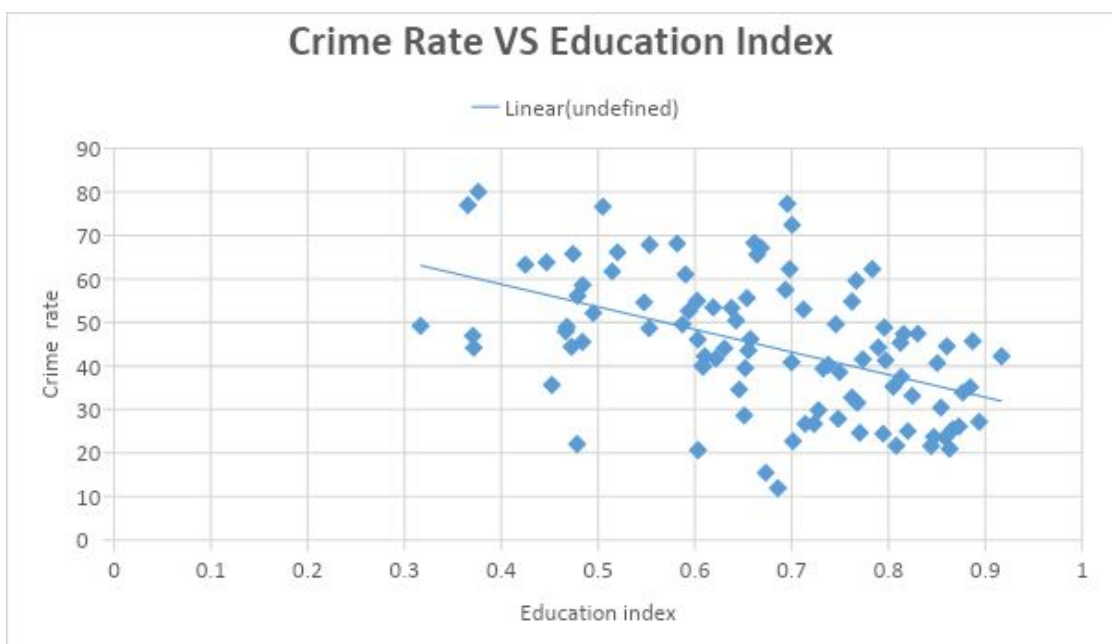


Again the regression line is a bad estimate of the data points as Coefficient of correlation is very small (0.0398)

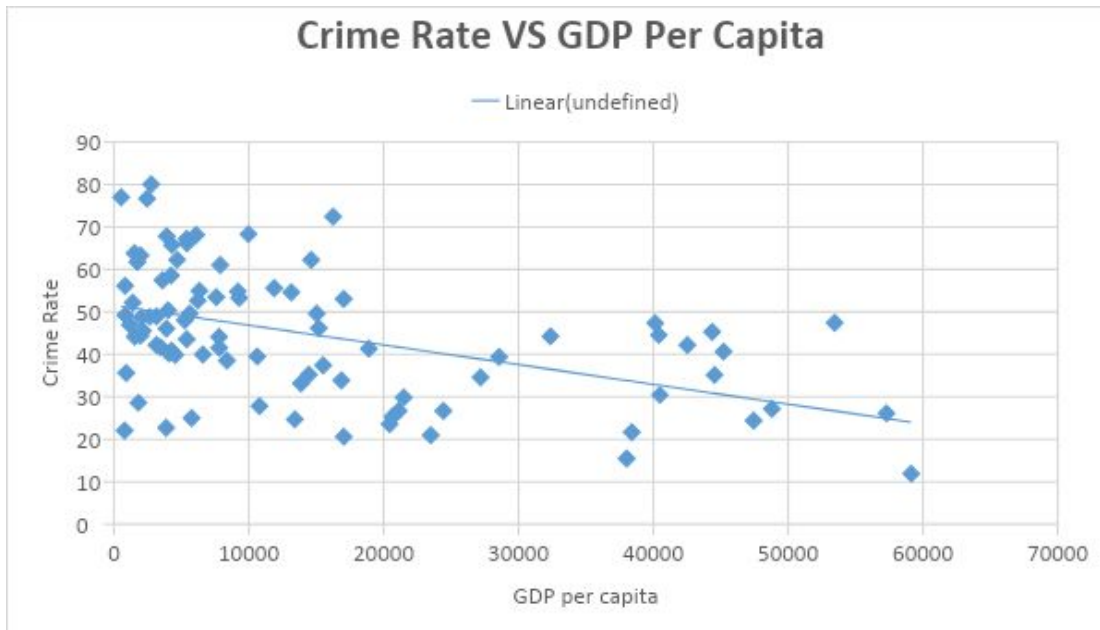
The slope is positive, so it kind of looks that with increase in unemployment rate crime rate is increasing which was expected



The regression line is a bad estimate of the data points as Coefficient of correlation is very small (0.1214)
 The slope is negative, so it kind of looks that with increase in life expectancy crime rate is decreasing



The regression line is a quite a good estimate of the data points as Coefficient of correlation is sufficiently big at least compared to previous three plots(0.2388)
 The slope is negative, so it looks like that with increase in Education index crime rate is decreasing



Again here the regression line is a quite a good estimate of the data points as Coefficient of correlation is sufficiently big at least compared to first three plots(0.2235)
The slope is negative, so it looks like that with increase in GDP per capita crime rate is decreasing