**ABSTRACT**

This study delves into the intricate dynamics of murder rates spanning the years 2015 to 2021, focusing on discerning trends, spatial patterns, and their correlation with socioeconomic factors. Utilizing a dataset encompassing these years, our analysis employs statistical methodologies to unearth meaningful insights.

We first scrutinize temporal trends, identifying fluctuations and patterns in murder rates over the specified period. By comparing these findings with historical data from the preceding decade, we offer a comprehensive understanding of the evolving landscape of lethal crimes.

Spatial analysis plays a pivotal role in our study, as we investigate geographical variations in murder rates. Through advanced mapping techniques, we delineate hotspots and coldspots, shedding light on the spatial distribution of violent incidents.

Furthermore, our research extends beyond mere descriptive analysis to uncover underlying correlations between murder rates and socioeconomic factors. By integrating data on poverty, education levels, and other pertinent metrics, we scrutinize the interplay between social conditions and crime prevalence. Through regression analysis and correlation studies, we aim to identify significant associations, providing valuable insights for policymakers and stakeholders.

**Introduction**

Crime is as old as human history itself. It has evolved into different forms, saw variations in it through cultural, temporal, and socio-economic changes. Through all this, the action of crime still persists in our society. Among all the crimes human commits, the crime of murder seems to be the most arrogant. Life is a precious gift to behold, regardless of its struggles, so there is little reasoning behind murder. It is a devastating act that not only ends a life but also leaves a lasting impact on families, communities, and society as a whole. Understanding the motives behind murders is crucial in addressing this serious issue and working towards creating safer communities.

This project delves into the motives behind murders in India over the past 7 years. By analyzing available data on murder cases and their motives, we aim to uncover insights that can inform strategies for crime prevention and intervention. Murders often stem from a variety of factors, including personal disputes, socio-economic inequalities, and systemic issues. By studying these motives, we hope to gain a deeper understanding of the root causes of violent crime and identify ways to address them effectively.

Our analysis will explore patterns and trends in murder motives, examining how they have evolved over time and vary across different regions and demographic groups. By examining the underlying factors contributing to homicides, we aim to provide actionable insights for law enforcement agencies, policymakers, and community organizations. Ultimately, our goal is to contribute to efforts aimed at reducing violent crime and promoting safety and well-being in Indian society.

Analysis on the murder counts in

Years (2015-2021)

We have the data of total murder counts of the year 2015-2021, so we look to plot them taking the years in the abscissa and the murder counts in the ordinate axis. We obtain the plot in Fig 1.1

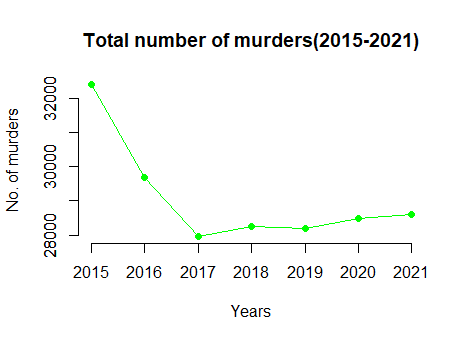


Fig 1.0 The time series plot of the murder counts

Through a cursory glance we see that the murder counts have seen a sharp decline in the period of 2015-2017 and saw its minimum at 2017. After 2017, the murder counts picked up but at a more humble rate than its previous decline in the former mentioned period. Now, we look to find a trend line to this time series plot.

**1.1 Fitting a trend line**

The most common trend line fitting is the linear one but the graph hints us to model it with a less common trend model but still a famous one. It would be really intuitive to model the trend of the murder counts of the period 2017-2021 with an exponential trend model to be specific the inverse exponential trend model. The trend model fitting of the linear type, the exponential type are shown in Fig 1.1.1.

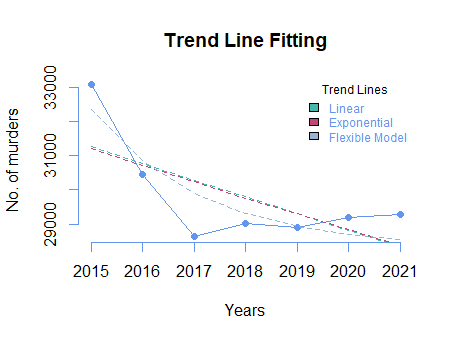


Fig 1.1.1 Fitting the linear, exponential and the inverse exponential trend models

We see that the linear trend line tries to give a representation of the trend but is not that useful, the exponential doesn’t improve the situation either. Now a more flexible model is used to capture the underlying trend. The form of the fitted model is provided below:

The estimates of and are calculated using the least squares method. The estimates along with the actual fitted equation is given below:

= 27580 and = 1012

So the model becomes,

Here the years variable used is the standardized version so as to obtain a non-zero estimate of . This model, however, quite efficient in capturing the trend has a huge downside to it, it is not interpretable. The coefficients in the model and doesn’t have an interpretable meaning. This is an example of the trade-off between prediction accuracy and model interpretability.

The linear trend equation is

And the exponential trend equation is

Keeping in mind the previously mentioned trade-off, we use the linear trend to interpret the trend of the murder counts. The linear fit gives us a decrease of about 488 murder counts as we go 1 year into the future from 2015. The p-value for the test is 0.09, so we can accept the alternate hypothesis of being non-zero at 10% level of significance based on the data provided. So murder counts have seen a yearly decline of 488 from 2015-2021 and we conclude this with sufficient statistical evidence.

**1.2 Spatial Analysis**

As we move deeper into the study, we look to bring out some spatial patterns or occurrences of murder in the period. Now we look to get a perspective of the murders in terms of their spatial occurrences. We take data on the state wise murder counts and look to dissect some information from it. We start by looking at a bar chart of the top 6 states in terms of murder counts in the period 2015-2021. The bar chart is provided in the Fig 1.2.1.

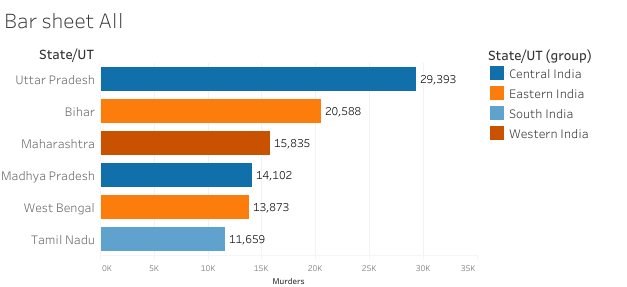


Fig 1.2.1 Top 6 States in terms of murder counts in 2015-2021

We also divided the states into six administrative blocks to get an idea of the location and culture of that state. The blocks are namely – Central, Eastern, Southern, Northern, North-Eastern and Western. We see that Uttar Pradesh has the most counts of murders in the period. There are two representatives of Central and Eastern India, and none from Northern and North-Eastern. The map with state wise murder counts is provided in Fig 1.2.2.

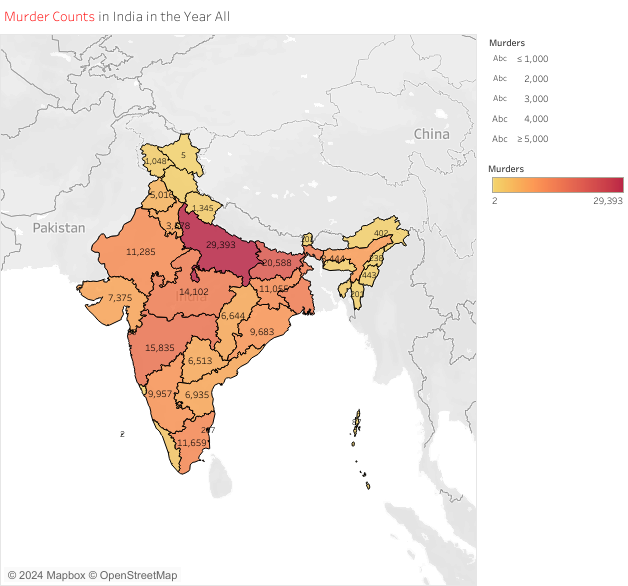


Fig 1.2.2 Map showing the state wise murder counts in 2015-2021

Through a cursory glance it might seem that states in Central and Eastern India are the most prone to murders in the period. But we are taking into account a big information, which can benefit our analysis. We must take into consideration the population of each state so as to obtain a much fairer perspective on the murder counts. As it is very logical that a state with high population is quite likely to have more murders. It doesn’t imply that it is a dangerous state. We have adjusted the murder counts with the population and created a new calculated field, namely,

The data on the population is collected from the crime report on India 2021. The plot so obtained is given in Fig 1.2.3.

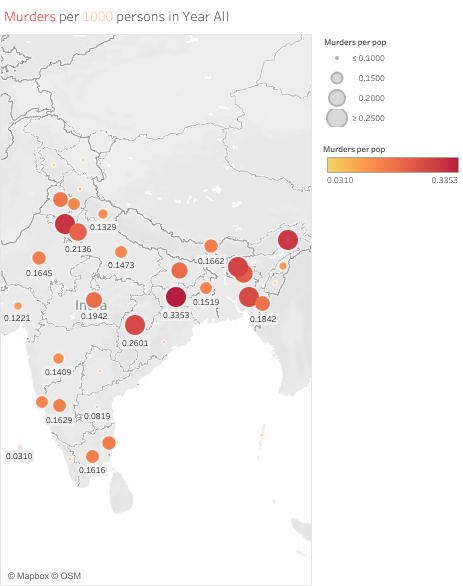


Fig 1.2.3 Map showing murders per 1000 people in 2015-2021

We see a very different scenario here after adjusting for population in our murder totals. Through visual inspection we see that there is a higher incidence of murder in the north eastern side of our country along with a few states in Central and Northern India. Let’s look closer at the top 6 states with worst murder incidence per population. The bar graph is given in Fig 1.2.4. At a closer look we see that there are 3 states from the North Eastern section of our country and two from Northern and one from Central and Eastern India each. This is quite a stark contrast from the aggregate murder counts. Now a question will naturally arise, is the murder incidence higher in the North Eastern area than in the rest of India.

So to solve this query we make use of statistical inference to find an answer with adequate statistical backing to the question.

CLAIM: Is murder per population is higher in North East India than in the rest of India?

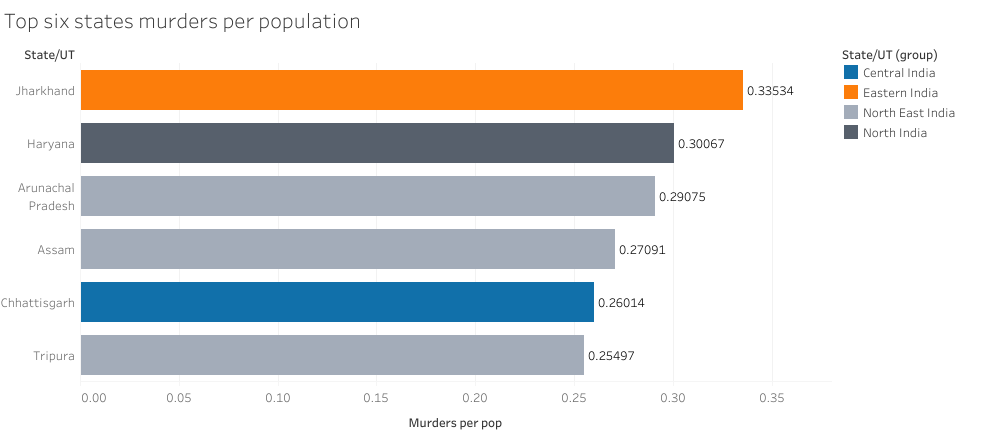


Fig 1.2.4 Top 6 states with highest murders per 1000 person

FORMULATION:

Now first we have to define properly by what we mean by “North East India”. Fig 1.2.5 gives the proper partition of the states of India into the two groups.

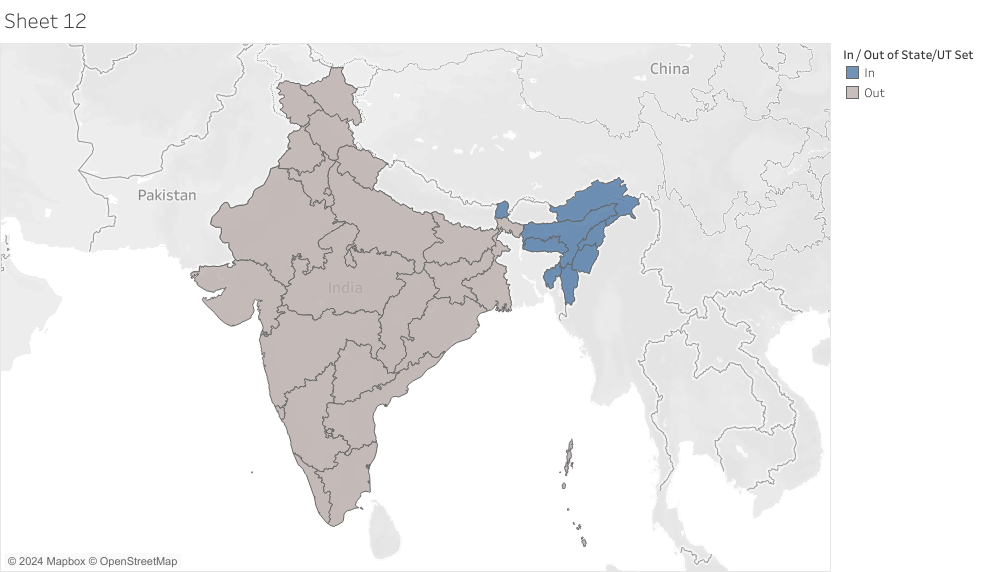


Fig 1.2.5 "North East and its neighbors" is in blue

Let us call this region NEI (North East India). Now we need to test whether the proportion of murder is higher in NEN than the Rest of India (ROI).

We do a binomial test for testing this claim. Let the murder proportion in NEI be defined by PNEI and the proportion for ROI is PROI . Let’s define our null hypothesis and the alternate.

Null Hypothesis :

H0 : PNEI = PROI

and the alternate is

H1 : PNEI > PROI

So this is a right tailed test. Now we apply two sample binomial test.

The test statistic for this test is

Z = ,

Here n1 = population of NEI region, n2 = population of ROI region.

where p is the proportion of successes for the combined sample and which under null follows standard normal.

We test this at 0.05 level of significance and calculate the p-value.

p-value = P(Z > Z0),

where Z0 is the observed value of the test statistic.

We reject null hypothesis if p-value < 0.05 in favour of the alternate.

The observed z-statistic value is 44.23 and the p-value of this test is approximately 0 . So we conclude that based on the data we have the murder proportion in the NEI region is greater than that of ROI region at 2.5% level of significance. This null hypothesis will be accepted at any value of α in (0,100).

Through visual inspection it might seem that the NEI region has a higher incidence of murder than ROI, and there is overwhelming statistical evidence to back that claim.

Taking the same route of action, we now try to divide the States and Union Territories of India into six administrative blocks, namely, Central, Western, Eastern, Northern, Southern and North Eastern. This classification is done on cultural and geographical factors. The given classification is visually represented in the Fig 1.2.6.

Now we look to find some patterns in the various regions of India on murder counts and rates.

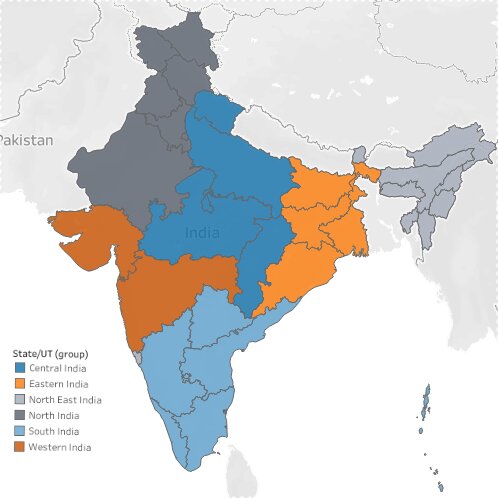


Fig 1.2.6 Six Administrative Blocks of India

The division is done based on the already made six blocks of administration. The states and UTs are divided in the following way:

**Central :** Chhattisgarh, Madhya Pradesh, Uttarakhand, Uttar Pradesh.

**Eastern** : West Bengal, Jharkhand, Odisha, and Bihar.

**North Eastern** :  Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim

**North :** Jammu & Kashmir, Ladakh, Chandigarh, Delhi, Haryana, Punjab, Rajasthan, Himachal Pradesh.

**South :** Puducherry, Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telangana.

**Western :** Maharashtra, Dadra and Nagar Haveli and Daman and Diu, Goa, Gujarat.

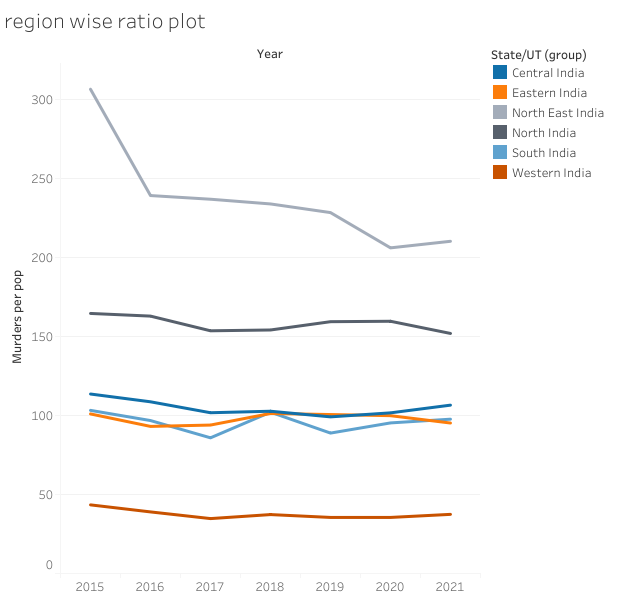


Fig 1.2.7 Plot of murders per million in blocks of India

Now the first task is to plot the murder counts for years 2015-2021 for these six administrative blocks of India.

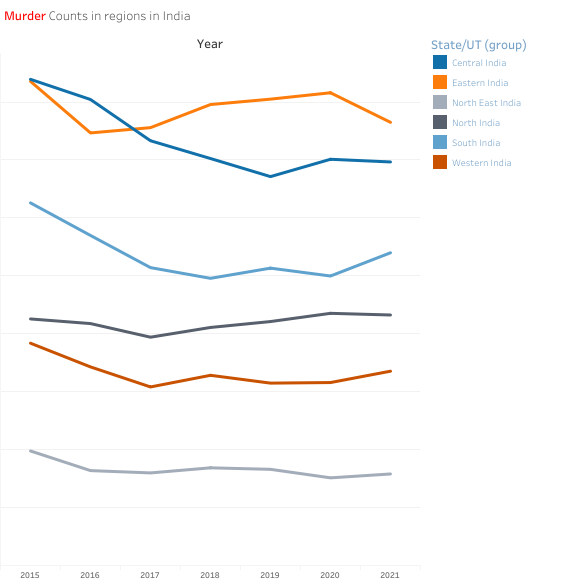


Fig 1.2.8 Murder Counts in the six regions in 2015-2021

Eastern India seems to have seen a sudden surge in murder counts from 2016, which has seen some decline after 2020. Central India has seen a gradual decrease in murder counts over the period. South India shows similar pattern as does central. North India has seen a steady increase in murder counts. Eastern India also has seen a similar steady increase.

We pose the following questions from this plot :

1. Is the murder counts in Central and Southern India correlated?
2. Is the increase of murder counts significant for Northern India?

CLAIM: There is a correlation between murder counts in Central and Southern India.

FORMULATION:

We look to test the correlation between crimes in Central and Southern India in the years 2015-2021. Here we take an assumption that the occurrences of murders in different years are independent to each other.

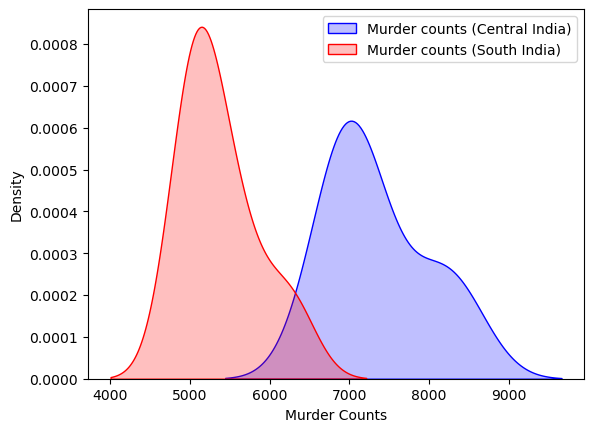
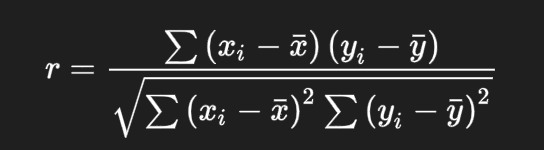


Fig 1.2.9 KDE Plot of murder counts in Central and Southern India

Let’s take a look at the kernel density plot of the two regions to get an idea of the underlying distribution. If the distribution is normal or close to normal we will apply the parametric Pearson correlation coefficient test otherwise we will apply the non-parametric Spearman rank order correlation test. The plot is given in Fig 1.2.8.

From this plot we see that the distributions are quite close to normal, we will not lose much differentiating power of the Pearson Correlation Coefficient test.

We calculate the correlation coefficient of X (murder counts in Central India) and Y (murder counts in Southern India) using the formula:



Let the correlation coefficient between the two regions be r then our null and alternative test are

H0 : r = 0

and

H1 : r ≠ 0

respectively. This is a two-tailed test and out test statistic is

t = , which under null

follows t distribution with n-2 degrees of freedom. We take the level of significance to be 5%. Here the value of the test statistic is 0.8877 and the p-value is 0.007 < 0.05. So we reject the null hypothesis.

So we can conclude that there is a positive correlation between crimes in Central and Southern India.

CLAIM: There is a significant increase in murders in Northern India.

We look to first plot the percentage crimes in various states of North India. Some of the states with very low contributions to the total were excluded for generating graph with proper scaling. The graph is provided in Fig 1.2.9.

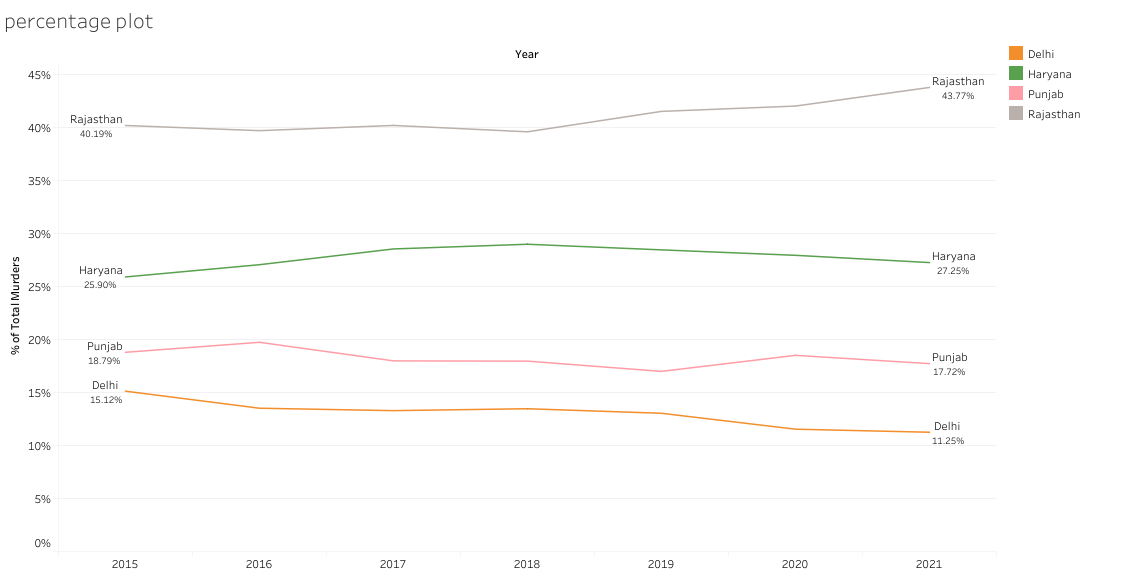
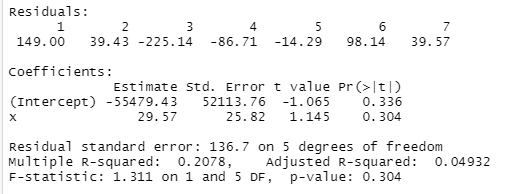


Fig 1.2.10 Percentage murders in states of North India

From the plot we see that Rajasthan is by far the most violent in terms of total murders and it has been increasing for few the past few years. The remaining states has remained almost constant. Now we check whether the increase in numbers has any statistical significance we fit a linear model to the data and test for the slope parameter. After fitting a simple linear model on the data we get the following summary:



Now as the p-value is quite high (0.304) we cannot conclude that there is significant increase in murder counts in Northern India. The trend fitting plot is given below :

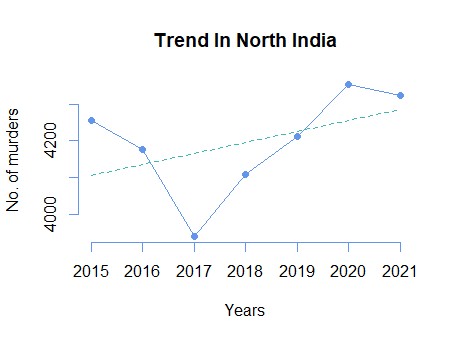


Fig 1.2.11 Trend line for North India

We see that although the hypothesis of non-zero increment is falsified, the murder counts of North India correlates not so heavily with the total India murder counts. To get a better idea of the correlations we plot a heat-map of the correlation matrix.

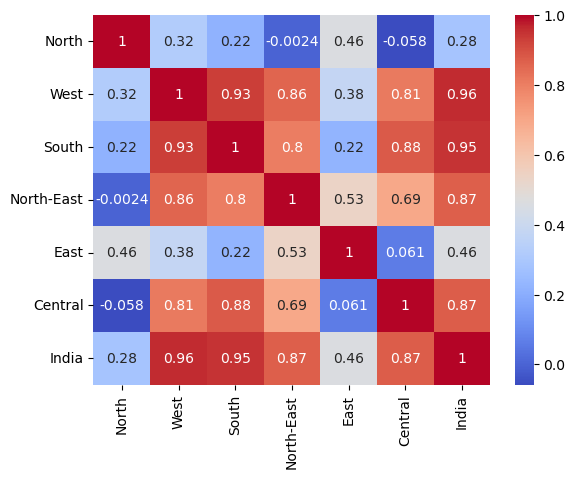


Fig 1.2.12 Heatmap of Correlation matrix

So we already have tested for the significance of correlation of crimes in Southern and Central India. Here we get a more detailed understanding of the correlations between various blocks of India. North India is the least correlated in terms of murder counts with Indian murder counts. We got a good idea about the correlations. Now let’s get an idea of the variance of the murder counts. For that reason we plot the box plots.

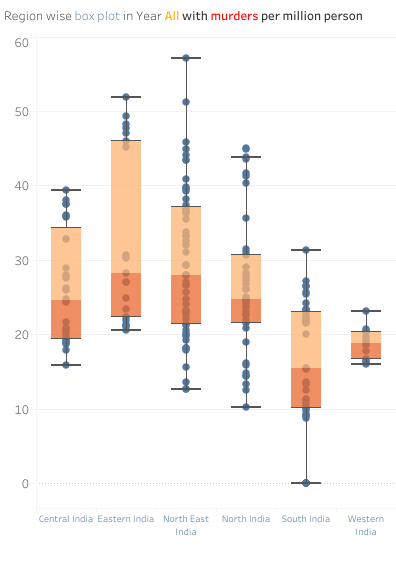
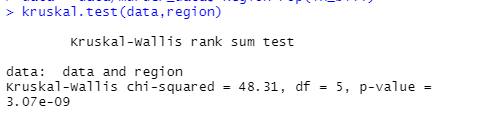
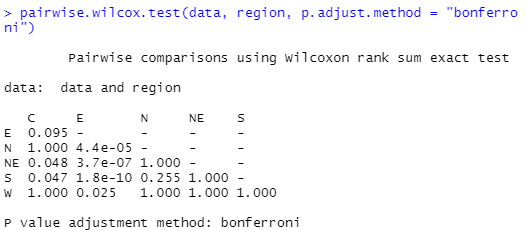


Fig 1.2.13 Region-wise box plot for murders per million

The murders per million in all the region is positively skewed. We see that North East India has the most scattered data, while western India has the least scattered data. So we can say there is more homogeneity in murder rates in the states of West India. South India has the least value of median. So South India on average is the safest region of India. We can also see that we cannot apply ANOVA testing here to check the differences in means as the groups are not homoscedastic. So we use the non-parametric equivalent of ANOVA i.e. Kruskal Wallis test. This is a test of equality of median of the populations. The test result is as follows:



We see that the p-value here is very close to 0. So we reject the null hypothesis that the median of the distributions of the different groups are same. Now for post hoc test we apply Pair-wise Wilcoxon Rank Sum Tests to see the pair wise differences and their statistical significance. The result is as follows:



As we see here NE is different from C, E. N is different from E and S is different from C. W is different from E. The parts, which show 1 as p-value, are those columns with ties so the p-value cannot be computed exactly.

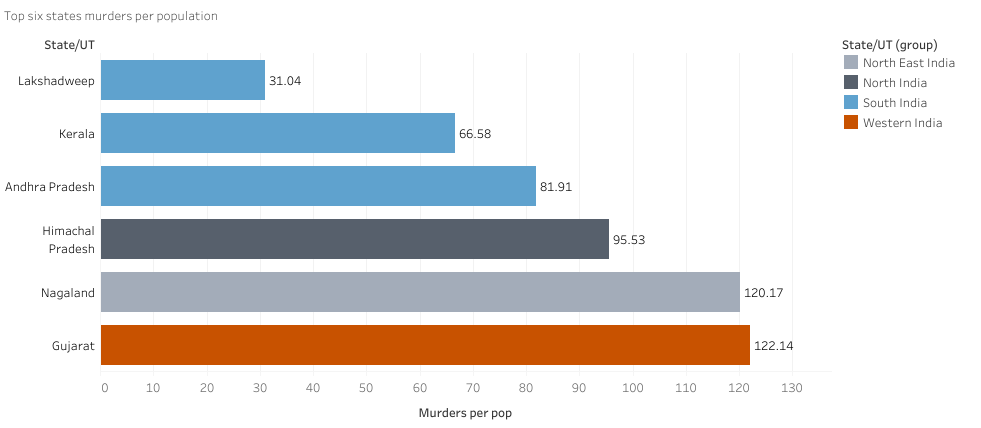


Fig 1.2.14 Top 6 states with least murders per million population

Now we just look at the top 6 states with least murders per million. We see that there are three states of South. So it also makes our claim that South India is the safest region in terms of murder incidence even stronger.

1.3 Critic

There are two concerns regarding areal spatial type of data. These issues are the sensitivity of spatial estimates to the areal units that are employed by the researcher and the effects of omitted areal units on spatial diagnosis and estimation. The first concern is known as the modifiable areal unit problem (MAUP) and the second is known as the boundary value problem. We are mainly concerned with MAUP here.

The MAUP comprises two distinct problems. The scale problem refers to the dependence of spatial correlation findings on the number of areal units into which a spatial plane is divided. A given plane may be divided into 5, 50, 500, or some other arbitrary number of polygons, with spatial autocorrelation results differing fundamentally depending on the n that is chosen. The aggregation problem refers to the dependence of spatial autocorrelation findings on the way that the spatial plane is divided into a particular set number of polygons.

Here we have taken administrative blocks divided in terms of cultural homogeneity and administrative benefits. This line of grouping might not be very effective in finding correlation between spatial units.

Motive Analysis

(2016-2021)

Motives are a major factor in murders, providing insight into the underlying causes and societal issues that lead to violent acts. Understanding these motives allows us to grasp the complex dynamics of human behavior and societal tensions. Motives can vary widely, from personal conflicts and financial desperation to ideological beliefs and psychological factors. By analyzing motives, we can identify patterns and trends in murder cases, shedding light on prevalent issues such as gang-related crime, and hate crimes.

It enables us to develop targeted interventions and preventive measures to address these underlying issues and reduce the occurrence of violent crime. Ultimately, analyzing murder motives provides valuable insights into our society's challenges and helps us devise strategies to create safer and more harmonious communities.

**2.1 Feature Engineering**

We have data of years 2016 to 2021 based on motives. On each of these datasets the motives have been uniquely divided into roughly 20 classes. The analysis based on all these classes will be really tedious and tough. So we have broadly classified the motives into 3 categories, namely, Gain or Greed, Love/Lust and Loath. We have excluded unintentional murders and murders due to other causes from the study as the data based on those motives were incomplete and the actual underlying motive is slippery. We look to analyze the data based on these grouped motives and find patterns and trends from them.

It was a huge task to group these motives into the following classes. The grouping was done by the judgement of the analyst. The idea of grouping these motives into the particular groups is supported by the paper given by [Peter Morrell](https://www.crimeandjustice.org.uk/sites/crimeandjustice.org.uk/files/09627250608553401.pdf). This makes our analysis much easier and lighter. We might have lost some information in grouping the smaller groups into larger ones but that was one risk we had to take to move forward with our analysis. As the stage of grouping the motives is successfully dealt with, now we move onto the actual work of analyzing the data to discern some patterns in it.

**2.2 Analysis**

We first look to plot the total number of murders through different motives in the time line 2016-2021. Here the three primary motives are Greed, Loath and Love/ Lust. The plot is provided in Fig 2.2.1.

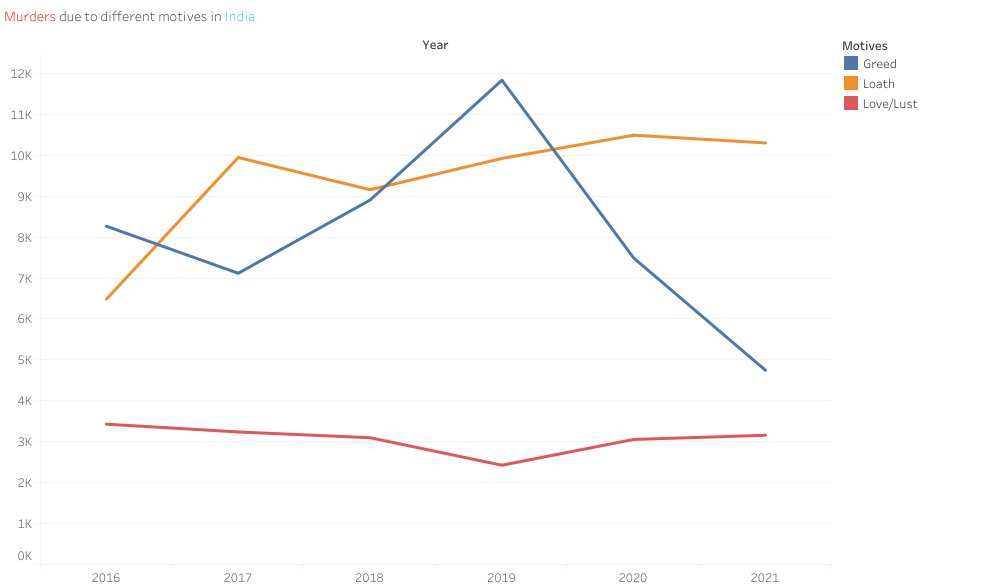


Fig 2.2.1 Time series plot of the total murders due to different motives

From this plot we see a concerning increase in the total number of murders due to loath. To give more context about these motives we provide the list of motives under the respective three primary motives in the table below.

|  |  |  |
| --- | --- | --- |
| Greed | Loath | Love/Lust |
| Political Causes | Disputes | Illicit Relationship |
| Gain | Honour Killing | Love Affairs |
| Robbery and Dacoity | Extremism | Rape |
| Sale of body parts | Casteism |  |
| Dowry | Class Conflicts |  |
| Property Dispute | Religious |  |
|  | Gang Rivalry |  |
|  | Communalism |  |

Now here we have excluded causes like Witchcraft, Lunacy, Blind Murder etc. as the murder counts in them are very low and not of much importance in our study. After making a good idea about what the respective motives are comprised of we move on with our analysis.

Next we plot a stacked bar chart of the total murders and their distribution in percentage among the three motives. The figure is provided in Fig 2.2.2

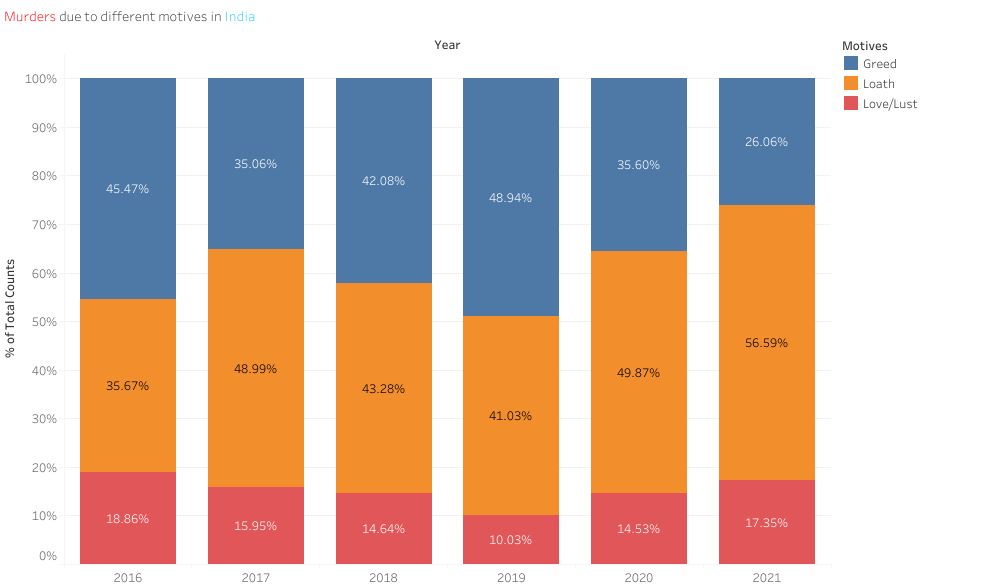


Fig 2.2.2 Stacked bar chart of murder percentage for three motives

We see that murder due to greed has declined by around 19% in the 6 year period. In 2016, majority of the murders have been due to greed at 45.47%. But after six years the landscape of Indian murder motives has changed drastically. In 2021, the majority of crimes has been attributed to Loath at 56.59%, which is a 21% increase from that of 2016. This gives us a clear indication that murders due to loath has increased in the past six years.

