Applied Econometrics & Time-Series Analysis Project Report

University of Texas, Dallas

MORE GUNS LESS CRIMES ?

Submitted to: Dr. Moran Blueshtein

Submitted by: Behab Patnaik

Helen Nguyen

Simran Gupta

Shivani Ramdas

Aishwarya Bhole

Table of contents

Introduction

More guns, less crime has always been a shaky proposition.

With increasing mass shootings and crime-ridden urban centers, many were compelled to believe that state laws enabling citizens to carry concealed handguns would reduce crime.

A Shall-issue law or right-to-carry law(RTC) obligates the government to issue concealed carry handgun permits to any applicant who meets the necessary criteria like the applicant must be an adult, have no significant criminal record, and no history of mental illness and successfully complete a course in firearms safety training

David R. Francis in an article for the National Bureau of economic research mentioned, "In theory, the effect of gun ownership on crime is ambiguous. If criminals are deterred from committing crimes when victims are likely to possess a firearm, then more gun ownership may lead to a reduction in criminal activity. If instead guns increase the payoff to criminal activity, or simply increase the likelihood that any particular confrontation will result in a victim's death, then an increase in gun ownership will tend to increase the crime rate. "

To elaborate, criminals may tend to think twice before attacking, thus opting for property crimes or other crimes that do not involve direct contact with the victims. This would work in the favor of our argument that more guns do lead to less crimes. But on the other hand, using guns for the purpose of defense against criminals is rare. Since aggressors are opportunistic, and retain the element of surprise, it makes it difficult for even the trained professionals to repel against such an armed assault. Also, study has proven that majority of the firearms used in crime are obtained either from burglaries or from the secondhand market. Thus, as the rate of gun ownership in the general population increases, the ease with which criminals can obtain a gun might also increase proving to be a dangerous byproduct of the RTC laws.

While a lot of different theories and scenarios have been thoroughly studied and put to test on the effects of concealed carry laws, none of them have yet been able to put a final word to this argument leaving it open for a wishful thinking.

Weaving the economic concepts and statistical procedures together, we have analyzed one such data piece in our project that spans 51 US states for 23 years (1977-1999) and try to choose a side based on our findings.

I. EXPLORATORY DATA ANALYSIS

We are working on a balanced panel dataset on 50 US states, plus the District of Columbia over a 23-year period of time. The total number of observations is 1173 and there's no null values in the dataset.

Before building our hypotheses and models, we will do descriptive analysis and some exploratory analysis to have an overview, as well as better insights about our data such as relationship between variables, trend, ... Besides 3 dependent variables (vio, rob, mur), we will also focus on 2 important explanatory variables - shall and incarc_rate to answer our main the question "Do more guns reduce crime?"

1. Descriptive statistics:

Variable	*vio*	*mur*	*rob*	*incarc_rate*	*shall*
Mean	503.07	7.67	161.82	226.58	0.24
Standard Error	9.76	0.22	4.98	5.22	0.01
Median	443.00	6.40	124.10	187.00	-
Mode	256.80	3.60	111.60	98.00	-
Standard Deviation	334.28	7.52	170.51	178.89	0.43
Sample Variance	111,741.24	56.59	29,073.65	32,000.95	0.18
Skewness	2.54	5.79	3.89	3.89	1.20
Range	2,874.80	80.40	1,628.70	1,894.00	1.00
Minimum	47.00	0.20	6.40	19.00	-
Maximum	2,921.80	80.60	1,635.10	1,913.00	1.00
Sum	590,106.60	8,991.20	189,815.10	265,778.00	285.00
Count	1,173.00	1,173.00	1,173.00	1,173.00	1,173.00
Confidence Level (95.0%)	19.15	0.43	9.77	10.25	0.02

For these 5 main variables (measured per 100,000 people), on average over 23 years and 51 states, violence rate is much higher compared to robbery and murder. However, the standard deviation of violence is smaller than standard deviation of murder and robbery compared to theirs mean. Violence is also less skewed than murder and robbery.

Although incarceration rate is also high, it's still much lower compared to total crime rate (violence and robbery and murder).

Variable	*pb1064*	*pw1064*	*pm1029*	*pop*	*avginc*	*density*
Mean	5.34	62.95	16.08	4.82	13.72	0.35
Standard Error	0.14	0.29	0.05	0.15	0.07	0.04
Median	4.03	65.06	15.90	3.27	13.40	0.08
Mode	#N/A	#N/A	#N/A	#N/A	11.66	#N/A
Standard Deviation	4.89	9.76	1.73	5.25	2.55	1.36

Sample Variance	23.87	95.29	3.00	27.58	6.53	1.84
Skewness	2.35	(2.23)	0.27	2.43	0.74	6.70
Range	26.73	54.75	10.14	32.74	15.09	11.10
Minimum	0.25	21.78	12.21	0.40	8.55	0.00
Maximum	26.98	76.53	22.35	33.15	23.65	11.10
Sum	6,259.38	73,834.99	18,863.16	5,649.57	16,099.19	412.94
Count	1,173.00	1,173.00	1,173.00	1,173.00	1,173.00	1,173.00
Confidence Level (95.0%)	0.28	0.56	0.10	0.30	0.15	0.08

2. Correlation between variables:

Below is the heatmap for correlation indexes between our variables. With blue is strong positive correlation, red is strong negative correlation.

	year	*vio*	*mur*	*rob*	*incarc_*	*pb1064*	*pw1064*	*pm1029*	*pop*	*avginc*	*density*	*stateid*	*shall*
year	1.00												
vio	0.12	1.00											
mur	-0.03	0.83	1.00										
rob	-0.01	0.91	0.80	1.00									
incarc_rate	0.50	0.70	0.71	0.57	1.00								
pb1064	0.07	0.57	0.60	0.58	0.53	1.00							
pw1064	-0.03	-0.57	-0.62	-0.58	-0.53	-0.98	1.00						
pm1029	-0.87	-0.17	0.01	-0.09	-0.45	0.02	-0.01	1.00					
pop	0.06	0.32	0.10	0.32	0.10	0.06	-0.07	-0.10	1.00				
avginc	0.53	0.41	0.22	0.41	0.46	0.26	-0.19	-0.53	0.22	1.00			
density	0.00	0.66	0.75	0.78	0.56	0.54	-0.56	-0.06	-0.08	0.34	1.00		
stateid	0.00	-0.32	-0.24	-0.25	-0.22	-0.31	0.31	0.01	-0.06	-0.20	-0.16	1.00	
shall	0.38	-0.21	-0.18	-0.21	0.04	-0.18	0.21	-0.28	-0.12	0.00	-0.11	0.19	1.00

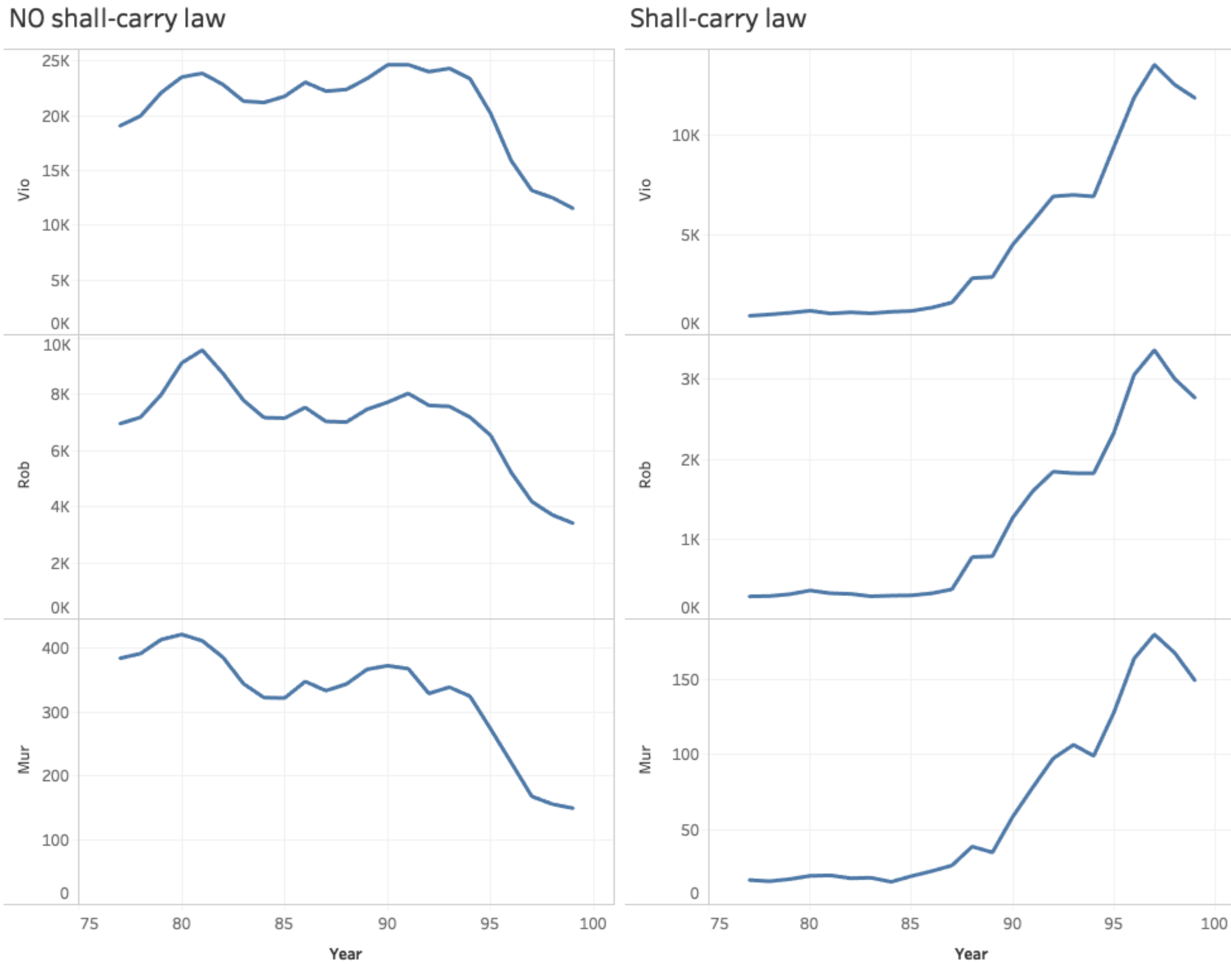
One of the important points we should notice is violence, murder and robbery are highly correlated (positive correlation), especially violence and robbery. Incarceration rate is highly correlated with violence and murder, but not that high with robbery. However, the shall-carry law is not correlated with crime rate, according to this overall analysis. This may not be accurate as the number of shall-carry law record is much lower than no shall-carry law in this dataset. We also use R to do pair plot for correlation of each variable - mur, vio, rob with the rest variables and they seem to have the same pattern. We will examine this further in our later parts.

There's also a positive correlation between density and crime rate, which makes sense to us. The percentage of white in the population is strongly negatively correlated with the percentage of black in the population. And, as time gone by, the percentage of young male in the population decrease.

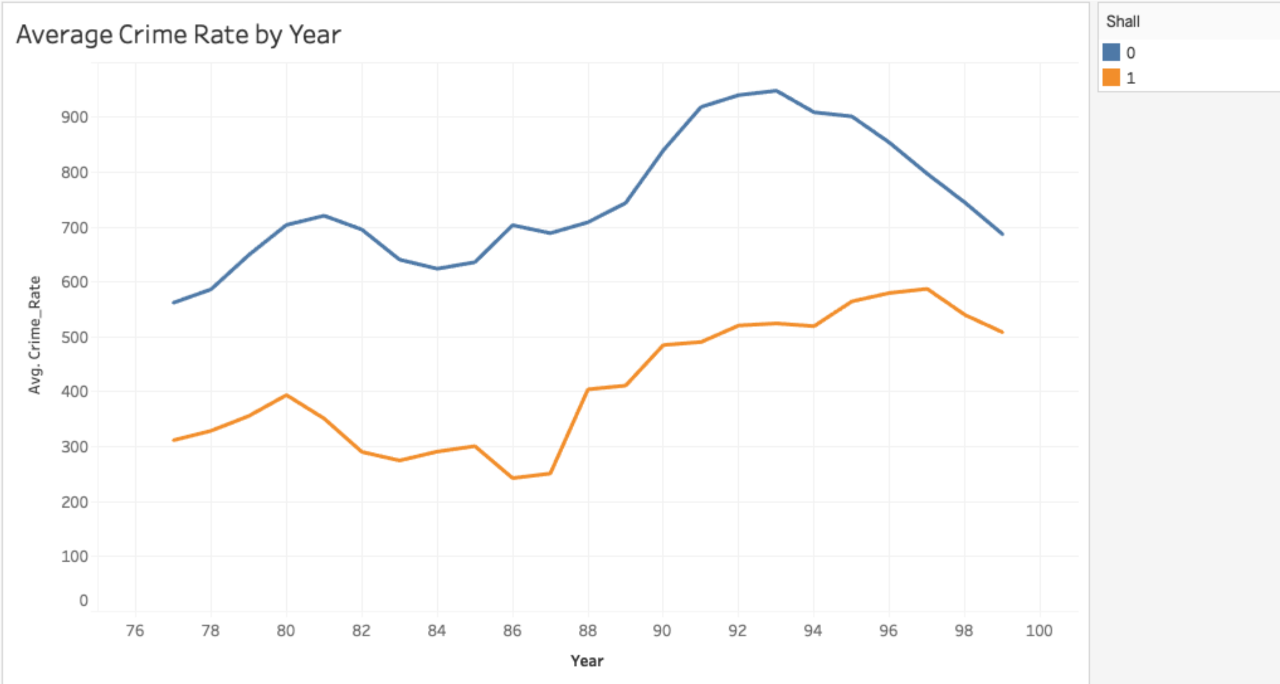
3. Variables studying in detail:

3.1. Dependent variable - Crime rate:

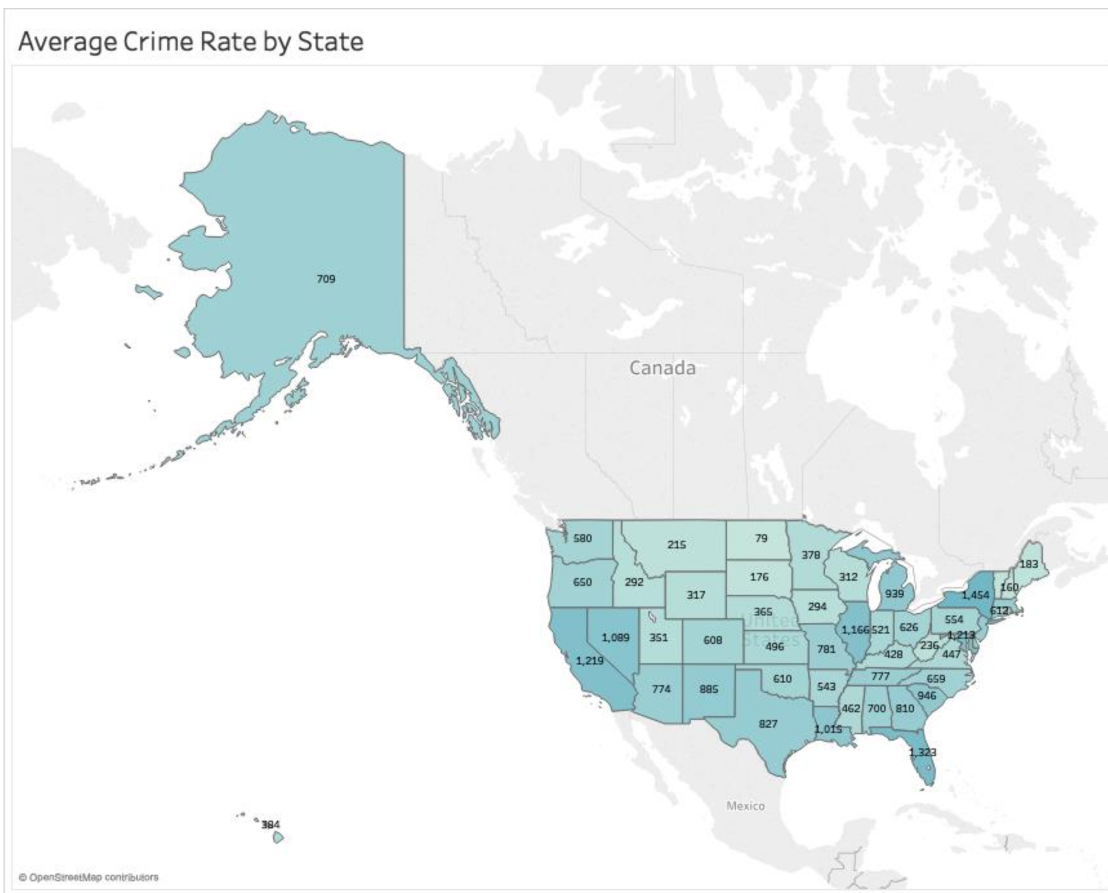
Below is the violent, robbery, murder crime rate respectively by year and by shall-carry law. These 3 variables have the same pattern in each case (shall-carry law and no shall-carry law). Additionally, shall-carry law had utility in reducing the violent crime up till 1989. (Since last decade the effect of Shall Carry Law is more or less similar to its counterpart.)



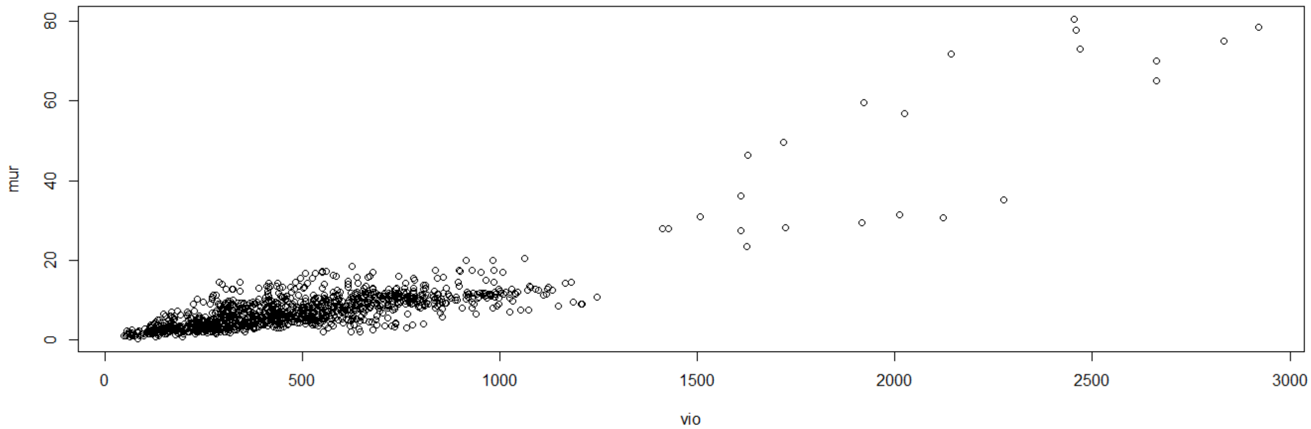
We have also known from the previous part that violence, murder and robbery are highly correlated. We will consider all of them as one, which is crime rate, by summing these 3 variables. From 1977 to 1999, the average crime rate has an upward trend, with shall-carry law has lower crime rate than none shall-carry law.

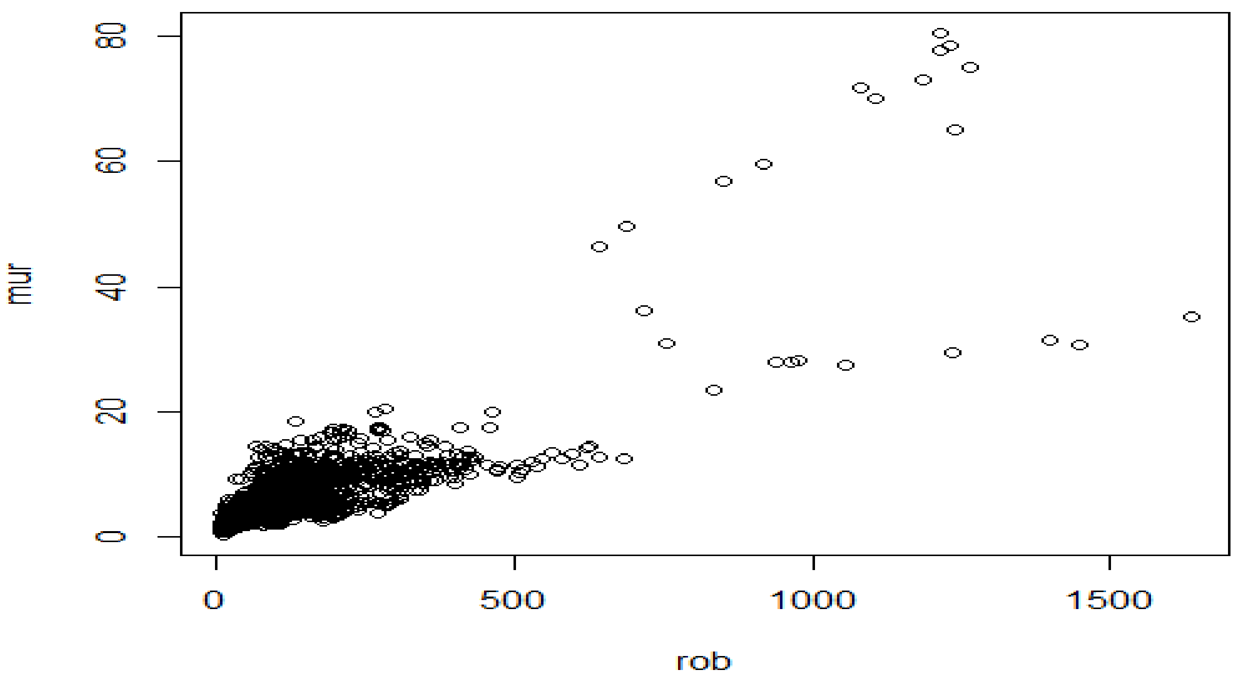


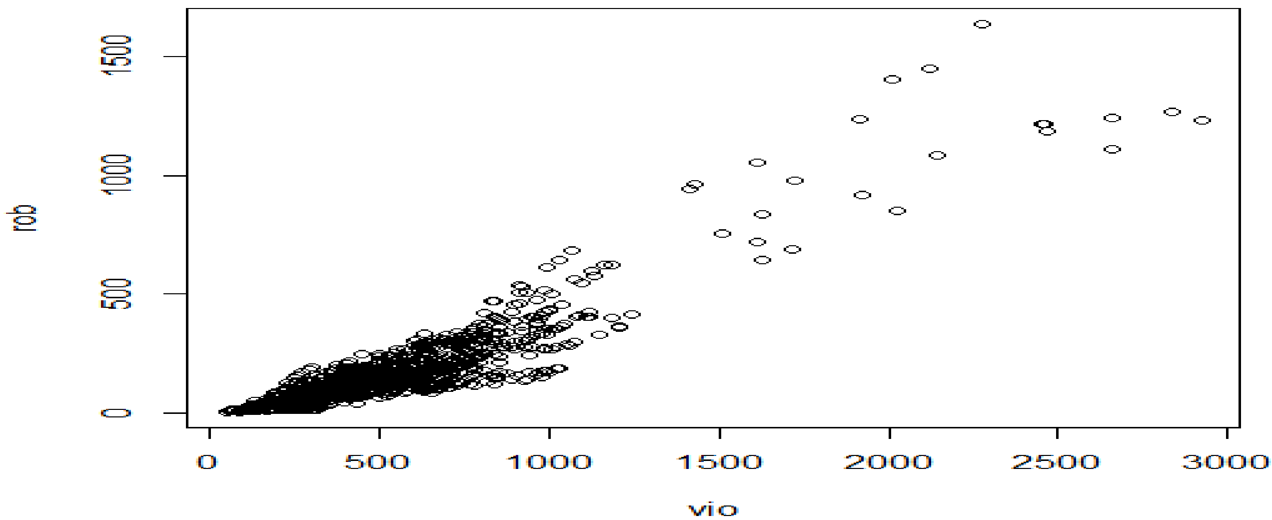
We can see from the below graph, the states with highest crime rate (regardless of applying shall-carry law or not) are District of Columbia, Florida, California, Illinois, and Maryland.



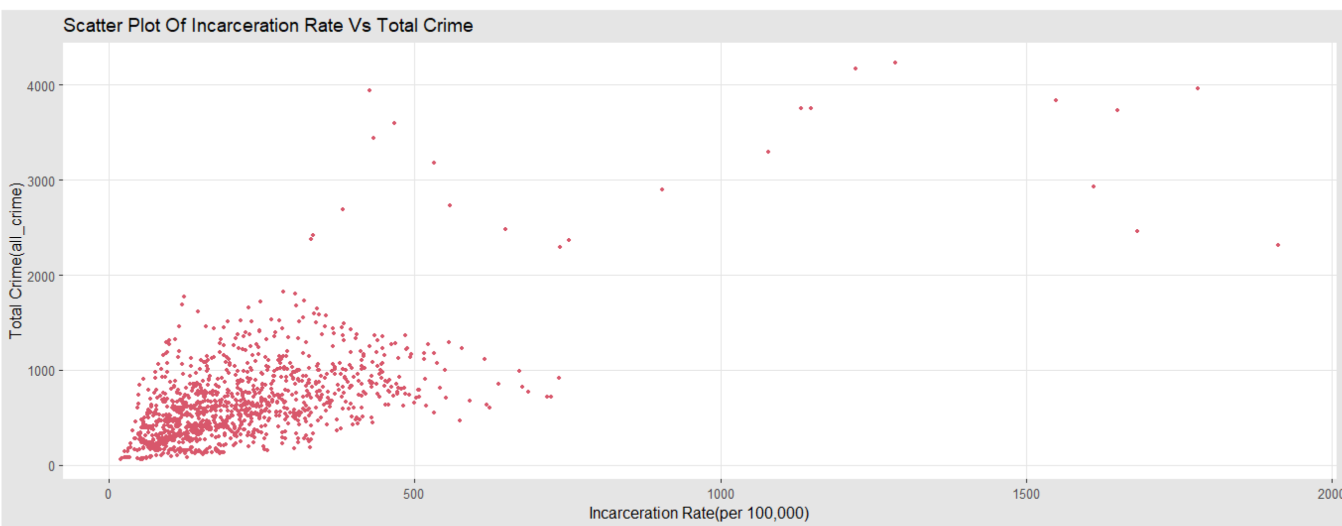
We see a strong positive correlation between the variables- mur,vio,rob which is justifiable since murder, violence and robbery rate are interrelated.


mur
0
20
40
60
80
vio
0
500
1000
1500
2000
2500
3000


mur
0
20
40
60
80
rob
0
500
1000
1500

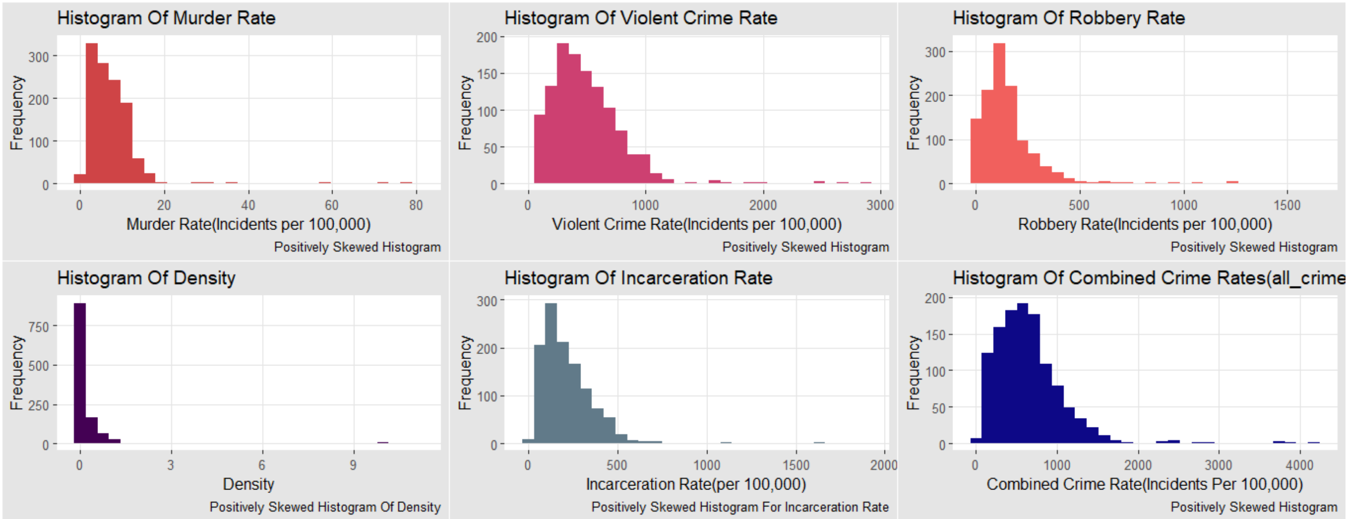


Hence we create another variable **all_crime** that sums up all 3 variables. There is a moderate to strong correlation between incarceration rate and these 3 variables individually. So we plot all_crime against incarc_rate to check whether our new variable captures this effect or not.

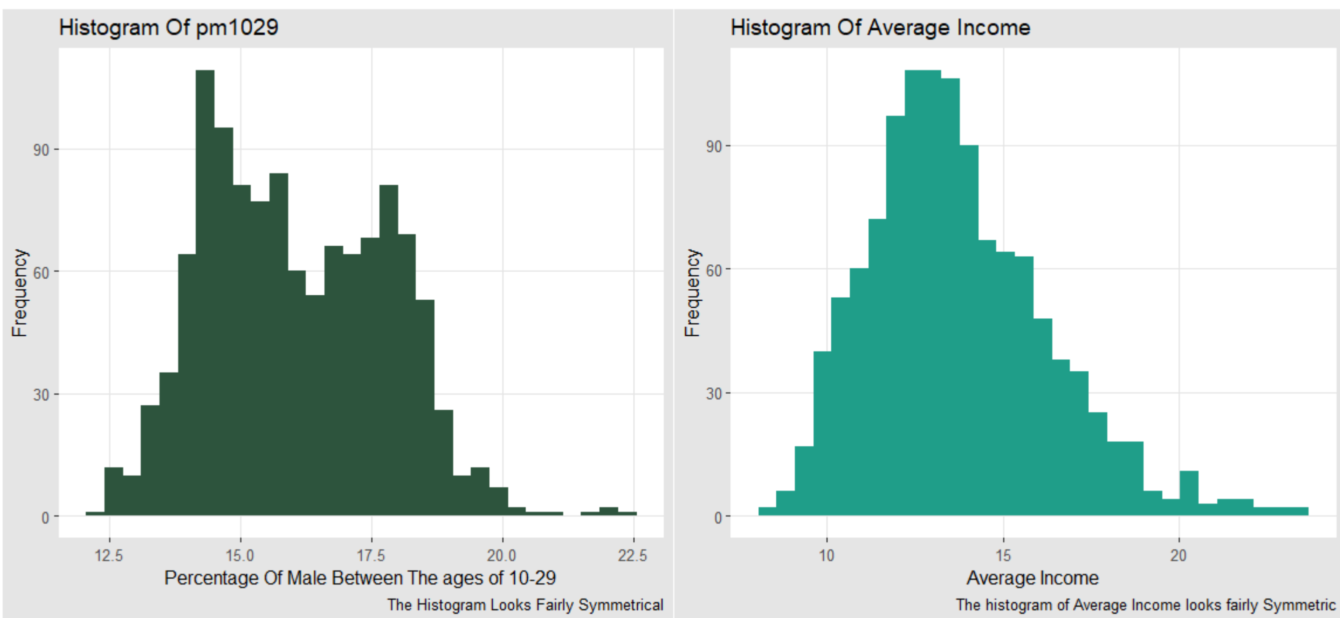


Histogram plots of dependent and explanatory variables.

Below we see the distribution of variables like mur,vio,rob,density,incarc_rate,all_crime,pm1029,avginc through the histogram plots.



All the above variables have a positively skewed histogram. Thus we use the ln() in order to take logarithmic values of these variables as these values will assist in dealing with issues like heteroscedasticity or non normal residuals if encountered further thus making our model estimates more efficient and improving the precision of the model.

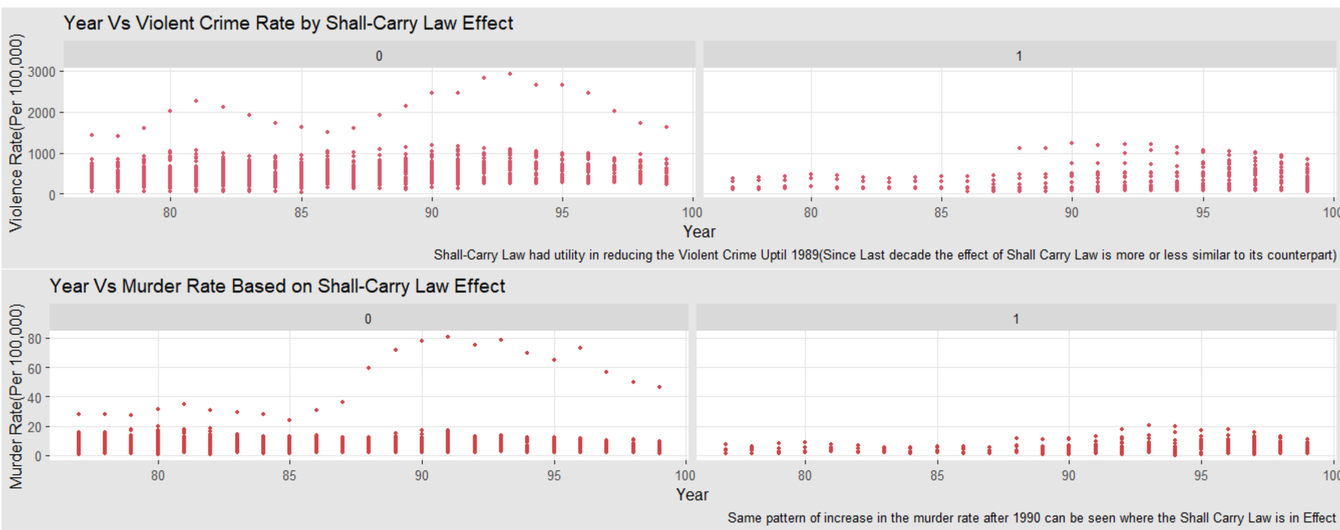


The percentage of males in the age group 10-29 and the average seem to have a fairly symmetrical distribution i.e no skewness observed and hence don't need any transformation.

Now in order to study the effect of the implementation of the Shall-Carry Law, we perform the following explanatory analysis.

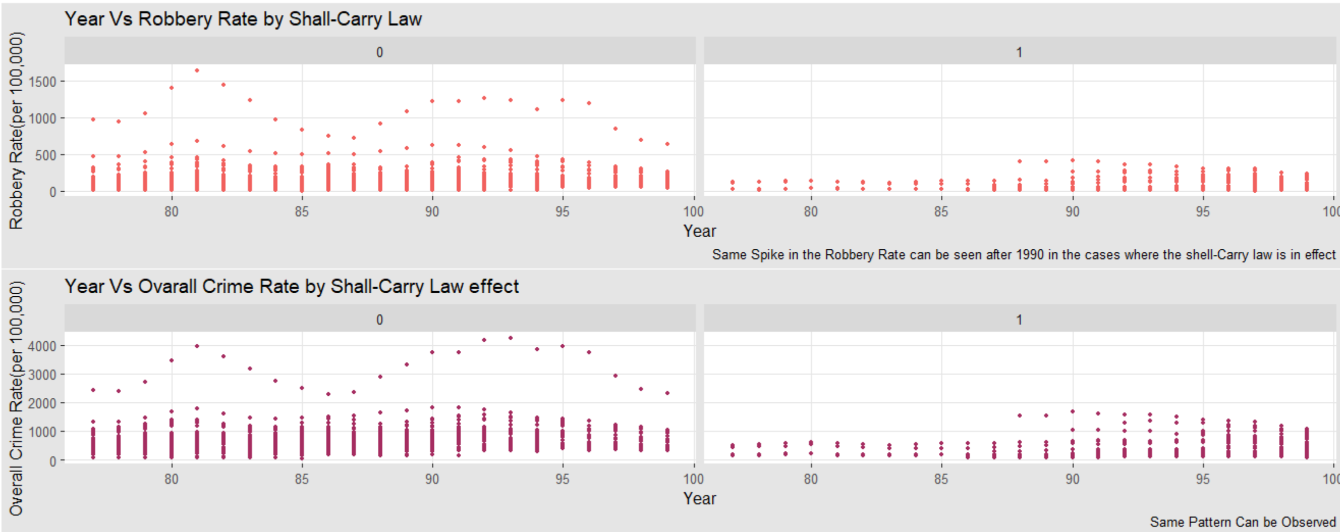
In the above boxplot, we can observe that the overall crime rate is less in states where the Shall-Carry Law is implemented as compared to the states without the Shall-Carry Law. Thus we can say here that the implementation of the Shall-Carry Law has a considerable effect on the overall crime rates.

Since the data is captured over a period, we can further investigate if the effect of Shall-Carry Law implementation varies with time.



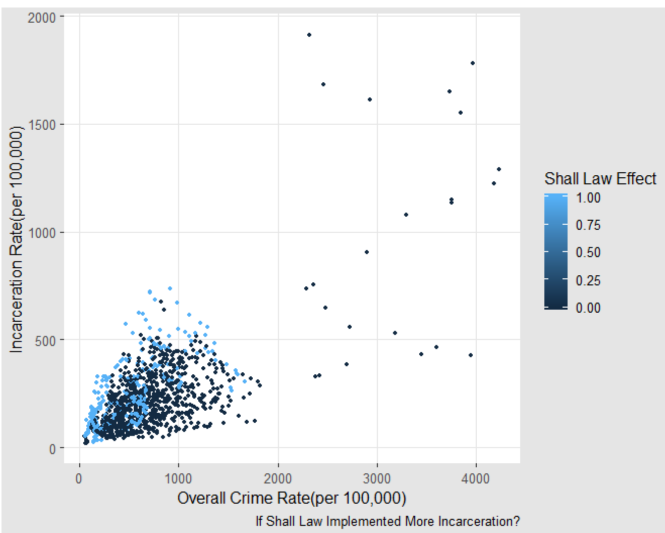
We observe that over the period of 1977-1999, the violent crime rate is less throughout in states with Shall-Carry Law in comparison with the states without the Shall-Carry Law. However we can note a trend in the states with Shall-Carry Law that the violent crime rate has increased in the last decade (after 1989) and thus the law seems less effective in later years.

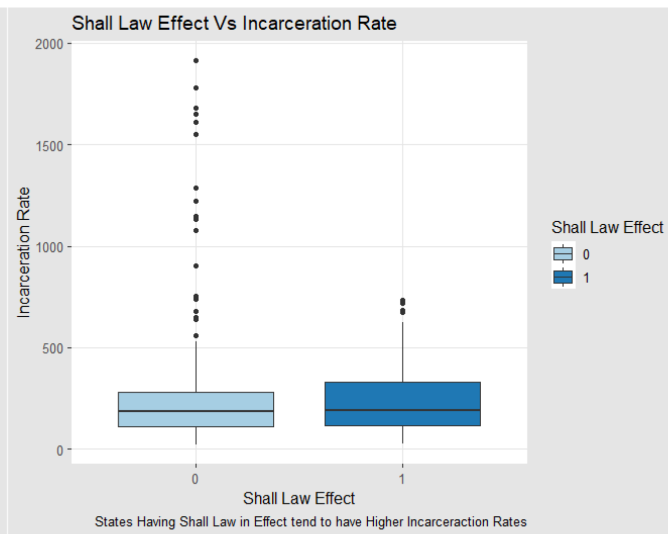
A similar trend is observed with murder rate from 1989 to 1999.



There exist a similar trend with Robbery rate over the period of time. Since all these variables are highly correlated, we can observe a similar trend in the overall crime rate.

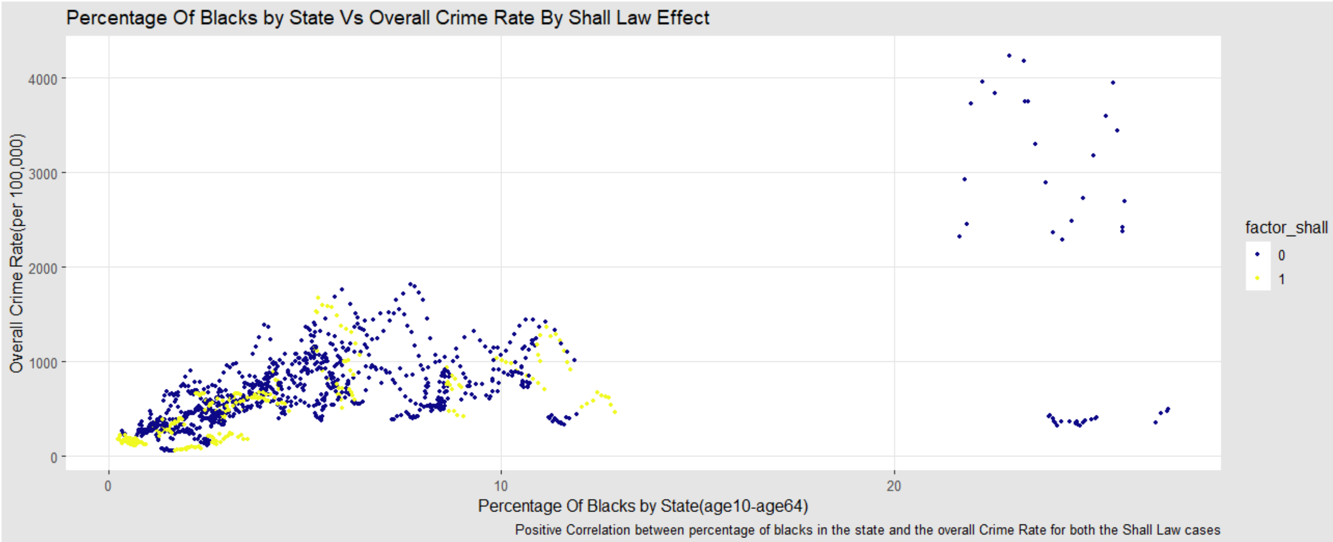
Thus we can deduce that with time, the effect of Shall-Carry Law has not been effective in reducing the Crime Rates.





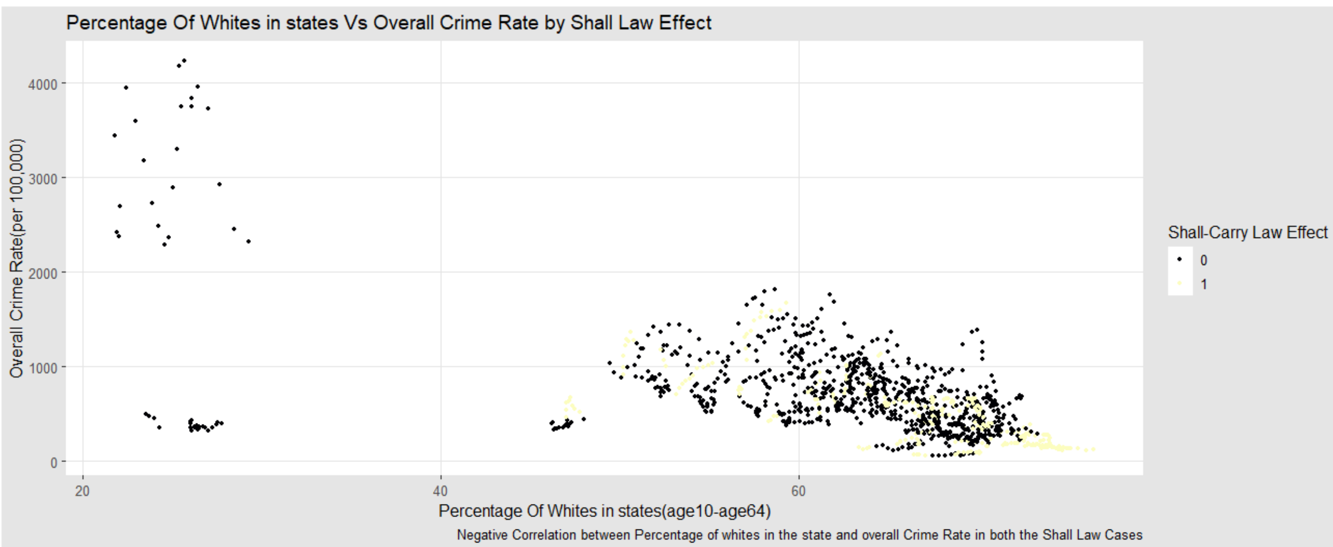
We see the trend between incarceration rate and over all crime rate with respect to Shall-Carry Law. There is a difference in the slope where the law is implemented and where its not. The states where the Law is implemented, the overall crime rate is less and the incarceration rate

Our data segregates the population by the percentage of blacks and whites in a region. Their effect on the Crime rates and the effect of Shall Law on this segregation can be viewed below.



There exist a trend between the percentage of blacks in a region and the overall crime rates. There is an increase in the overall crime along with an increase in the percentage of Blacks lying in the age group of 10-64. Thus there is a moderate-strong positive correlation between these two groups.

We can further see a bifurcation based on the Shall-Carry Law implementation. However there is no evidence that suffices to say that there exist a definite trend or pattern.

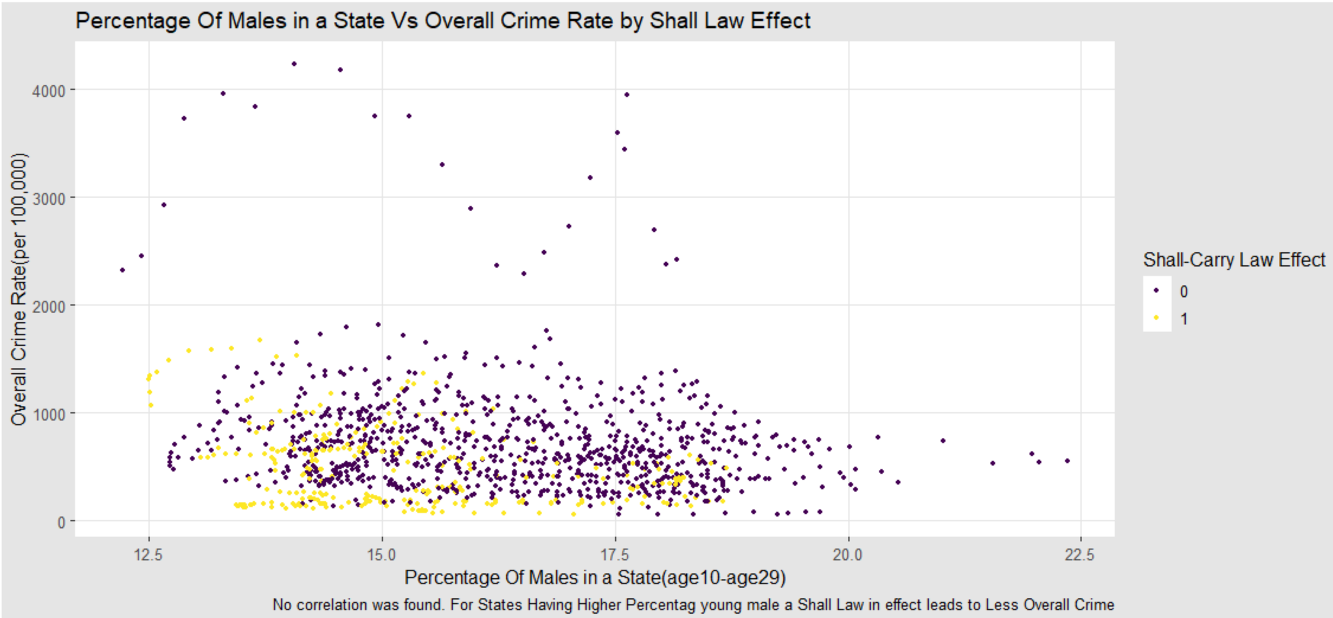


There is descending trend observed between the percentage of whites and the overall crime rate. The overall crime rate is on a decline as the percentage of whites in the state reduces. There is a negative correlation between these two groups.

As observed with Percentage of Blacks, there is no significant effect of implementation of Shall-Carry Law.

Intuitively, we can also justify the difference in the trends as the number of blacks in a region will be less where the population of whites is more and vice versa.

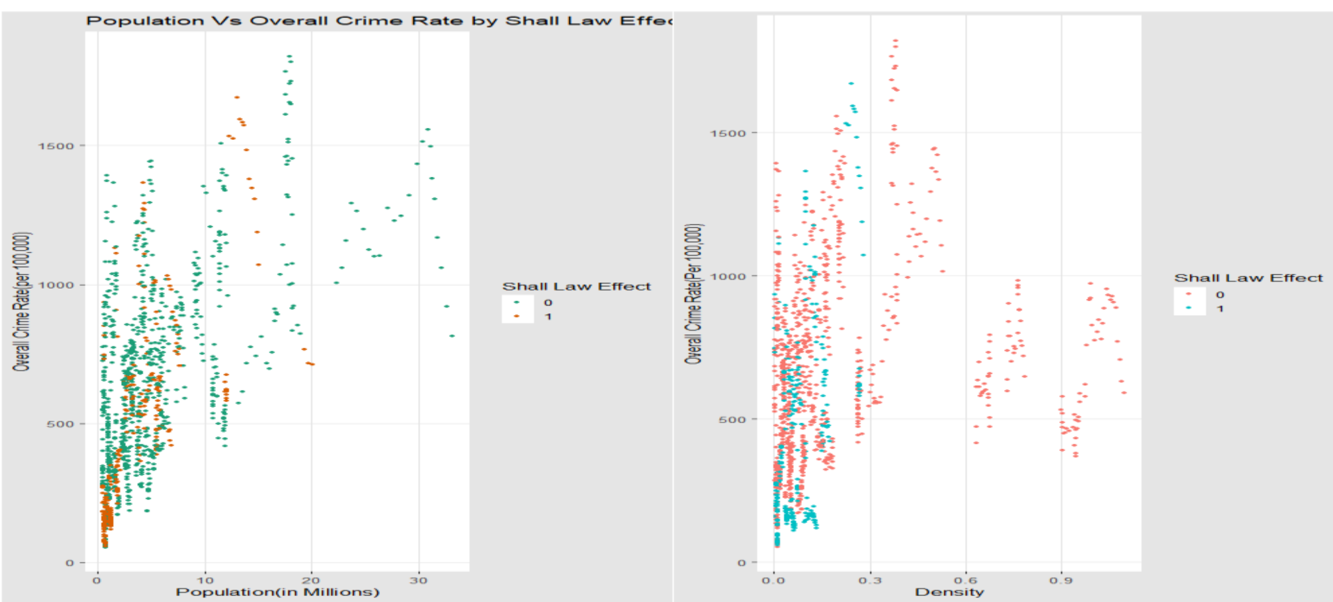
The data provides a bifurcation of percentage of young males(age group 10-29) of the entire population.



The overall crime rate doesn't fluctuate much with an increase in concentration of Young males in a state. The overall crime is almost constant for all Percentages of males (age 10-29). Thus, there is no correlation observed between the two groups.

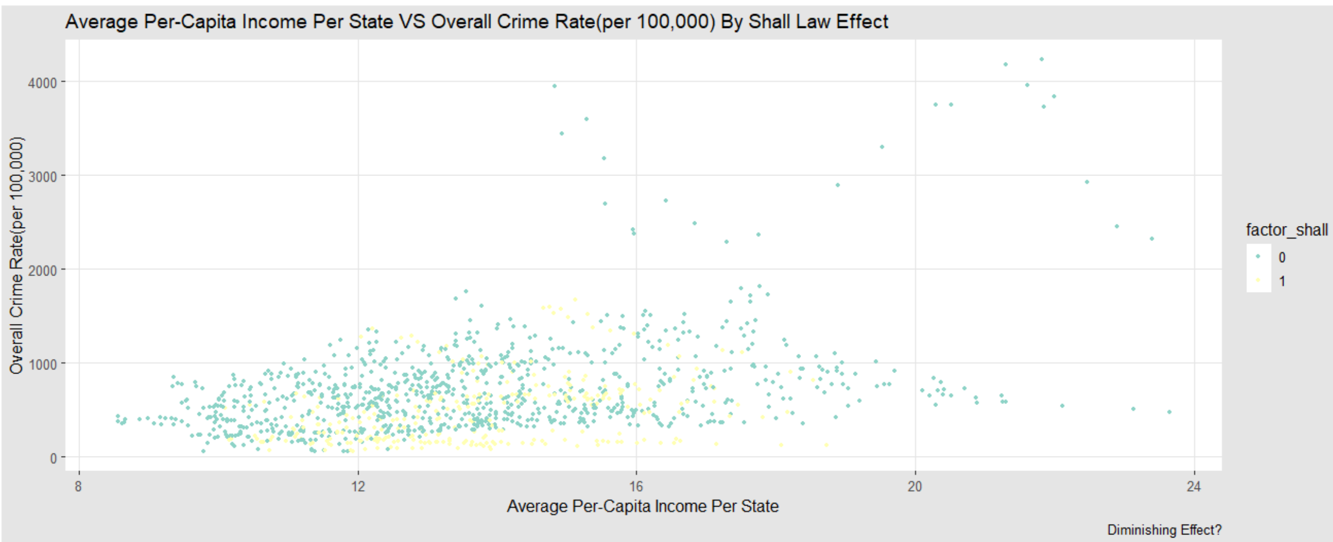
The implementation of Shall-Carry law doesn't provide any significant effect on the relation between the overall crime rate and the Young Male population. We can see that the implementation of Shall-Carry law has led to a decrease in the overall crime rate for states with high Percentage of young male population. Also we can see no implementation of Shall-Carry law where the Percentage of Male population. However this effect is not remarkable.

We can study the relation of population and density with the overall crime rate.



We see positive trend between Population and the Overall Crime rates. Similar trend can be seen between Density and the Overall Crime rate. Intuitively highly populated regions tend to have high crime rate as denser the region, higher scope for robbery, murders or other crimes to occur. Thus there is a strong positive correlation between these groups.

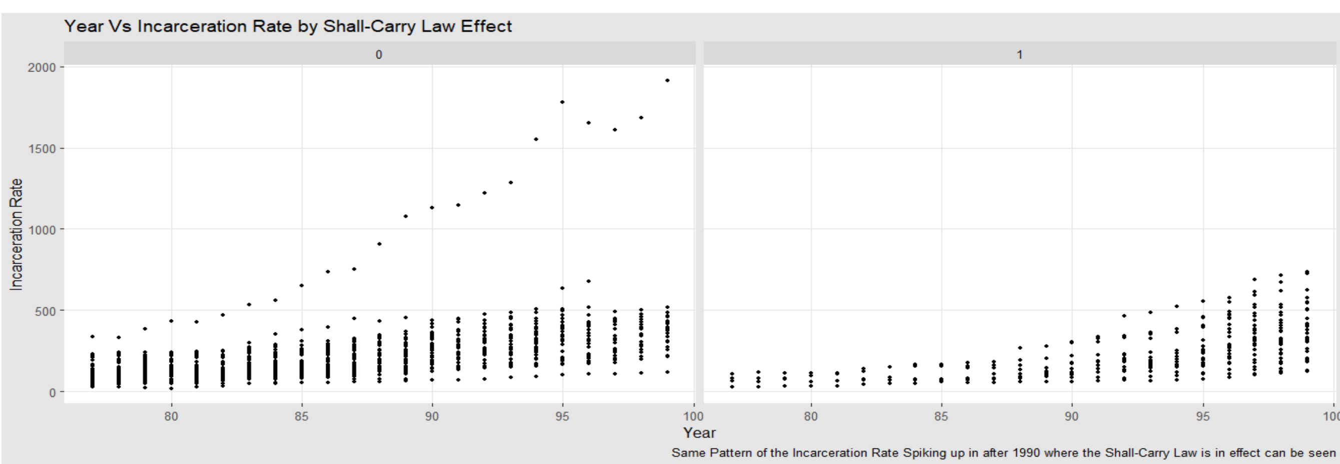
We can see that the implementation of Shall-Carry Law doesn't seem to have any significant effect. There is no remarkable trend with respect to low or highly dense states.



From the scatterplot above, we can observe that there is not strong linear relation between the Average Income of the population and the over all income. Thus we cannot say there is no correlation between these groups.

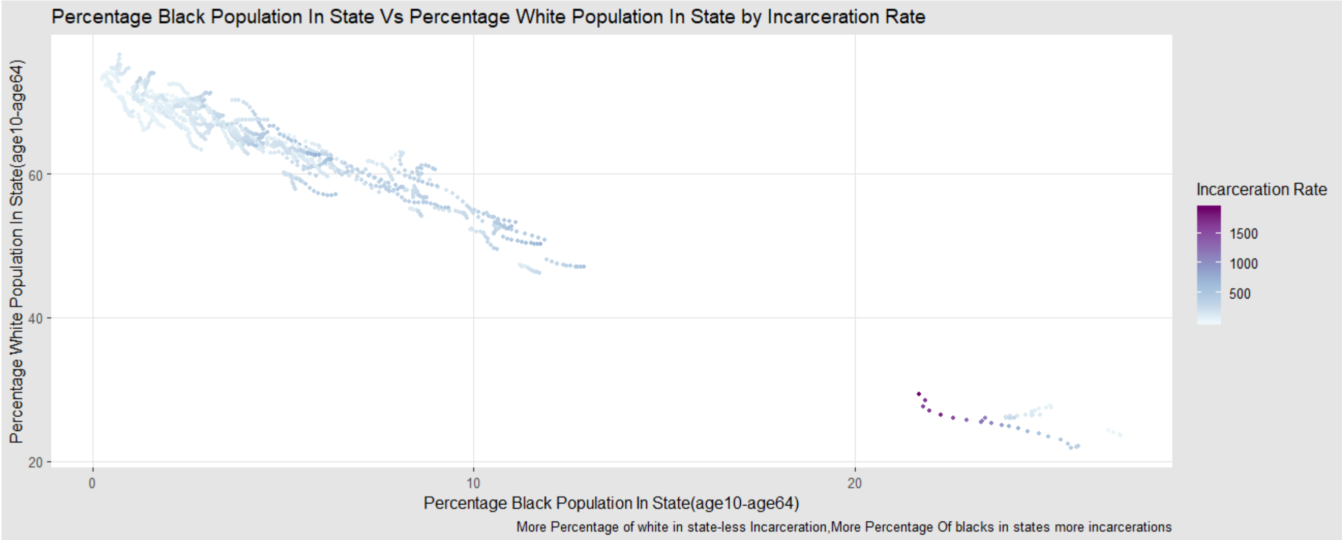
Also the implementation of the Shall-Carry Law doesn't provide any significant trend in the relation between the Overall crime rate and the average income of the state.

Further we study the effect of various variables on the incarceration rate.



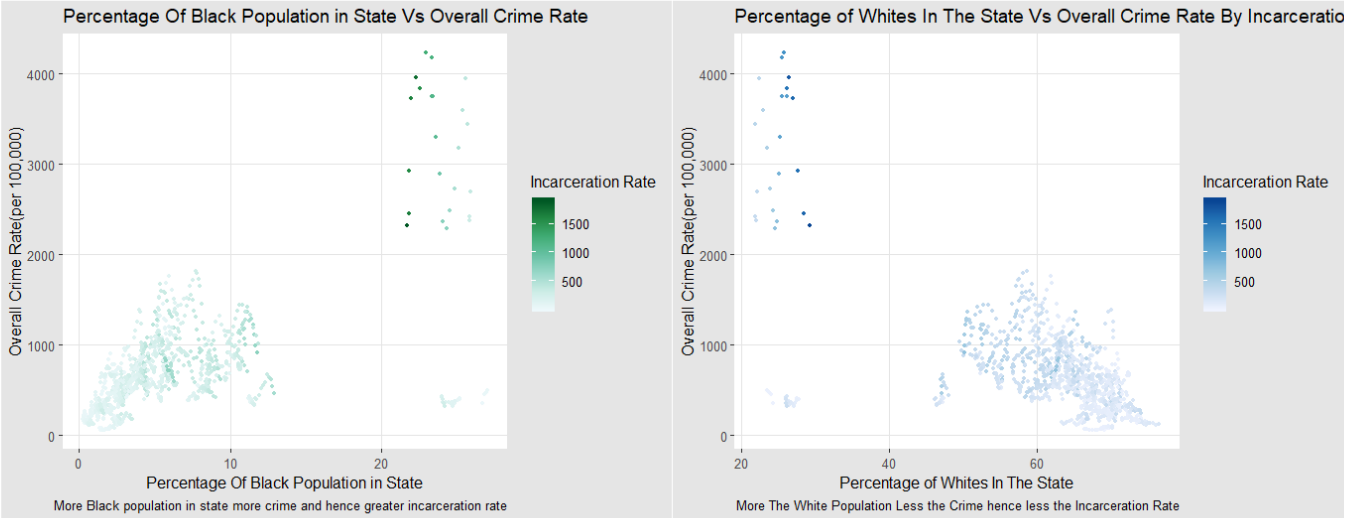
We can observe that there is an increase in the incarceration rate from 1977 to 1999 irrespective of the implementation of Shall Carry Law. However in states where the Law had been implemented, we can see that there has been a spike in the incarceration rate in the last decade i.e after 1990.

Similar effect was observed previously with overall crimes. There was a spike in the overall crime rate in the last decade. Intuitively, we can say that since there was an increase in the crime rate post 1990, with the implementation of the Shall-Carry Law, there is an increase in the incarceration rate as well.



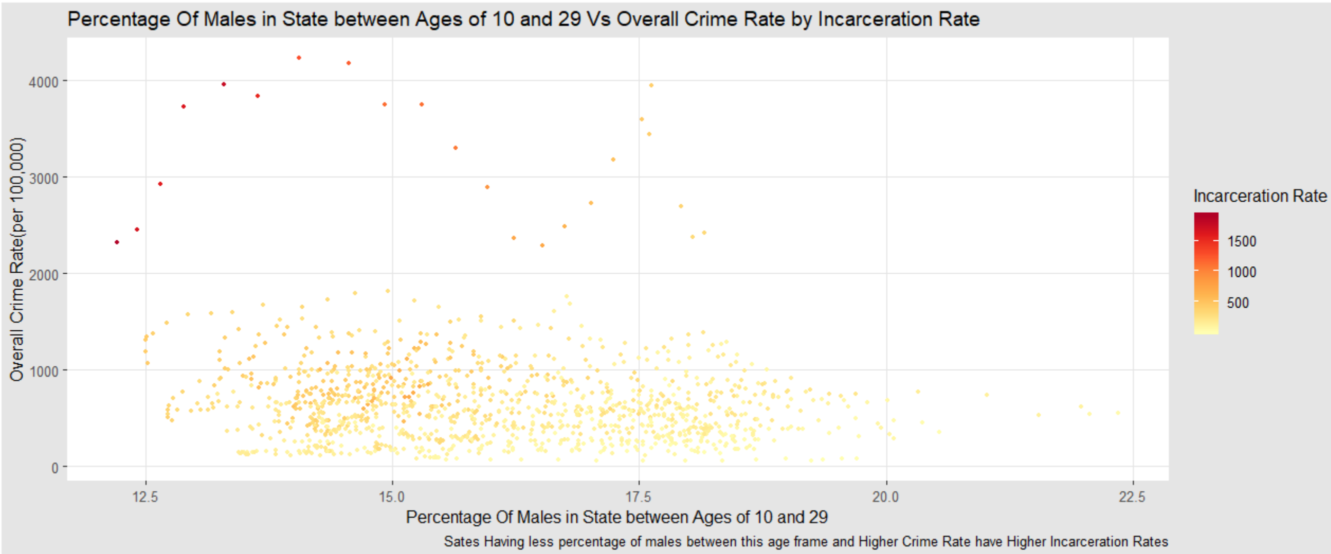
As discussed above, we can see the trend in states where the population of blacks is higher, the population of whites is lower and vice versa. There is an inverse relationship between the Population of Blacks and Whites of a state.

From the plot above we can see that states with higher percentage of white have lower incarceration rates and states having comparatively more percentage of Blacks have higher incarceration rate.

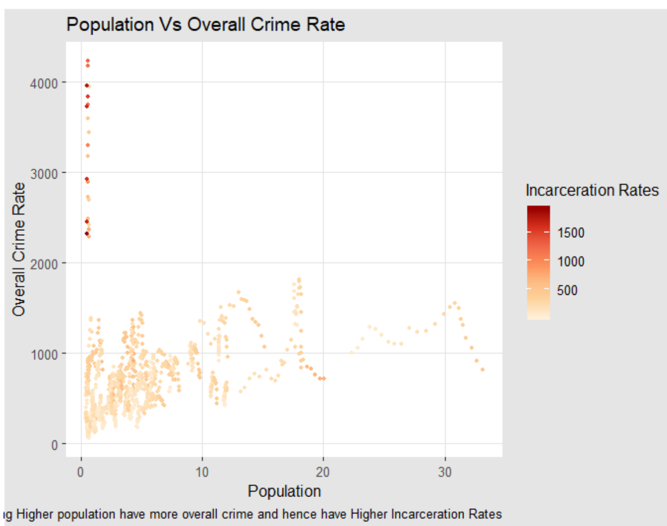


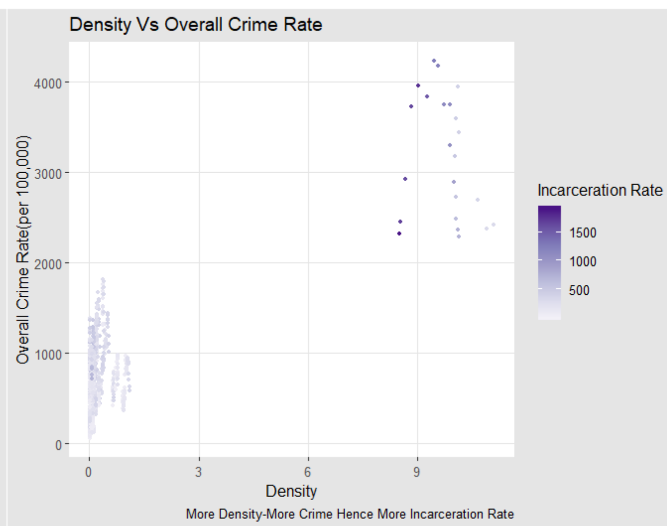
We can observe from these plots that states with low population blacks which also implies higher population of whites, the overall crime rate is lower which also explains the low incarceration rate.

States with high Population of Blacks i.e low Population of Whites have higher overall crime rate along with higher incarceration rates.

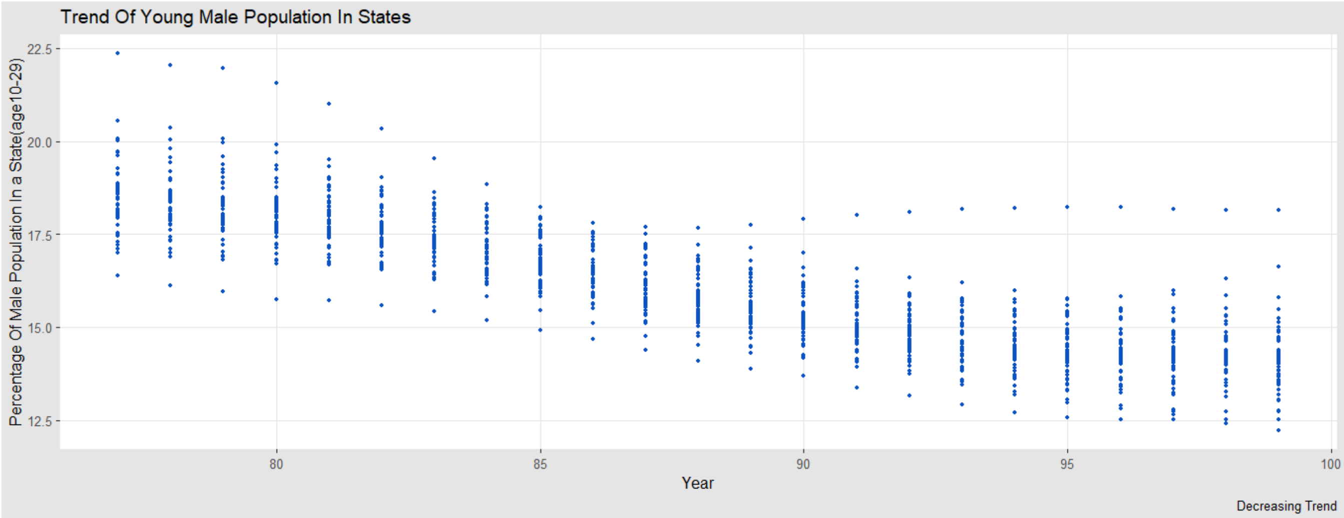


As observed earlier, there is no strong correlation between the Percentage of young males and the overall crime rates. However we can see that where there is low Percentage of Young Males with Higher Overall crime rates, the incarceration rate is also higher as compared to other scenarios.

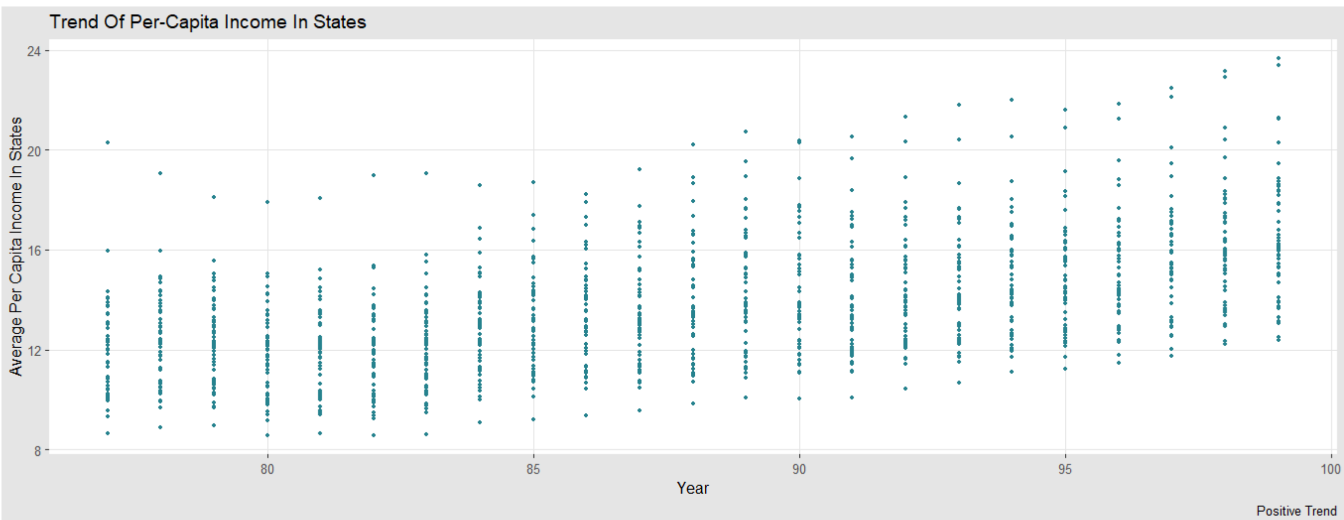




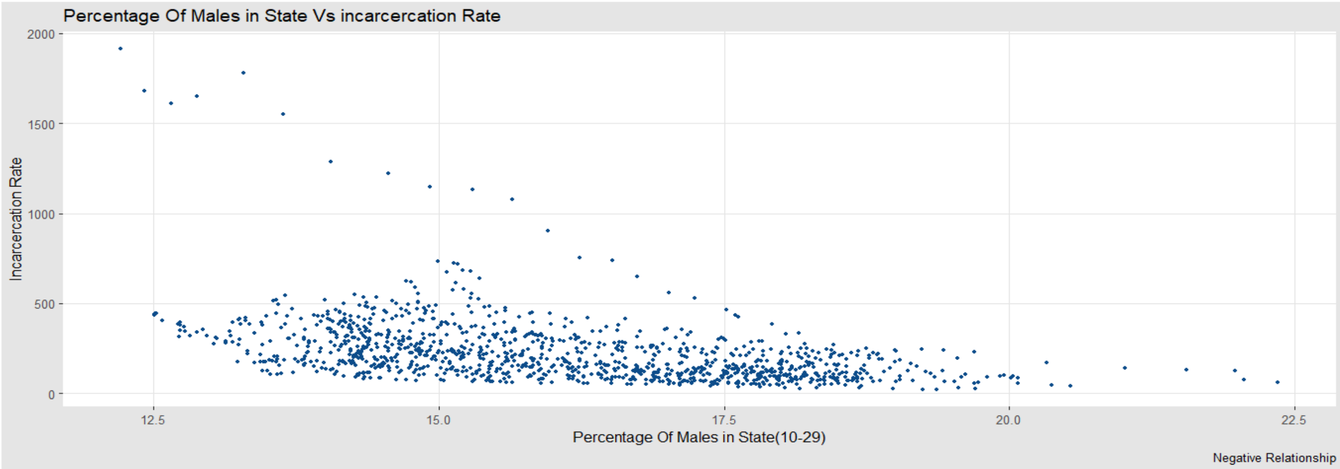
We know that population and density are correlated with the overall crime rate. The above plot shows us the variation of incarceration rate with these groups. No remarkable trend is observed however we can say that states with higher density of population, having higher Overall Crime rates tend to have higher Incarceration rates.



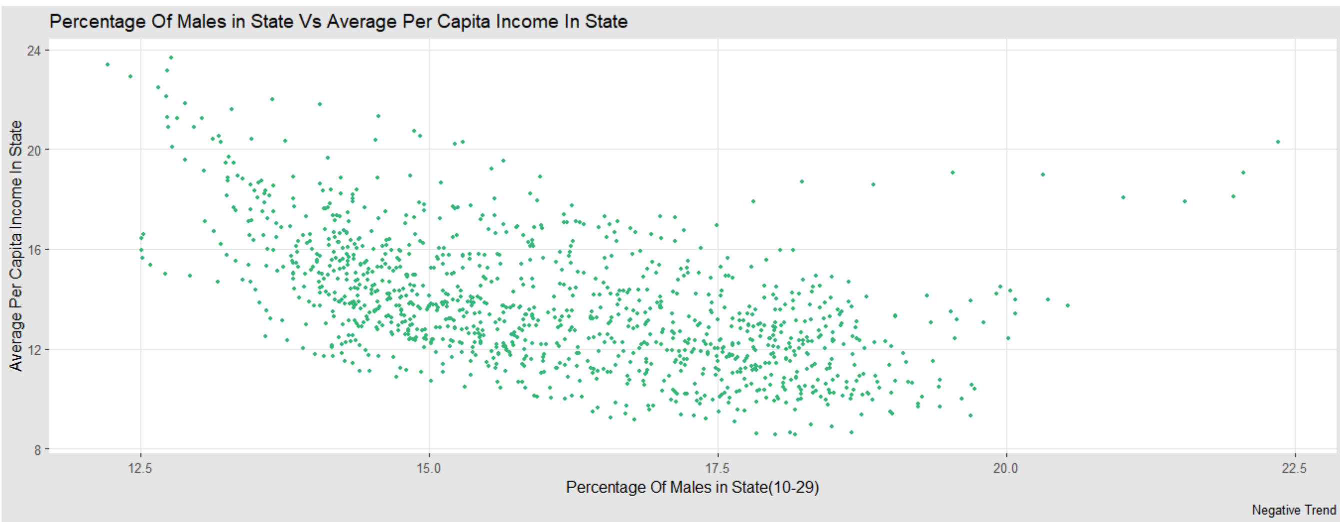
From the above plot we can infer that there is a decline in the Percentage of the Young Male Population over the years from 1977-1999. Time has an effect on the overall crime rates and the incarceration rates. They increase over the time and substantially increased in the last decade. Although descending, but we can see a similar effect or trend in the above plot.



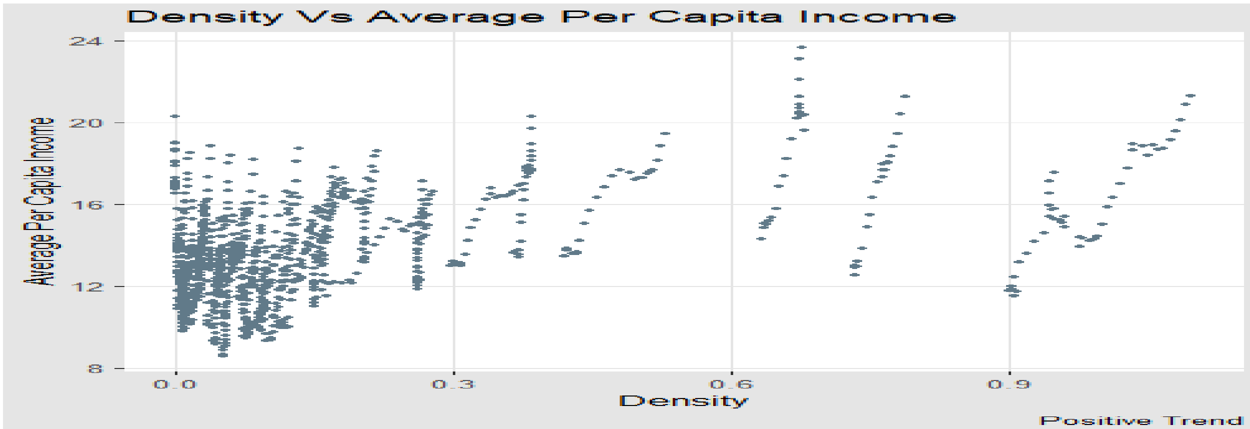
The Average per Capita Income have been seen to rise over the time period from 1977-1999. Thus a positive trend is observed.



There are no significant inferences or trends that can be drawn from the plot above. However we see that there is a slight decline in the incarceration rate as the Population of Young Males increases.

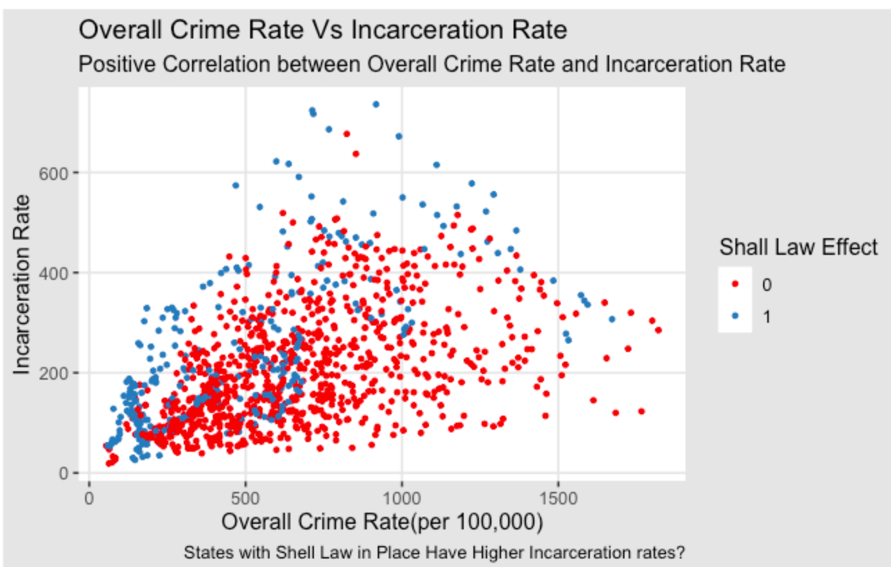


The Average per Capita Income is observed to be on a decline with increase in the Population of Young Males in the state. This can be justified intuitively. A person tends to have a lower income as compared to in the later years. Thus if the Population of Young Males in the state is higher, the average per capita income is lower. We observe a descending or negative trend here.

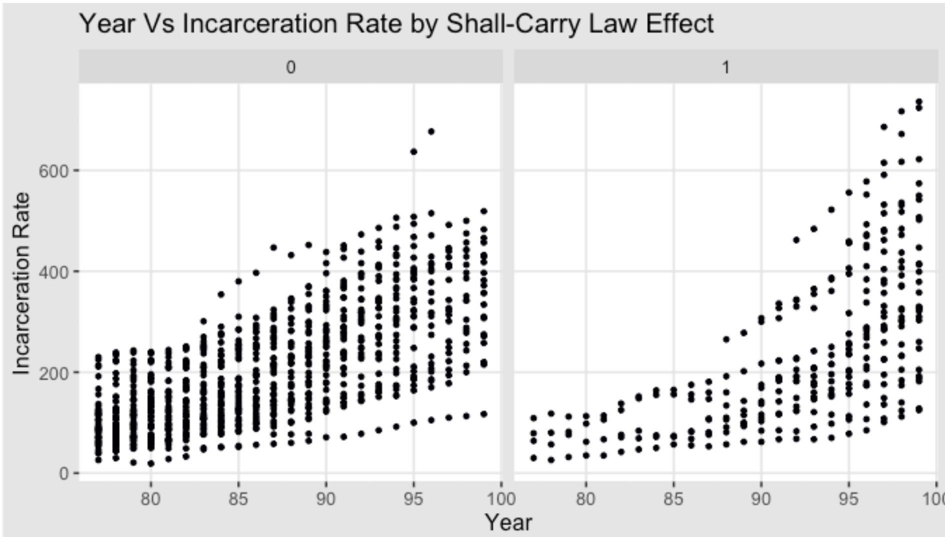


3.2. Incarceration rate:

By analyzing incarceration rate and overall crime rate, we can see that there is a positive correlation between crime rate and incarceration rate. It seems that states with shall law in place have higher incarceration rates.



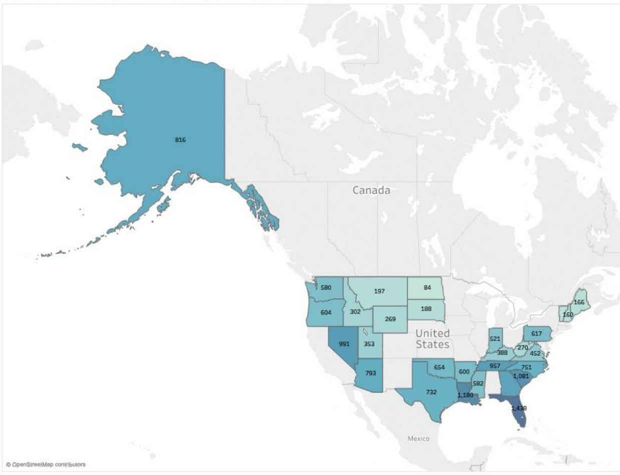
Same pattern of the incarceration rate spiking up in after 1990 where the shall-carry law is in effect can be seen.



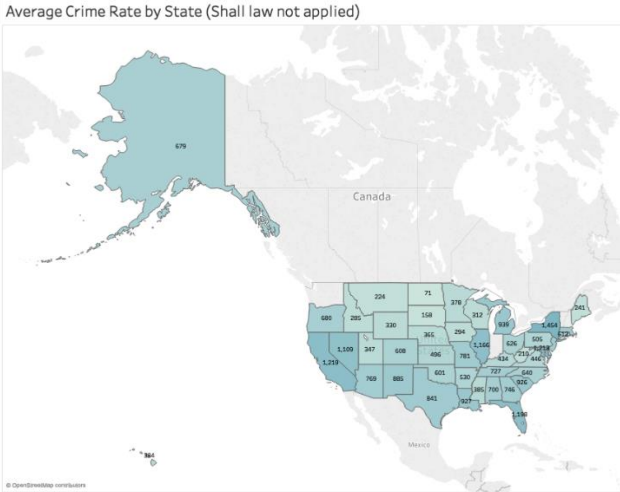
3.3. Shall-issues law:

Regarding shall-issues law, we can see that over 23 years, the total number of times these states adopted shall-carry law is much less than none shall-carry law (285 compared to 1,173). Also, there are some states that always applied shall-carry law throughout 23 years (Washington, New Hampshire, Vermont, ...), some states have never applied shall-carry law and some states (New Mexico, Colorado, Kansas, ...), and some states switch shall-carry law on and off. We will group these states and analyze in the next parts.

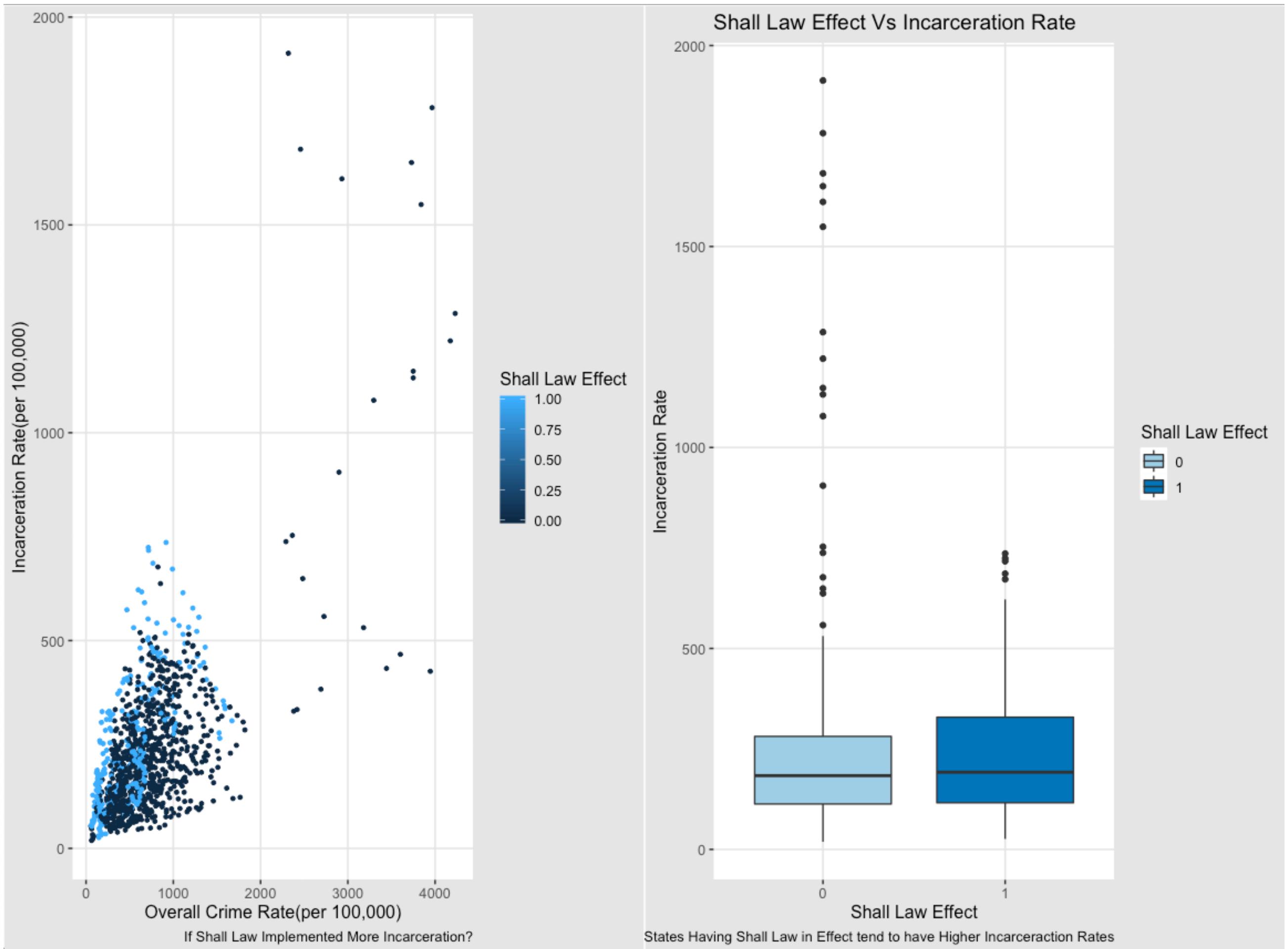
Average Crime Rate by State (Shall law applied)



Average Crime Rate by State (Shall law not applied)

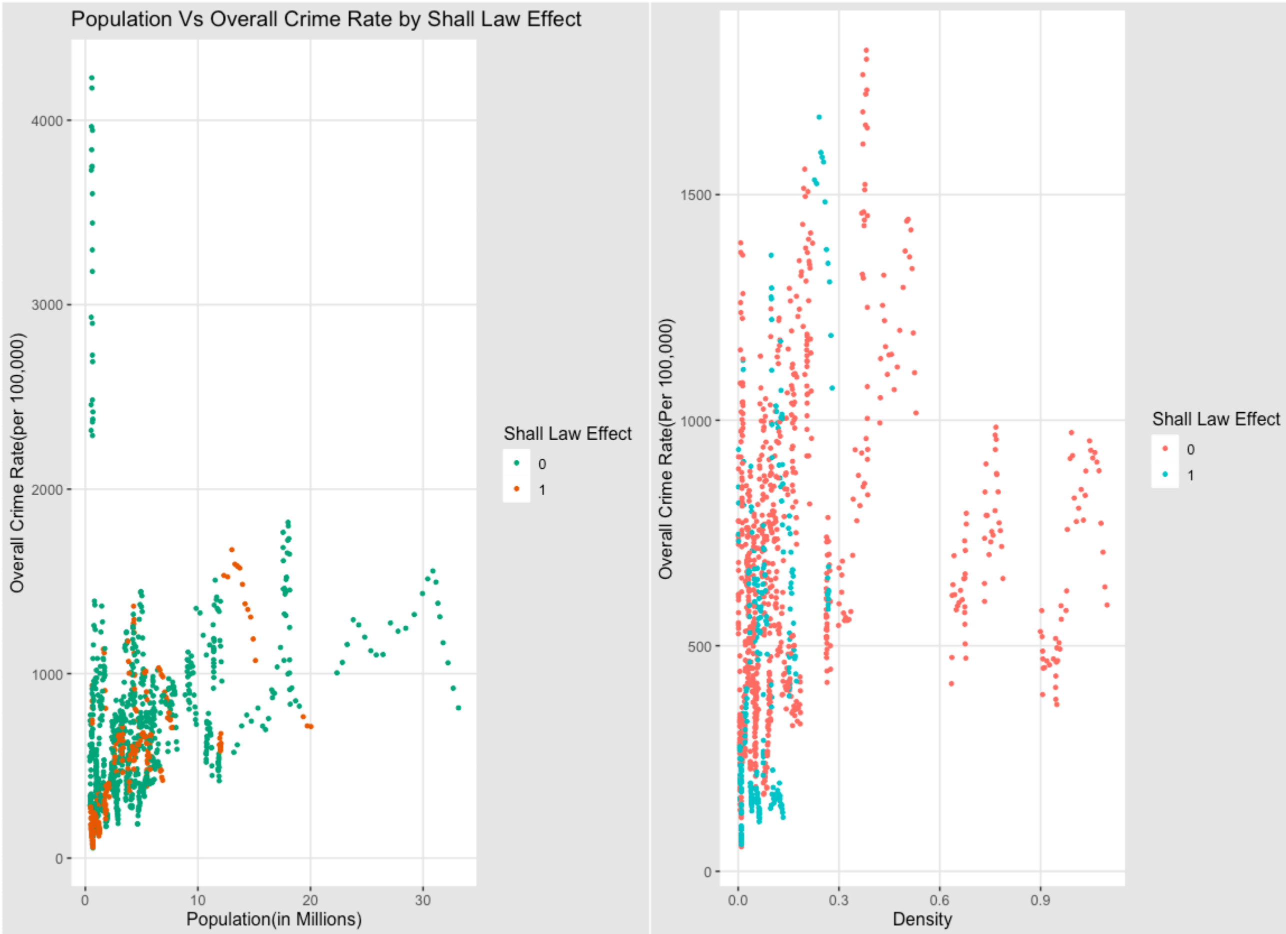


We also want to see if shall law implemented, will there be more incarceration? As we can see from the boxplot, states having shall-carry law in effect tend to have higher incarceration rates and less overall crime rates.

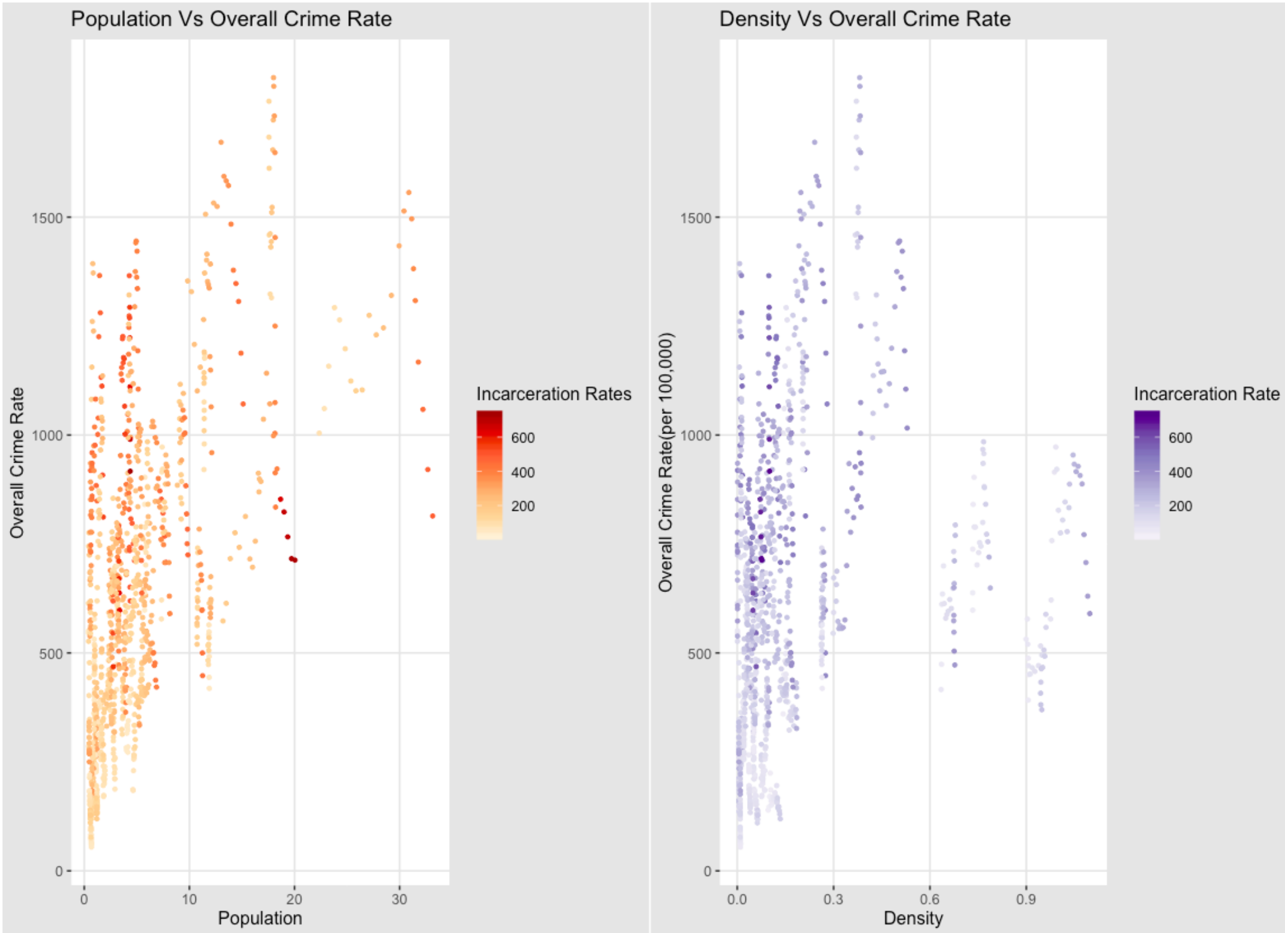


3.4. Density and Population:

Does higher density mean more crime? There is a positive correlation between overall crime and population as well as between density and overall crime rate. Also, densely populated states do not have shall law in effect.

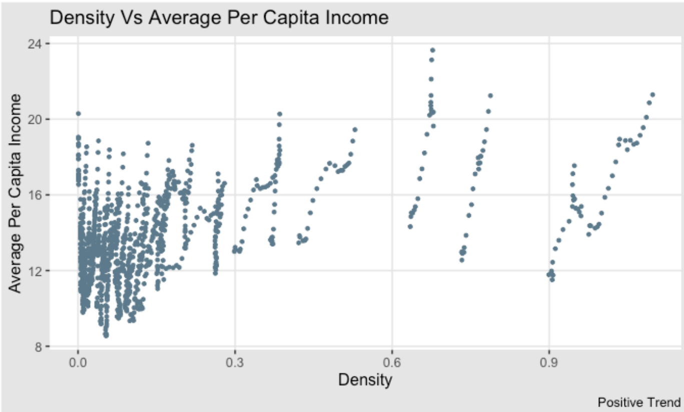


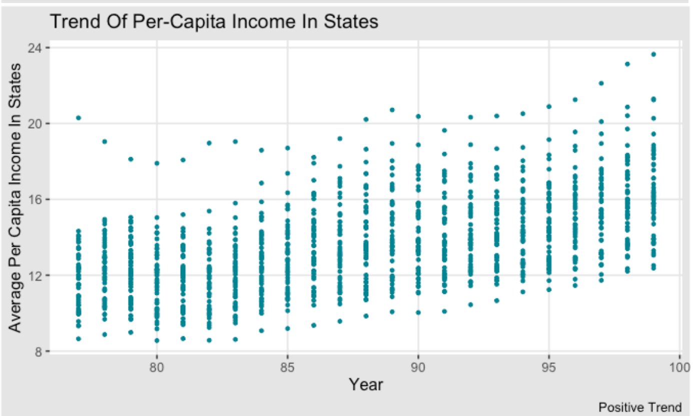
How about incarceration rate? We can see that states having higher population have more overall crime and hence have higher incarceration rates. Also, cities having high densityalso have more crime, hence higher incarceration rates.



3.5. Average Income (avginc):

People who live in states with higher density seems to have higher income. It is also obvious that average per capita income increases over year.





Average per capita of the states are rising with time indicating prosperity as we enter the modern era.

MODELS

Different Models that are apt for panel data framework have been run in this section to understand the significance and magnitude of various explanatory/control variables affecting the crime rate in the US.

We have looked into the dataset and found a very high correlation between the three dependent variables in our case such as **violence rate(vio), murder rate(mur)** and **robbery rate(rob)**, due to such high correlation between these variables we have come to a conclusion that these variables behave pretty much the same way under the influence of the explanatory variables, Therefore we have made an attempt to look at all those several types of crime under a uniform lens by creating another variable **overall crime rate(all_crime)** which captures the overall crime rate in the state, it is basically the summation of **violence rate(vio), murder rate(mur)** and **robbery rate(rob)**. By doing such a thing we get a variable overall crime rate which gives us number of crime per 100,000 individuals.

Our main agenda in this section is to:

1) Understand the effect of shall Carry law on the overall crime rate in the US
2) Understand how the incarceration rate affects the overall crime rate in the US
3) Understand which of the control variables have a significant part in affecting the overall crime rate of the US

To help us carry out the above goals we run **3 models** to estimate the relationship between the **shall carry law**, **incarceration rate** together with other **control variables** as **explanatory variables** with the **dependent variable overall crime rate**. Following three models have been run to estimate the relationship between the dependent and the explanatory variables:

1) The Pooled OLS Model
2) The Entity Fixed Effects Model
3) The Time and Entity Fixed Model

Looking at the distribution of certain variables such as overall crime rate, density and incarceration rate we have tried log transformations on them so that we can get a log-normal distribution for such variables which would give is better model fit which will eventually lead to more precise estimates.

Log transformations made:

Dependent Variables: Overall Crime Rate(all_crime) ➡ ln(all_crime)

Explanatory Variables: Density ➡ ln(density)

Incarceration Rate(incarc_rate) ➡ ln(incarc_rate)

<u>Table Of Expectations</u>

Variable	***Expected sign***	***Explain***
shall	(-)	States having shall-carry law in effect tend to have less overall crime rates.
incarc_rate	(-)	We expect that higher the incarceration rate less should be the overall crime rate
density	(+)	Higher population density increases overall crime rate
avginc	(-)	Higher average income reduces overall crime rate
pop	(+)	More people increase overall crime rate
pm1029	(+)	More the young male population greater should be the overall crime rate
pw1064	(-)	We expect a decrease in overall crime rate with higher population of white people
pb1064	(+)	We expect greater overall crime rate in states with higher population of black people
year	(+)	We expect that the overall crime rate is on the rise over the years

EXPLAIN EACH ONE OF THESE RELATIONSHIPS!!

The Pooled OLS Model

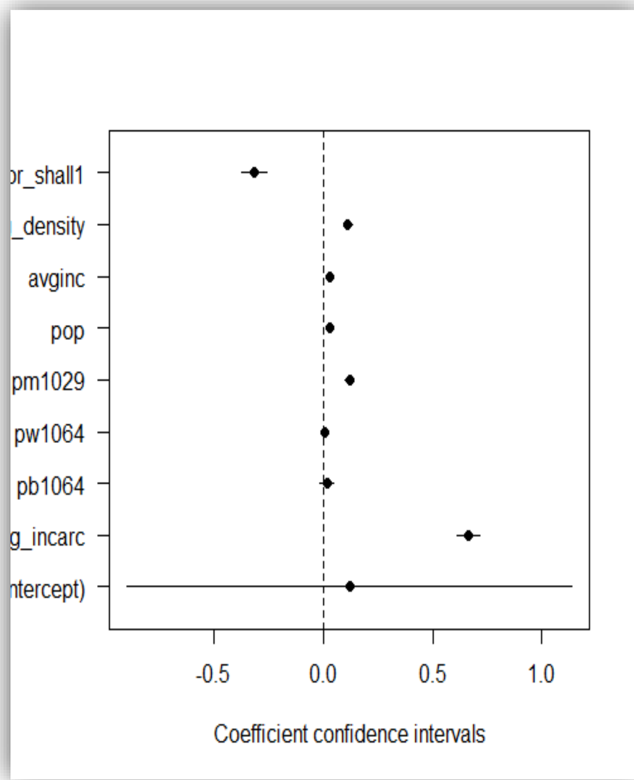
The dataset that we are dealing with is a panel data having 51 entities(states) that have been followed for 23 years hence our first intuition would be to go for the pooled OLS Model. We start with this model because this is the most rudimentary model in case of modelling panel data where the central idea is that the data for different individuals are pooled together, and the equation is estimated using least squares, in this particular case there is no provision for individual differences that might lead to different coefficients.

The following Pooled OLS Model has been estimated:

$$\log(\text{all}_{\text{\_crime}}) = \boldsymbol{\beta_1 + \beta_2 log\_incarc_{it} + \beta_3 pb1064_{it} + \beta_4 pw1064_{it} + \beta_5 pm1029_{it} + \beta_6 pop_{it} + \beta_7 avginc_{it} + \beta_8 + \beta_9 factor\_shall_{it} + \varepsilon_{it}}$$

The Regression output is as follows:

Pooled OLS Model	
Dependent Variable: ln(all_crime)	
(Intercept)	0.12
	(0.52)
log_incarc	0.66 ***
	(0.03)
pb1064	0.02
	(0.02)
pw1064	0.01
	(0.01)
pm1029	0.12 ***
	(0.01)
pop	0.03 ***
	(0.00)
avginc	0.03 ***
	(0.01)
log_density	0.11 ***
	(0.01)
factor_shall1	-0.31 ***
	(0.03)
nobs	1150
r.squared	0.65
adj.r.squared	0.64
statistic	260.68
p.value	0.00
deviance	171.95
df.residual	1141.00
*** p < 0.001; ** p < 0.01; * p < 0.05.	



Significant Variable

- **Factor_shall(shall_law)** at **all** significance level.
- **log_incarc(incarceration rate)** at **all** significance level
- **pm1029(percentage of young males)** at **all** significance level
- **pop(Population)** at **all** significance levels
- **avginc(Average Income)** at **all** significance level
- **log_density(Density)** at **all** significance level

Insignificant Variable

- **pb1064(% population of black people in state)**
- **pw1064(% population of white people in state)**

Interpretation Of The Pooled OLS Model

According to us the pooled OLS is not doing a good job in estimating the parameters in a proper fashion.

We see that the estimate of factor_shall(shall_law) is -0.31 which basically states that the places where the shall carry law is in place have 31% less overall crime rate as compared to the places where the shall law is not in effect on average, everything else kept constant. We feel that this estimate is blown out of proportion as reduction is crime by 31% where shall law seems to be in effect certainly is very large in magnitude in real world sense, and hence we feel that the estimate is certainly **downwardly biased.** The real effect is much smaller than this.

We feel that this bias is due to the unobserved heterogeneity that is hiding in the error term, this omitted variable like **Cultural values** which are different for different entities and are time invariant within entities are correlated with our explanatory variables such as factor_shall leading to a endogeneity problem thus in a process making it downwardly biased.

We expected that as the incarceration rates would go up the overall crime rate must come down(negative relationship), but the positive estimate of 0.66 which is significant at 5% significance level stating that 1% increase in the incarceration rate would lead to 0.66% increase in the overall crime rate, we suspect that this is a typical case of **Simultaneous Causality Bias** in which the states with higher overall crime rates are having more strict laws and better policing which is leading to higher incarceration rates here we can clearly see that there is both the explanatory variable incarceration rate could affect the dependent variable overall crime rate(as our expectation was more incarceration rate less overall crime) but the reversal the overall crime rate can also dictate the level of incarceration rate. Hence we feel that we have not estimated the ln(incar_rate) properly due to Simultaneous Causality Bias.

The estimate of population is consistent with our expectation that with increase in population the overall crime rate would also increase. We found the estimate of population to be 0.03 which states that with every million increase in the population the overall crime rate will increase by 3%. The estimate is significant at all the significance levels.

The estimate of average income is not consistent with our expectations which was that with increase in average income the overall crime rate should drop. But here we see a positive coefficient for the avginc variable which is 0.03 stating that if the per capita personal income is increased by $1000 this would lead to 3% more overall crime rate. We feel that this estimate is most certainly upwardly biased as the estimate dose not match with our expectations nor it makes much sense. We feel that the real estimate is much lower than what we have estimate in the pooled OLS Model. This bias is again due to the unobserved heterogeneity like cultural values of states that are hiding in our error term and is correlated to our explanatory variable and this making it bias.

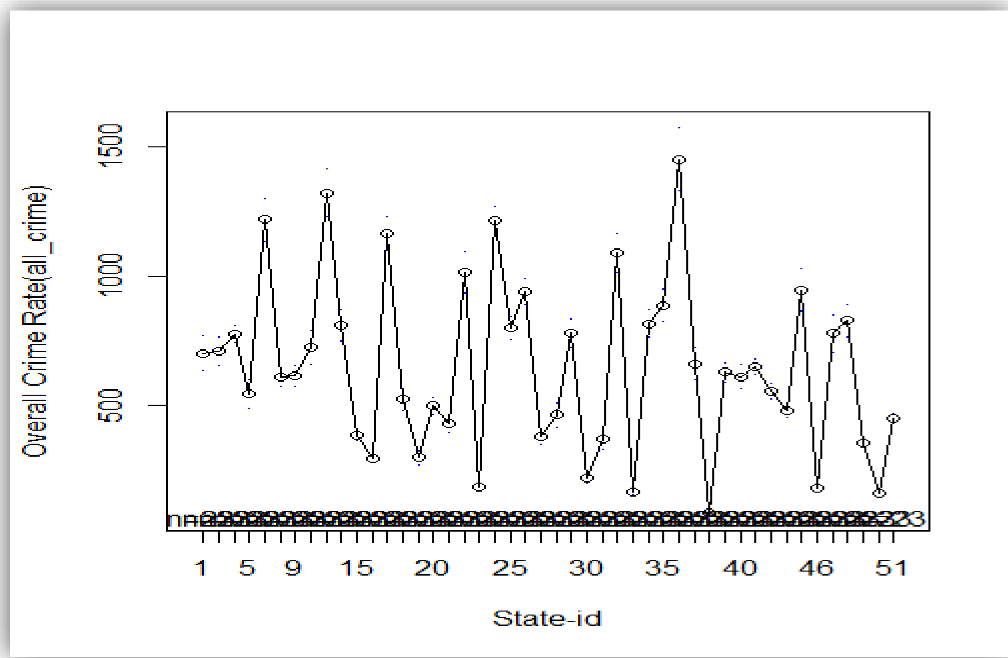
The estimate for density is consistent with our expectation that the increase in population density would lead to higher overall crime rates. Here we get an estimate for ln(density) as 0.11 stating that when the population density increases by 1% it leads to a 11% increase in the overall crime rate.

The estimate for percentage of young males in a state is also according to our expectations as we expect more overall crime when there are higher percentage of young males as males are more violent sex and have higher indulgence in crime which has time and again been established in the scientific literature. We get an estimate of 0.12 which states that if the percentage of young male increase by 1% it would lead to an increase in overall crime rate by 12% on average everything else kept constant, though we feel that this estimate is upwardly biased as we expect a lesser return form this variable. This bias is also due to the unobserved heterogeneity that is causing endogeneity and thus biasing our estimate.

The percentage of white males and black males though 2 important variables are found to be insignificant at significance level of even 10%. Hence we cannot trust the estimates that we receive in this regression output as the estimates are not significantly different than 0.

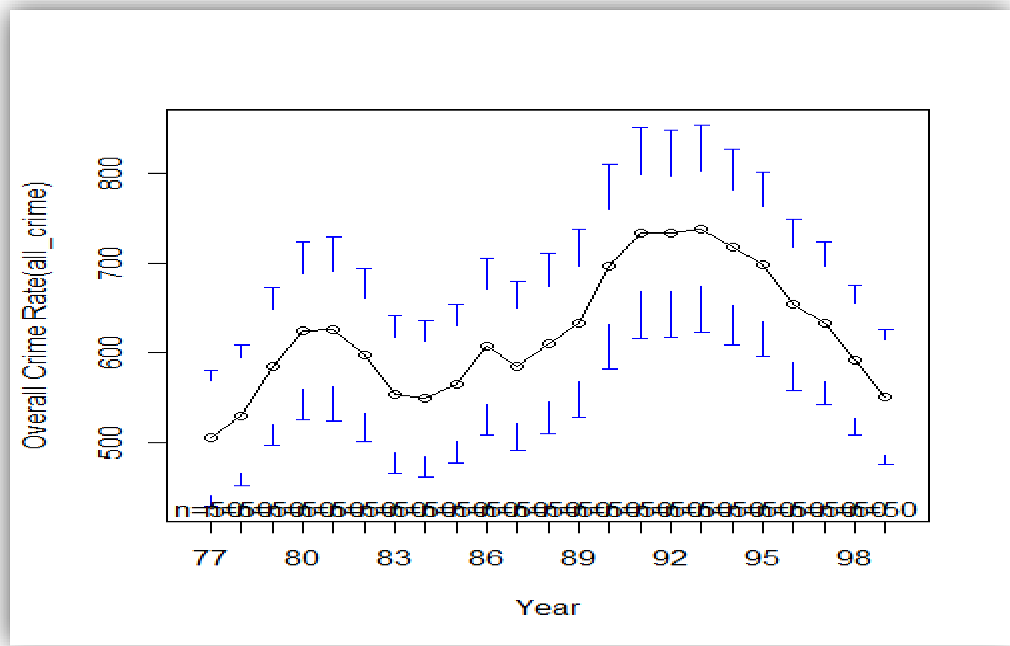
Heterogeneity across States

The Following is the plot of heterogeneity across States



Heterogeneity across Years

A trend of simultaneous increase and decrease of heterogeneity is observed with passing years.



Problem Of Heteroskedasticity and Serially Correlated errors

The dataset in which we are working on is a panel data and hence we think that there might be a case of heteroskedasticity in which the error term is not a constant and it might be different for different time periods the problem of serially correlated errors in which the omitted variables or the unobserved heterogeneity that would be in the error term would be serially correlated for the same entity(states).

Note: The error term for different states won't be correlated as the cultural aspects which we think resides in the error term will be different for different states and will only be correlated within entities.

If both the problems are indeed present in our model this would cause the following problems:

1) The model though still linear unbiased and consistent wont be the best or there would be another estimate which could give us variance lesser than that of least square estimator(Model is not efficient).
2) The Standard Errors that are computed by the least square are wrong or we have biased and inconsistent standard errors.

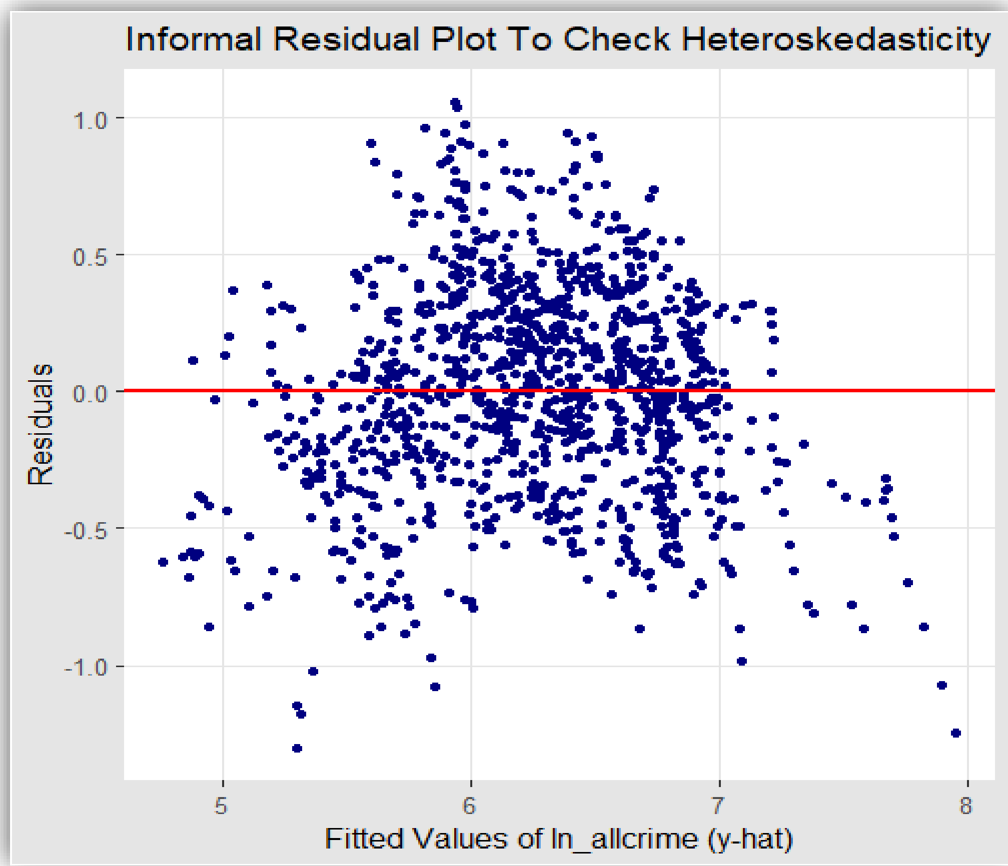
Testing for heteroskedasticity

For testing for weather there is heteroskedasticity in our model we take two approaches:

1) Informal approach: In which we plot the residuals against the fitted values of our dependent variable and look for any kind of pattern, if we find any kind of pattern we say that the informal test shows presence of heteroskedasticity.
2) Formal Whites Test For heteroskedasticity

Informal Approach To Check Heteroskedasticity

In this approach we plot the residuals against the fitted values of our dependent variable and look for any kind of pattern, if we find any kind of pattern we say that the informal test shows presence of heteroskedasticity



Here we can clearly see an inverted U shaped pattern formed by the residuals indicating there is something systematic about the variance of errors that we are not capturing in our model and therefore pointing towards presence of heteroskedasticity.

Formal Whites Test to Test Heteroskedasticity

For this test we assume a functional form for the variance of the error terms which is composed of function of explanatory variables and we regress it over the sum of squared residuals the regression equation looks something like this

$$\hat{e}_i^{\,2} = \alpha_1 + \alpha_2 z_2 + \alpha_3 z_3 + \alpha_4 z_4 + \alpha_n z_n + v_j$$

z's are the functions of the explanatory variables

we have considered all the linear form of all the explanatory variables as z's hence we have 8 z's for 8 explanatory variables.

Null Hypothesis: H_0: $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_n = 0$ (Homoskedasticity)

Alternative Hypothesis: H_1: At least on of the alpha is not equal to 0 (Heteroskedasticity)

We performed the white's test for heteroskedasticity using the Breusch-Pagan test for heteroskedasticity and get the following results:

Studentized Breusch-Pagan Test

statistic	p.value	parameter	method
84.48342	0	8	studentized Breusch-Pagan test

From the above test we get a very high statistic of 84.48 and a p-value of 0 hence we can easily reject the null hypothesis and conclude that there is indeed heteroskedasticity in our model.

These Heteroskedasticity and Serial Correlation are making our pooled least square estimator inefficient and with biased and inconsistent standard errors.

Pooled OLS With Cluster Robust Standard Errors

To make our standard errors correct so that we can use the standard errors for the purpose of statistical inference such as constructing confidence intervals and doing hypothesis testing we make use of cluster robust standard for the same Pooled OLS Estimator.

Generally we find that the standard errors that are computed by the cluster Robust standard errors are larger than that of the pooled OLS or we can say that the pooled OLS estimator overestimates the precision of the estimates.

The results of applying cluster robust standard errors on the pooled OLS estimator are as follows:

Pooled OLS With Cluster Robust Standard Errors

term	estimate	std.error	statistic	p.value
(Intercept)	0.1202039	2.1674206	0.0554594	0.9557822
log_incarc	0.6642183	0.0976091	6.8048786	0.0000000
pb1064	0.0161388	0.0577291	0.2795610	0.7798650
pw1064	0.0084857	0.0296319	0.2863695	0.7746471
pm1029	0.1200456	0.0297680	4.0327082	0.0000588
pop	0.0301421	0.0095642	3.1515684	0.0016662
avginc	0.0270854	0.0197379	1.3722573	0.1702531
log_density	0.1112226	0.0385331	2.8864145	0.0039700
factor_shall1	-0.3143095	0.0927582	-3.3884811	0.0007268

Comparing this to the pooled OLS estimates

Pooled OLS Model

	Estimate	Std. Error	t-value	Pr(>|t|)	
(Intercept)	0.1202039	0.5182524	0.2319	0.8166255	
log_incarc	0.6642183	0.0273128	24.3190	< 2.2e-16	***
pb1064	0.0161388	0.0156767	1.0295	0.3034729	
pw1064	0.0084857	0.0077561	1.0941	0.2741555	
pm1029	0.1200456	0.0111947	10.7235	< 2.2e-16	***
pop	0.0301421	0.0026111	11.5439	< 2.2e-16	***
avginc	0.0270854	0.0070720	3.8299	0.0001351	***
log_density	0.1112226	0.0100218	11.0980	< 2.2e-16	***
factor_shall1	-0.3143095	0.0298381	-10.5338	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

After using the cluster robust standard errors our estimator is still inefficient only the standard errors that are computed are correct.

Indeed the standard errors that we get for the pooled OLS with cluster robust standard errors are more than that of the pooled OLS estimator

We see that one of the estimates are no longer significant in the new pooled OLS model with cluster robust standard errors.

We find that the variable average income which was not consistent with our expectations is no longer significant.

We still feel this model is not the best model because we do not get rid of the unobserved heterogeneity form the model which are biasing our estimates of incarceration rates, shall law and percentage of young males in a state to name a few. Hence, we have to come up with a model that can get us the rid of the unobserved heterogeneity that would get us rid of the endogeneity problem so our estimates would become unbiased and consistent and its interpretation would make sense.

As biased and inconsistent estimates are of no use to us and this bias is also not going away with increase in the sample size, so we have to find another way to get rid of this particular problem of endogeneity.

Therefore, we conclude that pooled OLS estimator with cluster robust standard errors is still an unreliable estimator.

Entity Fixed Model

The entity fixed model accounts for deviations from the mean, the coefficient estimates depend only on the variation of the dependent and explanatory variable within individuals. We decided to use the Entity Fixed Model which accounts for unobserved heterogeneity that is time invariant or is constant over time but varies between states.

Our assumption in doing so being the individual characteristics of one state is not correlated with other states and the difference between states is fixed.

We think this assumption might be reasonable because the US states are culturally very diverse and we think each is different from other in a constant fashion.

The fixed effects model is less efficient than pooled and random effects model because we only take advantage of variation within entities(states) and not the variation between entities(states) hence we are limiting the amount of information we are making use of to estimate the model and hence our entity fixed model is less efficient.

One disadvantage of fixed effects model is that the time invariant variables cannot be estimated by the model and are not any more in the model or the variables that are not varying significantly in the model would not be properly estimated because we are not having significant changes in the x side to interpret the variation in the dependent variable y.

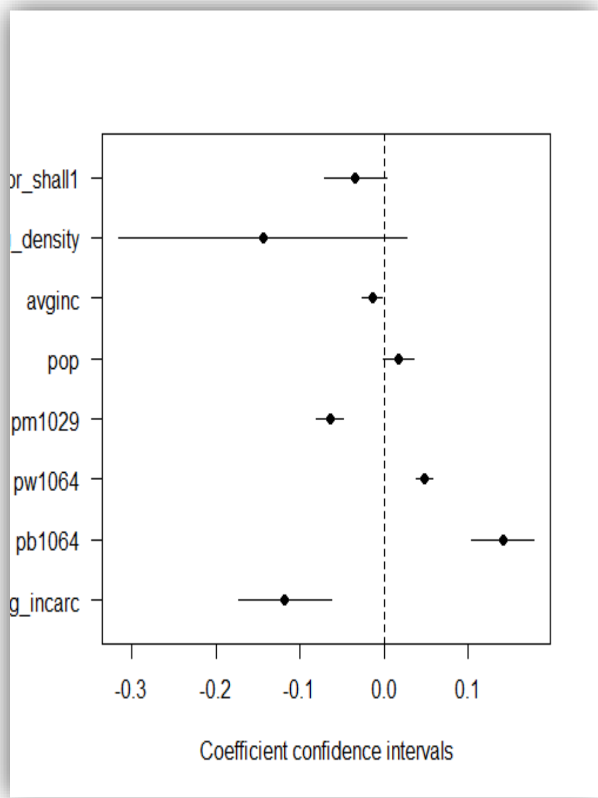
Though the entity fixed model removes most of the causes of endogeneity such as the variables that vary between entities but are constant with time. But it certainly can’t deal with the variables that could be causing endogeneity

The regression Equation that we estimate in the entity fixed model is as such:

$$\begin{aligned}\log\left(all_{crime_{it}}\right) &= \beta_{1i} + \beta_2 \log\_incarc_{it} + \beta_3 pb1064_{it} + \beta_4 pw1064_{it} + \beta_5 pm1029_{it} \\ &+ \beta_6 pop_{it} + \beta_7 avginc_{it} + \beta_8 \log\_density_{it} + \beta_9 factor\_shall_{it} + \varepsilon_{it}\end{aligned}$$

The regression output for the entity fixed model is as follows:

Entity Fixed Model	
Dependent Variable ln(all_crime)	
log_incarc	-0.12 ***
	(0.03)
pb1064	0.14 ***
	(0.02)
pw1064	0.05 ***
	(0.01)
pm1029	-0.06 ***
	(0.01)
pop	0.02
	(0.01)
avginc	-0.01 *
	(0.01)
log_density	-0.14
	(0.09)
factor_shall1	-0.03
	(0.02)
nobs	1150
r.squared	0.17
adj.r.squared	0.13
statistic	28.47
p.value	0.00
deviance	27.21
df.residual	1092.00
*** p < 0.001; ** p < 0.01; * p < 0.05.	



Significant Coefficients

- Incarceration Rate(ln(incar)) at all significance levels
- % population of black people in state at all significance levels
- % population of white people in state at all significance levels
- % population of young male in state at all significance levels
- Population(pop) in the state at 10% significance levels
- Average per capita Income(avginc) in the state at 5% significance levels
- Shall Law effect at 10% significance levels

Insignificant Coefficients

- Population Density(log_density) is not significant

Interpretation Of The Fixed Effects Model

We notice the unobserved heterogeneity that were biasing our estimates have been more or less been removed and hence the endogeneity in the model has been removed to a great extent and the estimates that we receive for the variables are unbiased and consistent.

We see that the coefficient of factor_shall is no longer significant as it estimate has become unbiased and consistent the current estimate is -0.03 form the previous estimate of -0.31 we can clearly see that the downward bias has been removed by controlling for the unobserved heterogeneity but we cannot reject the null hypothesis and say that this estimate of -0.03 is significantly different than 0. Hence, The shall law effect, is found to be insignificant in affecting the overall crime rate. However the estimate of -0.03 stated that the places where that shall law is in effect have 3% less crime as compared to the states that do not have shall law in place on average, everything else kept constant.

According to our expectations the simultaneous causality bias that was affecting the incarceration rate has been successfully removed and now we get a negative estimate for incarceration rate(ln(incar)) which is -0.12 which states that the overall crime rates fall be 0.12% if the incarceration rate increases by 1% on average everything else kept constant. Now this estimate is consistent with our expectation. We also find that this estimate is very significant at particularly every significance level.

According to our expectation we get a strong positive relationship between percentage of black people in the state and the overall crime rate. The estimate is also consistent with our assumption . The estimate that we are getting for this variable is of 0.14 which says that with 1% increase in percentage of black people in the state will increase the overall crime rate by 14%. This estimate is also very significant at particularly any significance level.

The estimate for percentage of white people is however not consistent with our expectations we thought that with higher amount of white people in a state will lead to lesser overall crime but this is in fact not the case we are getting an estimate as 0.05 which states the overall crime rate will increase by 5% for every 1% increase in the white population of white males in the state on average everything else kept constant. This estimate is found to be highly significant at almost very significance level, however we feel that with increase in the % of white people also the population is increasing and its that particular effect that is leading to the spike in the overall crime rate.

We get a negative relationship of % population of young male in the state and overall crime rate this is quite opposite to both our expectation and the pooled OLS model the estimate of -0.06 states that with 1% increase in the population of young male in the state the overall crime rate seems to decrease by 6%. The estimate is found to be significant at all significance level. However, we feel that this estimate is downwardly biased as young men are scientifically proven to being more violent and have significantly more participation in crimes and so we feel that maybe there are certain variables that are constant for all the states but changing with the time are correlated with this variable and causing the bias, such variables could be recession, federal laws etc. that would affect the overall crime rate.

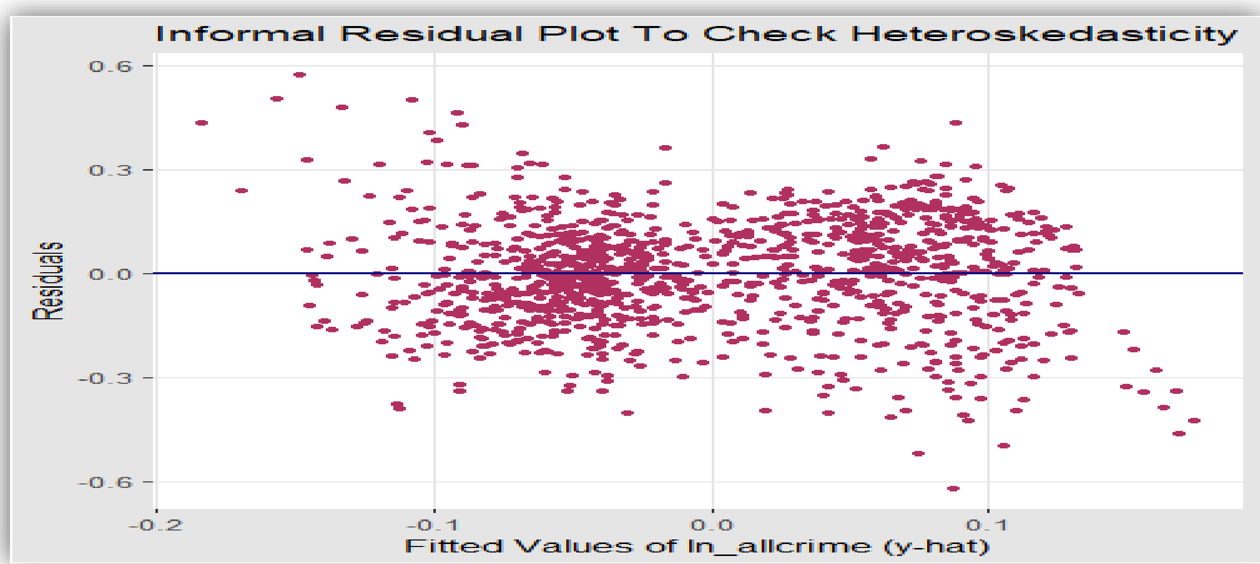
The estimate density (log(density)) is having a negative relationship which is opposite to what we had expected and also opposite to the pooled OLS model, however, the model is highly insignificant hence we could say that we could not estimate this particular variable properly. Our guess would be that as the population density not the population but population density is not changing dramatically in a state through the years and due to this we could not exploit this change and see the overall change in overall crime rate. Hence we can say that the variable was not properly estimated.

The population is only significant for the significance level of 10% but its estimate is according to both our expectations and the one we got for the Pooled OLS Model, however the estimate that we had received for the pooled OLS model was upwardly biased. The new estimate of 0.0172 states that with 1 million increase in the population the overall crime rate increases by 1.7%.

In the pooled OLS model we were getting an upwardly biased estimate for the average income variable which was also against our expectation as the relationship was positive and we had expected that the average per capita income to have a negative relationship with the overall crime rate. However, after applying the entity fixed effects model the bias seems to have been disappeared and now the estimate of -0.0140 is consistent with our expectations. The estimate is significant at 5% significance level.

We feel that we have received satisfactory estimates for the entity fixed model but we think that not all the heterogeneity seems to have been removed there could be some variables such as recession, federal laws affecting guns etc. that are actually constant for all the entities but vary with time. These variables are still in the error term as the entity fixed effects did not remove such kind of heterogeneity and we speculate these time variant and entity constant variables which now in error term could be correlated with some of our explanatory variables and could still be biasing our estimates.

We also see that plotting residual plot against fitted values of dependent variable overall crime rate(ln_all_crime) the heteroskedasticity and the seriel correlation has been removed by the Entity fixed effects model(no pattern exhibited by the residuals, no inverted u pattern).



Time And Entity Fixed Model

For eliminating the problems that are caused due to the unobserved heterogeneity such as those that are being caused by the variables such as recession and federal laws that affect the crime rate and which are constant for all the entities and change with time that are causing endogeneity in the model and thus could be causing biased estimates. To deal with such a model we come up with the Time and entity fixed model. This model also includes the time effects by adding the effect of time on the dependent variable overall crime rate by including the time dummies. This will capture the effect of time progression in years on the dependent variable overall crime rate(ln_all_crime)

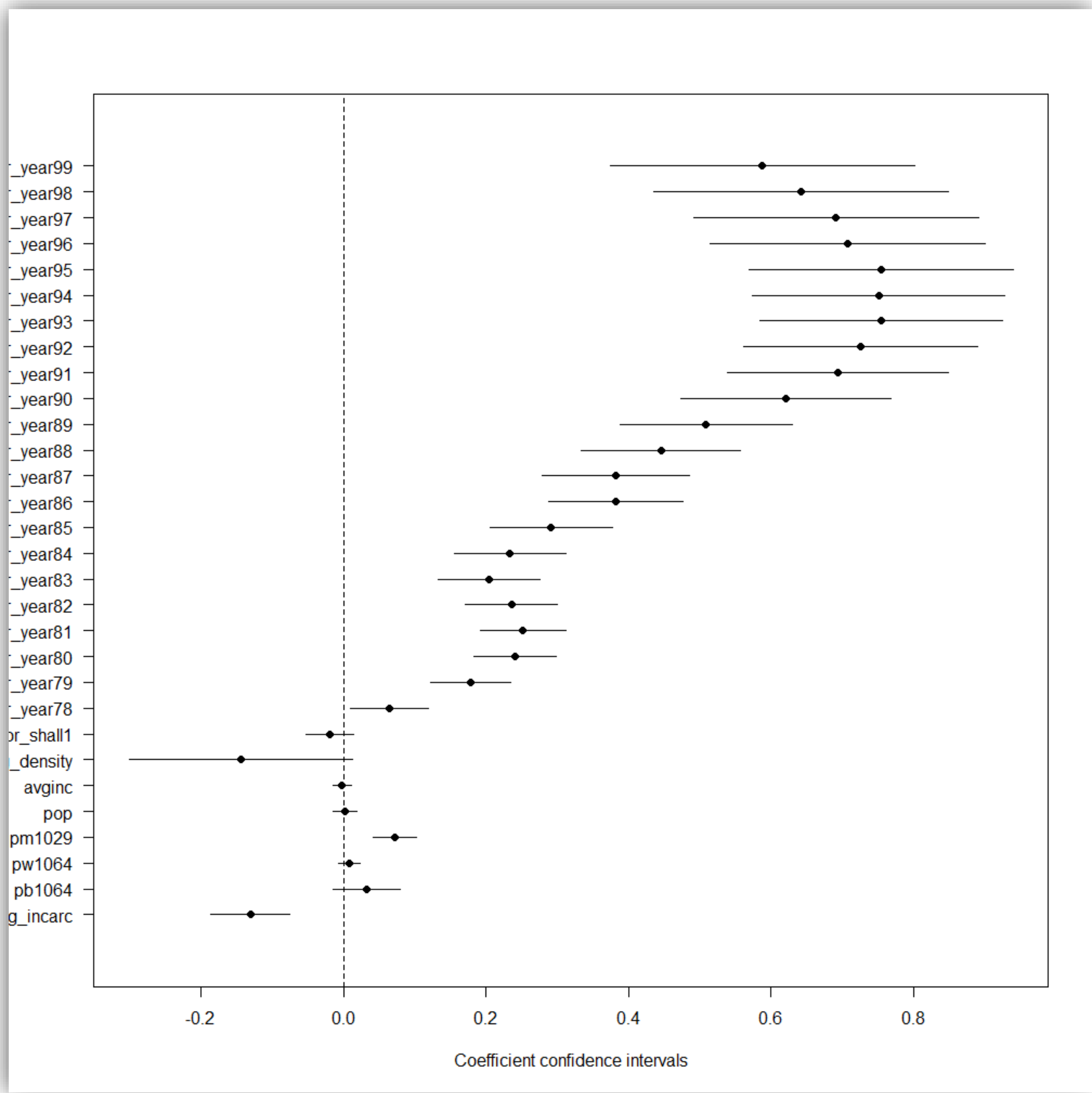
By adding the time dummies and including the time fixed effects along with the entity fixed we also get rid of the unobserved heterogeneity that are caused by variables in the error term that are constant for all entities and vary with time and are correlated with our explanatory variable and are causing endogeneity in our model.

The regression equation for Time and Entity fixed model looks something like this:

$$\begin{aligned}\log\left(all_{crime_{it}}\right) &= \beta_0 + \beta_1 \log\_incarc_{it} + \beta_2 pb1064_{it} + \beta_3 pw1064_{it} + \beta_4 pm1029_{it} \\ &+ \beta_5 pop_{it} + \beta_6 avginc_{it} + \beta_7 \log\_density_{it} + \beta_8 factor\_shall_{it} \\ &+ \delta_2 factor\_year78_{it} + \delta_3 factor\_year79_{it} + \delta_4 factor\_year80_{it} \\ &+ \delta_5 factor\_year81_{it} + \delta_6 factor\_year82_{it} + \delta_7 factor\_year83_{it} \\ &+ \delta_8 factor\_year84_{it} + \delta_9 factor\_year85_{it} + \delta_{10} factor\_year86_{it} \\ &+ \delta_{11} factor\_year87_{it} + \delta_{12} factor\_year88_{it} + \delta_{13} factor\_year89_{it} \\ &+ \delta_{14} factor\_year90_{it} + \delta_{15} factor\_year91_{it} + \delta_{16} factor\_year92_{it} \\ &+ \delta_{17} factor\_year93_{it} + \delta_{18} factor\_year94_{it} + \delta_{19} factor\_year95_{it} \\ &+ \delta_{20} factor\_year96_{it} + \delta_{21} factor\_year97_{it} + \delta_{22} factor\_year98_{it} \\ &+ \delta_{23} factor\_year99_{it} + u_{it}\end{aligned}$$

The regression output for the time fixed effect model are as follows

Time And Entity Fixed Model	
Dependent Variable: ln(all_crime)	
log_incarc	-0.13 ***
	(0.03)
pb1064	0.03
	(0.02)
pw1064	0.01
	(0.01)
pm1029	0.07 ***
	(0.02)
pop	0.00
	(0.01)
avginc	0.00
	(0.01)
log_density	-0.14
	(0.08)
factor_shall1	-0.02
	(0.02)
factor_year78	0.06 *
	(0.03)
factor_year79	0.18 ***
	(0.03)
factor_year80	0.24 ***
	(0.03)
factor_year81	0.25 ***
	(0.03)
factor_year82	0.24 ***
	(0.03)
factor_year83	0.20 ***
	(0.04)
factor_year84	0.23 ***
	(0.04)
factor_year85	0.29 ***
	(0.04)
factor_year86	0.38 ***
	(0.05)
factor_year87	0.38 ***
	(0.05)
factor_year88	0.45 ***
	(0.06)
factor_year89	0.51 ***
	(0.06)
factor_year90	0.62 ***
	(0.08)
factor_year91	0.69 ***
	(0.08)
factor_year92	0.73 ***
	(0.08)
factor_year93	0.75 ***
	(0.09)
factor_year94	0.75 ***
	(0.09)
factor_year95	0.75 ***
	(0.09)
factor_year96	0.71 ***
	(0.10)
factor_year97	0.69 ***
	(0.10)
factor_year98	0.64 ***
	(0.11)
factor_year99	0.59 ***
	(0.11)
nobs	1150
r.squared	0.38
adj.r.squared	0.33
statistic	21.86
p.value	0.00
deviance	20.39
df.residual	1070.00
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.	



_year99
_year98
_year97
_year96
_year95
_year94
_year93
_year92
_year91
_year90
_year89
_year88
_year87
_year86
_year85
_year84
_year83
_year82
_year81
_year80
_year79
_year78
r_shall1
_density
avginc
pop
pm1029
pw1064
pb1064
g_incarc
-0.2
0.0
0.2
0.4
0.6
0.8
Coefficient confidence intervals

Significant Coefficients:

- Incarceration rate(ln(incar)) is significant for any significance level
- % population of young male in state at all significance levels
- Year 78 is significant at 5 % significance level
- Year 79 – Year 99 are significant at all significance levels

Insignificant Coefficients

- % population of black people in state
- % population of white people in state
- Population
- Average per capita Income
- Density(log_density)
- Shall law(factor_shall)

Interpretation of time and entity fixed model

The interpretation of the shall law effect is still insignificant or has no effect on the overall crime rate this is consistent with our estimate that we got for the entity fixed model but the time and entity fixed model shows even higher evidence for its insignificance. The estimate that we receive for it is -0.02 which states that the states having shall law in effect have 2% less crimes than the ones who don't have shall law in effect however this estimate is highly insignificant as we cannot reject the null hypothesis that the estimate is significantly different than 0. Hence, we can say that shall law has no effect on the overall crime rate.

The incarceration rate(log(incar)) is found to be highly significant and as well as consistent with our expectations the estimate that we get for this variable is -0.13 which states that 1% increase in the incarceration rate will lead to 13% drop in the overall crime rates this estimate is 1% more than that of the entity fixed model its estimate being -0.12. The variable is found to be highly significant at every possible significance levels. This estimate also takes care of the simultaneous causality bias that we saw in the pooled OLS model in which saw that with increase in the incarceration rate the overall crime rate and hence it never made any sense. But now as almost all the unobserved and observed heterogeneity has been taken care of we can see the true effect of incarceration rate on overall crime rate which is negative and consistent with our expectation.

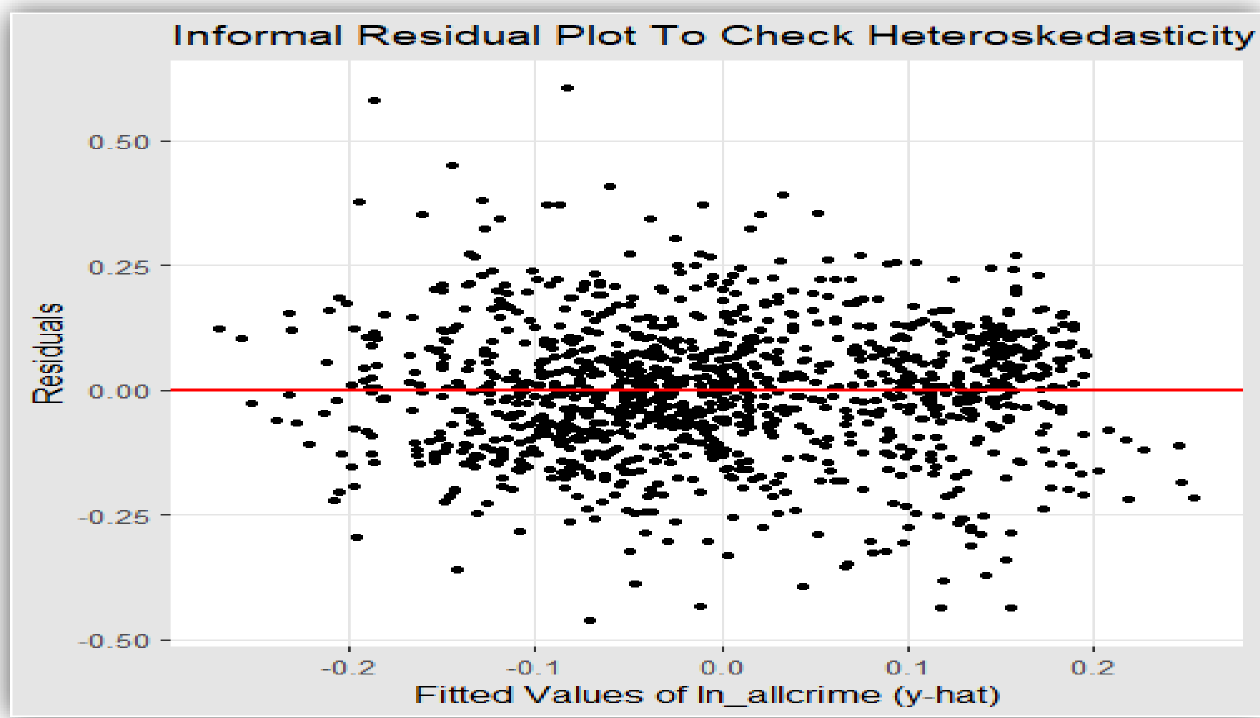
The variables % population of black people in state and % population of white people in state have both found to be insignificant or do not have any effect on the overall crime rate with coefficients of 0.03 and 0.01 respectively. We could not reject the null hypothesis and hence the estimates are not significantly different than 0. These two variables were found to be significant in the entity fixed model hence we can now say that the estimate of the entity fixed model was upwardly biased for both these variables and the actual effect that has been estimated by the time and entity fixed model is not significantly different than 0. This biased was removed by the time and entity fixed my further accounting for the unobserved heterogeneity that was due to variables in the error term and correlated with the explanatory variables that varied with time but were constant across entities like recession and federal laws for guns etc.

Like expected the % population of young male is found to be very significant and consistent with our expectation of having a positive relationship. This estimate was found to be downwardly biased in our entity fixed model but on removal of further unobserved heterogeneity that are caused by variables that are in the error term that are constant for entities but vary with time we got rid of the endogeneity that it was causing and hence got unbiased estimates. The estimate that we got for the time and entity fixed model is 0.07 stating with increase in young male population the overall crime rate increases by 7% on average everything else kept constant.

The estimates of population, average per capita income and density(log_density) have been found to be insignificant in affecting the overall crime rate the estimates of population and average per capita income is almost 0 and that of density(log_density) is -0.14 this estimate dose is not consistent with our expectation. But we cannot reject the null hypothesis of all these estimates hence we can say these estimates are not significantly different than 0 and are hence insignificant.

We see a trend of increase in the overall crime from the year 1977 to the year 1995 but after these years towards the end we do see slightly decreasing trend of the overall crime rate. All the time dummies are very significant and hence do affect the overall crime rate.

We also see that plotting residual plot against fitted values of dependent variable overall crime rate(ln_all_crime) the heteroskedasticity and the seriel correlation has been removed by the Time and Entity fixed effects model(no pattern exhibited by the residuals, no inverted u pattern).



Comparing between models and Choosing the best model

Now we must compare between the models by performing various statistical tests and choose the best possible model for our estimation process.

Before we do that here is the estimates that we receive for all the 3 models under one roof this will help us in comparing between the models.

Comparing Models

	Dependent variable:		
	log_all_crime		
	POOLED OLS	ENTITY FIXED	TIME AND ENTITY FIXED EFFECTS
	(1)	(2)	(3)
log_incarc	0.7***	-0.1***	-0.1***
	(0.03)	(0.03)	(0.03)
pb1064	0.02	0.1***	0.03
	(0.02)	(0.02)	(0.02)
pw1064	0.01	0.05***	0.01
	(0.01)	(0.01)	(0.01)
pm1029	0.1***	-0.1***	0.1***
	(0.01)	(0.01)	(0.02)
pop	0.03***	0.02*	0.002
	(0.003)	(0.01)	(0.01)
avginc	0.03***	-0.01**	-0.002
	(0.01)	(0.01)	(0.01)
log_density	0.1***	-0.1	-0.1*
	(0.01)	(0.1)	(0.1)
factor_shall1	-0.3***	-0.03*	-0.02
	(0.03)	(0.02)	(0.02)
factor_year78			0.1**
			(0.03)
factor_year79			0.2***
			(0.03)

factor_year80			0.2***
			(0.03)
factor_year81			0.3***
			(0.03)
factor_year82			0.2***
			(0.03)
factor_year83			0.2***
			(0.04)
factor_year84			0.2***
			(0.04)
factor_year85			0.3***
			(0.04)
factor_year86			0.4***
			(0.05)
factor_year87			0.4***
			(0.1)
factor_year88			0.4***
			(0.1)
factor_year89			0.5***
			(0.1)
factor_year90			0.6***
			(0.1)
factor_year91			0.7***
			(0.1)
factor_year92			0.7***
			(0.1)
factor_year93			0.8***
			(0.1)
factor_year94			0.8***
			(0.1)
factor_year95			0.8***
			(0.1)
factor_year96			0.7***
			(0.1)
factor_year97			0.7***
			(0.1)
factor_year98			0.6***
			(0.1)
factor_year99			0.6***
			(0.1)
Constant	0.1		
	(0.5)		
Observations	1,150	1,150	1,150
R2	0.6	0.2	0.4
Adjusted R2	0.6	0.1	0.3
F Statistic	260.7*** (df = 8; 1141)	28.5*** (df = 8; 1092)	21.9*** (df = 30; 1070)
Note:			*p<0.1; **p<0.05; ***p<0.01

Comparing Entity Fixed Model With Time And Entity Fixed Model

From the above output first we would like to compare the entity fixed model and the time and entity fixed model to do this we will conduct a simple F-Test or chi squared test in which we will assume that our null hypothesis be that all the coefficient estimates of the time dummies are equal to zero and the alternative that at least one of the coefficient of the time dummies is significant.

If we do not reject the null hypothesis we would mean that the time effects are irrelevant and hence should not be In our model and hence we would choose the Entity fixed model over the time and entity fixed model.

If we reject the null hypothesis it would mean that at least one of the coefficient of the time dummies is significant and hence we would prefer to add time dummies in our model and choose time and entity fixed model over the entity fixed model.

Null Hypothesis: H_0 : $\boldsymbol{\delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 = \delta_8 = \delta_9 = \delta_{10} = \delta_{11} = \delta_{12} = \delta_{13} = \delta_{14} = \delta_{15} = \delta_{16} = \delta_{17} = \delta_{18} = \delta_{19} = \delta_{20} = \delta_{21} = \delta_{22} = \delta_{23} = 0}$

Alternative Hypothesis: H_1 : At least one of them is not equal to zero

The results of the test are as follows:

```
Hypothesis:
factor_year78 = 0
factor_year79 = 0
factor_year80 = 0
factor_year81 = 0
factor_year82 = 0
factor_year83 = 0
factor_year84 = 0
factor_year85 = 0
factor_year86 = 0
factor_year87 = 0
factor_year88 = 0
factor_year89 = 0
factor_year90 = 0
factor_year91 = 0
factor_year92 = 0
factor_year93 = 0
factor_year94 = 0
factor_year95 = 0
factor_year96 = 0
factor_year97 = 0
factor_year98 = 0
factor_year99 = 0

Model 1: restricted model
Model 2: log_all_crime ~ log_incarc + pb1064 + pw1064 + pm1029 + pop +
    avginc + log_density + factor_shall + factor_year

  Res.Df Df  Chisq Pr(>Chisq)
1   1092
2   1070 22 358.01  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We get a very high chi squared statistic of 358.01 and a p-value of practically 0 hence we can easily reject the null hypothesis in favor of the alternative stating that we have found significant evidence for relevant time effects in our model and hence we will choose the Time and entity fixed model as our preferred model when compared to the entity fixed model.

Comparing Time And Entity Fixed Model With Pooled OLS Model

To compare the time and entity fixed model with the pooled OLS Model we need to another statistical test which is the F-test for individual effects. In this test we want to see weather there is individual fixed effects or weather there should be individual intercepts for the entities.

We want to test the definition of the fixed effect model we want to see weather the entities differ form each other in a fixed way or not.

The null hypothesis of this test would be all the i intercepts for the i entities are the same which basically means that there is no fixed effect for different entities. In this particular case if we are unable to reject the null hypothesis we will conclude that the Pooled OLS Model is a better model because there is no fixed effect difference in entities was observed or in simpler terms there is only 1 intercept term and not n intercept term.

The alternative hypothesis of this test will be that there will be i different intercepts for i different entities which basically means that if we reject the null hypothesis then we can say that there is indeed individual effects or the slopes for every entity is different and hence fixed effects will be our model of choice

The null and alternative hypothesis will be as follows

Null Hypothesis: H_0 : $\boldsymbol{\beta_{1,1} = \beta_{1,2} = \beta_{1,3} = \beta_{1,4} = \cdots = \beta_{1,i}}$ **(Pooled OLS)**

Alternative Hypothesis H_1 : At least one of the above betas is different (Fixed Effect)

The output of the following test is the following

```
F Test For Individual Effects
| df1|  df2| statistic| p.value|method                         |alternative        |
|---:|----:|---------:|-------:|:------------------------------|:------------------|
|  71| 1070|  112.0454|       0|F test for individual effects  |significant effects |
```

We see that foe the test we get a very high t-statistic of 112.04 and a p-value of 0 so we can easily reject the null hypothesis in favor of the alternate and conclude that there is evidence for significant individual effects which means there is evidence for the fixed effects or simply different intercepts for different entities.

Hence we reject the null hypothesis and conclude that our model of choice would still be Time and Entity Fixed Model.

Is Random Effects Model Needed?

According to the assumption for the intercept term it is only sensible to use the random effects model for a sample that is randomly selected from a population. If the case is so then using random effects makes more sense as it can also estimate time invariant variables as well as is the best linear unbiased estimator(efficient)

But as we know we are not dealing with a sample here but here we have accounted for the whole population in the dataset which is all the 51 states of United States of America hence according to economic theory using a random effect model dose not make any sense here and as we are dealing with the entire population we will prefer the Time And Entity Fixed Effects Model as our final model.

Hence, The Final model of choice for estimating the overall crime rate would be **Time and Entity Fixed Model.**

Conclusion

While working with the dataset multiple problems were encountered like heteroskedasticity, serially correlated errors, endogeneity etc. After the exploration process we could understand the relationship between all the variables with the dependent variable overall crime rate and with other variables. After gaining several insights we utilized the knowledge that we gained to come up with several models that would help us understand the real world relation between our dependent and independent variables. We used Pooled OLS Model, Entity Fixed Model and Time and entity Fixed Model in our analysis. We tried to go from one model to another with the aim of removing the errors or problems that we were encountering at a particular model.

Finally after in depth analysis of the models we came to the conclusion that the **Time and Entity fixed model** performs the best to explain the relationship of overall crime rate with our explanatory variables. The model performs optimally getting rid of all the problems of heteroskedasticity, serially correlated errors, simultaneous causality bias and endogeneity to name a few.

We have received the following insights and conclusions from our preferred model:

- ✓ Shall law effect is insignificant or it does not have any effect on the overall crime rate, Though its coefficient is negative we never found enough evidence to show that it was significantly different than 0.
- ✓ Incarceration rate was found to be very significant or it had a huge effect on the overall crime rate. With increase in the Incarceration rate the overall crime rate decreased.
- ✓ Percentage of young male in the states between the ages of 10-29 was highly significant or had an huge effect on the overall crime rate. With higher percentage of young males in the states between the ages of 10-29 the overall crime rate increased.
- ✓ A general trend of increase in the overall crime rate was found from the year 1977- year 1995 but a slight decline in the overall crime rates were observed for the later years from 1996-1999.

Shall law was found to be insignificant or it did not seem to affect the overall crime rate, however we might think that there could be other variables which could affect the overall crime rate like cultural attitudes of different states ect.

Incarceration rate however, was found to be extremely significant in our analysis meaning that it significantly affected the overall crime rate. It had a negative relationship with the overall crime rate which gives us an insight that by tightening the laws and improving the policing there is a lot that can be done to reduce the overall crime rate. Fast track courts and strict federal laws and more intensive policing would be our advice to reduce the overall crime rate.

Appendix-I

While doing this project there are several tools and packages that were made used of. We would like to take this opportunity to mention a list of those tools and packages.

- **Rstudio**: This has been our primary platform in which all the analysis was done including both the exploration, modelling and the statistical tests.
- **Tableau**: For some of the visualizations in the exploration we had made use of Tableau. To name a few visualization that were made using tableau: 1) Trend chart of overall crime rate 2) Spatial Visualization of the Overall Crime Rate
- **Package 'foreign'**: It was used to upload the dataset into our R environment from the .dta format.
- **Package 'moments'**: Gave us functionality to find the skewness and kurtosis of various distribution of our variables for better understanding of the dataset.
- **Package 'dplyr**': For manipulating the dataset, for example adding new variable, filtering, sorting etc.
- **Package 'corrplot':** For making the correlation matrix
- **Package 'ggplot2':** Majority of the visualizations that were made in the exploration part of the project were made using this package
- **Package 'ggthemes':** This package was used to enhance our visualization by adding appropriate themes to our visualizations
- **Package 'gridextra':** To visualize multiple plot on a single frame.
- **Package 'broom' and 'stargazer':** To table the regressions output in a tabular framework
- **Package 'devtools' and 'sjplot':** To visualize our regression outputs.
- **Package 'lmtest':** For conducting multiple hypothesis testing.
- **Package 'estimatr'** For conducting the White's test

Appendix-2

(Code)

Importing Libraries

```
library(margins)
library(esquisse)
library(tidyverse)
library(moments)
library(dplyr)
library(ggplot2)
library(corrplot)
library(ISLR)
library(car)
library(foreign)
library(ggthemes)
library(colorRamps)
library(foreign)
library(multcomp)
library(survey)
library(lmtest)
library(car)
library(estimatr)
library(gridExtra)
library(gplots)
library(plm)
library(estimatr)
library(lmtest)
source("https://www.r-statistics.com/wp-content/uploads/2010/07/coefplot.r.txt")
library(broom)
```

```
library(knitr)
library(sjPlot)
library(stargazer)
install.packages("jtools")
install.packages("devtools")
library(jtools)
library(devtools)
install.packages("huxtable")
library(huxtable)
```

Importing the dataset

```
setwd("C:/Users/behab/Documents/R Documents")
guns_data<- read.dta("guns.dta")
View(guns_data)
attach(guns_data)
na.omit(guns_data)
```

Summary Statistics of Dataset

```
summary(guns_data)
```

Data Exploration

```
cor(guns_data[,-15])
corrplot.mixed(cor(guns_data[,c(-(15:39))]))
corrplot(cor(guns_data[,c(-(15:39))]),method = "shade")

pairs(mur~year+incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density)
pairs(rob~year+incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density)
pairs(vio~year+incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density)
pairs(all_crime~year+incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density)

plot(vio,mur)
plot(vio,rob)
```

```
plot(rob,mur)



guns_data$all_crime<-vio+rob+mur

View(guns_data)

attach(guns_data)


stargazer(guns_data[c("vio","mur","rob","incarc_rate","pb1064","pw1064","pm1029","pop","avginc","de
nsity","all_crime")], type = "text", title="Descriptive statistics", digits=1, out="descriptive.dox")

names(guns_data)

library(ggplot2)



guns_data$factor_shall<- as.factor(shall)

attach(guns_data)

library(ggplot2)


ggplot(p_dat) +

 aes(x = incarc_rate, y = all_crime) +

 geom_point(size = 1L, colour = "#d8576b") +

 labs(x = "Incarceration Rate(per 100,000)", y = "Total Crime(all_crime)", title = "Scatter Plot Of
Incarceration Rate Vs Total Crime") +

 theme_igray()


# Histogram for Vio,mur,rob, all_crime

library(ggplot2)


q<-ggplot(guns_data) +

 aes(x = mur) +

 geom_histogram(bins = 30L, fill = "#cf4446") + labs(x = "Murder Rate(Incidents per 100,000)", y =
"Frequency", title = "Histogram Of Murder Rate", caption = "Positively Skewed Histogram") +
```

```
 theme_igray()


library(ggplot2)


y<-ggplot(guns_data) +
 aes(x = vio) +
 geom_histogram(bins = 30L, fill = "#cd4071") +
 labs(x = "Violent Crime Rate(Incidents per 100,000)", y = "Frequency", title = "Histogram Of Violent Crime Rate", caption = "Positively Skewed Histogram") +
 theme_igray()


library(ggplot2)


v<-ggplot(guns_data) +
 aes(x = rob) +
 geom_histogram(bins = 30L, fill = "#f1605d") +
 labs(x = "Robbery Rate(Incidents per 100,000)", y = "Frequency", title = "Histogram Of Robbery Rate", caption = "Positively Skewed Histogram") +
 theme_igray()




# Histograms For Density and Incarceration



library(ggplot2)


k1<-ggplot(p_dat) +
 aes(x = density) +
 geom_histogram(bins = 30L, fill = "#440154") +
```

```
 labs(x = "Density", y = "Frequency", title = "Histogram Of Density", caption = "Positively Skewed Histogram Of Density") +

 theme_igray()


library(ggplot2)


k2<-ggplot(p_dat) +

 aes(x = incarc_rate) +

 geom_histogram(bins = 30L, fill = "#617a89") +

 labs(x = "Incarceration Rate(per 100,000)", y = "Frequency", title = "Histogram Of Incarceration Rate", caption = "Positively Skewed Histogram For Incarceration Rate") +

 theme_igray()


grid.arrange(k1,k2,nrow=1)

library(ggplot2)


k3<-ggplot(guns_data) +

 aes(x = all_crime) +

 geom_histogram(bins = 30L, fill = "#0d0887") +

 labs(x = "Combined Crime Rate(Incidents Per 100,000)", y = "Frequency", title = "Histogram Of Combined Crime Rates(all_crime)", caption = "Positively Skewed Histogram") +

 theme_igray()


# Plot of ALL SKEWED HISTOGRAMS

grid.arrange(q,y,v,k1,k2,k3,nrow=2)


library(ggplot2)


q1<-ggplot(p_dat) +
```

```
 aes(x = pm1029) +
 geom_histogram(bins = 30L, fill = "#2d543d") +
 labs(x = "Percentage Of Male Between The ages of 10-29", y = "Frequency", title = "Histogram Of pm1029", caption = "The Histogram Looks Fairly Symmetrical") +
 theme_igray()

library(ggplot2)

q2<-ggplot(p_dat) +
 aes(x = avginc) +
 geom_histogram(bins = 30L, fill = "#1f9e89") +
 labs(x = "Average Income", y = "Frequency", title = "Histogram Of Average Income", caption = "The histogram of Average Income looks fairly Symmetric") +
 theme_igray()

grid.arrange(q1,q2,nrow=1)

skewness(all_crime)

skewness(log_all_crime)
# Box-Plot
ggplot(guns_data) +
  aes(x = factor_shall, y = all_crime, fill = factor_shall) +
  geom_boxplot() +
  scale_fill_viridis_d(option = "magma") +
  labs(x = "Shall-Carry Law Effect", y = "Overall Crime Rate(per100,000)", fill = "Shall-Carry Effect") +
  theme_igray()

# Violence Histogram - Shall Law
ggplot(guns_data) +
 aes(x = vio) +
```

```
 geom_histogram(bins = 30L, fill = "#d8576b") +
 labs(x = "Violent crime rate(Incedents per 100,000)", y = "Frequency", title = "Histogram Of Violent Crime Rates Based on The Shall-Carry Law Effect", caption = "Sharp decline in Violent Crime rates for the cases wher Shall-Carry Law is in place") +
 theme_igray() +
 facet_wrap(vars(shall))

skewness(vio)

library(ggplot2)
# Murder Histogram - Shall Law
ggplot(guns_data) +
 aes(x = mur) +
 geom_histogram(bins = 30L, fill = "#cf4446") +
 labs(x = "Murder Rate(Incedents per 100,000)", y = "Frequency", title = "Murder Rate histogram as per Shall-Carry Law effect", caption = "Sharp Decline in the Murder Rate can be seen in case of Shall-Carry Law being in place") +
 theme_igray() +
 facet_wrap(vars(shall))

skewness(mur)

# Robbery Histogram - Shall Law

library(ggplot2)

ggplot(guns_data) +
 aes(x = rob) +
```

```
 geom_histogram(bins = 30L, fill = "#f1605d") +
 labs(x = "Robbery Rate(Incidents per 100,000)", y = "Frequency", title = "Histogram of Robbery Rate as per weather the Shall-Carry Law is in effect") +
 theme_igray() +
 facet_wrap(vars(shall))

skewness(rob)

library(ggplot2)
# Year Vs Violent Crime Rate based on Shall Carry Law
ggplot(guns_data) +
 aes(x = year, y = vio) +
 geom_point(size = 1L, colour = "#d8576b") +
 labs(x = "Year", y = "Violence Rate(Per 100,000)", title = "Year Vs Violent Crime Rate by Shall-Carry Law Effect", caption = "Shall-Carry Law had utility in reducing the Violent Crime Uptil 1989(Since Last decade the effect of Shall Carry Law is more or less similar to its counterpart)") +
 theme_igray() +
 facet_wrap(vars(shall))

cor(vio,year)
cor(new_data$vio,new_data$year)

library(ggplot2)
# Year Vs Murder Crime Rate based on Shall Carry Law
ggplot(guns_data) +
 aes(x = year, y = mur) +
```

```
 geom_point(size = 1L, colour = "#cf4446") +
 labs(x = "Year", y = "Murder Rate(Per 100,000)", title = "Year Vs Murder Rate Based on Shall-Carry Law Effect", caption = "Same pattern of increase in the murder rate after 1990 can be seen where the Shall Carry Law is in Effect") +
 theme_igray() +
 facet_wrap(vars(shall))


cor(mur,year)
cor(new_data$mur,new_data$year)




library(ggplot2)
# Year Vs Robbery Crime Rate based on Shall Carry Law
ggplot(guns_data) +
 aes(x = year, y = rob) +
 geom_point(size = 1L, colour = "#f1605d") +
 labs(x = "Year", y = "Robbery Rate(per 100,000)", title = "Year Vs Robbery Rate by Shall-Carry Law", caption = "Same Spike in the Robbery Rate can be seen after 1990 in the cases where the shell-Carry law is in effect") +
 theme_igray() +
 facet_wrap(vars(shall))


cor(rob,year)
cor(new_data$rob,new_data$year)




library(ggplot2)
```

```
guns_data$all_crime<-vio+rob+mur

View(guns_data)

attach(guns_data)




library(ggplot2)

# Over all Crime Rate Vs Year By shall-carry law effect

ggplot(guns_data) +

 aes(x = year, y = all_crime) +

 geom_point(size = 1L, colour = "#a52c60") +

 labs(x = "Year", y = "Overall Crime Rate(per 100,000)", title = "Year Vs Ovarall Crime Rate by Shall-Carry Law effect", caption = "Same Pattern Can be Observed") +

 theme_igray() +

 facet_wrap(vars(shall))


cor(all_crime,year)




library(ggplot2)

# Incar Crime

t1<-ggplot(guns_data) +

 aes(x = all_crime, y = incarc_rate, colour = shall) +

 geom_point(size = 1L) +

 scale_color_gradient() +

 labs(x = "Overall Crime Rate(per 100,000)", y = "Incarceration Rate(per 100,000)", caption = "If Shall Law Implemented More Incarceration?", color = "Shall Law Effect") +

 theme_igray()


t2<-ggplot(guns_data) +
```

```
  aes(x = factor_shall, y = incarc_rate, fill = factor_shall) +
  geom_boxplot() +
  scale_fill_brewer(palette = "Paired") +
  labs(x = "Shall Law Effect", y = "Incarceration Rate", title = "Shall Law Effect Vs Incarceration Rate", caption = "States Having Shall Law in Effect tend to have Higher Incarceraction Rates", fill = "Shall Law Effect") +
  theme_igray()

grid.arrange(t1,t2,nrow=1)

# pb,pw,pm Variables Vs Overall Crime Rates by Shall Law Effect

library(ggplot2)

ggplot(guns_data) +
 aes(x = pb1064, y = all_crime, colour = factor_shall) +
 geom_point(size = 1L) +
 scale_color_viridis_d(option = "plasma") +
 labs(x = "Percentage Of Blacks by State(age10-age64)", y = "Overall Crime Rate(per 100,000)", title = "Percentage Of Blacks by State Vs Overall Crime Rate By Shall Law Effect", caption = "Positive Correlation between percentage of blacks in the state and the overall Crime Rate for both the Shall Law cases") +
 theme_igray()

library(ggplot2)

ggplot(guns_data) +
```

```
 aes(x = pw1064, y = all_crime, colour = factor_shall) +

 geom_point(size = 1L) +

 scale_color_viridis_d(option = "magma") +

 labs(x = "Percentage Of Whites in states(age10-age64)", y = "Overall Crime Rate(per 100,000)", title = "Percentage Of Whites in states Vs Overall Crime Rate by Shall Law Effect", color = "Shall-Carry Law Effect",caption = "Negative Correlation between Percentage of whites in the state and overall Crime Rate in both the Shall Law Cases") +

 theme_igray()




library(ggplot2)


ggplot(guns_data) +

 aes(x = pm1029, y = all_crime, colour = factor_shall) +

 geom_point(size = 1L) +

 scale_color_viridis_d(option = "viridis") +

 labs(x = "Percentage Of Males in a State(age10-age29)", y = "Overall Crime Rate(per 100,000)", title = "Percentage Of Males in a State Vs Overall Crime Rate by Shall Law Effect", caption = "No correlation was found. For States Having Higher Percentag young male a Shall Law in effect leads to Less Overall Crime", color = "Shall-Carry Law Effect") +

  theme_igray()




# Density

# Population
```

```
library(ggplot2)


p<-ggplot(guns_data) +
 aes(x = pop, y = all_crime, colour = factor_shall) +
 geom_point(size = 1L) +
 scale_color_brewer(palette = "Dark2") +
 labs(x = "Population(in Millions)", y = "Overall Crime Rate(per 100,000)", title = "Population Vs Overall Crime Rate by Shall Law Effect", caption = "Positive Correlation between Overall Crime and Population", color = "Shall Law Effect") +
 theme_igray()


guns_data <- guns_data %>%
 filter(density >= 0L & density <= 5L)
d<-ggplot(guns_data) +
 aes(x = density, y = all_crime, colour = factor_shall) +
 geom_point(size = 1L) +
 scale_color_hue() +
 labs(x = "Density", y = "Overall Crime Rate(Per 100,000)", caption = "Positive Correlation Between Population Density and Overall Crime Rate. Densely Populated States do Not have Shall Law in Effect. Afraid that More Density More Crime?", color = "Shall Law Effect") +
 theme_igray()


grid.arrange(p,d,nrow=1)
```

```
library(ggplot2)
```

```
library(ggplot2)

ggplot(guns_data) +
 aes(x = all_crime, y = incarc_rate, colour = factor_shall) +
 geom_point(size = 1L) +
 scale_color_brewer(palette = "Set1") +
 labs(x = "Overall Crime Rate(per 100,000)", y = "Incarceration Rate", title = "Overall Crime Rate Vs Incarceration Rate", subtitle = "Positive Correlation between Overall Crime Rate and Incarceration Rate", caption = "States with Shell Law in Place Have Higher Incarceration rates?", color = "Shall Law Effect")
+
 theme_igray()

# avg Income Vs Overall Crime Rate by Shall Law Effect
ggplot(guns_data) +
  aes(x = avginc, y = all_crime, colour = factor_shall) +
  geom_point(size = 1L) +
  scale_color_brewer(palette = "Set3")  +
  labs(x = "Average Per-Capita Income Per State", y = "Overall Crime Rate(per 100,000)", title = "Average Per-Capita Income Per State VS Overall Crime Rate(per 100,000) By Shall Law Effect",caption = "Diminishing Effect?") +
  theme_igray()

# Incarceration

# Year Vs Incarceration Rate by Shall Law

ggplot(guns_data) +
  aes(x = year, y = incarc_rate) +
```

```
  geom_point(size = 1L, colour = "#000004") +
  labs(x = "Year", y = "Incarceration Rate", title = "Year Vs Incarceration Rate by Shall-Carry Law Effect", caption = "Same Pattern of the Incarceration Rate Spiking up in after 1990 where the Shall-Carry Law is in effect can be seen") +
  theme_igray() +
  facet_wrap(vars(shall))


# Incar Percentage of blacks and whites
ggplot(guns_data) +
  aes(x = pb1064, y = pw1064, colour = incarc_rate) +
  geom_point(size = 1L) +
  scale_color_distiller(palette = "BuPu",direction = 1) +
  labs(x = "Percentage Black Population In State(age10-age64)", y = "Percentage White Population In State(age10-age64)", title = "Percentage Black Population In State Vs Percentage White Population In State by Incarceration Rate", caption = "More Percentage of white in state-less Incarceration,More Percentage Of blacks in states more incarcerations", color = "Incarceration Rate") +
  theme_igray()




library(ggplot2)


w<-ggplot(guns_data) +
 aes(x = pw1064, y = all_crime, colour = incarc_rate) +
 geom_point(size = 1L) +
 scale_color_distiller(palette = "Blues",direction = 1) +
 labs(x = "Percentage of Whites In The State", y = "Overall Crime Rate(per 100,000)", title = "Percentage of Whites In The State Vs Overall Crime Rate By Incarceration Rate", caption = "More The White Population Less the Crime hence less the Incarceration Rate", color = "Incarceration Rate") +
 theme_igray()
```

```
library(ggplot2)


b<-ggplot(guns_data) +
 aes(x = pb1064, y = all_crime, colour = incarc_rate) +
 geom_point(size = 1L) +
 scale_color_distiller(palette = "BuGn",direction = 1) +
 labs(x = "Percentage Of Black Population in State", y = "Overall Crime Rate(per 100,000)", title = "Percentage Of Black Population in State Vs Overall Crime Rate", caption = "More Black population in state more crime and hence greater incarceration rate", color = "Incarceration Rate") +
 theme_igray()


grid.arrange(b,w,nrow=1)




# Incar percentage of males
library(ggplot2)


ggplot(guns_data) +
 aes(x = pm1029, y = all_crime, colour = incarc_rate) +
 geom_point(size = 1L) +
 scale_color_distiller(palette = "YlOrRd",direction = 1) +
 labs(x = "Percentage Of Males in State between Ages of 10 and 29", y = "Overall Crime Rate(per 100,000)", title = "Percentage Of Males in State between Ages of 10 and 29 Vs Overall Crime Rate by Incarceration Rate", caption = "Sates Having less percentage of males between this age frame and Higher Crime Rate have Higher Incarceration Rates", color = "Incarceration Rate") +
 theme_igray()


# Incar Pop-dENSITY
library(ggplot2)
```

```
d2<-ggplot(guns_data) +
 aes(x = density, y = all_crime, colour = incarc_rate) +
 geom_point(size = 1L) +
 scale_color_distiller(palette = "Purples",direction = 1) +
 labs(x = "Density", y = "Overall Crime Rate(per 100,000)", title = "Density Vs Overall Crime Rate", caption = "More Density-More Crime Hence More Incarceration Rate", color = "Incarceration Rate") +
 theme_igray()


library(ggplot2)

p2<-ggplot(guns_data) +
 aes(x = pop, y = all_crime, colour = incarc_rate) +
 geom_point(size = 1L) +
 scale_color_distiller(palette = "OrRd",direction = 1) +
 labs(x = "Population", y = "Overall Crime Rate", title = "Population Vs Overall Crime Rate", caption = "States Hving Higher population have more overall crime and hence have Higher Incarceration Rates", color = "Incarceration Rates") +
 theme_igray()

grid.arrange(p2,d2,nrow=1)


library(ggplot2)

ggplot(guns_data) +
 aes(x = factor_shall, y = incarc_rate, fill = factor_shall) +
 geom_boxplot() +
 scale_fill_brewer(palette = "Paired") +
```

```
 labs(x = "Shall Law Effect", y = "Incarceration Rate", title = "Shall Law Effect Vs Incarceration Rate", caption = "States Having Shall Law in Effect tend to have Higher Incarceraction Rates", fill = "Shall Law Effect") +

 theme_igray()

library(ggplot2)

# Trend of Young Men In States

ggplot(guns_data) +

 aes(x = year, y = pm1029) +

 geom_point(size = 1L, colour = "#0b53c1") +

 labs(x = "Year", y = "Percentage Of Male Population In a State(age10-29)", title = "Trend Of Young Male Population In States", caption = "Decreasing Trend") +

 theme_igray()

library(ggplot2)

# Trend of Avg Income in states

library(ggplot2)

ggplot(guns_data) +

 aes(x = year, y = avginc) +

 geom_point(size = 1L, colour = "#26828e") +

 labs(x = "Year", y = "Average Per Capita Income In States", title = "Trend Of Per-Capita Income In States", caption = "Positive Trend") +

 theme_igray()

# Pm1029 Vs Incar
```

```
library(ggplot2)

ggplot(guns_data) +
 aes(x = pm1029, y = incarc_rate) +
 geom_point(size = 1L, colour = "#0c4c8a") +
 labs(x = "Percentage Of Males in State(10-29)", y = "Incarcercation Rate", title = "Percentage Of Males in State Vs incarcercation Rate", caption = "Negative Relationship") +
 theme_igray()

# Pm1029 Vs Avg Income

library(ggplot2)

ggplot(guns_data) +
 aes(x = pm1029, y = avginc) +
 geom_point(size = 1L, colour = "#35b779") +
 labs(x = "Percentage Of Males in State(10-29)", y = "Average Per Capita Income In State", title = "Percentage Of Males in State Vs Average Per Capita Income In State", caption = "Negative Trend") +
 theme_igray()

# Average Income Vs Density

library(ggplot2)

ggplot(guns_data) +
 aes(x = density, y = avginc) +
 geom_point(size = 1L, colour = "#617a89") +
```

labs(x = "Density", y = "Average Per Capita Income", title = "Density Vs Average Per Capita Income", caption = "Positive Trend") +

theme_igray()

Modelling

#Heteroginity across year

View(guns_data)

plotmeans(all_crime~year,xlab = "Year",ylab = "Overall Crime Rate(all_crime)")

lin_model1<-lm(all_crime~year)

summary(lin_model1)

m<-resid(lin_model1)

plot(year,m)

#heteroginity across states

num_stateid<-as.numeric(stateid)

attach(guns_data)

l<- filter(guns_data,num_stateid<=51)

lin_model2<-lm(all_crime~num_stateid,data = l)

summary(lin_model2)

k<-resid(lin_model2)

plot(num_stateid,k)

attach(l)

plotmeans(all_crime~stateid,xlab = "State-id",ylab = "Overall Crime Rate(all_crime)")

attach(p_dat)

attach(guns_data)

guns_data$factor_year<-as.factor(year)

Panel Data Model

POOLED OLS FINAL REGRESSION(FOR REPORT)

attach(p_dat)

```
ols_model_log_all_crime<-plm(log_all_crime~log_incarc+pb1064+pw1064+pm1029+pop+avginc+log_density+factor_shall,data = p_dat,model = "pooling")

summary(ols_model_log_all_crime)

kable(tidy(ols_model_log_all_crime))

coefplot(ols_model_log_all_crime)

a<-resid(ols_model_log_all_crime)

b<-fitted(ols_model_log_all_crime)


# Residual PLOT PROOF FOR HETEROSKEDASTICITY INFORMAL WAY

ggplot(mapping=aes(x=b,y=a))+geom_point(color="navyBlue")+geom_abline(slope = 0,intercept = 0,color="red",size=1)+xlab("Fitted Values of ln_allcrime (y-hat)")+ylab("Residuals")+theme_igray()+ggtitle("Informal Residual Plot To Check Heteroskedasticity")


ggplot(mapping=aes(x=year,y=a))+geom_point(color="black")+geom_abline(slope = 0,intercept = 0,color="red",size=1)+xlab("Fitted Values of y (y-hat)")+ylab("Residuals")+theme_igray()+ggtitle("Informal Residual Plot To Check Heteroskedasticity")


ggplot(mapping=aes(x=stateid,y=a))+geom_point(color="darkgreen")+geom_abline(slope = 0,intercept = 0,color="red",size=1)+xlab("State-ID")+ylab("Residuals")+theme_igray()+ggtitle("Informal Residual Plot across entities To Check Heteroskedasticity across entities")

# Whites Test For Heteroskedasticity

kable(tidy(bptest(ols_model_log_all_crime,~log_incarc+pb1064+pw1064+pm1029+pop+avginc+log_density+factor_shall,data = p_dat)))

# P-vale practically 0.009737 hence we reject the null hypothesis and conclude that there is indeed heteroskedasticity

x1<-coeftest(ols_model_log_all_crime, vcov=vcovHC(ols_model_log_all_crime, cluster="group"))


kable(tidy(x1))
```

Entity Fixed Model

```
fixed_entity_model<-plm(log_all_crime~log_incarc+pb1064+pw1064+pm1029+pop+avginc+log_density+factor_shall,data = p_dat,model = "within")

summary(fixed_entity_model)
```

```
coefplot(fixed_entity_model)

entity_fixed_resid<- resid(fixed_entity_model)
entity_fixed_fitted<- fitted(fixed_entity_model)

ggplot(mapping=aes(x=entity_fixed_fitted,y=entity_fixed_resid))+geom_point(color="maroon")+geom_abline(slope = 0,intercept = 0,color="navyblue",size=1)+xlab("Fitted Values of ln_allcrime (y-hat)")+ylab("Residuals")+theme_igray()+ggtitle("Informal Residual Plot To Check Heteroskedasticity")

fixed_time_entity_model<-
plm(log_all_crime~log_incarc+pb1064+pw1064+pm1029+pop+avginc+log_density+factor_shall+factor_year,data = p_dat,model = "within")
coefplot(fixed_time_model)
summary(fixed_time_entity_model)
coefplot(fixed_time_entity_model)
export_summs(fixed_time_entity_model, scale = TRUE, transform.response = TRUE,model.names = "Time and Entity Fixed Model",to.file = "xlsx",file.name = "Time and Entity Fixed_table.xlsx")
residual_fixed<- resid(fixed_time_entity_model)
fitted_fixed<- fitted(fixed_time_entity_model)

ggplot(mapping=aes(x=fitted_fixed,y=residual_fixed))+geom_point(color="black")+geom_abline(slope = 0,intercept = 0,color="red",size=1)+xlab("Fitted Values of ln_allcrime (y-hat)")+ylab("Residuals")+theme_igray()+ggtitle("Informal Residual Plot To Check Heteroskedasticity")

export_summs(fixed_entity_model, scale = TRUE, transform.response = TRUE,model.names = "Entity Fixed Model",to.file = "xlsx",file.name = "Entity_Fixed_Model_table.xlsx")

# ENTITY-FIXED VS TIME & ENTITY FIXED
null2<-
c("factor_year78=0","factor_year79=0","factor_year80=0","factor_year81=0","factor_year82=0","factor_year83=0","factor_year84=0","factor_year85=0","factor_year86=0","factor_year87=0","factor_year88=0","factor_year89=0","factor_year90=0","factor_year91=0","factor_year92=0","factor_year93=0","factor_year94=0","factor_year95=0","factor_year96=0","factor_year97=0","factor_year98=0","factor_year99=0")
```

linearHypothesis(fixed_time_entity_model,null2)

kable(linearHypothesis(fixed_time_entity_model,null2))

pFtest(fixed_time_entity_model,fixed_entity_model)

Choosing between entity and time Fixed Effect and OLS Pooling Model

#Null Hypothesis: Ols Model Better than entity and timeFixed effect Model(B1,1=B1,2=B1,3=....=B1,90)

#Alternative Hypothesis: entity and time Fixed effect Model is better than Ols Model(NOT ALL THE INTERCEPTS ARE EQUAL)

pFtest(fixed_time_entity_model,ols_model_log_all_crime)

We Reject the null hypothesis and conclude that the entity and time fixed effect model is indeed better than the ols pooling model

Comparing Between Fixed Effects and Pooled OLS Estimator

kable(tidy(pFtest(fixed_time_entity_model,ols_model_log_all_crime)))

fixef(fixed_time_entity_model)

kable(fixef(fixed_entity_model))

The Entity and Time fixed Effects model with cluster robust standard errors

coeftest(fixed_time_model, vcov=vcovHC(fixed_time_model, cluster="group"))

hist(vio)

hist(all_crime)

hist(rob)

hist(mur)

skewness(all_crime)

skewness(vio)

skewness(rob)

```
skewness(mur)

guns_data$log_vio<- log(vio)
guns_data$log_mur<- log(mur)
guns_data$log_all_crime<- log(all_crime)
guns_data$log_rob<- log(rob)

View(guns_data)

attach(guns_data)

hist(log_all_crime)
skewness(log_all_crime)

hist(log_mur)
skewness(log_mur)
hist(log_vio)
skewness(log_vio)
skewness(log_rob)
hist(log_rob)

View(p_dat)
names(guns_data)

p_dat$log_incarc<- log(incarc_rate)

p_dat$log_all_crime<- log(all_crime)
p_dat$log_vio<- log(vio)
p_dat$log_mur<- log(mur)
p_dat$log_rob<- log(rob)
```

```
p_dat$log_density<- log(density)

View(p_dat)


stargazer(ols_model_log_all_crime,fixed_entity_model,fixed_time_entity_model,
column.labels=c("POOLED OLS","ENTITY FIXED","TIME AND ENTITY FIXED EFFECTS"),

        type="text",align = TRUE,title="Comparing Models", digits = 1, out="Comparing.docx")
```