BUAN 6312 ECONOMETRICS AND TIME SERIES

GROUP PROJECT

Vaibhav Rajan Thakur – VXT210009
Tanmayee Kodumagulla –TXK210004
Navneet Kumar Singh – NXS200079
Aditya Oad – AXO200007

Introduction:

Do more Guns reduce Crime?

In the United States of America, Gun culture is most commonly seen, and it had created a lot of sensation in the news. When it was taken to public opinion, Many Americans had believed the opposite. They claim that gun's ownership stops crimes and that belief drives laws which made it easy to won and keep firearms. Guns took more than 36,000 lives in 2015. This and other alarming statistics have led many to ask whether our nation would be better off with firearms in fewer hands.

And as a result, the NRA (National Rifle Association) has realized that the murder rates are not plummeted. There for shall-issue law has been introduced. It means, as long as theapplicant passes the requirements set out by the state law, the issuing authority (police department, state police, etc...)is compelled to issue a permit. An individual has to meet minimal requirements such as, no criminal record of that individual, no mental illness and also has to finish the course as per state law and guidance.

Problem Statement:

Do shall-issues law reduce crime-or not?

Variable	Definition
vio	violent crime rate (incidents per 100,000 members of the population)
rob	robbery rate (incidents per 100,000)
mur	murder rate (incidents per 100,000)
shall	= 1 if the state has a shall-carry law in effect in that year = 0 otherwise
incarc_rate	incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year)
density	population per square mile of land area, divided by 1000
avginc	real per capita personal income in the state, in thousands of dollars
pop	state population, in millions of people
pm1029	percent of state population that is male, ages 10 to 29
pw1064	percent of state population that is white, ages 10 to 64
pb1064	percent of state population that is black, ages 10 to 64
stateid	ID number of states (Alabama = 1, Alaska = 2, etc.)
year	Year (1977-1999)

The panel data for all the 51 states from 1977 to 1999.

	year	vio	mur	rob	incarc_rate	pb1064	pw1064	pm1029	pop	avginc	density	stateid	shall
1	77	414.4	14.2	96.8	83	8.384873	55.12291	18.17441	3.780403	9.563148	0.0745524019	1	0
2	78	419.1	13.3	99.1	94	8.352101	55.14367	17.99408	3.831838	9.932000	0.0755667314	1	0
3	79	413.3	13.2	109.5	144	8.329575	55.13586	17.83934	3.866248	9.877028	0.0762453228	1	0
4	80	448.5	13.2	132.1	141	8.408386	54.91259	17.73420	3.900368	9.541428	0.0768288076	1	0
5	81	470.5	11.9	126.5	149	8.483435	54.92513	17.67372	3.918531	9.548351	0.0771865770	1	0
6	82	447.7	10.6	112.0	183	8.514000	54.89621	17.51052	3.925229	9.478919	0.0773185119	1	0
7	83	416.0	9.2	98.4	215	8.545608	54.83936	17.35089	3.934103	9.783000	0.0774933100	1	0
8	84	431.2	9.4	96.1	243	8.559511	54.77876	17.11902	3.951826	10.357200	0.0778424144	1	0
9	85	457.5	9.8	105.4	256	8.562801	54.67899	16.85875	3.972520	10.725858	0.0782500431	1	0
10	86	558.0	10.1	111.6	267	8.566521	54.51791	16.57609	3.991562	11.091618	0.0786251277	1	0
11	87	559.2	9.3	112.2	283	8.592103	54.38770	16.28230	4.015257	11.323824	0.0790918767	1	0
12	88	558.6	9.9	117.8	307	8.618144	54.23505	15.99270	4.023848	11.654964	0.0792610943	1	0
13	89	590.8	10.2	133.9	300	8.638031	54.06622	15.67523	4.030224	11.963897	0.0793866888	1	0
14	90	708.6	11.6	143.7	328	8.699674	56.07016	15.38070	4.048508	12.063984	0.0797735602	1	0
15	91	844.2	11.5	152.8	370	8.771641	55.97353	15.18314	4.091025	12.087816	0.0806113258	1	0
16	92	871.7	11.0	164.9	394	8.877969	55.80952	15.02558	4.139269	12.398020	0.0815619528	1	0
17	93	780.4	11.6	159.5	407	8.972758	55.66076	14.86296	4.193114	12.395800	0.0826229304	1	0
18	94	683.7	11.9	171.2	431	9.047583	55.50783	14.67744	4.232965	12.673920	0.0834081769	1	0
19	95	632.4	11.2	185.8	450	9.094921	55.33187	14.54549	4.262731	12.872682	0.0839946941	1	0

Showing 1 to 20 of 1,173 entries, 13 total columns

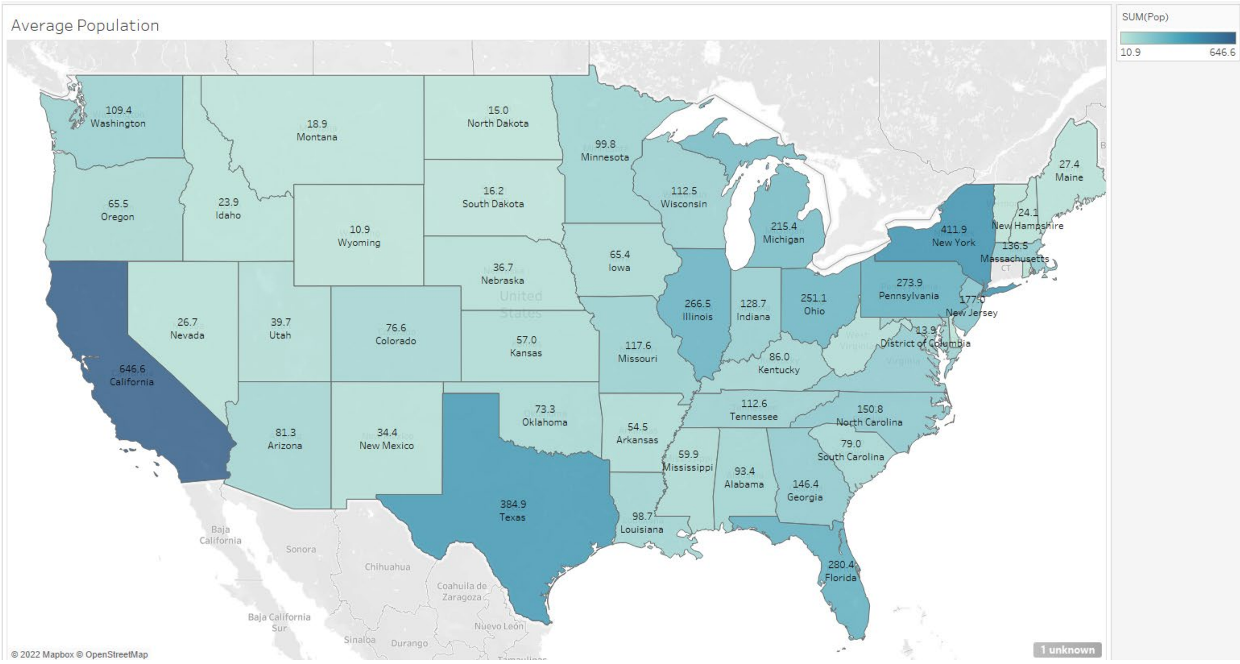
Each state has a specific state ID:

State ID	***State Name***	***Abbreviation***	***State ID***	***State Name***	***Abbreviation***
1	Alabama	AK	30	Nebraska	NC
2	Alaska	AL	31	Nevada	ND
3	American Samoa	AS	32	New Hampshire	NE
4	Arizona	AR	33	New Jersey	NH
5	Arkansas	AZ	34	New Mexico	NJ
6	California	CA	35	New York	NM
7	Colorado	CO	36	North Carolina	NV
8	Connecticut	CT	37	North Dakota	NY
9	Delaware	DE	38	Northern Mariana Islands	MP
10	District of Columbia	DC	39	Ohio	OH
11	Florida	FL	40	Oklahoma	OK
12	Georgia	GA	41	Oregon	OR
13	Guam	GU	42	Pacific Trust Territory	
14	Hawaii	HI	43	Panama Canal Zone	
15	Idaho	IA	44	Pennsylvania	PA
16	Illinois	ID	45	Puerto Rico	PR
17	Indiana	IL	46	Rhode Island	RI
18	Iowa	IN	47	South Carolina	SC
19	Kansas	KS	48	South Dakota	SD
20	Kentucky	KY	49	Tennessee	TN
21	Louisiana	LA	50	Texas	TX
22	Maine	MA	51	U.S. Virgin Islands	US VI
23	Maryland	MD	52	Utah	UT
24	Massachusetts	ME	53	Vermont	VT
25	Michigan	MI	54	Virginia	VA
26	Minnesota	MN	55	Washington	WA
27	Mississippi	MO	56	West Virginia	WI
28	Missouri	MS	57	Wisconsin	WV
29	Montana	MT	58	Wyoming	WY

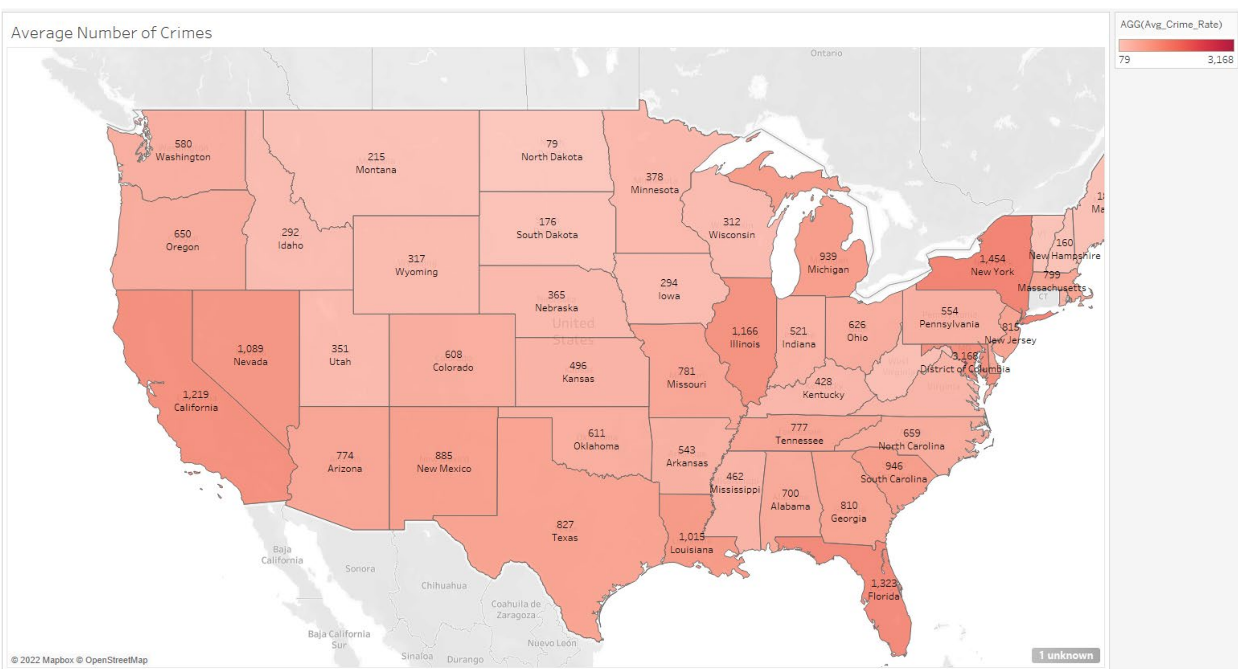
Based on the economic theory, each variable indicates:

- SHALL: Introduction of Shall-carry law should reduce the crime rate and therefore will have an inverse relation with the crime rate.
- INCARC_RATE: Increase in Incarceration rate should reduce the crime rate and therefore will have an inverse relation with the crime rate.
- DENSITY: The role of population density in the generation or suppression of crime has been the subject of debate for decades. So, we can say that it can increase but it is uncertain.
- AVGINC: The real per capita personal income in the state should reduce the crime rate, therefore an inverse relation.
- POP: More the state population, more the chances violent crime rate. So, POP will have a positive relation with VIO.
- PM1029: Having more male population between ages 10 and 29 increase the chances of crime. Therefore, it will have positive relation with crime rate.
- PW1064 and PB1064: The effect of these two variables, according to economic theory, are highly contrasting. The difference because of different racial groups effects the crime rate differently and are debatable. The effect of population of blacks increases the crime rate as compared to population of whites. Competitive society in which there is an inequality in the distribution of goods, those groups with limited or restricted access to goods will be more likely to turn to crime.

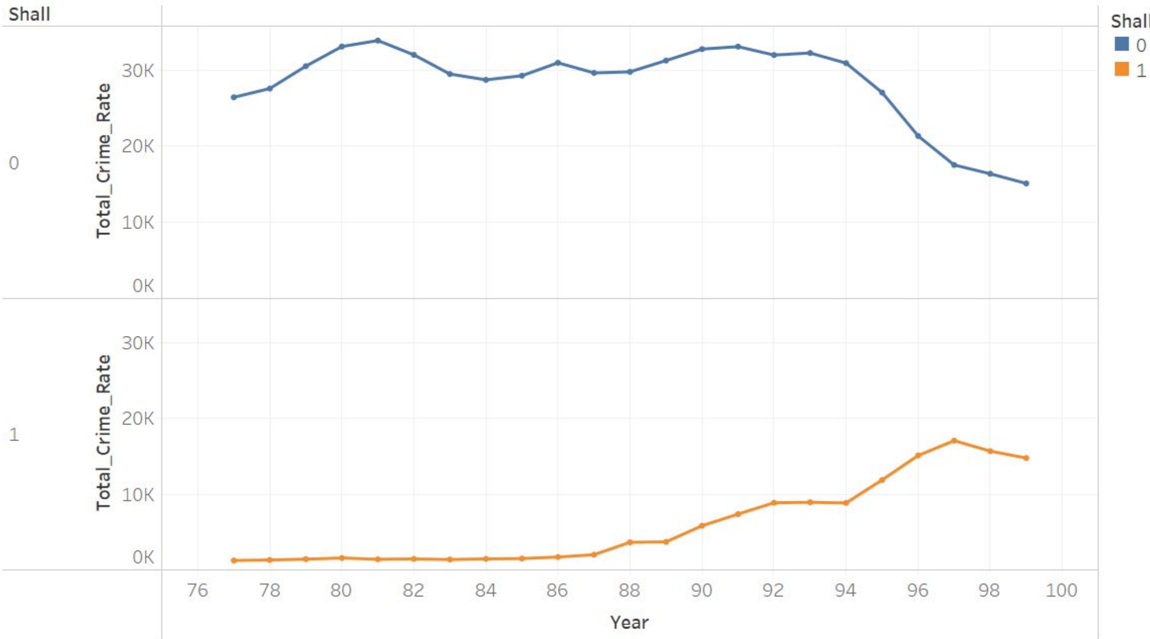
Population Across The United States



Average Crime Rate

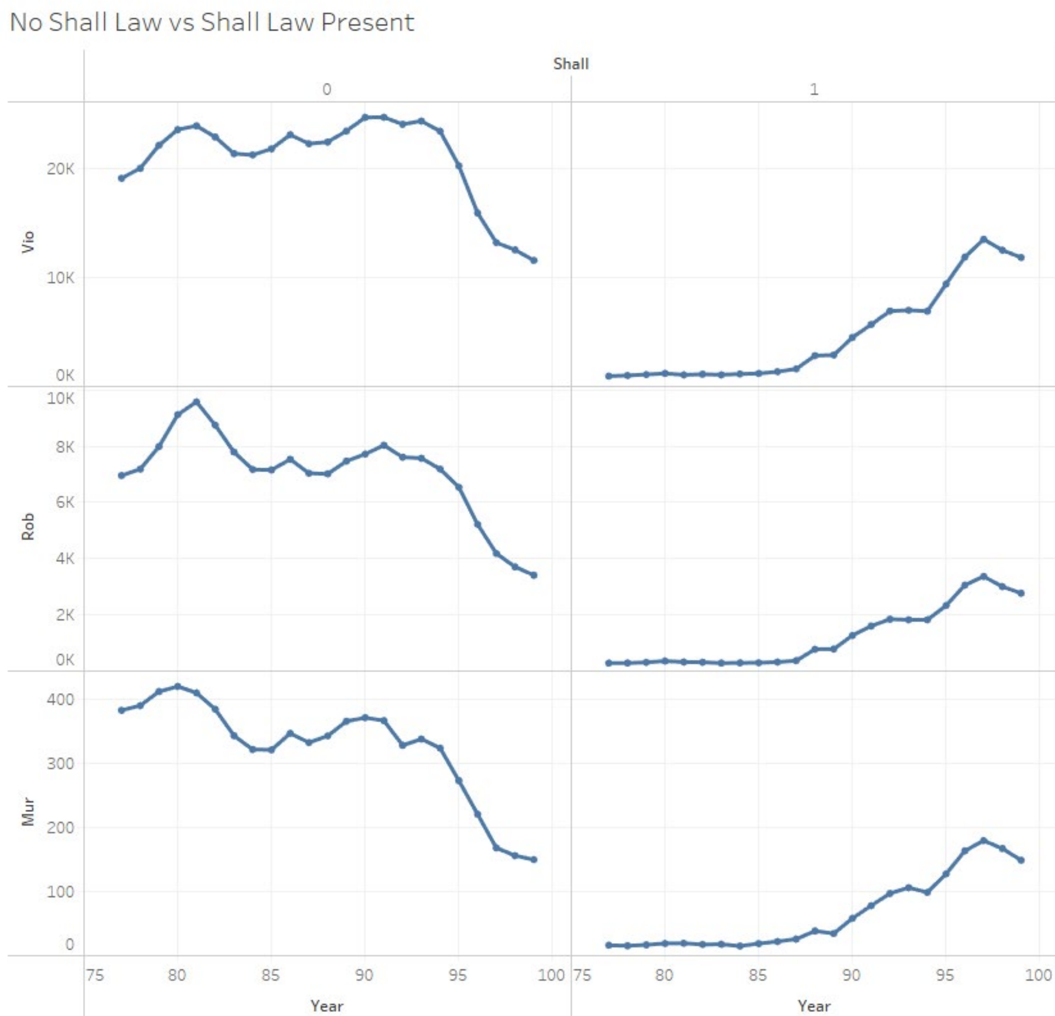


Presence Of Shall Law On Total Crime Rates



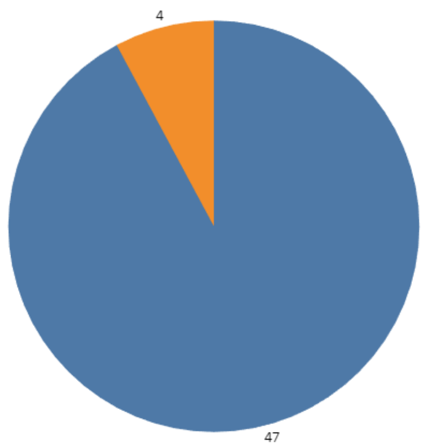
The trend of Total_Crime_Rate for Year broken down by Shall. Color shows details about Shall.

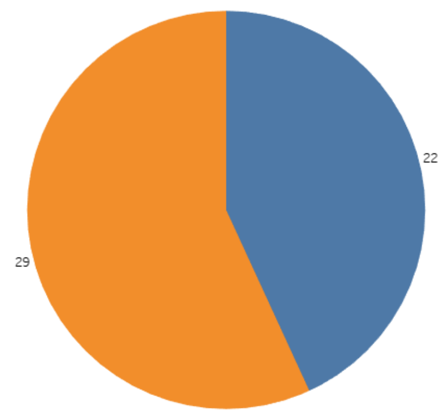
The crime rate dropped significantly after the Shall law was implemented. Hence, it's a success and should be implemented in every state.



After implementing the shall law, we can see a overall decrease in the crime rates.

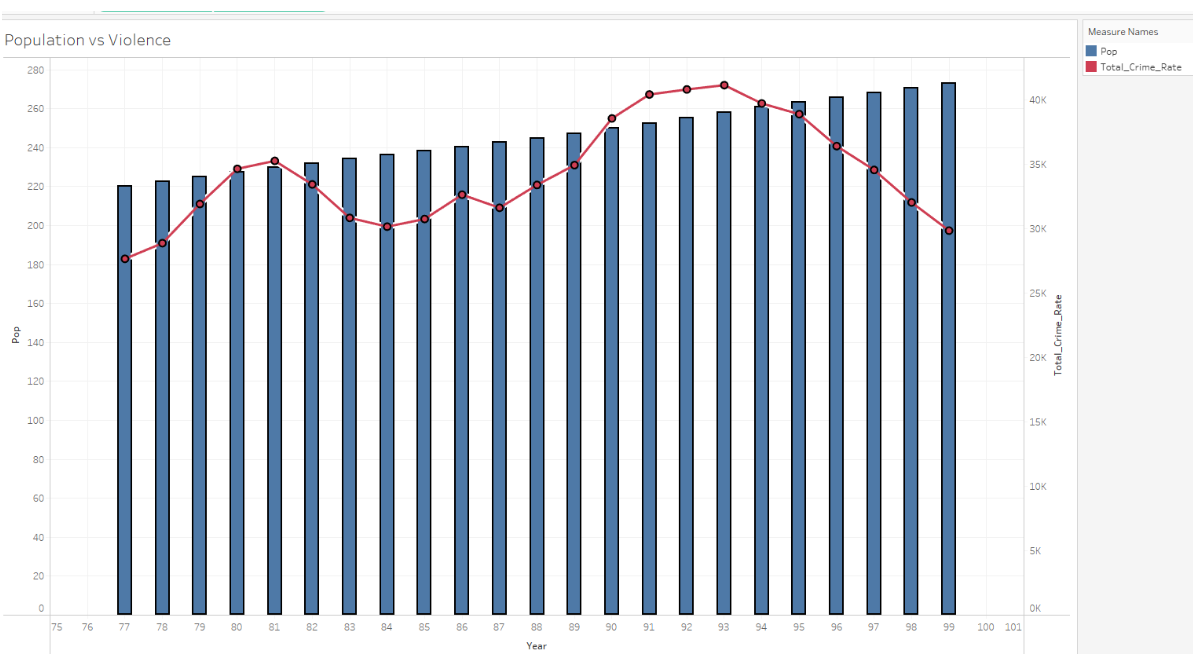
However, we can also see here that the states which had the shall law present had a decreasing trend with respect to the crimes.





Number of states having shall law in the year 1977 Number of states having shall law in 1999

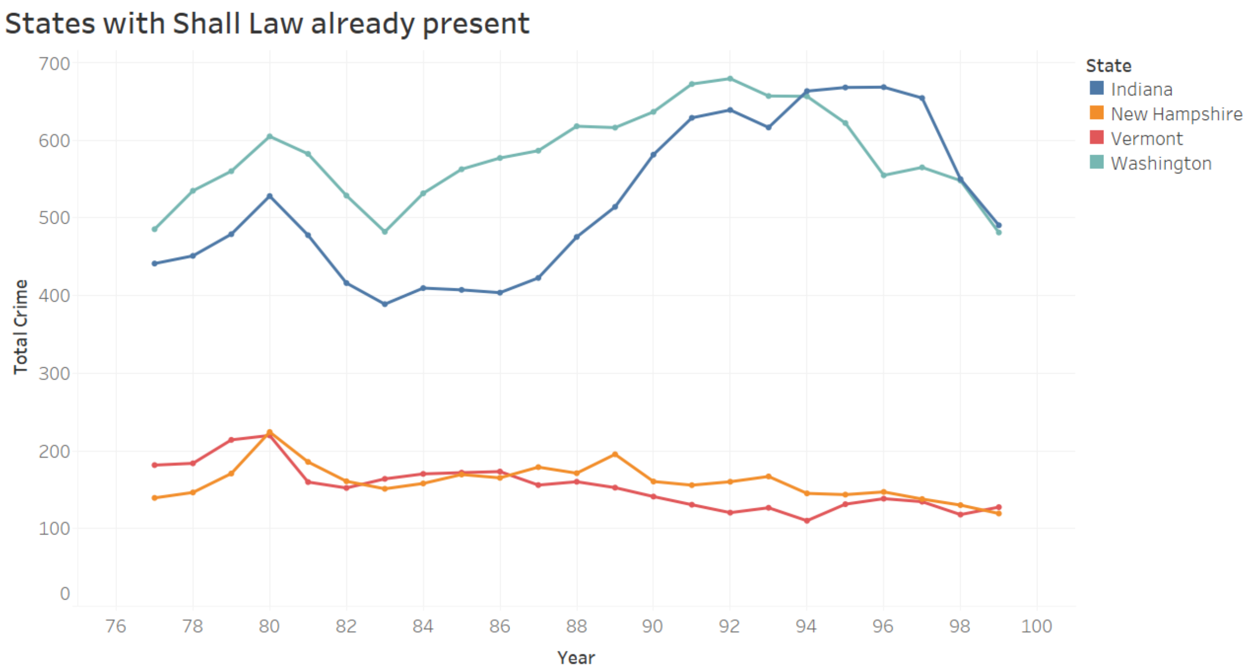
EFFECT OF POPULATION ON TOTAL NUMBER OF CRIMES



We can see that as the population increases the corresponding crime rate also increases in a particular year.

However, this is only true till the year 1993. After that we can see a decline in the crime rates

States Where Shall Law Was Already Present



The trend of Total Crime for Year. Color shows details about State. The view is filtered on State, which keeps Indiana, New Hampshire, Vermont and Washington.

The above graphs tell us about the crime rate in four states where Shall Law was implemented in 1977.

Washington –

In 1977, the average crime rate was 485, which grew to 605 in 1980. For the next three years, there was a slight decline in the crime rate. After 1983, the crime rate increased continuously and peaked in 1992, with the crime rate at 679. After that, there has been a continuous decline, with 1999 recording the lowest ever crime rate of 481.

Indiana –

In 1977, the crime rate was 441. When it increased to 528 in 1980, there was a sharp decline for the next three years, and it hit an all-time low of 389 in 1983. Since then, it has continuously grown and reached an all-time high of 663 in 1994 and remained relatively constant for the next three years. But, there was a sharp decline of 18% in the next year, and the crime rate declined considerably.

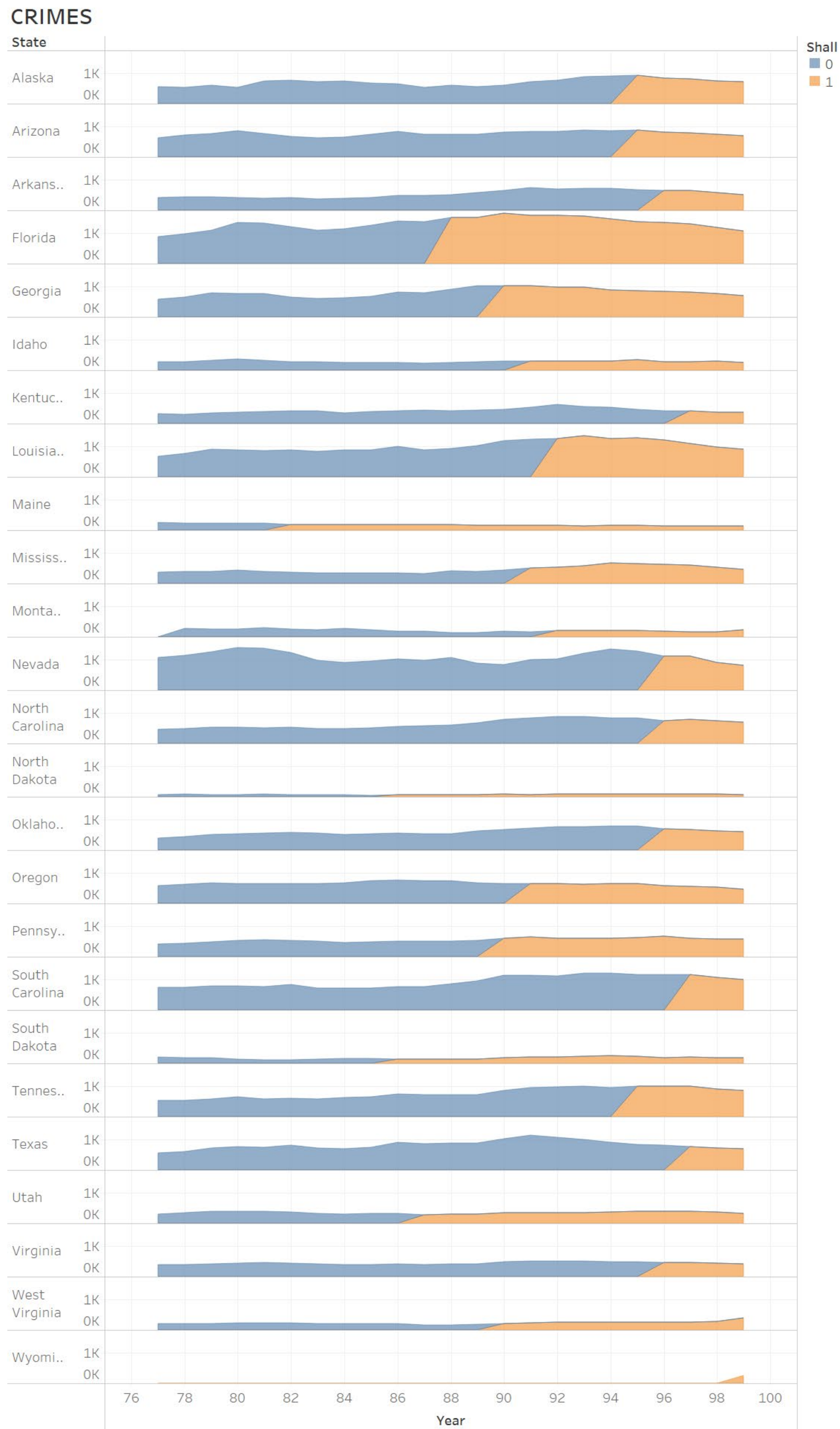
Vermont –

The crime rate in Vermont has been low compared to the other three states. The highest crime rate was recorded in 1980 and was 225. Since then, the crime rate has continuously declined and has been under 150 since 1990.

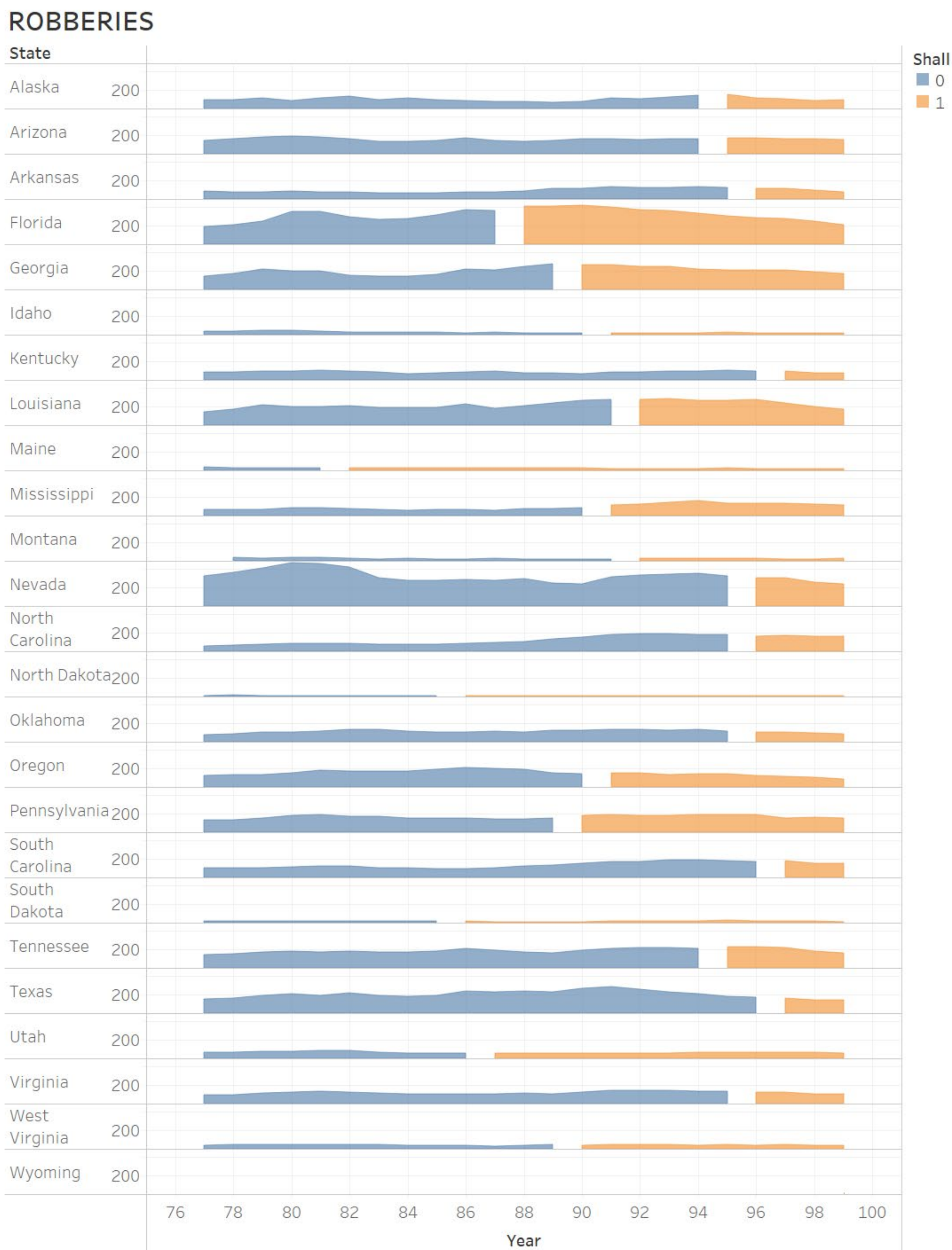
New Hampshire –

The crime rates in New Hampshire and Vermont are close and peaked in the same year, with New Hampshire recording 224 cases. Since then, the crime rate has continuously declined and has been under 200, with 1999 recording just 119 cases.

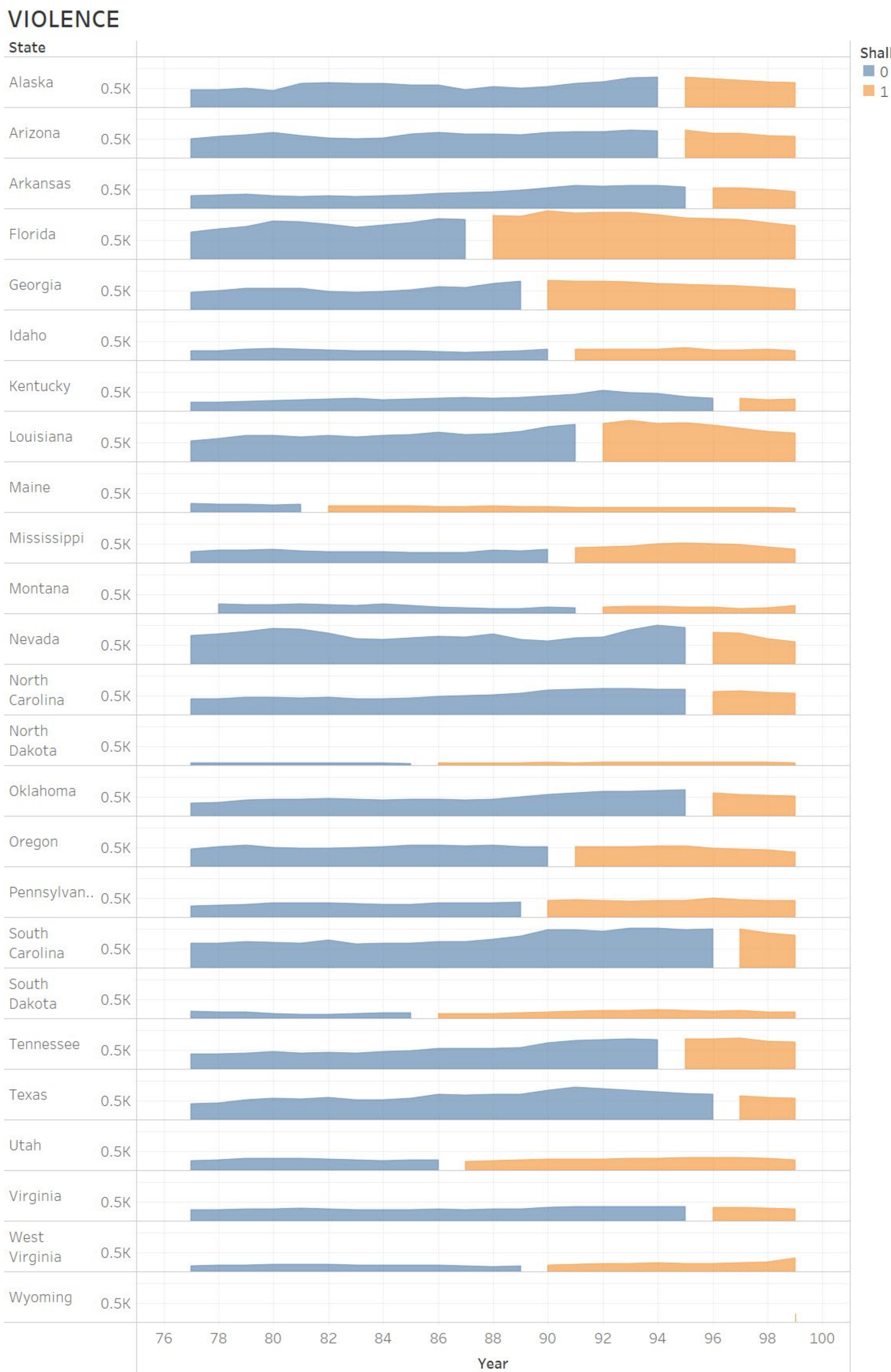
States Where Shall Law Was Implemented Over The Years



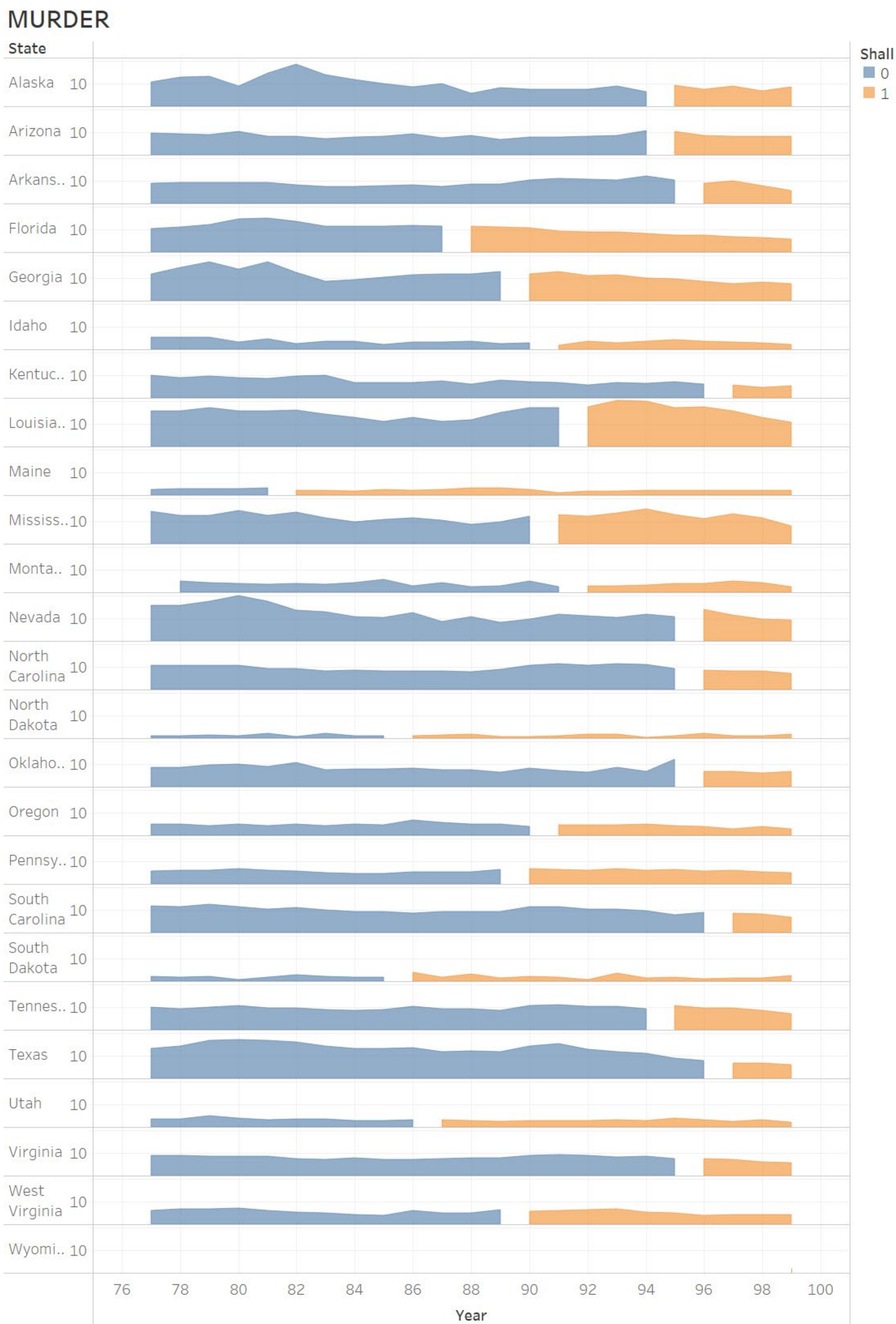
The plot of Total Crime for Year broken down by State. Color shows details about Shall.



The plot of sum of Rob for Year broken down by State. Color shows details about Shall.



The plot of sum of Vio for Year broken down by State. Color shows details about Shall.



The plot of sum of Mur for Year broken down by State. Color shows details about Shall.

Crimes –

Looking at the above visualization, we can conclude that the implementation of the Shall Law was successful and has led to a decline in crimes in several states, excluding Mississippi, where crimes increased after the enforcement.

Robberies –

The number of robberies reduced in every state except Mississippi, where 162 thefts were reported in 1994, the same year when crimes increased despite Shall law enforcement.

Violence –

Gun-related violence cases declined in most states except Florida and Mississippi, where the number of cases increased after implementing the Shall Law. Florida saw a significant decline in cases after 1996, and the number of cases was lower than the times when the law was not implemented.

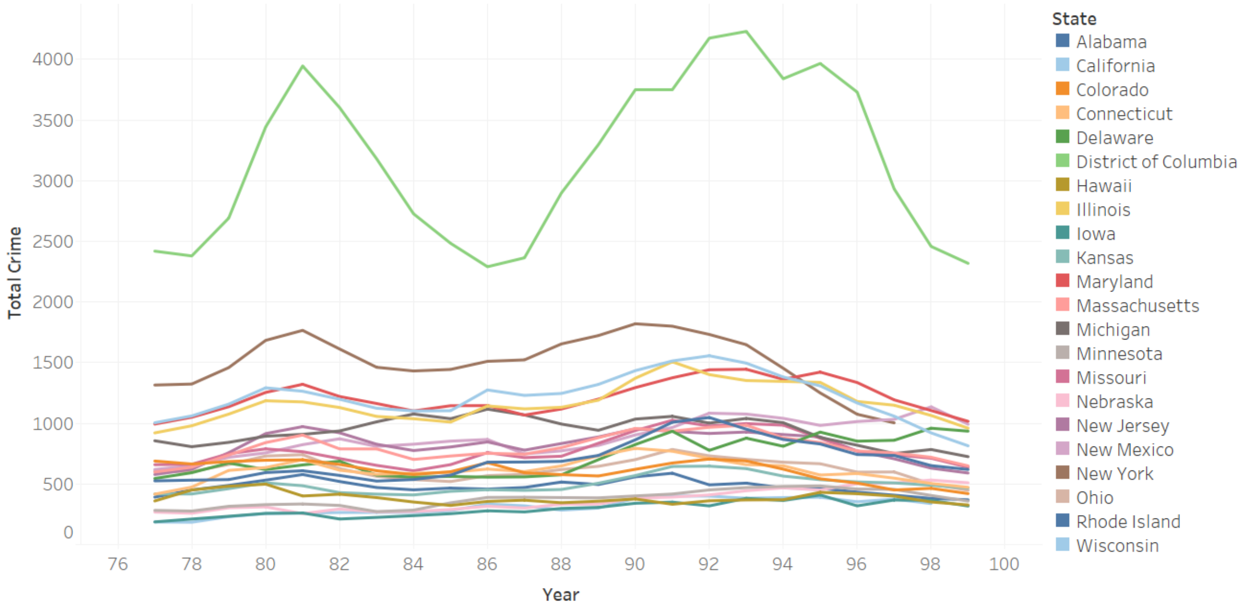
Murder –

The murder rate declined in many states but increased in Mississippi in 1994, the same year when violence, robbery, and crime rates increased.

We can conclude that the enforcement of The Shall law was successful and has led to control in gun-related notorious activities.

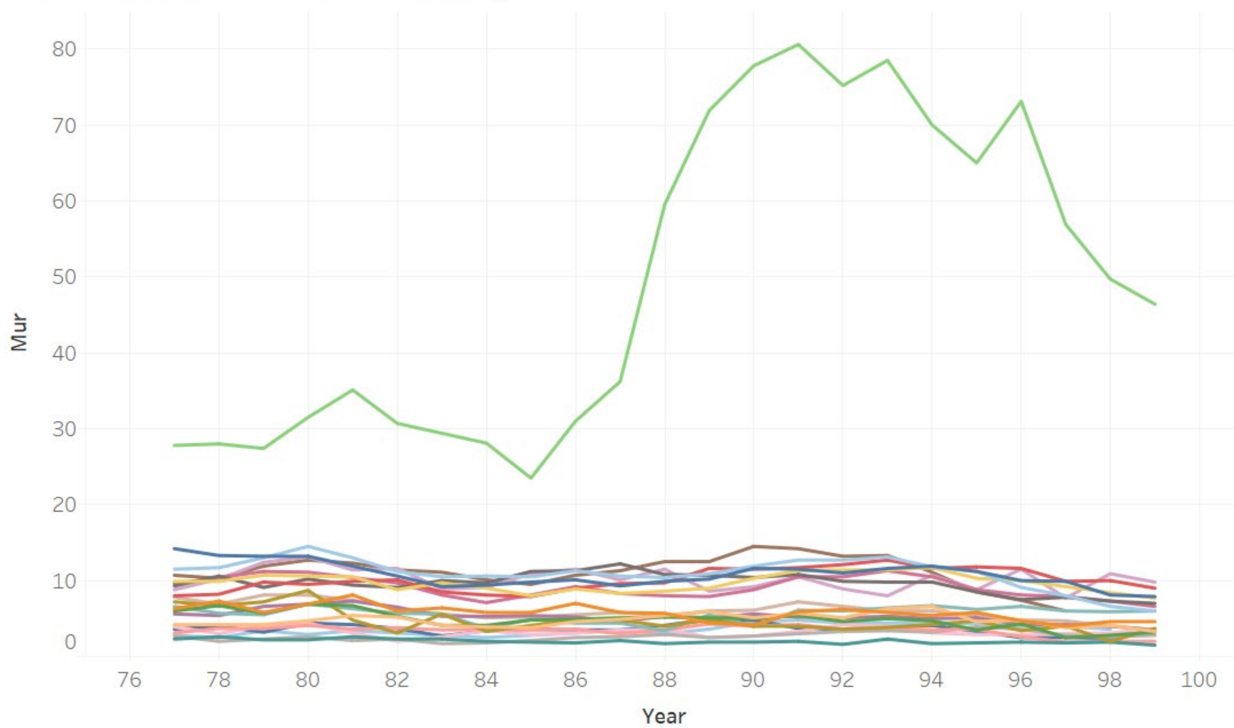
States Where Shall Law Was Not Implemented

Shall Law 0 Total Crime Rate Throughout



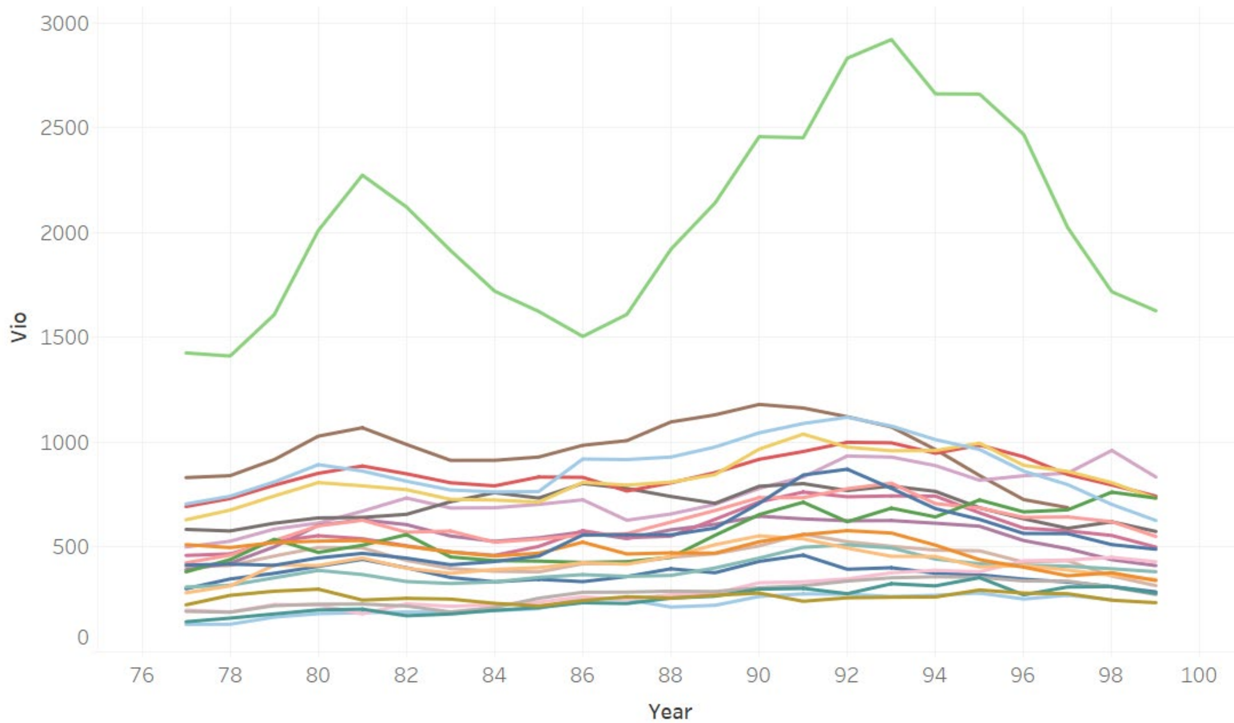
The trend of Total Crime for Year. Color shows details about State. The view is filtered on State, which keeps 22 of 52 members.

Shall Law 0 Murder Rate throughout



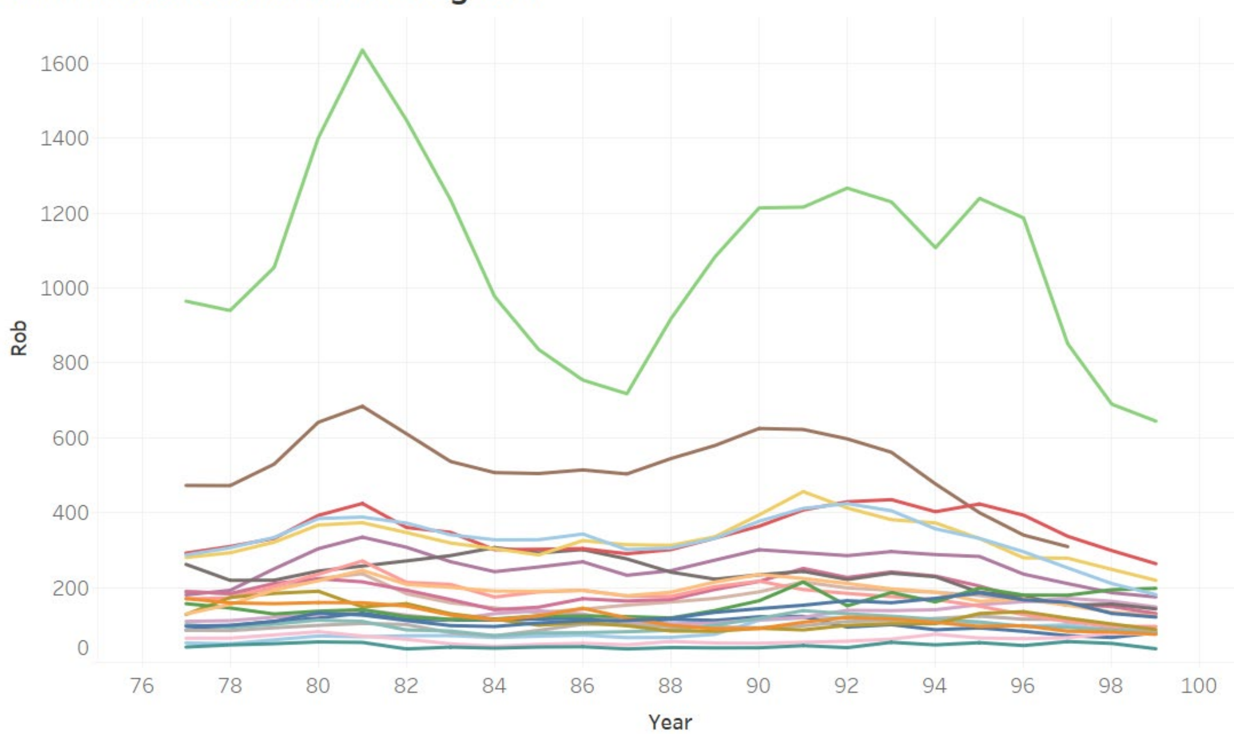
The trend of sum of Mur for Year. Color shows details about State. The view is filtered on State, which keeps 22 of 52 members.

Shall Law 0 Violent Crimes throughout



The trend of sum of Vio for Year. Color shows details about State. The view is filtered on State, which keeps 22 of 52 members.

Shall Law 0 Robberies throughout

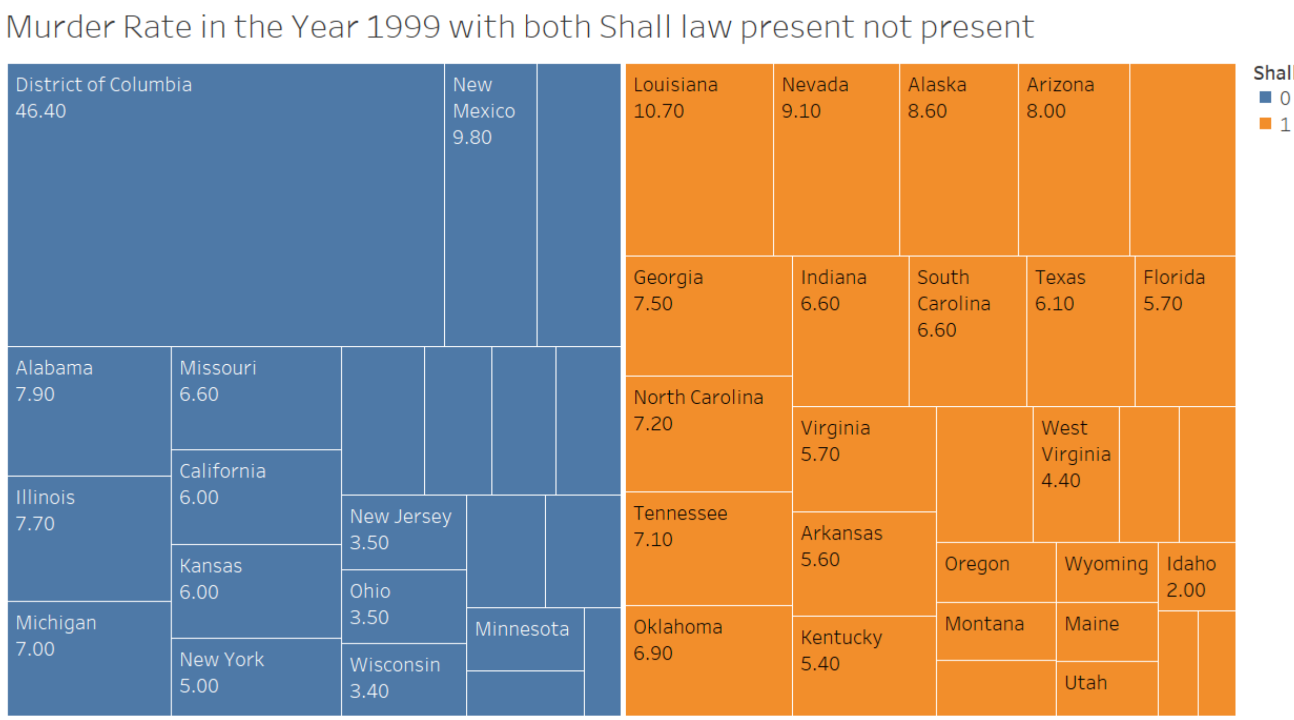


The trend of sum of Rob for Year. Color shows details about State. The view is filtered on State, which keeps 22 of 52 members.

We can see that the District of Columbia has the highest crime rate amongst any other state where Shall Law was not implemented.

In case of robbery, the state of New York seems to have a fluctuation,

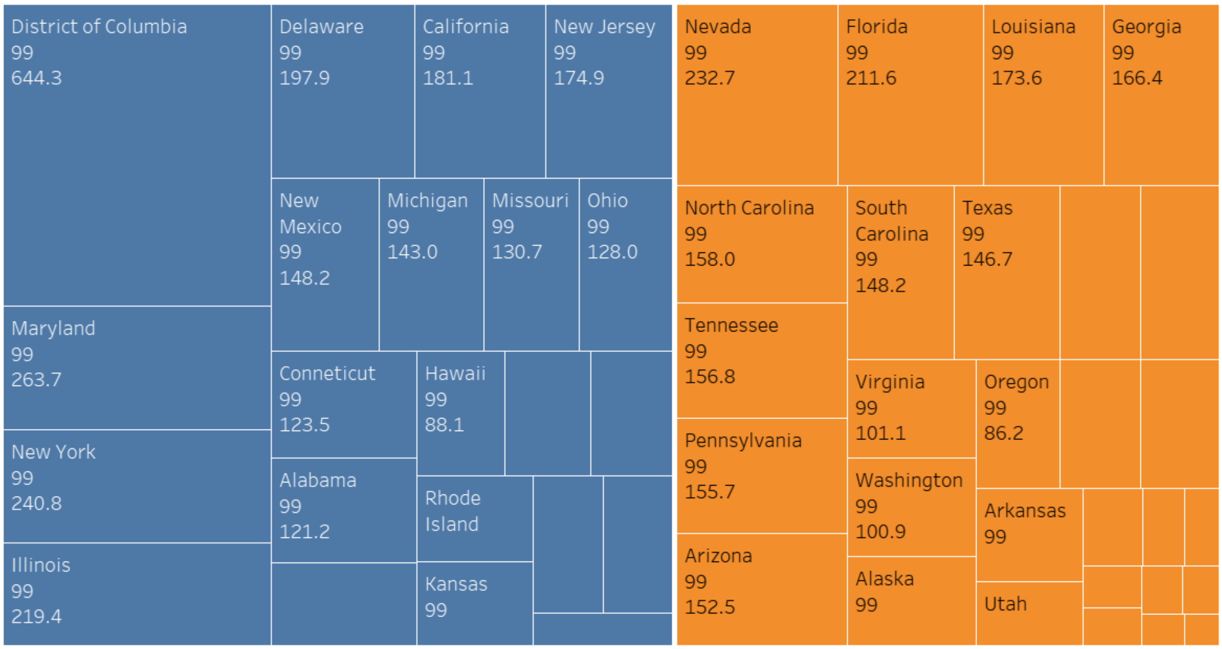
The rest of the states are constant throughout without any noticeable fluctuations.



State and sum of Mur. Color shows details about Shall. Size shows sum of Mur. The marks are labeled by State and sum of Mur. The data is filtered on Year, which ranges from 99 to 99.

As evident from the above tree map, the District of Columbia had the highest murder rate before the implementation of the Shall law whereas Louisiana had the highest murder rate after the implementation of the law.

Avg Rob in the Year 1999 Shall law present/not present

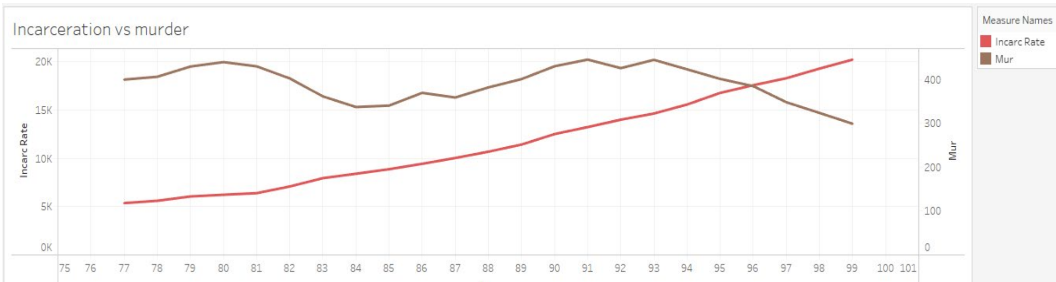


The District of Columbia had the highest number of robberies before the law was implemented whereas Nevada had the highest number of robberies after the law was implemented but still it was less than half of what the district of Columbia recorded before the law was implemented.

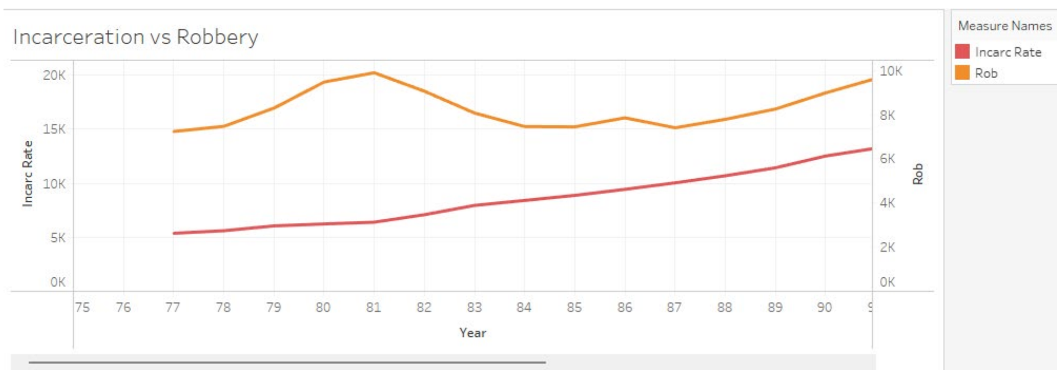
Effect Of Incarceration Rate On Crimes

Incarceration is the state of being confined in prison, imprisonment.
Most of the people who have committed a crime and are in imprisoned, are hesitant to commit a crime again. So as the incarceration rate increases, we should expect a decline in the crime rate across all the 50 states.

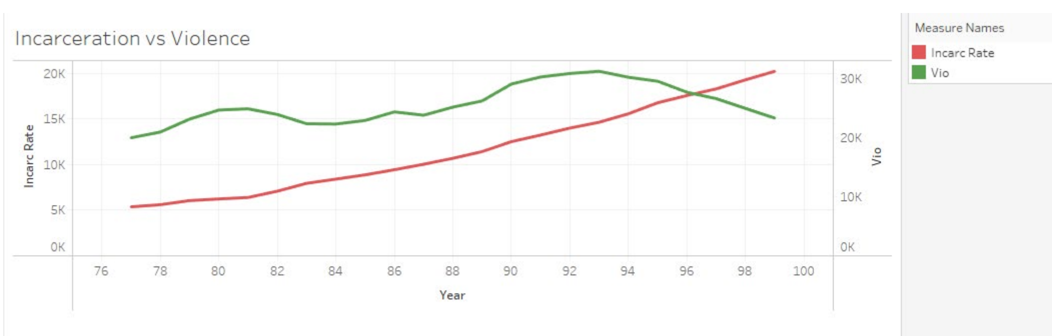
Let us take a look at the following graphs which show the incarceration rate along with murder rate and robbery rate, and violent crime rates.



Incarceration vs Murder



Incarceration vs Robbery



Incarceration vs Violence

The line plot of Incarceration rate vs Violence and Murder follow a very similar trend. We can see that till the year 1995, as the incarceration rate increased, murder and violence also increased.

After the year 1996, we could see a fall in these crimes. However, the Robbery rate has been somewhat constant over the years.

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 five main variables (measured per 100,000 people), on average, over 23 years and 51 states, the violence rate is much higher compared to robbery and murder. However, the standard deviation of violence is smaller than the standard deviation of murder and robbery compared to their mean. Violence is also less skewed than murder and robbery.Although the incarceration rate is also high, it is still much lower than the total crime rate (violence and robbery, and murder).

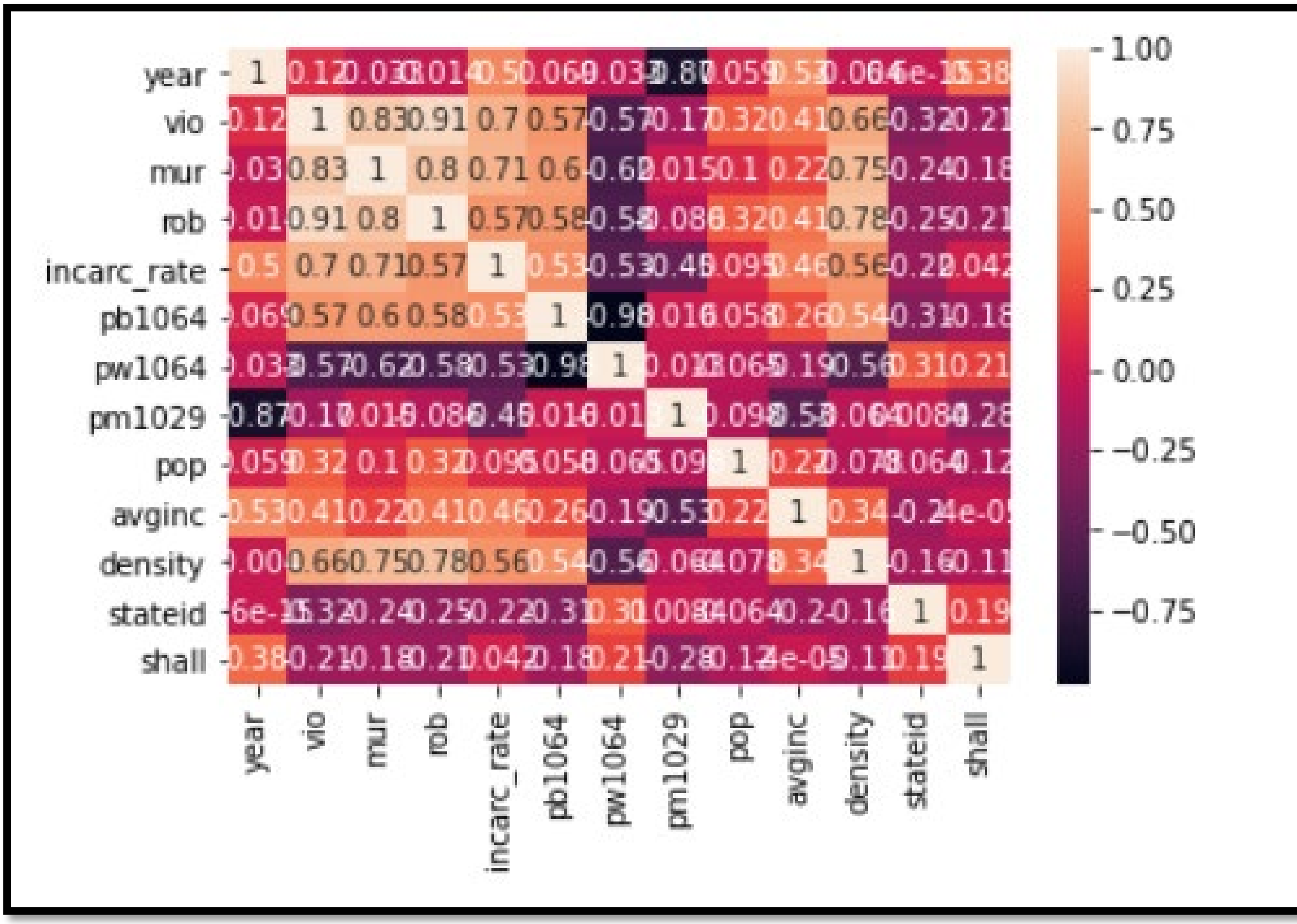
Variable	*pb1064*	*pw1064*	*pm1029*	*pop*	*avginc*	*density*
Mean	5.34	62.95	16.08	4.82	13.72	0.35
Standard Error	0.14	0.29	0.05	0.15	0.07	0.04
Median	4.03	65.06	15.90	3.27	13.40	0.08
Mode		#N/A	#N/A	#N/A	11.66	#N/A
Standard Deviation	4.89	9.76	1.73	5.25	2.55	1.36
Sample Variance	23.87	95.29	3.00	27.58	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

Data Pre-Processing

Determining the Dependent variable

From the dataset, we can clearly see that there are three dependent variables which are vio, rob and mur. Hence, we cannot use any of these 3 variables as independent as these would lead to misleading results.

First, we try to understand the relationship of these variables with all other variables. For this we plot the correlation plot.



One of the important points we should notice is violence, murder and robbery are highly correlated (positive correlation), especially violence and robbery.

Out first option was that we can run 3 sets of models for each of these dependent variables one by one.

There is another possibility that we can create a dependent variable which can be a summation of all these variables.

Also, we need to understand that murder and robbery may also be considered into violent crimes in real life hence there is a possibility that violence variable itself also accounts for these two crimes. But we are not sure of this as we are not sure how the data has been collected.

Hence our approach is that we will be that we will create 2 additional dependent variables as below:

all_crime = vio + rob + mur

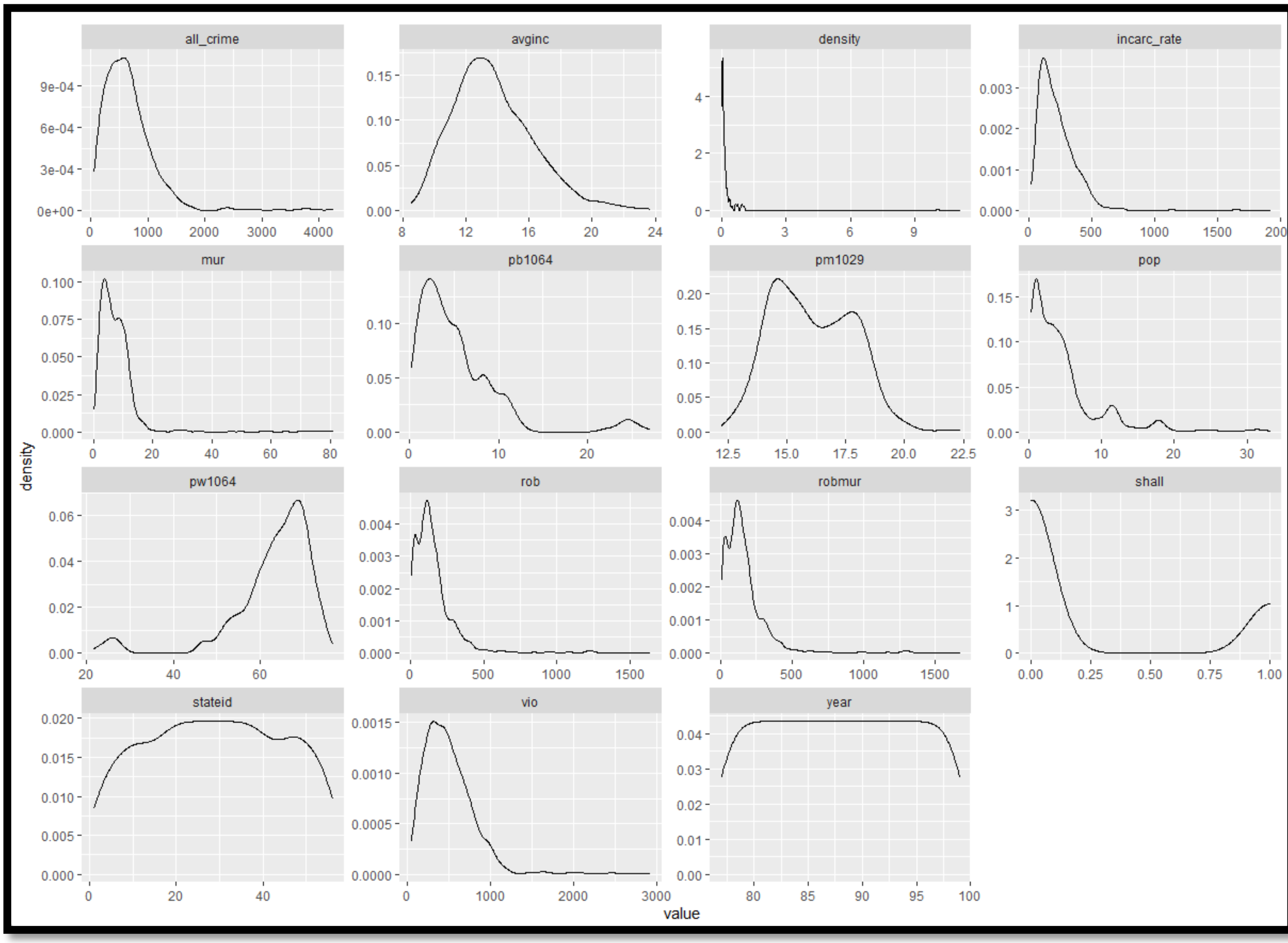
robmur = rob + mur

Our approach will be to run 5 sets of normal OLS models on the data with these different dependent variables and then try to figure the best two or three dependent variables.

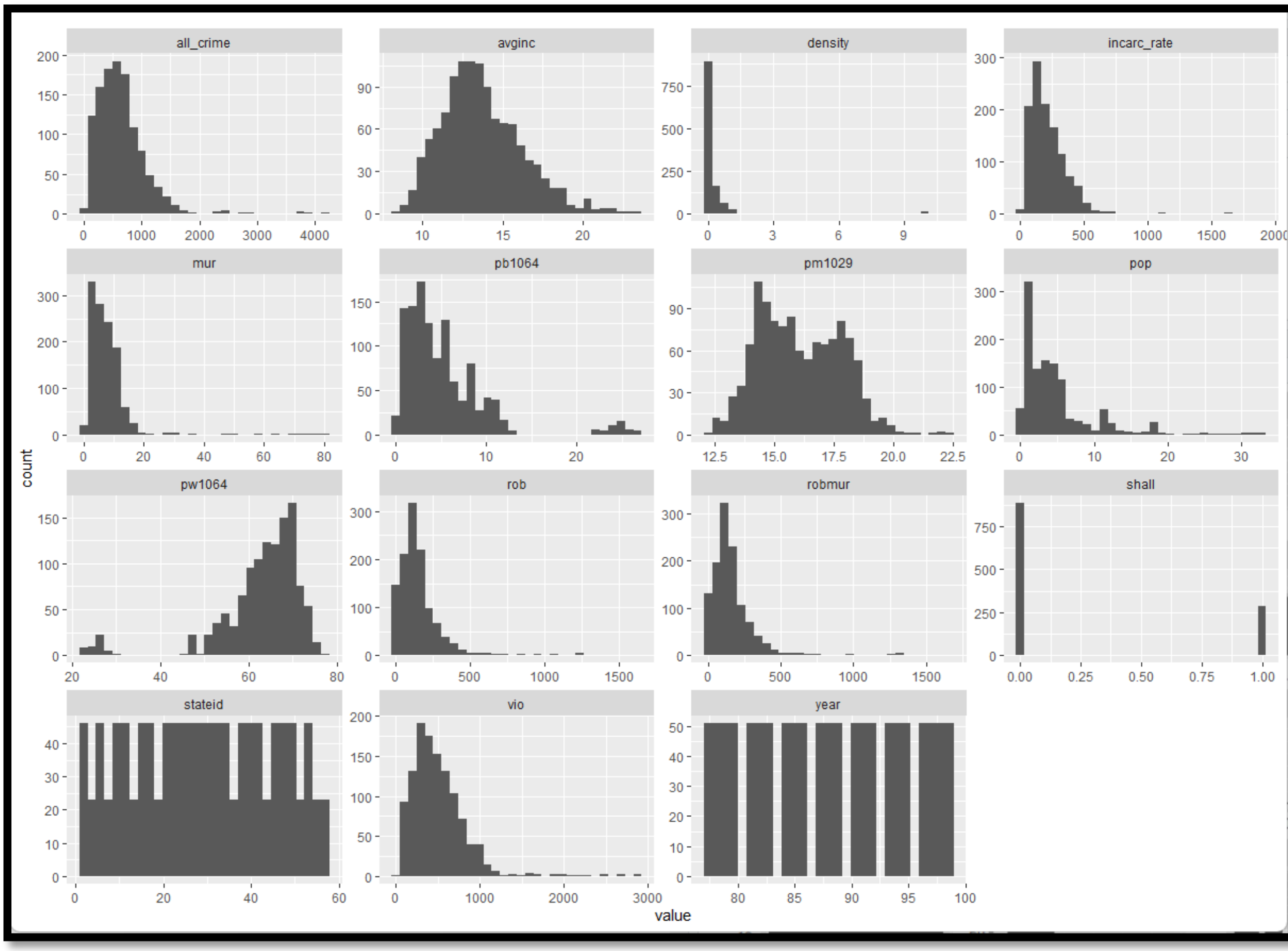
Checking the distribution of dependent and explanatory variables.

We first do a density plot and histogram plot of all the variables in the dataset.

Density Plot:



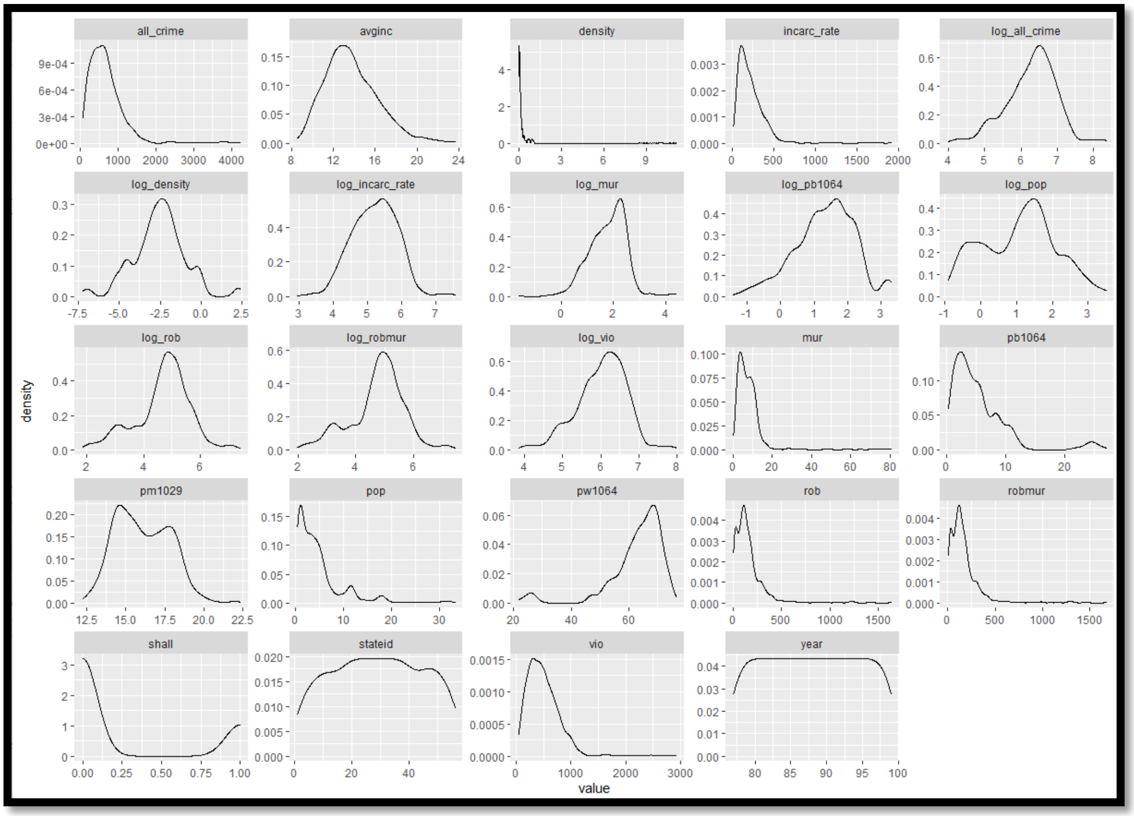
Histogram Plot :

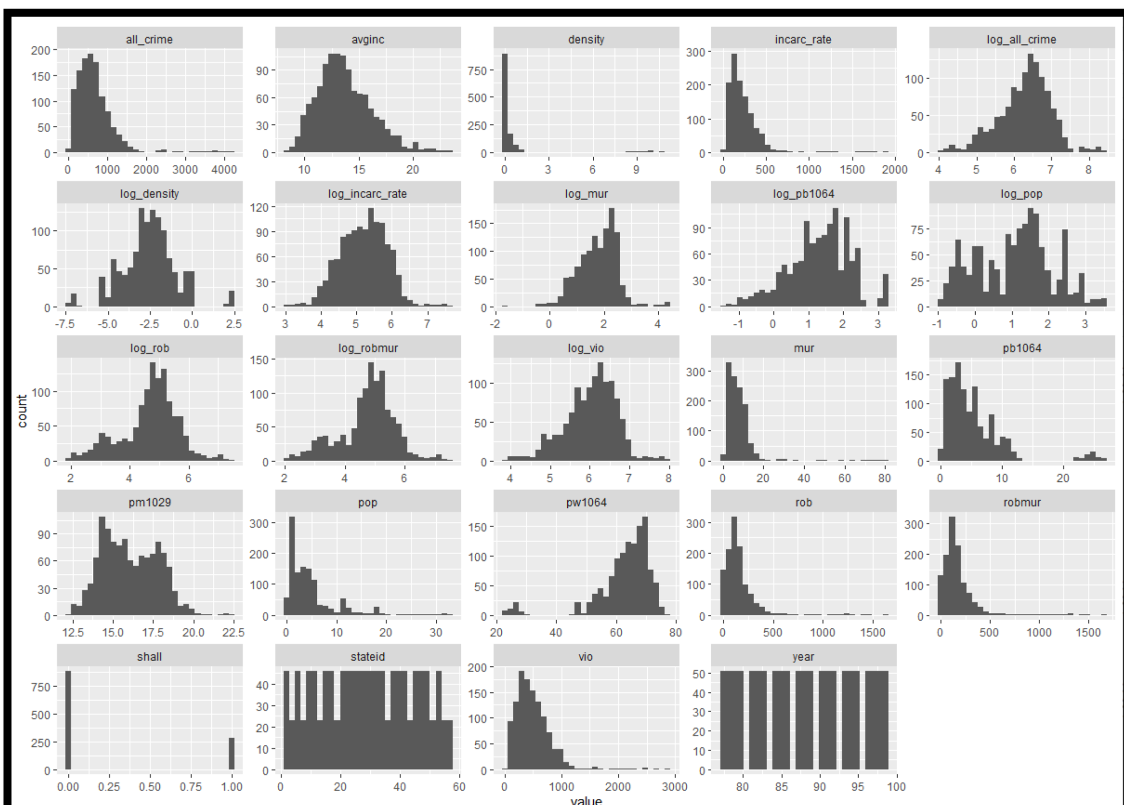


From the above we can clearly see that for the variables **vio, all_crime, incar_rate, robmur, rob, mur, pb1064** are positively skewed. Thus, we use the ln() in order to take logarithmic transform of these variables as this will assist in dealing with this issue and help us bring the distribution of these variables to a somewhat normal distribution. This would further help us handle issues of 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.

Below are the plots of the variables after taking a logarithmic transform of these variables.


all_crime
avginc
density
incarc_rate
log_all_crime
log_density
log_incarc_rate
log_mur
log_pb1064
log_pop
log_rob
log_robmur
log_vio
mur
pb1064
pm1029
pop
pw1064
rob
robmur
shall
stateid
vio
year
density
value


all_crime
avginc
density
incarc_rate
log_all_crime
log_density
log_incarc_rate
log_mur
log_pb1064
log_pop
log_rob
log_robmur
log_vio
mur
pb1064
pm1029
pop
pw1064
rob
robmur
shall
stateid
vio
year
count
value

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.

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 variables . 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

Log Transformations Made:

```
guns_data$log_vio <- log(guns_data$vio)

guns_data$log_all_crime <- log(guns_data$all_crime)

guns_data$log_density <- log(guns_data$density)

guns_data$log_incarc_rate <- log(guns_data$incarc_rate)

guns_data$log_mur <- log(guns_data$mur)

guns_data$log_pop <- log(guns_data$pop)

guns_data$log_rob <- log(guns_data$rob)

guns_data$log_robmur <- log(guns_data$robmur)

guns_data$log_pb1064 <- log(guns_data$pb1064)
```

Expected relationship between the dependent and explanatory variables:

Variable	Expected Sign	Reasoning
shall	-ve	States having shall-carry law in effect tend to have less overall crime rates. This is supported by the fact that if shall law is implemented in a State, people would be able to protect them as gun would be a self-defense, and the crime rate would decrease.
Incarc_rate	-ve	We expect that higher the incarceration rate less should be the overall crime rate. This is supported by the fact that as more criminals would be put in the prison , the crime rate is expected to decrease.
density	+ve	Higher population density increases overall crime rate. This is supported by the fact that as more people would be living in a particular place, there are more chances of them to get into some conflicts and hence , the crime rate would increase.
Avginc	-ve	Higher average income reduces overall crime rate. This is supported by the fact that if a state has people with more average income, it implies that more educated people live there and hence, less chances of criminals in that state.
Pop	+ve	More people increase overall crime rate. This is supported by the fact that as more people would be living in a state, there are more chances of crimes happening and hence , the crime rate would increase.
pm1029	+ve	More the young male population greater should be the overall crime rate. This is supported by the fact that younger the population, between age 10-29, they are more aggressive and sensitive towards certain situations, there are more chances of crimes happening by this population and hence , the crime rate would increase.
pw1064	-ve	We expect a decrease in overall crime rate with higher population of white people.
pb1064	+ve	We expect greater overall crime rate in states with higher population of black people. This is supported by the fact that black people do not have enough money for survival, and they therefore involve in crimes in attempt to maintain their survival. And hence , the crime rate would increase.
year	+ve	We expect that the overall crime rate is on the rise over the years. This is supported by the fact that as the year passes, the population increases and hence as stated earlier , the crime rate would increase.

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.

Initially we run different Pooled OLS Models for each of the dependent variables separately to see if there is a common pattern amongst all those or few of them at least. We are using the untransformed data initially.

Model 1 : Dependent variable = all_crime

```
> ols_model2<-lm(all_crime~incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+year,data = guns_data)
> summary(ols_model2)

Call:
lm(formula = all_crime ~ incarc_rate + pb1064 + pw1064 + pm1029 +
    pop + avginc + density + shall + year, data = guns_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1639.38  -141.46   -33.54   134.42  1280.48

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1852.21855  392.26850   4.722 2.62e-06 ***
incarc_rate    1.13926    0.06431  17.714  < 2e-16 ***
pb1064        43.07677    9.66997   4.455 9.21e-06 ***
pw1064        15.25025    4.82638   3.160 0.001620 **
pm1029       -51.00511    9.47109  -5.385 8.74e-08 ***
pop           28.30154    1.43179  19.766  < 2e-16 ***
avginc         7.67851    4.34524   1.767 0.077472 .
density      155.36919    7.89506  19.679  < 2e-16 ***
shall        -71.95048   18.95742  -3.795 0.000155 ***
year         -23.70777    2.68145  -8.841  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 237.9 on 1163 degrees of freedom
Multiple R-squared:  0.7757,    Adjusted R-squared:  0.774
F-statistic: 446.9 on 9 and 1163 DF,  p-value: < 2.2e-16
```

Model 2 : Dependent variable = robmur

```
> ols_model3<-lm(robmur~incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+year,data = guns_data)
> summary(ols_model3)

Call:
lm(formula = robmur ~ incarc_rate + pb1064 + pw1064 + pm1029 +
    pop + avginc + density + shall + year, data = guns_data)

Residuals:
    Min      1Q  Median      3Q     Max
-494.95  -39.47  -12.97   28.52  617.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 769.6697   126.2708   6.095 1.48e-09 ***
incarc_rate   0.2019     0.0207   9.754  < 2e-16 ***
pb1064       17.7950     3.1128   5.717 1.38e-08 ***
pw1064        6.2250     1.5536   4.007 6.55e-05 ***
pm1029      -23.3175     3.0487  -7.648 4.25e-14 ***
pop          10.5672     0.4609  22.928  < 2e-16 ***
avginc        4.5237     1.3987   3.234  0.00125 **
density      75.5506     2.5414  29.728  < 2e-16 ***
shall        -7.3643     6.1024  -1.207  0.22775
year        -10.1767     0.8632 -11.790  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 76.59 on 1163 degrees of freedom
Multiple R-squared:  0.8133,	Adjusted R-squared:  0.8118
F-statistic: 562.8 on 9 and 1163 DF,  p-value: < 2.2e-16
```

Model 3 : Dependent variable = vio

```
> ols_model4<-lm(vio~incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+year,data = guns_data)
> summary(ols_model4)

Call:
lm(formula = vio ~ incarc_rate + pb1064 + pw1064 + pm1029 + pop +
    avginc + density + shall + year, data = guns_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1144.43  -108.82   -27.42    98.45   662.97

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1082.54885  284.93386   3.799 0.000153 ***
incarc_rate    0.93734    0.04672  20.065  < 2e-16 ***
pb1064        25.28174    7.02402   3.599 0.000332 ***
pw1064         9.02527    3.50576   2.574 0.010164 *
pm1029       -27.68758    6.87956  -4.025 6.08e-05 ***
pop           17.73433    1.04002  17.052  < 2e-16 ***
avginc         3.15483    3.15627   1.000 0.317739
density       79.81855    5.73477  13.918  < 2e-16 ***
shall        -64.58613   13.77019  -4.690 3.05e-06 ***
year         -13.53112    1.94774  -6.947 6.19e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 172.8 on 1163 degrees of freedom
Multiple R-squared:  0.7347,	Adjusted R-squared:  0.7327
F-statistic: 357.9 on 9 and 1163 DF,  p-value: < 2.2e-16
```

Model 4 : Dependent variable = rob

```
> ols_model5<-lm(rob~incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+year,data = guns_data)
> summary(ols_model5)

Call:
lm(formula = rob ~ incarc_rate + pb1064 + pw1064 + pm1029 + pop +
    avginc + density + shall + year, data = guns_data)

Residuals:
    Min      1Q  Median      3Q     Max
-475.98  -38.05  -11.59   27.94  620.51

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 747.38210  123.75844   6.039 2.08e-09 ***
incarc_rate   0.17480    0.02029   8.615  < 2e-16 ***
pb1064       17.41777    3.05082   5.709 1.44e-08 ***
pw1064        6.10083    1.52269   4.007 6.55e-05 ***
pm1029      -23.41135    2.98808  -7.835 1.05e-14 ***
pop          10.42024    0.45172  23.068  < 2e-16 ***
avginc        4.81072    1.37090   3.509 0.000467 ***
density      73.42286    2.49084  29.477  < 2e-16 ***
shall        -6.98472    5.98096  -1.168 0.243116
year         -9.84109    0.84598 -11.633  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 75.07 on 1163 degrees of freedom
Multiple R-squared:  0.8076,    Adjusted R-squared:  0.8062
F-statistic: 542.6 on 9 and 1163 DF,  p-value: < 2.2e-16
```

Model 5 : Dependent variable = mur

```
> ols_model6<-lm(mur~incarc_rate+pb1064+pw1064+pm1029+pop+avginc+density+shall+year,data = guns_data)
> summary(ols_model6)

Call:
lm(formula = mur ~ incarc_rate + pb1064 + pw1064 + pm1029 + pop +
    avginc + density + shall + year, data = guns_data)

Residuals:
     Min       1Q   Median       3Q      Max
-18.9732  -1.8033  -0.1032   1.5856  29.4444

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.2875944  5.6262357   3.961 7.91e-05 ***
incarc_rate  0.0271237  0.0009224  29.405  < 2e-16 ***
pb1064       0.3772531  0.1386946   2.720  0.00662 **
pw1064       0.1241368  0.0692239   1.793  0.07319 .
pm1029       0.0938199  0.1358422   0.691  0.48992
pop          0.1469746  0.0205360   7.157 1.46e-12 ***
avginc      -0.2870428  0.0623230  -4.606 4.56e-06 ***
density      2.1277742  0.1132374  18.790  < 2e-16 ***
shall       -0.3796352  0.2719029  -1.396  0.16292
year        -0.3355558  0.0384596  -8.725  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.413 on 1163 degrees of freedom
Multiple R-squared:  0.7958,    Adjusted R-squared:  0.7942
F-statistic: 503.5 on 9 and 1163 DF,  p-value: < 2.2e-16
```

From the above results, we can see that for model 4 and 5, the coefficients for shall law are not significant. The same is observed for model 2 as well. Now as all three of there models are related to Robbery, Murder, and a combination of both, and are showing similar results , hence we decided to keep only 1 one of these three. We decided to go with the model 2 (dependent variable robmur) as it is covering both robbery and murder.

For model 1 and mode 3, we see very different results, hence we decided to explore more on these as well.

To further explore these models, we decided to run model 1, 2 and 3 with the transformed variables to see if the transform is helping us get better or significant results.

Model 1 : Dependent variable = log_all_crime

```
> ols_model7<-lm(log_all_crime~log_incarc_rate+log_pb1064+pw1064+pm1029+log_pop+avginc+log_density+shall,data = guns_data)
> summary(ols_model7)

Call:
lm(formula = log_all_crime ~ log_incarc_rate + log_pb1064 + pw1064 + 
    pm1029 + log_pop + avginc + log_density + shall, data = guns_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2353 -0.2114  0.0052  0.2440  1.0176 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.202996   0.312541   0.650 0.516142    
log_incarc_rate  0.544610   0.028345  19.213  < 2e-16 ***
log_pb1064       0.210511   0.032095   6.559 8.13e-11 ***
pw1064           0.009783   0.002533   3.863 0.000118 ***
pm1029           0.120695   0.009531  12.663  < 2e-16 ***
log_pop          0.146575   0.014647  10.007  < 2e-16 ***
avginc           0.041325   0.005464   7.563 7.96e-14 ***
log_density      0.098975   0.009695  10.208  < 2e-16 ***
shall           -0.229697   0.027817  -8.258 4.00e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.365 on 1164 degrees of freedom
Multiple R-squared:  0.7228,	Adjusted R-squared:  0.7208 
F-statistic: 379.3 on 8 and 1164 DF,  p-value: < 2.2e-16
```

Model 2 : Dependent variable = log_robmur

```
> ols_model8<-lm(log_robmur~log_incarc_rate+log_pb1064+pw1064+pm1029+log_pop+avginc+log_density+shall,data = guns_data)
> summary(ols_model8)

Call:
lm(formula = log_robmur ~ log_incarc_rate + log_pb1064 + pw1064 + 
    pm1029 + log_pop + avginc + log_density + shall, data = guns_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.42626 -0.27531 -0.00696  0.28343  1.59484 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -1.598302   0.391637  -4.081 4.79e-05 ***
log_incarc_rate  0.407243   0.035519  11.466  < 2e-16 ***
log_pb1064       0.329585   0.040218   8.195 6.55e-16 ***
pw1064           0.008793   0.003174   2.770  0.00569 ** 
pm1029           0.156339   0.011944  13.090  < 2e-16 ***
log_pop          0.263710   0.018354  14.368  < 2e-16 ***
avginc           0.069929   0.006847  10.213  < 2e-16 ***
log_density      0.177913   0.012149  14.644  < 2e-16 ***
shall           -0.265063   0.034856  -7.604 5.87e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4573 on 1164 degrees of freedom
Multiple R-squared:  0.7579,	Adjusted R-squared:  0.7562 
F-statistic: 455.4 on 8 and 1164 DF,  p-value: < 2.2e-16
```

Model 3 : Dependent variable = log_vio

```
> ols_model9<-lm(log_vio~log_incarc_rate+log_pb1064+pw1064+pm1029+log_pop+avginc+log_density+shall,data = guns_data)
> summary(ols_model9)

Call:
lm(formula = log_vio ~ log_incarc_rate + log_pb1064 + pw1064 + 
    pm1029 + log_pop + avginc + log_density + shall, data = guns_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.18347 -0.21169  0.01828  0.23995  1.05270 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.003363   0.304667  -0.011    0.991    
log_incarc_rate  0.580641   0.027631  21.014  < 2e-16 ***
log_pb1064       0.188064   0.031287   6.011 2.46e-09 ***
pw1064           0.010954   0.002469   4.437 1.00e-05 ***
pm1029           0.109664   0.009291  11.803  < 2e-16 ***
log_pop          0.119069   0.014278   8.339  < 2e-16 ***
avginc           0.031646   0.005326   5.941 3.73e-09 ***
log_density      0.078625   0.009451   8.319 2.46e-16 ***
shall           -0.220354   0.027116  -8.126 1.12e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3558 on 1164 degrees of freedom
Multiple R-squared:  0.6985,	Adjusted R-squared:  0.6964 
F-statistic: 337.1 on 8 and 1164 DF,  p-value: < 2.2e-16
```

Interpretation Of the Pooled OLS Model

From the above model results , we have concluded that the Pooled OLS Model, is not doing a very good job in estimating the coefficients.

We see that the estimate of shall(shall law) is -0.22 in model 1 and 3 and -0.26 in model 2 which basically states that the places where the shall carry law is in place have around 22% 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 22% 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 as per our expectations. This bias is due to the unobserved heterogeneity that is hiding in the error term, this omitted variable like cultural attitude towards the shall law w which are different for different entities and are time invariant within entities are correlated with our explanatory variables such as shall leading to a endogeneity problem thus in a process making it downwardly biased.

Also, regarding the other estimates , even though all of them are very significant in all the three models, but in all these models, the sign of the coefficients for some explanatory variables is different from our expectation. For example, we expected that as the incarceration rates would go up the overall crime rate must come down(negative relationship), but the positive estimate of 0.4 – 0.6 in all the models.

Similarly, **avginc** and **pw1064** also do not have estimates of the expected sign.

Pooled OLS With Cluster Robust Standard Errors

The above models are unbiased and consistent. But they are not efficient and also, they have incorrect Standard Errors.

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

```
> x1<-coeftest(ols_model7, vcov=vcovHC(ols_model7, cluster="group"))
> x1

t test of coefficients:

                 Estimate Std. Error t value  Pr(>|t|)    
(Intercept)     0.2029960  0.3271021  0.6206 0.5349916    
log_incarc_rate 0.5446102  0.0309285 17.6087 < 2.2e-16 ***
log_pb1064      0.2105111  0.0314658  6.6902 3.455e-11 ***
pw1064          0.0097835  0.0025578  3.8250 0.0001377 ***
pm1029          0.1206951  0.0097655 12.3594 < 2.2e-16 ***
log_pop         0.1465749  0.0143828 10.1910 < 2.2e-16 ***
avginc          0.0413249  0.0051150  8.0792 1.619e-15 ***
log_density     0.0989747  0.0092715 10.6752 < 2.2e-16 ***
shall          -0.2296973  0.0272404 -8.4322 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> x2<-coeftest(ols_model8, vcov=vcovHC(ols_model8, cluster="group"))
> x2

t test of coefficients:

                 Estimate Std. Error t value  Pr(>|t|)    
(Intercept)    -1.5983023  0.4284181 -3.7307 0.0002001 ***
log_incarc_rate 0.4072434  0.0405229 10.0497 < 2.2e-16 ***
log_pb1064      0.3295850  0.0396576  8.3108 2.622e-16 ***
pw1064          0.0087925  0.0031392  2.8009 0.0051801 ** 
pm1029          0.1563393  0.0121521 12.8652 < 2.2e-16 ***
log_pop         0.2637104  0.0181281 14.5470 < 2.2e-16 ***
avginc          0.0699294  0.0062599 11.1710 < 2.2e-16 ***
log_density     0.1779132  0.0111427 15.9669 < 2.2e-16 ***
shall          -0.2650633  0.0344469 -7.6948 3.009e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> x3<-coeftest(ols_model9, vcov=vcovHC(ols_model9, cluster="group"))
> x3

t test of coefficients:

                  Estimate Std. Error t value  Pr(>|t|)    
(Intercept)     -0.0033633  0.3124797 -0.0108    0.9914    
log_incarc_rate  0.5806415  0.0294240 19.7336 < 2.2e-16 ***
log_pb1064       0.1880636  0.0306829  6.1293 1.207e-09 ***
pw1064           0.0109542  0.0024499  4.4713 8.534e-06 ***
pm1029           0.1096643  0.0095284 11.5092 < 2.2e-16 ***
log_pop          0.1190692  0.0137382  8.6670 < 2.2e-16 ***
avginc           0.0316459  0.0050409  6.2779 4.835e-10 ***
log_density      0.0786248  0.0089523  8.7827 < 2.2e-16 ***
shall           -0.2203540  0.0268593 -8.2040 6.103e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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 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.

Problem Of Heteroskedasticity and Serially Correlated errors

The dataset on 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 won't 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.

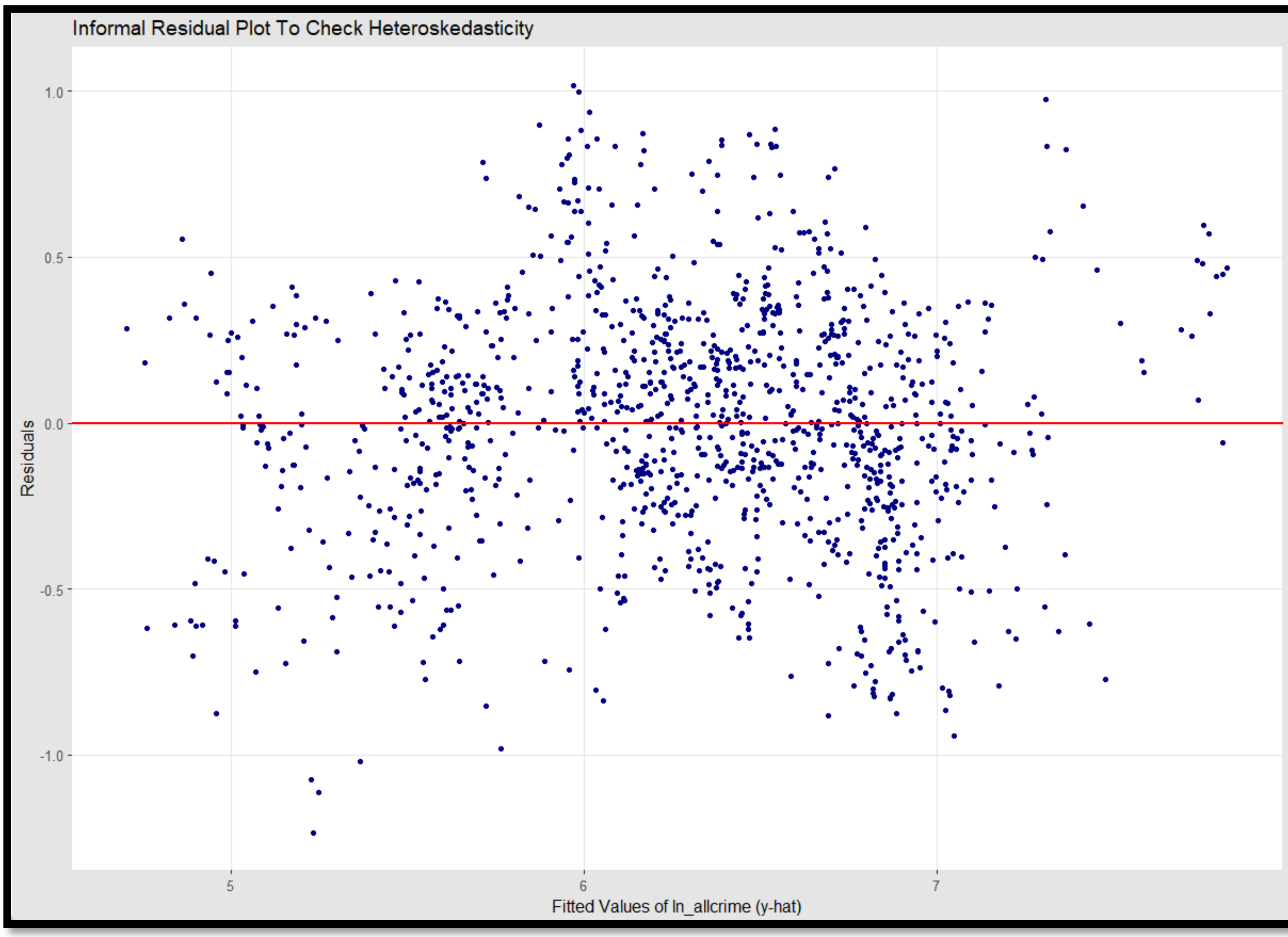
Testing for heteroskedasticity

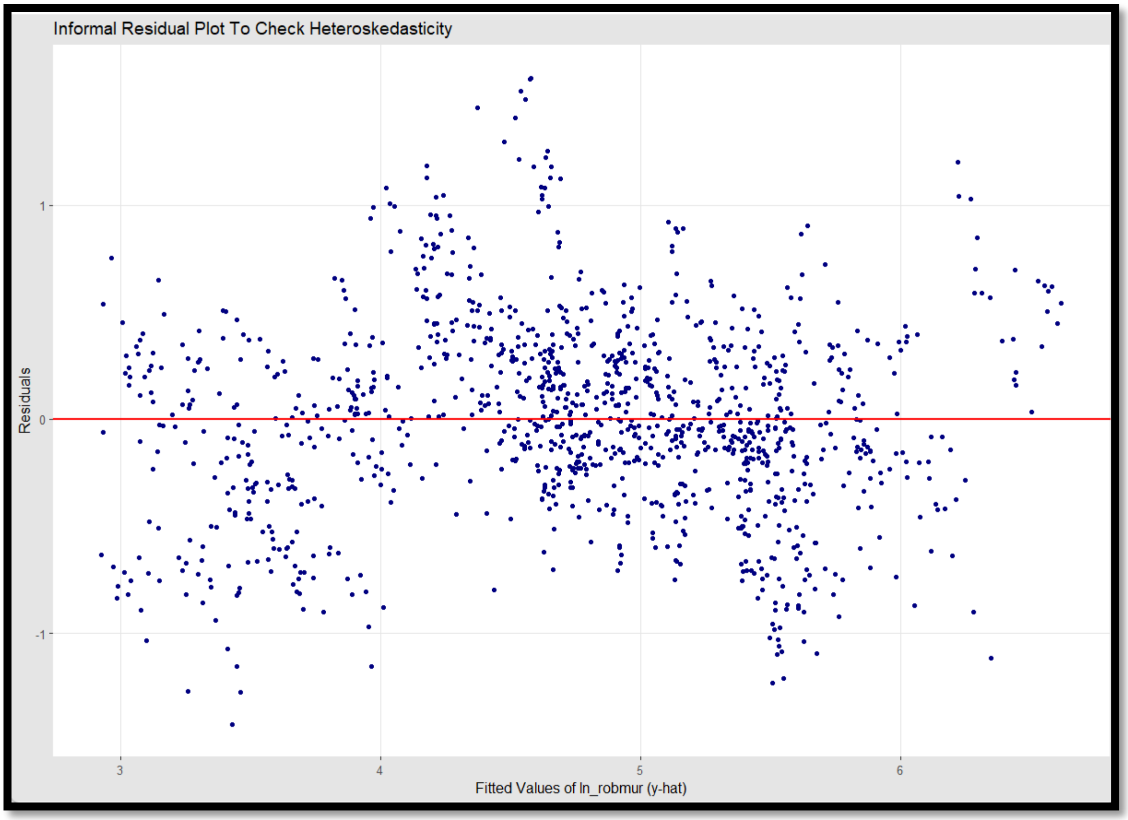
For testing for weather there is heteroskedasticity in our model we take two approaches:

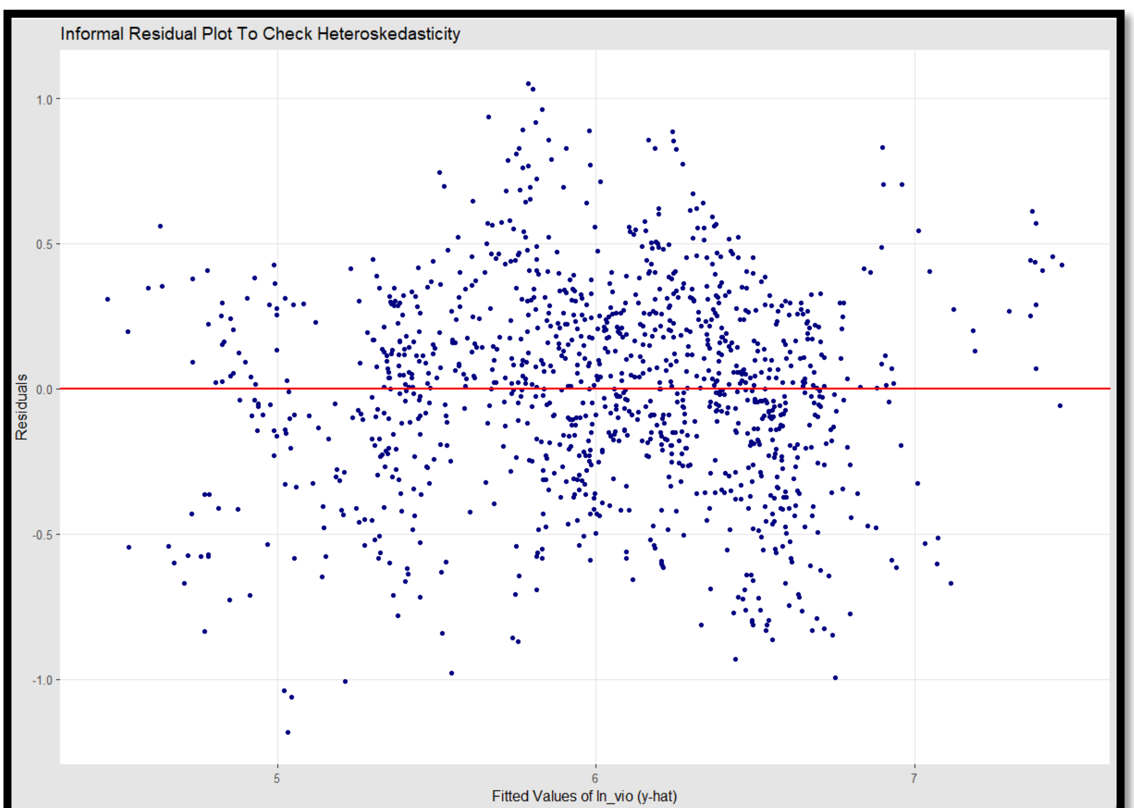
1) Informal approach: In which we plot the residuals against the fitted values of our dependent variable and look for any kind of pattern, if we find any kind of pattern, we say that the informal test shows presence of heteroskedasticity.

2) Formal Whites Test For heteroskedasticity

First, we check informally using residual plots for our three models:




Informal Residual Plot To Check Heteroskedasticity
Residuals
1
0
-1
3
4
5
6
Fitted Values of ln_robmur (y-hat)


Informal Residual Plot To Check Heteroskedasticity
Residuals
1.0
0.5
0.0
-0.5
-1.0
5
6
7
Fitted Values of ln_vio (y-hat)

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.

To confirm the same, now we will do the Formal White Test to confirm our informal analysis.

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

Null Hypothesis: H0: All the alphas (coefficient of z) in the error form equation are 0 (Homoskedasticity)

Alternative Hypothesis: H1: At least one of the alphas 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:

```
> kable(tidy(bptest(ols_model7,~log_incarc_rate+log_pb1064+pw1064+pm1029+log_pop+avginc+log_density+shall,data = guns_data)))


| statistic|  p.value| parameter|method                          |
|---------:|--------:|---------:|:-------------------------------|
|  32.95842| 6.27e-05|         8|studentized Breusch-Pagan test |
> kable(tidy(bptest(ols_model8,~log_incarc_rate+log_pb1064+pw1064+pm1029+log_pop+avginc+log_density+shall,data = guns_data)))


| statistic| p.value| parameter|method                          |
|---------:|-------:|---------:|:-------------------------------|
|  84.36971|       0|         8|studentized Breusch-Pagan test |
> kable(tidy(bptest(ols_model9,~log_incarc_rate+log_pb1064+pw1064+pm1029+log_pop+avginc+log_density+shall,data = guns_data)))


| statistic|  p.value| parameter|method                          |
|---------:|--------:|---------:|:-------------------------------|
|  31.32495| 0.000123|         8|studentized Breusch-Pagan test |
>
```

From the above test we get a very high statistic of 30-80 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.

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.

NOTE: As we can see that all the 3 different models related to the different dependent variables are giving similar results hence its safe to assume that we can continue with only one of these models for further analysis as separate analysis will not give us anything different.

In order to cover all the crimes hence we have decided to go with the all_crime dependent variable.

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.

Dependent variable = log_all_crime

```
Oneway (individual) effect Within Model

Call:
plm(formula = log_all_crime ~ log_incarc_rate + log_pb1064 +
    pw1064 + pm1029 + log_pop + avginc + log_density + factor_shall,
    data = p_dat, model = "within")

Balanced Panel: n = 51, T = 23, N = 1173

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.5725640 -0.1056221  0.0028253  0.1098533  0.5226958

Coefficients:
                   Estimate  Std. Error t-value  Pr(>|t|)
log_incarc_rate -0.02700857  0.02811450 -0.9607  0.336930
log_pb1064      -0.18386311  0.05596666 -3.2852  0.001051 **
pw1064           0.02000643  0.00423583  4.7231 2.618e-06 ***
pm1029          -0.05586563  0.00842540 -6.6306 5.200e-11 ***
log_pop         -0.82063209  1.50158643 -0.5465  0.584825
avginc           0.00060575  0.00599908  0.1010  0.919589
log_density      0.84400576  1.51898969  0.5556  0.578571
factor_shall1    0.00250021  0.01885629  0.1326  0.894539
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    33.904
Residual Sum of Squares: 29.721
R-Squared:      0.12337
Adj. R-Squared: 0.077729
F-statistic: 19.5971 on 8 and 1114 DF, p-value: < 2.22e-16
```

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 its estimates has become unbiased and consistent the current estimate is in the range of -0.02 to -0.04 form the previous estimate of -0.22, 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.

Most of the variables are insignificant in all the models. Variables like Incarceration Rate which are so important are also only significant in case of Model 3 and not in other two models.

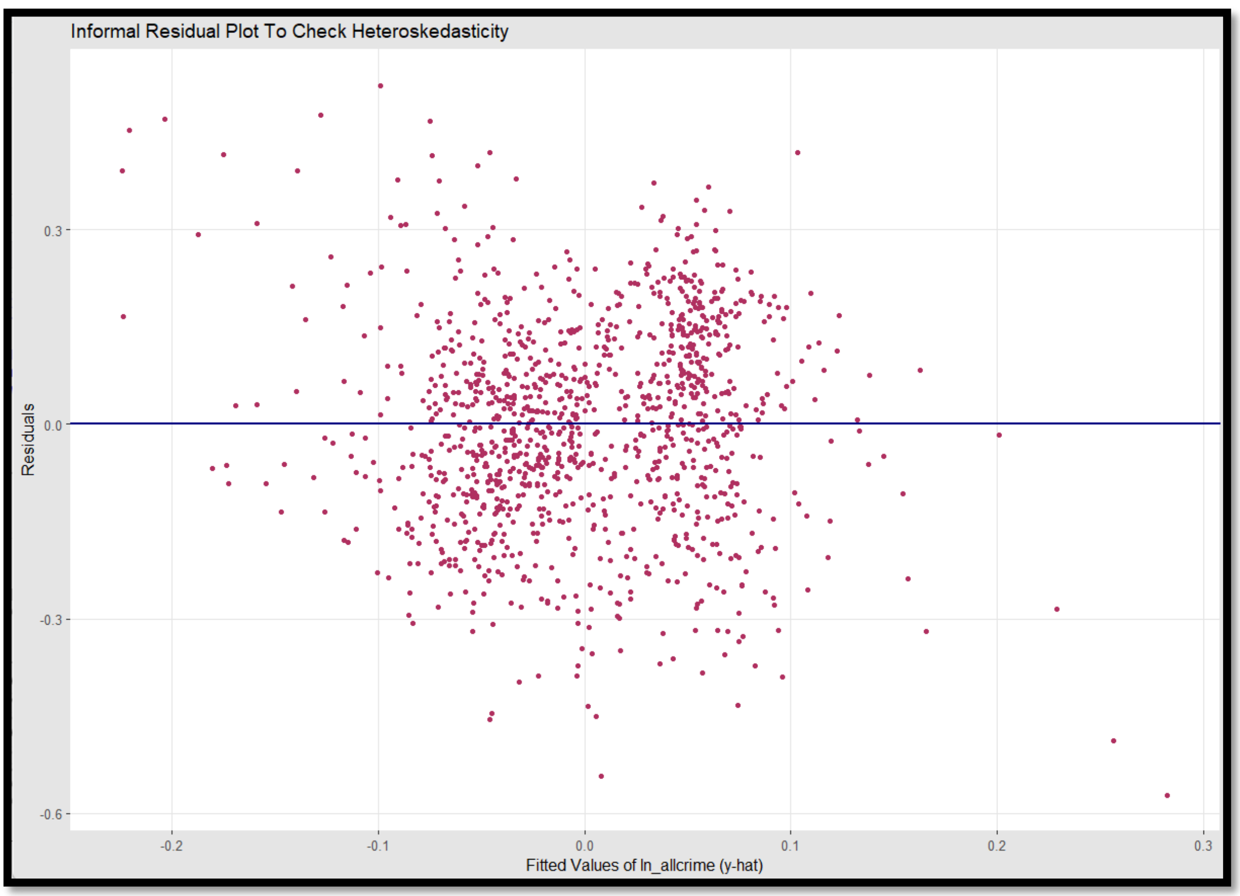
According to our expectation we get a strong positive relationship between percentage of black people in the state and the overall crime rate. The estimate is also consistent with our assumption . The estimate that we are getting for this variable is of 0.18 which says that with 1% increase in percentage of black people in the state will increase the overall crime rate by 18%. 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 2% 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 it's 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 -0.055 states that with 1% increase in the population of young male in the state the overall crime rate seems to decrease by 5.5%. 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.

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 this 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 serial 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.

log_all_crime ~ log_incarc_rate + log_pb1064 + pw1064 + pm1029 + log_pop + avginc + log_density + factor_shall + factor_year

```
Balanced Panel: n = 51, T = 23, N = 1173

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.4693447 -0.0729632  0.0051797  0.0785097  0.5714834

Coefficients:
                 Estimate Std. Error t-value  Pr(>|t|)
log_incarc_rate -0.1343271  0.0268409 -5.0046 6.522e-07 ***
log_pb1064      -0.4355061  0.0486190 -8.9575 < 2.2e-16 ***
pw1064          -0.0061273  0.0040212 -1.5237   0.12787
pm1029           0.0940555  0.0112362  8.3707 < 2.2e-16 ***
log_pop          1.3153974  1.2731373  1.0332   0.30174
avginc           0.0057866  0.0060079  0.9632   0.33568
log_density     -1.5224591  1.2880725 -1.1820   0.23748
factor_shall1   -0.0320782  0.0168313 -1.9059   0.05693 .
factor_year78    0.0719761  0.0267365  2.6920   0.00721 **
factor_year79    0.1995189  0.0271352  7.3528 3.801e-13 ***
factor_year80    0.2869910  0.0275182 10.4291 < 2.2e-16 ***
factor_year81    0.3138947  0.0283184 11.0845 < 2.2e-16 ***
factor_year82    0.3103378  0.0300156 10.3392 < 2.2e-16 ***
factor_year83    0.2880747  0.0323384  8.9081 < 2.2e-16 ***
factor_year84    0.3222106  0.0350740  9.1866 < 2.2e-16 ***
factor_year85    0.3880864  0.0378747 10.2466 < 2.2e-16 ***
factor_year86    0.4866720  0.0412924 11.7860 < 2.2e-16 ***
factor_year87    0.4992379  0.0446438 11.1827 < 2.2e-16 ***
factor_year88    0.5780354  0.0482130 11.9892 < 2.2e-16 ***
factor_year89    0.6551287  0.0516065 12.6947 < 2.2e-16 ***
factor_year90    0.8116477  0.0564930 14.3672 < 2.2e-16 ***
factor_year91    0.8963382  0.0592556 15.1266 < 2.2e-16 ***
factor_year92    0.9418520  0.0624393 15.0843 < 2.2e-16 ***
factor_year93    0.9822369  0.0647288 15.1746 < 2.2e-16 ***
factor_year94    0.9872180  0.0673688 14.6539 < 2.2e-16 ***
factor_year95    1.0026555  0.0701888 14.2851 < 2.2e-16 ***
factor_year96    0.9664963  0.0729428 13.2501 < 2.2e-16 ***
factor_year97    0.9542514  0.0754353 12.6499 < 2.2e-16 ***
factor_year98    0.9074010  0.0781692 11.6082 < 2.2e-16 ***
factor_year99    0.8626670  0.0803259 10.7396 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    33.904
Residual Sum of Squares: 19.474
R-Squared:      0.4256
Adj. R-Squared: 0.38352
F-statistic: 26.9708 on 30 and 1092 DF, p-value: < 2.22e-16
> coefplot(fixed_time_entity_model1)
>
```

Significant Coefficients:

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

Insignificant Coefficients

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

<u>Interpretation of Time and Entity Fixed Model</u>

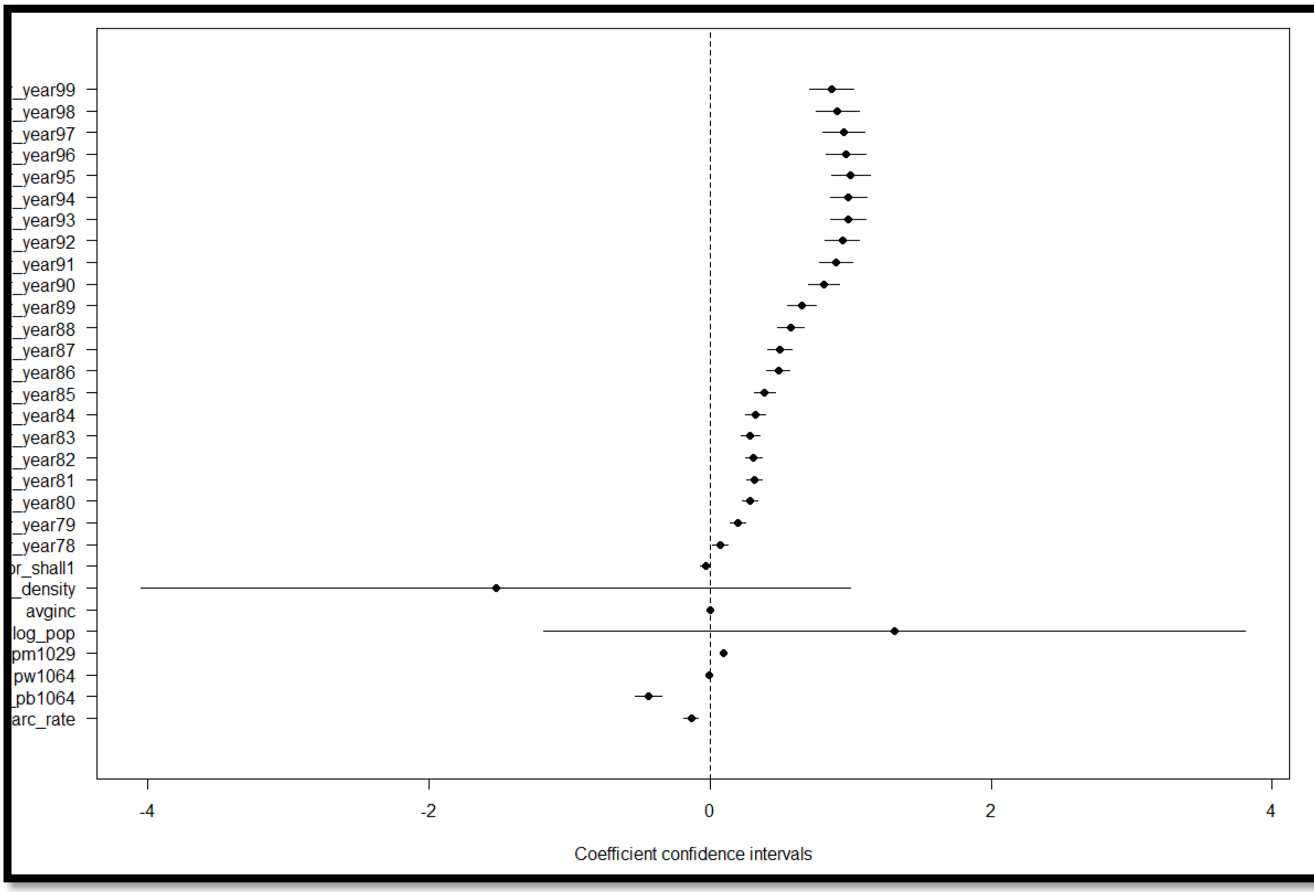
The interpretation of the shall law effect is still insignificant or has no effect on the overall crime rate this is consistent with our estimate that we got for the entity fixed model, but the time and entity fixed model shows even higher evidence for its insignificance. The estimate that we receive for it is -0.03 which states that the states having shall law in effect have 3% 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. The variable is found to be highly significant at every possible significance levels.

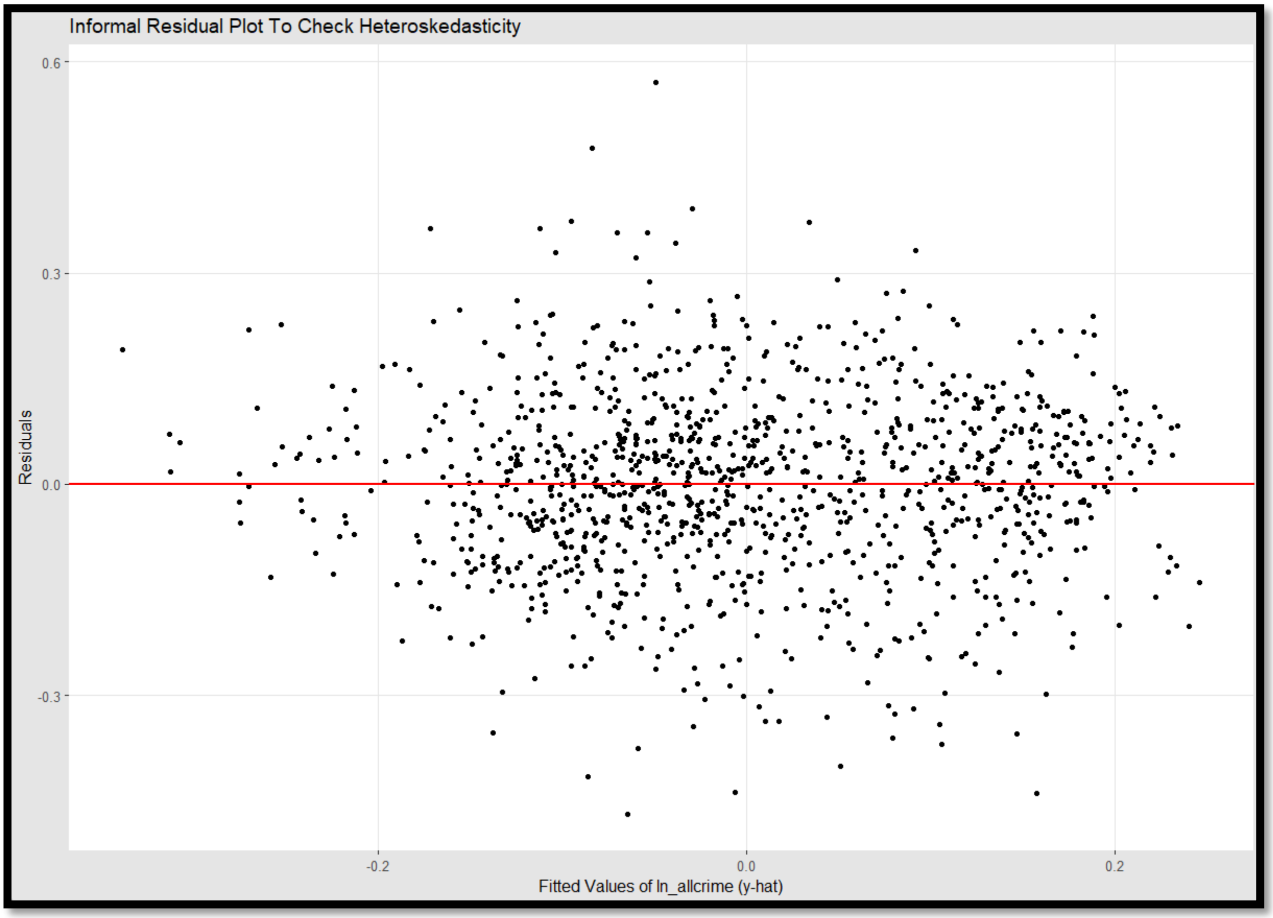
The variable % population of white people in state has found to be insignificant or do not have any effect on the overall crime rate with coefficients of -0.006 which is very insignificant magnitude wise as well.

Like expected the % population of young male is found to be very significant and consistent with our expectation of having a positive relationship. The estimate that we got for the time and entity fixed model is 0.09 stating with increase in young male population the overall crime rate increases by 9% on average everything else kept constant.

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 serial correlation has been removed by the Time and Entity fixed effects model(no pattern exhibited by the residuals, no inverted u pattern).


Informal Residual Plot To Check Heteroskedasticity
0.6
0.3
0.0
-0.3
Residuals
-0.2
0.0
0.2
Fitted Values of ln_allcrime (y-hat)

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 coefficients 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 will mean that at least one of the coefficients 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 : All the time coefficients are zero.

Alternative Hypothesis: At least one of the time coefficients is not zero.

The results of the test are as follows:

```
> linearHypothesis(fixed_time_entity_model1,null2)
Linear hypothesis test

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_rate + log_pb1064 + pw1064 + pm1029 +
    log_pop + avginc + log_density + factor_shall + factor_year

  Res.Df Df  Chisq Pr(>Chisq)
1   1114
2   1092 22 574.58  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We get a very high chi squared statistic of 574.58 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.

Conclusion

Heteroskedasticity, serially correlated errors, endogeneity etc.. were encountered when working with multiple datasets. We now understand the relationship b/w the dependent variable overall crime rate, and 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 remove errors in a particular model by comparing one model to another.

Finally, after a depth analysis of the models, we concluded that the Time and Entity fixed model performs the best in explaining the relationship of the 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.

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

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

Shall law being insignificant, it did not affect the overall crime rate. We might think that cultural attitudes of different states etc. could affect the overall crime rate

However, the incarceration rate was highly 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, a lot 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