Final Project-

Guns in 51 US States Analysis

Khushboo Rathore-KXR190005

Shrestha Jha- SXJ180082

Naishal Kanubhai Thakkar-NXT180016

Yanyi Tang-YXT190011

Contents

1. Objective

The main objective of our project is to analyze the data to gain insights about the understanding of how shall-issues laws affect crime rate across 51 states in U.S from 1977 to 1999.

2. Variable Definition

- The total number of observations in our dataset are 1173 which is spread across 51 states within 23 years.
- Violent is the dependent variable which is a quantitative variable and we have 12 independent variables which are- rob, mur, shall, incarc_rate, density, avginc, pop, pm1029, pw1064, pb1064, stateid, year
- There is no missing data and so we can conclude that this is a balanced panel data.

Variable	type	Definition
Vio	*numerical*	violent crime rate (incidents per 100,000 members of the population)
Rob	*numerical*	robbery rate (incidents per 100,000)
Mur	*numerical*	murder rate (incidents per 100,000)
Shall	*Binary*	= 1 if the state has a shall-carry law in effect in that year = 0 otherwise
incarc_rate	*numerical*	incarceration rate in the state in the previous year (sentenced prisoners per 100,000 residents; value for the previous year)
Density	*numerical*	population per square mile of land area, divided by 1000
Avginc	*numerical*	real per capita personal income in the state, in thousands of dollars

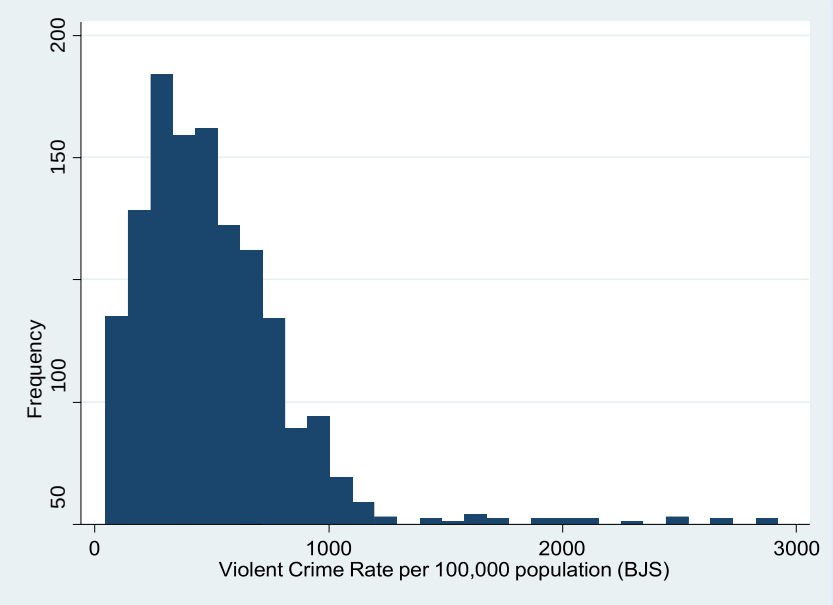
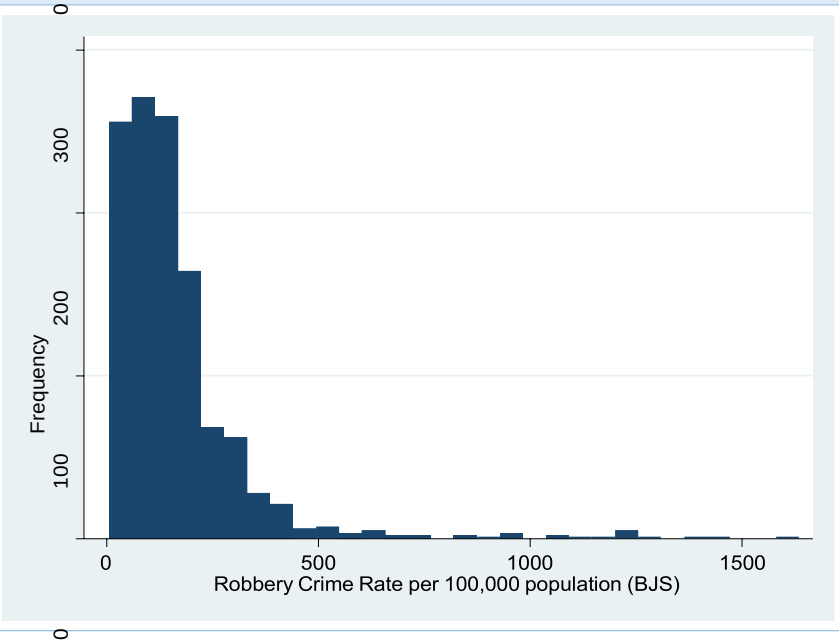
Pop	*numerical*	state population, in millions of people
pm1029	*numerical*	percent of state population that is male, ages 10 to 29
pw1064	*Numerical*	percent of state population that is white, ages 10 to 64
pb1064	*numerical*	percent of state population that is black, ages 10 to 64
Stated	*Character*	ID number of states (Alabama = 1, Alaska = 2, etc.)
Year	*Character*	Year (1977-1999)

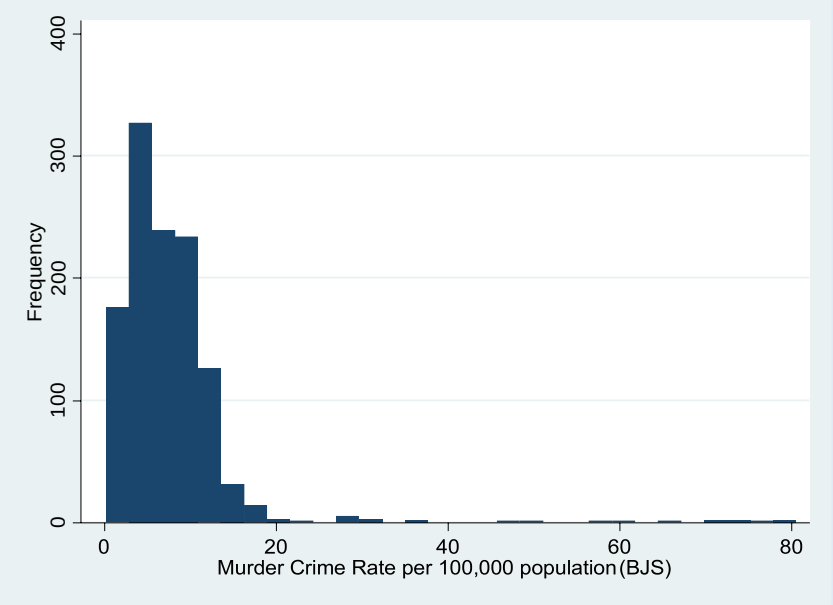
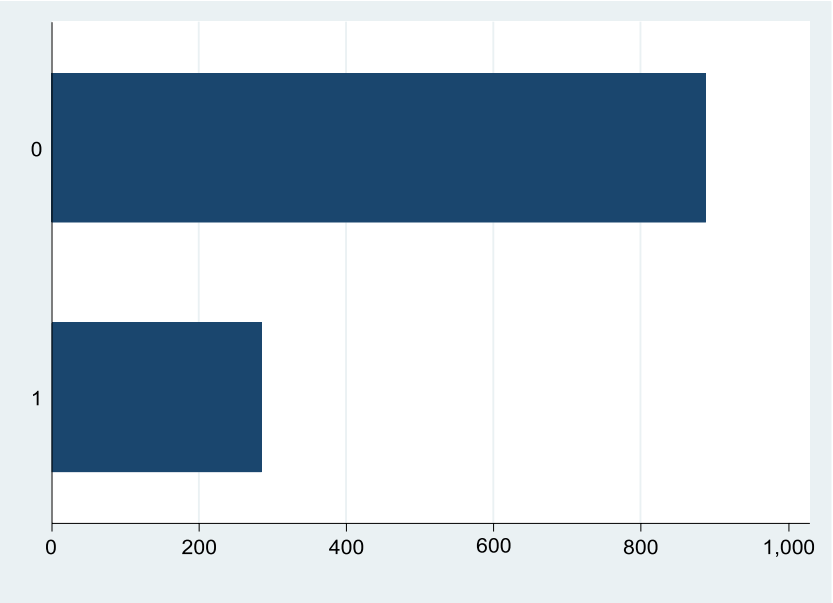
3. Summary Statistics and Distributions

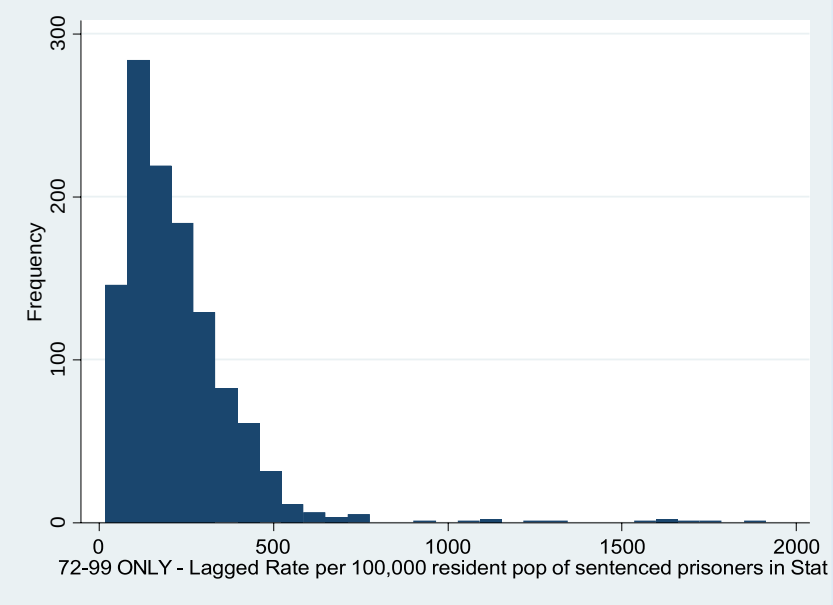
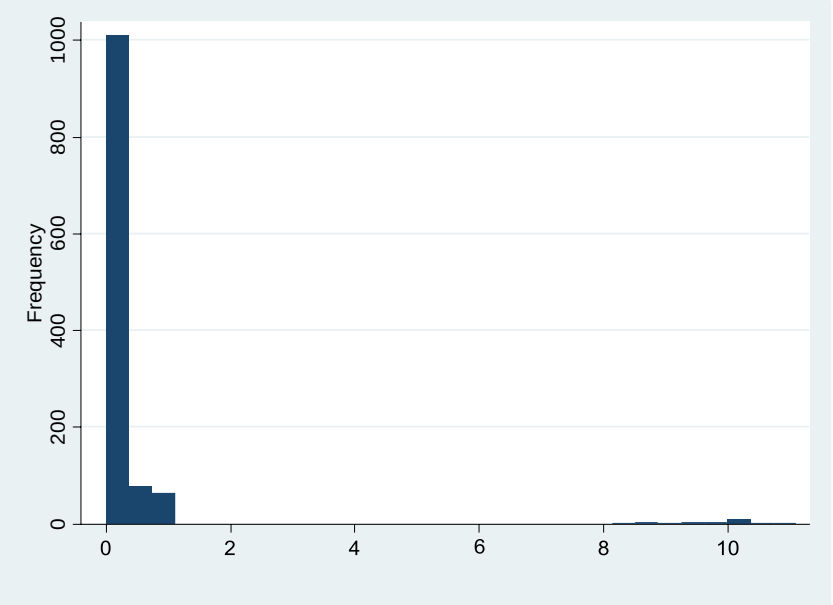
Summary Statistic-

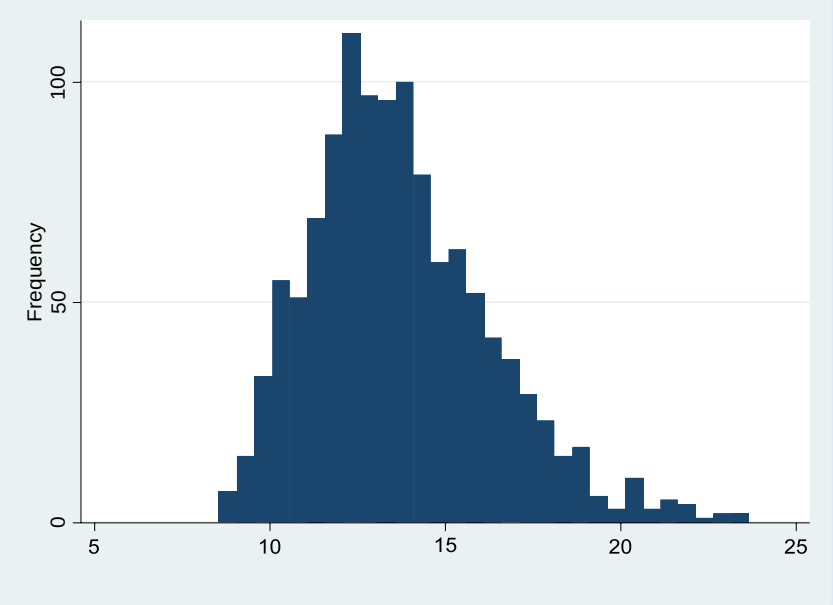
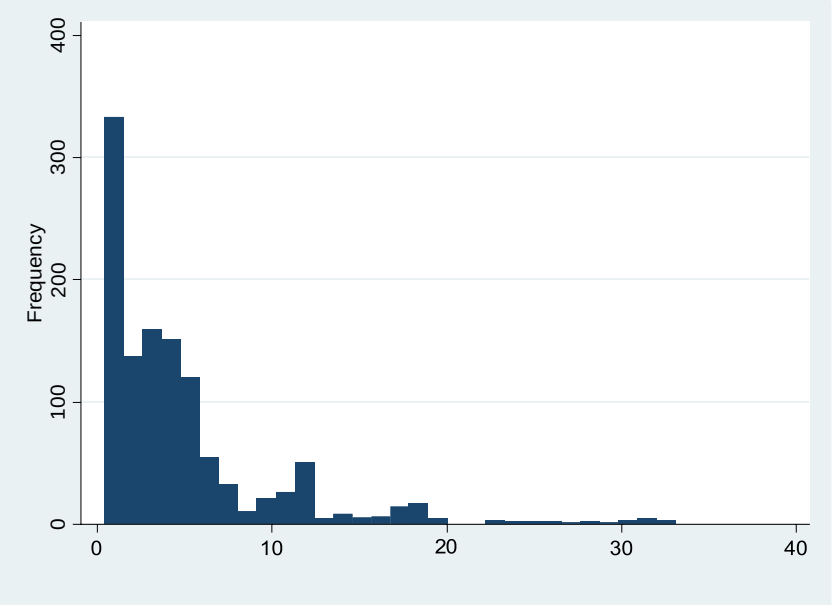
Variable	Mean	Median	Std	Min	Max
Vio	503.07	439.45	334.28	47.00	2,921.80
rob	161.82	107.45	170.51	6.40	1,635.10
mur	7.67	11.25	7.52	0.20	80.60
incarc_rate	226.58	166.00	178.89	19.00	1,913
density	0.35	0.08	1.36	0.00	11.10
avginc	13.72	9.83	2.56	8.55	23.65
pop	4.82	3.92	5.25	0.40	33.15
pm1029	16.08	17.59	1.73	12.21	22.35
pw1064	62.95	54.90	9.76	21.78	76.53
pb1064	5.34	8.50	4.89	0.25	26.98

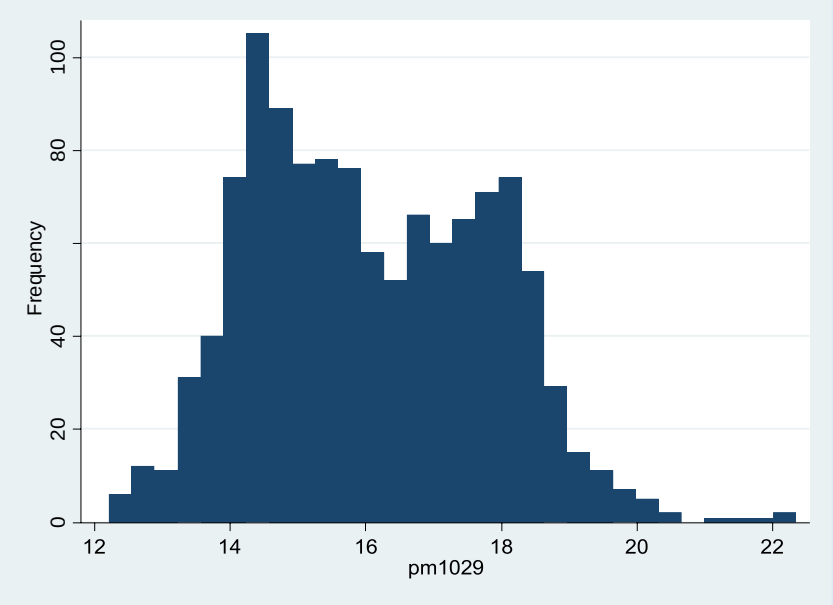
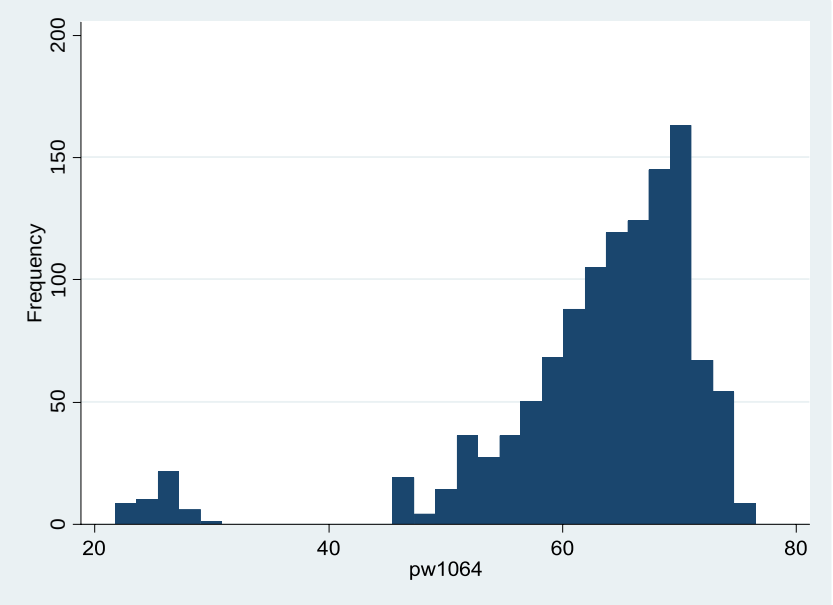
Distributions-

Variable	Distribution
vio	
rob	

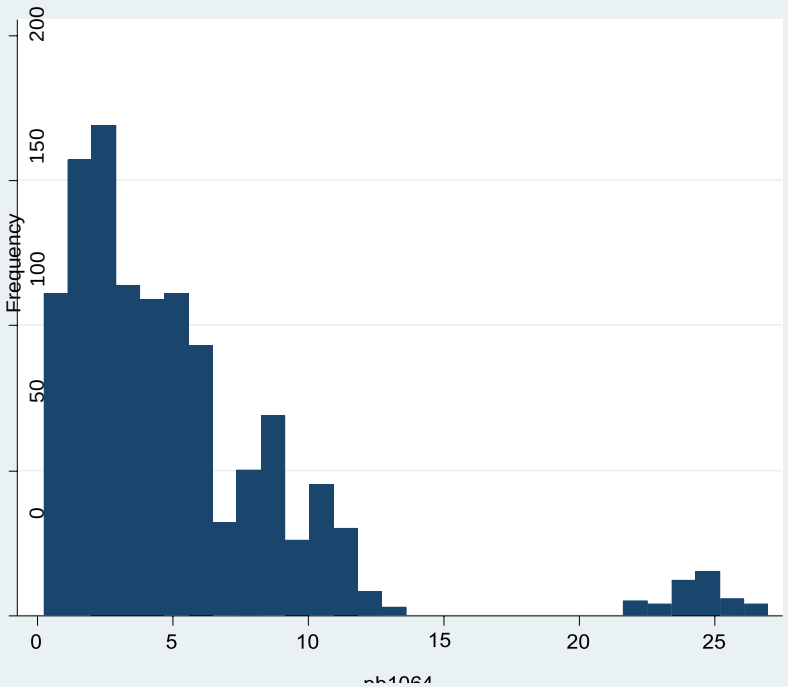
mur	 Frequency 0 100 200 300 400 0 20 40 60 80 Murder Crime Rate per 100,000 population (BJS)
shall	 0 1 0 200 400 600 800 1,000

incarc_rate	
density	

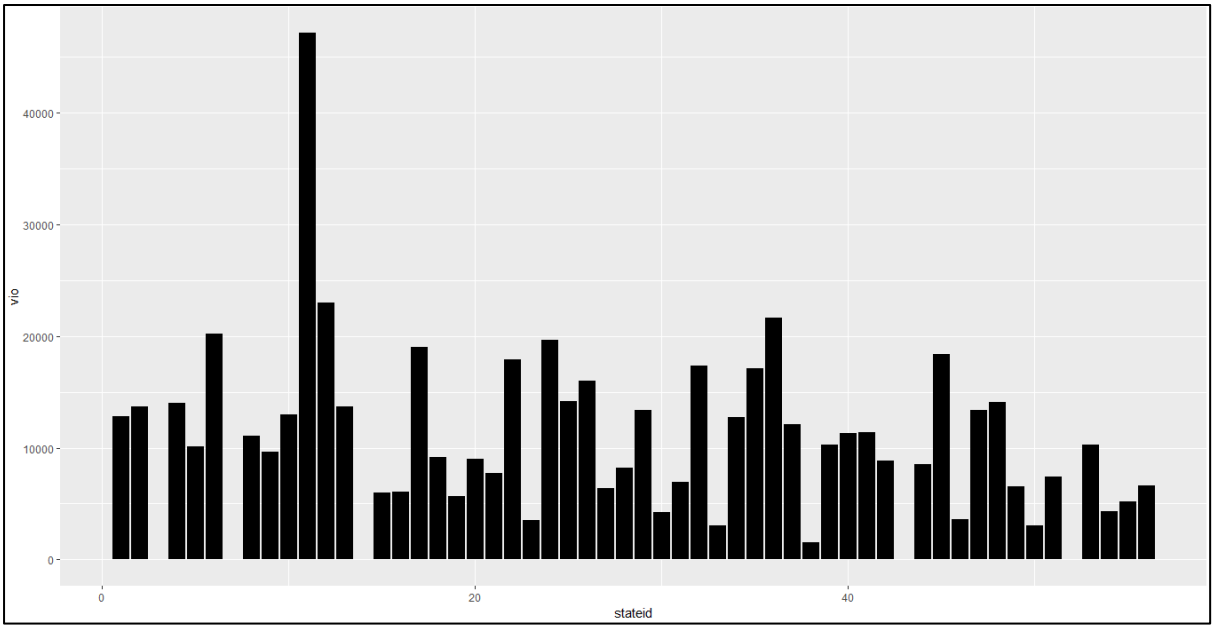
avginc	 Frequency 0 50 100 5 10 15 20 25
pop	 Frequency 0 100 200 300 400 0 10 20 30 40

pm1029	
pw1064	

pb1064



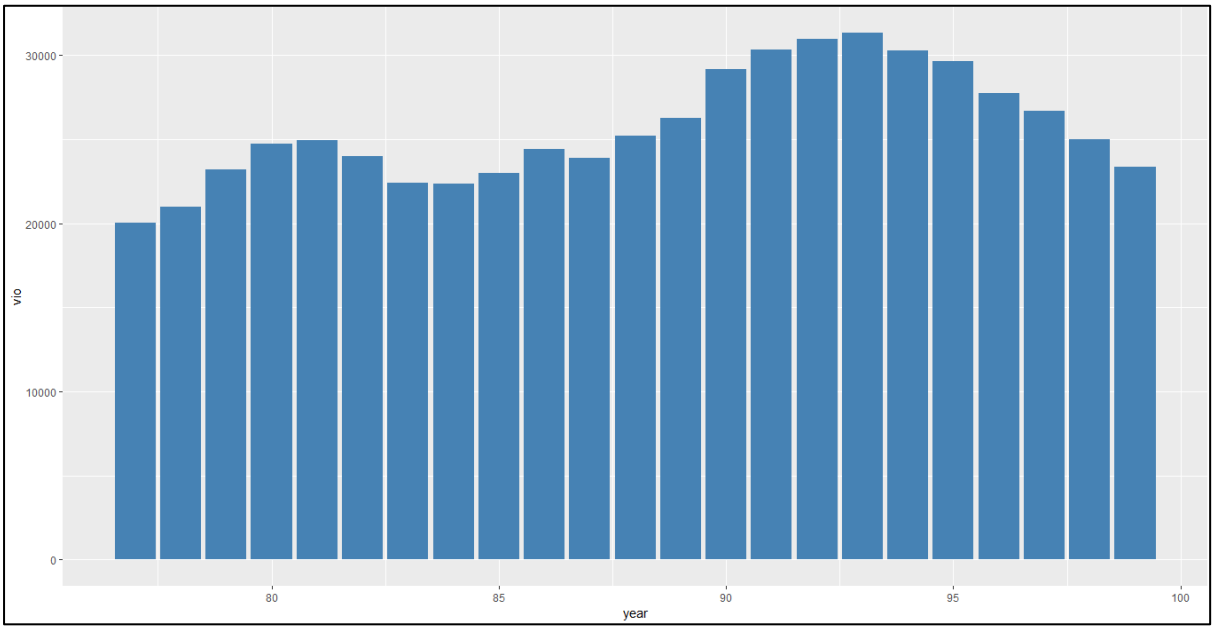
Plot of stateid vs violence



(Note: No data for stateid 3, 7, 14, 43, 52)

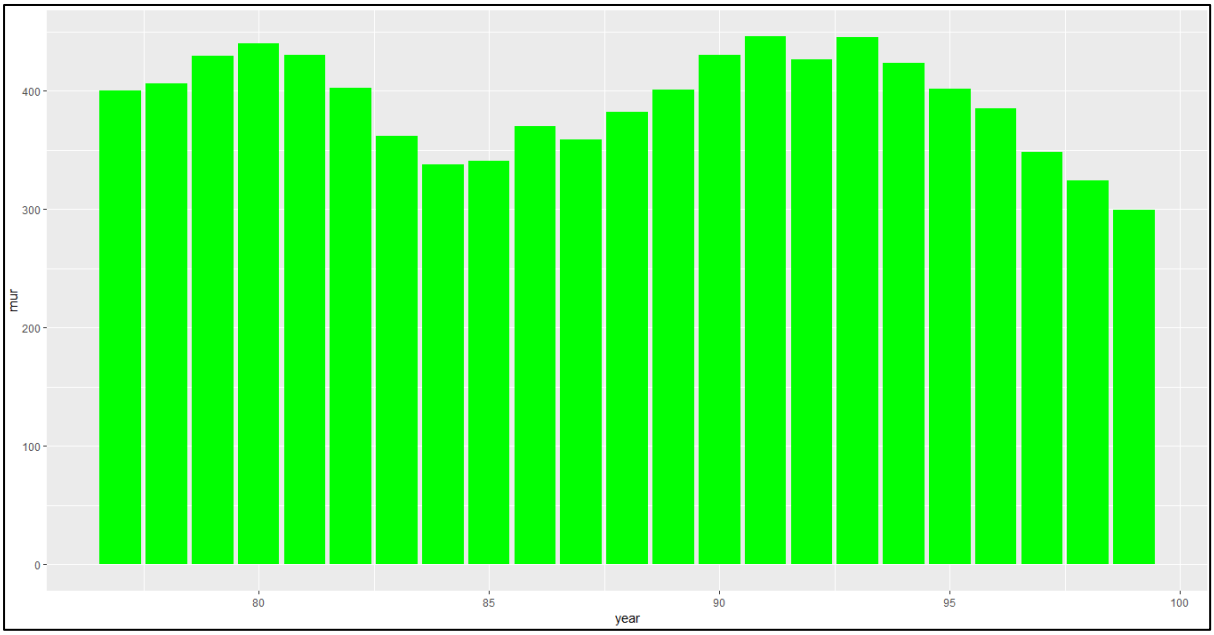
On plotting sum of violence against each stateid, we find that stateid 11 has the maximum violence compared to all the other states. Also, state with id 38 has the least violence rate.

Plot of Year vs Violence



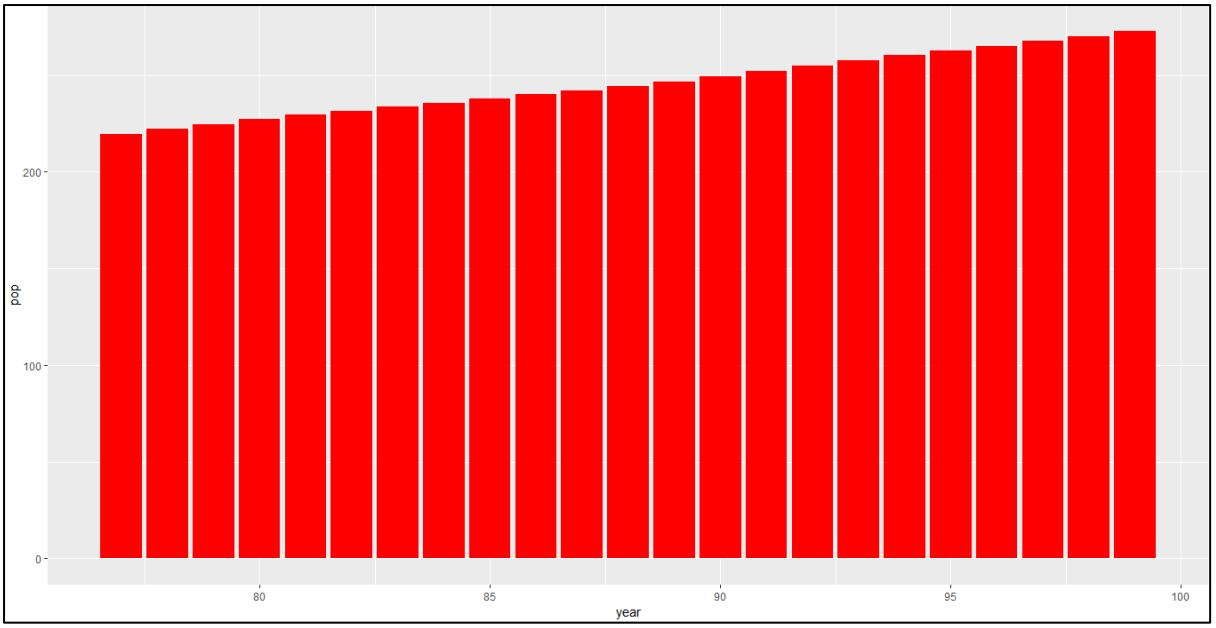
We can see from the figure above that year 1993 had the highest number of violence cases and year 1977 has the least violent cases.

Plot of Year Vs Murder



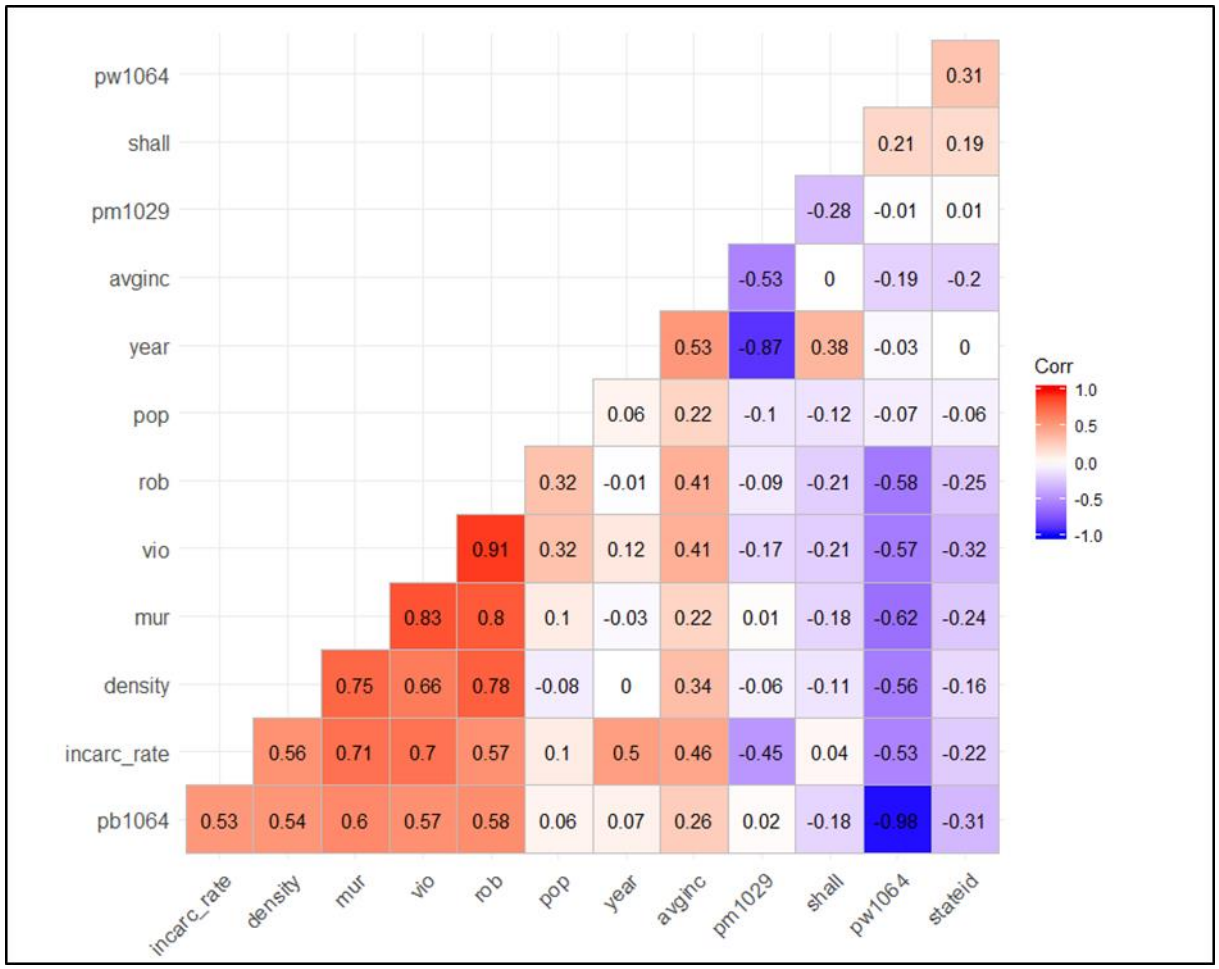
On plotting sum of murder against each year, we find that year 1991 had the highest number of murder cases and lowest was in 1999.

Plot for Year Vs Population



On plotting sum of population against year we see that population has increased with every passing year.

4. Correlation Matrix



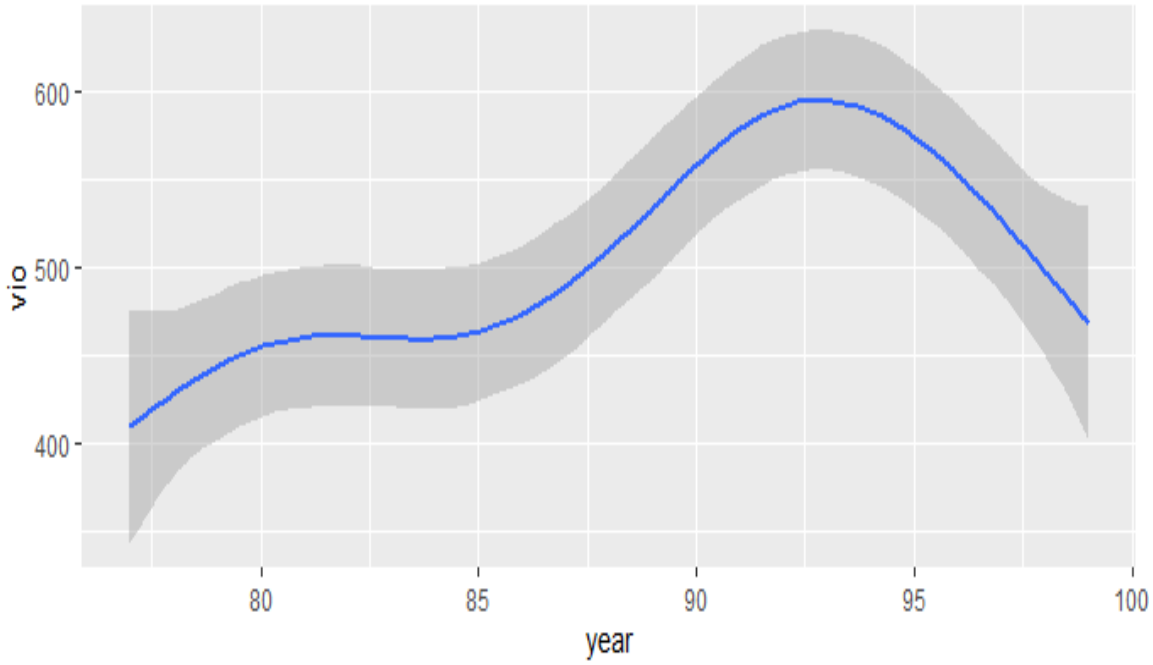
INFERENCE-

We have plotted a correlation matrix to find the correlation between response variable (Violent Crime Rate) and quantitative explanatory variables if they have a positive or negative or no association. In some case, the increase in some variables can lead to increase in the crime rate leading to a positive correlation, similarly if the crime rate decreases with the rise in some variables then they have a negative correlation. We can come to know which the variables that influence our dependent variable are and how strongly which can help us draw important conclusion. We have found

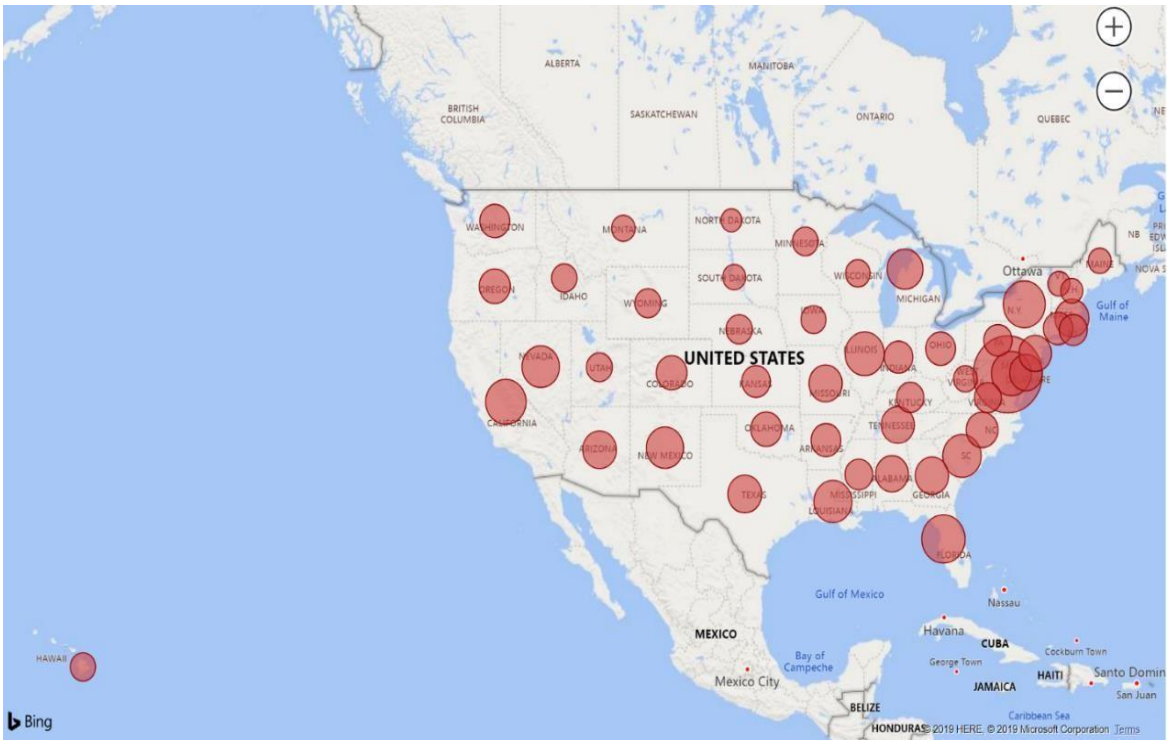
the following inferences about our variables after studying the correlation matrix.

- Areas that have higher density of population tend to have a higher crime rate.
- Higher the population of black people, higher the crime rate.
- Higher percentage of white populations are more likely to live in areas which tend to have lower population density.
- Higher the population of white people, lower the crime rate.
- When the percentage of number of males increase, crime rate decreases.

-Violent crime rate across different years



-Violent crime rate across different states


ALBERTA
MANITOBA
BRITISH COLUMBIA
SASKATCHEWAN
ONTARIO
QUEBEC
Ottawa
Gulf of Maine
UNITED STATES
MEXICO
Gulf of Mexico
Nassau
Havana
CUBA
Cockburn Town
George Town
HAITI
JAMAICA
Santo Domin
San Juan
Caribbean Sea
Mexico City
Bay of Campeche
BELIZE
HONDURAS
HAWAII
Bing
© 2019 HERE, © 2019 Microsoft Corporation Terms


Beaufort Sea
ALASKA
YUKON TERRITORY
Norton Sound
Bristol Bay
Gulf of Alaska

5. Explanatory Variables Expectation-

a. **Rob**: It is natural for the crime rate to increase as the number of robberies increase as robbery is a crime which would be exponentially increasing and hence, we expect "Rob" is positive.

b. **Mur:** As the number of murders increases, the crime rate would also increase as murder is a crime by itself and hence, we can expect "Mur" to be positive.

c. **Shall:** This is the variable on which we have an argument claim that if the state allows citizens to carry concealed guns, the violent crime rate would decrease. As there is a negative association between "crime rates" and "shall" we expect "shall" to be negative.

d. **incarc_rate:** As the new law will be enforced, people will be less likely to involve in criminal activity and so we can expect this variable to be negative.

e. **Density**: If the population of an area is higher, they are more likely to have a higher criminal rate as there are more chances of criminals as compared to areas with less density of population and so we can expect "Density" to be positive.

f. **Avginc:** People who have low income area more likely to involve themselves in crime to get some more money, hence "Avginc" can be expected to be a negative as when the income of people in an area is low the crime rate is expected to be higher.

g. **Pop**: As the population increases, there can be a chance for the crime rate to increase or decrease deepening on the type of people, hence we cannot expect anything about "Pop" variable with much confidence.

h. **Pm1029:** The "Males" are more likely to be involving in illegal/criminal activities and hence we can expect this variable to be positive.

i. **Pw1064:** As the percentage of state population (white) goes up, we expect the violence rate to go up. However, the expected signs of variables are debatable.

j. **Pb1064:** The Blacks are more likely to be involved in illegal/criminal activities and hence we expect this variable to be positive.

k. **Year:** After observing the trend of crime rate over the span of years we can infer that though a greater number of states have enforced the shall-law with time but the overall crime rate has slightly increased over the

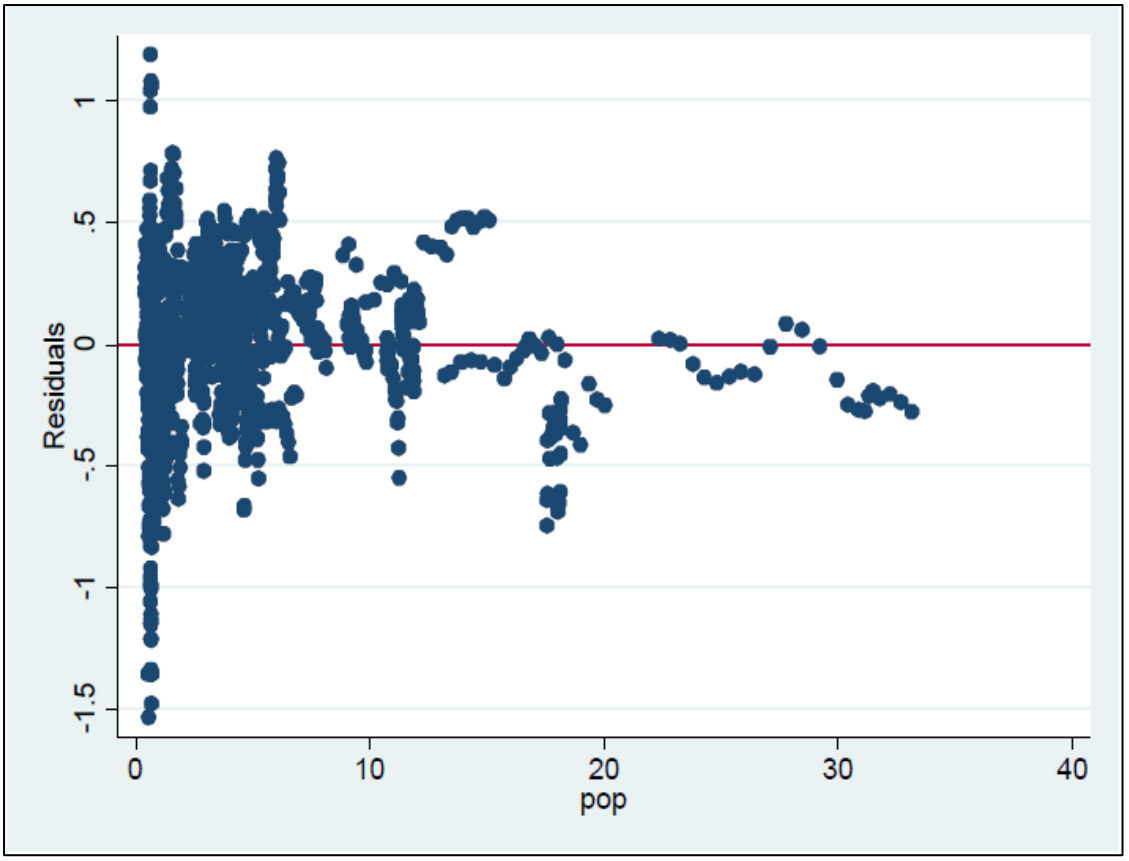
years i.e. from 1977 to 1999

6. Approach

1. We conduct explanatory data analysis which is scatterplots, heatmaps, correlation matrix and other statistical and distributions analysis.
2. Check for the presence of Heteroskedasticity by plotting the **"Residuals vs Fitted"** plot.
3. Applied the Pooled ordinary least squares on our dataset without cluster robust error.
4. Run Pooled ordinary least squares with cluster robust error on our dataset.
5. Run the Entity Fixed Effect model with cluster robust error on the dataset.
6. Run the Entity and Time Fixed Effect model with cluster robust error on the dataset.
7. Run Random effect model on our dataset.
8. Use the **Hausman** test to check for the **"Endogeneity Problem"** in our dataset**.**
9. Compare the outcomes after running each of the models to come to the conclusion of which is the best model for our dataset.

7. Models

Checking for Heteroskedasticity using Residual Plot:



INFERENCE-

1. After plotting all the independent variables against the residual values, we found that the variable “population” has heteroskedasticity.
2. Therefore, the standard errors of the pooled OLS without robust standard errors are incorrect.
3. Since pooled OLS model ignores the panel nature of the data, we cannot trust the results of this model.

Model 1- Pooled OLS model Without Cluster Robust Standard Errors

```
. reg lvio  mur rob incarc_rate pop avginc density pb1064 pw1064 pm1029 i.shall

      Source |       SS           df       MS      Number of obs   =     1,173
-------------+----------------------------------   F(10, 1162)     =    300.35
       Model |  352.32434         10  35.232434   Prob > F        =    0.0000
    Residual |  136.307219     1,162  .117303975   R-squared       =    0.7210
-------------+----------------------------------   Adj R-squared   =    0.7186
       Total |  488.631558     1,172  .416921125   Root MSE        =     .3425

------------------------------------------------------------------------------
        lvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mur |   .0034401   .0032353     1.06   0.288    -.0029076    .0097878
         rob |   .0031608   .0001437    22.00   0.000     .0028789    .0034427
 incarc_rate |   .0012571   .0001126    11.16   0.000     .0010361     .001478
         pop |   .0074291   .0024851     2.99   0.003     .0025534    .0123049
      avginc |  -.0084354   .0063823    -1.32   0.187    -.0209574    .0040867
     density |  -.2481097    .015181   -16.34   0.000    -.2778949   -.2183245
      pb1064 |   .0578545   .0133718     4.33   0.000     .0316189     .08409
      pw1064 |   .0265646   .0067147     3.96   0.000     .0133903    .0397389
      pm1029 |  -.0056762   .0091565    -0.62   0.535    -.0236412    .0122889
     1.shall |  -.2779262   .0263605   -10.54   0.000    -.3296458   -.2262066
       _cons |   3.549915   .4360367     8.14   0.000     2.694408    4.405422
------------------------------------------------------------------------------
```

INFERENCE:

1. The coefficient of Shall equals -0.277, which suggests that shall-issue laws reduce the violent crime rate by 28%, for every percent increase in Shall.
2. This is a huge decline in the crime rate and the result is highly significant result as well and we can infer this by the P value.
3. As expected, all the other variables are positively contributing to crime except pm1029, density, avginc etc. All the variables are statistically significant at 5% significance level except pm1029, avginc and mur
4. We will now run a regression with the Robust Cluster Standard error to remove the problem of heteroskedasticity to get accurate standard errors, inefficient, unbiased and consistent but our estimates will still be inefficient

Model 2- Pooled OLS model with Robust Cluster Standard Errors

```
. reg lvio  mur rob incarc_rate pop avginc density pb1064 pw1064 pm1029 i.shall, vce(cluster stateid)

Linear regression                               Number of obs     =      1,173
                                                F(10, 50)         =      95.32
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7210
                                                Root MSE          =      .3425

                                   (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
        lvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mur |   .0034401   .0120341     0.29   0.776    -.0207312    .0276114
         rob |   .0031608   .0005113     6.18   0.000     .0021338    .0041878
 incarc_rate |   .0012571   .0003247     3.87   0.000     .0006048    .0019093
         pop |   .0074291   .0084814     0.88   0.385    -.0096063    .0244646
      avginc |  -.0084354   .0213671    -0.39   0.695    -.0513525    .0344818
     density |  -.2481097   .0746818    -3.32   0.002    -.3981125   -.0981069
      pb1064 |   .0578545   .0541324     1.07   0.290    -.0508737    .1665826
      pw1064 |   .0265646   .0251827     1.05   0.297    -.0240164    .0771456
      pm1029 |  -.0056762   .0291031    -0.20   0.846    -.0641314    .0527791
     1.shall |  -.2779262   .0941361    -2.95   0.005    -.4670041   -.0888482
       _cons |   3.549915   1.539068     2.31   0.025     .4586061    6.641224
------------------------------------------------------------------------------
```

INFERENCE:

1. As we can observe there is a significant increase in the standard errors of the Shall variable which have increased to 0.094 from 0. 0264.This is a strong indication for the presence of heteroskedasticity in the data.
2. Other variables also show a significant increase in standard errors.
3. There can be a problem of endogeneity in our model as there could be some unobserved heterogeneity i.e. omitted variables bias or simultaneous causality bias which got added in the error term and will lead to biased and inconsistent estimators. One such omitted variable could be attitude of people towards gun.
4. There is large difference in standard errors of the pooled models with and without using robust standard error. This implies that there may be heteroskedasticity and/or auto-correlation. This can be tested using White-test.
5. We will proceed with fixed effects model as that will confirm the existence of endogeneity, however, we show endogeneity using Hausman test as

well. We reject the null hypothesis of no endogeneity.

Model 3- Fixed Effects Model – Entity Fixed (adjusted for cluster robust errors)

```
Fixed-effects (within) regression               Number of obs      =       1,173
Group variable: stateid                         Number of groups   =          51

R-sq:                                           Obs per group:
     within  = 0.4392                                         min =          23
     between = 0.6465                                         avg =        23.0
     overall = 0.6240                                         max =          23

                                                F(10,50)           =       65.55
corr(u_i, Xb)  = 0.3810                         Prob > F           =      0.0000

                                (Std. Err. adjusted for 51 clusters in stateid)
------------------------------------------------------------------------------
             |               Robust
       lnvio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mur |   .0050171   .0016618     3.02   0.004     .0016793     .008355
         rob |   .0016168   .0005563     2.91   0.005     .0004993    .0027342
 incarc_rate |    .000118   .0002174     0.54   0.590    -.0003186    .0005546
         pop |   .0089291   .0093866     0.95   0.346    -.0099244    .0277826
      avginc |  -.0000537   .0117846    -0.00   0.996    -.0237238    .0236163
     density |  -.0372724   .0988035    -0.38   0.708     -.235725    .1611803
      pb1064 |   .0752598    .024667     3.05   0.004     .0257147    .1248049
      pw1064 |   .0321464   .0127295     2.53   0.015     .0065784    .0577143
      pm1029 |  -.0403235   .0195701    -2.06   0.045    -.0796311   -.0010159
     1.shall |  -.0556272   .0353181    -1.58   0.122    -.1265657    .0153112
       _cons |   3.908221   .6842916     5.71   0.000     2.533781    5.282661
-------------+----------------------------------------------------------------
     sigma_u |  .40730413
     sigma_e |  .13620755
         rho |  .89941675   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

INFERENCE-

- The results change when we run the fixed effects model with fixed states.
- The absolute effect of Shall decreases to 5.6% from 27.79% which is a significant drop. The effect of shall issue laws on the violent crime rate is no longer statistically significantly different from zero due to insignificant p values and this makes our estimate not very reliable.

- This can be due to unobserved heterogeneity i.e. omitted variable bias which could have caused to overstate the effect of state in the previous model
- The regression model with fixed effects is more credible because this controls for unobserved heterogeneity that vary between states but that are constant over time.

Model 4- Fixed Effects Model – Entity and Time Fixed

```
Fixed-effects (within) regression               Number of obs      =       1,173
Group variable: stateid                         Number of groups   =          51

R-sq:                                           Obs per group:
     within  = 0.5318                                         min =          23
     between = 0.6406                                         avg =        23.0
     overall = 0.5673                                         max =          23

                                                F(32,50)           =       68.28
corr(u_i, Xb)  = 0.5248                         Prob > F           =      0.0000
```

lnvio	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
mur	-.0001777	.0020285	-0.09	0.931	-.004252	.0038965
rob	.0013728	.0004597	2.99	0.004	.0004495	.0022961
incarc_rate	.0002365	.000194	1.22	0.228	-.0001531	.0006261
pop	-.0048344	.0106941	-0.45	0.653	-.0263141	.0166453
avginc	.0042625	.0159587	0.27	0.790	-.0277916	.0363165
density	-.0454542	.1043341	-0.44	0.665	-.2550153	.1641069
pb1064	.0220232	.0444321	0.50	0.622	-.0672213	.1112677
pw1064	.0116374	.0214253	0.54	0.589	-.0313966	.0546714
pm1029	.0572108	.0463707	1.23	0.223	-.0359274	.150349
1.shall	-.0478819	.0357701	-1.34	0.187	-.1197282	.0239643
year						
78	.0480007	.0147777	3.25	0.002	.0183188	.0776825
79	.1283473	.0236324	5.43	0.000	.0808802	.1758144
80	.1496631	.0333752	4.48	0.000	.082627	.2166992
81	.1354229	.0378097	3.58	0.001	.05948	.2113658
82	.130421	.0424324	3.07	0.003	.045193	.215649
83	.1136524	.0510845	2.22	0.031	.0110461	.2162587
84	.1565921	.0660406	2.37	0.022	.0239456	.2892386
85	.2016936	.0794325	2.54	0.014	.0421487	.3612385
86	.2641283	.094051	2.81	0.007	.0752212	.4530353
87	.2699695	.1071642	2.52	0.015	.054724	.4852151
88	.3155064	.1200197	2.63	0.011	.0744396	.5565731
89	.3518629	.1314078	2.68	0.010	.0879226	.6158032
90	.4233356	.1686763	2.51	0.015	.0845393	.7621318
91	.4550355	.1757298	2.59	0.013	.1020718	.8079993
92	.4865985	.1858883	2.62	0.012	.1132309	.859966
93	.506119	.1925554	2.63	0.011	.1193601	.8928779
94	.4962866	.1993631	2.49	0.016	.095854	.8967192
95	.4880726	.2067133	2.36	0.022	.0728768	.9032684
96	.4456522	.2160079	2.06	0.044	.0117876	.8795167
97	.4412763	.2229722	1.98	0.053	-.0065766	.8891292
98	.4017626	.2344254	1.71	0.093	-.0690946	.8726198
99	.3511401	.2445952	1.44	0.157	-.1401439	.8424241
_cons	3.681353	1.017483	3.62	0.001	1.637679	5.725027
sigma_u	.4827741					
sigma_e	.12570543					
rho	.93650627	(fraction of variance due to u_i)				

INFERENCE-

- The results change slightly when we run the fixed effects model with both states and time fixed as compared to our previous model.
- There is a slight change in the effect of shall-carry law on violence rate further reduced from 5.5% to 4.8% in the entity and time fixed effect model.
- The coefficient is still not significantly different from zero as we can infer this from the P values.

- After running the test to check if the time variables are statistically significant, we can infer that the time variables are significant. Hence we can infer that the time and state fixed model is a better fit than the only state fixed model.

F-Test for significance of time variables

```
. do "C:\Users\AAJ170~1\AppData\Local\Temp\20\STD4f78_000000.tmp"

. . testparm i.year

 ( 1)  78.year = 0
 ( 2)  79.year = 0
 ( 3)  80.year = 0
 ( 4)  81.year = 0
 ( 5)  82.year = 0
 ( 6)  83.year = 0
 ( 7)  84.year = 0
 ( 8)  85.year = 0
 ( 9)  86.year = 0
 (10)  87.year = 0
 (11)  88.year = 0
 (12)  89.year = 0
 (13)  90.year = 0
 (14)  91.year = 0
 (15)  92.year = 0
 (16)  93.year = 0
 (17)  94.year = 0
 (18)  95.year = 0
 (19)  96.year = 0
 (20)  97.year = 0
 (21)  98.year = 0
 (22)  99.year = 0

       F( 22,    50) =    8.30
            Prob > F =    0.0000
```

INFERENCE-

As we can see that the F stat-value is very high, hence we can infer that time variables are statistically significant and should be kept in our model for better results.

Model 5- Random Effect Model

The dataset used includes data from all the states therefore it is not random data. However, we can prove that random effect model is not appropriate for this data using Hausman Test.

```
. * xtreg lwage educ exper exper2 tenure tenure2 black south union, re
. xtreg lvio  mur rob incarc_rate pop avginc density pb1064 pw1064 pm1029 i.shall

Random-effects GLS regression                   Number of obs      =     1,173
Group variable: stateid                         Number of groups   =        51

R-sq:                                           Obs per group:
     within  = 0.4378                                         min =        23
     between = 0.6572                                         avg =      23.0
     overall = 0.6373                                         max =        23

                                                Wald chi2(10)      =    973.74
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

lvio	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mur	.0057858	.0016522	3.50	0.000	.0025475	.0090241
rob	.0016633	.0000968	17.19	0.000	.0014737	.001853
incarc_rate	.0001352	.0000656	2.06	0.039	6.76e-06	.0002637
pop	.0168583	.0055339	3.05	0.002	.006012	.0277046
avginc	-.001379	.0050625	-0.27	0.785	-.0113013	.0085434
density	-.0541179	.0326346	-1.66	0.097	-.1180806	.0098449
pb1064	.0862816	.011607	7.43	0.000	.0635322	.1090309
pw1064	.0326171	.0043166	7.56	0.000	.0241566	.0410776
pm1029	-.0378299	.0054399	-6.95	0.000	-.0484919	-.0271679
1.shall	-.0644369	.0160769	-4.01	0.000	-.0959471	-.0329268
_cons	3.750413	.3245209	11.56	0.000	3.114364	4.386463
sigma_u	.29737993					
sigma_e	.13620755					
rho	.82659164	(fraction of variance due to u_i)				

Hausman Test for endogeneity

	Coefficients (b) fixed	(B) random	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
mur	.0050171	.0057858	-.0007687	.0005205
rob	.0016168	.0016633	-.0000466	9.45e-06
incarc_rate	.000118	.0001352	-.0000172	.0000506
pop	.0089291	.0168583	-.0079292	.0050227
avginc	-.0000537	-.001379	.0013252	.0013101
density	-.0372724	-.0541179	.0168455	.0650716
pb1064	.0752598	.0862816	-.0110218	.01065
pw1064	.0321464	.0326171	-.0004707	.0004893
pm1029	-.0403235	-.0378299	-.0024936	.0015072
1.shall	-.0556272	-.0644369	.0088097	.

```
                b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

              chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =   -31.14     chi2<0 ==> model fitted on these
                                     data fails to meet the asymptotic
                                     assumptions of the Hausman test;
                                     see suest for a generalized test

.
end of do-file
```

OBSERVATION-

We can see from the above table that there are significant differences in co-efficient between the fixed effect and random effect models. Therefore, we can conclude that random effect model is not appropriate for this data.

CONCLUSION-

Based on the above analysis, we can conclude that entity and time fixed effect model is the best.

LIMITATION- However, there are some limitations which are:

- Simultaneously causality bias: the variable incarc_rate (incarceration rate) and violence rate effect each other causing bias in the estimate of incarc_rate.
- Inefficient: estimates are computed only using variations within entities.
- Time variant factors: Unobserved heterogeneity which is time variant is not controlled in the fixed effect model. This causes biased results.

8. Conclusion

- There is a large estimated effect of concealed weapons laws in pooled OLS models.
- This effect is however due to omitted variable bias and unobserved characteristics because the effect disappears when state and time effects are added.
- So, the model with both time and state fixed effects is the best model as it is robust to endogeneity.
- We can conclude that there is no significant effect of concealed weapon laws on the violent crime rate. The shall law, when enforced reduces the crime rate nearly by 5%.

9. Limitations of analysis

- Even the FE estimate could have following bias:

    - It captures only within variation
    - It captures only variables which are constant

- FE estimator is still biased if the unobserved heterogeneity changing over time, and correlated with the regressors.
- This means that if some variable such as attitude of people towards gun changes over time then it will not be captured here.
- We will need more data to check for that effect but it is really difficult to get such data

THANK YOU