Applied Econometrics and Time Series Analysis

Final Project on Guns data

Mentored By

Dr. Moran Blueshtein

Project Members:

Name	Section	NetID
Amit Sawant	001	axs173331
Guru Shanmukha Sahitya Bulusu	003	gxb170430
Malvika Bodh	003	mxb172830
Sahith Reddy Karnakanti	003	kxr171530
Keerthi Danda	003	kxd172230

Contents

1. Introduction

The report analyzes the data to shed light on the following question, “Do shall-issues law reduce crime or not?” The analysis provided gives a description on the statistics behind crime activity in the United States over the span of 23 years. By understanding the trends of violence, robberies, incarceration rate, and murder as well as data on the demographic the states, one can understand the impact that concealed handgun laws have on crime.

Background Information on Shall-Law: Many states in the United States have passed right-to-carry laws (also known as shall-issue laws). A Shall-issue law is one that requires that governments issue concealed carry handgun permits to any applicant who meets the necessary criteria. These criteria are that the applicant must be an adult, have no significant criminal record and no history of mental illness and successfully complete a course in firearms safety training (if required by law). If these criteria are met, then the granting authority has no discretion in the awarding of the license.

Guns is a balanced panel of data on 50 US states, plus the District of Columbia (for a total of 51 “states”), by year for 1977 – 1999. Each observation is a given state in a given year. There are a total of 51 states × 23 years = 1173 observations

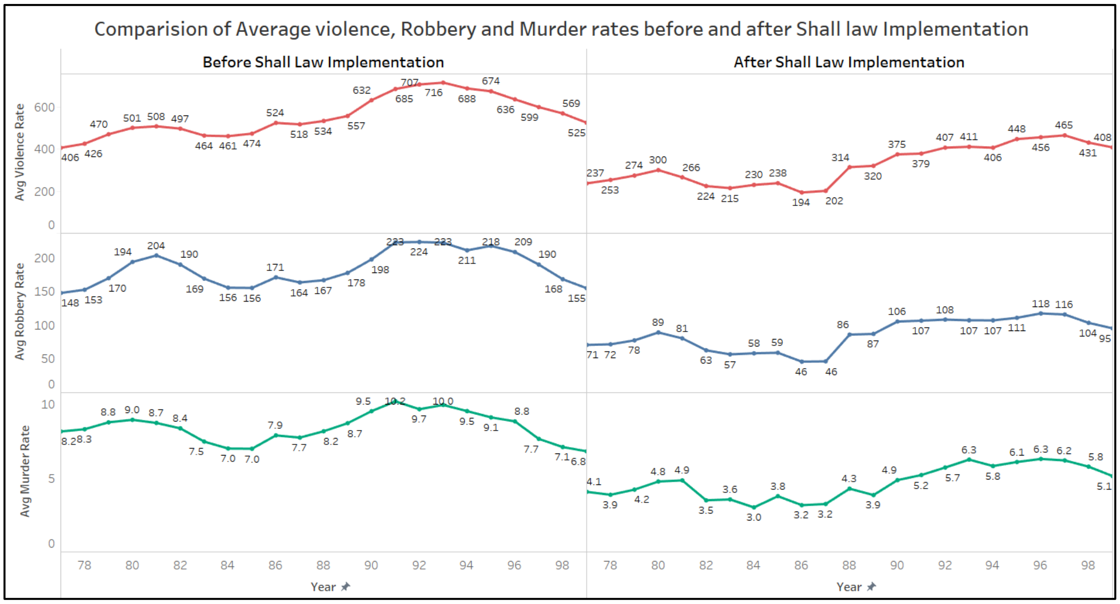
Variable Definitions

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)

2. Descriptive Analytics

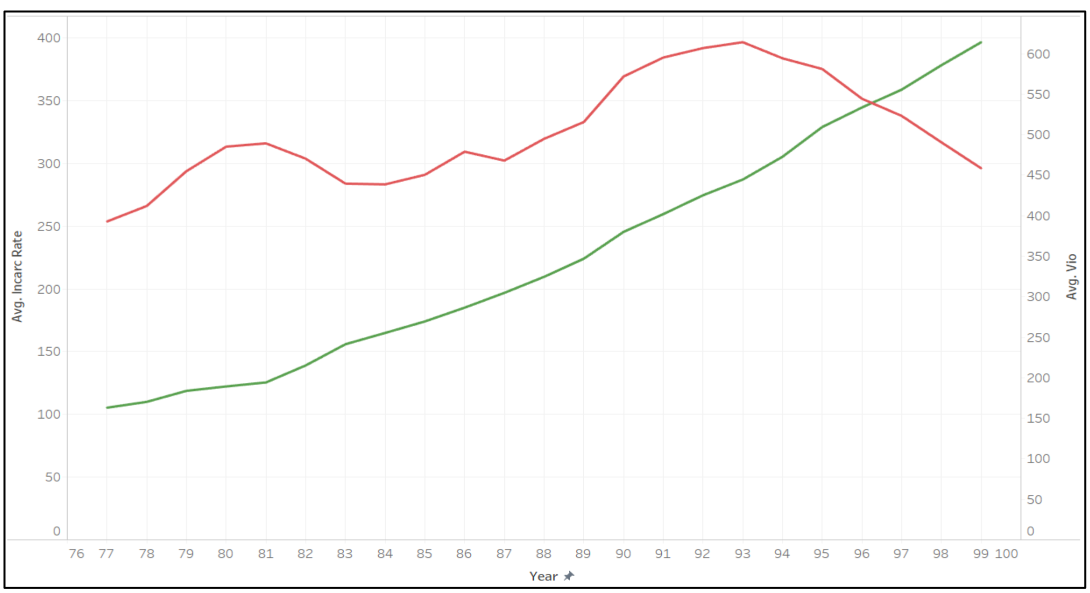
Before we entered the modelling stage, we did descriptive analysis to understand the summaries and trends of different variables in the given dataset.

In the first step, we tried understanding the impact on violence, robbery and murder rates before and after Shall Law implementation. The below picture clearly depicts that the average violence, robbery and murder rates have come down as the shall laws have been implemented.

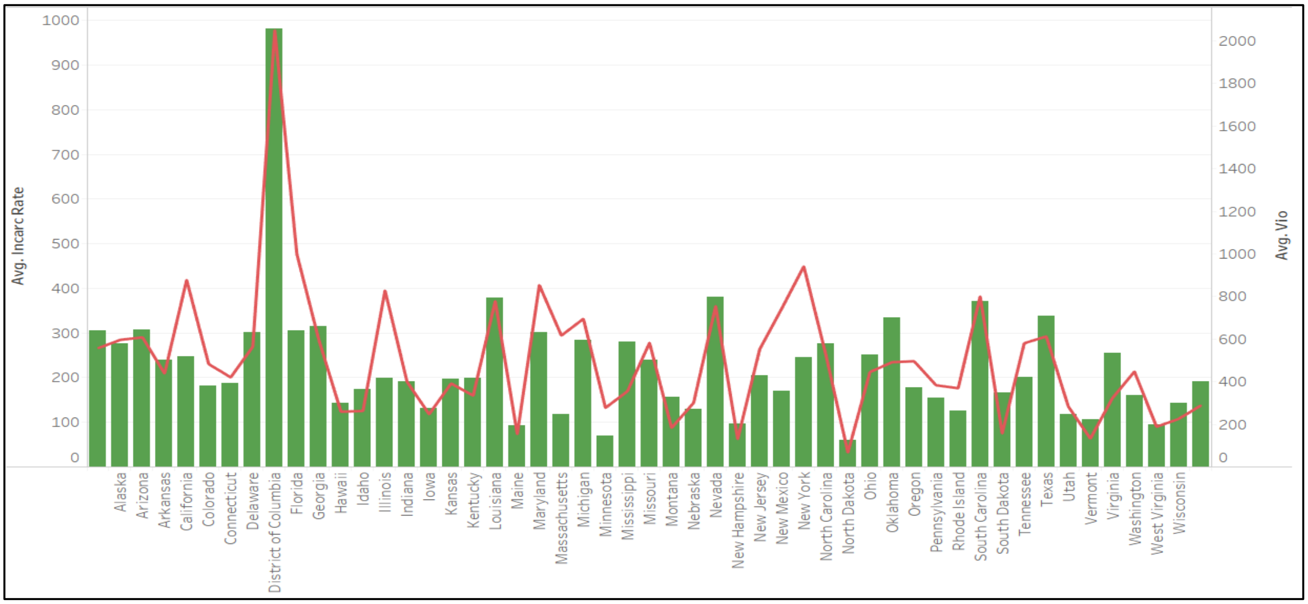


Also, from the plot above we observe the variations in violent crime rate, Murder rate and Robbery rates over the years. We observe that for the periods 1980-82 and 1990-94 there is a slight increase for all the three variables.

Next, we looked at incarceration rate trends. The below graph clearly depicts that there is an increasing trend observed w.r.t incarceration rate. Although clear conclusions couldn't be drawn about the impact of incarceration rate on Violence rate. (Red – violence rate, Green - Incarceration rate)



Also, it is observed that states DC have the highest average violence rates and North Dakota has the lowest average violence rates. Also, the same two states have higher incarceration rates as observed in the below image. (Red line – violence rate, Green bars – Average Incarceration rate).



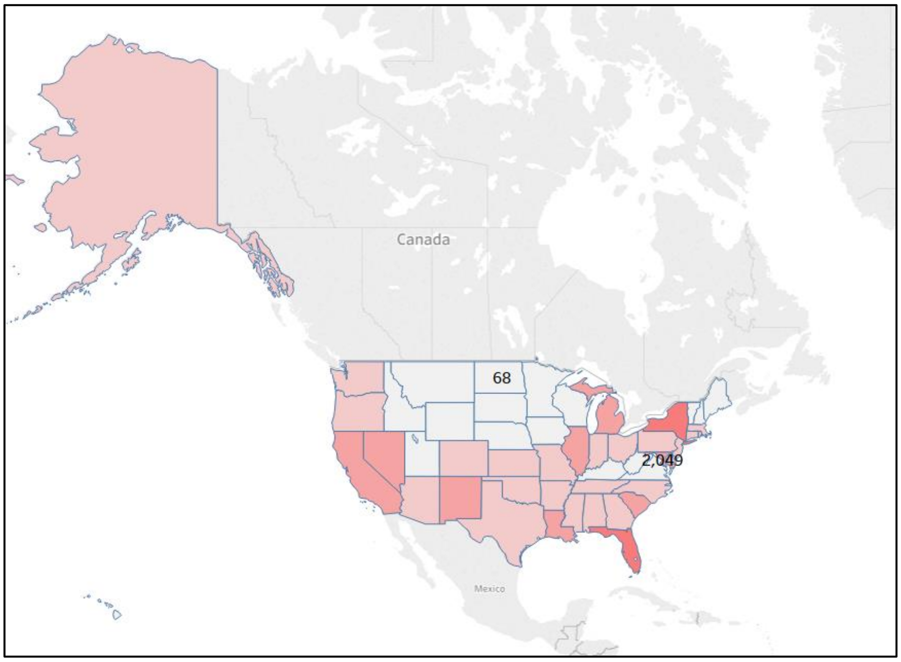
Next, we had a look at the correlation matrix below to understand the correlations between the variables. We observe that population density, Black population and Income have positive correlations with violence rate.

	Vio	*Year*	*Mur*	*Rob*	*Incarc Rate*	*Pb1064*	*Pw1064*	*Pm1029*	*Pop*	*Avginc*	*Density*
Vio	1.00										
Year	0.12	1.00									
Mur	0.83	-0.03	1.00								
Rob	0.91	-0.01	0.80	1.00							
Incarc Rate	0.70	0.50	0.71	0.57	1.00						
Pb1064	0.57	0.07	0.60	0.58	0.53	1.00					
Pw1064	-0.57	-0.03	-0.62	-0.58	-0.53	-0.98	1.00				
Pm1029	-0.17	-0.87	0.01	-0.09	-0.45	0.02	-0.01	1.00			
Pop	0.32	0.06	0.10	0.32	0.10	0.06	-0.07	-0.10	1.00		
Avginc	0.41	0.53	0.22	0.41	0.46	0.26	-0.19	-0.53	0.22	1.00	
Density	0.66	0.00	0.75	0.78	0.56	0.54	-0.56	-0.06	-0.08	0.34	1.00

We tend to observe some obvious correlations above like Murder rate and Robbery rates have higher correlations with Violence rate. So we cannot use all the three variables in the same model. As a result, we need to run models with these three variables separately.

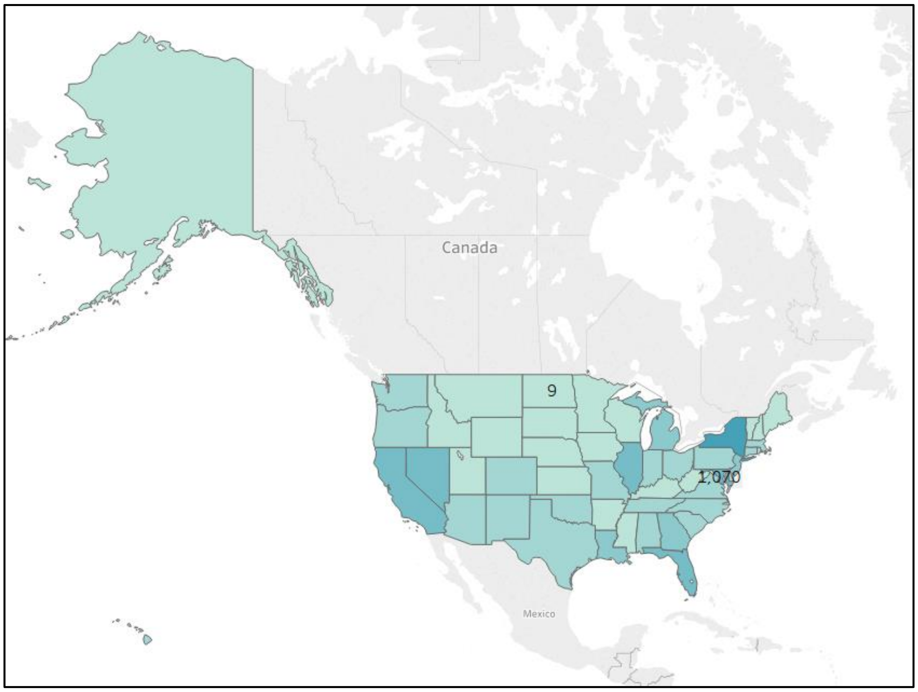
Average violence rates observed over states:

It is observed that the states North Dakota and District of Columbia have the lowest and highest average violence rates. Both the states with their average violence rate values are highlighted in the picture below. States like Florida, New York and California also tend to have higher violence rates as observed in the picture below.



Average Robbery rates observed over states:

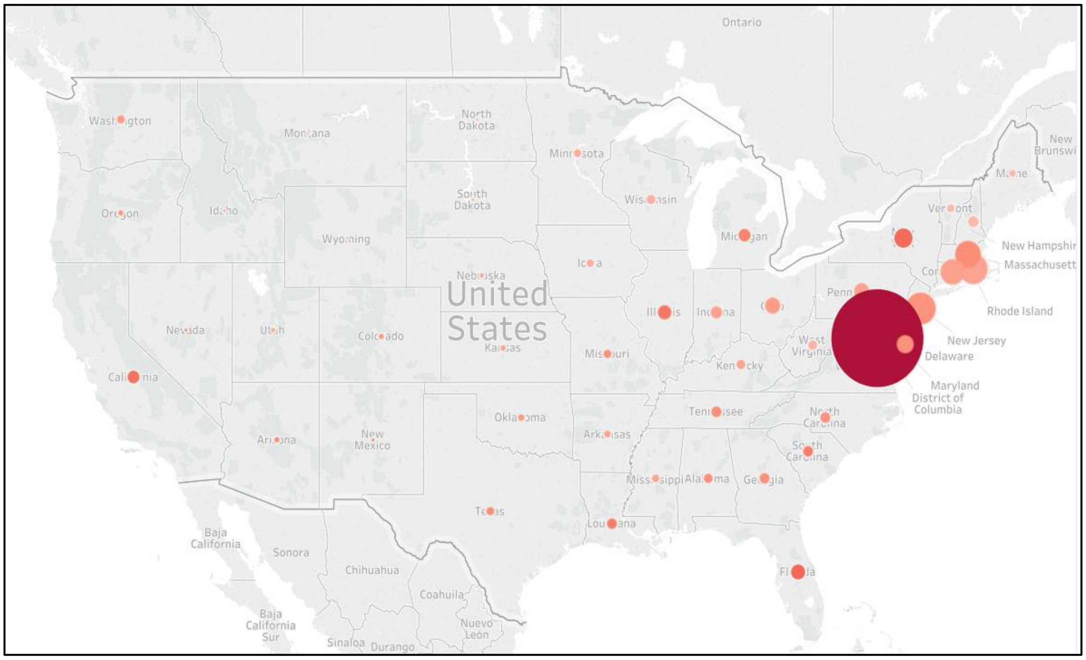
It is observed that the states North Dakota and District of Columbia have the lowest and highest average violence rates. Both the states with their average robbery rate values are highlighted in the picture below. States like New York, California, Illinois and Nevada also have higher average robbery rates.



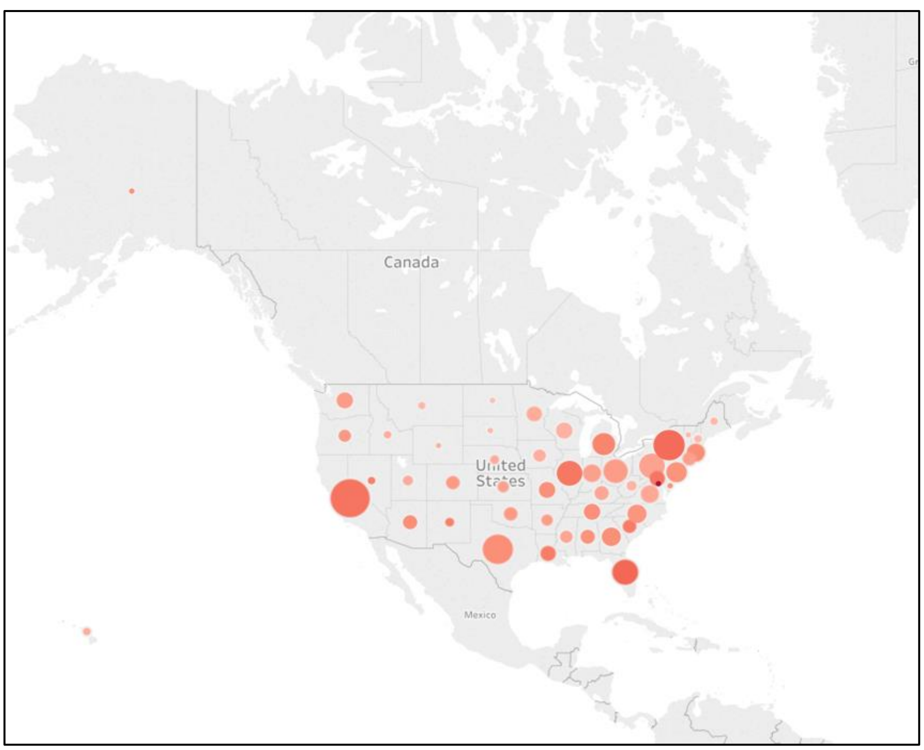
Impact of population density on Violence rate:

It is commonly thought that gun violence is higher in bigger cities and metros which have higher populations and population density. From the below graph (Average violence rates colour coded in Red, Population density by size of the circle), we observe that DC has the highest density and highest average violence rate. Also states like Florida, New York and California also have higher violence rates although their population densities are not as high as DC.

<u>Average Population Density Vs Average Violence rates</u>



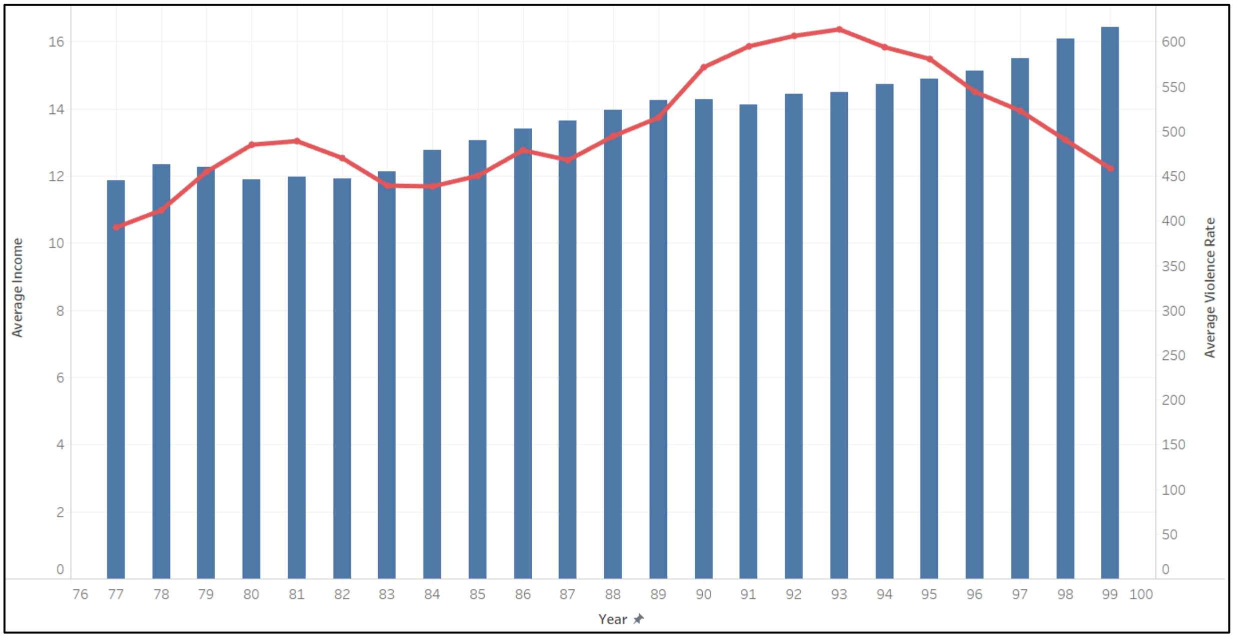
<u>Average Population (in Mn) Vs Violence rates</u>



Impact of Income on Violence rate:

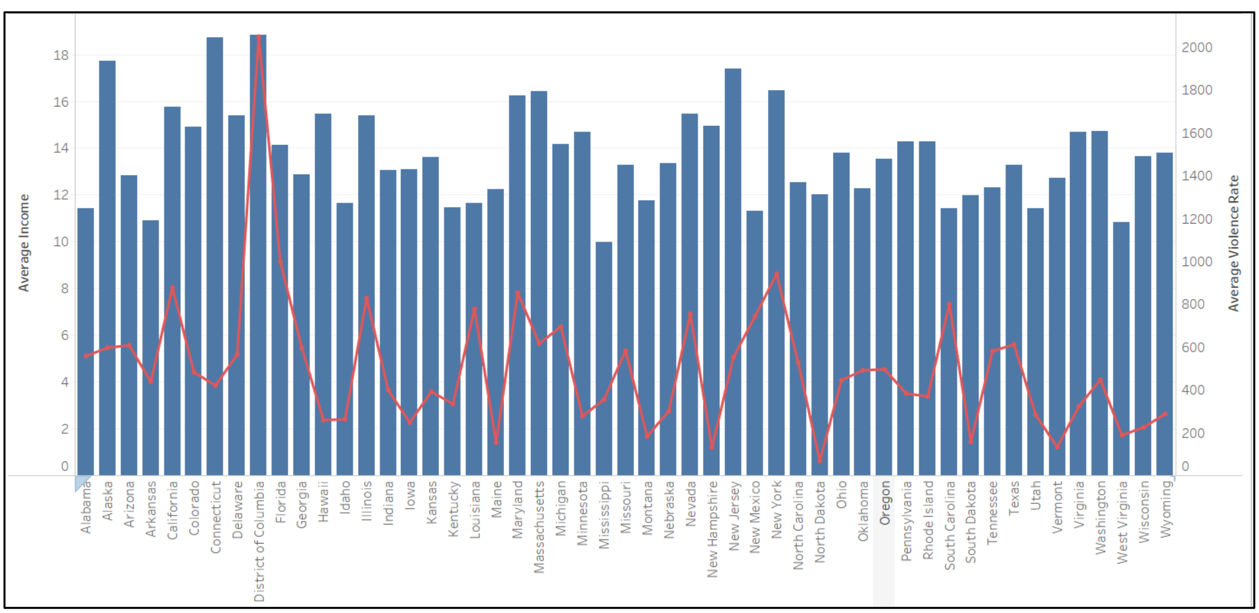
As per economic theory the state with high average income should be associated with a low average crime rate as low income/poverty can be one of the motivations for committing crimes. During the periods 77-79 and 83-89 there is a good relationship observed between Avg income and Avg violence rates. During the other periods there is no clear trend observed between the variables. May be some other variables contributed to violence rates during those periods.

Average Income (blue)/ Violence rate (red) Year wise



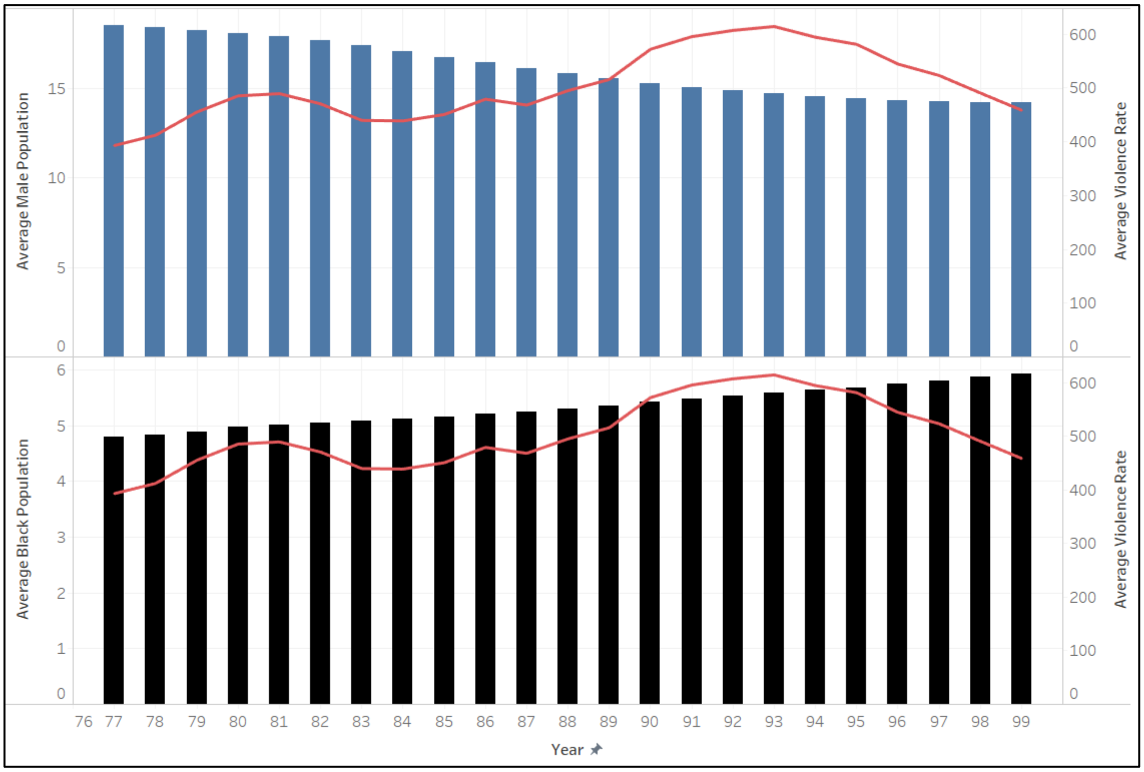
From the plot below, state DC has the highest Average income and highest violence rate. Also, it is observed that states like Hawaii, New Jersey and Connecticut etc have higher average income, but their average violence rates are not so high.

Average Income/Average Violence Rate State wise

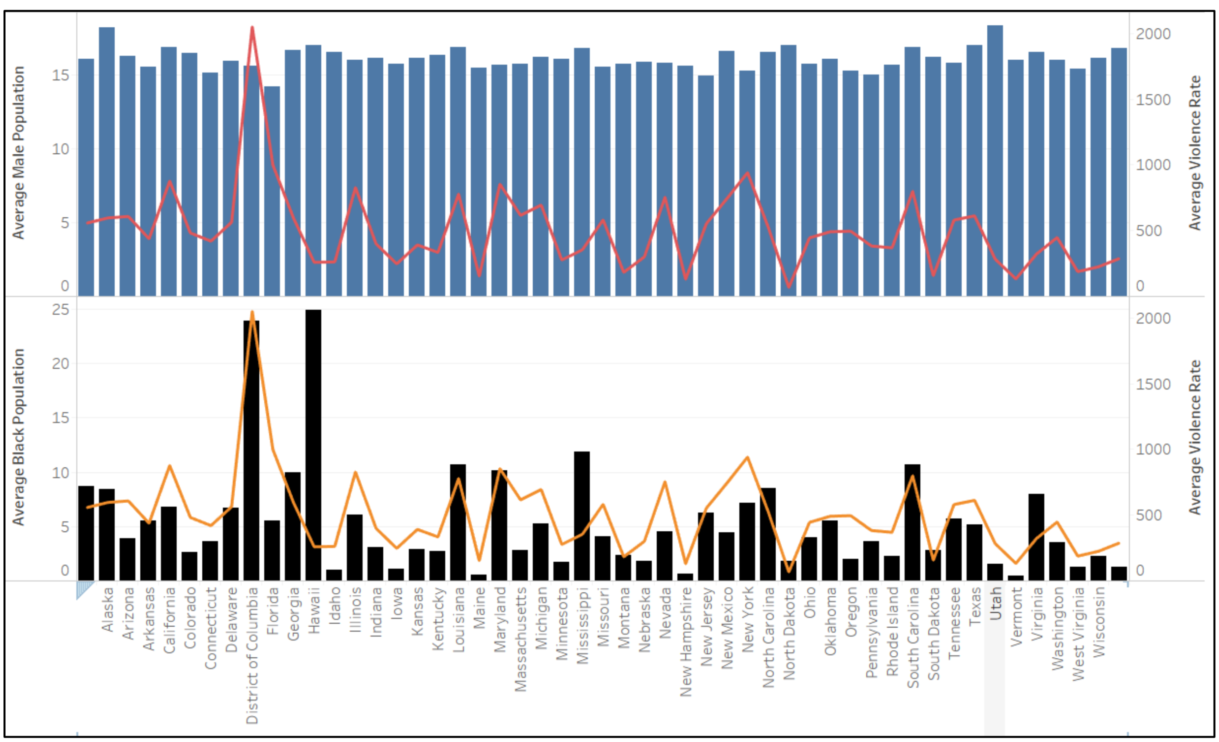


Impact of Male (Aged 10-19) and Black population on violence rate:

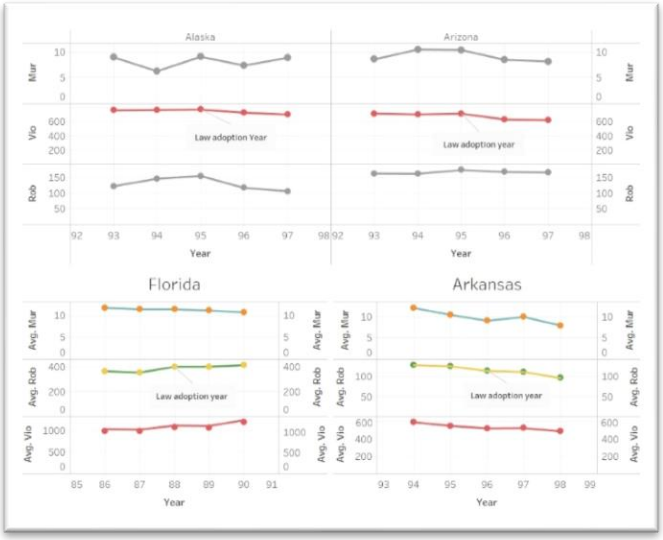
It is observed from the below plot that average male population (aged 10-29) is having a declining trend over the years. At the same time average black population is rising over the years. We don't see a proper correlation between male population and violence rate.

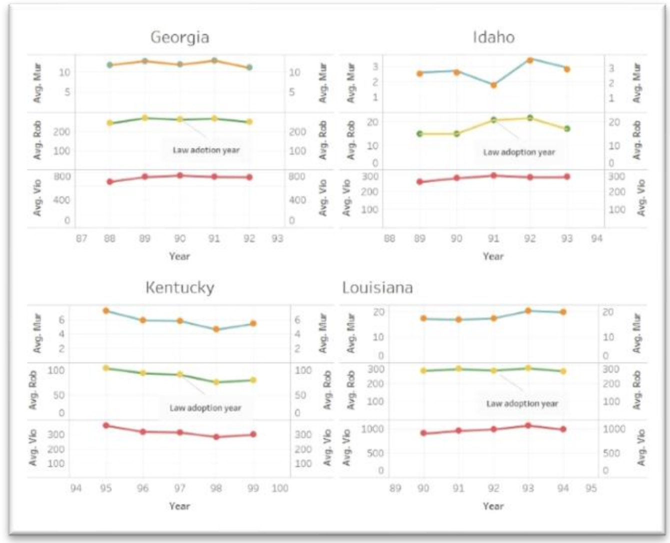


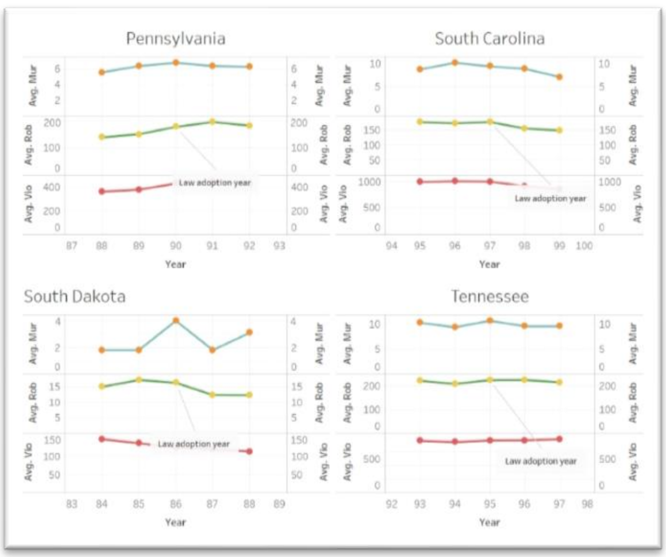
In the plot given below we observe that the in most states with higher average black population have higher violence rates like DC, Illinois, South Carolina etc.

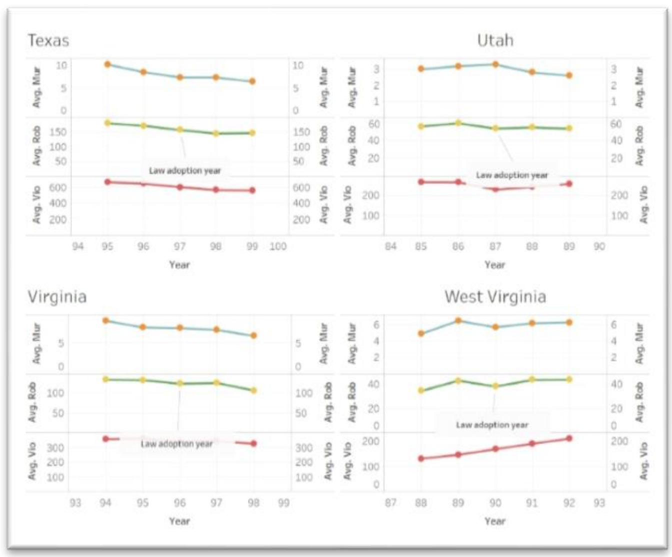


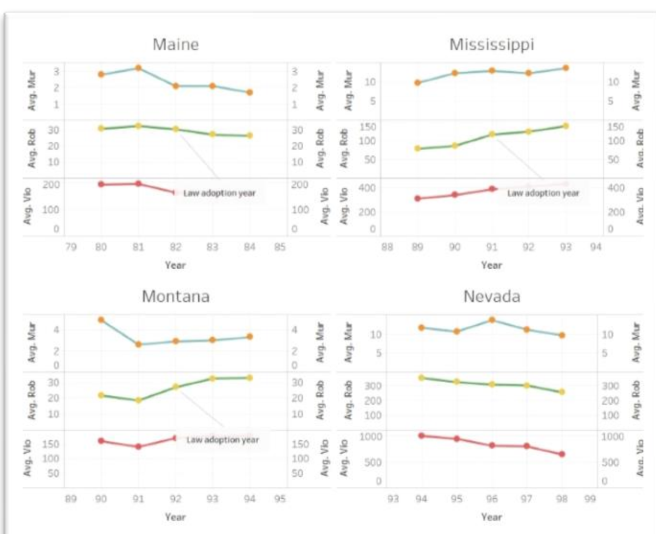
To further look at how these rates have changed after shall law adoption for the states that have adopted shall law, we have created line charts. Here we look at the rates 2 years before and 2 years after the law adoption.

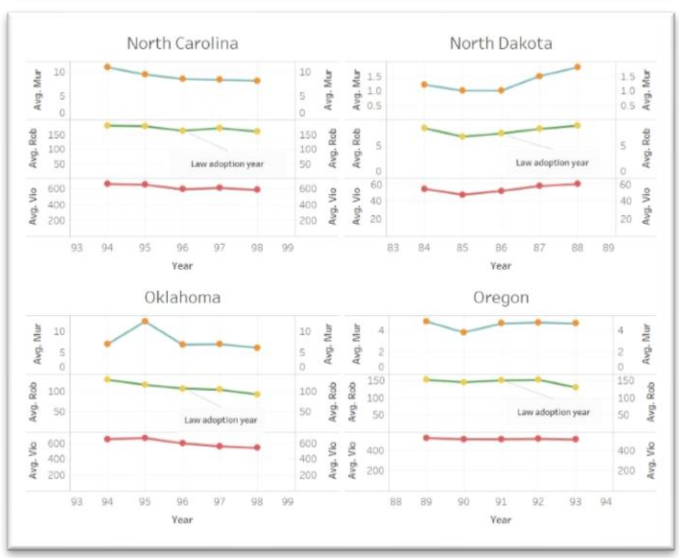






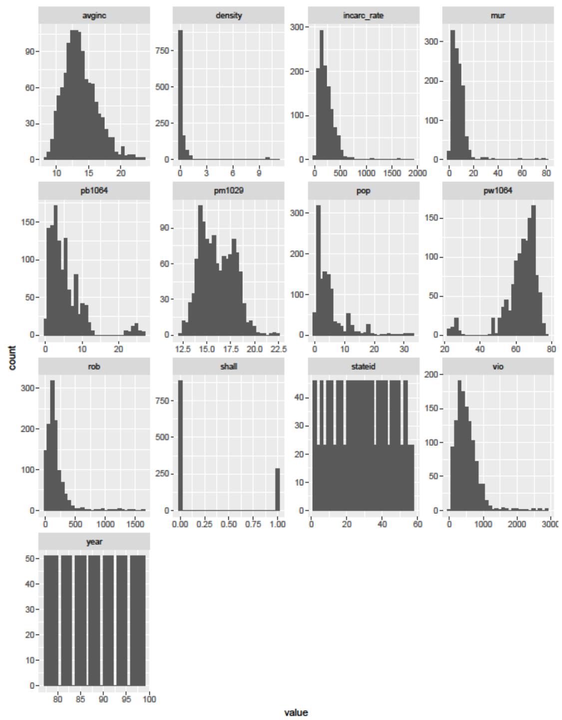




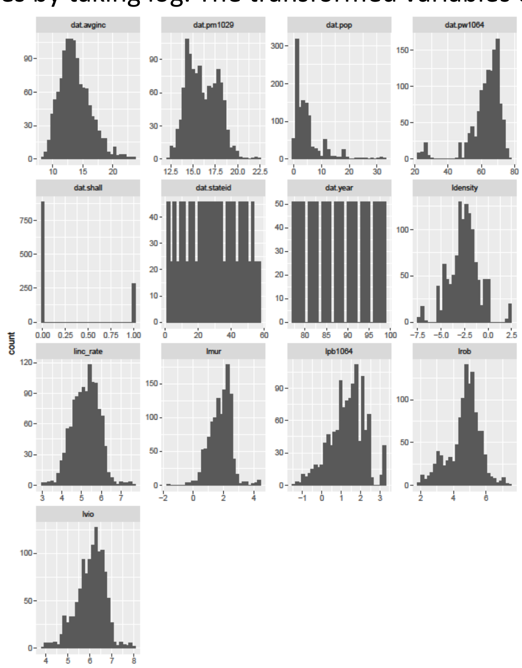


From these graphs above, we noticed that for different states, the trend is different. For e.g. in West Virginia there was a decline in the law adoption year for murder and robbery rate but since the law adoption we haven't noticed a decline. For states like Texas, there has been a steady decline. We can conclude from this descriptive analysis that there is some variable other than shall law that plays an important role in affecting these crime rates. We will investigate this claim further later in our report.

Before moving on to the modelling part of our analysis, we looked at the distributions of different discrete variables that we were provided with in the dataset.



After looking at the distributions of all these discrete variables, we noticed that variables like density, incarceration rate, murder rate, pb1064, violent crime rate and robbery rates are positively skewed. Hence, we transformed these variables by taking log. The transformed variables distributions are as below.



3. Violent Crime Rate

LINEAR REGRESSION:

In the first step we started with a simple model having only one independent variable (shall).

model1 <- lm(log(vio)~shall, data=dat)

summary(model1)

```
Call:
lm(formula = log(vio) ~ shall, data = dat)

Residuals:
     Min       1Q    Median       3Q      Max
-2.28477 - 0.42748  0.04655  0.42172  1.84504

Coefficients:
            Estimate Std. Error t value   Pr(>|t|)
(Intercept)  6.13492    0.02072  296.13   <2e-16 ***
shall       -0.44296    0.04203  -10.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6174 on 1171 degrees of freedom
Multiple R-squared:  0.08664,  Adjusted R-squared:  0.08586
F-statistic: 111.1 on 1 and 1171 DF, p-value: < 2.2e-16
```

Analysis:

In the model above, we are trying to determine the relationship between violent crime rate on a logarithmic scale and shall law. The estimate of shall is -0.44296 which indicates that there is 44% decrease in violent crime rate in a state having shall law. The estimate is significant at 0.1% confident interval. However, this is a very crude model as we are neglecting the panel structure of our data.

MODEL2: MULTIPLE REGRESSION

model2 <- lm(log(vio)~shall+log(incarc_rate), dat)

summary(model2)

```
Call:
lm(formula = log(vio) ~ shall + log(incarc_rate), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.51173 -0.31167  0.01675  0.30325  1.13191

Coefficients:
                 Estimate Std. Error t value   Pr(>|t|)
(Intercept)       2.78287    0.10304   27.01   <2e-16 ***
shall            -0.49104    0.03035  -16.18   <2e-16 ***
log(incarc_rate)  0.64653    0.01966   32.88   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4453 on 1170 degrees of freedom
Multiple R-squared:  0.5253,   Adjusted R-squared:  0.5245
F-statistic: 647.3 on 2 and 1170 DF,  p-value: < 2.2e-16
```

Analysis:

In step 2 we are have added one more independent variable incarc_rate on a logarithmic scale.

Shall law: The estimate of shall law is -0.49104 which indicates that there is 49% decrease in violent crime rate in a state having shall law. The estimate is significant at 1% confident interval.

Incarceration Rate: The estimate of shall law is 0.64653 which indicates if incarceration rate increases by 1% will lead to 0.64653% in violent crime rate. The estimate is significant at 1% confident interval. There is a serious potential for simultaneous causality bias: On one hand, increased incarceration reduces violent rate.

MODEL3: LINEAR REGRESSION

model3 <- lm(log(vio)~log(incarc_rate), dat)

summary(model3)

```
Call:
lm(formula = log(vio) ~ log(incarc_rate), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.60273 -0.28607  0.02391  0.32976  1.19426 

Coefficients:
                 Estimate Std. Error t value    Pr(>|t|)    
(Intercept)       2.74331    0.11390   24.08   <2e-16 ***
log(incarc_rate)  0.63120    0.02172   29.06   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4924 on 1171 degrees of freedom
Multiple R-squared:  0.4191,   Adjusted R-squared:  0.4186 
F-statistic: 844.7 on 1 and 1171 DF,  p-value: < 2.2e-16
```

Analysis:

In the above model we are checking whether incarceration has any effect on violent crime rate on a logarithmic scale. From the results the logarithmic estimate of incarceration rate is 0.63120 which indicates that if incarceration rate increases by 1% there is 0.63% increase in the violent crime rate. This result is inconsistent with our understanding that higher incarceration rate should decrease the robbery rate.

MODEL4: POOLED OLS MODEL

model4 <-
plm(log(vio)~shall+log(incarc_rate)+log(pb1064)+pw1064+pop+pm1029+avginc+log(density),index=c("stateid","year"), model="pooling",data=dat)

summary(model4)

```
Call:
plm(formula = log(vio) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + pop + pm1029 + avginc + log(density), data = dat,
    model = "pooling", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.239620 -0.210630  0.011983  0.251015  1.011706

Coefficients:
                   Estimate Std. Error t-value  Pr(>|t|)
(Intercept)       0.0659098  0.3098492  0.2127 0.8315859
shall            -0.2248926  0.0273296 -8.2289 5.016e-16 ***
log(incarc_rate)  0.5646460  0.0277901 20.3182 < 2.2e-16 ***
log(pb1064)       0.2603349  0.0282742  9.2075 < 2.2e-16 ***
pw1064            0.0167455  0.0022086  7.5818 6.939e-14 ***
pop               0.0165731  0.0023497  7.0533 2.990e-12 ***
pm1029            0.0981486  0.0093172 10.5341 < 2.2e-16 ***
avginc            0.0200272  0.0052307  3.8287 0.0001357 ***
log(density)      0.0988473  0.0086572 11.4179 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    488.63
Residual Sum of Squares: 149.73
R-Squared:      0.69358
Adj. R-Squared: 0.69148
F-statistic: 329.342 on 8 and 1164 DF, p-value: < 2.22e-16
```

Analysis:

After including all other variables, the coefficient of shall has decreased. We can interpret it as
Shall Law: The coefficient equals -0.2248956, which suggests that shall-issue laws reduce the violent crime rate by 22%. This is a large effect. This a large effect and may be overestimated. Also here the shall variable is highly significant.
Incarceration Rate: If the incarceration rate increases by 1% will lead to increase in violent crime rates by 0.56%. There is a serious potential for simultaneous causality bias: On one hand, increased incarceration reduces violent rate. On the other hand, if violent rate goes up and the police do their job, there will be more prisoners.
Pm1029: For every 1% increase in pm1029 will lead to increase in violent crime rate by 9.8% which is obvious.

Density:
Denser the area more violent rate will be there. So for every 1% increase in density will lead to 9.8% increase in violent rate. In this model all the variables are significant at 0.1% confident interval.

White corrected Pooled Model

summary(model4, vcov = vcovHC)

```
Call:
plm(formula = log(vio) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + pop + pm1029 + avginc + log(density), data = dat,
    model = "pooling", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.239620 -0.210630  0.011983  0.251015  1.011706

Coefficients:
                  Estimate Std. Error   t-value  Pr(>|t|)
(Intercept)       0.0659098  1.1511911  0.0573 0.9543530
shall            -0.2248926  0.0666756 -3.3729 0.0007682 ***
log(incarc_rate)  0.5646460  0.1150240  4.9089 1.046e-06 ***
log(pb1064)       0.2603349  0.0924918  2.8147 0.0049648 **
pw1064            0.0167455  0.0082249  2.0360 0.0419805 *
pop               0.0165731  0.0073201  2.2641 0.0237534 *
pm1029            0.0981486  0.0301878  3.2513 0.0011818 **
avginc            0.0200272  0.0150095  1.3343 0.1823661
log(density)      0.0988473  0.0327134  3.0216 0.0025692 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    488.63
Residual Sum of Squares: 149.73
R-Squared:      0.69358
Adj. R-Squared: 0.69148
F-statistic: 34.4201 on 8 and 50 DF, p-value: < 2.22e-16
```

Analysis:

The white corrected model shows that the significance of the variable density, average income has decreased. log(density), log(pm1064), pm1029 seems to be significant at 10 % level. Pw1064, pop are not significant even at 10 % level. The variable avginc is not having any significant impact in this model.

F TEST TO CHECK WHETHER THE VARIABLE AVGINC IS SIGNIFICANT OR NOT

linearHypothesis(model4, "avginc", vcov = vcovHC)

```
Linear hypothesis test

Hypothesis:
avginc = 0

Model 1: restricted model
Model 2: log(vio) ~ shall + log(incarc_rate) + log(pb1064) + pw1064 +
    pop + pm1029 + avginc + log(density)

Note: Coefficient covariance matrix supplied.

  Res.Df  Df  Chisq   Pr(>Chisq)
1   1165
2   1164  1  1.7804     0.1821
```

Analysis:

The p-value is .1821. Hence, we fail to reject the null and conclude that average income's coefficient i s not significantly different from zero and that it is an irrelevant variable in our model.

MODEL5: FIXED EFFECTS MODEL

model5<-plm(log(vio)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+pop+log(density)+pm1029,index=c("st ateid","year"),model="within",data=dat)

summary(model5)

```
Call:
plm(formula = log(vio) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + avginc + pop + log(density) + pm1029, data = dat,
    model = "within", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.575624 -0.100739   0.005564  0.103467  0.551996

Coefficients:
                   Estimate Std. Error t-value  Pr(>|t|)
shall            -0.0095591  0.0186818 -0.5117 0.6089781
log(incarc_rate) -0.0092714  0.0282222 -0.3285 0.7425833
log(pb1064)      -0.1818298  0.0550045 -3.3057 0.0009775 ***
pw1064            0.0231878  0.0043157  5.3729 9.429e-08 ***
avginc            0.0040626  0.0058890  0.6899 0.4904215
pop               0.0251254  0.0094149  2.6687 0.0077257 **
log(density)     -0.1449091  0.0868631 -1.6682 0.0955477 .
pm1029           -0.0660520  0.0084332 -7.8323 1.111e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    36.789
Residual Sum of Squares: 29.305
R-Squared:      0.20343
Adj. R-Squared: 0.16196
F-statistic: 35.562 on 8 and 1114 DF, p-value: < 2.22e-16
```

Analysis:

From the above results we can see that shall, log(incar_rate), avginc, are not significant even at 10% confident interval. We can interpret pm1029 as 1% increase in the percent of population of males between 10 to 29 with de crease the crime rate by 6.6%. log(pb1064), pw1064, pm1029 are still significant at 0.1% confident interval where as log(density) is significant at 10% confident interval. The value of Adj R squared is also very less.

summary(model5, vcov = vcovHC)

```
Oneway (individual) effect Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = log(vio) ~ shall + log(incarc_rate) + log(pb1064) + 
    pw1064 + avginc + pop + log(density) + pm1029, data = dat, 
    model = "within", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max. 
-0.575624 -0.100739  0.005564  0.103467  0.551996 

Coefficients:
                   Estimate Std. Error t-value Pr(>|t|)   
shall            -0.0095591  0.0422335 -0.2263 0.820980   
log(incarc_rate) -0.0092714  0.0687138 -0.1349 0.892693   
log(pb1064)      -0.1818298  0.2138604 -0.8502 0.395382   
pw1064            0.0231878  0.0118924  1.9498 0.051450 . 
avginc            0.0040626  0.0134431  0.3022 0.762548   
pop               0.0251254  0.0106883  2.3507 0.018909 * 
log(density)     -0.1449091  0.1751258 -0.8275 0.408155   
pm1029           -0.0660520  0.0226627 -2.9146 0.003633 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    36.789
Residual Sum of Squares: 29.305
R-Squared:      0.20343
Adj. R-Squared: 0.16196
F-statistic: 8.98565 on 8 and 50 DF, p-value: 1.5421e-07
```

Analysis:

From the above results we can see that pm1029 is significant at 1% confident interval and pop is significant at 5%. Rest of all the variables are not having any significant effect. The estimate of shall has decreased drastically.

MODEL6: TIME AND ENTITY FIXED EFFECTS

```
model6<- plm(log(vio)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+pop+log(density)+pm1029+factor(yea
r),data=dat,index=c("stateid","year"), model ="within")
coeftest(model6, vcovHC)
```

```
t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)    
shall            -0.0448508  0.0355792 -1.2606   0.20773    
log(incarc_rate) -0.1072195  0.0619256 -1.7314   0.08366 .  
log(pb1064)      -0.4288153  0.1873959 -2.2883   0.02231 *  
pw1064           -0.0036306  0.0132245 -0.2745   0.78373    
avginc            0.0063910  0.0143064  0.4467   0.65516    
pop              -0.0025200  0.0107455 -0.2345   0.81463    
log(density)     -0.2299976  0.1800161 -1.2776   0.20164    
pm1029            0.0906899  0.0330317  2.7455   0.00614 ** 
factor(year)78    0.0746987  0.0128736  5.8025 8.554e-09 ***
factor(year)79    0.2023756  0.0219187  9.2330 < 2.2e-16 ***
factor(year)80    0.2790765  0.0302812  9.2162 < 2.2e-16 ***
factor(year)81    0.2976852  0.0321859  9.2489 < 2.2e-16 ***
factor(year)82    0.3018164  0.0405524  7.4426 1.996e-13 ***
factor(year)83    0.2891444  0.0495218  5.8387 6.931e-09 ***
factor(year)84    0.3385867  0.0598821  5.6542 1.998e-08 ***
factor(year)85    0.4062637  0.0704736  5.7648 1.063e-08 ***
factor(year)86    0.5037355  0.0826340  6.0960 1.506e-09 ***
factor(year)87    0.5223179  0.0949048  5.5036 4.636e-08 ***
factor(year)88    0.6038152  0.1052204  5.7386 1.236e-08 ***
factor(year)89    0.6783878  0.1149711  5.9005 4.830e-09 ***
factor(year)90    0.8306506  0.1355799  6.1266 1.251e-09 ***
factor(year)91    0.9046770  0.1420593  6.3683 2.810e-10 ***
factor(year)92    0.9556750  0.1501192  6.3661 2.849e-10 ***
factor(year)93    0.9958397  0.1572354  6.3334 3.497e-10 ***
factor(year)94    0.9981378  0.1634739  6.1058 1.420e-09 ***
factor(year)95    1.0117607  0.1725895  5.8622 6.043e-09 ***
factor(year)96    0.9753904  0.1812976  5.3801 9.108e-08 ***
factor(year)97    0.9704780  0.1868621  5.1936 2.460e-07 ***
factor(year)98    0.9289378  0.1941139  4.7855 1.939e-06 ***
factor(year)99    0.8849205  0.1996877  4.4315 1.030e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis:

The absolute value of the coefficient on shall falls further to -0.044, the coefficient is not significantly different from zero. The time effects are jointly statistically significant, so this regression seems better specified than the regression using Fixed Effect model. From the above results we can see that pm1029 is significant at 1% confident interval and log(pw1064) is significant at 5%. Rest of all the variables are not having any significant effect. The estimate of shall has decreased drastically.

MODEL7: RANDOM EFFECTS

```
model7<- plm(log(vio)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+pop+log(density)+pm1029+factor(yea
r),data=dat,index=c("stateid","year"), model ="random")
coeftest(model7, vcovHC)
```

```
t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)    
(Intercept)       4.7052918  0.7769276  6.0563 1.888e-09 ***
shall            -0.0364334  0.0400212 -0.9104 0.3628285    
log(incarc_rate)  0.0046763  0.0686533  0.0681 0.9457060    
log(pb1064)      -0.0370559  0.1297469 -0.2856 0.7752353    
pw1064           -0.0047792  0.0118529 -0.4032 0.6868681    
avginc            0.0089268  0.0144262  0.6188 0.5361765    
pop               0.0076908  0.0112074  0.6862 0.4927112    
log(density)      0.1120305  0.0618706  1.8107 0.0704464 .  
pm1029            0.0864204  0.0342251  2.5251 0.0117024 *  
factor(year)78    0.0564547  0.0129098  4.3730 1.337e-05 ***
factor(year)79    0.1636309  0.0211384  7.7409 2.164e-14 ***
factor(year)80    0.2216461  0.0283419  7.8204 1.191e-14 ***
factor(year)81    0.2229821  0.0305777  7.2923 5.682e-13 ***
factor(year)82    0.2021522  0.0388203  5.2074 2.270e-07 ***
factor(year)83    0.1650150  0.0492618  3.3498 0.0008353 ***
factor(year)84    0.1953737  0.0620557  3.1484 0.0016844 ** 
factor(year)85    0.2453433  0.0740208  3.3145 0.0009468 ***
factor(year)86    0.3234316  0.0876613  3.6896 0.0002352 ***
factor(year)87    0.3236376  0.0993173  3.2586 0.0011525 ** 
factor(year)88    0.3862589  0.1087880  3.5506 0.0004000 ***
factor(year)89    0.4415131  0.1203744  3.6678 0.0002559 ***
factor(year)90    0.5708032  0.1360884  4.1944 2.948e-05 ***
factor(year)91    0.6277867  0.1439535  4.3610 1.411e-05 ***
factor(year)92    0.6593618  0.1512847  4.3584 1.428e-05 ***
factor(year)93    0.6828564  0.1568794  4.3527 1.465e-05 ***
factor(year)94    0.6692786  0.1622554  4.1248 3.979e-05 ***
factor(year)95    0.6642768  0.1681153  3.9513 8.249e-05 ***
factor(year)96    0.6097658  0.1773551  3.4381 0.0006069 ***
factor(year)97    0.5885795  0.1835104  3.2073 0.0013769 ** 
factor(year)98    0.5299168  0.1922427  2.7565 0.0059354 ** 
factor(year)99    0.4703716  0.1994210  2.3587 0.0185075 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis:

The absolute value of the coefficient on shall falls further to -0.0364, the coefficient is not significantly different from zero. The time effects are jointly statistically significant, so this regression seems better specified than the regression using Fixed Effect model. From the above results we can see that pm1029 is significant at 5% confident interval. Rest of all the variables are not having any significant effect. The estimate of shall has decreased drastically.

HAUSMAN TEST:

phtest(model6, model7)

```
        Hausman Test

data:  log(vio) ~ shall + log(incarc_rate) + log(pb1064) + pw1064 +  ...
chisq = 180.97, df = 30, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

Analysis:

According to Hausman test, the p-value is less 5%. So, we reject the null hypothesis that there is no endogeneity in the model and we conclude that the random effects model is not the correct mod el for us. Also, since we have data on all the 51 states we do not have a random sample and we should avoid using our Random Effects.

We conclude that entity time fixed effects model is the best model for our data.

4. Murder Rate

Simple Linear Regression:

model2.1 <- lm(log(mur)~shall, data=dat)

Output:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.89756    0.02261   83.93   <2e-16 ***
shall       -0.47337    0.04587  -10.32   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6737 on 1171 degrees of freedom
Multiple R-squared:  0.08337,	Adjusted R-squared:  0.08259 
F-statistic: 106.5 on 1 and 1171 DF,  p-value: < 2.2e-16
```

Interpretation:

We can see that the coefficient on Shall is highly significant (1% significance level), with an estimate of -0.47337. It can be interpreted that if shall law is introduced; the murder rate reduces by 47.33 %. From this above linear regr ession, we can say that the shall law reduces the Murder rate.

Simple Linear Regression:

model2.3 <- lm(log(mur)~log(incarc_rate), dat)

Output:

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -1.25698    0.13597  -9.245   <2e-16 ***
log(incarc_rate)  0.58422    0.02593  22.535   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5877 on 1171 degrees of freedom
Multiple R-squared:  0.3025,	Adjusted R-squared:  0.3019 
F-statistic: 507.8 on 1 and 1171 DF,  p-value: < 2.2e-16
```

Interpretation:
From the results, we can see that incarceration rate is a very significant variable from its p value. To interpret, for every 1% percent increase in incarceration rate, the murder rate increases by 0.58%. This result is inconsistent wit h our understanding.

Multiple Linear Regression:

model2.2 <- lm(log(mur)~shall+log(incarc_rate), dat)

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -1.21525    0.12598  -9.647   <2e-16 ***
shall            -0.51802    0.03710 -13.961   <2e-16 ***
log(incarc_rate)  0.60039    0.02404  24.974   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5444 on 1170 degrees of freedom
Multiple R-squared:  0.4021,	Adjusted R-squared:  0.4011 
F-statistic: 393.4 on 2 and 1170 DF,  p-value: < 2.2e-16
```

After adding shall to the previous model, we can see that there is change in the coefficients. We can interpret this as follows. We can interpret this as: When shall is adopted, the murder rate decreases by 51.8%. Thus, we can see that shall is highly significant variable and the effect of shall law adoption has caused the murder rate to decrease further.

Pooled OLS Model:

model2.4 <-plm(log(mur)~shall+log(incarc_rate)+log(pb1064)+pw1064+pop+pm1029+avginc +log(density),index=c("stateid","year"), model="pooling",data=dat)

The coefficients are:

```
Coefficients :
                   Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)      -3.8736996  0.3524756 -10.9900 < 2.2e-16 ***
shall            -0.1438363  0.0310894  -4.6265 4.135e-06 ***
log(incarc_rate)  0.6205142  0.0316132  19.6283 < 2.2e-16 ***
log(pb1064)       0.2577878  0.0321639   8.0148 2.664e-15 ***
pw1064            0.0034797  0.0025125   1.3850    0.1663
pop               0.0147782  0.0026729   5.5288 3.977e-08 ***
pm1029            0.1579261  0.0105990  14.9001 < 2.2e-16 ***
avginc           -0.0367214  0.0059503  -6.1713 9.335e-10 ***
log(density)      0.0795709  0.0098482   8.0797 1.612e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    579.9
Residual Sum of Squares: 193.75
R-Squared:      0.66588
Adj. R-Squared: 0.66359
F-statistic: 289.977 on 8 and 1164 DF, p-value: < 2.22e-16
```

After including all the other explanatory variables, we can see that the coefficient of shall variable decreased drastically. We can interpret as: if shall law is adopted the murder rate reduces by 14.38% only. But as we can see it is highly significant. All the variables are highly significant at 5% level except pw1064. Its p value is 0.1663 which is not even significant at 10 % level. The adjusted R squared is quite good.

White Corrected Pooled Model:

summary(model2.4, vcov = vcovHC)

```
Coefficients :
                   Estimate Std. Error t-value  Pr(>|t|)
(Intercept)      -3.8736996  1.5939210 -2.4303 0.0152368 *
shall            -0.1438363  0.0756981 -1.9001 0.0576629 .
log(incarc_rate)  0.6205142  0.1513113  4.1009 4.401e-05 ***
log(pb1064)       0.2577878  0.1290855  1.9970 0.0460543 *
pw1064            0.0034797  0.0122210  0.2847 0.7759018
pop               0.0147782  0.0081725  1.8083 0.0708217 .
pm1029            0.1579261  0.0405085  3.8986 0.0001023 ***
avginc           -0.0367214  0.0226350 -1.6223 0.1050043
log(density)      0.0795709  0.0406815  1.9559 0.0507102 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    579.9
Residual Sum of Squares: 193.75
R-Squared:      0.66588
Adj. R-Squared: 0.66359
F-statistic: 35.0922 on 8 and 50 DF, p-value: < 2.22e-16
```

The white corrected model shows that the significance of the variable shall has decreased. Shall,pop,log(density) seems to be significant at 10 % level. Pw1064, avginc are not significant even at 10 % level. Rest all variables are significant.

Fixed Effects Model:

model2.5<-plm(log(mur)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+pop+log(density)
+pm1029,index=c("stateid","year"),model="within",data=dat)

```
Coefficients :
                   Estimate  Std. Error t-value  Pr(>|t|)    
shall             -0.06359977  0.02539854 -2.5041  0.012419 *  
log(incarc_rate) -0.16322229  0.03836899 -4.2540 2.276e-05 ***
log(pb1064)       -0.07001010  0.07478033 -0.9362  0.349368    
pw1064             0.01885083  0.00586730  3.2129  0.001352 ** 
avginc             0.03840862  0.00800630  4.7973 1.827e-06 ***
pop               -0.01043381  0.01279988 -0.8151  0.415161    
log(density)      -0.37588839  0.11809316 -3.1830  0.001498 ** 
pm1029             0.00063193  0.01146526  0.0551  0.956055    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    63.314
Residual Sum of Squares: 54.165
R-Squared:      0.1445
Adj. R-Squared: 0.099957
F-statistic: 23.52 on 8 and 1114 DF, p-value: < 2.22e-16
```

From this above Fixed effects model, we can see that except for log(incarc_rate), pop, pm1029 rest all variables are significant. According to this model, shall law decreases murder rate by 6.35%. From this we can see that the results differ from pooling model. And we can conclude that there is observed and unobserved heterogeniety in our data that was unaccounted for in the pooling model.

Time and Entity Fixed Effects with white corrected standard error:

model2.6<- plm(log(mur)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+pop+log(density)
+pm1029+factor(year),data=dat,index=c("stateid","year"), model ="within")

coeftest(model2.6, vcovHC)

```
t test of coefficients:

                  Estimate Std. Error t value Pr(>|t|)   
shall           -0.0225671  0.0379884 -0.5941 0.552601   
log(incarc_rate) -0.0826066  0.0685446 -1.2052 0.228406   
log(pb1064)     -0.0835812  0.1912257 -0.4371 0.662139   
pw1064           0.0105280  0.0175585  0.5996 0.548900   
avginc           0.0723970  0.0238146  3.0400 0.002422 **
pop             -0.0194031  0.0268317 -0.7231 0.469749   
log(density)    -0.2575235  0.1955396 -1.3170 0.188119   
pm1029           0.0384717  0.0325341  1.1825 0.237263   
factor(year)78  -0.0029502  0.0320568 -0.0920 0.926691   
factor(year)79   0.0668937  0.0297725  2.2468 0.024850 * 
factor(year)80   0.1149502  0.0394924  2.9107 0.003679 **
factor(year)81   0.1279939  0.0444923  2.8768 0.004096 **
factor(year)82   0.0564654  0.0485080  1.1640 0.244660   
factor(year)83   0.0042069  0.0606181  0.0694 0.944684   
factor(year)84  -0.1103576  0.0653995 -1.6874 0.091805 . 
factor(year)85  -0.0668263  0.0828783 -0.8063 0.420235   
factor(year)86   0.0023609  0.0859665  0.0275 0.978095   
factor(year)87  -0.0182743  0.0967169 -0.1889 0.850170   
factor(year)88  -0.0109228  0.1038479 -0.1052 0.916251   
```

```
factor(year)90   0.0484605  0.1424821  0.3401 0.733834   
factor(year)91   0.0994693  0.1503439  0.6616 0.508359   
factor(year)92   0.0621263  0.1632926  0.3805 0.703678   
factor(year)93   0.1512735  0.1640240  0.9223 0.356594   
factor(year)94   0.0402407  0.1817128  0.2215 0.824782   
factor(year)95   0.0550843  0.1875078  0.2938 0.768989   
factor(year)96  -0.0126120  0.1949836 -0.0647 0.948439   
factor(year)97  -0.1182865  0.2073494 -0.5705 0.568477   
factor(year)98  -0.1855494  0.2188789 -0.8477 0.396776   
factor(year)99  -0.2561003  0.2302589 -1.1122 0.266285   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After adding the fixed time effects, shall issue law further reduce the murder rate. If shall law is interpreted, the murder rate reduces by 2.25%. And that is not significant in the model even at 10 % level. This states that the shall law has no effect on murder rate. Most of the variables became non-significant.

We cannot use Random Effect model because we are considering entire population in the data set (51 states) and not a sample from population. We can prove this by building a Random effects model and using Hausman test to verify which model should be used.

Random Effects model with White Corrected Standard Error:

model2.7<- plm(log(mur)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+pop+log(density)

+pm1029+factor(year),data=dat,index=c("stateid","year"), model ="random")

coeftest(model2.7, vcovHC)

```
t test of coefficients:

                  Estimate Std. Error t value  Pr(>|t|)
(Intercept)      -0.4861526  0.8172107 -0.5949 0.5520331
shall            -0.0064290  0.0402318 -0.1598 0.8730687
log(incarc_rate)  0.0813820  0.0979714  0.8307 0.4063333
log(pb1064)       0.3515156  0.0999419  3.5172 0.0004533 ***
pw1064            0.0035584  0.0092548  0.3845 0.7006851
avginc            0.0585605  0.0285797  2.0490 0.0406879 *
pop              -0.0012429  0.0099255 -0.1252 0.9003720
log(density)      0.0207929  0.0460825  0.4512 0.6519235
pm1029            0.0393689  0.0278715  1.4125 0.1580711
factor(year)78   -0.0151875  0.0325391 -0.4667 0.6407708
factor(year)79    0.0292481  0.0326178  0.8967 0.3700725
factor(year)80    0.0499234  0.0380750  1.3112 0.1900591
factor(year)81    0.0451211  0.0423877  1.0645 0.2873346
factor(year)82   -0.0588143  0.0532824 -1.1038 0.2699025
factor(year)83   -0.1381523  0.0632501 -2.1842 0.0291487 *
factor(year)84   -0.2647937  0.0800673 -3.3071 0.0009719 ***
factor(year)85   -0.2369904  0.1041921 -2.2746 0.0231175 *
factor(year)86   -0.1862804  0.1136745 -1.6387 0.1015476
factor(year)87   -0.2251987  0.1266441 -1.7782 0.0756369 .
factor(year)88   -0.2350627  0.1352082 -1.7385 0.0823880 .
factor(year)89   -0.2530818  0.1586254 -1.5955 0.1108843
factor(year)90   -0.2118095  0.1727627 -1.2260 0.2204460
factor(year)91   -0.1838652  0.1826280 -1.0068 0.3142563
factor(year)92   -0.2389632  0.1964158 -1.2166 0.2240007
factor(year)93   -0.1680904  0.2000665 -0.8402 0.4009874
factor(year)94   -0.2932702  0.2213513 -1.3249 0.1854666
factor(year)95   -0.2982383  0.2255789 -1.3221 0.1863989
factor(year)96   -0.3840658  0.2338408 -1.6424 0.1007774
factor(year)97   -0.5033770  0.2517000 -1.9999 0.0457466 *
factor(year)98   -0.5815275  0.2731487 -2.1290 0.0334696 *
factor(year)99   -0.6646000  0.2842422 -2.3381 0.0195516 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We can see that by using the random effects model for murder rate, the shall variable has still reduced to a lower rate and is insignificant even at 10% level. Using Random Effects, If the shall law has been implemented, the murder rate decreases by 0.64%. The shall has no effect on murder rate using, random effects model.

Hausman Test:

phtest(model2.6, model2.7)

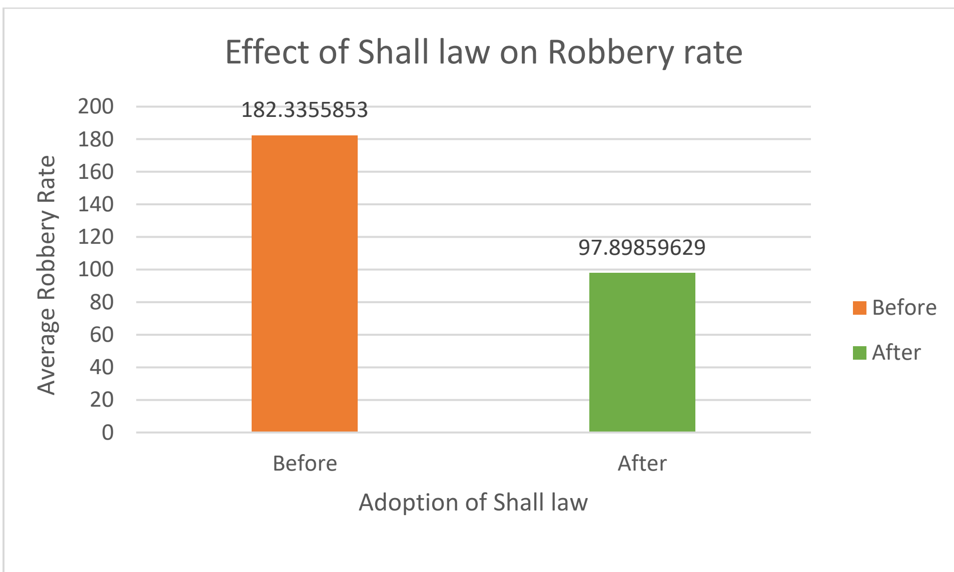
```
> phtest(model2.6, model2.7)

        Hausman Test

data:  log(mur) ~ shall + log(incarc_rate) + log(pb1064) + pw1064 +  ...
chisq = 189.59, df = 30, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent
```

According to Hausman test, the p-values is less 5%. So, we reject the null hypothesis that there is no endogeneity in the model.

5. Robbery Rate



By looking at the above chart we can say that the average robbery rate has decreased after Shall law was adopted. In order to verify the same, let us find out the regression results by developing models.

Simple Linear Regression

```
Call:
lm(formula = log(rob) ~ shall, data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-3.01675 -0.52148  0.05493  0.61216  2.52641 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.87305    0.03005  162.16   <2e-16 ***
shall       -0.77332    0.06096  -12.69   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8955 on 1171 degrees of freedom
Multiple R-squared:  0.1208,	Adjusted R-squared:  0.1201 
F-statistic: 160.9 on 1 and 1171 DF,  p-value: < 2.2e-16
```

We can see that the coefficient on Shall is highly significant (1% significance level), with an estimate of -0.77332. It can be interpreted that if shall law is introduced; the robbery rate reduces by 77.3%. This interpretation is in accordance with the conclusion from the chart above. For now, we can say that shall law reduces the robbery rate. Let us investigate further.

Multiple Linear Regression

```
Call:
lm(formula = log(rob) ~ shall + log(incarc_rate), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.27070 -0.48483  0.03145  0.52293  1.92305 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)        1.1407     0.1756   6.498  1.2e-10 ***
shall             -0.8268     0.0517 -15.992  < 2e-16 ***
log(incarc_rate)   0.7199     0.0335  21.489  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7586 on 1170 degrees of freedom
Multiple R-squared:  0.3696,	Adjusted R-squared:  0.3685 
F-statistic:   343 on 2 and 1170 DF,  p-value: < 2.2e-16
```

After adding incarceration rate to the mode, we can see that there is change in shall coefficient. We can interpret this as follows. When shall law is adopted, the robbery rate decreases by 82.7%. Thus, we can see that shall is highly significant variable and the effect of shall law adoption has caused the robbery rate to decrease further. This is again in accordance with our assumption that robbery rate should decrease with the adoption of shall law.

Simple Linear Regression

```
Call:
lm(formula = log(rob) ~ log(incarc_rate), data = dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.45081 -0.43911  0.03359  0.54745  2.12314 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       1.07405    0.19365   5.546  3.6e-08 ***
log(incarc_rate)  0.69408    0.03692  18.798  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.837 on 1171 degrees of freedom
Multiple R-squared:  0.2318,	Adjusted R-squared:  0.2312 
F-statistic: 353.4 on 1 and 1171 DF,  p-value: < 2.2e-16

> 
```

In the above model we check if incarceration rate has any effect on the robbery rate. From the results, we can see that incarceration rate is a very significant variable. For 1% percent increase in incarceration rate, the robbery rate increases by 0.69%. This result is inconsistent with our understanding that higher incarceration rate should decrease the robbery rate.

There is a serious potential for simultaneous causality bias. As per our understanding, the increased incarceration rate should reduce robbery rate but the model shows that increased incarceration rate causes increased robbery. Also, if there is more robbery, that will cause the police to increase the number of arrests and thus increase in incarceration rate. This can be handled by introducing Instrumental variable.

Introducing other explanatory variables in the model to avoid omitted variable bias.

Pooled Regression

```
Pooling Model

Call:
plm(formula = log(rob) ~ shall + log(incarc_rate) + log(pb1064) + 
    pw1064 + log(pop) + pm1029 + avginc + log(density), data = dat, 
    model = "pooling", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max. 
-1.471515 -0.277373 -0.010839  0.294331  1.641646 

Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)    
(Intercept)      -1.7484342  0.4056630 -4.3101 1.770e-05 ***
shall            -0.2725615  0.0361045 -7.5492 8.810e-14 ***
log(incarc_rate)  0.3995703  0.0367907 10.8606 < 2.2e-16 ***
log(pb1064)       0.3364677  0.0416579  8.0769 1.647e-15 ***
pw1064            0.0093288  0.0032875  2.8377  0.004623 ** 
log(pop)          0.2772829  0.0190109 14.5854 < 2.2e-16 ***
pm1029            0.1571636  0.0123714 12.7038 < 2.2e-16 ***
avginc            0.0754670  0.0070920 10.6411 < 2.2e-16 ***
log(density)      0.1851789  0.0125843 14.7151 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1068
Residual Sum of Squares: 261.19
R-Squared:      0.75545
Adj. R-Squared: 0.75377
F-statistic: 449.475 on 8 and 1164 DF, p-value: < 2.22e-16
```

After adding all the other explanatory variables, we can see that that coefficient for Shall variable decreased. It is still highly significant. The Pooled model states that, if shall law is adopted, the robbery rate decreases by 27.2%. All the other variables are also highly significant at 5% level.

Now, we need to check if any heteroskedasticity exists in the model. So, we perform a test to check heteroskedasticity.

```
> # heteroskedasicity
> lmtest::bptest(model3.4)  # Breusch-Pagan test

	studentized Breusch-Pagan test

data:  model3.4
BP = 87.945, df = 8, p-value = 1.214e-15

> |
```

According to the test, we see that p-values is less than 5%. This means that we will reject the null that the variance is constant(homoskedasticity) and thus, we can conclude that heteroskedasticity exists in the model.

In order to handle the heteroskedasticity, we need to carry out white robust corrected test for the pooled model.

White Corrected Standard Error Pooled model

```
Pooling Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = log(rob) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + log(pop) + pm1029 + avginc + log(density), data = dat,
    model = "pooling", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.471515 -0.277373 -0.010839  0.294331  1.641646

Coefficients:
                   Estimate Std. Error t-value  Pr(>|t|)
(Intercept)      -1.7484342  1.5574675 -1.1226 0.2618331
shall            -0.2725615  0.0782320 -3.4840 0.0005123 ***
log(incarc_rate)  0.3995703  0.1567815  2.5486 0.0109437 *
log(pb1064)       0.3364677  0.1410535  2.3854 0.0172202 *
pw1064            0.0093288  0.0110383  0.8451 0.3982138
log(pop)          0.2772829  0.0671835  4.1272 3.933e-05 ***
pm1029            0.1571636  0.0386325  4.0682 5.058e-05 ***
avginc            0.0754670  0.0195987  3.8506 0.0001242 ***
log(density)      0.1851789  0.0375388  4.9330 9.274e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1068
Residual Sum of Squares: 261.19
R-Squared:      0.75545
Adj. R-Squared: 0.75377
F-statistic: 47.3817 on 8 and 50 DF, p-value: < 2.22e-16
```

The white corrected model shows that the significance of the variable shall has decreased but it is still significant at 5% level. The variable pw1064 is not significant even at the 10% level after the white correction. Rest all the variables are significant. The main disadvantage of using the Pooled OLS model is that it does not take into account individual observed and unobserved characteristics. We will now run regression using the Fixed effect model as it accounts for observed and unobserved characteristics.

Fixed Effects Model

```
Oneway (individual) effect Within Model

Call:
plm(formula = log(rob) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + avginc + log(pop) + log(density) + pm1029, data = dat,
    model = "within", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.613936 -0.139740 -0.001763  0.141689  0.721649

Coefficients:
                   Estimate Std. Error t-value  Pr(>|t|)
shall             0.0160099  0.0244330  0.6553  0.512439
log(incarc_rate) -0.1105353  0.0364293 -3.0342  0.002467 **
log(pb1064)      -0.4110240  0.0725188 -5.6678 1.841e-08 ***
pw1064            0.0123486  0.0054886  2.2499  0.024651 *
avginc            0.0076718  0.0077733  0.9869  0.323886
log(pop)         -5.0145153  1.9456801 -2.5773  0.010087 *
log(density)      5.3733273  1.9682303  2.7300  0.006433 **
pm1029           -0.0227200  0.0109172 -2.0811  0.037651 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    53.526
Residual Sum of Squares: 49.901
R-Squared:      0.067732
Adj. R-Squared: 0.019194
F-statistic: 10.117 on 8 and 1114 DF, p-value: 1.0546e-13
```

From the Fixed Effects (Within) model we can see that the variables shall and avginc are not significant. According to this model shall law have no effect on Robbery rate. This also suggests that pooling model results were not consistent.

White Corrected Standard Error Fixed Effects model

```
Oneway (individual) effect Within Model

Note: Coefficient variance-covariance matrix supplied: vcovHC

Call:
plm(formula = log(rob) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + avginc + log(pop) + log(density) + pm1029, data = dat,
    model = "within", index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.613936 -0.139740 -0.001763  0.141689  0.721649

Coefficients:
                   Estimate Std. Error t-value Pr(>|t|)
shall             0.0160099  0.0539380  0.2968   0.7667
log(incarc_rate) -0.1105353  0.0940568 -1.1752   0.2402
log(pb1064)      -0.4110240  0.2558320 -1.6066   0.1084
pw1064            0.0123486  0.0145703  0.8475   0.3969
avginc            0.0076718  0.0212395  0.3612   0.7180
log(pop)         -5.0145153  4.7118409 -1.0642   0.2875
log(density)      5.3733273  4.7143425  1.1398   0.2546
pm1029           -0.0227200  0.0329331 -0.6899   0.4904

Total Sum of Squares:    53.526
Residual Sum of Squares: 49.901
R-Squared:      0.067732
Adj. R-Squared: 0.019194
F-statistic: 2.45047 on 8 and 50 DF, p-value: 0.0254
>
```

The white corrected Fixed effects model states that none of the variables are significant. This could be because Shall being a law changes over time and not across entities. We have not factored variables which vary over time but not across States into this model. Therefore, we will now model the data using time and entity fixed effects.

Time and Entity Fixed Effects with White Corrected Standard Error

```
> # MODEL 6: Time and Entity Fixed Effects
> model3.6<- plm(log(rob)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+log(pop)+log(density)+pm1029+factor(year),data=da
t,index=c("stateid","year"), model ="within")
> coeftest(model3.6, vcovHC)

t test of coefficients:

                  Estimate Std. Error t value  Pr(>|t|)    
shall            -0.021051   0.041849 -0.5030 0.6150535    
log(incarc_rate) -0.232016   0.094826 -2.4467 0.0145719 *  
log(pb1064)      -0.712148   0.237573 -2.9976 0.0027827 ** 
pw1064           -0.021410   0.016521 -1.2959 0.1952826    
avginc            0.023306   0.020491  1.1374 0.2556184    
log(pop)         -1.114633   3.497543 -0.3187 0.7500222    
log(density)      1.156277   3.540395  0.3266 0.7440364    
pm1029            0.128500   0.045393  2.8308 0.0047280 ** 
factor(year)78    0.054074   0.022278  2.4272 0.0153753 *  
factor(year)79    0.192001   0.032583  5.8927 5.055e-09 ***
factor(year)80    0.329828   0.041795  7.8915 7.220e-15 ***
factor(year)81    0.386755   0.043199  8.9529 < 2.2e-16 ***
factor(year)82    0.374309   0.054723  6.8400 1.316e-11 ***
factor(year)83    0.319285   0.075861  4.2088 2.778e-05 ***
factor(year)84    0.301736   0.087851  3.4346 0.0006157 ***
factor(year)85    0.362630   0.101359  3.5777 0.0003618 ***
factor(year)86    0.470437   0.118776  3.9607 7.957e-05 ***
factor(year)87    0.468935   0.128820  3.6402 0.0002852 ***
factor(year)88    0.535067   0.146241  3.6588 0.0002655 ***
factor(year)89    0.621347   0.162425  3.8254 0.0001379 ***
factor(year)90    0.782594   0.184422  4.2435 2.387e-05 ***
factor(year)91    0.931012   0.192998  4.8240 1.607e-06 ***
factor(year)92    0.955839   0.204156  4.6819 3.199e-06 ***
factor(year)93    0.997233   0.216681  4.6023 4.668e-06 ***
factor(year)94    1.034095   0.226175  4.5721 5.380e-06 ***
factor(year)95    1.065701   0.235212  4.5308 6.522e-06 ***
factor(year)96    1.036428   0.247633  4.1853 3.076e-05 ***
factor(year)97    0.991254   0.255250  3.8835 0.0001092 ***
factor(year)98    0.914865   0.262173  3.4896 0.0005031 ***
factor(year)99    0.858468   0.270623  3.1722 0.0015551 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
```

Comparing with the previous models, we can see that the coefficient of Shall has decreased to -0.021051 indicating that shall law implementation reduces robbery rate by 2.1%. This could be due to the effect of time. However, the variable shall is not significant. This states that shall law has no effect on robbery rate.

We cannot use Random Effect model because we are considering entire population in the data set (51 states) and not a sample from population. We can prove this by building a Random effects model and using Hausman test to verify which model should be used.

Random Effects model with White Corrected Standard Error

```
> # MODEL 7: Random Effects
> model3.7<- plm(log(rob)~shall+log(incarc_rate)+log(pb1064)+pw1064+avginc+log(pop)+log(density)+pm1029+factor(year),data=da
t,index=c("stateid","year"), model ="random")
> coeftest(model3.7, vcovHC)

t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)    
(Intercept)       4.4070461  1.2383427  3.5588 0.0003878 ***
shall            -0.0061118  0.0473557 -0.1291 0.8973324    
log(incarc_rate) -0.0828983  0.1028958 -0.8057 0.4206109    
log(pb1064)      -0.1885567  0.1616324 -1.1666 0.2436247    
pw1064           -0.0259455  0.0139604 -1.8585 0.0633537 .  
avginc            0.0315649  0.0191544  1.6479 0.0996439 .  
log(pop)          0.3181021  0.1434270  2.2179 0.0267600 *  
log(density)      0.1925171  0.0917962  2.0972 0.0361932 *  
pm1029            0.1236346  0.0472835  2.6148 0.0090468 ** 
factor(year)78    0.0277960  0.0214555  1.2955 0.1954023    
factor(year)79    0.1389621  0.0290016  4.7915 1.873e-06 ***
factor(year)80    0.2545896  0.0369516  6.8898 9.213e-12 ***
factor(year)81    0.2886707  0.0366086  7.8853 7.290e-15 ***
factor(year)82    0.2431774  0.0475133  5.1181 3.619e-07 ***
factor(year)83    0.1542719  0.0674764  2.2863 0.0224186 *  
factor(year)84    0.1081190  0.0806180  1.3411 0.1801459    
factor(year)85    0.1437539  0.0970167  1.4817 0.1386842    
factor(year)86    0.2238205  0.1156080  1.9360 0.0531107 .  
factor(year)87    0.1964872  0.1268556  1.5489 0.1216819    
factor(year)88    0.2359120  0.1409980  1.6732 0.0945699 .  
factor(year)89    0.2950091  0.1608740  1.8338 0.0669454 .  
factor(year)90    0.4339347  0.1769540  2.4522 0.0143456 *  
factor(year)91    0.5605461  0.1854671  3.0223 0.0025641 ** 
factor(year)92    0.5581886  0.1944170  2.8711 0.0041661 ** 
factor(year)93    0.5774821  0.2039005  2.8322 0.0047045 ** 
factor(year)94    0.5924907  0.2108382  2.8102 0.0050361 ** 
factor(year)95    0.5987253  0.2153081  2.7808 0.0055119 ** 
factor(year)96    0.5442942  0.2238059  2.4320 0.0151689 *  
factor(year)97    0.4760208  0.2319655  2.0521 0.0403857 *  
factor(year)98    0.3745132  0.2355865  1.5897 0.1121781    
factor(year)99    0.2964981  0.2445826  1.2123 0.2256630    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hausman test

```
> #Hausman Test for Endogeneity : One model is inconsistent. Since our data is not randomly sampled hence we reject random e
ffects model
> phtest(model3.6, model3.7)

	Hausman Test

data:  log(rob) ~ shall + log(incarc_rate) + log(pb1064) + pw1064 +  ...
chisq = 94.443, df = 30, p-value = 1.372e-08
alternative hypothesis: one model is inconsistent
```

According to Hausman test, the p-values is less 5%. So, we reject the null hypothesis that there is no endogeneity. As a result, we should use the Entity and Time effects model. Thus, the Entity and Time effects model is the best model for our panel data.

```
> summary(model3.6)
Oneway (individual) effect Within Model

Call:
plm(formula = log(rob) ~ shall + log(incarc_rate) + log(pb1064) +
    pw1064 + avginc + log(pop) + log(density) + pm1029 + factor(year),
    data = dat, model = "within", index = c("stateid", "year"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.6247414 -0.1168678  0.0045862  0.1110545  0.6015440

Coefficients:
                   Estimate Std. Error  t-value  Pr(>|t|)
shall            -0.0210508  0.0227838  -0.9239   0.35572
log(incarc_rate) -0.2320160  0.0363334  -6.3857 2.518e-10 ***
log(pb1064)      -0.7121481  0.0658135 -10.8207 < 2.2e-16 ***
pw1064           -0.0214100  0.0054434  -3.9332 8.910e-05 ***
avginc            0.0233062  0.0081327   2.8657   0.00424 **
log(pop)         -1.1146335  1.7233922  -0.6468   0.51792
log(density)      1.1562771  1.7436094   0.6632   0.50737
pm1029            0.1285002  0.0152100   8.4484 < 2.2e-16 ***
factor(year)78    0.0540739  0.0361921   1.4941   0.13544
factor(year)79    0.1920013  0.0367318   5.2271 2.062e-07 ***
factor(year)80    0.3298278  0.0372502   8.8544 < 2.2e-16 ***
factor(year)81    0.3867550  0.0383334  10.0892 < 2.2e-16 ***
factor(year)82    0.3743093  0.0406308   9.2124 < 2.2e-16 ***
factor(year)83    0.3192852  0.0437751   7.2938 5.781e-13 ***
factor(year)84    0.3017361  0.0474782   6.3553 3.050e-10 ***
factor(year)85    0.3626300  0.0512694   7.0730 2.703e-12 ***
factor(year)86    0.4704366  0.0558958   8.4163 < 2.2e-16 ***
factor(year)87    0.4689345  0.0604325   7.7596 1.948e-14 ***
factor(year)88    0.5350668  0.0652639   8.1985 6.772e-16 ***
factor(year)89    0.6213473  0.0698576   8.8945 < 2.2e-16 ***
factor(year)90    0.7825938  0.0764722  10.2337 < 2.2e-16 ***
factor(year)91    0.9310118  0.0802118  11.6069 < 2.2e-16 ***
factor(year)92    0.9558388  0.0845214  11.3088 < 2.2e-16 ***
factor(year)93    0.9972329  0.0876207  11.3812 < 2.2e-16 ***
factor(year)94    1.0340954  0.0911943  11.3395 < 2.2e-16 ***
factor(year)95    1.0657013  0.0950117  11.2165 < 2.2e-16 ***
factor(year)96    1.0364282  0.0987396  10.4966 < 2.2e-16 ***
factor(year)97    0.9912539  0.1021135   9.7074 < 2.2e-16 ***
factor(year)98    0.9148654  0.1058143   8.6460 < 2.2e-16 ***
factor(year)99    0.8584677  0.1087337   7.8951 7.024e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    53.526
Residual Sum of Squares: 35.684
R-Squared:      0.33333
Adj. R-Squared: 0.28449
F-statistic: 18.1996 on 30 and 1092 DF, p-value: < 2.22e-16
```

Also, looking at the F-statistics and p-value of the Entity and Time Fixed effects model, we can say that it is a good fit for our panel data

6. Conclusion

We need to consider both state fixed effects and time fixed effects that accounts for omitted variable bias, unobserved heterogeneity and time effects in our panel data. The Entity and Time effect model is the best model for our panel data.

We conclude based on the above results and interpretations that there is no significant effect of shall-law on the violent crime rate, murder rate and robbery rate.

Also, Incarceration Rate causes endogeneity in the model because of simultaneous causality bias as discussed earlier. So we can remove these endogeneity by introducing an appropriate Instrumental variables in the model.

We cannot use Random Effects model since the data we have is not random and consists of data from 50 US states + the District of Columbia. Random effects model, hence cannot be used to solve our problem.