BUAN 6312.002

Applied Econometrics and Time Series Analysis

Group 4

Members:

Parth Hitesh Shah - PXS220058

Jay Patel - JAP210006

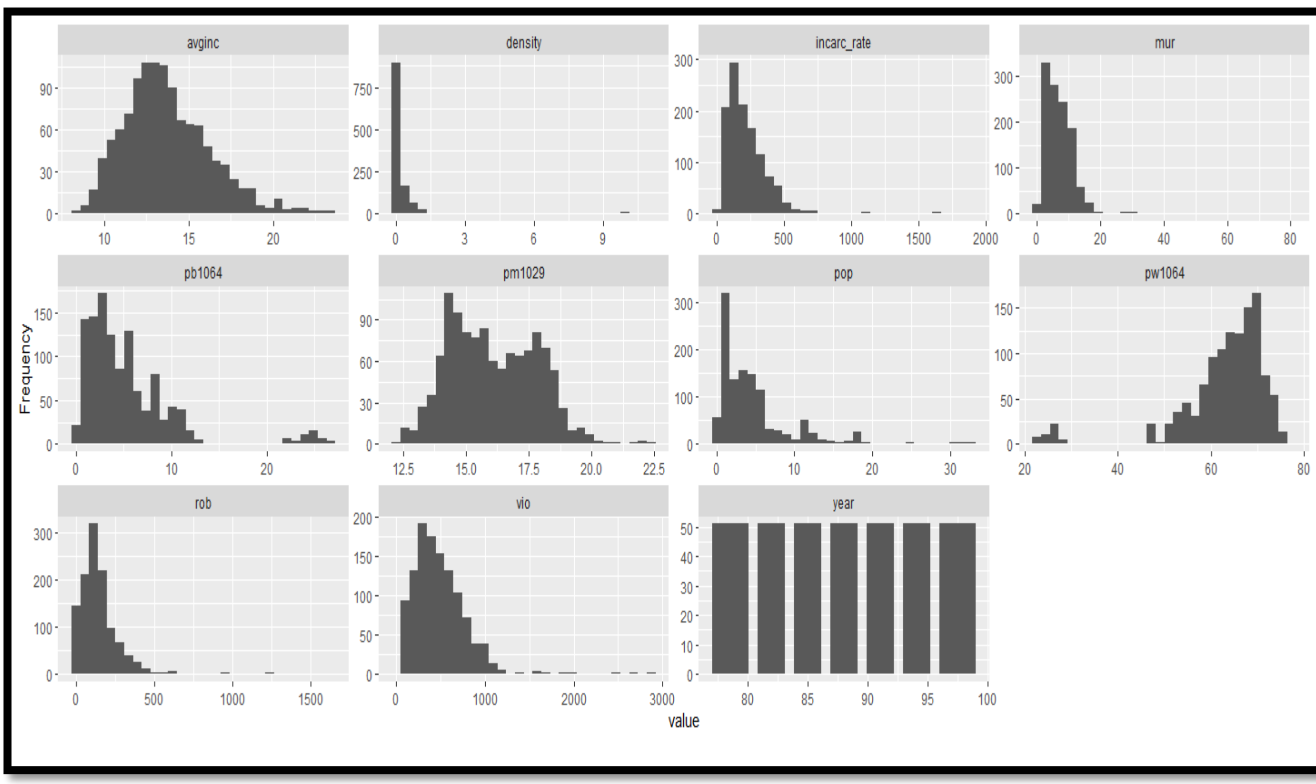
Mayur Kumar Tikmani - MXT210030

Sai Venkata Ramanamurty Yedavilli - SXY220046

Exploratory Data Analysis

- Our data has a total of 1137 observations, and the data description shows 13 columns.
- All the columns are numeric. Year, Incarc_rate, State ID, and Shall are the integer type, and the rest of the float types.
- As the state id and shall are not numeric variables, we have converted them to categorical variables. State id columns have 51 categories and shall have two categories of 1 or 0.
- We have converted the date as the ordinal variable and labeled it as the "year1" column.
- Our data has no missing value in any of the columns.

Univariate analysis



Year:

- We have the data for each year, starting from the year 1977 to the year 1999, which is 23 years. Each year has the same number of observations.

Violent crime rate:

- The violent crime rate ranges from a minimum of 47 crimes per 100,000 to a maximum of 2922 crimes per 100,000. The mean crime across the states and years is 503 crimes per 100,000, and the median is 443 crimes per 100,000.
- The data is left-skewed, and 75% of the data fall under the vio rate of less than 651 crimes per 100,000.

Murder Rate:

- The murder rate ranges from 0.2 to 80.6 per 100,000 people across the state and the year.
- The data is highly leftward skewed, and the median murder per 100,000 across the state and the year is 6.4, while the mean for the same is 7.6.

Robbery Rate:

- The Robbery Rates very from the 6.4 robbery per 100,000 population in one year across the years and states to the 1635.1 robbery per 100000.
- The data is highly left skewed and the we can see that the median robbery per 100000 is the 124.1 and the mean robbery per 100000 is the 124.1.

Incarceration rate:

- It is the rate of the people per 100,000 who were sentenced in the previous year. The vales varies between the range of the 19 people to the 1913 people per 100,000 residents across the state and year.
- The data is the left skewed and the median incarnation rate is the 187, while the mean incarnation rate is the 226.6 people out of the 100,000 residents.

PB1064

- The percentage of the state population is black between the ages of the 10 to 64 varies in the range of the 0.24% to the 26.97%.
- The median percentage population, which belongs to the African-American community of between the age of the 10 to 64 is the 4.02% and the mean for the same is 5.33%
- The data is the left skewed.

PW1064

- It has the approximately apposite distribution than the PB1064.
- The data is rightward skewed and the minimum white population percentage between the age of 10 to 64 is the 21.76%, while the maximum is the 76.53%.
- The median and mean for the same is 65% and 62% respectively.

PM1029

- The male population between the ages of 10 to 29 years percentagewise varies from the 12.21 % to the 22.25%.
- The median percentage of the male between the age of the 10 to 29 are the 15.90 and mean for the same is the 16.08%.

Population

- State population varies from the 0.4027 million to the 33.14 million Average population among the state the years. The median population was 3.2713 million across the state and years and the mean population was the 4.81 million.

Average Income

- The average per capita income among the state across the years varies from the 8.5 thousand to the 23,64 thousand. The mean per capita income across the state and year was the 13.72 thousand and the median was the 13.40 thousand dollars.
- The graph shows that the distribution is approximately follows the normal distribution and not near to symmetric.

Density

- Population density among the states varies from the 0.707 per square mile to the 11102.11 people per square mile across the state and across the years.
- The median populations density is the 81.5 people and the mean density is the 352 people per states.
- Data is very highly left skewed.

State ID

- Each state has the unique ID, and the data for the 23 years. There are 51 states.

Shall

- Whether the state has the shall low or not is reflected by the 0 and 1 in the data. If the state has the shall low in that particular year has been labeled as the 1 and the state has not, then it is labeled as the 0.
- The number of observations with the shall is equal to 1 are 285 and the number of the shall is equal to 0 are the 888.

Descriptive Analysis with District of Columbia

```
> data.frame(mean = sapply(guns, mean, na.rm = TRUE)
+            , median = sapply(guns, median, na.rm = TRUE)
+            , min = sapply(guns, min, na.rm = TRUE)
+            , max = sapply(guns, max, na.rm = TRUE)
+            , sd = sapply(guns, sd, na.rm = TRUE)
+            , miss.val = sapply(guns, function(x)
+              sum(is.na(x))))
                  mean       median          min        max          sd miss.val
year        88.0000000  88.00000000 7.700000e+01   99.00000   6.6360789        0
vio        503.0746806 443.00000000 4.700000e+01 2921.80005 334.2771946        0
mur          7.6651321   6.40000010 2.000000e-01   80.60000   7.5227099        0
rob        161.8202044 124.09999847 6.400000e+00 1635.09998 170.5099609        0
incarc_rate 226.5797101 187.00000000 1.900000e+01 1913.00000 178.8880945        0
pb1064       5.3362169   4.02621317 2.482066e-01   26.97957   4.8856876        0
pw1064      62.9454322  65.06127930 2.178043e+01   76.52575   9.7615273        0
pm1029      16.0811272  15.89516926 1.221368e+01   22.35269   1.7321433        0
pop          4.8163414   3.27133203 4.027530e-01   33.14512   5.2521152        0
avginc      13.7247960  13.40155125 8.554884e+00   23.64671   2.5545425        0
density      0.3520382   0.08156896 7.070804e-04   11.10212   1.3554718        0
stateid     28.9607843  29.00000000 1.000000e+00   56.00000  15.6835220        0
shall        0.2429668   0.00000000 0.000000e+00    1.00000   0.4290581        0
```

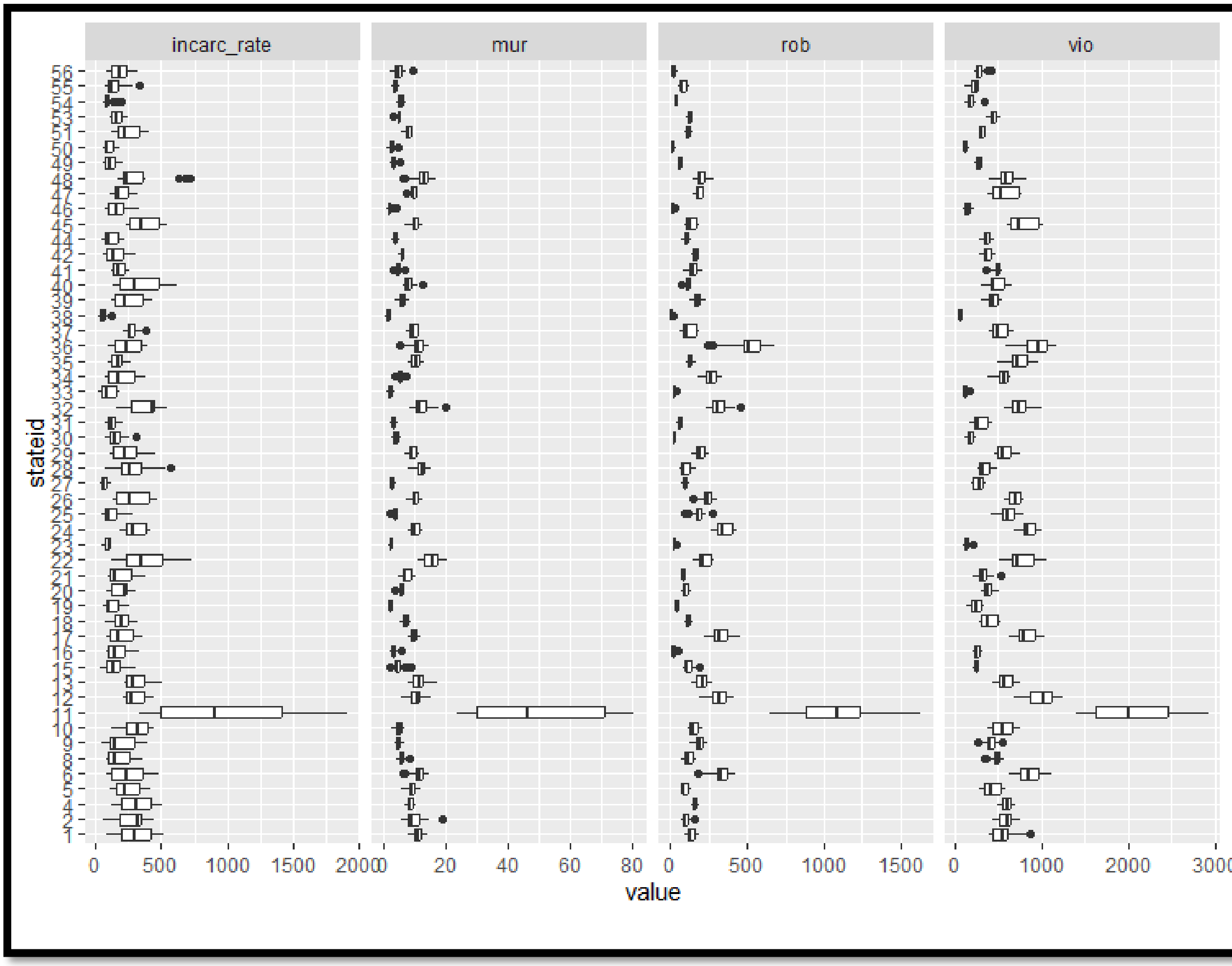
Descriptive Analysis without District of Columbia

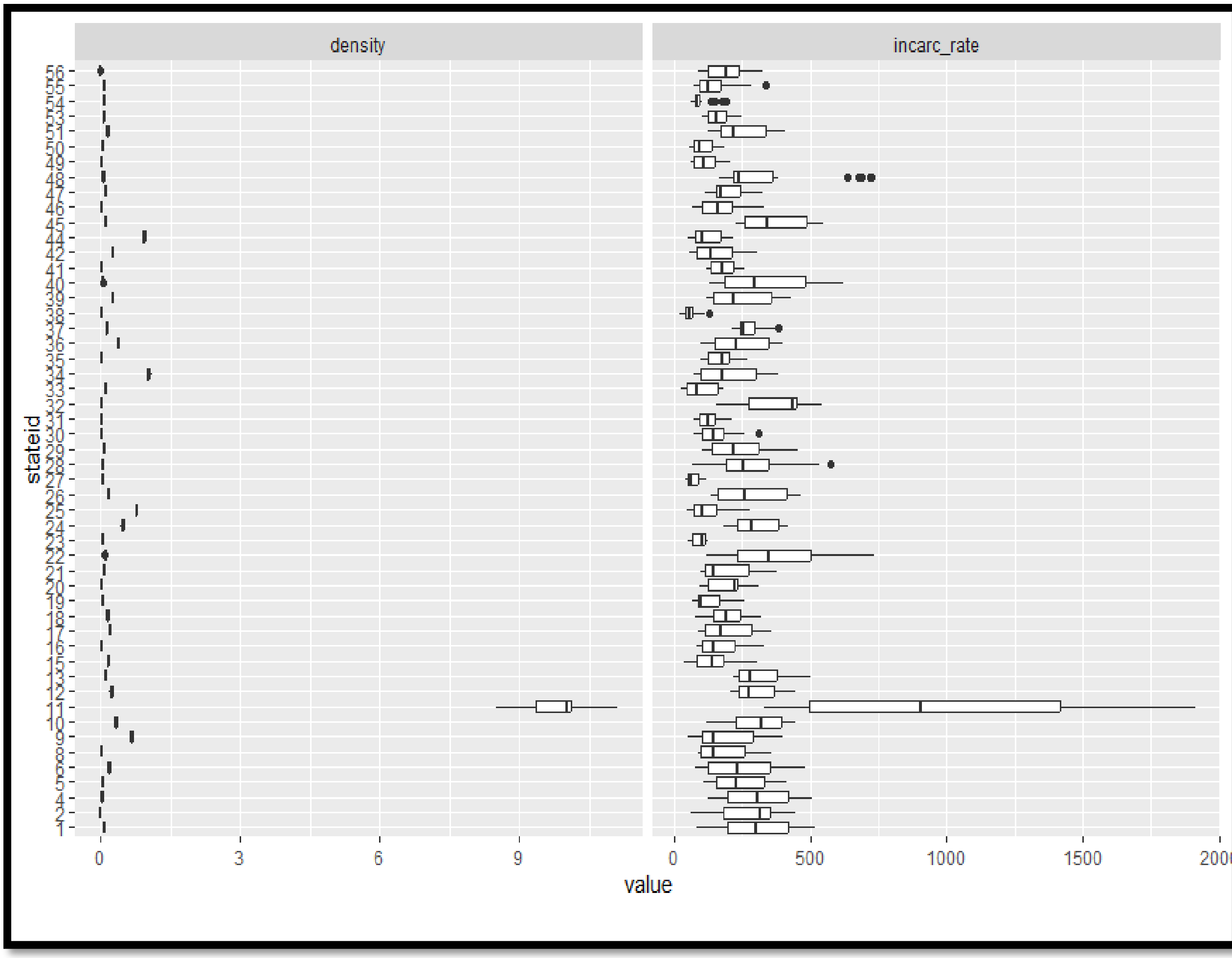
```
> data.frame(mean = sapply(df11, mean, na.rm = TRUE)
+            , median = sapply(df11, median, na.rm = TRUE)
+            , min = sapply(df11, min, na.rm = TRUE)
+            , max = sapply(df11, max, na.rm = TRUE)
+            , sd = sapply(df11, sd, na.rm = TRUE)
+            , miss.val = sapply(df11, function(x)
+              sum(is.na(x))))
                  mean       median          min         max          sd miss.val
year        88.0000000  88.00000000 7.700000e+01   99.000000   6.6361355        0
vio        472.1566089 437.50000000 4.700000e+01 1244.300049 247.0173349        0
mur          6.8329565   6.30000019 2.000000e-01   20.299999   3.7541113        0
rob        143.6603477 122.40000153 6.400000e+00  684.000000 107.4703019        0
incarc_rate 211.4939130 183.00000000 1.900000e+01  736.000000 125.4978786        0
pb1064       4.9652110   3.96507454 2.482066e-01   26.979570   4.1571277        0
pw1064      63.7069907  65.33569717 2.351738e+01   76.525749   8.2160921        0
pm1029      16.0912845  15.88927507 1.250534e+01   22.352686   1.7258659        0
pop          4.9005699   3.28884804 4.027530e-01   33.145123   5.2701721        0
avginc      13.6228364  13.31595230 8.554884e+00   23.646713   2.4407813        0
density      0.1636141   0.07958012 7.070804e-04    1.097643   0.2287866        0
stateid     29.3200000  29.50000000 1.000000e+00   56.000000  15.6304205        0
shall        0.2478261   0.00000000 0.000000e+00    1.000000   0.4319381        0
```

- As we can see there is a huge difference in the maximum values of violent crime rate, murder rate, robbery rate and incarceration rate. The state District of Columbia has been removed in our further analysis as it acts as an outlier.
- Performing the descriptive analysis without the District of Columbia also helps the standard deviation. The standard deviation is now lower as compared to the standard deviation with District of Columbia.

Multivariate analysis

To find out the reason behind the outliner or skewness in the data, we created graphs of all the variables per state id.





As we can see form the both of the above graphs that for the state id = 11, Washington dc, the Murder rate, violation rate and robbery rate is very high and cause skewness and outliners. Also, For the density and the incarnation rate, for the state id 11 has the outliners and very high values.

We are thinking that the outline effect for the national capital (state id = 11) is coming because of it is city state, small in size and high population density. To avoid the wrong analysis, we have decided to remove the state no 11 from the further analysis.

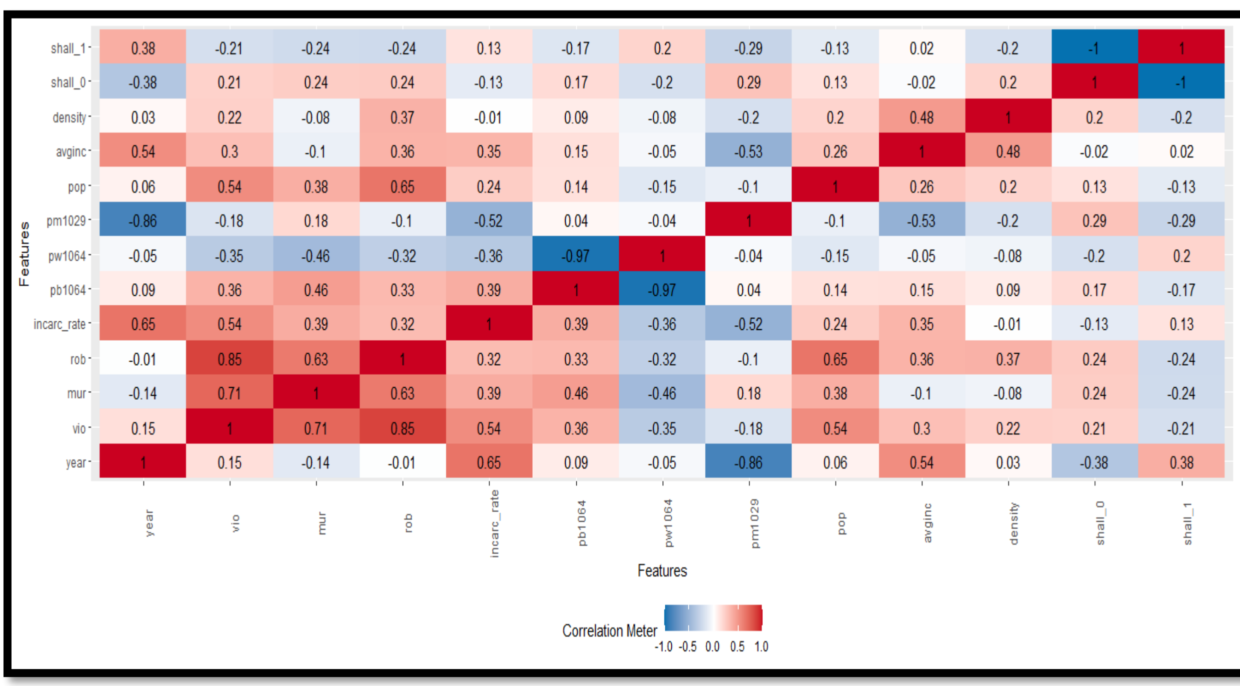
Correlation Matrix

Correlation Matrix with District of Columbia

```
> round(cor(guns),2)
             year   vio   mur   rob incarc_rate pb1064 pw1064 pm1029   pop avginc density stateid shall
year         1.00  0.12 -0.03 -0.01        0.50   0.07  -0.03  -0.87  0.06   0.53    0.00    0.00  0.38
vio          0.12  1.00  0.83  0.91        0.70   0.57  -0.57  -0.17  0.32   0.41    0.66   -0.32 -0.21
mur         -0.03  0.83  1.00  0.80        0.71   0.60  -0.62   0.01  0.10   0.22    0.75   -0.24 -0.18
rob         -0.01  0.91  0.80  1.00        0.57   0.58  -0.58  -0.09  0.32   0.41    0.78   -0.25 -0.21
incarc_rate  0.50  0.70  0.71  0.57        1.00   0.53  -0.53  -0.45  0.10   0.46    0.56   -0.22  0.04
pb1064       0.07  0.57  0.60  0.58        0.53   1.00  -0.98   0.02  0.06   0.26    0.54   -0.31 -0.18
pw1064      -0.03 -0.57 -0.62 -0.58       -0.53  -0.98   1.00  -0.01 -0.07  -0.19   -0.56    0.31  0.21
pm1029      -0.87 -0.17  0.01 -0.09       -0.45   0.02  -0.01   1.00 -0.10  -0.53   -0.06    0.01 -0.28
pop          0.06  0.32  0.10  0.32        0.10   0.06  -0.07  -0.10  1.00   0.22   -0.08   -0.06 -0.12
avginc       0.53  0.41  0.22  0.41        0.46   0.26  -0.19  -0.53  0.22   1.00    0.34   -0.20  0.00
density      0.00  0.66  0.75  0.78        0.56   0.54  -0.56  -0.06 -0.08   0.34    1.00   -0.16 -0.11
stateid      0.00 -0.32 -0.24 -0.25       -0.22  -0.31   0.31   0.01 -0.06  -0.20   -0.16    1.00  0.19
shall        0.38 -0.21 -0.18 -0.21        0.04  -0.18   0.21  -0.28 -0.12   0.00   -0.11    0.19  1.00
```

Correlation Matrix without District of Columbia

```
> round(cor(guns_non_DC),2)
             year   vio   mur   rob incarc_rate pb1064 pw1064 pm1029   pop avginc density stateid shall
year         1.00  0.15 -0.14 -0.01        0.65   0.09  -0.05  -0.86  0.06   0.54    0.03    0.00  0.38
vio          0.15  1.00  0.71  0.85        0.54   0.36  -0.35  -0.18  0.54   0.30    0.22   -0.29 -0.21
mur         -0.14  0.71  1.00  0.63        0.39   0.46  -0.46   0.18  0.38  -0.10   -0.08   -0.24 -0.24
rob         -0.01  0.85  0.63  1.00        0.32   0.33  -0.32  -0.10  0.65   0.36    0.37   -0.21 -0.24
incarc_rate  0.65  0.54  0.39  0.32        1.00   0.39  -0.36  -0.52  0.24   0.35   -0.01   -0.18  0.13
pb1064       0.09  0.36  0.46  0.33        0.39   1.00  -0.97   0.04  0.14   0.15    0.09   -0.27 -0.17
pw1064      -0.05 -0.35 -0.46 -0.32       -0.36  -0.97   1.00  -0.04 -0.15  -0.05   -0.08    0.27  0.20
pm1029      -0.86 -0.18  0.18 -0.10       -0.52   0.04  -0.04   1.00 -0.10  -0.53   -0.20    0.00 -0.29
pop          0.06  0.54  0.38  0.65        0.24   0.14  -0.15  -0.10  1.00   0.26    0.20   -0.08 -0.13
avginc       0.54  0.30 -0.10  0.36        0.35   0.15  -0.05  -0.53  0.26   1.00    0.48   -0.17  0.02
density      0.03  0.22 -0.08  0.37       -0.01   0.09  -0.08  -0.20  0.20   0.48    1.00   -0.03 -0.20
stateid      0.00 -0.29 -0.24 -0.21       -0.18  -0.27   0.27   0.00 -0.08  -0.17   -0.03    1.00  0.18
shall        0.38 -0.21 -0.24 -0.24        0.13  -0.17   0.20  -0.29 -0.13   0.02   -0.20    0.18  1.00
```



Violent crime rate, Murder Rate and Robbery Rate

- Violent Crime Rate is positively correlated to both murder rate and robbery rate as it is close to 1.
- The correlation between violent crime rate and murder rate is 0.71 and correlation between violent crime rate and robbery rate is 0.85 and this is without the state Washington DC.

Violent Crime Rate with other variables

- Apart from murder rate and robbery rate, violent crime rate is moderately correlated with variables incarnation rate and population with 0.54 correlation.

Murder Rate with other variables

- Murder Rate is positively correlated to robbery rate with a value of 0.63.
- Apart from that we don’t see a linear trend of murder rate with any other variable.

Robbery Rate with other variables

- Apart from the violent rate and murder rate where robbery is highly correlated. There is also a variable population which is highly correlated with population that means that as the population increases the robbery rate is bound to increase.

Incarceration Rate and Average Income with Year

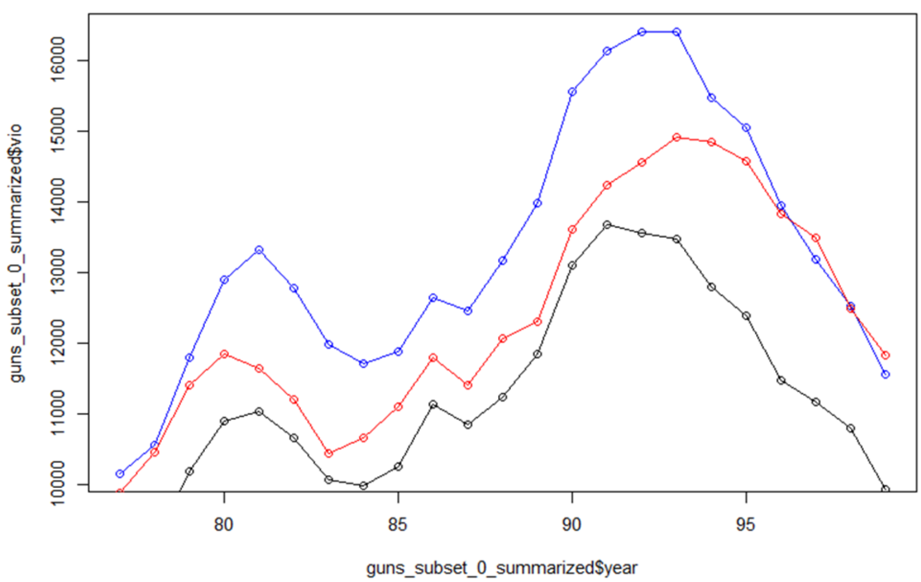
- This is an obvious positive correlation (0.65) as with year passing the prisoners that get sentenced is bound to increase.
- The same with average income, the per capita personal income in the state increases with year

PB1064 and PW1064

- If we compare the correlation of percentage of state population that is white with the percentage of state population that is black, we can see that there is a high correlation of violent crime rate, murder rate, robbery rate and incarceration rate in states where the percentage of African American is more as compared to the white population.

The average violence rate was lower for the years before the Shall laws were passed.

The above graph represents yearly violence. Based on this, we can see that violence increased from 90s, but it is difficult to attribute to Shall laws as there's a rise in states where Shall laws have not been passed.
Blue - States with no Shall laws
Red - States with Shall laws
Black - States with no Shall laws barring Washington DC.



Average violence rates before and after Gun Laws were passed at state level

	Violence rate	
State ID/Gun Laws	**0**	**1**
12	893.5	1096.2
22	704.8	916.6
45	783.3	913.5
47	533.9	749.0
32	765.4	705.9
2	569.0	697.0
13	543.9	662.8
4	604.9	619.6
37	515.3	579.1
48	618.2	575.8
40	479.1	551.1
5	427.2	491.6
41	510.8	473.9
28	302.0	438.7
42	344.4	434.3
51	321.1	331.7
21	339.8	300.5
49	273.8	291.0
16	252.0	279.2
56	298.5	247.8
54	162.1	223.4
46	138.5	169.5
30	192.4	166.8
23	204.8	140.0
38	59.2	73.7

Takeaway – For most states, average violence rate is higher post gun laws were passed.

Data Modelling

- To model the data, we must find the dependent and independent variables. As per the problem statement, we must determine whether the shall-issue law reduces crime or not?
- In That case, The hypothesis here will be
    - Null H_0: $\beta_{Shall-issue\ Laws} \geq 0$
    - Alternative H_1: $\beta_{Shall-issue\ Laws} < 0$
- In our case, the dependent variable can be the either violent crime rate, murder rate or the robbery rate, but as shown in the multivariate analysis the correlation among them is very high, and we cannot use the three dependent variables and that's why we have created a new variable "crime rate", which is simply the sum of the murder rate, robbery rate and the violent crime rate.
- crime rate = violent crime rate + murder rate + robbery rate

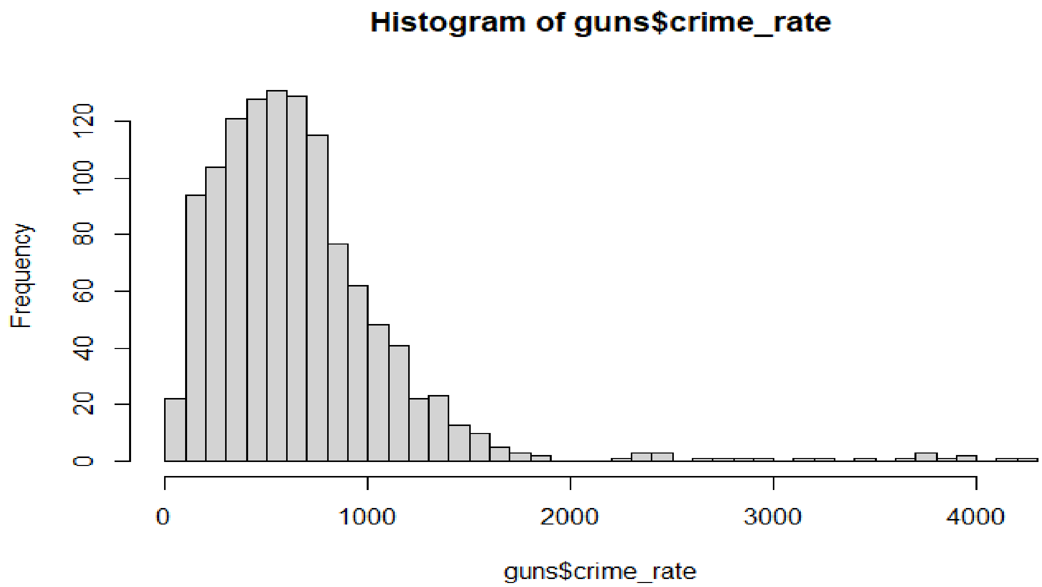
The first model we build was using all the variables, as below

- **crime_rate = year+incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density + shall + stated**
- As the state id is the categorical variable, we used the dummy variable for it using the state id =1 as base.
- The regression model has a very high adjusted R-square of 90%.
- Next page has the regression output.

```
> model1 <- lm(crime_rate ~ year+ incarc_rate + pb1064 + pw1064
+              + avginc + density + shall + stateid, data = df1
> summary(model1)

Call:
lm(formula = crime_rate ~ year + incarc_rate + pb1064 + pw1064 +
    pm1029 + pop + avginc + density + shall + stateid, data = df

Residuals:
     Min       1Q   Median       3Q      Max
-1163.27   -62.06    -1.22    54.77  1002.50

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.627e+02  3.865e+02   1.197 0.231447
year        -7.283e+00  3.763e+00  -1.935 0.053191 .
incarc_rate  8.881e-03  8.588e-02   0.103 0.917656
pb1064       5.489e+01  1.888e+01   2.908 0.003710 **
pw1064       2.247e+01  5.630e+00   3.991 7.00e-05 ***
pm1029      -5.130e+01  1.180e+01  -4.348 1.50e-05 ***
pop          1.733e+01  7.794e+00   2.223 0.026397 *
avginc      -6.286e+00  5.743e+00  -1.094 0.273992
density     -2.103e+02  7.541e+01  -2.788 0.005389 **
shall1      -3.882e+00  1.708e+01  -0.227 0.820277
stateid2     8.032e+01  6.446e+01   1.246 0.213015
stateid4     1.382e+02  8.000e+01   1.728 0.084340 .
stateid5    -1.011e+02  6.066e+01  -1.666 0.095958 .
stateid6     1.643e+02  2.057e+02   0.799 0.424594
stateid8    -7.321e+01  9.938e+01  -0.737 0.461468
stateid9     6.356e+01  1.198e+02   0.530 0.595874
stateid10    1.319e+02  6.811e+01   1.937 0.053051 .
stateid11    4.429e+03  7.075e+02   6.261 5.45e-10 ***
stateid12    5.347e+02  1.016e+02   5.265 1.68e-07 ***
stateid13    6.676e+01  4.919e+01   1.357 0.174974
stateid15   -4.113e+02  2.377e+02  -1.730 0.083834 .
stateid16   -2.664e+02  1.144e+02  -2.329 0.020052 *
stateid17    3.972e+02  9.516e+01   4.174 3.22e-05 ***
stateid18   -1.456e+02  9.844e+01  -1.479 0.139444
stateid19   -3.096e+02  1.162e+02  -2.664 0.007832 **
stateid20   -1.117e+02  9.217e+01  -1.212 0.225748
stateid21   -2.281e+02  9.491e+01  -2.403 0.016406 *
stateid21   -2.281e+02  9.491e+01  -2.403 0.016406 *
stateid22    3.418e+02  5.057e+01   6.759 2.24e-11 ***
stateid23   -4.494e+02  1.196e+02  -3.759 0.000179 ***
stateid24    5.251e+02  6.196e+01   8.475  < 2e-16 ***
stateid25    2.439e+02  1.315e+02   1.855 0.063844 .
stateid26    1.916e+02  8.759e+01   2.188 0.028891 *
stateid27   -2.648e+02  1.122e+02  -2.361 0.018393 *
stateid28   -1.800e+02  6.833e+01  -2.635 0.008539 **
stateid29    9.832e+01  8.178e+01   1.202 0.229551
stateid30   -4.166e+02  9.550e+01  -4.362 1.41e-05 ***
stateid31   -2.208e+02  1.047e+02  -2.108 0.035228 *
stateid32    3.964e+02  7.676e+01   5.164 2.86e-07 ***
stateid33   -4.538e+02  1.241e+02  -3.656 0.000268 ***
stateid34    2.280e+02  1.158e+02   1.969 0.049176 *
stateid35    2.596e+02  7.202e+01   3.604 0.000327 ***
stateid36    5.690e+02  1.334e+02   4.267 2.15e-05 ***
stateid37   -7.944e+01  4.892e+01  -1.624 0.104716
stateid38   -4.495e+02  1.070e+02  -4.199 2.89e-05 ***
stateid39   -1.314e+02  1.110e+02  -1.184 0.236653
stateid40   -5.244e+01  6.102e+01  -0.859 0.390373
stateid41   -2.301e+01  1.024e+02  -0.225 0.822197
stateid42   -2.292e+02  1.212e+02  -1.891 0.058878 .
stateid44    5.635e+01  1.397e+02   0.403 0.686701
stateid45    2.558e+02  5.115e+01   5.001 6.62e-07 ***
stateid46   -3.802e+02  9.277e+01  -4.099 4.46e-05 ***
stateid47    6.702e+01  6.358e+01   1.054 0.292114
stateid48   -3.103e+01  1.281e+02  -0.242 0.808612
stateid49   -1.038e+02  1.166e+02  -0.890 0.373731
stateid50   -4.512e+02  1.217e+02  -3.707 0.000220 ***
stateid51   -2.962e+02  5.618e+01  -5.272 1.62e-07 ***
stateid53   -1.169e+02  9.146e+01  -1.278 0.201637
stateid54   -4.138e+02  1.105e+02  -3.744 0.000190 ***
stateid55   -3.327e+02  1.055e+02  -3.155 0.001647 **
stateid56   -2.568e+02  1.114e+02  -2.304 0.021388 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 141.8 on 1113 degrees of freedom
Multiple R-squared:  0.9238,    Adjusted R-squared:  0.9197
F-statistic: 228.6 on 59 and 1113 DF,  p-value: < 2.2e-16
```

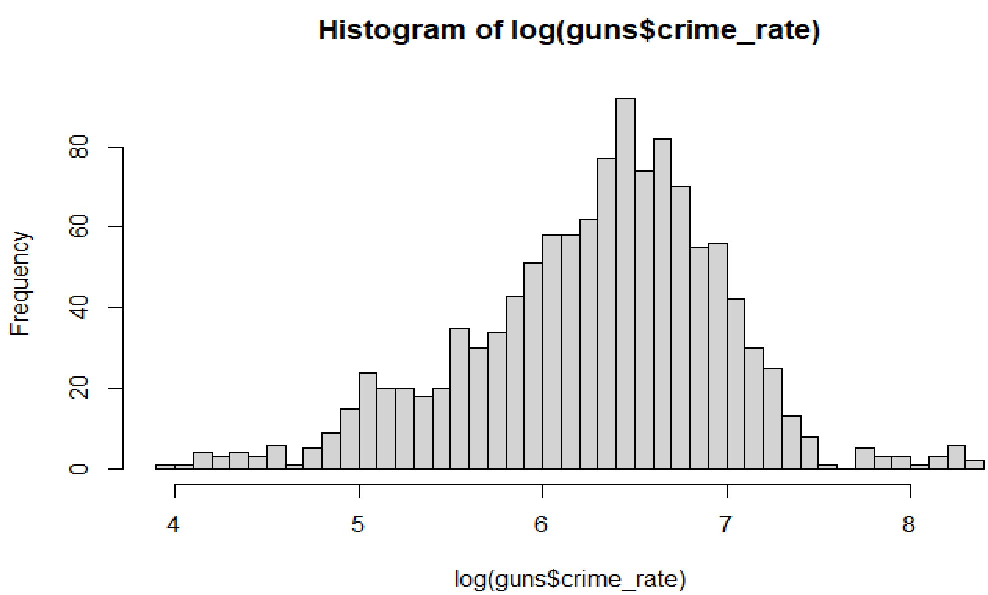
Justification why we are running Log based models

The below histogram shows that the dependent variable crime rate is highly left skewed.



For converting it to an approximately normal distribution, we applied the natural log function to the dependent variable crime rate.

After applying the log function to the dependent variable crime rate, we got the below normal distribution.



Pooled Models

- Pooled Model having incarc_rate, pb1064, pw1064, pm1029, pop, avginc, density and shall as independent variables and Log(crime_rate) as the dependent variable.

```
> model_pooled_l <- plm(crime_rate_ln ~ incarc_rate + pb1064 + pw1064 + pm1029 + pop
+                       + avginc + density + shall
+                       , data = df11, model = "pooling", index = c("stateid", "year"))
> summary(model_pooled_l)
Pooling Model

Call:
plm(formula = crime_rate_ln ~ incarc_rate + pb1064 + pw1064 +
    pm1029 + pop + avginc + density + shall, data = df11, model = "pooling",
    index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.734616 -0.291571  0.055076  0.313877  1.097294

Coefficients:
               Estimate  Std. Error  t-value              Pr(>|t|)
(Intercept)  2.98036479  0.56264414   5.2971          0.0000001406 ***
incarc_rate  0.00146093  0.00011095  13.1673 < 0.00000000000000022 ***
pb1064       0.08906287  0.01724127   5.1657          0.0000002816 ***
pw1064       0.03227731  0.00867437   3.7210             0.0002079 ***
pm1029       0.01239508  0.01115534   1.1111             0.2667401
pop          0.04971090  0.00264948  18.7625 < 0.00000000000000022 ***
avginc       0.00808222  0.00805585   1.0033             0.3159372
density      0.04612028  0.01363454   3.3826             0.0007419 ***
shall1      -0.39677045  0.03372117 -11.7662 < 0.00000000000000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    559.19
Residual Sum of Squares: 228.27
R-Squared:      0.59179
Adj. R-Squared: 0.58898
F-statistic: 210.93 on 8 and 1164 DF, p-value: < 0.000000000000000222
```

- Based on this model, we can see that shall carry law is the very significant variable at the 5% significant level. It shows that with the shall laws in effect, the crime rate drops by 40%.

- This is not practically justifiable in real life. The mere introduction of the shall laws cannot decrease the crime rate by 40%. Also, the PM1029 and Avginc are not significant in the above model. Therefore, we decided to run a model by removing the insignificant variables: PM1029 and Average Income.

The below model shows the updated model.

Dropping insignificant variables - pm1029 and avginc

```
Pooling Model

Call:
plm(formula = crime_rate_ln ~ incarc_rate + pb1064 + pw1064 +
    pop + density + shall, data = df11, model = "pooling", index = c("stateid",
    "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.719526 -0.294138  0.055877  0.316881  1.113135

Coefficients:
                Estimate   Std. Error  t-value              Pr(>|t|)
(Intercept)  2.917615658  0.526583846   5.5406     0.000000037229833 ***
incarc_rate  0.001426410  0.000095221  14.9800 < 0.00000000000000022 ***
pb1064       0.099814735  0.014173426   7.0424     0.000000000003222 ***
pw1064       0.037385730  0.007188333   5.2009     0.000000234101991 ***
pop          0.050302748  0.002542249  19.7867 < 0.00000000000000022 ***
density      0.051992322  0.012657863   4.1075     0.000042786750544 ***
shall1      -0.409201936  0.032002411 -12.7866 < 0.00000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    559.19
Residual Sum of Squares: 228.55
R-Squared:      0.59128
Adj. R-Squared: 0.58918
F-statistic: 281.134 on 6 and 1166 DF, p-value: < 0.000000000000000222
```

- In the above model, shall is significant as well as have the approximately same effect on the dependent variable crime rate. This is again not practically possible that the mere introduction of the crime rate can reduce the crime rate by 40%.

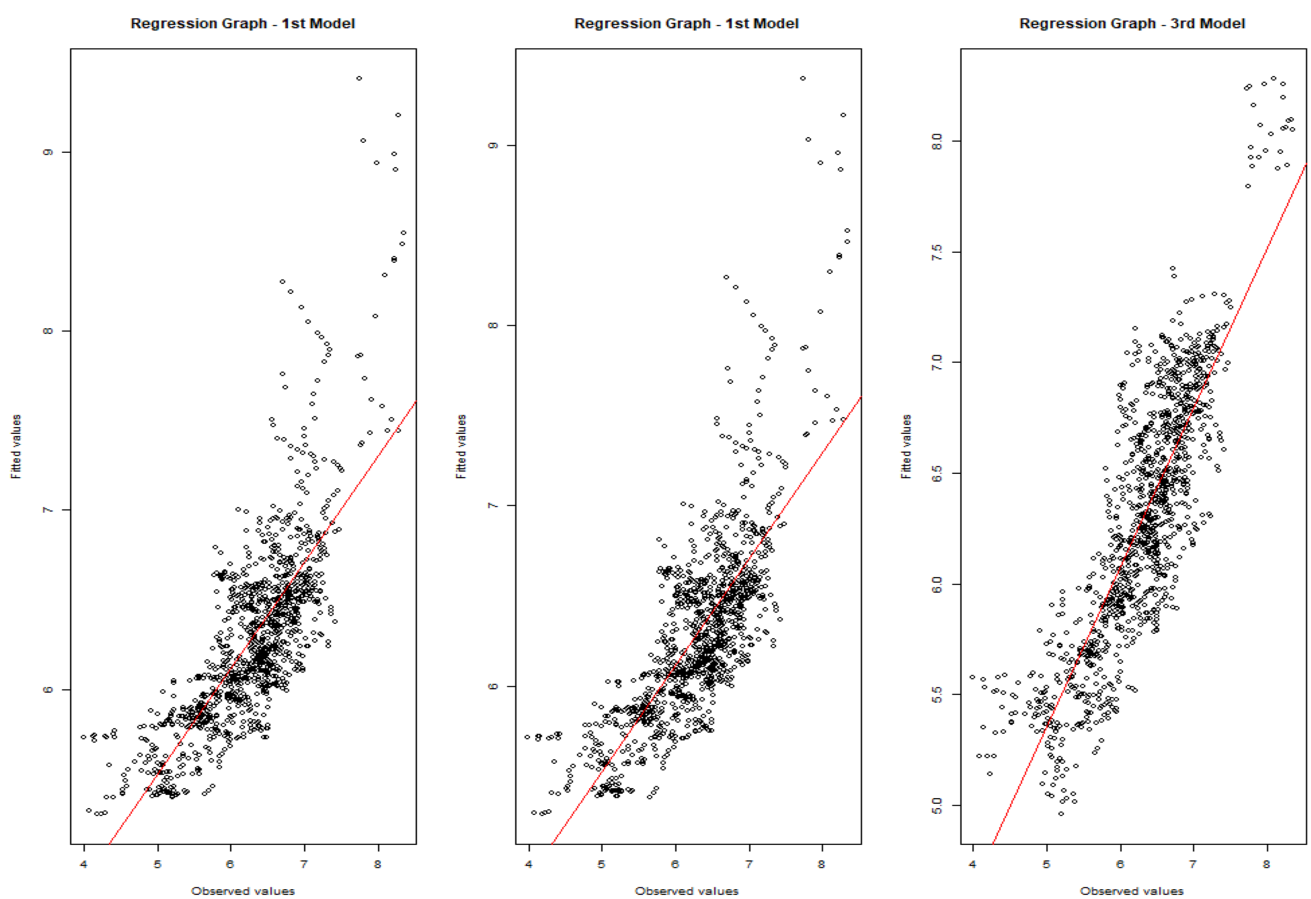
- We feel that here there is omitted variable biased, and therefore, we decided to dig further into individual effect of each of the variable on the crime rate. Based on our analysis, we found below insights.

- Based on EDA, we figured that Incarceration Rate, pb1064, pop, avginc have diminishing effect on crime rate. Hence, we added squared terms for these variables in our model. We have also included density and density^2 term.

The below model shows the updated model with the squared terms for the above-mentioned variables.

```
Pooling Model

Call:
plm(formula = crime_rate_ln ~ incarc_rate + I(incarc_rate^2) + 
    pb1064 + pw1064 + pop + I(pop^2) + I(pb1064^2) + density + 
    shall + avginc + I(avginc^2), data = df11, model = "pooling", 
    index = c("stateid", "year"))

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max. 
-1.578008 -0.201164  0.028155  0.223505  0.935060 

Coefficients:
                       Estimate     Std. Error  t-value              Pr(>|t|)    
(Intercept)       1.28509519989  0.49376722701   2.6026             0.0093689 ** 
incarc_rate       0.00111299195  0.00017726926   6.2785      0.000000000481969 ***
I(incarc_rate^2) -0.00000034171  0.00000012143  -2.8139             0.0049765 ** 
pb1064            0.19539368112  0.01569219369  12.4516 < 0.00000000000000022 ***
pw1064            0.02403794886  0.00714326959   3.3651             0.0007901 ***
pop               0.07415568469  0.00605758747  12.2418 < 0.00000000000000022 ***
I(pop^2)         -0.00188595754  0.00022591377  -8.3481 < 0.00000000000000022 ***
I(pb1064^2)      -0.00565273009  0.00040202159 -14.0608 < 0.00000000000000022 ***
density           0.19240260521  0.01299416766  14.8068 < 0.00000000000000022 ***
shall1           -0.35577089647  0.02762349487 -12.8793 < 0.00000000000000022 ***
avginc            0.31152057239  0.04380740044   7.1111      0.000000000002007 ***
I(avginc^2)      -0.01025804015  0.00145995525  -7.0263      0.000000000003608 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    559.19
Residual Sum of Squares: 156.35
R-Squared:      0.7204
Adj. R-Squared: 0.71776
F-statistic: 271.948 on 11 and 1161 DF, p-value: < 0.000000000000000222
```

Fitted vs Observed Values comparison for the above 3 pooled Models



- We can see that Adjusted R-Squared value has increased from 0.59 to 0.72 after adding the squared terms of the variables that showed diminishing effect on crime rate.
- Looking at the fitted values vs observed value graph, we can see that the fitted line for the 3rd model (Dropping insignificant variables - pm1029 and avginc) is much closer to the observed values as compared to the first two models.
- Also, all the variables are significant at the 5% significant level. The coefficient estimates of the shall has decreased from approximately 40% to 35%. Still in real life, the effect of shall on the crime rate is very high and we cannot justify such a large effect.
- As we have panel data, in pooled OLS we are not considering the information given for each entity and the time. We can create the model to consider the fixed effect of entity and time.

First, we run the below entity fixed effect model.

<u>Entity Fixed Effect Model</u>

- Fixed Effects Model having incarc_rate, pb1064, pw1064, pm1029, pop, avginc, density and shall as independent variables and Log(crime_rate) as dependent variables.

```
> model_fe_1 <- plm(crime_rate_ln ~ incarc_rate + pb1064 + pm1029 + pop + pw1064
+                   + avginc + density + shall
+                   , data = df11, model = "within", index = c("stateid"))
> summary(model_fe_1)
Oneway (individual) effect Within Model

Call:
plm(formula = crime_rate_ln ~ incarc_rate + pb1064 + pm1029 +
    pop + pw1064 + avginc + density + shall, data = df11, model = "within",
    index = c("stateid"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.6028613 -0.1000265  0.0041989  0.1083594  0.5442413

Coefficients:
                Estimate   Std. Error t-value             Pr(>|t|)
incarc_rate -0.000110943  0.000093628 -1.1849              0.23630
pb1064       0.116240315  0.017761369  6.5446 0.000000000090827860 ***
pm1029      -0.040581353  0.006405524 -6.3354 0.000000000342986947 ***
pop          0.010459987  0.008726324  1.1987              0.23091
pw1064       0.041147649  0.005075922  8.1064 0.000000000000001363 ***
avginc      -0.014775041  0.005909937 -2.5000              0.01256 *
density     -0.193207956  0.085060095 -2.2714              0.02331 *
shall1      -0.035903910  0.018872128 -1.9025              0.05737 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    33.904
Residual Sum of Squares: 28.793
R-Squared:      0.15075
Adj. R-Squared: 0.10653
F-statistic: 24.7175 on 8 and 1114 DF, p-value: < 0.000000000000000222
```

- As we can see that, the effect of the shall on the crime rate has reduced drastically from 35% in the pooled OLS model to just 3.6%. Based on this we can infer that by keeping other factor constant, the introduction of the shall law reduces the crime rate by 3.6%.

- Also, the shall law variable is not significant at the 5% significant level, but it is significant at the 10% significant level.

Adding squared terms for incarc_rate, pop, pb1064 and avginc

```
Oneway (individual) effect Within Model

Call:
plm(formula = crime_rate_ln ~ incarc_rate + I(incarc_rate^2) +
    pb1064 + pw1064 + pm1029 + pop + I(pop^2) + I(pb1064^2) +
    density + shall + avginc + I(avginc^2) + I(pb1064^2), data = df11,
    model = "within", index = c("stateid"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.5013811 -0.0963656  0.0012866  0.1072396  0.5090279

Coefficients:
                       Estimate     Std. Error t-value              Pr(>|t|)
incarc_rate      -0.00021676795  0.00014668751 -1.4778               0.13976
I(incarc_rate^2)  0.00000018160  0.00000010844  1.6746               0.09429 .
pb1064            0.16044020574  0.03562179458  4.5040 0.0000073737430552382 ***
pw1064            0.02312071725  0.00536779780  4.3073 0.0000179885032969598 ***
pm1029           -0.01594490873  0.00774462360 -2.0588               0.03974 *
pop               0.05093012023  0.02195071621  2.3202               0.02051 *
I(pop^2)         -0.00114124881  0.00045880475 -2.4874               0.01301 *
I(pb1064^2)      -0.00159897627  0.00114772539 -1.3932               0.16385
density          -0.28221761592  0.11581978822 -2.4367               0.01498 *
shall1           -0.07860973705  0.01885426877 -4.1693 0.0000329275636484034 ***
avginc            0.22469902186  0.02757183921  8.1496 0.0000000000000009773 ***
I(avginc^2)      -0.00788856156  0.00088637104 -8.8998 < 0.00000000000000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    33.904
Residual Sum of Squares: 26.53
R-Squared:      0.21748
Adj. R-Squared: 0.17377
F-statistic: 25.7082 on 12 and 1110 DF, p-value: < 0.000000000000000222
```

- As discussed previously, the incarc_rate, pop, pb1064 and avginc have the diminishing effect and therefore, we introduce the square term here.

- After adding the square term, we found that shall become significant at 5% significance level and the coefficient value for the same increase from 3.6% to 7.8%.

- But at the same time other variables, such as incarc_rate, incarc_rate^2, pb1064^2 are insignificant and therefore, we decided to remove them and run the model again.

Dropping insignificant terms – incarc_rate, incarc_rate^2, pv_1064^2

```
Oneway (individual) effect Within Model

Call:
plm(formula = crime_rate_ln ~ pb1064 + pw1064 + pm1029 + pop +
    I(pop^2) + density + shall + avginc + I(avginc^2), data = df11,
    model = "within", index = c("stateid"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.4935137 -0.0949985  0.0011281  0.1070806  0.5082022

Coefficients:
              Estimate  Std. Error t-value              Pr(>|t|)
pb1064       0.10994786  0.01695029  6.4865   0.000000000131892501 ***
pw1064       0.02263141  0.00527520  4.2902   0.000019405990478779 ***
pm1029      -0.01726122  0.00591870 -2.9164               0.003612 **
pop          0.05293436  0.02177349  2.4311               0.015209 *
I(pop^2)    -0.00113768  0.00045493 -2.5008               0.012534 *
density     -0.41862058  0.06809266 -6.1478   0.000000001093087202 ***
shall       -0.07903298  0.01865741 -4.2360   0.000024632828116187 ***
avginc       0.20762101  0.02595036  8.0007   0.000000000000003087 ***
I(avginc^2) -0.00723222  0.00080284 -9.0083 < 0.00000000000000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    33.904
Residual Sum of Squares: 26.636
R-Squared:      0.21436
Adj. R-Squared: 0.17272
F-statistic: 33.7428 on 9 and 1113 DF, p-value: < 0.000000000000000222
```

- By removing the factors which were not significant previously, we can see that the effect of shall law on the crime rate has not changed much at the same time they shall is still the significant at the 5% significant level.

- Entity fixed effect models do not consider the factors which have changed across the entity but constant at the same time, such as a factor which has nationwide effect.

- This factors, might be inflation, GDP growth rate, nationwide law change in the other factors, nationwide unemployment and technological improvement in the guns or the police surveillance system.

Therefore, we decided to use the Entity and Time fixed effect model.

Entity and Time fixed effect model

- Fixed Effects and Time Model having incarc_rate, pb1064, pw1064, pm1029, pop, avginc, density and shall as independent variables and Log(crime_rate) as dependent variables.

```
> model_fet_l <- plm(crime_rate_ln ~ incarc_rate + pb1064 + pm1029 + pop + pw1064
+                   + avginc + density + shall
+                   , data = df11, model = "within", index = c("stateid", "year"), effect = "twoways")
> summary(model_fet_l)
Twoways effects Within Model

Call:
plm(formula = crime_rate_ln ~ incarc_rate + pb1064 + pm1029 +
    pop + pw1064 + avginc + density + shall, data = df11, effect = "twoways",
    model = "within", index = c("stateid", "year"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.4493435 -0.0816470  0.0024252  0.0869655  0.6113997

Coefficients:
                Estimate    Std. Error t-value   Pr(>|t|)
incarc_rate  0.000040373  0.000089988  0.4487    0.65377
pb1064       0.043432057  0.022617569  1.9203    0.05508 .
pm1029       0.071796902  0.015562675  4.6134 0.00000443 ***
pop         -0.005027838  0.007841696 -0.6412    0.52155
pw1064       0.010398004  0.007835905  1.3270    0.18480
avginc      -0.000196330  0.006413817 -0.0306    0.97559
density     -0.098314242  0.076031798 -1.2931    0.19626
shall1      -0.013777332  0.017101525 -0.8056    0.42064
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    22.598
Residual Sum of Squares: 21.271
R-Squared:      0.058718
Adj. R-Squared: -0.010241
F-statistic: 8.51493 on 8 and 1092 DF, p-value: 0.000000000027183
```

- The Entity and Time fixed effect model has a coefficient of shall value of 1.4%. Which is very low compared to fixed effect model. In the fixed effect model, the effect of shall law on the crime rate was approximately 7.9%.

- Also, here the shall law is not significant even at 10% significant level. Also, in this model except pm1029, no other variable is significant at the 5% significant level and therefore we decided to remove the variable incarc_rate and add the square term of the pop and avg income based on our learning from the previous model.

<u>Dropping insignificant term – incarc_rate and adding squared terms for pop and avginc</u>

```
>
> model_fet_l_2 <- plm(crime_rate_ln ~ pb1064 + pm1029 + pop + I(pop^2)
+                       + density + shall + avginc + I(avginc^2)
+                     , data = df11, model = "within", index = c("stateid", "year"), effect = "twoways")
> summary(model_fet_l_2)
Twoways effects Within Model

Call:
plm(formula = crime_rate_ln ~ pb1064 + pm1029 + pop + I(pop^2) +
    density + shall + avginc + I(avginc^2), data = df11, effect = "twoways",
    model = "within", index = c("stateid", "year"))

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

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.4642439 -0.0787499 -0.0023998  0.0797421  0.5890593

Coefficients:
               Estimate  Std. Error t-value              Pr(>|t|)
pb1064       0.05614576  0.01397450  4.0177       0.0000627871620 ***
pm1029       0.05247060  0.01146711  4.5757       0.0000052886705 ***
pop          0.03690734  0.01900274  1.9422              0.052368 .
I(pop^2)    -0.00095006  0.00040136 -2.3671              0.018102 *
density     -0.37770757  0.06006996 -6.2878       0.0000000004648 ***
shall1      -0.04906149  0.01693564 -2.8969              0.003843 **
avginc       0.22556632  0.02623357  8.5984 < 0.00000000000000022 ***
I(avginc^2) -0.00701904  0.00076732 -9.1475 < 0.00000000000000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    22.598
Residual Sum of Squares: 19.57
R-Squared:      0.13397
Adj. R-Squared: 0.070524
F-statistic: 21.1157 on 8 and 1092 DF, p-value: < 0.000000000000000222
```

- Based on the above model, we can see that the shall carry law has been significant at the 5% significant level. Apart from it all other variables are also significant at the 5% significant level except pop. Population is significant at the 10% significant level.

- Our hypothesis was:
  - Null $H_0: \beta_{Shall-issue\ Laws} \geq 0$
  - Alternative $H_1: \beta_{Shall-issue\ Laws} \leq 0$
- Based on the P-Value form the above model, we can infer that the 5% significant level we can reject the null and accept the alternative hypothesis that shall carry law reduces the crime rate by 4.9% on an average.

<u>pFTest</u>

We ran an F test on Fixed Effects Model and Fixed and Time Effects Model.

model_fet_l_2 <- plm(crime_rate_ln ~ pb1064 + pm1029 + pop + I(pop^2) + density + shall + avginc + I(avginc^2) , data = df11, model = "within", index = c("stateid", "year"), effect = "twoways")

model_fe_l_2_test <- plm(crime_rate_ln ~ pb1064 + pm1029 + pop + I(pop^2) + density + shall + avginc + I(avginc^2) , data = df11, model = "within", index = c("stateid"))

```
> pFtest(model_fet_l_2,model_fe_l_2_test)

	F test for twoways effects

data:  crime_rate_ln ~ pb1064 + pm1029 + pop + I(pop^2) + density +  ...
F = 19.038, df1 = 22, df2 = 1092, p-value < 0.00000000000000022
alternative hypothesis: significant effects
```

As the p-value is almost 0 (<0.05), we can say that there’s time varying effect and Fixed & Time Effect Model should be used to interpret the impact Shall laws.
Additionally, since the data is not taken randomly, there’s no point in running the Random Effects Model.

Conclusion

Based on our analysis, we started with a pooled OLS model with log-transformed crime rate as the dependent variable and several independent variables. However, we found that the introduction of the shall law variable had an unreasonably large effect on the crime rate. Therefore, we identified omitted variable bias and added squared terms of variables that had a diminishing effect on crime rate, such as incarceration rate, population, and average income.

We then moved on to entity fixed effect models and found that the effect of the shall law variable on crime rate reduced significantly compared to the pooled OLS model. However, these models did not consider nationwide factors that could impact crime rates. Therefore, we introduced a fixed effects and time model that considered both entity and time-fixed effects.

We found that the effect of the shall law variable reduced even further in the entity and time-fixed effects model and was not statistically significant. Additionally, only one variable (pm1029) was significant at the 5% significance level, and we further refined the model by removing insignificant variables and adding squared terms for population and average income.

Overall, our analysis suggests that the introduction of shall laws does influence crime rates, but the effect is significantly lower than initially estimated.