Econometrics Project

Karthik Mahanth Kattula, Akshata Bodhankar, Garima Tuteja, Syamala Anisha Katta

11/21/2020

```
guns<-read.dta("guns.dta",convert.factors = TRUE,
                missing.type = FALSE,
                convert.underscore = FALSE, warn.missing.labels = TRUE)
```

```
str(guns)
```

```
## 'data.frame':    1173 obs. of  13 variables:
##  $ year       : int  77 78 79 80 81 82 83 84 85 86 ...
##  $ vio        : num  414 419 413 448 470 ...
##  $ mur        : num  14.2 13.3 13.2 13.2 11.9 ...
##  $ rob        : num  96.8 99.1 109.5 132.1 126.5 ...
##  $ incarc_rate: int  83 94 144 141 149 183 215 243 256 267 ...
##  $ pb1064     : num  8.38 8.35 8.33 8.41 8.48 ...
##  $ pw1064     : num  55.1 55.1 55.1 54.9 54.9 ...
##  $ pm1029     : num  18.2 18 17.8 17.7 17.7 ...
##  $ pop        : num  3.78 3.83 3.87 3.9 3.92 ...
##  $ avginc     : num  9.56 9.93 9.88 9.54 9.55 ...
##  $ density    : num  0.0746 0.0756 0.0762 0.0768 0.0772 ...
##  $ stateid    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ shall      : int  0 0 0 0 0 0 0 0 0 0 ...
##  - attr(*, "datalabel")= chr ""
##  - attr(*, "time.stamp")= chr " 5 Sep 2014 22:29"
##  - attr(*, "formats")= chr [1:13] "%9.0g" "%9.0g" "%9.0g" "%9.0g" ...
##  - attr(*, "types")= int [1:13] 251 254 254 254 252 254 254 254 254 254 ...
##  - attr(*, "val.labels")= chr [1:13] "" "" "" "" ...
##  - attr(*, "var.labels")= chr [1:13] "" "Violent Crime Rate per 100,000 population (BJS)" "Murder Cri
##  - attr(*, "expansion.fields")=List of 5
##   ..$ : chr [1:3] "_dta" "ReS_Xij" "@age1019 @age2029 @age3039 @age4049 @age5064 @ageo65"
##   ..$ : chr [1:3] "_dta" "ReS_str" "1"
##   ..$ : chr [1:3] "_dta" "ReS_j" "demog"
##   ..$ : chr [1:3] "_dta" "ReS_ver" "v.2"
##   ..$ : chr [1:3] "_dta" "ReS_i" "fipsstat year"
##  - attr(*, "version")= int 12
```

#The given balanced panel data observes the guns related data across different states in USA over the years 1977-1999 .From structure of data, we see that variables vio,mur,rob,incarc_rate,avginc,pop,density,pb1064,pw1064,pm1029 are continuous and the variables year,stateid and shall should be indicator variables.

```
summary(guns)
```

```
##       year            vio              mur             rob
##  Min.   :77    Min.   :  47.0   Min.   : 0.200   Min.   :   6.4
##  1st Qu.:82    1st Qu.: 283.1   1st Qu.: 3.700   1st Qu.:  71.1
##  Median :88    Median : 443.0   Median : 6.400   Median : 124.1
##  Mean   :88    Mean   : 503.1   Mean   : 7.665   Mean   : 161.8
##  3rd Qu.:94    3rd Qu.: 650.9   3rd Qu.: 9.800   3rd Qu.: 192.7
##  Max.   :99    Max.   :2921.8   Max.   :80.600   Max.   :1635.1
##   incarc_rate        pb1064           pw1064          pm1029
##  Min.   :  19.0   Min.   : 0.2482   Min.   :21.78   Min.   :12.21
##  1st Qu.: 114.0   1st Qu.: 2.2022   1st Qu.:59.94   1st Qu.:14.65
##  Median : 187.0   Median : 4.0262   Median :65.06   Median :15.90
##  Mean   : 226.6   Mean   : 5.3362   Mean   :62.95   Mean   :16.08
##  3rd Qu.: 291.0   3rd Qu.: 6.8507   3rd Qu.:69.20   3rd Qu.:17.53
##  Max.   :1913.0   Max.   :26.9796   Max.   :76.53   Max.   :22.35
##       pop            avginc          density          stateid
##  Min.   : 0.4027   Min.   : 8.555   Min.   : 0.000707   Min.   : 1.00
##  1st Qu.: 1.1877   1st Qu.:11.935   1st Qu.: 0.031911   1st Qu.:16.00
##  Median : 3.2713   Median :13.402   Median : 0.081569   Median :29.00
##  Mean   : 4.8163   Mean   :13.725   Mean   : 0.352038   Mean   :28.96
##  3rd Qu.: 5.6856   3rd Qu.:15.271   3rd Qu.: 0.177718   3rd Qu.:42.00
##  Max.   :33.1451   Max.   :23.647   Max.   :11.102116   Max.   :56.00
##      shall
##  Min.   :0.000
##  1st Qu.:0.000
##  Median :0.000
##  Mean   :0.243
##  3rd Qu.:0.000
##  Max.   :1.000
```

```r
mydata <- guns[, c(2,3,4,5,6,7,8,9,10,11)]
cormat <- round(cor(mydata),2)
melted_cormat <- melt(cormat)

get_lower_tri<-function(cormat){
    cormat[upper.tri(cormat)] <- NA
    return(cormat)
}

get_upper_tri <- function(cormat){
    cormat[lower.tri(cormat)]<- NA
    return(cormat)
}

upper_tri <- get_upper_tri(cormat)

melted_cormat <- melt(upper_tri, na.rm = TRUE)
# Heatmap
library(ggplot2)
ggplot(data = melted_cormat, aes(Var2, Var1, fill = value))+
 geom_tile(color = "white")+
 scale_fill_gradient2(low = "blue", high = "red", mid = "white",
   midpoint = 0, limit = c(-1,1), space = "Lab",
   name="Pearson\nCorrelation") +
  theme_minimal()+
```
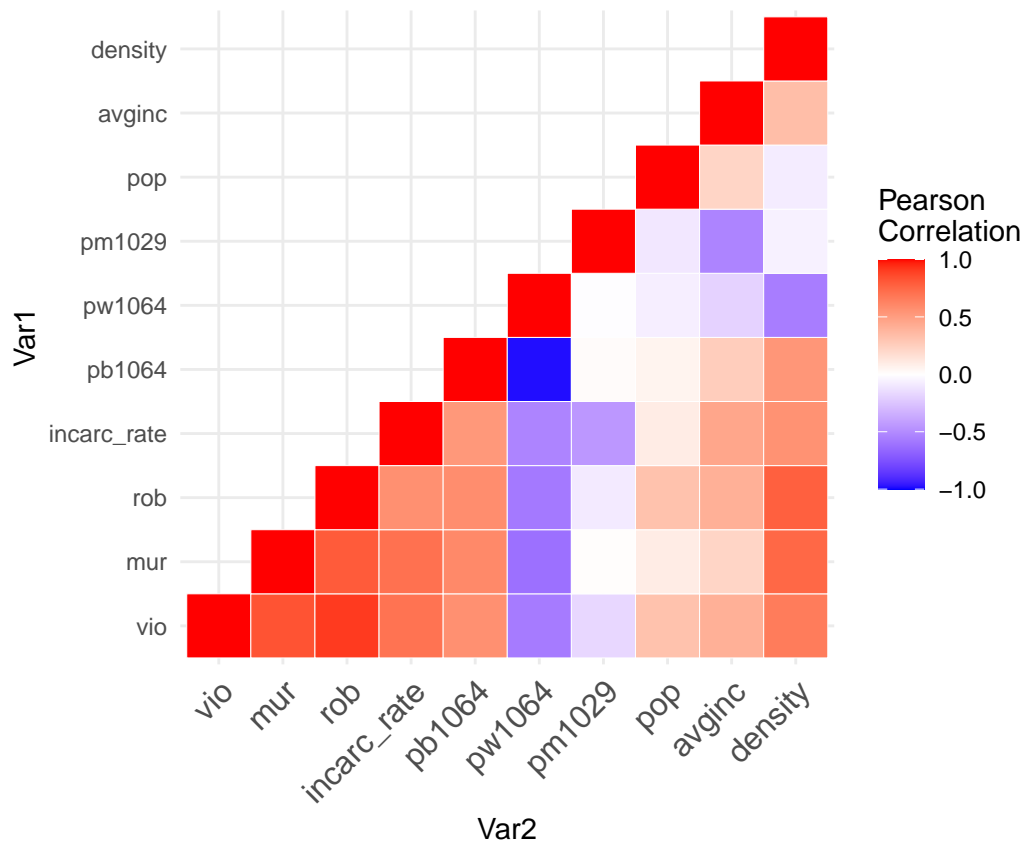
```
theme(axis.text.x = element_text(angle = 45, vjust = 1,
    size = 12, hjust = 1))+
coord_fixed()
```


#There are significant strong correlations between violent rate , murder rate, robbery rate

#Check if there are any missing values in data

```
sapply(guns, function(x) sum(is.na(x)))
```

```
##        year         vio         mur         rob incarc_rate      pb1064
##           0           0           0           0           0           0
##      pw1064      pm1029         pop      avginc     density      stateid
##           0           0           0           0           0           0
##       shall
##           0
```

#Average crime rate across each state in United States over the period 1977-1999

```
shall<-data.frame(guns%>%group_by(stateid)%>%summarise(avgcrime=mean(vio),.groups='drop')%>%arrange(des
names(shall)<-c('stateid','avgcrime')
head(shall)
```

```
##   stateid  avgcrime
## 1      11 2048.9783
```

3

```
## 2        12   999.2348
## 3        36   941.3174
## 4         6   877.4826
## 5        24   853.6348
## 6        17   828.1652
```

#From the above analysis we see that State with Id 11 is having highest crime rate whereas State with Id 38 is having lowest crime rate. Further will see if there are any significant differences across these two states

#Subset the two states with Id 38 and 11 and analyse if we can get any specific information

```r
crimes<-guns[guns$stateid=='38' | guns$stateid=='11' | guns$stateid=='12',]
```

```r
blacks<-data.frame(crimes%>%group_by(stateid)%>%summarise(blacks=mean(pb1064),whites=mean(pw1064),.group
names(blacks)<-c('stateid','blacks','whites')
blacks
```

```
##   stateid    blacks    whites
## 1      11 23.886516 24.86751
## 2      12  5.481352 58.68433
## 3      38  1.813590 68.73256
```

```r
cols <- c('red','blue');
ylim <- c(0,max(blacks[c('blacks','whites')])*1.8);
par(lwd=6);
barplot(
    t(blacks[c('blacks','whites')]),
    beside=T,
    ylim=ylim,
    border=cols,
    col='white',
    names.arg=blacks$stateid,
    xlab='StateId',
    ylab='PopulationPercentage',
    legend.text=c('AvgPercentageBlacks','AvgPercentageWhites'),
    args.legend=list(text.col=cols,col=cols,border=cols,bty='n')
);
box();
```
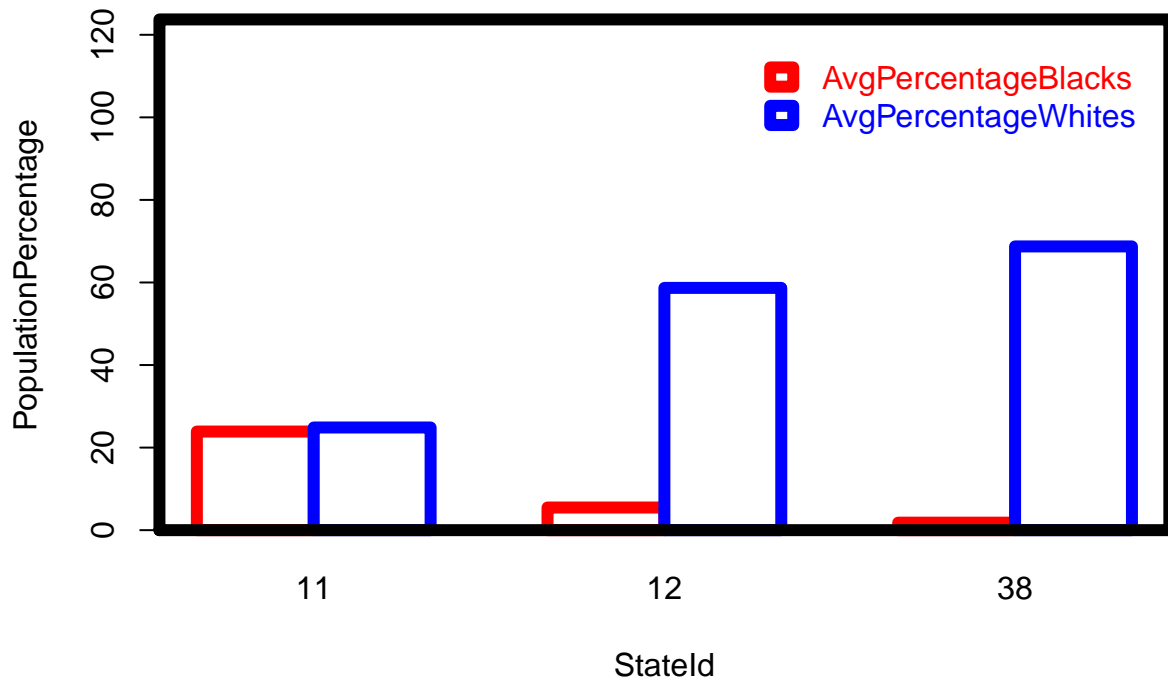
#State Id 38 is having lowest crime rate and we see that Average percentage of Blacks living in State 38 is very less when compared with state which is having highest crime rate State 11. However in State Id 11, there are equal Blacks and Whites. So, we need to further analyse if blacks were the main cause for higher crime rate in State Id 11. Also we observed that crime rate in State id 11 is very high when compared with all other states.

```
shall
```

```
##    stateid   avgcrime
## 1       11 2048.97827
## 2       12  999.23478
## 3       36  941.31739
## 4        6  877.48261
## 5       24  853.63478
## 6       17  828.16522
## 7       45  800.26522
## 8       22  778.43913
## 9       32  755.03478
## 10      35  743.86957
## 11      26  695.21304
## 12      25  616.71739
## 13      48  612.65652
## 14       4  608.09565
## 15       2  596.79565
## 16      13  595.57826
## 17      29  582.65218
## 18      47  580.68261
```

```
## 19        10  564.35652
## 20         1  558.17392
## 21        34  552.82174
## 22        37  526.39565
## 23        41  496.33913
## 24        40  491.58696
## 25         8  482.78695
## 26        53  447.72173
## 27        39  445.45217
## 28         5  438.40000
## 29         9  420.67391
## 30        18  399.86956
## 31        20  391.70870
## 32        42  383.45652
## 33        44  369.40000
## 34        28  355.52609
## 35        21  334.69565
## 36        51  322.95218
## 37        31  300.73478
## 38        56  287.50000
## 39        49  283.50869
## 40        27  277.26522
## 41        16  262.64348
## 42        15  259.43478
## 43        19  247.60000
## 44        55  225.28261
## 45        54  188.76087
## 46        30  183.47826
## 47        46  157.36087
## 48        23  154.04783
## 49        50  133.20870
## 50        33  130.84783
## 51        38   68.00435
```

#The average crime rate in State Id 11 is 2048 which is double than that of all the other states. In state id 12 the crime rate is 999 which is half that of the average crime rate in state id 11.

```
k<-crimes[crimes$stateid==11,]
unique(k$shall)
```

```
## [1] 0
```

```
k<-crimes[crimes$stateid==12,]
unique(k$shall)
```

```
## [1] 0 1
```

```
k<-crimes[crimes$stateid==38,]
unique(k$shall)
```

```
## [1] 0 1
```

#The shall issue laws were not passed in the State Id 11. It might be one of the reason that crime rate is very high when compared with other states and shall issue laws were passed in states 12 and 38. Also crime rate is very high in state 12 when compared to state 38, though the shall issue laws were passed in both of these two states. Percentage of blacks living in State 12 is 5.48 which is high when compared with State 38 1.48%. So, shall issue laws and percentage of blacks living in the state are some of the important features in this sample
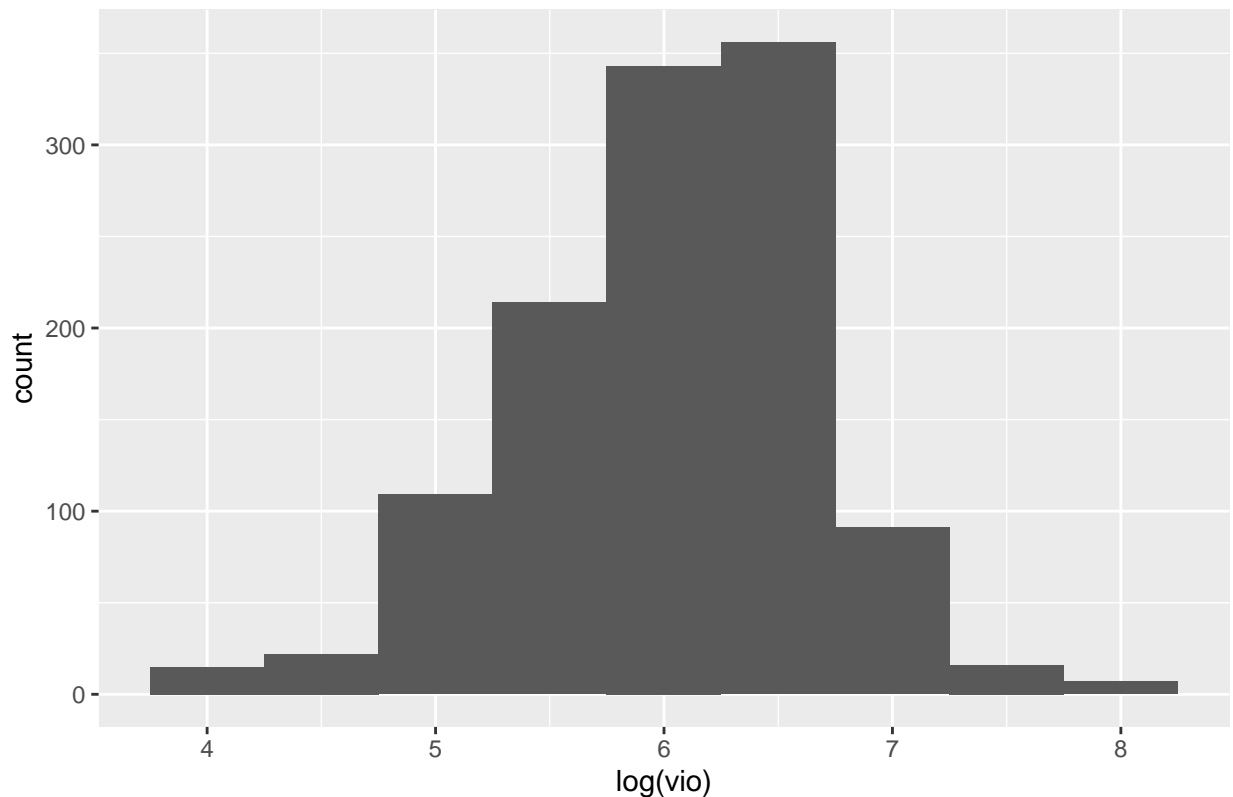
#Histogram of Violent crime rate

```
ggplot(guns,aes(x=vio)) + geom_histogram(binwidth = 40) +ggtitle("Violent Crime rate across United State
```

Violent Crime rate across United States
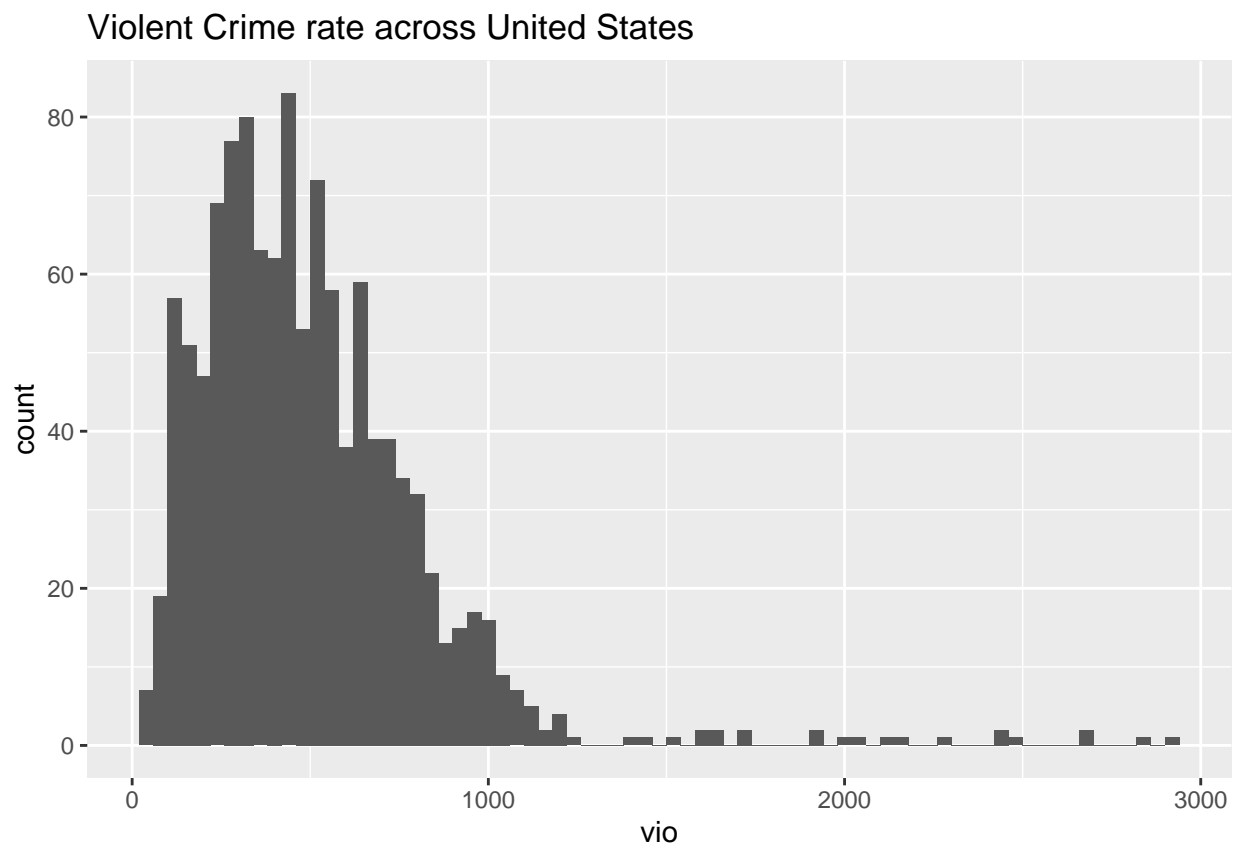

```
ggplot(guns,aes(x=log(vio))) + geom_histogram(binwidth =0.5) + ggtitle("Violent crime rate across United
```

Violent crime rate across United States

#Before applying log transformation, the distribution of crime rate across United States in heavily skewed towards right and poitively skewed distribution. Most of the statistical analysis or statistical models require Normal Distribution because of its significant statistical properties such as constant mean and constant variance across the data. After applying necessary log transformation on data the distribution of crime rate became approximately symmetrical or atleast weakly skewed but not heavily skewed

```
guns$vio<-log(guns$vio)
```

#The given guns data is a balanced panel of data on 50 United States plus the District of Columbia (for a total of 51 states) by year from 1977-1999. #Ordinary Least Squares Estimation:

```
ols <- lm(vio~mur + rob + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density + shall,
          data = guns)
tidy(ols)
```

```
## # A tibble: 11 x 5
##    term         estimate std.error statistic  p.value
##    <chr>           <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)  3.55      0.436        8.14  9.99e-16
##  2 mur          0.00344   0.00324      1.06  2.88e- 1
##  3 rob          0.00316   0.000144    22.0   6.19e-90
##  4 incarc_rate  0.00126   0.000113    11.2   1.50e-27
##  5 pb1064       0.0579    0.0134       4.33  1.64e- 5
##  6 pw1064       0.0266    0.00671      3.96  8.08e- 5
##  7 pm1029      -0.00568   0.00916     -0.620 5.35e- 1
```

8

```
##  8 pop         0.00743  0.00249    2.99  2.85e- 3
##  9 avginc     -0.00844  0.00638   -1.32  1.87e- 1
## 10 density    -0.248    0.0152    -16.3  3.34e-54
## 11 shall      -0.278    0.0264    -10.5  6.96e-25
```

#Check Linear Regression Assumptions

```
par(mfrow=c(2,2))
plot(ols)
```
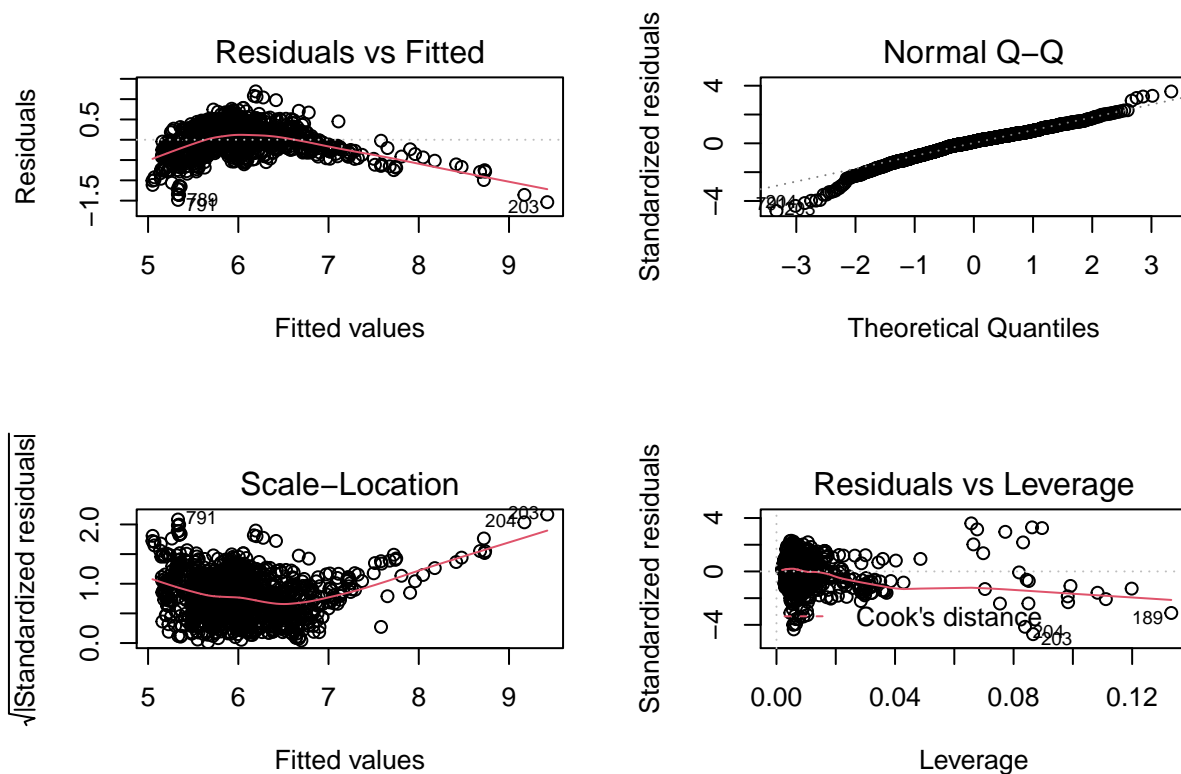

#From the Ordinary Least Squares Regression Assumption plots, Expected value of random error is not zero(Residuals versus Fitted), Variance across the error terms are not constant(Scale-Location) i.e. Heteroskedasticity of Variances and Covariance between any two pair of error terms are not zero(Q-Q) error terms are not normally distributed

#Pooles OLS Estimation: In case of Pooled OLS model we ignore the panel nature of data and estimate it using Ordinary Least squares method. However we need to specify arguments to ignore the panel nature of data i.e. the intercepts and slope coefficients do not vary across states and time

```
pooled_ols <- plm(vio~mur + rob + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density + shal
                  data = guns,index = c("stateid","year"))
tidy(pooled_ols)
```

```
## # A tibble: 11 x 5
##    term        estimate std.error statistic  p.value
##    <chr>          <dbl>     <dbl>     <dbl>    <dbl>
```

```
##  1 (Intercept)  3.55      0.436        8.14  9.99e-16
##  2 mur          0.00344   0.00324      1.06  2.88e- 1
##  3 rob          0.00316   0.000144    22.0   6.19e-90
##  4 incarc_rate  0.00126   0.000113    11.2   1.50e-27
##  5 pb1064       0.0579    0.0134       4.33  1.64e- 5
##  6 pw1064       0.0266    0.00671      3.96  8.08e- 5
##  7 pm1029      -0.00568   0.00916     -0.620 5.35e- 1
##  8 pop          0.00743   0.00249      2.99  2.85e- 3
##  9 avginc      -0.00844   0.00638     -1.32  1.87e- 1
## 10 density     -0.248     0.0152     -16.3   3.34e-54
## 11 shall       -0.278     0.0264     -10.5   6.96e-25
```

#The least squares estimation equation using Pooled OLS estimation is $\log(vio)= 3.549 + 0.0034 mur + 0.0031 rob + 0.0012 incarc\_rate + 0.057 pb1064 + 0.026 pw1064 -0.0056 pm1029 + 0.0074 pop -0.0084 avginc -0.2481 density -0.2779 shall$ #71.86 percentage of variation in crime rate across United States over time period 1977-1999 is explained by explanatory variables murder rate, robbery rate incarceration rate, balcks and whites living in states with age 10-64, average income, density and shall issue laws.
#Except murder rate,percentage of male population in states with age 10-29 and average real per capita income all the explanatory variables are significant at 1% #The coefficient of Indicator variable shall issue laws is negative and the interpretation is violent crime rate is going to reduce by 28% approximately when not having shall issue laws. Reducing the crime rate by approximately 28% by implementing shall issue laws is a good impact and shall issue laws are playing a major role in reducing the crime rate and also we can observe that the p-value is very very less and this coefficient is highly significant

```r
coeftest(pooled_ols, vcov=vcovHC(pooled_ols,type="HC0",cluster="group"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)  3.54991496  1.51738916  2.3395 0.0194788 *
## mur          0.00344009  0.01186464  0.2899 0.7719100
## rob          0.00316078  0.00050411  6.2700 5.080e-10 ***
## incarc_rate  0.00125707  0.00032017  3.9263 9.135e-05 ***
## pb1064       0.05785445  0.05336990  1.0840 0.2785773
## pw1064       0.02656463  0.02482802  1.0699 0.2848660
## pm1029      -0.00567616  0.02869313 -0.1978 0.8432182
## pop          0.00742914  0.00836194  0.8884 0.3744843
## avginc      -0.00843535  0.02106618 -0.4004 0.6889195
## density     -0.24810972  0.07362984 -3.3697 0.0007772 ***
## shall       -0.27792618  0.09281014 -2.9946 0.0028065 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#The estimators obtained using Ordinary Least squares estimation and pooled ols estimation are same. However the basic assumptions of linear regression are violated. Heteroskedasticity exists in data. The estimators are linear unbiased and consistent but they are no longer the best and standard erros are incorrect. Expected value of error term is zero is violated, we can relax this assumption for now because we are having consistent estimators. Having large sample sizes will converge to true population parameter. We can fix standard errors or calculate correct standard errors using Cluster Robust Standard Errors. The standard errors are now large using Cluster Robust Standard errors. The standard errors obtained using pooled OLS with least squares standard errors is very very less and overstating the pooled ols model. Here we are ignoring the individual correlation within entities state over time. With cluster robust standard errors the standard errors and confidence intervals obtained are correct

#In Pooled OLS model, estimators are same for all states and across all time periods. This seems to be not a good approach because there will be states where population of people living there are nice and having low crime rate and there will be states where more people in those states tends to commit crime always. So the nature of people living in these states in an unobservable characteristic and this might be correlated with error term resulting in endogeneity. So the variables such as percent of population living in states might be correlated with error term. This unobserved heterogeneity leads to biased and inconsistent estimators. Panel Data can control this unobserved heterogeneity using Fixed effects model

#Entity Fixed Effects: Entities are States: The omitted variable which is hiding in the error term results in biased estimators known as Omitted Variable bias. These omitted variables might vary over entities but not across time. In such a case we should use Entity fixed effects model

```
fixedeffects_state <- plm(vio~mur + rob + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + densi
tidy(fixedeffects_state)
```

```
## # A tibble: 10 x 5
##    term            estimate std.error statistic  p.value
##    <chr>              <dbl>     <dbl>     <dbl>    <dbl>
##  1 mur            0.00502   0.00173      2.90   3.85e- 3
##  2 rob            0.00162   0.0000972   16.6    1.23e-55
##  3 incarc_rate    0.000118  0.0000828    1.43   1.54e- 1
##  4 pb1064         0.0753    0.0158       4.78   2.01e- 6
##  5 pw1064         0.0321    0.00434      7.40   2.69e-13
##  6 pm1029        -0.0403    0.00564     -7.14   1.64e-12
##  7 pop            0.00893   0.00747      1.19   2.32e- 1
##  8 avginc        -0.0000537 0.00523     -0.0103 9.92e- 1
##  9 density       -0.0373    0.0728      -0.512  6.09e- 1
## 10 shall         -0.0556    0.0161      -3.46   5.59e- 4
```

#The estimators for explanatory variables are same across all entities and across all time periods. However the intercepts vary for each entity and if we observe now that passing shall issue laws across United States will reduce the crime rate by 5.5% approximately. In case of Pooled OLS we got an estimator which estimates that shall issue laws will reduce crime rate by 28%. 28% reduce in crime rate is huge and there may be chance that people will pretend as good till they get licence and start doing crime once they have guns. The nature or characteristics of people across different states is unobservable heterogeneity and this unobservable heterogeneity is varied across each state so the intercept varies across each state. The effect of entity fixed effects across each state varies and shown below

```
fixef(fixedeffects_state)
```

```
##      1      2      4      5      6      8      9     10     11     12     13
## 4.1823 4.2649 4.3279 4.0970 4.1104 4.0759 3.8645 4.1889 3.8987 4.5034 4.0991
##     15     16     17     18     19     20     21     22     23     24     25
## 3.2860 3.8206 4.2385 3.9785 3.6656 4.0143 3.8086 4.3815 3.2247 4.1759 4.2820
##     26     27     28     29     30     31     32     33     34     35     36
## 4.2182 3.6473 3.8780 4.1902 3.3619 3.8240 4.1279 3.0539 3.9676 4.5509 3.9752
##     37     38     39     40     41     42     44     45     46     47     48
## 4.1097 2.5398 3.8853 4.1442 4.1405 3.7826 3.9823 4.5220 3.3333 4.1322 4.0866
##     49     50     51     53     54     55     56
## 3.9512 3.1070 3.6122 4.0579 3.3587 3.4563 3.8327
```

```r
summary(fixedeffects_state)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = vio ~ mur + rob + incarc_rate + pb1064 + pw1064 +
##     pm1029 + pop + avginc + density + shall, data = guns, model = "within",
##     index = "stateid")
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##        Min.    1st Qu.     Median    3rd Qu.        Max.
## -0.6022695 -0.0706786  0.0068197  0.0780594  0.5381769
##
## Coefficients:
##                  Estimate  Std. Error t-value  Pr(>|t|)
## mur            5.0171e-03  1.7323e-03  2.8963 0.0038501 **
## rob            1.6168e-03  9.7213e-05 16.6311 < 2.2e-16 ***
## incarc_rate    1.1800e-04  8.2803e-05  1.4251 0.1544210
## pb1064         7.5260e-02  1.5753e-02  4.7776 2.011e-06 ***
## pw1064         3.2146e-02  4.3443e-03  7.3997 2.685e-13 ***
## pm1029        -4.0324e-02  5.6448e-03 -7.1435 1.642e-12 ***
## pop            8.9291e-03  7.4734e-03  1.1948 0.2324268
## avginc        -5.3739e-05  5.2293e-03 -0.0103 0.9918025
## density       -3.7272e-02  7.2797e-02 -0.5120 0.6087481
## shall         -5.5627e-02  1.6074e-02 -3.4606 0.0005594 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    36.789
## Residual Sum of Squares: 20.63
## R-Squared:      0.43922
## Adj. R-Squared: 0.40897
## F-statistic: 87.0967 on 10 and 1112 DF, p-value: < 2.22e-16
```

```r
phtest(pooled_ols,fixedeffects_state)
```

```
##
##   Hausman Test
##
## data:  vio ~ mur + rob + incarc_rate + pb1064 + pw1064 + pm1029 + pop +  ...
## chisq = 764.17, df = 10, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

#p-value is less than 0.05. So we reject the Null Hypothesis and conclude that the fixed effects model is better one

#In case of entity fixed effects, we assumed unobservable heterogeneity is not varying over time and it is varying only across entities. However if the unobservable hetrogeneity is changing over time then the estimators are still biased only. There is a chance that nature of people change over time because if the government pass strict laws to reduce crime then crime rate will reduce automatically and the nature of people will change.

12

#Time Fixed Effects:

```
fixedeffects_time <- plm(vio~mur + rob + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density
tidy(fixedeffects_time)
```

```
## # A tibble: 10 x 5
##    term        estimate std.error statistic  p.value
##    <chr>          <dbl>     <dbl>     <dbl>    <dbl>
##  1 mur          0.00141  0.00335      0.421 6.74e- 1
##  2 rob          0.00323  0.000152    21.3   8.81e-85
##  3 incarc_rate  0.00130  0.000127    10.2   1.84e-23
##  4 pb1064       0.0581   0.0151       3.85  1.26e- 4
##  5 pw1064       0.0274   0.00752      3.64  2.85e- 4
##  6 pm1029       0.0167   0.0149       1.13  2.61e- 1
##  7 pop          0.00764  0.00252      3.03  2.47e- 3
##  8 avginc      -0.0109   0.00673     -1.62  1.05e- 1
##  9 density     -0.243    0.0154     -15.8   7.62e-51
## 10 shall       -0.282    0.0277     -10.2   2.29e-23
```

#With time fixed effects shall issue laws will reduce crime rate by 28% approximately

```
fixef(fixedeffects_time)
```

```
##     77     78     79     80     81     82     83     84     85     86     87
## 3.0816 3.1148 3.1464 3.1065 3.0675 3.0740 3.0709 3.1265 3.1570 3.1965 3.2041
##     88     89     90     91     92     93     94     95     96     97     98
## 3.2195 3.2174 3.2146 3.2079 3.2358 3.2362 3.2169 3.2027 3.1861 3.2082 3.1834
##     99
## 3.1330
```

#The estimator of shall issue laws seems to be biased because it is overstating that passing shall issue laws reduce crime rate by 28 %. This might be a problem because may be the government passed laws across some states which are more prone to crimes. In such a case as we are ignoring this nature it might be correlated with explanatory variable, and the exact interpretation of reduce in crime rate estimation , how much it is coming from error terma nd how much is coming from explanatory variable cannot be seperated. As a result the estimator will be biased. In such a case use fixed effects model varying across each state and also over time

```
summary(fixedeffects_time)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = vio ~ mur + rob + incarc_rate + pb1064 + pw1064 +
##     pm1029 + pop + avginc + density + shall, data = guns, model = "within",
##     index = "year")
##
## Balanced Panel: n = 23, T = 51, N = 1173
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -1.527529 -0.195365  0.037854  0.208111  1.135199
```

```
## 
## Coefficients:
##               Estimate  Std. Error  t-value  Pr(>|t|)
## mur          0.00140940  0.00334996   0.4207 0.6740373
## rob          0.00322683  0.00015175  21.2647 < 2.2e-16 ***
## incarc_rate  0.00129558  0.00012694  10.2059 < 2.2e-16 ***
## pb1064       0.05807788  0.01509196   3.8483 0.0001256 ***
## pw1064       0.02737456  0.00751994   3.6403 0.0002846 ***
## pm1029       0.01673708  0.01487321   1.1253 0.2606914
## pop          0.00763562  0.00251673   3.0339 0.0024684 **
## avginc      -0.01091003  0.00672663  -1.6219 0.1050975
## density     -0.24283666  0.01539637 -15.7723 < 2.2e-16 ***
## shall       -0.28193312  0.02768758 -10.1827 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Total Sum of Squares:    474.53
## Residual Sum of Squares: 134.42
## R-Squared:      0.71674
## Adj. R-Squared: 0.70879
## F-statistic: 288.457 on 10 and 1140 DF, p-value: < 2.22e-16
```

#Entity time fixed effects:

```
fixedeffects_entity_time <- plm(vio~mur + rob + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc +
tidy(fixedeffects_entity_time)
```

```
## # A tibble: 10 x 5
##    term         estimate std.error statistic  p.value
##    <chr>           <dbl>     <dbl>     <dbl>    <dbl>
##  1 mur          0.00502  0.00173       2.90  3.85e- 3
##  2 rob          0.00162  0.0000972    16.6   1.23e-55
##  3 incarc_rate  0.000118 0.0000828     1.43  1.54e- 1
##  4 pb1064       0.0753   0.0158        4.78  2.01e- 6
##  5 pw1064       0.0321   0.00434       7.40  2.69e-13
##  6 pm1029      -0.0403   0.00564      -7.14  1.64e-12
##  7 pop          0.00893  0.00747       1.19  2.32e- 1
##  8 avginc      -0.0000537 0.00523     -0.0103 9.92e- 1
##  9 density     -0.0373   0.0728       -0.512  6.09e- 1
## 10 shall       -0.0556   0.0161       -3.46   5.59e- 4
```

#With entity time fixed effects the estimator of shall seems to be reliable and also it is highly significant. Also the estimator is unbiased and consistent. However the standard errors are not correct. We can estimate or we can calculate correct standard errors using Robust Stnadrd errors

```
summary(fixedeffects_entity_time)
```

```
## Oneway (individual) effect Within Model
## 
## Call:
## plm(formula = vio ~ mur + rob + incarc_rate + pb1064 + pw1064 +
##     pm1029 + pop + avginc + density + shall, data = guns, model = "within",
```

```
##       index = c("stateid", "year"))
##
## Balanced Panel: n = 51, T = 23, N = 1173
##
## Residuals:
##        Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.6022695 -0.0706786  0.0068197  0.0780594  0.5381769
##
## Coefficients:
##                 Estimate  Std. Error t-value  Pr(>|t|)
## mur           5.0171e-03  1.7323e-03  2.8963 0.0038501 **
## rob           1.6168e-03  9.7213e-05 16.6311 < 2.2e-16 ***
## incarc_rate   1.1800e-04  8.2803e-05  1.4251 0.1544210
## pb1064        7.5260e-02  1.5753e-02  4.7776 2.011e-06 ***
## pw1064        3.2146e-02  4.3443e-03  7.3997 2.685e-13 ***
## pm1029       -4.0324e-02  5.6448e-03 -7.1435 1.642e-12 ***
## pop           8.9291e-03  7.4734e-03  1.1948 0.2324268
## avginc       -5.3739e-05  5.2293e-03 -0.0103 0.9918025
## density      -3.7272e-02  7.2797e-02 -0.5120 0.6087481
## shall        -5.5627e-02  1.6074e-02 -3.4606 0.0005594 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    36.789
## Residual Sum of Squares: 20.63
## R-Squared:      0.43922
## Adj. R-Squared: 0.40897
## F-statistic: 87.0967 on 10 and 1112 DF, p-value: < 2.22e-16
```

#We cannot use Random Effects for this model because these entities are not coming from Random Population. All these entities are states in the country United states and they are not random in nature

#As of now we observed that if shall issue laws were affecting the crime rate. However the robbery and murder rates are also considered as crimes and we further analyse that if the shall issue laws are playing any significant role in affecting these two

#Robbery Rate:

```
ggplot(guns,aes(x=rob)) + geom_histogram() + ggtitle("Average Robbery Rate across United States per 100
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

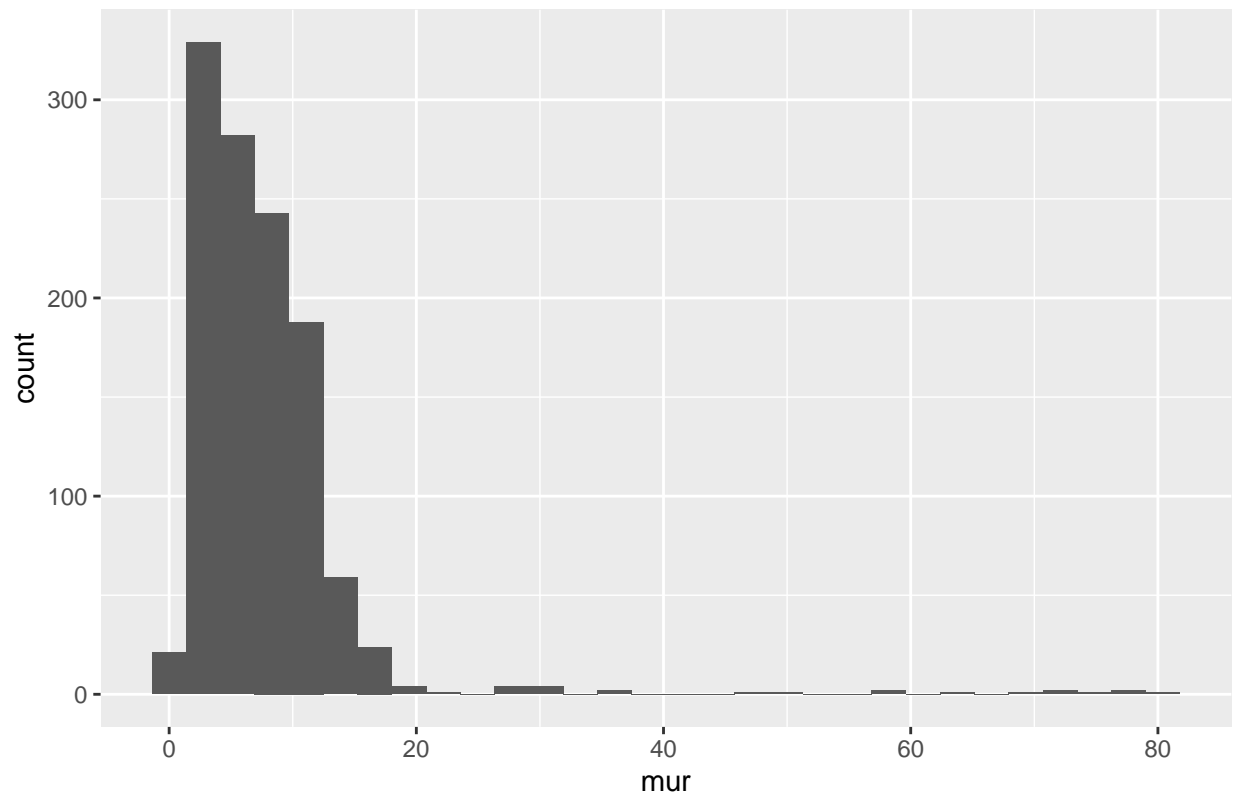## Average Robbery Rate across United States per 100,000 incidents


```r
ggplot(guns,aes(x=mur)) + geom_histogram() + ggtitle("Average Murder Rate across United States per 100,0
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
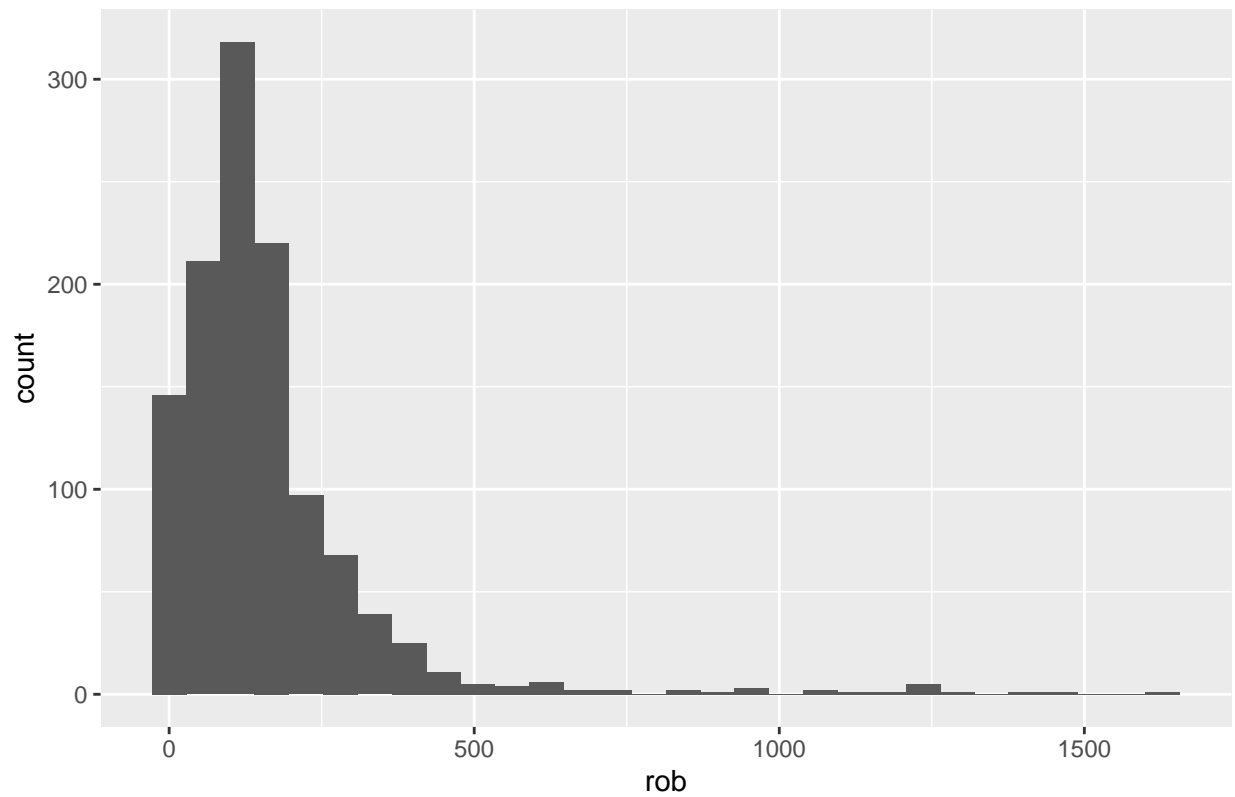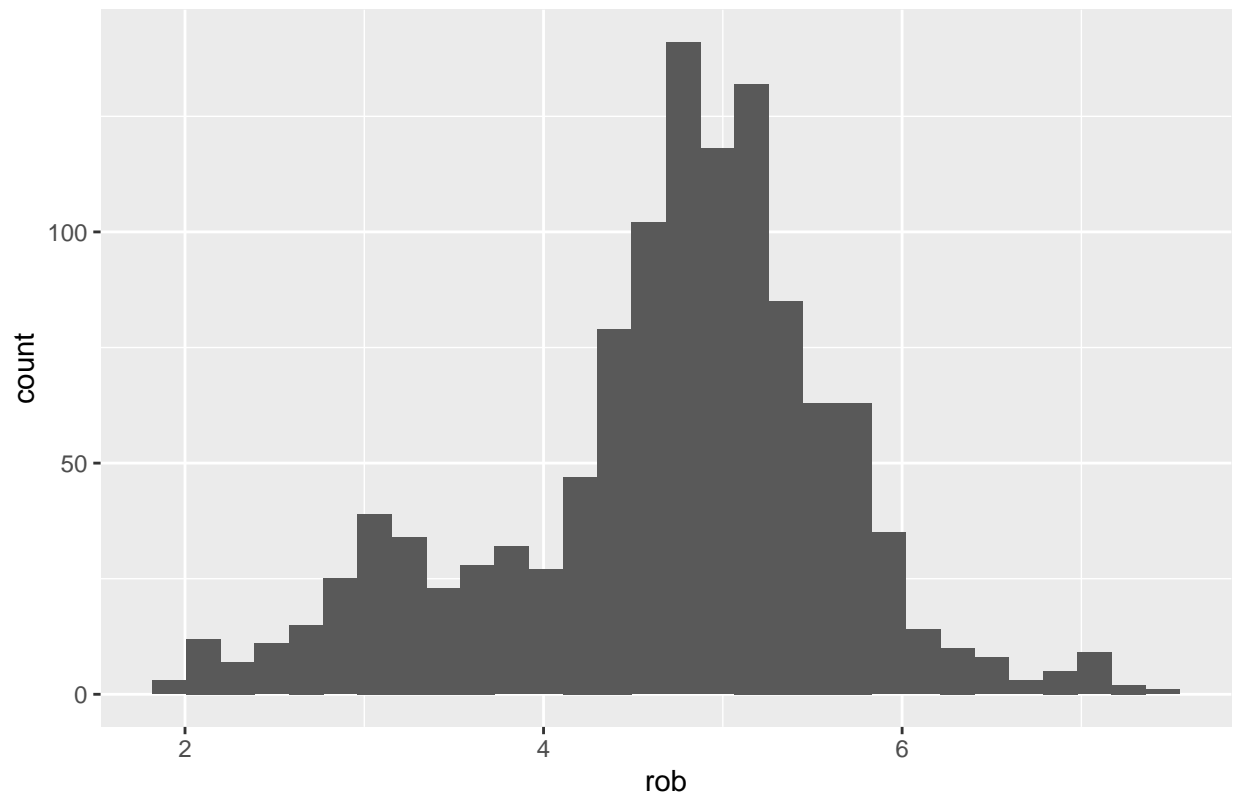
Average Murder Rate across United States per 100,000 incidents


```
guns$rob<-log(guns$rob)
guns$mur<-log(guns$mur)
ggplot(guns,aes(x=rob)) + geom_histogram() + ggtitle("Average Robbery Rate across United States per 100
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
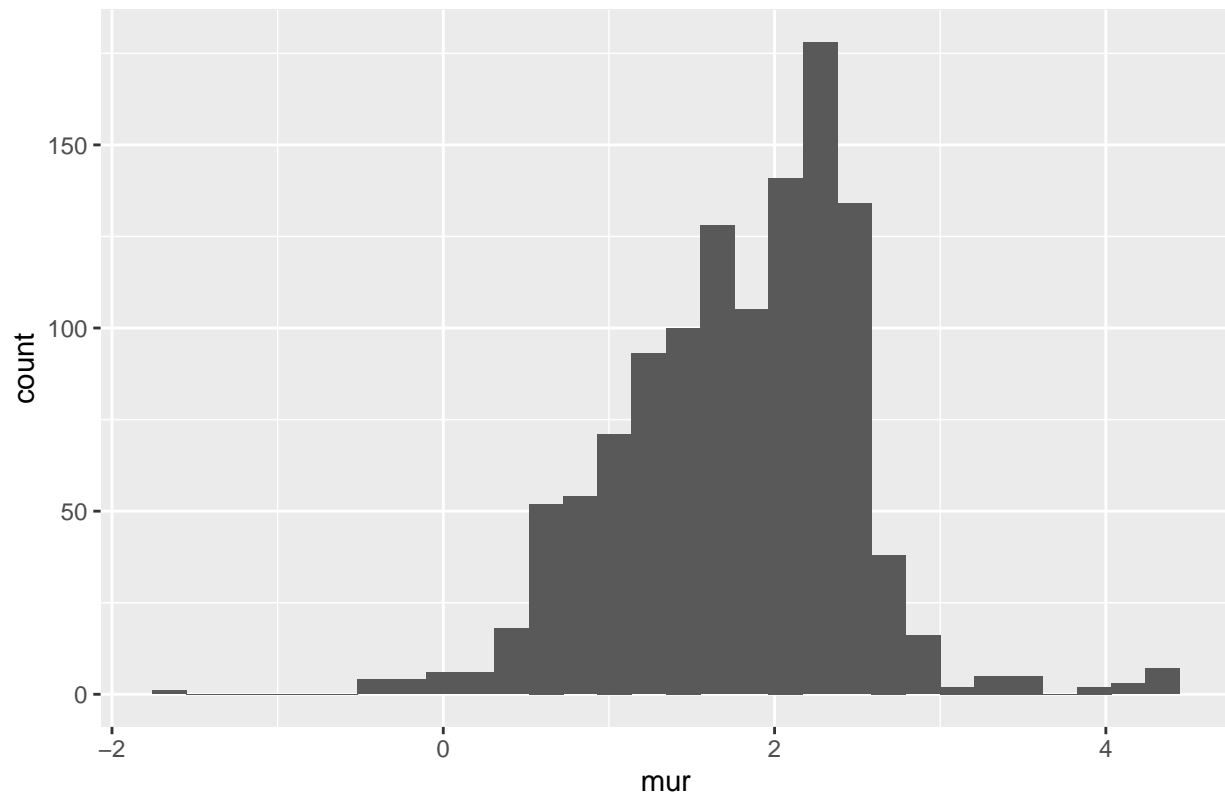
Average Robbery Rate across United States per 100,000 incidents


```
ggplot(guns,aes(x=mur)) + geom_histogram() + ggtitle("Average Murder Rate across United States per 100,0
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Average Murder Rate across United States per 100,000 incidents


```
skewness(guns$rob)
```

```
## [1] -0.5196271
```

```
skewness(guns$mur)
```

```
## [1] -0.135789
```

```
fixedeffectsrob_entity_time <- plm(rob~mur + incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + der
tidy(fixedeffectsrob_entity_time)
```

```
## # A tibble: 9 x 5
##   term          estimate std.error statistic  p.value
##   <chr>            <dbl>     <dbl>     <dbl>    <dbl>
## 1 mur           0.399      0.0269     14.8   1.29e-45
## 2 incarc_rate   0.0000672  0.000115    0.584 5.59e- 1
## 3 pb1064        0.0993     0.0217      4.57  5.49e- 6
## 4 pw1064        0.0231     0.00622     3.71  2.17e- 4
## 5 pm1029       -0.00446    0.00791    -0.564 5.73e- 1
## 6 pop           0.0266     0.0107      2.48  1.31e- 2
## 7 avginc       -0.0272     0.00726    -3.75  1.87e- 4
## 8 density       0.0813     0.106       0.769 4.42e- 1
## 9 shall         0.0164     0.0231      0.710 4.78e- 1
```

#Generally , we observe that the robbery rate also will reduce if we pass any strict laws. However our model is estimating that passing shall issue laws will increase the robbery rate by 1.6%. However this seems strange and lets do Hypothesis testing. p-value is 0.47 which is gretaer than 0.05. So we are unable to reject Null Hypothesis and conclude that there is no sufficient evidence to conclude that shall issue laws increase robbery rate and the estimator is not significant and it is as expected.

#Further lets see, if shall issue laws are impacting murder rate as well

```
fixedeffectsmur_entity_time <- plm(mur~incarc_rate + pb1064 + pw1064 + pm1029 + pop + avginc + density
tidy(fixedeffectsmur_entity_time)
```

```
## # A tibble: 8 x 5
##   term          estimate std.error statistic      p.value
##   <chr>            <dbl>     <dbl>     <dbl>        <dbl>
## 1 incarc_rate  -0.000360  0.000128     -2.82 0.00493
## 2 pb1064        0.0307    0.0242        1.27 0.206
## 3 pw1064        0.0103    0.00693       1.49 0.136
## 4 pm1029        0.0392    0.00874       4.49 0.00000793
## 5 pop          -0.0257    0.0119       -2.16 0.0311
## 6 avginc        0.0243    0.00807       3.01 0.00264
## 7 density      -0.671     0.116        -5.78 0.00000000985
## 8 shall        -0.0608    0.0258       -2.36 0.0184
```

#From the above analysis, we see that passing shall issue laws will reduce the murder rate by 6% and it is significant at 5%

#So, Based on these analysis, shall issue laws will reduce the crime rate in United States by approximately 5%