How do drunk driving laws impact traffic deaths?

Econometrics Project

<u>Objective</u>:

The objective of this project is to estimate the impact of drunk driving laws on traffic deaths.

<u>Dataset Description</u>:

The dataset contains **balanced panel data** from the "lower 48" U.S. states (excluding Alaska and Hawaii), annually from 1982 through 1988. Each state has the same number of observations.

<u>Dependent Variable:</u>

The traffic fatality rate is that the number of traffic deaths per 10,000 people for a given state during a given year. By looking the variables within the dataset, "mrall", which is that the Vehicle fatality rate (VFR), would represent this number best in comparison to other variables within the dataset. Additionally, "mralln", which is that the night-time VFR, are going to be used as a variable to check the effect of alcohol laws on changes in night-time VFR.

<u>Independent Variables:</u>

Our objective is to estimate the impact of the drinking-and-driving laws on traffic deaths. The variables "jaild" and "comserd" describe the states' way of handling drinking-and-driving cases and are therefore considered essential to incorporate within the model. Additionally, there are other factors we must consider which will or might not have a significant effect on the VFR. These are listed below:

Variable	Descriptions
jaild	Mandatory Jail Sentence: one if state requires jail time.
comserd	Mandatory Community Service: one if state requires community service
spircons	Per Capita Pure Alcohol Consumption (Annual, Gallons)
unrate	State Unemployment Rate (%)
perinc	Per Capita Personal Income ($)
beertax	Tax on Case of Beer ($)
sobapt	% Southern Baptist
mormon	% Mormon
mlda	Minimum Legal Drinking Age (years)
dry	% Residing in Dry Counties: A dry county is a county whose government forbids the sale of any kind of alcoholic beverages. Some prohibit off-premises
yngdrv	% of Drivers Aged 15-24
vmiles	Ave. Mile per Driver
pop	Population
miles	total vehicle miles (millions)
gspch	GSP Rate of Change: This is a measure of economic growth

About the variables:

1. Jaild: This variable tells us about the punishment which describe the state's minimum sentencing requirements for an initial drunk driving conviction. Equals one, if the state requires jail time or otherwise equals zero.
2. Comserd: This variable tells us about the punishment which describe the state's minimum sentencing requirements for an initial drunk driving conviction. Equals one, if the state requires Mandatory Community Service or otherwise equals zero.
3. Spircons: This variable tells us about the per capita annual Pure Alcohol Consumption in gallons across years for each state.
4. Unrate: This variable tells us about the % State Unemployment Rate for each state across years.
5. Perinc: This variable tells us about the Per Capita Personal Income for each state across years.
6. Beertax: This variable tells us about the Tax on Case of Beer in dollars for each state across years.
7. Sobapt: This variable tells us about the % Southern Baptists for each state across years.
8. Mormon: This variable tells us about the % Mormon for each state across years.
9. Mlda: This variable tells us about the minimum legal drinking age in years for each state across years.
10. Dry: This variable tells us about the % of population residing in dry counties for each state across years.
11. Yngdrv: This variable tells us about the % of drivers aged 15-24 age for each state across years.
12. Vmiles: This variable tells us about the average miles per driver for each state across years.
13. Pop: This variable tells us about the population for each state across years.
14. Miles: This variable tells us about the total vehicle miles in millions for each state across years.
15. Gspch: This variable tells us about the GSP rate of change for each state across years.

Exploratory Data Analysis:

I have done exploratory analysis on the data to understand the trends in the data over the years.

I focused mainly on the variables "mrall", "mralln" which constitute the major category of Vehicle Fatality rate. I wanted to explore its impact on different states over the year using visualizations and descriptive analysis.

Descriptive Statistics: The following table shows the descriptive statistics of our data.

```
In [19]: data.describe()
Out[19]:
```

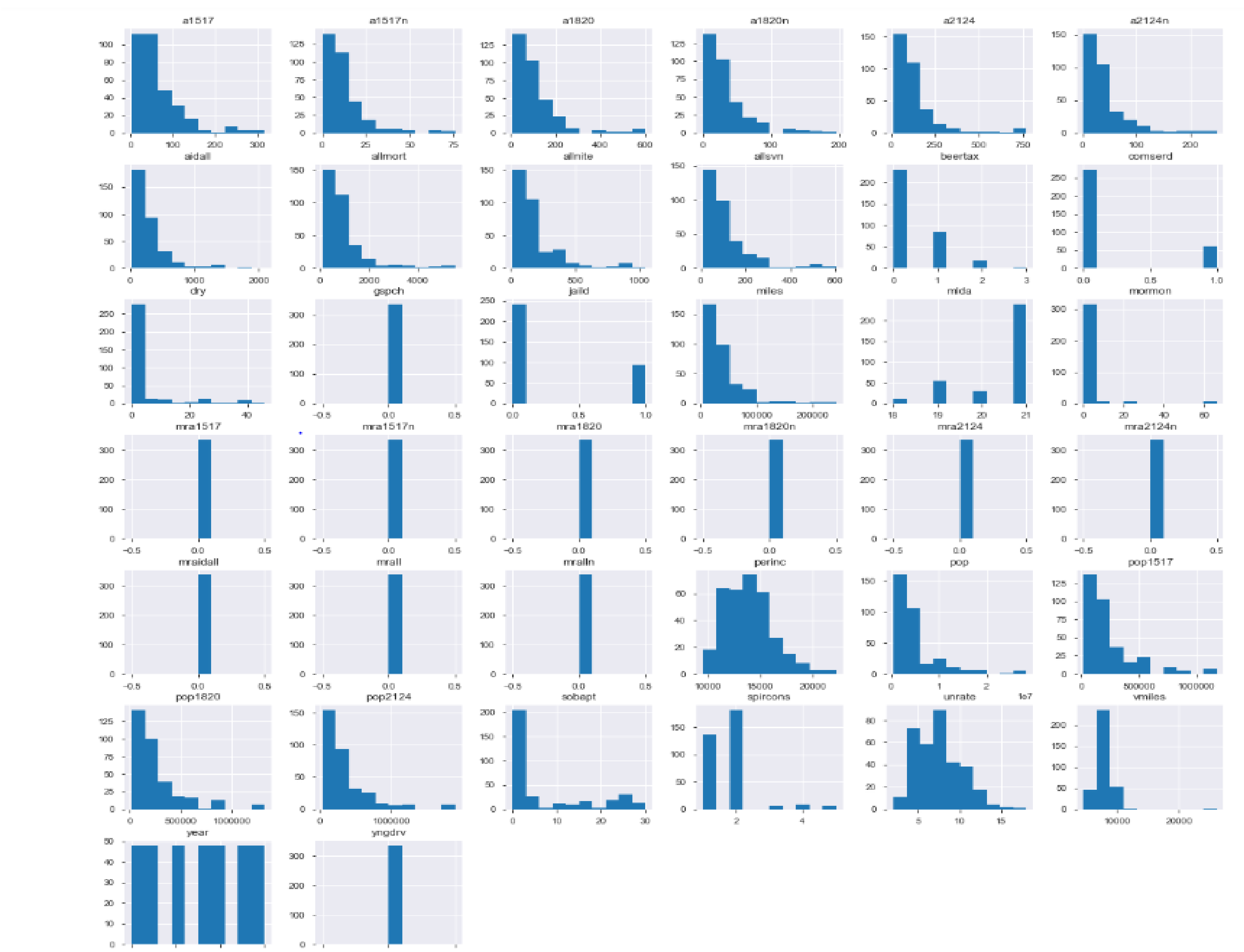
	year	spircons	unrate	perinc	beertax	sobapt	mormon	mlda	dry	yngdrv	...	a2124n	mra2124n	
count	336.000000	336.000000	336.000000	336.000000	336.000000	336.000000	336.000000	336.000000	336.000000	336.0	...	336.000000	336.0	3
mean	1985.000000	1.711310	7.380952	13880.175595	0.386905	7.154762	2.657738	20.473214	4.238095	0.0	...	41.377976	0.0	2
std	2.002983	0.770728	2.564533	2253.041090	0.631753	9.733149	9.702383	0.890680	9.518082	0.0	...	42.930315	0.0	3
min	1982.000000	1.000000	2.000000	9514.000000	0.000000	0.000000	0.000000	18.000000	0.000000	0.0	...	1.000000	0.0	[illegible]
25%	1983.000000	1.000000	5.750000	12085.750000	0.000000	1.000000	0.000000	20.000000	0.000000	0.0	...	13.000000	0.0	[illegible]
50%	1985.000000	2.000000	7.000000	13763.000000	0.000000	2.000000	0.000000	21.000000	0.000000	0.0	...	30.000000	0.0	2
75%	1987.000000	2.000000	9.000000	15174.750000	1.000000	13.250000	1.000000	21.000000	2.250000	0.0	...	49.000000	0.0	3
max	1988.000000	5.000000	18.000000	22193.000000	3.000000	30.000000	66.000000	21.000000	46.000000	0.0	...	249.000000	0.0	20

8 rows x 38 columns

```
In [23]: pd.isnull(data).any()

Out[23]: state        False
         year         False
         spircons     False
         unrate       False
         perinc       False
         beertax      False
         sobapt       False
         mormon       False
         mlda         False
         dry          False
         yngdrv       False
         vmiles       False
         jaild        False
         comserd      False
         allmort      False
         mrall        False
         allnite      False
         mralln       False
         allsvn       False
         a1517        False
         mra1517      False
         a1517n       False
         mra1517n     False
         a1820        False
         a1820n       False
         mra1820      False
         mra1820n     False
         a2124        False
         mra2124      False
         a2124n       False
         mra2124n     False
         aidall       False
```

It can be seen that the data set is clean and has no missing values. For our analysis we focus on mostly mrall and mralln and try to derive inference from their statistics.

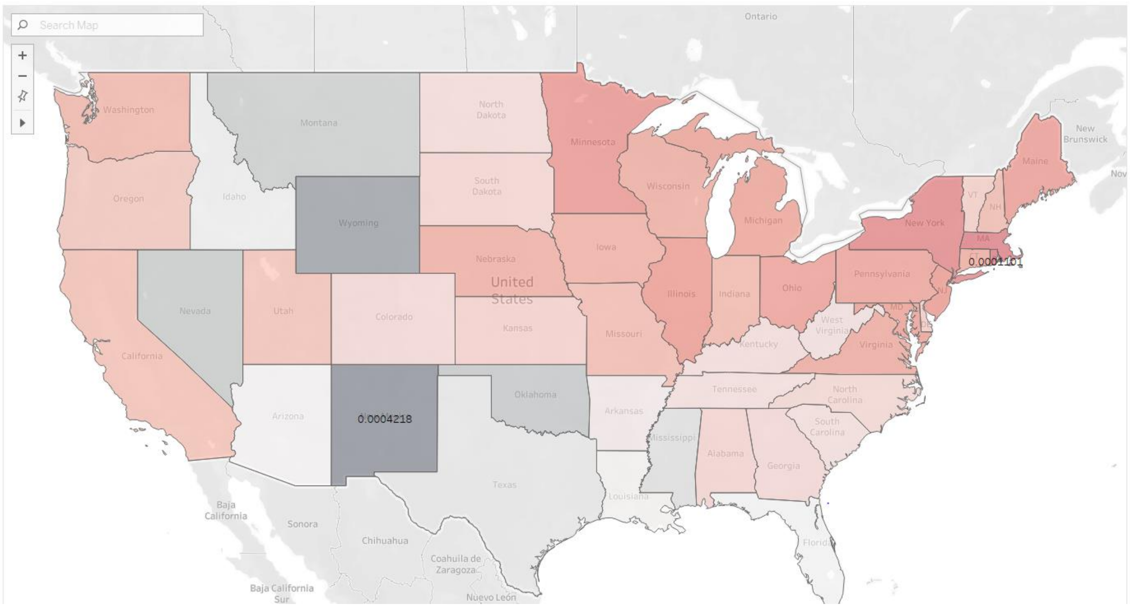
Further the histogram for the distribution of all other variables can be seen below:



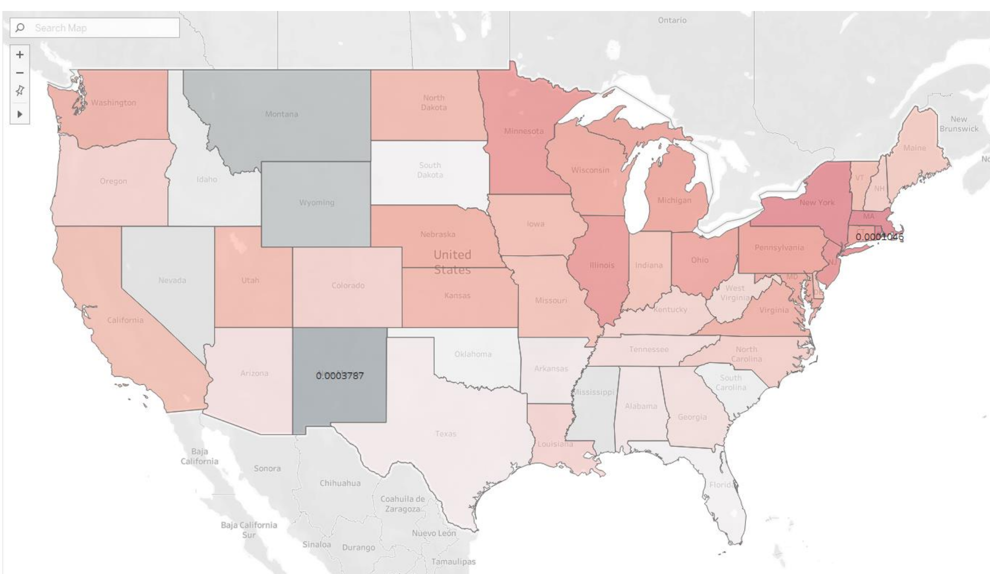
Data Visualization:

Vehicle Fatality Rate (VFR):

The following map shows the general vehicle fatality rate on a state-by-state basis. From the graph, we will see that New Mexico has the very best average VFR of 3.653 per 10,000 people and Rhode Island has lowest average VFR of 1.11 per 10,000 people. States like New Mexico, Wyoming, Montana, South Carolina, Nevada, Arizona have high VFRs whereas states like Massachusetts, Minnesota, New York, New Jersey and Illinois have low VFRs.



1982 Year



1983 Year

Correlation Matrix:

```
corr.style.background_gradient(cmap='coolwarm')
```

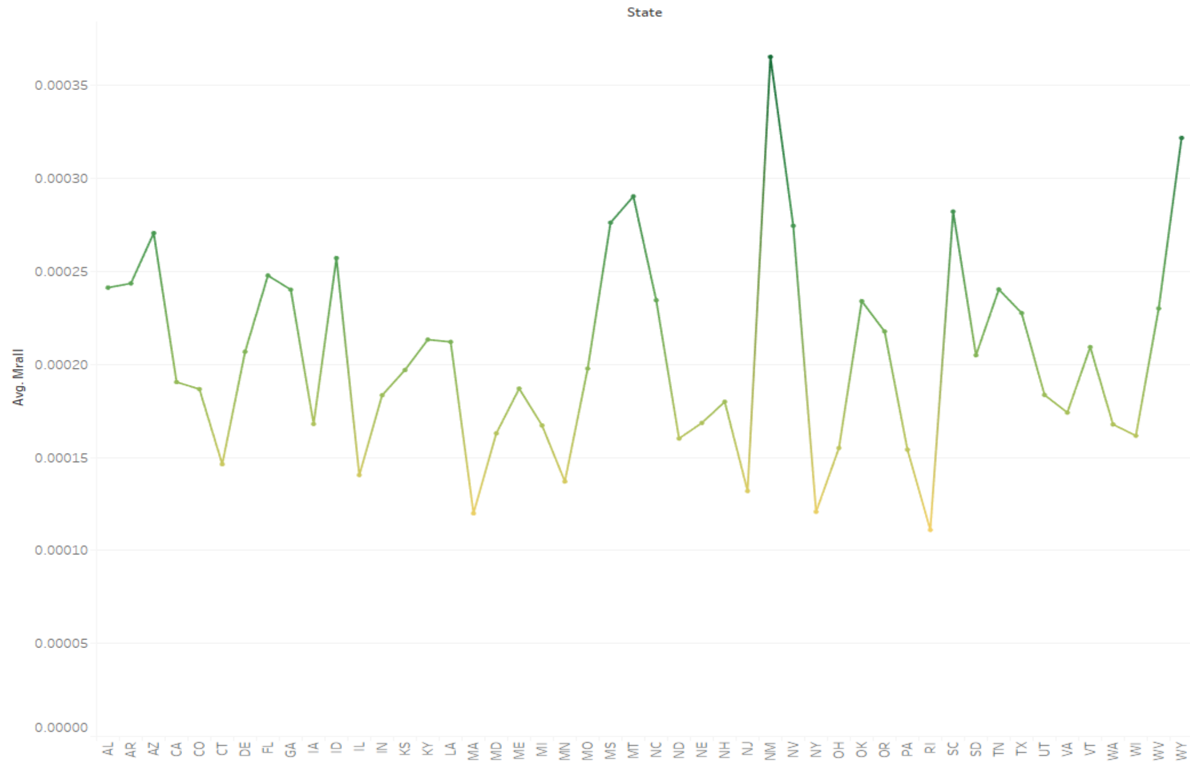
Out[53]:

	jaild	comserd	spircons	unrate	perinc	beertax	sobapt	mormon	mlda	dry	vmiles	pop	pop'
jaild	1	0.523551	0.0105011	0.13884	-0.15074	-0.0471163	-0.0798951	0.217752	-0.106982	-0.209173	0.0715823	-0.251484	-0.25
mserd	0.523551	1	0.139174	-0.0632016	0.0521705	0.0966622	-0.036776	0.247252	0.0324496	-0.18785	-0.022817	-0.103863	-0.12
ircons	0.0105011	0.139174	1	-0.23718	0.397338	0.0155091	-0.319927	-0.12742	-0.0569567	-0.297417	-0.0493831	-0.0741762	-0.083
ınrate	0.13884	-0.0632016	-0.23718	1	-0.555099	0.0266719	0.274361	-0.00878058	-0.256889	0.258222	-0.284492	0.0798847	0.1
perinc	-0.15074	0.0521705	0.397338	-0.555099	1	-0.354969	-0.480291	-0.211287	0.198725	-0.343278	-0.0801897	0.365592	0.34
eertax	-0.0471163	0.0966622	0.0155091	0.0266719	-0.354969	1	0.588321	0.0265386	-0.0557974	0.207532	0.139549	-0.115851	-0.10
obapt	-0.0798951	-0.036776	-0.319927	0.274361	-0.480291	0.588321	1	-0.143926	0.0510968	0.570285	0.135819	0.000755554	0.019
ormon	0.217752	0.247252	-0.12742	-0.00878058	-0.211287	0.0265386	-0.143926	1	0.011544	-0.0935338	0.00457905	-0.150709	-0.15
mlda	-0.106982	0.0324496	-0.0569567	-0.256889	0.198725	-0.0557974	0.0510968	0.011544	1	0.13984	0.0568166	0.0586122	0.039
dry	-0.209173	-0.18785	-0.297417	0.258222	-0.343278	0.207532	0.570285	-0.0935338	0.13984	1	-0.0828148	0.0255565	0.043
vmiles	0.0715823	-0.022817	-0.0493831	-0.284492	-0.0801897	0.139549	0.135819	0.00457905	0.0568166	-0.0828148	1	-0.241396	-0.26
pop	-0.251484	-0.103863	-0.0741762	0.0798847	0.365592	-0.115851	0.000755554	-0.150709	0.0586122	0.0255565	-0.241396	1	0.99
p1517	-0.258303	-0.126165	-0.0834077	0.12091	0.343274	-0.104245	0.0190509	-0.158188	0.0398907	0.0436359	-0.263244	0.994531	
p1820	-0.258137	-0.121882	-0.0790848	0.116792	0.336737	-0.100287	0.0222156	-0.155126	0.0398769	0.0408909	-0.258541	0.99494	0.9
p2124	-0.251105	-0.113438	-0.0697343	0.113684	0.338212	-0.106373	0.0141582	-0.147051	0.0354554	0.0310056	-0.250103	0.994244	0.99
miles	-0.247996	-0.092579	-0.0957178	0.0384506	0.345614	-0.0742667	0.0661559	-0.148399	0.0814845	0.0281362	-0.103473	0.973689	0.95

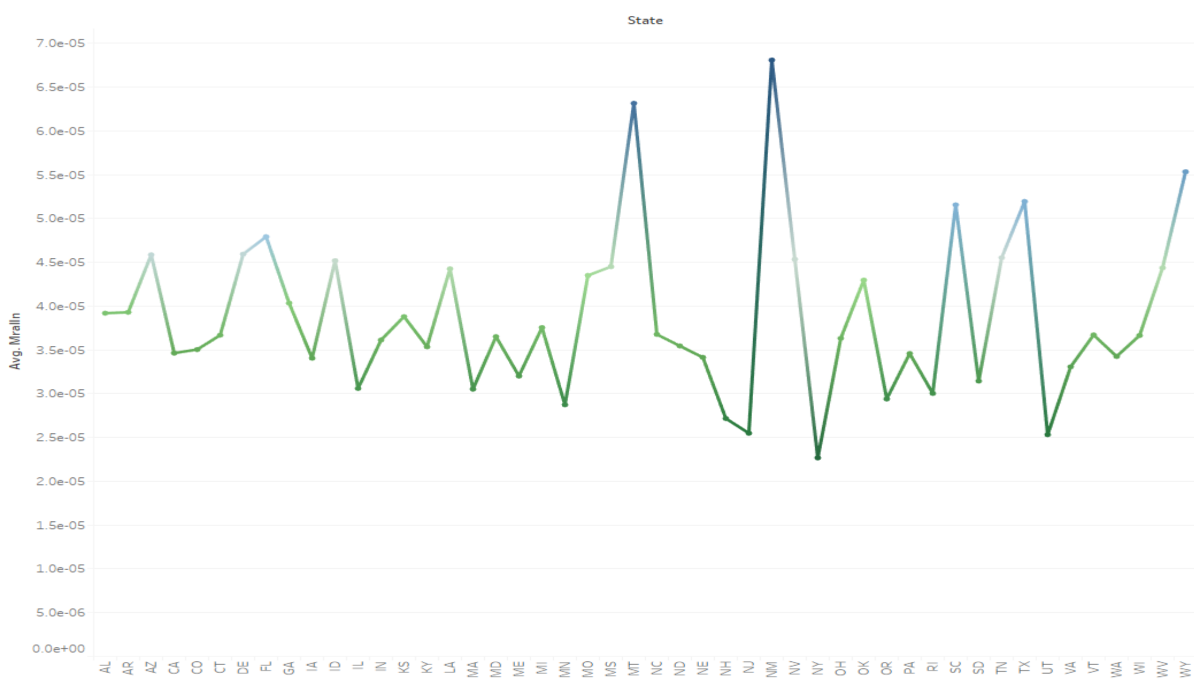
We can infer the following observations from the above correlation matrix:

- As we will see from the above matrix, the variables for population and total vehicle miles (in millions) are highly correlated which is expected: because as the population of a state increases then the number of people driving in that state also increases which results in higher total vehicle miles.
- From the correlation matrix, we can observe per capita income and unemployment are negatively correlated, which aligns with practical theory, because as the per capita income increases, unemployment goes down.
- From the correlation matrix, we can observe that Per Capita Pure Alcohol Consumption (Annual, Gallons) spircons is negatively correlated with states with high % Southern Baptist and % Mormom in the state, which is as excepted.
- From the correlation matrix, we can observe that, the variable spircons are negatively correlated with dry variables, as it makes sense that alcohol consumption in dry county is very less compared to other counties.
- Lastly, from the correlation matrix, we can observe that it appears that either states with high populations of Southern Baptists tend to have higher beer taxes as seen by the high correlation between the two variables.

Average Vehicle Fatality Rate across states:



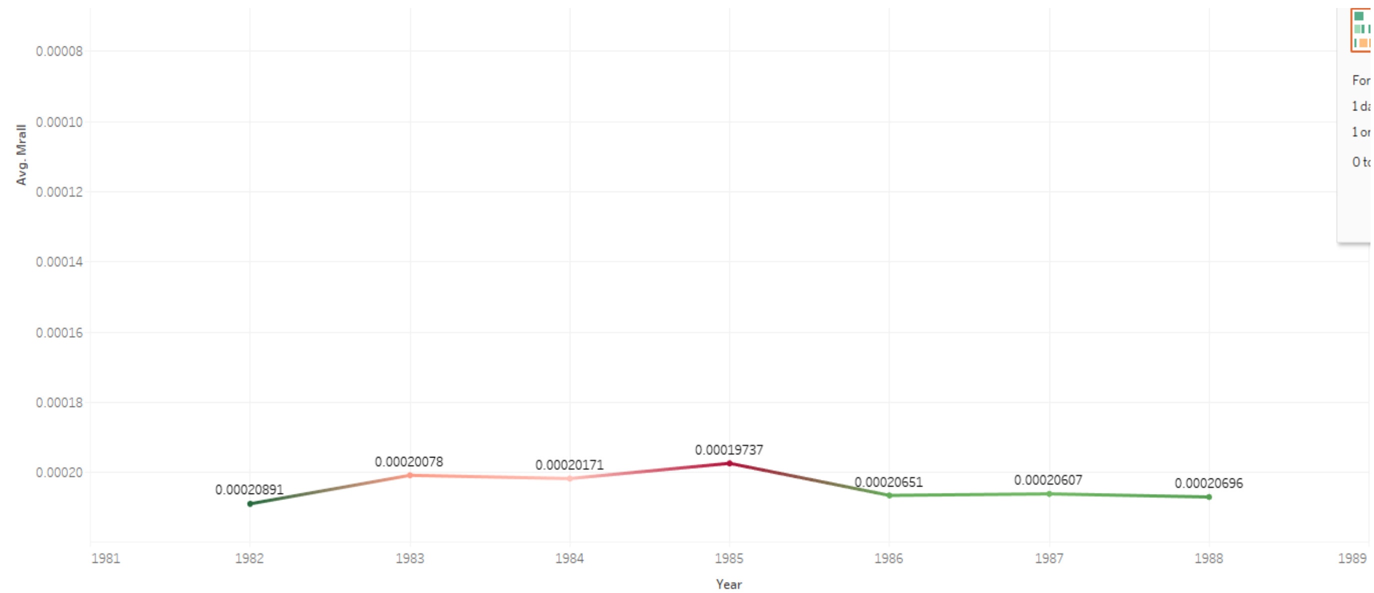
Average Mrall rate across States



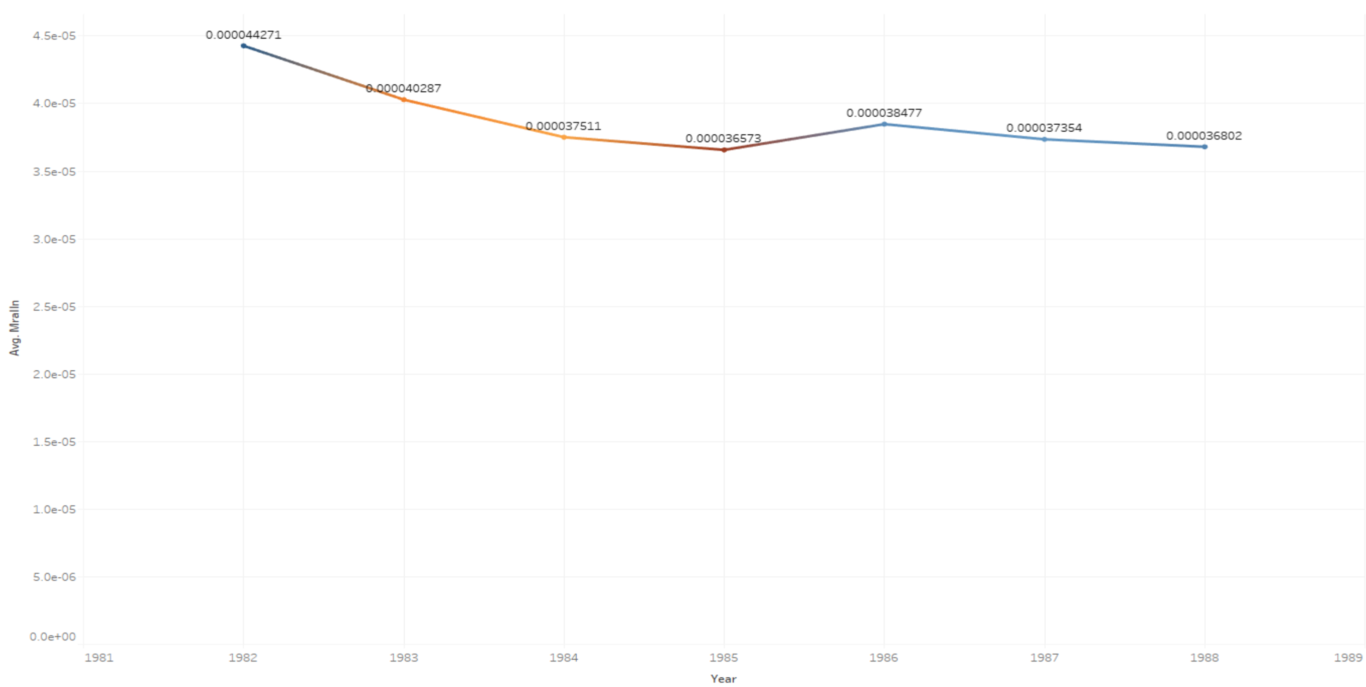
Average Mrall rate at night across States

The graphs above display the average of Mrall and Mralln across states. It indicates that the state New Mexico has the highest Vehicle Fatality Rate(Mrall and Mralln) as compared to others.

Average Vehicle Fatality Rate across years:



Average Mrall across Years

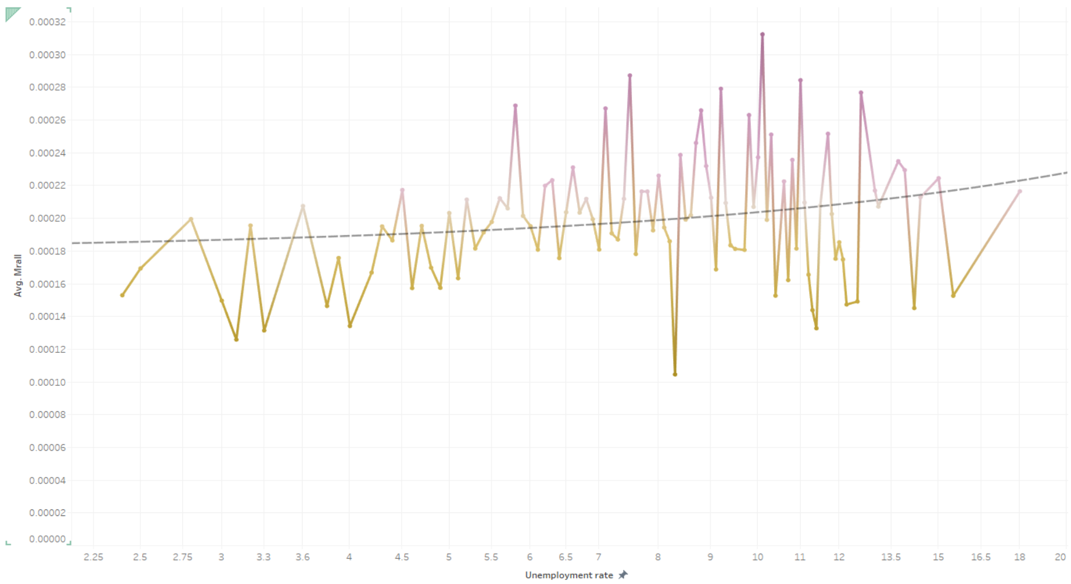


Average Mrall rate at night across Years

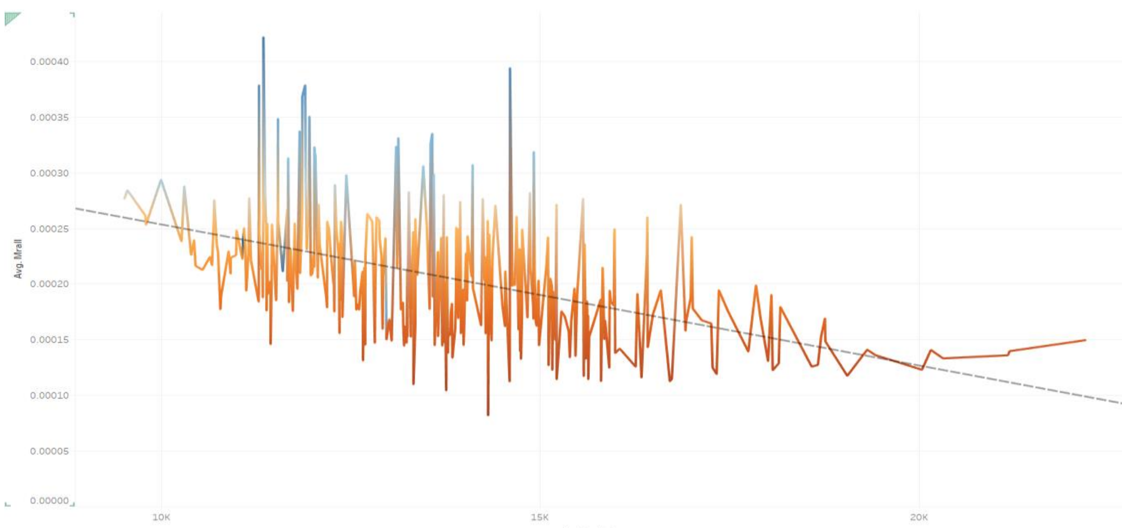
The graphs above display the crime rate trends across the years 1982 - 1988. It indicates that Mrall is uniform across years peaking in year 1985, lowest being 1988 and surprisingly Mralln has a low in 1985 as well, highest being 1982 and lowest being 1988.

Unemployment Rate vs VFR:

Using my knowledge of economic theory, I postulate that as the unemployment rate increases people tend to consume more alcohol due to more stress including, but not limited to, family pressure and personal emotional stress etc. Additionally, people who are unemployed now have more idle time in which they can drink and are more prone to vehicle accidents. This has been proven by the graph below.

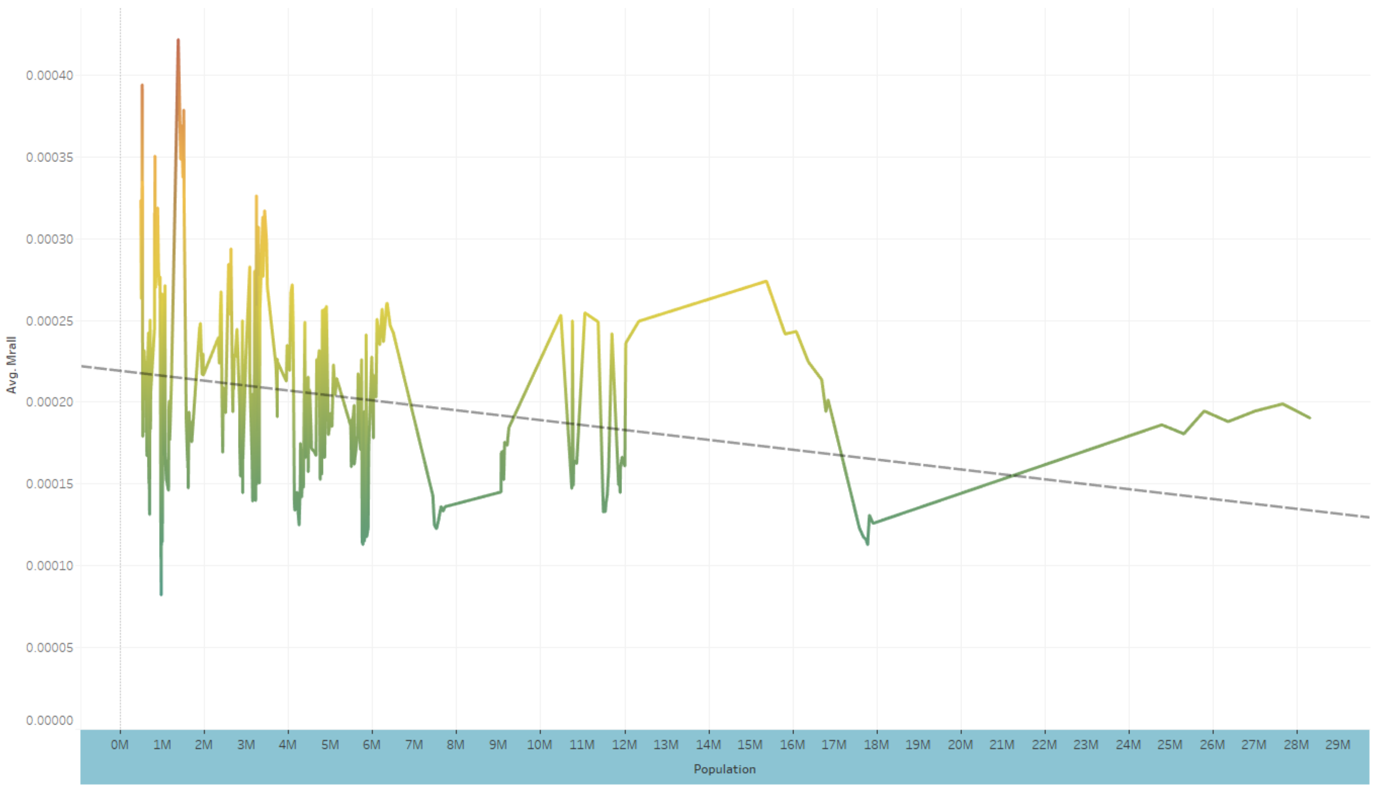


Per Capita Income vs VFR:



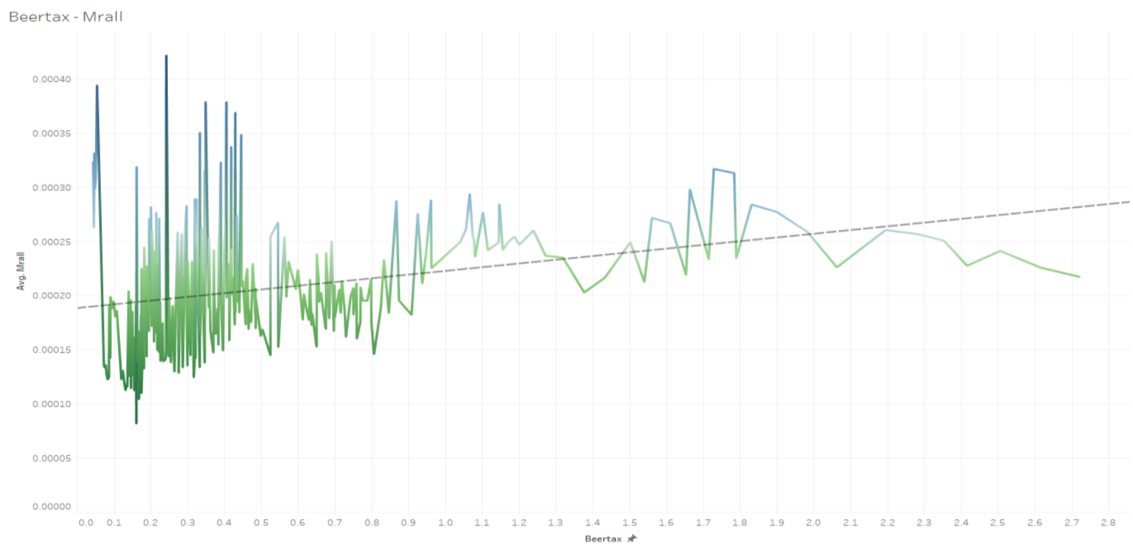
Based on my knowledge of economic theory, I postulate, that the relationship between the per capita income and vehicle fatality rate (VFR) is related to due to per capita income and alcohol consumption. As per my assumption, people with low income, due to the lower income experienced by this group of people which includes loans repayment, family costs, general frustration about inflation and higher levels of stress likely to spend more money on alcohol. It is then easy to ascertain that because the per capita income increases, people tend to possess lower consumption of alcohol as we assume that higher per capita income suggests folks that are more educated, responsible, and usually less stressed than those with low per capita income. What could also be less evident is that the concept once per capita income reaches a particular point, alcohol consumption begins again to extend and that we assume this is often thanks to the very fact that the people on this end of the spectrum are experiencing overly high levels of occupational and potentially familial stress. As they assert, “More money, more problems”! The graph presented below appears to prove our theory correct; it's possible that additional data points with higher per capita income will add more concrete proof to our theory.

Population vs VFR:



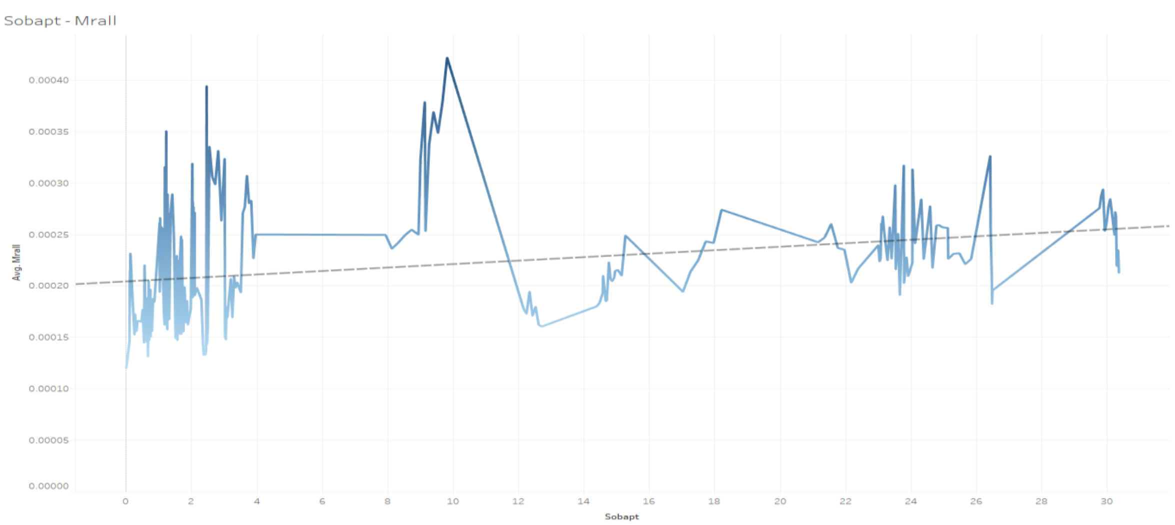
Using my knowledge of economic theory, I postulate that as the population of a region increases, the Vehicle Fatality rate decreases. To support my assumption, we can see that **VFR across years graph** above, that VFR is uniform across years. But when you see that population increases, but VFR being constant, there will be a drop in Mrall rate with increases in population rate because here, numerator will be constant (Mrall) but denominator (Population) will be increasing. This assumption holds true as we can see from the above graph.

BeerTax vs VFR:



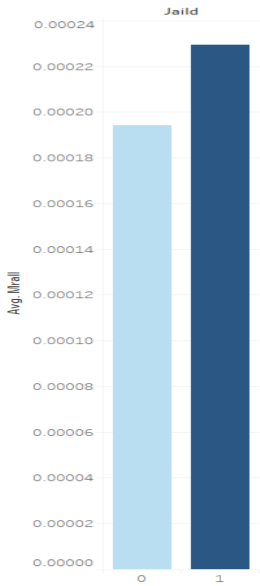
Using my knowledge of economic theory, I postulate that with low beer tax, the alcohol consumption will be very high because many people can afford it at cheaper rates, with higher alcohol consumption, there can be more people driving vehicles in drunken state which can lead to higher VFR. As, we can see from the graph that states with low beer tax have higher VFR as compared to the states with higher beer tax which has low VFR. This assumption holds true as we can see from the above graph.

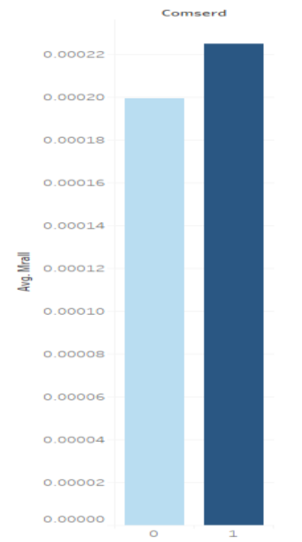
% Southern Baptist vs VFR:



According to my assumption, I postulate that regions with more % of Southern Baptists and Mormons who are teetotalers, the alcohol consumption should be very low, as a result the regions with more % of Southern Baptists should have lower VFR as compared to regions with less % of Southern Baptists. This assumption holds true as we can see from the above graph.

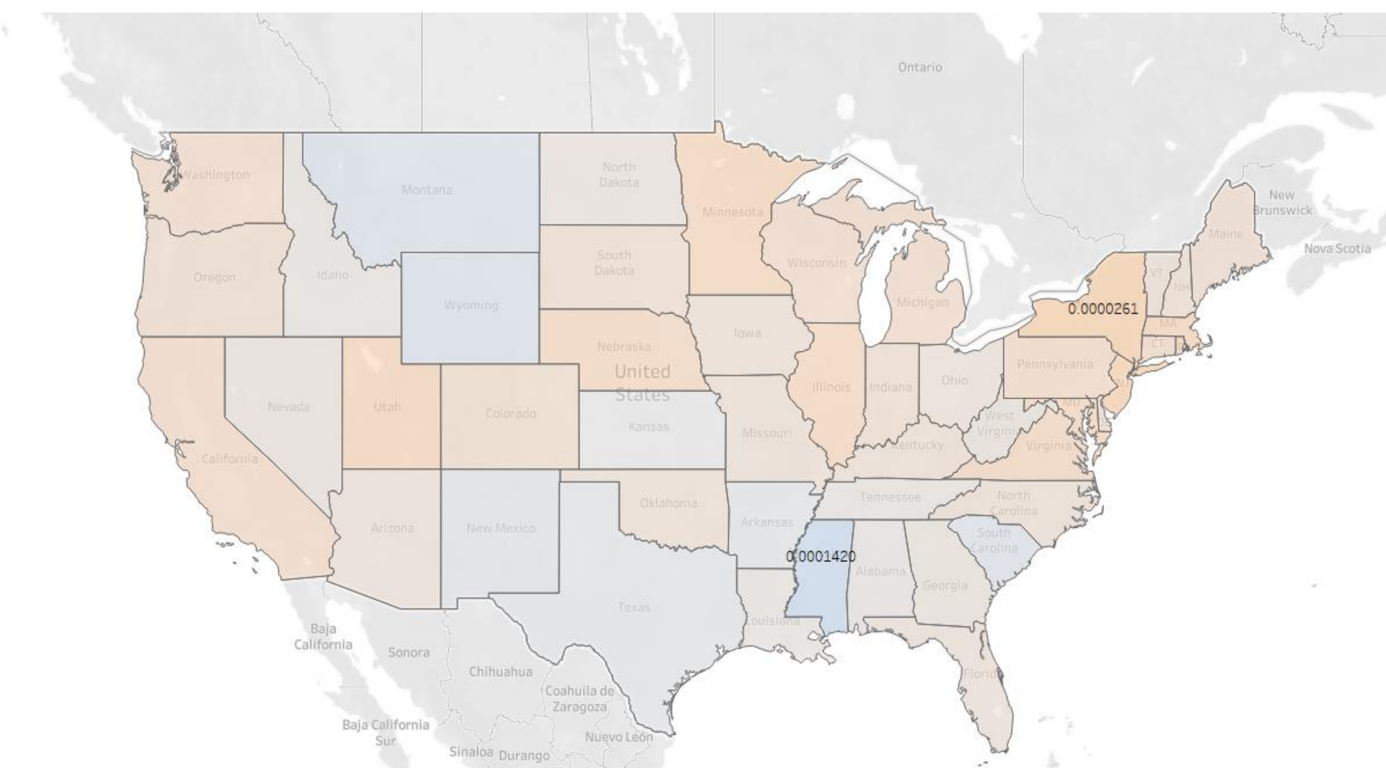
Mandatory Jail Sentence and Mandatory Community Service vs VFR:





According to my assumption, for states that have enacted mandatory jail sentences and/or mandatory community service as a way to discourage drinking and driving, we might expect the effect of those laws to be a decrease in VFR. However, from the above graphs, the effect appears to be completely opposite. this might flow from to reverse causality where mandatory jail sentencing and mandatory community service aren't causing the VFR to extend but instead a higher VFR causes the states to impose mandatory jail sentencing and mandatory community service.

Alcohol-Involved VFR across states:



The graphs above display the average of mraidall across states. It indicates that the state Mississippi has the highest Alcohol-Involved VFR (mraidall) as compared to others.

Various Regressions Models For MRALL:

Pooled Least Square Model:

Running a pooled model corresponding to given data gives following output.

```
.
. reg mrall spircons unrate perinc beertax sobapt mormon mlda dry yngdrv vmiles jaild comserd miles pop gspch

      Source |       SS       df       MS              Number of obs =     335
-------------+------------------------------           F(15, 319)    =   24.56
       Model |  5.8357e-07    15  3.8905e-08           Prob > F      =  0.0000
    Residual |  5.0540e-07   319  1.5843e-09           R-squared     =  0.5359
-------------+------------------------------           Adj R-squared =  0.5141
       Total |  1.0890e-06   334  3.2604e-09           Root MSE      = 4.0e-05

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000227   4.09e-06     5.54   0.000     .0000146    .0000307
      unrate |  -2.15e-06   1.36e-06    -1.59   0.114    -4.83e-06    5.19e-07
      perinc |  -1.28e-08   1.91e-09    -6.71   0.000    -1.66e-08   -9.06e-09
     beertax |  -.0000194   6.95e-06    -2.79   0.006    -.0000331   -5.74e-06
      sobapt |   2.46e-06   4.04e-07     6.07   0.000     1.66e-06    3.25e-06
      mormon |   1.13e-07   2.63e-07     0.43   0.669    -4.05e-07    6.31e-07
        mlda |  -2.60e-07   2.71e-06    -0.10   0.924    -5.58e-06    5.06e-06
         dry |  -5.05e-07   3.19e-07    -1.58   0.115    -1.13e-06    1.23e-07
      yngdrv |   .0000411   .0001112     0.37   0.712    -.0001777      .00026
      vmiles |   1.09e-08   2.00e-09     5.45   0.000     6.97e-09    1.49e-08
       jaild |   .0000224   6.49e-06     3.45   0.001     9.65e-06    .0000352
     comserd |   .0000138   7.23e-06     1.91   0.057    -4.27e-07     .000028
       miles |   3.22e-10   3.45e-10     0.93   0.352    -3.57e-10    1.00e-09
         pop |  -1.58e-12   2.57e-12    -0.61   0.541    -6.63e-12    3.48e-12
       gspch |  -.0000907   .0000631    -1.44   0.152    -.0002148    .0000334
       _cons |    .000253    .000072     3.52   0.001     .0001115    .0003946
------------------------------------------------------------------------------
```

From the output above it is clear that many of the variables are insignificant. Unemployment rate should have good contribution towards VFR but it is insignificant according to this model. Removing all the highly insignificant variables and rerunning the regression gives following results.

Pooled Least Square Model With Only Significant Variables:

```
.
. reg mrall spircons perinc beertax sobapt vmiles jaild comserd

      Source |       SS       df       MS              Number of obs =     335
-------------+------------------------------           F(7, 327)     =   50.96
       Model |  5.6814e-07     7  8.1163e-08           Prob > F      =  0.0000
    Residual |  5.2083e-07   327  1.5927e-09           R-squared     =  0.5217
-------------+------------------------------           Adj R-squared =  0.5115
       Total |  1.0890e-06   334  3.2604e-09           Root MSE      = 4.0e-05

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000202   3.69e-06     5.46   0.000     .0000129    .0000274
      perinc |  -1.14e-08   1.26e-09    -9.03   0.000    -1.38e-08   -8.89e-09
     beertax |  -.0000149   6.23e-06    -2.39   0.017    -.0000272   -2.66e-06
      sobapt |   2.05e-06   3.13e-07     6.53   0.000     1.43e-06    2.66e-06
      vmiles |   1.29e-08   1.51e-09     8.58   0.000     9.98e-09    1.59e-08
       jaild |   .0000207   6.05e-06     3.42   0.001     8.77e-06    .0000326
     comserd |   .0000166   6.94e-06     2.39   0.017     2.94e-06    .0000302
       _cons |   .0002084   .0000211     9.86   0.000     .0001668    .0002499
------------------------------------------------------------------------------
```

Since this is panel data, there is a possibility of serial correlation over time. The variance can also vary in several time periods (heteroskedasticity). As we all know, if serially correlated errors and heteroskedasticity exist, the least squares estimator remains a linear unbiased estimator, but it's not best for analysis.

Interpretation:

- Spircons: For every additional gallon consumed by a person annually, the VFR (mrall) results in 20 more Vehicle Fatalities.
- Perinc: For every $1000 increase in per capita income, the VFR (mrall) decreases by 114 units.
- Beertax: For every 10 cents increase in BeerTax, the VFR (mrall) decreases by 1.49 units.
- Sobapt: For every 1% increase in Southern Baptist Population in a state, the Vehicle Fatality Rate increases by 0.20 units.
- Vmiles: When average miles per driver increases by 10000, the VFR (mrall) increases by 1.29 units.
- Jaild: The states where there is mandatory Jail sentence (Jaild = 1), the VF rate is 0.2 higher when compared to the states where there is no mandatory Jail sentence.
- Comserd: The states where there is mandatory Community Service sentence (comserd = 1), the VF rate is 0.16 higher when compared to the states where there is no Community Service sentence.

Test for Heteroskedasticity:

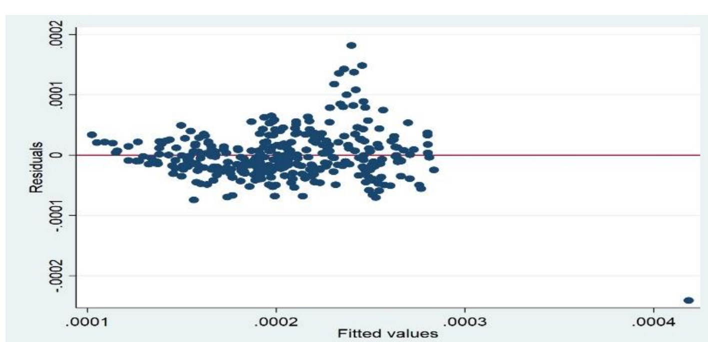
```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of mrall

         chi2(1)      =   154.37
         Prob > chi2  =   0.0000
```

From the above test, we can conclude based on the values that null hypothesis can be rejected and say that there is heteroskedasticity in the data.

Here, Null Hypothesis is that there is Homoskedasticity in the data.

Alternate Hypothesis is that there is Heteroskedasticity in the data.

As, there is a form of heteroskedasticity in the data, one of the methods which can be used to treat it is to use robust standard errors with cluster method.



Pooled Least Square Model with Cluster Robust Standard Errors:

```
.
. reg mrall spircons perinc beertax sobapt vmiles jaild comserd, vce(cluster state)

Linear regression                               Number of obs     =        335
                                                F(6, 47)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.5217
                                                Root MSE          =    4.0e-05

                                  (Std. Err. adjusted for 48 clusters in state)
```

mrall	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
spircons	.0000202	5.48e-06	3.68	0.001	9.14e-06	.0000312
perinc	-1.14e-08	2.60e-09	-4.37	0.000	-1.66e-08	-6.13e-09
beertax	-.0000149	.0000115	-1.30	0.201	-.0000381	8.24e-06
sobapt	2.05e-06	5.61e-07	3.64	0.001	9.17e-07	3.17e-06
vmiles	1.29e-08	7.01e-09	1.85	0.071	-1.16e-09	2.70e-08
jaild	.0000207	.0000158	1.31	0.197	-.0000111	.0000525
comserd	.0000166	.0000145	1.14	0.260	-.0000127	.0000458
_cons	.0002084	.0000554	3.76	0.000	.000097	.0003198

In response to the incorrect standard errors, we are ready to address this issue by running a regression using robust standard errors as seen above. As we see, using robust standard errors, although now correct, yield standard errors that more than double in value compared to the first standard errors. this means that the model is inefficient and thus still not best. Additionally, to the higher standard errors, we some of the variables are not any longer significant like Jaild, Comserd and Beertax. Furthermore, one other issue that both models have is that of omitted variable bias like external variables.

```
xtset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 1982 to 1988
                delta:  1 unit
```

With the above command, we state in stata that this is panel data, where state is the panel variable and year is the time variable.

The Fixed Effects Model:

```
. xtreg mrall spircons unrate perinc beertax sobapt mormon mlda dry yngdrv vmiles jaild comserd miles pop gspch, fe

Fixed-effects (within) regression               Number of obs      =       335
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.3772                                          min =         6
     between = 0.1090                                          avg =       7.0
     overall = 0.0841                                          max =         7

                                                F(15,272)          =     10.98
corr(u_i, Xb)  = -0.9084                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |    .000082   .0000102     8.02   0.000     .0000618    .0001021
      unrate |  -3.47e-06   1.04e-06    -3.32   0.001    -5.52e-06   -1.41e-06
      perinc |   9.60e-09   2.15e-09     4.47   0.000     5.37e-09    1.38e-08
     beertax |  -.0000402   .0000197    -2.04   0.043     -.000079   -1.33e-06
      sobapt |  -4.27e-06   5.97e-06    -0.71   0.475     -.000016    7.48e-06
      mormon |  -8.70e-07   4.38e-06    -0.20   0.843    -9.48e-06    7.74e-06
        mlda |   1.97e-06   1.82e-06     1.08   0.280    -1.61e-06    5.55e-06
         dry |   2.73e-06   1.33e-06     2.06   0.041     1.17e-07    5.34e-06
      yngdrv |   .0000481   .0000781     0.62   0.539    -.0001057    .0002018
      vmiles |   5.68e-10   1.35e-09     0.42   0.675    -2.10e-09    3.23e-09
       jaild |   1.26e-06   .0000121     0.10   0.917    -.0000226    .0000251
     comserd |  -1.54e-06    .000014    -0.11   0.912    -.0000291     .000026
       miles |   3.73e-10   6.12e-10     0.61   0.543    -8.32e-10    1.58e-09
         pop |  -8.06e-12   1.18e-11    -0.68   0.494    -3.12e-11    1.51e-11
       gspch |  -.0000534   .0000326    -1.64   0.103    -.0001176    .0000108
       _cons |  -.0000321   .0000945    -0.34   0.734    -.0002182     .000154
-------------+----------------------------------------------------------------
     sigma_u |  .00012721
     sigma_e |  .00001571
         rho |  .98496885   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:  F(47, 272) =    37.76              Prob > F = 0.0000
```

After reviewing the output, we see that the results are not as we expected as it appears that a majority of the explanatory variables are insignificant. We begin to drop insignificant variables one by one based on p-values. After dropping the insignificant variables, we run again fixed effects model and the output is below. In this model, the least squares estimator is still linear and unbiased, but it is not best as the standard errors are once again incorrect.

```
. xtreg mrall spircons unrate perinc beertax dry gspch, fe

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.3672                                          min =         7
     between = 0.0999                                          avg =       7.0
     overall = 0.0724                                          max =         7

                                                F(6,282)           =     27.28
corr(u_i, Xb)  = -0.8620                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000817   7.87e-06    10.39   0.000     .0000662    .0000972
      unrate |  -3.60e-06   9.74e-07    -3.69   0.000    -5.51e-06   -1.68e-06
      perinc |   1.02e-08   2.05e-09     4.99   0.000     6.18e-09    1.42e-08
     beertax |  -.0000469   .0000162    -2.90   0.004    -.0000788   -.0000151
         dry |   2.50e-06   1.28e-06     1.95   0.052    -2.32e-08    5.02e-06
       gspch |  -.0000507   .0000314    -1.62   0.107    -.0001126    .0000111
       _cons |  -.0000399   .0000421    -0.95   0.345    -.0001228     .000043
-------------+----------------------------------------------------------------
     sigma_u |  .00010556
     sigma_e |  .00001556
         rho |  .97874445   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:  F(47, 282) =    60.46              Prob > F = 0.0000
```

The Random Effect Model:

Because the fixed effect model doesn't work well with data that are slow changing or that remain constant within states, I decided I would try the random effect model also despite the very fact that the random effect model works better for data during which observations are randomly selected and our data set didn't involve any kind of random selection process. The output of random effect model is given below.

```
. xtreg mrall spircons unrate perinc beertax sobapt mormon mlda dry yngdrv vmiles jaild comserd miles pop gspch, re

Random-effects GLS regression                   Number of obs      =       335
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.2746                                          min =         6
     between = 0.2578                                          avg =       7.0
     overall = 0.2559                                          max =         7

                                                Wald chi2(13)      =         .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =         .

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000305   6.90e-06     4.42   0.000     .0000169     .000044
      unrate |  -5.23e-06   1.09e-06    -4.81   0.000    -7.36e-06   -3.10e-06
      perinc |   4.98e-10   2.05e-09     0.24   0.808    -3.51e-09    4.51e-09
     beertax |  -.0000287   .0000126    -2.28   0.022    -.0000533   -4.07e-06
      sobapt |   4.36e-06   8.66e-07     5.04   0.000     2.66e-06    6.06e-06
      mormon |   1.20e-06   6.10e-07     1.97   0.049     4.60e-09    2.39e-06
        mlda |   1.04e-06   1.95e-06     0.54   0.592    -2.78e-06    4.87e-06
         dry |   3.50e-08   6.70e-07     0.05   0.958    -1.28e-06    1.35e-06
      yngdrv |   .0002305   .0000764     3.02   0.003     .0000807    .0003803
      vmiles |   9.85e-10   1.24e-09     0.79   0.428    -1.45e-09    3.42e-09
       jaild |   .0000194   .0000101     1.93   0.054    -2.95e-07    .0000392
     comserd |  -.0000116   .0000116    -1.00   0.318    -.0000343    .0000112
       miles |   5.66e-10   4.08e-10     1.39   0.165    -2.34e-10    1.37e-09
         pop |  -5.92e-12   3.27e-12    -1.81   0.070    -1.23e-11    4.90e-13
       gspch |   -.000075   .0000364    -2.06   0.039    -.0001464   -3.67e-06
       _cons |    .000097   .0000566     1.71   0.087     -.000014    .0002079
-------------+----------------------------------------------------------------
     sigma_u |  .00003288
     sigma_e |  .00001571
         rho |  .81402826   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

After reviewing the output, we see that the results are not as we expected as it appears that a majority of the explanatory variables are insignificant. We begin to drop insignificant variables one by one based on p- values. After dropping the insignificant variables, we run again random effects model and the output is shown on the following page.

```
.
. xtreg mrall unrate beertax mormon sobapt yngdrv jaild pop gspch, re

Random-effects GLS regression                   Number of obs      =       335
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.1904                                          min =         6
     between = 0.2469                                          avg =       7.0
     overall = 0.2422                                          max =         7

                                                Wald chi2(7)       =         .
corr(u_i, X)    = 0 (assumed)                   Prob > chi2        =         .

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      unrate |  -5.63e-06   8.05e-07    -7.00   0.000    -7.21e-06   -4.05e-06
     beertax |  -.0000324   .0000133    -2.43   0.015    -.0000584   -6.30e-06
      mormon |   6.44e-07   6.94e-07     0.93   0.354    -7.17e-07    2.00e-06
      sobapt |   3.81e-06   8.05e-07     4.74   0.000     2.24e-06    5.39e-06
      yngdrv |   .0003337   .0000666     5.01   0.000     .0002032    .0004642
       jaild |   7.03e-06   6.21e-06     1.13   0.258    -5.15e-06    .0000192
         pop |  -2.35e-12   1.31e-12    -1.80   0.072    -4.92e-12    2.07e-13
       gspch |  -.0000548   .0000358    -1.53   0.126    -.0001248    .0000153
       _cons |   .0001819   .0000159    11.41   0.000     .0001507    .0002132
-------------+----------------------------------------------------------------
     sigma_u |  .00004457
     sigma_e |  .00001755
         rho |  .86576771   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Upon inspection, the variables beertax and perinc have become insignificant which is completely opposite to what we considered in our statistical analysis. This leads us to question the effectiveness of the random effect model as compared with the fixed effect model. Additionally, it is possible that the random effect model may suffer from the problem of endogeneity wherein one of the explanatory variables may be highly correlated with an omitted variable that is part of the error term. If this is in fact the case, we should opt to settle with the fixed effect model and its estimators. To determine if there is the presence of endogeneity in the random effect model, we perform a Hausman test.

<u>Hausman Test:</u>

```
. hausman fixed .

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
      unrate |   -3.60e-06    -5.63e-06        2.04e-06        5.48e-07
     beertax |   -.0000469    -.0000324       -.0000146        9.23e-06
       gspch |   -.0000507    -.0000548        4.02e-06               .
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       14.87
                Prob>chi2 =      0.0019
                (V_b-V_B is not positive definite)
```

Based on the results of the Hausman test, we will reject the null hypothesis at any significance level that there's no endogeneity and instead choose the alternate hypothesis that an explanatory variable could also be correlated with an omitted variable within the error term. due to this, we decide that the fixed effect model as the suitable method for the given data.

Fixed Effect model with Robust Standard Error:

```
. xtreg mrall spircons unrate perinc beertax dry gspch, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.3672                                          min =         7
     between = 0.0999                                          avg =       7.0
     overall = 0.0724                                          max =         7

                                                F(6,47)            =     22.49
corr(u_i, Xb)  = -0.8620                        Prob > F           =    0.0000

                                   (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000817   .0000127     6.43   0.000     .0000561    .0001072
      unrate |  -3.60e-06   9.83e-07    -3.66   0.001    -5.57e-06   -1.62e-06
      perinc |   1.02e-08   3.45e-09     2.96   0.005     3.27e-09    1.72e-08
     beertax |  -.0000469   .0000224    -2.10   0.041    -.0000919   -1.90e-06
         dry |   2.50e-06   1.20e-06     2.08   0.043     7.96e-08    4.92e-06
       gspch |  -.0000507   .0000235    -2.16   0.036    -.0000979   -3.56e-06
       _cons |  -.0000399   .0000719    -0.55   0.582    -.0001846    .0001048
-------------+----------------------------------------------------------------
     sigma_u |  .00010556
     sigma_e |  .00001556
         rho |  .97874445   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The above model yields good results as all variables are significant and therefore the standard errors are corrected, although still not efficient. While the results yielded are good, we might still wish to check out the regression model with time fixed effects since is useful when an omitted variable varies over time but not across states.

Entity and Time Fixed Effect Model:

```
. xtreg mrall spircons unrate perinc beertax dry gspch i.year, fe vce(cluster state)

Fixed-effects (within) regression               Number of obs     =        336
Group variable: state                           Number of groups  =         48

R-sq:                                           Obs per group:
     within  = 0.4595                                         min =          7
     between = 0.1010                                         avg =        7.0
     overall = 0.0693                                         max =          7

                                                F(12,47)          =      13.90
corr(u_i, Xb)  = -0.8608                        Prob > F          =     0.0000

                                 (Std. Err. adjusted for 48 clusters in state)
```

mrall	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
spircons	.0000817	.000012	6.81	0.000	.0000575	.0001058
unrate	-5.10e-06	1.24e-06	-4.13	0.000	-7.59e-06	-2.61e-06
perinc	8.71e-09	3.19e-09	2.73	0.009	2.30e-09	1.51e-08
beertax	-.0000443	.0000245	-1.81	0.076	-.0000935	4.91e-06
dry	1.91e-06	9.75e-07	1.96	0.056	-5.27e-08	3.87e-06
gspch	.0000396	.0000356	1.11	0.272	-.000032	.0001112
year						
1983	-7.18e-06	4.07e-06	-1.76	0.084	-.0000154	1.02e-06
1984	-.0000186	5.08e-06	-3.66	0.001	-.0000288	-8.36e-06
1985	-.0000203	4.95e-06	-4.10	0.000	-.0000303	-.0000103
1986	-5.50e-06	6.48e-06	-0.85	0.400	-.0000185	7.53e-06
1987	-9.99e-06	7.34e-06	-1.36	0.180	-.0000247	4.77e-06
1988	-.0000134	8.46e-06	-1.59	0.120	-.0000304	3.61e-06
_cons	1.81e-06	.0000632	0.03	0.977	-.0001253	.0001289
sigma_u	.00010588					
sigma_e	.00001453					
rho	.98150969	(fraction of variance due to u_i)				

Upon reviewing the output, we realize that while the model does include time variation, it fails to yield significant results. The sign of the variable “gspch” switched from negative to positive and other variables became even smaller. Hence, this model cannot be used and that we can conclude that there are not any omitted variables that change over time but not across states. we'll be using Fixed Effect model with robust standard errors.

Models & Limitations:

Pooled model

OLS when applied to panel data is referred to as Pooled OLS. When Pooled method is used on panel data, we observed heteroskedasticity and serial correlation. With these two, the model is unbiased and consistent but not efficient and the standard errors are incorrect. Along with this due to unobserved heterogeneity we also face endogeneity problem which causes our estimates to be biased and inconsistent. In order to resolve these issues, we decided to use Random or Fixed effects model

Random Effects Model

This is the most efficient model for panel data but due to the random individual differences present in the error term we frequently face endogeneity problem. I conducted Hausman test to verify that the correlation between the explanatory variables and the error term is zero. Since we rejected the null hypothesis in Hausman test and our data is not random data, we decided to go forward with the Fixed Effects model.

Entity and Time Fixed Effects model & Fixed Effect model with Robust Standard Error:

Upon reviewing the output from both the models, we realize that while the Entity and Time Fixed Effects model does include time variation, it fails to yield significant results. The sign of the variable “gspch” switched from negative to positive and other variables became even smaller. Hence, Entity and Time Fixed Effects model cannot be used and that we can conclude that there are not any omitted variables that change over time but not across states. we'll be using Fixed Effect model with robust standard errors.

Various Regressions Models For MRALLN:

Pooled Least Square Model:

Running a pooled model corresponding to given data gives following output.

```
.
. reg mralln spircons unrate perinc beertax sobapt mormon mlda dry yngdrv vmiles jaild comserd miles pop gspch

      Source |       SS           df       MS      Number of obs   =       335
-------------+----------------------------------   F(15, 319)      =     12.24
       Model |  1.4675e-08        15  9.7832e-10   Prob > F        =    0.0000
    Residual |  2.5500e-08       319  7.9938e-11   R-squared       =    0.3653
-------------+----------------------------------   Adj R-squared   =    0.3354
       Total |  4.0175e-08       334  1.2028e-10   Root MSE        =   8.9e-06

------------------------------------------------------------------------------
      mralln |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   2.55e-06   9.19e-07     2.77   0.006     7.37e-07    4.35e-06
      unrate |  -3.31e-08   3.05e-07    -0.11   0.914    -6.33e-07    5.67e-07
      perinc |  -1.44e-09   4.29e-10    -3.36   0.001    -2.28e-09   -5.95e-10
     beertax |  -5.77e-06   1.56e-06    -3.69   0.000    -8.83e-06   -2.70e-06
      sobapt |   4.29e-07   9.08e-08     4.72   0.000     2.50e-07    6.07e-07
      mormon |  -1.56e-07   5.91e-08    -2.64   0.009    -2.72e-07   -3.97e-08
        mlda |  -9.08e-07   6.08e-07    -1.49   0.136    -2.10e-06    2.88e-07
         dry |  -2.65e-07   7.17e-08    -3.69   0.000    -4.06e-07   -1.24e-07
      yngdrv |   .0000705    .000025     2.82   0.005     .0000213    .0001196
      vmiles |   7.89e-10   4.50e-10     1.75   0.080    -9.63e-11    1.67e-09
       jaild |   3.41e-06   1.46e-06     2.34   0.020     5.44e-07    6.28e-06
     comserd |   2.96e-07   1.62e-06     0.18   0.855    -2.90e-06    3.49e-06
       miles |   1.90e-10   7.74e-11     2.46   0.015     3.78e-11    3.43e-10
         pop |  -1.24e-12   5.78e-13    -2.15   0.032    -2.38e-12   -1.07e-13
       gspch |  -.0000344   .0000142    -2.43   0.016    -.0000623   -6.54e-06
       _cons |   .0000542   .0000162     3.35   0.001     .0000224     .000086
------------------------------------------------------------------------------
```

From the output above it is clear that many of the variables are insignificant. Unemployment rate should have good contribution towards VFR but it is insignificant according to this model. Removing all the highly insignificant variables and rerunning the regression gives following results.

Pooled Least Square Model With Only Significant Variables:

```
.
. reg mralln spircons perinc beertax sobapt dry yngdrv vmiles jaild miles pop gspch

      Source |       SS           df       MS      Number of obs   =       335
-------------+----------------------------------   F(11, 323)      =     15.43
       Model |  1.3839e-08        11  1.2581e-09   Prob > F        =    0.0000
    Residual |  2.6336e-08       323  8.1534e-11   R-squared       =    0.3445
-------------+----------------------------------   Adj R-squared   =    0.3222
       Total |  4.0175e-08       334  1.2028e-10   Root MSE        =   9.0e-06

------------------------------------------------------------------------------
      mralln |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   3.24e-06   8.85e-07     3.66   0.000     1.49e-06    4.98e-06
      perinc |  -1.38e-09   3.68e-10    -3.74   0.000    -2.10e-09   -6.52e-10
     beertax |  -5.88e-06   1.52e-06    -3.87   0.000    -8.88e-06   -2.89e-06
      sobapt |   4.81e-07   8.80e-08     5.47   0.000     3.08e-07    6.54e-07
         dry |  -2.80e-07   7.18e-08    -3.90   0.000    -4.21e-07   -1.38e-07
      yngdrv |   .0000647   .0000242     2.68   0.008     .0000172    .0001122
      vmiles |   8.19e-10   4.35e-10     1.88   0.061    -3.75e-11    1.68e-09
       jaild |   3.13e-06   1.19e-06     2.63   0.009     7.92e-07    5.48e-06
       miles |   1.75e-10   7.78e-11     2.25   0.025     2.20e-11    3.28e-10
         pop |  -1.11e-12   5.77e-13    -1.92   0.056    -2.24e-12    2.68e-14
       gspch |  -.0000405   .0000138    -2.94   0.003    -.0000676   -.0000134
       _cons |   .0000336   8.60e-06     3.91   0.000     .0000167    .0000505
------------------------------------------------------------------------------
```

Since this is panel data, there is a possibility of serial correlation over time. The variance can also vary in several time periods (heteroskedasticity). As we all know, if serially correlated errors and heteroskedasticity exist, the least squares estimator remains a linear unbiased estimator, but it's not best for analysis.

Interpretation:

- Spircons: For every additional gallon consumed by a person annually, the VFR (mralln) results in 3.24 more Vehicle Fatalities.
- Perinc: For every 1000$ increase in per capita income, the VFR (mralln) decreases by 138 units.
- Beertax: For every 10 cents increase in BeerTax, the VFR (mralln) decreases by 0.58 units.
- Sobapt: For every 1% increase in Southern Baptist Population in a state, the Vehicle Fatality Rate increases by 0.048 units.
- Dry: For every 1% increase in Dry Counties in a State, the Vehicle Fatality Rate decreases by 2.8 %.
- Yngdrv: For every 1 % increase in Yngdrv in a state, the Vehicle Fatality increases by 64.7 units.
- Vmiles: When average miles per driver increases by 10,000, the VFR (mralln) increases by 8.19 units.
- Jaild: The states where there is mandatory Jail sentence (Jaild = 1), the VF rate is 0.03 higher when compared to the states where there is no mandatory Jail sentence.
- Miles: For every unit increase in miles in a state, the Vehicle Fatality rate increases by 1.75 units.
- Pop: For every unit increase in population in a state, the Vehicle Fatality rate decreases by 0.01 units.
- Gspch: For every 1% increase in Gspch in a state, the Vehicle Fatality rate decreases by 0.405 units.

Test for Heteroskedasticity:

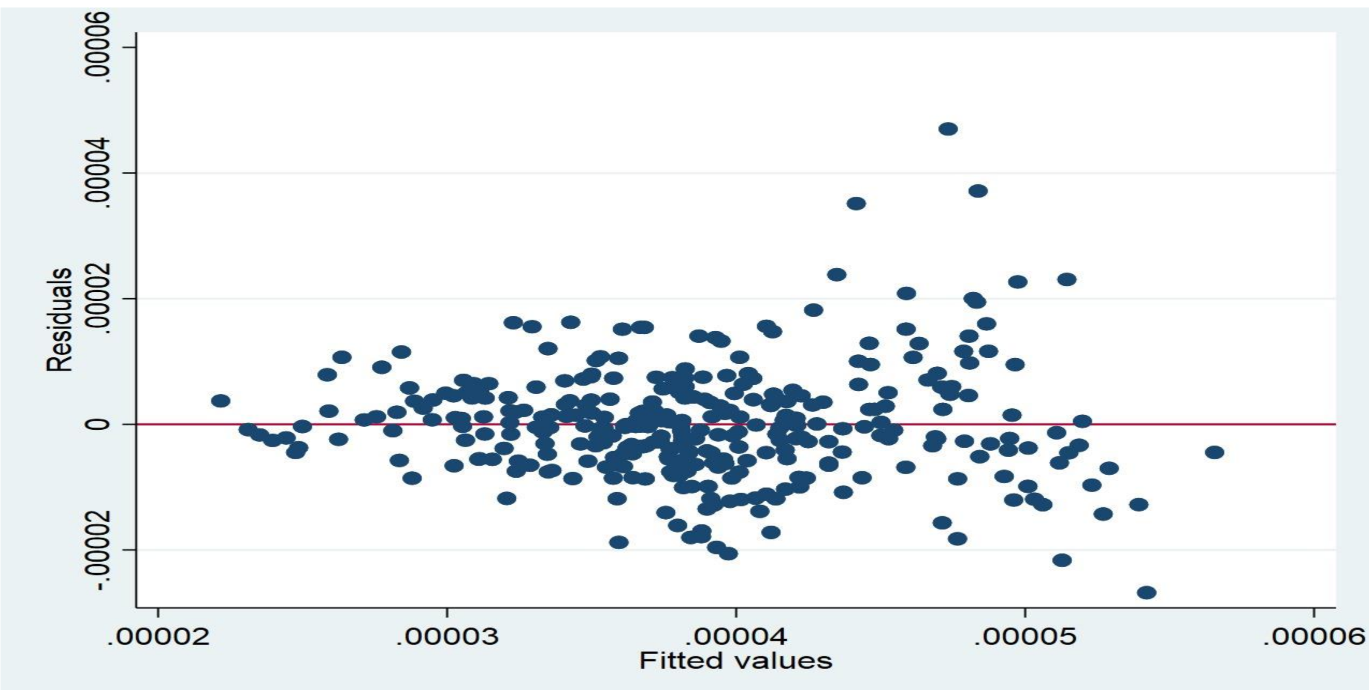
```
.
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of mralln

         chi2(1)      =    62.77
         Prob > chi2  =   0.0000
```

From the above test, we can conclude based on the values that null hypothesis can be rejected and say that there is heteroskedasticity in the data.

Here, Null Hypothesis is that there is Homoskedasticity in the data.

Alternate Hypothesis is that there is Heteroskedasticity in the data.

As, there is a form of heteroskedasticity, one of the methods which can be used to treat it is to use robust standard errors with cluster method.



Pooled Least Square Model with Cluster Robust Standard Errors:

```
. reg mralln spircons perinc beertax sobapt dry yngdrv vmiles jaild miles pop gspch, vce(cluster state)

Linear regression                               Number of obs     =        335
                                                F(8, 47)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.3445
                                                Root MSE          =    9.0e-06

                                   (Std. Err. adjusted for 48 clusters in state)
```

mralln	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
spircons	3.24e-06	1.49e-06	2.18	0.035	2.46e-07	6.22e-06
perinc	-1.38e-09	4.88e-10	-2.82	0.007	-2.36e-09	-3.95e-10
beertax	-5.88e-06	2.74e-06	-2.15	0.037	-.0000114	-3.73e-07
sobapt	4.81e-07	1.77e-07	2.72	0.009	1.25e-07	8.37e-07
dry	-2.80e-07	1.18e-07	-2.36	0.022	-5.18e-07	-4.15e-08
yngdrv	.0000647	.0000342	1.89	0.065	-4.19e-06	.0001336
vmiles	8.19e-10	1.12e-09	0.73	0.470	-1.44e-09	3.08e-09
jaild	3.13e-06	2.46e-06	1.28	0.208	-1.81e-06	8.08e-06
miles	1.75e-10	9.21e-11	1.90	0.063	-1.02e-11	3.60e-10
pop	-1.11e-12	6.80e-13	-1.63	0.110	-2.48e-12	2.60e-13
gspch	-.0000405	.0000191	-2.12	0.040	-.000079	-2.00e-06
_cons	.0000336	.000013	2.59	0.013	7.47e-06	.0000598

In response to the incorrect standard errors, we are ready to address this issue by running a regression using robust standard errors as seen above. One other issue that both models have is that of omitted variable bias like external variables.

```
.
. xtset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 1982 to 1988
                delta:  1 unit
```

With the above command, we state in stata that this is panel data, where state is the panel variable and year is the time variable.

Fixed Effects Model:

```
. xtreg mralln spircons unrate perinc beertax sobapt mormon mlda dry yngdrv vmiles jaild comserd miles pop gspch, fe

Fixed-effects (within) regression               Number of obs      =       335
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.1525                                          min =         6
     between = 0.0410                                          avg =       7.0
     overall = 0.0356                                          max =         7

                                                F(15,272)          =      3.26
corr(u_i, Xb)  = -0.9172                        Prob > F           =    0.0001

------------------------------------------------------------------------------
      mralln |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000139   4.04e-06     3.44   0.001     5.96e-06    .0000219
      unrate |  -1.10e-07   4.13e-07    -0.27   0.791    -9.22e-07    7.03e-07
      perinc |   8.40e-10   8.50e-10     0.99   0.324    -8.33e-10    2.51e-09
     beertax |  -.0000128   7.80e-06    -1.64   0.103    -.0000281    2.60e-06
      sobapt |   9.67e-07   2.36e-06     0.41   0.682    -3.68e-06    5.61e-06
      mormon |  -1.03e-06   1.73e-06    -0.59   0.553    -4.43e-06    2.38e-06
        mlda |   3.95e-07   7.20e-07     0.55   0.584    -1.02e-06    1.81e-06
         dry |   4.74e-07   5.24e-07     0.90   0.367    -5.58e-07    1.51e-06
      yngdrv |   .0000258   .0000309     0.83   0.405     -.000035    .0000866
      vmiles |  -4.39e-10   5.35e-10    -0.82   0.413    -1.49e-09    6.15e-10
       jaild |   5.21e-07   4.79e-06     0.11   0.913    -8.91e-06    9.96e-06
     comserd |  -4.31e-06   5.54e-06    -0.78   0.437    -.0000152    6.60e-06
       miles |   1.47e-10   2.42e-10     0.61   0.544    -3.30e-10    6.24e-10
         pop |  -4.53e-12   4.66e-12    -0.97   0.331    -1.37e-11    4.64e-12
       gspch |  -.0000203   .0000129    -1.57   0.117    -.0000457    5.11e-06
       _cons |   .0000124   .0000374     0.33   0.740    -.0000612     .000086
-------------+----------------------------------------------------------------
     sigma_u |  .00002352
     sigma_e |  6.216e-06
         rho |  .93469247   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(47, 272) = 8.25                     Prob > F = 0.0000
```

After reviewing the output, we see that the results are not as we expected as it appears that a majority of the explanatory variables are insignificant.

The Random Effect Model:

Because the fixed effect model doesn't work well with data that are slow changing or that remain constant within states, I decided to try random effect model also despite the very fact that the random effect model works better for data during which observations are randomly selected and our data set didn't involve any kind of random selection process. The output of random effect model is given below.

```
.
. xtreg mralln spircons unrate perinc beertax sobapt mormon mlda dry yngdrv vmiles jaild comserd miles pop gspch, re

Random-effects GLS regression                   Number of obs      =       335
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.0941                                         min =         6
     between = 0.3986                                         avg =       7.0
     overall = 0.3035                                         max =         7

                                                Wald chi2(13)      =         .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =         .

------------------------------------------------------------------------------
      mralln |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   3.42e-06   1.65e-06     2.08   0.038     1.91e-07    6.65e-06
      unrate |  -4.46e-07   3.50e-07    -1.28   0.202    -1.13e-06    2.39e-07
      perinc |  -1.05e-09   6.10e-10    -1.72   0.086    -2.24e-09    1.47e-10
     beertax |  -6.69e-06   2.95e-06    -2.27   0.023    -.0000125   -9.07e-07
      sobapt |   6.08e-07   1.86e-07     3.27   0.001     2.43e-07    9.73e-07
      mormon |  -7.34e-08   1.24e-07    -0.59   0.553    -3.16e-07    1.69e-07
        mlda |  -2.22e-07   6.51e-07    -0.34   0.733    -1.50e-06    1.05e-06
         dry |  -2.88e-07   1.44e-07    -2.00   0.046    -5.71e-07   -5.45e-09
      yngdrv |   .0000638   .0000251     2.55   0.011     .0000147    .0001129
      vmiles |  -1.38e-10   4.16e-10    -0.33   0.740    -9.52e-10    6.77e-10
       jaild |   2.95e-06   2.55e-06     1.16   0.248    -2.05e-06    7.94e-06
     comserd |  -3.14e-06   2.93e-06    -1.07   0.284    -8.89e-06    2.60e-06
       miles |   1.10e-10   1.18e-10     0.93   0.352    -1.22e-10    3.42e-10
         pop |  -8.64e-13   8.97e-13    -0.96   0.336    -2.62e-12    8.95e-13
       gspch |  -.0000274   .0000125    -2.19   0.029    -.0000519   -2.88e-06
       _cons |   .0000455   .0000178     2.56   0.011     .0000106    .0000804
-------------+----------------------------------------------------------------
     sigma_u |  6.877e-06
     sigma_e |  6.216e-06
         rho |   .5503421   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

After reviewing the output, we see that the results are not as we expected as it appears that a majority of the explanatory variables are insignificant. We begin to drop insignificant variables one by one based on p- values. After dropping the insignificant variables, we run again random effects model and the output is shown on the following page.

```
.
. xtreg mralln spircons perinc beertax sobapt dry yngdrv gspch, re

Random-effects GLS regression                   Number of obs      =       336
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.0919                                          min =         7
     between = 0.3353                                          avg =       7.0
     overall = 0.2590                                          max =         7

                                                Wald chi2(7)       =     50.31
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      mralln |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   3.11e-06   1.55e-06     2.01   0.044     7.72e-08    6.14e-06
      perinc |  -6.59e-10   4.43e-10    -1.49   0.137    -1.53e-09    2.10e-10
     beertax |  -6.91e-06   2.97e-06    -2.33   0.020    -.0000127   -1.10e-06
      sobapt |   6.74e-07   1.82e-07     3.71   0.000     3.18e-07    1.03e-06
         dry |  -3.25e-07   1.46e-07    -2.23   0.026    -6.11e-07   -3.99e-08
      yngdrv |   .0000523   .0000238     2.19   0.028     5.59e-06     .000099
       gspch |   -.000021   .0000115    -1.83   0.068    -.0000434    1.52e-06
       _cons |   .0000334   9.27e-06     3.60   0.000     .0000152    .0000515
-------------+----------------------------------------------------------------
     sigma_u |  7.219e-06
     sigma_e |  6.159e-06
         rho |  .57872153   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Upon inspection, the variables perinc have become insignificant which is completely opposite to what we considered in our statistical analysis. This leads us to question the effectiveness of the random effect model as compared with the fixed effect model. To determine if there is the presence of endogeneity in the random effect model, we perform a Hausman test.

<u>Hausman Test:</u>

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
    spircons |    .0000139     3.11e-06        .0000108        3.74e-06
      perinc |    8.40e-10    -6.59e-10        1.50e-09        7.25e-10
     beertax |   -.0000128    -6.91e-06       -5.86e-06        7.22e-06
      sobapt |    9.67e-07     6.74e-07        2.93e-07        2.35e-06
         dry |    4.74e-07    -3.25e-07        8.00e-07        5.04e-07
      yngdrv |    .0000258     .0000523       -.0000265        .0000196
       gspch |   -.0000203     -.000021        6.74e-07        5.90e-06
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       12.53
                Prob>chi2 =      0.0511
```

Based on the results of the Hausman test, we can’t reject the null hypothesis at 5% significance level that there's endogeneity and we choose the alternate hypothesis that an explanatory variable could also be correlated with an omitted variable within the error term. due to this, we decide that the random effects model as the suitable method for the given data.

Random model with Robust Standard Error:

```
. xtreg mralln spircons perinc beertax sobapt dry yngdrv gspch, re vce(cluster stat

Random-effects GLS regression                   Number of obs      =       336
Group variable: state                           Number of groups   =        48

R-sq:                                           Obs per group:
     within  = 0.0919                                         min =          7
     between = 0.3353                                         avg =        7.0
     overall = 0.2590                                         max =          7

                                                Wald chi2(6)       =         .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =         .

                                  (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      mralln |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   3.11e-06   1.40e-06     2.22   0.026     3.68e-07    5.85e-06
      perinc |  -6.59e-10   4.05e-10    -1.63   0.104    -1.45e-09    1.35e-10
     beertax |  -6.91e-06   2.60e-06    -2.66   0.008     -.000012   -1.82e-06
      sobapt |   6.74e-07   1.55e-07     4.36   0.000     3.71e-07    9.77e-07
         dry |  -3.25e-07   1.18e-07    -2.76   0.006    -5.57e-07   -9.41e-08
      yngdrv |   .0000523   .0000263     1.99   0.047     6.85e-07    .0001039
       gspch |   -.000021   .0000124    -1.69   0.091    -.0000453    3.37e-06
       _cons |   .0000334   .0000107     3.13   0.002     .0000124    .0000543
-------------+----------------------------------------------------------------
     sigma_u |  7.219e-06
     sigma_e |  6.159e-06
         rho |  .57872153   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The above model yields good results as all variables are significant and therefore the standard errors are corrected, although still not efficient, we'll be using Random Effect model with robust standard errors.

Models & Limitations:

Pooled model

OLS when applied to panel data is referred to as Pooled OLS. When Pooled method is used on panel data, we observed heteroskedasticity and serial correlation. With these two, the model is unbiased and consistent but not efficient and the standard errors are incorrect. Along with this due to unobserved heterogeneity we also face endogeneity problem which causes our estimates to be biased and inconsistent. In order to resolve these issues, we decided to use Random or Fixed effects model

Random Effects model with Robust Standard Error:

This is the most efficient model for panel data but due to the random individual differences present in the error term we frequently face endogeneity problem. We conducted Hausman test to verify that the correlation between the explanatory variables and the error term is zero. Since we didn’t reject the null hypothesis in Hausman test, I decided to go forward with the Random Effects model with Robust Standard Error.

Summary for both Mrall and Mralln:

Without recognizing the existence of serial correlation or testing for heteroskedasticity, our initial pooled least square estimator was a linear unbiased estimator but not best. By introducing the cluster robust standard errors in our regression, we corrected the quality errors. However, our model was still inefficient. so as to enhance our model, we introduced fixed effects using an estimator with entity and time fixed effects.

Since a random effects model exposes our model to possible endogenous regressors, we decided to match the coefficient estimates from both the fixed effects and random effects models. From the Hausman test, we conclude the random effects estimator is inconsistent and that we choose to stick with either the fixed effects estimator for Mrall and Mralln.

Conclusion:

The fixed effect estimators indicate that increasing beer tax will reduce the Vehicle Fatality rate in states whereas state laws against drinking and driving are ineffective thanks to reverse causality. I would like to return up with alternative ways to deal with this problem for instance by running several awareness programs to bring down the alcohol consumption and educate people on the results of drinking and driving. Other factors like the unemployment rate, GSP, and per capita income can't be controlled immediately, in order that they just indicate how they affect Vehicle Fatality rate. I saw that unemployment may be a major factor and states’ policies to scale back unemployment will help reduce the strain and depression among people which successively will decrease Vehicle Fatality rate.