PROJECT REPORT

Applied Econometrics and Time Series Analysis (BUAN 6312)

How do Drunk driving Laws affect Traffic Deaths?

Group Members:

John Paul Nokki (jxn190014)
Raghava Prasad Sridar (rxs180135)

Abstract

This Project focuses on understanding the effect of Drinking laws on Fatalities caused on the roads. The dataset used in this project consists of data on laws, drinking habits, economic conditions, socio-cultural practices, driving practices, The data is recorded for a time period of 6 years from 1982-1988 for 48 Lower states in The United States of America. This is found to be clearly a classic case of panel data. The process is thus carried out through the methods specified for analyzing panel data and results have been postulated in the report.

Dataset Description

The dataset is balanced consisting of data for the lower 48 states (excluding Alaska and Hawaii) annually from 1982 through1988.

Dependent variable

We intend to find how the traffic fatality rates were affected by the drinking laws and other factors from the given dataset.
The traffic fatality related data is provided by the US Department of Transportation Fatal Accident Reporting System.
The dependent variable is the vehicle fatality rate (**mrall**), It is the number of vehicle fatalities for every 10000 people living in the state.

Explanatory Variables

We are considering several significant factors that might be explaining the cause of Vehicle fatalities in general and drinking law specific variables are also studied and taken into consideration. All the other variables in the data are mentioned below

Jaild - Mandatory Jail sentence, **comserd** - Mandatory community service, **mlda** - Minimum legal drinking age looks like the strategies made by the state to regularize alcohol consumption of its residents and fatalities occurred due to them.
So, we are considering them as explanatory variables. However, we need consider several other factors like **beertax**, **spircons**, which are not exactly laws but can be considered as a strong influence for understanding drinking practices of different states.
We could also consider the **unrate - State** Unemployment rate, **perinc** - Per capita personal Income as a socio-economic factor which could influence drinking habits and hence it has its effect on traffic deaths.

Data Pre-Processing:

Correlation Matrix:

This is a primary data analysis method used to understand the effect of all the variables with each other. We cannot include two correlated Explanatory variables, since they could cause Multicollinearity problems in our analysis.

	spircons	*unrate*	*perinc*	*beertax*	*sobapt*	*mormon*	*dry*	*vmiles*	*mrall*	*mralln*	*allsvnrate*	*mra1517*	*mra1517n*	*mra1820*	*mra1820n*	*mra2124*	*mra2124n*	*mraidall*	*gspch*
spircons	1																		
unrate	-0.23918	1																	
perinc	0.45214	-0.55362	1																
beertax	-0.08947	0.056772	-0.39752	1															
sobapt	-0.29428	0.263981	-0.47453	0.633053	1														
mormon	-0.17882	-0.00783	-0.21771	0.005202	-0.14801	1													
dry	-0.26806	0.256416	-0.34366	0.177618	0.571674	-0.09077	1												
vmiles	-0.05649	-0.27815	-0.0802	0.141951	0.135363	0.000442	-0.08272	1											
mrall	-0.06478	0.175345	-0.49881	0.305553	0.440098	0.084292	0.126658	0.396831	1										
mralln	-0.04718	0.25788	-0.35548	0.12267	0.276159	-0.07689	-0.02319	0.225992	0.766548	1									
allsvnrate	-0.03134	0.215298	-0.38231	0.123504	0.232958	-0.07676	-0.00357	0.237952	0.691871	0.893659	1								
mra1517	-0.22804	0.012803	-0.43116	0.246049	0.271059	0.193993	0.116347	0.412941	0.675399	0.451831	0.431823	1							
mra1517n	-0.07339	0.025789	-0.2411	0.039751	0.085729	0.030111	-0.06594	0.253988	0.422366	0.503377	0.476947	0.526085	1						
mra1820	-0.003	0.134883	-0.36257	0.067024	0.202364	0.070199	0.017471	0.332384	0.819707	0.676975	0.64337	0.58348	0.346515	1					
mra1820n	-0.0428	0.156709	-0.2268	-0.04986	0.071378	0.014522	-0.05691	0.17179	0.495928	0.652601	0.578681	0.320469	0.256248	0.709803	1				
mra2124	0.037171	0.071354	-0.34187	0.153997	0.280776	-0.02812	0.037837	0.341916	0.865334	0.732513	0.69745	0.541133	0.403783	0.703492	0.420711	1			
mra2124n	0.052972	0.014568	-0.08929	-0.01274	0.057253	-0.20251	-0.0624	0.149572	0.405655	0.64422	0.613724	0.217779	0.244076	0.398922	0.318121	0.588638	1		
mraidall	-0.11711	0.281985	-0.54414	0.290951	0.460143	-0.06402	0.219859	0.209787	0.749882	0.683637	0.671264	0.470944	0.29492	0.628268	0.464546	0.666694	0.353964	1	
gspch	0.267518	-0.43283	0.391126	0.112859	-0.02469	-0.05315	0.007607	-0.00423	-0.18183	-0.29897	-0.29833	-0.11424	-0.14563	-0.21785	-0.31158	-0.13251	-0.14116	-0.26985	1

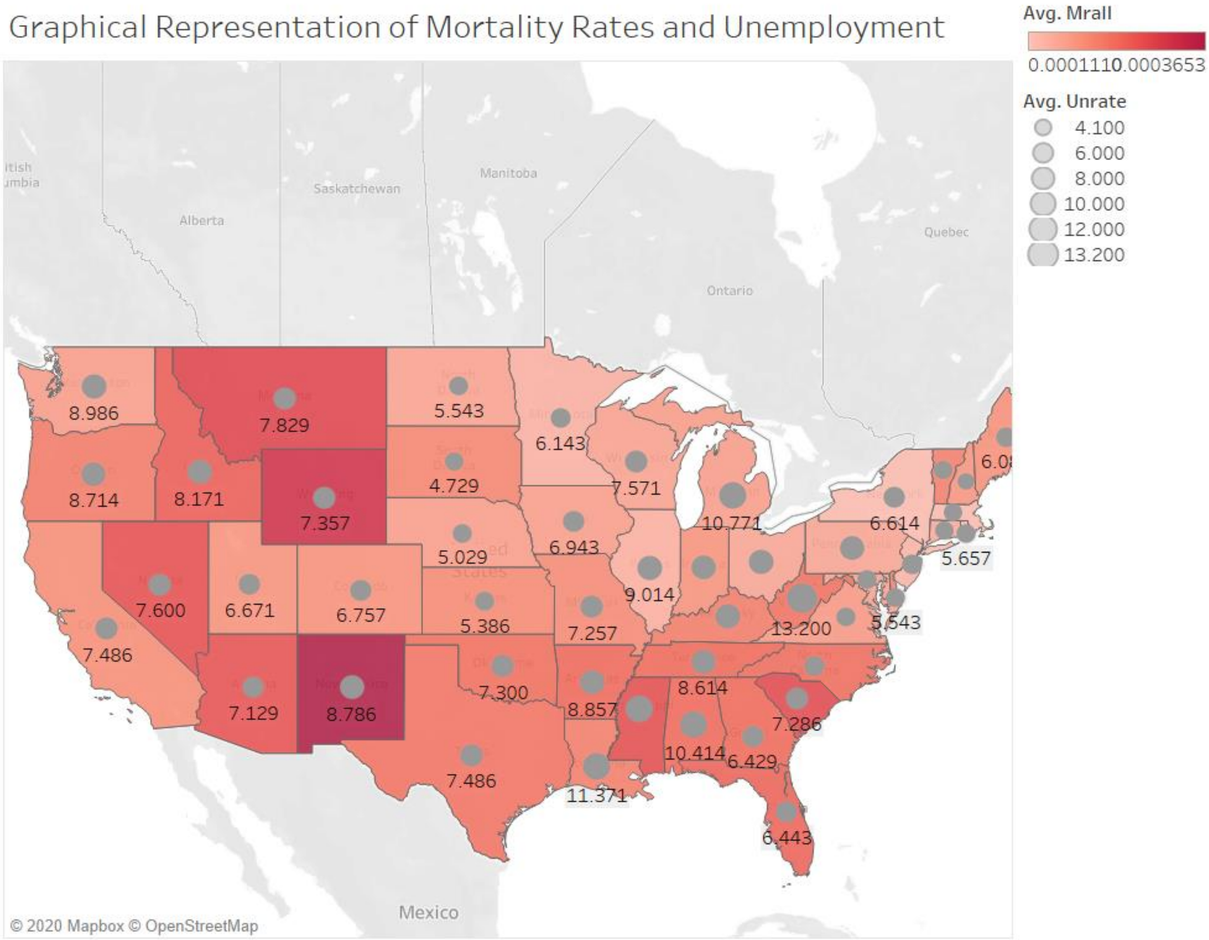
We can see that unemployment rate and per capita income are negatively correlated and the economic theory agrees with it. Also, we see that dry areas are present consistently in areas where the presence of southern Baptists is high. We see that single vehicle fatality rates are highly correlated with the night vehicle fatalities.

Based on the correlation matrix, we would like to look at

a) Vehicle fatality rates vs unemployment grouped by state.
b) Alcohol Involved Vehicle fatality rates vs spirit consumption grouped by state.
c) Vehicle fatality rate vs per capita income.
d) Alcohol involved VFR vs Beer tax.

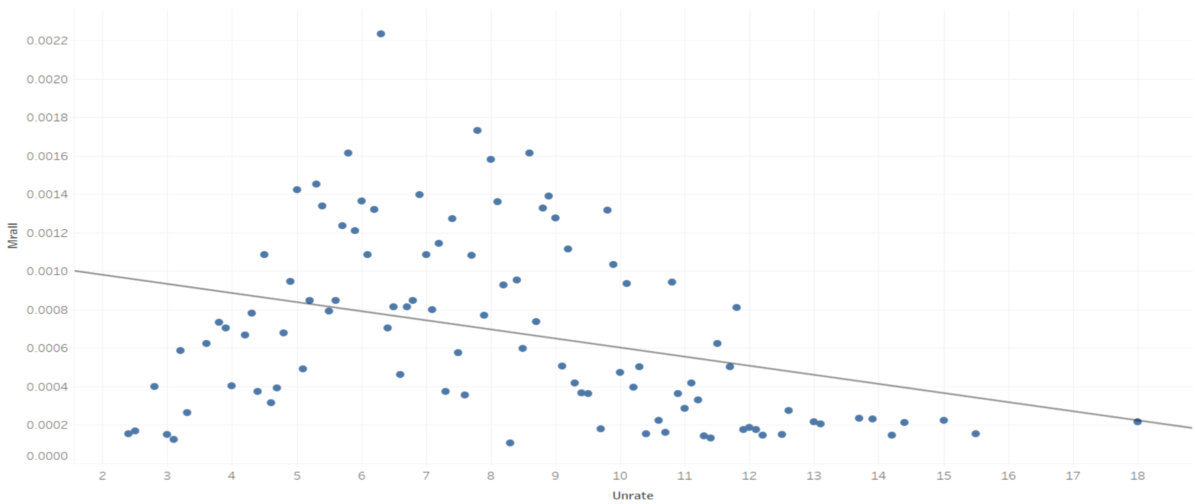
Vehicle Fatality Rates (VFR) and Unemployment

From the graph below we can see that the states with higher vehicle fatality rates like New Mexico, Wyoming, South Carolina, Mississippi, Idaho, Nevada, Montana have moderate to high unemployment rates. While there is huge variance in unemployment rates in other states.



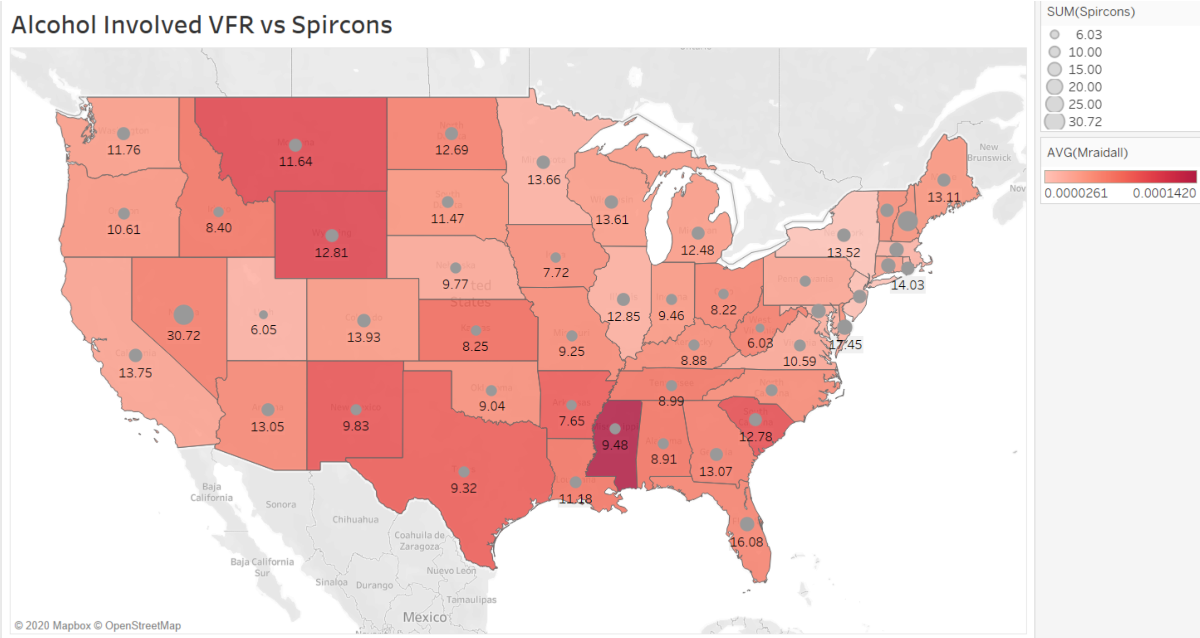
VFR vs Unemployment:

We would like to look into this relationship because it is believed that Alcohol consumption is higher when there is higher unemployment rate.

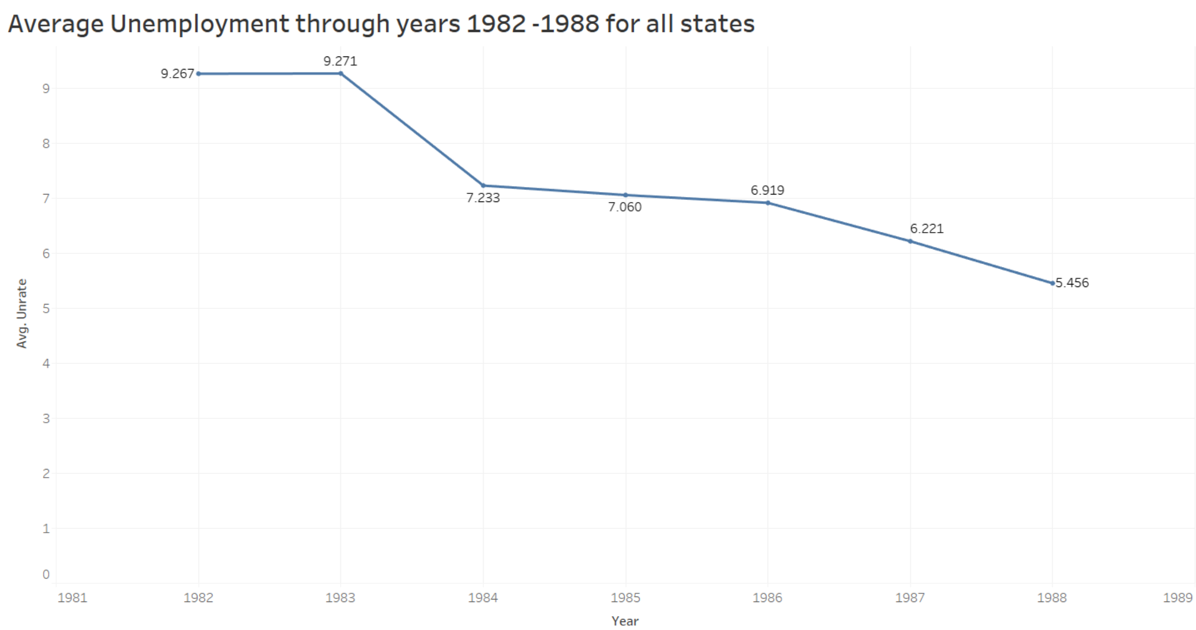


Alcohol Induced VFR vs Spircons

From the graph below, we can see that the states with higher vehicle fatality rates like Wyoming, Montana, New Mexico, Nevada, Arizona, South Carolina have moderate to high spirit consumption rates, and ironically Mississippi, the state with highest Alcohol aided VFR has lesser spirit consumption rate.

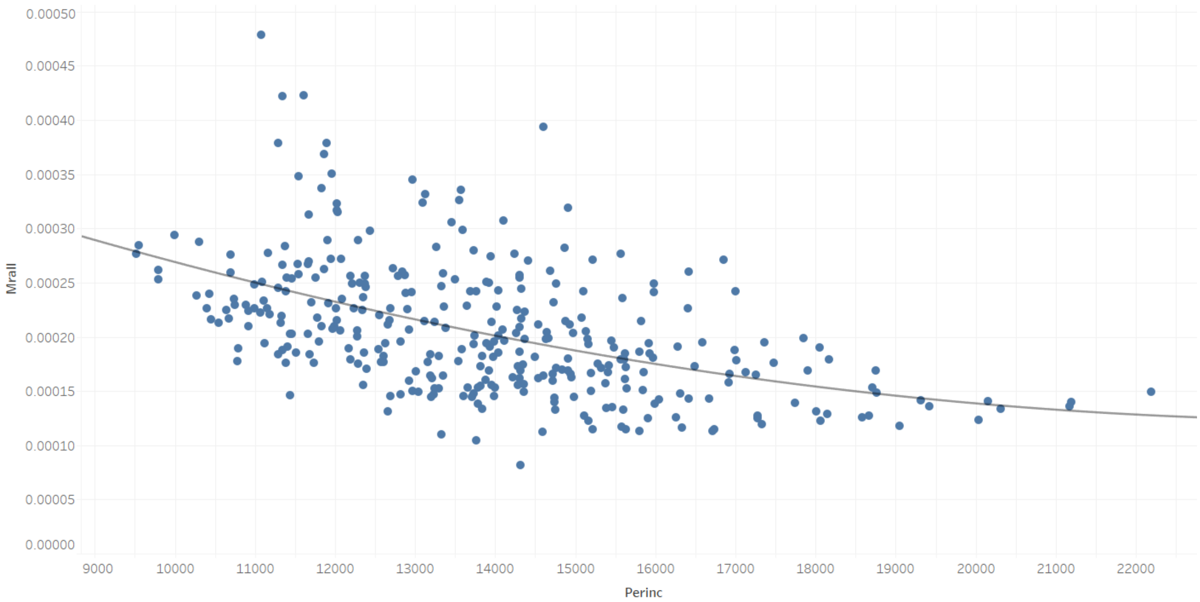


Average Unemployment along years and states



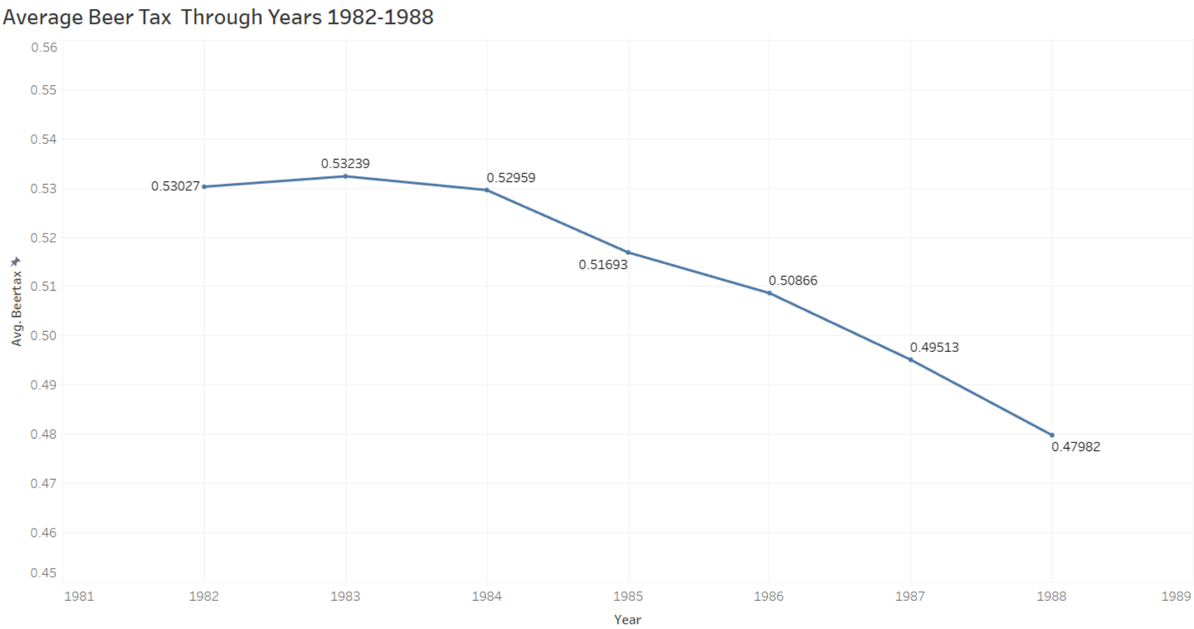
The above graph clearly depicts the trend of Unemployment along the time period. We see that the rate of unemployment decreasing at a constant rate, which also could infer an increase in per capita income of people across the years.

VFR vs Per Capita Income



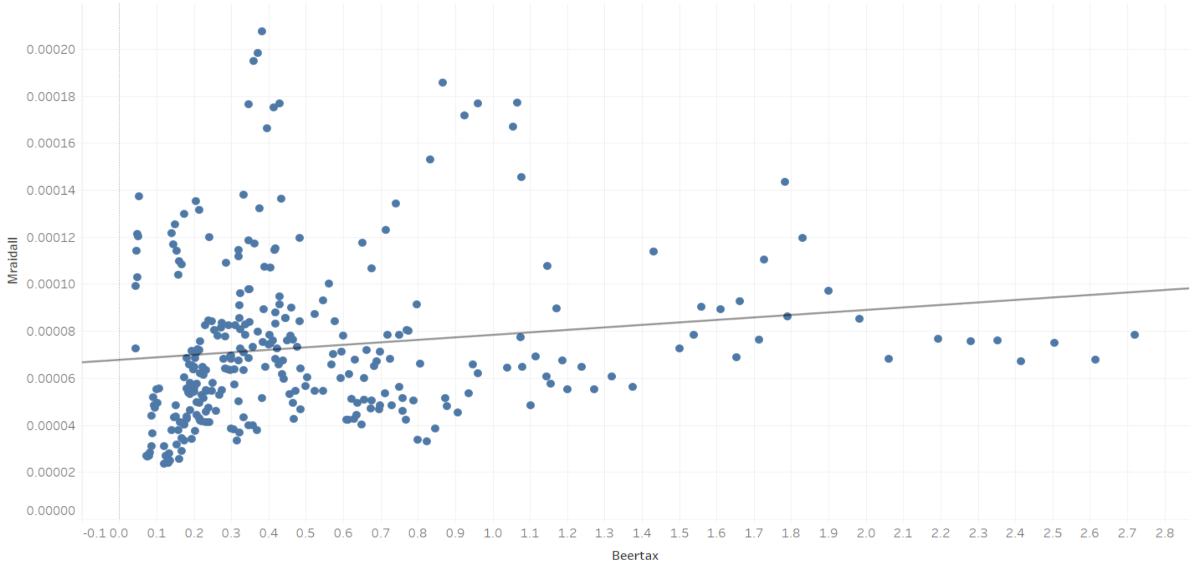
The above graph shows us a depiction of the effect of per capita income on vehicle fatality rates. We fit this as a quadratic equation due to higher variances in the perinc variable. We got a denser data for per capita income range of 11000 to 150000, but there is significantly lesser data for higher per capita income. The exact relationship between the high income and VFR hence could not be clearly understood.

Average Beer Tax across the years 1982-1988 for all states



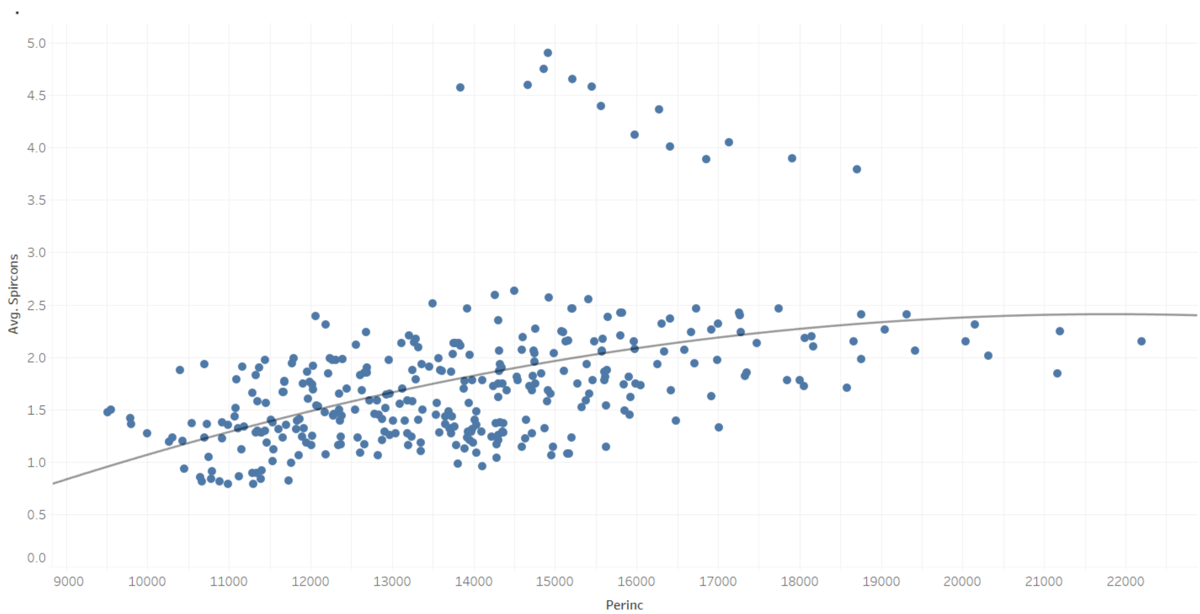
The above graph clearly informs us that the beer tax is steadily decreasing across years. However, the fluctuation pattern for individual states across years could vary. But going with a general trend could help draw conclusions on a nation-wide basis and its effects on Vehicle Fatality Rates.

Alcohol Involved VFR vs Beer Tax



Alcohol involved VFR is expected to be negatively affected by increasing beer taxes, but this graph explains a contradictory point. We could see a positive relationship between them. But this can be because we have less data on cases where the beer tax is higher. And also, the variance is huge. This ought to be studied further and an individual state's beer tax over years would give us a much better idea on its effects on VFR.

Spircons Vs Perinc



This graph shows us the causation of personal income and consumption of alcohol. We expected this graph to have a positive relationship as higher income could cause people to consume more alcohol. But due disparity in data for higher per capita income, we could not clearly explain the effect of Per capita income for spircons. We could also see few outliers in the data.

Regression Models

Initially we are not sure about the significance of any explanatory variables, so we are implementing a kitchen-sink model, where we are including all the explanatory variables and throwing them into the regression to understand their significance.

All Variables

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

Source	SS	df	MS
Model	5.7944e-07	13	4.4572e-08
Residual	5.0953e-07	321	1.5873e-09
Total	1.0890e-06	334	3.2604e-09

Number of obs	=	335
F(13, 321)	=	28.08
Prob > F	=	0.0000
R-squared	=	0.5321
Adj R-squared	=	0.5132
Root MSE	=	4.0e-05

mrall	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
spircons	.0000208	3.90e-06	5.34	0.000	.0000131	.0000285
unrate	-1.49e-06	1.26e-06	-1.18	0.238	-3.96e-06	9.89e-07
perinc	-1.15e-08	1.68e-09	-6.87	0.000	-1.48e-08	-8.24e-09
beertax	-.0000185	6.93e-06	-2.67	0.008	-.0000322	-4.89e-06
mlda	-3.93e-07	2.70e-06	-0.15	0.885	-5.71e-06	4.92e-06
sobapt	2.58e-06	3.95e-07	6.54	0.000	1.81e-06	3.36e-06
mormon	1.32e-07	2.63e-07	0.50	0.618	-3.86e-07	6.49e-07
dry	-5.48e-07	3.18e-07	-1.72	0.086	-1.17e-06	7.77e-08
vmiles	1.16e-08	1.68e-09	6.93	0.000	8.34e-09	1.50e-08
jaild	.0000202	6.32e-06	3.20	0.002	7.76e-06	.0000326
comserd	.0000145	7.22e-06	2.01	0.045	3.12e-07	.0000287
gspch	-.0000815	.0000629	-1.30	0.196	-.0002052	.0000422
yngdrv	.0000254	.0001107	0.23	0.818	-.0001924	.0002433
_cons	.0002367	.000071	3.33	0.001	.000097	.0003764

Here we could see several variables which we considered to be effective on explaining the VFR seems to be insignificant. Unrate, mlda, mormon, comserd, yngdrv are insignificant and we will exclude ‘yngdrv’ because it is highly insignificant, and we decided to exclude ‘mormon’ because the % of mormons with respect to the entire population is negligible. We chose not to remove ‘mlda’ because it is a law directly related to drinking. We also chose to remove ‘gspch’ due to its high insignificance and also that is highly sensitive to other imported products and not just alcohol. And we want to run a regression again to see if there is any improvement in the significance of any other variables.

Significant Variables

We ran a regression with ‘spircons’ ‘unrate’ ‘mlda’ ‘perinc’ ‘beertax’ ‘sobapt’ ‘dry’ ‘vmiles’ ‘jaild’ ‘comserd’. The variables ‘dry’, ‘comserd’ which were insignificant in the first model seem to have become significant in this model. And significance of ‘mlda’ decreased this could be due to Omitted variable bias. We can clearly see the differences in standard errors and Coefficients of the variables.

```
. reg mrall spircons      unrate  mlda perinc      beertax  sobapt dry      vmiles jaild comserd

      Source |       SS           df       MS      Number of obs   =       335
-------------+----------------------------------   F(10, 324)      =     36.39
       Model |  5.7603e-07        10  5.7603e-08   Prob > F        =    0.0000
    Residual |  5.1294e-07       324  1.5831e-09   R-squared       =    0.5290
-------------+----------------------------------   Adj R-squared   =    0.5144
       Total |  1.0890e-06       334  3.2604e-09   Root MSE        =   4.0e-05

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |     .00002   3.75e-06     5.32   0.000     .0000126    .0000273
      unrate |  -1.10e-06   1.19e-06    -0.92   0.356    -3.45e-06    1.24e-06
        mlda |  -6.95e-07   2.66e-06    -0.26   0.794    -5.93e-06    4.54e-06
      perinc |  -1.24e-08   1.51e-09    -8.23   0.000    -1.54e-08   -9.45e-09
     beertax |  -.0000202   6.69e-06    -3.02   0.003    -.0000334   -7.08e-06
      sobapt |   2.54e-06   3.82e-07     6.64   0.000     1.78e-06    3.29e-06
         dry |  -6.25e-07   3.13e-07    -2.00   0.046    -1.24e-06   -1.02e-08
      vmiles |   1.18e-08   1.67e-09     7.08   0.000     8.53e-09    1.51e-08
       jaild |   .0000192   6.20e-06     3.09   0.002     6.99e-06    .0000314
     comserd |   .0000157   7.00e-06     2.24   0.026     1.91e-06    .0000295
       _cons |    .000257   .0000626     4.10   0.000     .0001338    .0003803
------------------------------------------------------------------------------
```

Now, instead of running a regression model with all variables, we have taken an approach to analyze the variables based on their effect towards vehicle fatality rates by categorizing them into intuitive groups.

Group - 1: VFR & Nighttime VFR & Alcohol Aided VFR

a) Drinking Laws:

We intended to analyze the factors that were legal and expected to be affecting VFR directly. We included the variables ‘beertax’, ‘mlda’, ‘jaild’ and ‘comserd’ which are legal actions taken by the government in the United States of America.

Drinking Laws vs Total VFR

```
. xtreg mrall beertax mlda jaild comserd ,fe vce (cluster state)

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

R-sq:                                           Obs per group:
     within  = 0.0450                                         min =          6
     between = 0.1136                                         avg =        7.0
     overall = 0.0953                                         max =          7

                                                F(3,47)           =          .
corr(u_i, Xb)  = -0.6819                        Prob > F          =          .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0000639     .00003    -2.13   0.038    -.0001242   -3.53e-06
        mlda |   1.22e-06   2.66e-06     0.46   0.649    -4.13e-06    6.57e-06
       jaild |  -5.52e-06   3.49e-07   -15.83   0.000    -6.23e-06   -4.82e-06
     comserd |   .0000113    .000013     0.87   0.391    -.0000149    .0000374
       _cons |   .0002115   .0000622     3.40   0.001     .0000864    .0003365
-------------+----------------------------------------------------------------
     sigma_u |  .00007083
     sigma_e |  .00001908
         rho |  .93235972   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Here, we took All vehicle Fatality rates as the dependent variable and carried out FE model, with fixed state effects. We could see that only 'mlda' and 'comserd' are insignificant.

Drinking laws vs Nighttime VFR

```
. xtreg mralln beertax mlda jaild comserd , fe vce(cluster state)

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

R-sq:                                           Obs per group:
     within  = 0.0397                                         min =          6
     between = 0.0104                                         avg =        7.0
     overall = 0.0018                                         max =          7

                                                F(3,47)            =         .
corr(u_i, Xb)  = -0.4083                        Prob > F           =         .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      mralln |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -4.92e-06   7.79e-06    -0.63   0.531    -.0000206    .0000108
        mlda |  -1.40e-06   5.80e-07    -2.42   0.020    -2.57e-06   -2.36e-07
       jaild |  -2.57e-07   9.07e-08    -2.83   0.007    -4.39e-07   -7.44e-08
     comserd |  -5.70e-06   3.03e-06    -1.88   0.066    -.0000118    3.88e-07
       _cons |   .0000711   .0000142     5.01   0.000     .0000426    .0000997
-------------+----------------------------------------------------------------
     sigma_u |  .00001015
     sigma_e |  6.487e-06
         rho |  .70986444   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

In this regression output, we could see that 'beertax' is insignificant and others are significant at 10%.

Drinking Laws vs Alcohol aided VFR

```
. xtreg mraidall beertax mlda jaild comserd , fe vce (cluster state)

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

R-sq:                                           Obs per group:
     within  = 0.0396                                         min =          6
     between = 0.0001                                         avg =        7.0
     overall = 0.0002                                         max =          7

                                                F(3,47)            =         .
corr(u_i, Xb)  = -0.4583                        Prob > F           =         .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
    mraidall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0000134   .0000242    -0.56   0.581     -.000062    .0000352
        mlda |  -3.37e-06   1.58e-06    -2.14   0.038    -6.55e-06   -1.99e-07
       jaild |   .0000199   2.81e-07    70.90   0.000     .0000194    .0000205
     comserd |  -.0000223   8.77e-06    -2.55   0.014      -.00004   -4.70e-06
       _cons |   .0001404   .0000388     3.62   0.001     .0000624    .0002185
-------------+----------------------------------------------------------------
     sigma_u |  .00002586
     sigma_e |  .00001352
         rho |  .78542016   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

In this Regression output, we could see that the ‘beertax’ is insignificant and the others are significant at 5%.

From the above Regressions we can see that ‘beertax’ is significant for the VFR in general but is seen as insignificant for nighttime VFR and aided VFR , ‘jaild’ is definitely significant, and ‘mlda’ followed the same scenario as of ‘beertax’. As for ‘comserd’, it was insignificant for total VFR but improved for nighttime VFR and aided VFR. This can be explained because the law id directly related to alcohol induced fatality.

b) Social & Economic Variables:

We intended to analyze the factors that are Social and Economic and expected to be affecting VFR directly. We included the variables ‘unrate’, ‘perinc’, ‘dry’, ‘gspch’, ‘sobapt’ and ‘mormon’ which are the social variables and economic variables across the United States of America.

Social & Economic Variables vs Total VFR

```
. xtreg mrall unrate perinc sobapt mormon 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  = 0.1221                                         min =          7
     between = 0.0720                                         avg =        7.0
     overall = 0.0605                                         max =          7

                                                F(6,47)           =       7.18
corr(u_i, Xb)  = -0.8861                        Prob > F          =     0.0000

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      unrate |  -3.71e-06   1.18e-06    -3.14   0.003    -6.09e-06   -1.33e-06
      perinc |   4.62e-10   2.92e-09     0.16   0.875    -5.40e-09    6.33e-09
      sobapt |  -6.54e-06    .000011    -0.60   0.553    -.0000286    .0000155
      mormon |   6.18e-06   3.75e-06     1.65   0.106    -1.36e-06    .0000137
         dry |   2.96e-06   1.86e-06     1.59   0.119    -7.88e-07    6.71e-06
       gspch |  -.0000475   .0000275    -1.72   0.091    -.0001029    7.93e-06
       _cons |    .000243   .0001014     2.40   0.021     .0000391    .0004469
-------------+----------------------------------------------------------------
     sigma_u |  .00011464
     sigma_e |  .00001832
         rho |  .97508732   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

In the regression output, we can see that all the variables are insignificant except for ‘unrate’. While ‘gspch’ and ‘dry’ are close to significance at 10%.

Social & Economic Variables vs Nighttime VFR

In the regression output below, we can see that all the variables are insignificant at 5% significance but ‘mormon’, ‘gspch’ are significant at 10% and ‘perinc’ is close to significant at 10 % significance.

```
. xtreg mralln unrate perinc sobapt mormon 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  = 0.0512                                         min =          7
     between = 0.0173                                         avg =        7.0
     overall = 0.0140                                         max =          7

                                                F(6,47)           =       3.10
corr(u_i, Xb)  = -0.8883                        Prob > F          =     0.0122

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
      mralln |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      unrate |  -1.61e-07   3.03e-07    -0.53   0.598    -7.71e-07    4.49e-07
      perinc |  -1.16e-09   6.93e-10    -1.68   0.101    -2.56e-09    2.33e-10
      sobapt |   8.63e-07   3.15e-06     0.27   0.785    -5.47e-06    7.19e-06
      mormon |   1.48e-06   7.73e-07     1.92   0.061    -7.23e-08    3.04e-06
         dry |   5.47e-07   7.30e-07     0.75   0.458    -9.23e-07    2.02e-06
       gspch |  -.0000242    .0000121   -2.00   0.051    -.0000486    1.40e-07
       _cons |    .000044    .0000272    1.62   0.112    -.0000107    .0000987
-------------+----------------------------------------------------------------
     sigma_u |  .00002004
     sigma_e |  6.460e-06
         rho |  .90585846   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Social & Economic Variables vs Alcohol aided VFR

```
. xtreg mraidall unrate perinc sobapt mormon 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  = 0.0270                                         min =          7
     between = 0.0737                                         avg =        7.0
     overall = 0.0523                                         max =          7

                                                F(6,47)           =       1.83
corr(u_i, Xb)  = -0.7957                        Prob > F          =     0.1134

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
    mraidall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      unrate |  -1.21e-06   7.70e-07    -1.57   0.124    -2.75e-06    3.43e-07
      perinc |  -4.07e-09   1.44e-09    -2.83   0.007    -6.97e-09   -1.18e-09
      sobapt |  -2.25e-06   8.97e-06    -0.25   0.803    -.0000203     .0000158
      mormon |   1.13e-06   1.33e-06     0.85   0.401    -1.55e-06    3.80e-06
         dry |   6.44e-07   2.33e-06     0.28   0.784    -4.05e-06    5.34e-06
       gspch |  -8.20e-06    .0000225   -0.36   0.717    -.0000534      .000037
       _cons |   .0001417    .0000786    1.80   0.078    -.0000164     .0002998
-------------+----------------------------------------------------------------
     sigma_u |  .00003659
     sigma_e |  .00001364
         rho |  .87805824   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

In this regression output, we can see that all variables except ‘perinc’ are insignificant.

From the above regressions on Social & Economic variables, we can see that no variable is consistently significant across all three VFRs. But we believe that these Social & Economic variables have a role to play when they are interacted with the laws imposed by the government.

Group - 2: VFR across Age Groups (15-17, 18-20, 21-24):

From the data we observed that age groups between 15-24 are highly involved in the vehicle fatalities. So, we wanted to know the effect of different variables on the VFRs of the particular age groups. This might give us an insight if a certain group specific VFRs are more sensitive to certain laws or other variables or not.

a) Drinking Laws:

We intended to analyze the factors that were legal and expected to be affecting VFR across age groups. So, we initially included the variables 'beertax', 'mlda', 'jaild' and 'comserd' which are legal actions taken by the government in the United States of America.

Drinking Laws vs VFR (15-17)

```
. xtreg mra1517 beertax mlda jaild comserd , fe vce (cluster state)

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

R-sq:                                           Obs per group:
     within  = 0.1043                                         min =          6
     between = 0.1246                                         avg =        7.0
     overall = 0.0454                                         max =          7

                                                F(3,47)           =          .
corr(u_i, Xb)  = -0.7883                        Prob > F          =          .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
     mra1517 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0001403   .0000493    -2.85   0.007    -.0002395   -.0000412
        mlda |   .0000228   5.70e-06     4.00   0.000     .0000114    .0000343
       jaild |  -.0000222   5.74e-07   -38.61   0.000    -.0000233    -.000021
     comserd |   .0000778   .0000315     2.47   0.017     .0000145    .0001411
       _cons |  -.0000993    .000124    -0.80   0.427    -.0003487      .00015
-------------+----------------------------------------------------------------
     sigma_u |  .00011956
     sigma_e |  .00006027
         rho |  .79735678   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From the output of the regression above, we can see that all the variables, 'beertax', 'mlda', 'jaild', 'comserd' are highly significant at 5% significance level. This can be explained because the age group is not eligible to drink and drive in certain states. The fear of apprehension and the amount of money at hand might play a huge role on the variables. So, they are significant.

Drinking Laws vs VFR (18-20)

```
. xtreg mra1820 beertax mlda jaild comserd , fe vce (cluster state)

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

R-sq:                                           Obs per group:
     within  = 0.0347                                         min =          6
     between = 0.0099                                         avg =        7.0
     overall = 0.0049                                         max =          7

                                                F(3,47)            =          .
corr(u_i, Xb)  = -0.6718                        Prob > F           =          .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
     mra1820 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0002229   .0001183    -1.88   0.066     -.000461    .0000151
        mlda |  -.0000138   8.34e-06    -1.65   0.106    -.0000305    3.03e-06
       jaild |  -.0000217   1.38e-06   -15.79   0.000    -.0000245    -.000019
     comserd |  -.0000135   .0000557    -0.24   0.809    -.0001256    .0000985
       _cons |   .0008774   .0002039     4.30   0.000     .0004671    .0012876
-------------+----------------------------------------------------------------
     sigma_u |  .00017727
     sigma_e |   .0000853
         rho |  .81198133   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From the output of the regression above, we can see that all the variables, ‘mlda’, ‘comserd’ are insignificant at 10% significance level.

Drinking Laws vs VFR (21-24)

```
. xtreg mra2124 beertax mlda jaild comserd , fe vce (cluster state)

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

R-sq:                                           Obs per group:
     within  = 0.0272                                         min =          6
     between = 0.0416                                         avg =        7.0
     overall = 0.0241                                         max =          7

                                                F(3,47)            =          .
corr(u_i, Xb)  = -0.6347                        Prob > F           =          .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
     mra2124 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0001245   .0000899    -1.38   0.173    -.0003055    .0000564
        mlda |   .0000104   7.59e-06     1.37   0.178    -4.88e-06    .0000256
       jaild |   .0000206   1.05e-06    19.67   0.000     .0000185    .0000227
     comserd |  -.0000501   .0000525    -0.95   0.345    -.0001558    .0000556
       _cons |   .0002643   .0001762     1.50   0.140    -.0000902    .0006188
-------------+----------------------------------------------------------------
     sigma_u |  .00013353
     sigma_e |  .00007042
         rho |  .78242365   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From the output of the regression above, we can see that all the variables are insignificant except ‘jaild’ even at 10 % significance level.

From the above regressions, We see that 'jaild' is consistently significant across all age groups and 'beertax' and 'mlda' became insignificant with increase in age. But we would like to explore the age specific VFR by including the social and economic variables to measure the true effect and also to remove omitted variable bias.

b) Laws & Social & Economic variables:

We intended to analyze all the significant Social, Economic and legal factors that are expected to be affecting VFR across age groups directly. We included the variables, 'beertax', 'mlda', 'jaild', 'comserd', 'spircons', 'unrate', 'perinc', 'sobapt', 'dry', 'vmiles' against the age specific fatalities.

vs VFR (15-17)

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

R-sq:                                           Obs per group:
     within  = 0.1660                                         min =          6
     between = 0.0086                                         avg =        7.0
     overall = 0.0169                                         max =          7

                                                F(9,47)            =          .
corr(u_i, Xb)  = -0.7641                        Prob > F           =          .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
     mra1517 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |   -.000088   .0000645    -1.37   0.179    -.0002177    .0000417
        mlda |   .0000156   6.31e-06     2.47   0.017     2.89e-06    .0000283
       jaild |   -.000012   3.04e-06    -3.94   0.000    -.0000181   -5.87e-06
     comserd |   .0000468   .0000322     1.46   0.152    -.0000179    .0001115
    spircons |   .0000446   .0000328     1.36   0.180    -.0000213    .0001106
      unrate |  -5.77e-06   3.79e-06    -1.52   0.134    -.0000134    1.85e-06
      perinc |   7.48e-09   8.35e-09     0.90   0.375    -9.31e-09    2.43e-08
      sobapt |   .0000112   .0000131     0.86   0.393     -.000015    .0000375
         dry |   2.69e-06   3.78e-06     0.71   0.480    -4.92e-06    .0000103
      vmiles |   7.70e-09   2.36e-09     3.27   0.002     2.96e-09    1.24e-08
       _cons |  -.0002676   .0002177    -1.23   0.225    -.0007056    .0001705
-------------+----------------------------------------------------------------
     sigma_u |  .00011877
     sigma_e |  .00005879
         rho |   .8032051   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From the regression output above, we can see that 'mlda', 'jaild', 'vmiles' are highly significant for the fatalities due to the age group 15- 17.

vs VFR (18-20)

From the regression output below, we see that 'jaild', 'spircons', 'unrate', 'perinc' are found to be significant at 10% significance level.

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

R-sq:                                           Obs per group:
     within  = 0.1259                                         min =          6
     between = 0.0673                                         avg =        7.0
     overall = 0.0427                                         max =          7

                                                F(9,47)            =         .
corr(u_i, Xb)  = -0.9570                        Prob > F           =         .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
     mra1820 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0001529   .0001048    -1.46   0.151    -.0003637    .0000578
        mlda |  -.0000151   9.65e-06    -1.56   0.126    -.0000345    4.37e-06
       jaild |  -.0000131   3.09e-06    -4.24   0.000    -.0000193   -6.89e-06
     comserd |   -.000042   .0000465    -0.90   0.371    -.0001355    .0000515
    spircons |   .0001873   .0000595     3.15   0.003     .0000677    .0003069
      unrate |  -6.95e-06   3.95e-06    -1.76   0.085    -.0000149    1.00e-06
      perinc |   2.59e-08   1.06e-08     2.43   0.019     4.48e-09    4.73e-08
      sobapt |   -.000026   .0000311    -0.84   0.407    -.0000887    .0000366
         dry |   2.19e-06   3.65e-06     0.60   0.551    -5.15e-06    9.52e-06
      vmiles |  -3.86e-09   3.86e-09    -1.00   0.323    -1.16e-08    3.91e-09
       _cons |   .0004419    .000433     1.02   0.313    -.0004293     .001313
-------------+----------------------------------------------------------------
     sigma_u |   .0004487
     sigma_e |  .00008205
         rho |  .96764724   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

vs VFR (21-24)

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

R-sq:                                           Obs per group:
     within  = 0.1377                                         min =          6
     between = 0.1152                                         avg =        7.0
     overall = 0.0731                                         max =          7

                                                F(9,47)            =         .
corr(u_i, Xb)  = -0.9598                        Prob > F           =         .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
     mra2124 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.0000401   .0000729    -0.55   0.585    -.0001868    .0001065
        mlda |   6.55e-06   7.59e-06     0.86   0.393    -8.73e-06    .0000218
       jaild |   .0000316   3.17e-06     9.99   0.000     .0000253     .000038
     comserd |   -.000085   .0000494    -1.72   0.092    -.0001843    .0000144
    spircons |   .0001482   .0000422     3.51   0.001     .0000633    .0002332
      unrate |  -.0000105   3.52e-06    -2.97   0.005    -.0000176   -3.38e-06
      perinc |   1.53e-08   1.08e-08     1.42   0.164    -6.43e-09    3.70e-08
      sobapt |  -.0000265   .0000215    -1.23   0.224    -.0000698    .0000168
         dry |   7.01e-06   3.87e-06     1.81   0.076    -7.76e-07    .0000148
      vmiles |  -6.45e-10   1.17e-09    -0.55   0.582    -2.99e-09    1.70e-09
       _cons |   .0000731   .0003556     0.21   0.838    -.0006423    .0007885
-------------+----------------------------------------------------------------
     sigma_u |  .00036334
     sigma_e |  .00006701
         rho |  .96710274   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

From the regression output above, we see that ‘jaild’, ‘comserd’, ‘spircons’, ‘unrate’ and ‘dry’ are significant at 10% significance levels.

From all regressions above, we can see that ‘jaild’ is consistently significant across all age groups. ‘beertax’, ‘mlda’ became insignificant as the age increased. But ‘unrate’ became significant with the increase in age.

Fixed Effects Model :

So far we’ve looked at the effects of specific variables across different categories of VFRs. Now, we would like to run a fixed effects regression across all the variables for the Total VFR grouped by state to remove any bias and to remove any serial correlation that might be induced due to time. And we use Cluster robust standard errors because although the estimators are linear and unbiased, the standard errors are not robust. And they can yield different results for confidence intervals and hypothesis testing. So, whenever we use a fixed effects model, we use cluster robust standard errors. We have used cluster robust standard errors in all the above fixed effects models as well.

Fixed effect model with robust standard errors for all variables:

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

R-sq:                                           Obs per group:
     within  = 0.3761                                         min =          6
     between = 0.1594                                         avg =        7.0
     overall = 0.1242                                         max =          7

                                                F(12,47)          =          .
corr(u_i, Xb)  = -0.9064                        Prob > F          =          .

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000832   .0000146     5.71   0.000     .0000539    .0001125
      unrate |  -3.59e-06   1.10e-06    -3.26   0.002    -5.80e-06   -1.37e-06
        mlda |   1.94e-06   2.26e-06     0.86   0.394    -2.60e-06    6.48e-06
      perinc |   9.60e-09   3.32e-09     2.89   0.006     2.92e-09    1.63e-08
     beertax |  -.0000414   .0000305    -1.36   0.182    -.0001028      .00002
      sobapt |  -3.21e-06   7.78e-06    -0.41   0.681    -.0000189    .0000124
      mormon |  -5.27e-07   5.29e-06    -0.10   0.921    -.0000112    .0000101
         dry |   2.70e-06   1.28e-06     2.11   0.040     1.24e-07    5.27e-06
      vmiles |   1.19e-09   7.68e-10     1.54   0.129    -3.59e-10    2.73e-09
       jaild |   1.22e-06   1.27e-06     0.96   0.342    -1.33e-06    3.77e-06
     comserd |  -1.45e-06   .0000143    -0.10   0.920    -.0000303    .0000274
       gspch |  -.0000543   .0000235    -2.31   0.025    -.0001016   -7.05e-06
      yngdrv |   .0000399   .0000617     0.65   0.521    -.0000843    .0001641
       _cons |  -.0000695   .0001298    -0.54   0.595    -.0003306    .0001916
-------------+----------------------------------------------------------------
     sigma_u |  .00012293
     sigma_e |  .00001567
         rho |  .98401103   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

We see many variables which are highly significant and we start a stepwise regression by removing different insignificant variables.

Final Fixed effects model :

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

R-sq:                                           Obs per group:
     within  = 0.3650                                         min =          6
     between = 0.0983                                         avg =        7.0
     overall = 0.0715                                         max =          7

                                                F(7,47)           =      15.30
corr(u_i, Xb)  = -0.8658                        Prob > F          =     0.0000

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000842   .0000136     6.21   0.000      .000057    .0001115
      unrate |  -2.91e-06   9.70e-07    -3.00   0.004    -4.86e-06   -9.56e-07
        mlda |   2.20e-06   2.20e-06     1.00   0.323    -2.23e-06    6.62e-06
      perinc |   1.01e-08   3.22e-09     3.12   0.003     3.58e-09    1.65e-08
     beertax |  -.0000485   .0000233    -2.08   0.043    -.0000954   -1.65e-06
         dry |   2.64e-06   1.27e-06     2.09   0.042     9.39e-08    5.19e-06
       jaild |  -6.62e-07   9.88e-06    -0.07   0.947    -.0000205    .0000192
       _cons |   -.000093   .0000996    -0.93   0.355    -.0002933    .0001074
-------------+----------------------------------------------------------------
     sigma_u |  .00010704
     sigma_e |  .00001564
         rho |  .97909889   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

After running models of different variable combinations, we have arrived at the following model. Despite the high insignificance of 'jaild' variable, we chose to include it in the model because it proved to highly significant for the age group of 15 – 24. And also, we decided to have 'mlda' in the regression because removing that variable is introducing high bias in other significant variables.

Time Fixed Effects Model :

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

R-sq:                                           Obs per group:
     within  = 0.4602                                         min =          6
     between = 0.0970                                         avg =        7.0
     overall = 0.0658                                         max =          7

                                                F(13,47)          =      13.15
corr(u_i, Xb)  = -0.8576                        Prob > F          =     0.0000

                                 (Std. Err. adjusted for 48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       mrall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000813    .000012     6.80   0.000     .0000572    .0001053
      unrate |  -5.49e-06   1.20e-06    -4.57   0.000    -7.90e-06   -3.07e-06
        mlda |   1.35e-06   2.15e-06     0.63   0.533    -2.97e-06    5.67e-06
      perinc |   8.52e-09   3.17e-09     2.68   0.010     2.13e-09    1.49e-08
     beertax |  -.0000456   .0000243    -1.88   0.067    -.0000945    3.25e-06
         dry |   1.99e-06   1.01e-06     1.97   0.055    -4.67e-08    4.02e-06
       jaild |   4.29e-06   9.81e-06     0.44   0.664    -.0000154     .000024
             |
        year |
       1983  |  -5.58e-06   3.06e-06    -1.82   0.075    -.0000117    5.82e-07
       1984  |  -.0000166   4.34e-06    -3.82   0.000    -.0000253   -7.84e-06
       1985  |  -.0000202   4.90e-06    -4.11   0.000      -.00003   -.0000103
       1986  |  -5.63e-06   6.40e-06    -0.88   0.383    -.0000185    7.24e-06
       1987  |  -.0000106   7.37e-06    -1.45   0.155    -.0000255    4.17e-06
       1988  |   -.000014   8.42e-06    -1.67   0.102     -.000031    2.90e-06
             |
       _cons |  -.0000197   .0000951    -0.21   0.837    -.0002109    .0001715
-------------+----------------------------------------------------------------
     sigma_u |  .00010503
     sigma_e |  .00001458
         rho |   .9811049   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The output of this model reduces the estimators of the regression, while it changes the sign of jaild from negative to positive. We have accounted for time variation, but we could not arrive at a best estimator the explanatory variables. So, it is practical not to use this regression result for further analysis, and hence we stick with Fixed effects with cluster robust standard errors.

Random Effects model

```
. *Random Effects
. xtreg mrall spircons unrate mlda perinc beertax  dry  jaild  , re

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

R-sq:                                           Obs per group:
     within  = 0.2522                                         min =          6
     between = 0.0012                                         avg =        7.0
     overall = 0.0000                                         max =          7

                                                Wald chi2(7)      =      59.97
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       mrall |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spircons |   .0000354   6.81e-06     5.19   0.000      .000022    .0000487
      unrate |  -4.81e-06   9.78e-07    -4.92   0.000    -6.73e-06   -2.89e-06
        mlda |   1.07e-06   1.98e-06     0.54   0.590    -2.81e-06    4.95e-06
      perinc |  -6.47e-10   2.00e-09    -0.32   0.746    -4.56e-09    3.26e-09
     beertax |   2.06e-07   .0000121     0.02   0.986    -.0000235    .0000239
         dry |   1.93e-06   6.93e-07     2.79   0.005     5.75e-07    3.29e-06
       jaild |   6.77e-06   6.29e-06     1.08   0.282    -5.56e-06    .0000191
       _cons |   .0001543   .0000548     2.81   0.005     .0000468    .0002618
-------------+----------------------------------------------------------------
     sigma_u |  .00004196
     sigma_e |  .00001564
         rho |  .87799392   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The regression output, says us the variables beertax and perinc are insignificant, which we considered to be a significant factors during our explanatory data analysis. We are not so confident about the random effect regression effect , due to the fact that the regression has been executed post removing few variables, which might have caused endogeneity problem and also, since the model is random, it is not expected to function ideally for variables that are varying slowly across time. So, we expect Fixed model effects with Cluster robust standard error to provide us a proper estimates compared to all other analyzed models.

Hausman Test

To determine if exists any heterogeneity problem, we conduct Hausman test between Fixed and Random Effects models.

```
. *Hausman Test
. hausman fixed random

Note: the rank of the differenced variance matrix (6) does not equal the number
        of coefficients being tested (7); be sure this is what you expect, or
        there may be problems computing the test.  Examine the output of your
        estimators for anything unexpected and possibly consider scaling your
        variables so that the coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
    spircons |    .0000842     .0000354        .0000489        4.69e-06
      unrate |   -2.91e-06    -4.81e-06        1.90e-06               .
        mlda |    2.20e-06     1.07e-06        1.13e-06               .
      perinc |    1.01e-08    -6.47e-10        1.07e-08        6.02e-10
     beertax |   -.0000485     2.06e-07       -.0000487         .000011
         dry |    2.64e-06     1.93e-06        7.07e-07        1.09e-06
       jaild |   -6.62e-07     6.77e-06       -7.43e-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)
                          =      119.62
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

Based on the result of the test, be reject the null hypothesis and conclude that the Fixed Effects model with cluster robust standard error is the best model for the given data and across all analyzed models.

Conclusion:

We saw the significance of ‘jaild’ in general for all categories of VFRs, but it became insignificant mostly due to omitted variable bias. Beer tax became insignificant with the increase in age and with the increase in time. But this could be because, as age increased, the probability of having more income is high, this can explain its insignificance with increase to age. As for the insignificance with time might probably because of the decline in the average beer tax with time. Unemployment became highly significant for the VFRs with increase in age. So, this must be given more importance too. There are a lot of other factors that are involved in vehicle fatalities like the weather conditions, condition of the roads, size of the roads, accidents on highways vs accidents in cities. Based on our findings, our laws probably might have to revised based on age, per capita alcohol consumption and unemployment. But there is no generalized outcome of the changes in laws across the entire nation.