Lab 3: Panel Models

US Traffic Fatalities: 1980 - 2004

Contents

1 U.S. traffic fatalities: 1980-2004

In this lab, we are asking you to answer the following **causal** question:

> **"Do changes in traffic laws affect traffic fatalities?"**

To answer this question, please complete the tasks specified below using the data provided in `data/driving.Rdata`. This data includes 25 years of data that cover changes in various state drunk driving, seat belt, and speed limit laws.

Specifically, this data set contains data for the 48 continental U.S. states from 1980 through 2004. Various driving laws are indicated in the data set, such as the alcohol level at which drivers are considered legally intoxicated. There are also indicators for "per se" laws—where licenses can be revoked without a trial—and seat belt laws. A few economics and demographic variables are also included. The description of the each of the variables in the dataset is also provided in the dataset.

```
load(file="./data/driving.RData")

## please comment these calls in your work
glimpse(data)
```

```
## Rows: 1,200
## Columns: 56
## $ year          <int> 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 198~
## $ state         <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ sl55          <dbl> 1.000, 1.000, 1.000, 1.000, 1.000, 1.000, 1.000, 0.542, 0~
## $ sl65          <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.458, 1~
## $ sl70          <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0~
## $ sl75          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

```
## $ slnone       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ seatbelt     <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, ~
## $ minage       <dbl> 18, 18, 18, 18, 18, 20, 21, 21, 21, 21, 21, 21, 21, 21, 2~
## $ zerotol      <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0~
## $ gdl          <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.0~
## $ bac10        <dbl> 1.000, 1.000, 1.000, 1.000, 1.000, 1.000, 1.000, 1.000, 1~
## $ bac08        <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0~
## $ perse        <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0~
## $ totfat       <int> 940, 933, 839, 930, 932, 882, 1080, 1111, 1024, 1029, 112~
## $ nghtfat      <int> 422, 434, 376, 397, 421, 358, 500, 499, 423, 418, 466, 47~
## $ wkndfat      <int> 236, 248, 224, 223, 237, 224, 279, 300, 226, 247, 271, 27~
## $ totfatpvm    <dbl> 3.200, 3.350, 2.810, 3.000, 2.830, 2.510, 3.177, 2.970, 2~
## $ nghtfatpvm   <dbl> 1.437, 1.558, 1.259, 1.281, 1.278, 1.019, 1.471, 1.334, 1~
## $ wkndfatpvm   <dbl> 0.803, 0.890, 0.750, 0.719, 0.720, 0.637, 0.821, 0.802, 0~
## $ statepop     <int> 3893888, 3918520, 3925218, 3934109, 3951834, 3972527, 399~
## $ totfatrte    <dbl> 24.14, 24.07, 21.37, 23.64, 23.58, 22.20, 27.08, 27.67, 2~
## $ nghtfatrte   <dbl> 10.84, 11.08, 9.58, 10.09, 10.65, 9.01, 12.53, 12.43, 10.~
## $ wkndfatrte   <dbl> 6.060000, 6.330000, 5.710000, 5.670000, 6.000000, 5.64000~
## $ vehicmiles   <dbl> 29.37500, 27.85200, 29.85765, 31.00000, 32.93286, 35.1394~
## $ unem         <dbl> 8.8, 10.7, 14.4, 13.7, 11.1, 8.9, 9.8, 7.8, 7.2, 7.0, 6.9~
## $ perc14_24    <dbl> 18.9, 18.7, 18.4, 18.0, 17.6, 17.3, 17.0, 16.6, 16.2, 15.~
## $ sl70plus     <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0~
## $ sbprim       <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ sbsecon      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, ~
## $ d80          <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d81          <int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d82          <int> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d83          <int> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d84          <int> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d85          <int> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d86          <int> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d87          <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d88          <int> 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d89          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d90          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d91          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ~
## $ d92          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, ~
## $ d93          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, ~
## $ d94          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, ~
## $ d95          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, ~
## $ d96          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
## $ d97          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, ~
## $ d98          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, ~
## $ d99          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d00          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d01          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d02          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d03          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ d04          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ vehicmilespc <dbl> 7543.874, 7107.785, 7606.622, 7879.802, 8333.562, 8845.61~
```

```
desc
```

```
##       variable                                              label
## 1         year                                  1980 through 2004
```

```
## 2         state           48 continental states, alphabetical
## 3          sl55                           speed limit == 55
## 4          sl65                           speed limit == 65
## 5          sl70                           speed limit == 70
## 6          sl75                           speed limit == 75
## 7        slnone                              no speed limit
## 8      seatbelt     =0 if none, =1 if primary, =2 if secondary
## 9        minage                        minimum drinking age
## 10      zerotol                          zero tolerance law
## 11          gdl                graduated drivers license law
## 12        bac10                      blood alcohol limit .10
## 13        bac08                      blood alcohol limit .08
## 14        perse administrative license revocation (per se law)
## 15       totfat                      total traffic fatalities
## 16      nghtfat                    total nighttime fatalities
## 17      wkndfat                      total weekend fatalities
## 18    totfatpvm         total fatalities per 100 million miles
## 19   nghtfatpvm     nighttime fatalities per 100 million miles
## 20   wkndfatpvm       weekend fatalities per 100 million miles
## 21     statepop                              state population
## 22    totfatrte        total fatalities per 100,000 population
## 23   nghtfatrte    nighttime fatalities per 100,000 population
## 24   wkndfatrte       weekend accidents per 100,000 population
## 25   vehicmiles              vehicle miles traveled, billions
## 26         unem                    unemployment rate, percent
## 27    perc14_24          percent population aged 14 through 24
## 28     sl70plus                          sl70 + sl75 + slnone
## 29       sbprim                    =1 if primary seatbelt law
## 30      sbsecon                  =1 if secondary seatbelt law
## 31          d80                               =1 if year == 1980
## 32          d81
## 33          d82
## 34          d83
## 35          d84
## 36          d85
## 37          d86
## 38          d87
## 39          d88
## 40          d89
## 41          d90
## 42          d91
## 43          d92
## 44          d93
## 45          d94
## 46          d95
## 47          d96
## 48          d97
## 49          d98
## 50          d99
## 51          d00
## 52          d01
## 53          d02
## 54          d03
## 55          d04                               =1 if year == 2004
```

```
## 56 vehicmilespc
```

```r
print("seatbelt values not 0, 1 or 2")
```

```
## [1] "seatbelt values not 0, 1 or 2"
```

```r
summary(data$seatbelt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   1.000   1.116   2.000   2.000
```

```r
sb<-data$seatbelt[ (data$seatbelt != 1) & (data$seatbelt != 2) & (data$seatbelt != 0)]
prop_sb<-length(sb)/length(data$seatbelt)
head(sb)
```

```
## integer(0)
```

```r
class(data$seatbelt)
```

```
## [1] "integer"
```

```r
unique(data$seatbelt)
```

```
## [1] 0 2 1
```

```r
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

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

```r
Variable <- c("Seatbelt Values")
Proportion <- c(prop_sb)

prop_df <- data.frame(Variable, Proportion)


print("minim age values not 18 or 21")
```

```
## [1] "minim age values not 18 or 21"
```

```r
summary(data$minage)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    18.0    21.0    21.0    20.6    21.0    21.0
```

```r
sb<-data$minage[ (data$minage != 18) & (data$minage != 21) & (data$minage != 19) ]
head(sb,5)
```

```
## [1] 20.0 19.5 18.5 19.5 20.0
```

```r
class(data$minage)
```

```
## [1] "numeric"
```

```r
unique(data$minage)
```

```
##  [1] 18.0 20.0 21.0 19.5 18.5 20.5 19.0 18.7 19.7 20.7 19.8 18.6
```

```r
prop_sb<-length(sb)/ length(data$minage)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.04916667
```

```r
Variable <- c("Minimum Age")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)

print("zero tolerance values not 0 or 1")
```

```
## [1] "zero tolerance values not 0 or 1"
```

```r
sb<-data$zerotol[ (data$zerotol != 0) & (data$zerotol != 1) ]
unique(data$zerotol)
```

```
##  [1] 0.000 0.667 1.000 0.250 0.583 0.500 0.167 0.417 0.083 0.333 0.750
```

```r
class(data$zerotol)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$zerotol)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.0325
```

```r
Variable <- c("Zero Tolerance")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)


print("gdl values not 0 or 1")
```

```
## [1] "gdl values not 0 or 1"
```

```r
unique(data$gdl)
```

```
## [1] 0.000 0.750 1.000 0.500 0.250 0.167 0.670 0.833
```

```r
sb<-data$gdl[ (data$gdl != 0) & (data$gdl != 1) ]
class(data$gdl)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$gdl)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.01666667
```

```r
Variable <- c("gdl")
Proportion <- c(prop_sb)
```

```r
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)



print("perse values not 0 or 1")
```

```
## [1] "perse values not 0 or 1"
```

```r
unique(data$perse)
```

```
## [1] 0.000 0.417 1.000 0.500 0.167 0.250 0.333 0.750 0.083
```

```r
sb<-data$perse[ (data$perse != 0) & (data$perse != 1) ]
class(data$perse)
```

```
## [1] "numeric"
```

```r
unique(data$perse)
```

```
## [1] 0.000 0.417 1.000 0.500 0.167 0.250 0.333 0.750 0.083
```

```r
prop_sb<-length(sb)/length(data$perse)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.0225
```

```r
Variable <- c("perse")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)






print("sl65 values not 0 or 1")
```

```
## [1] "sl65 values not 0 or 1"
```

```r
sb<-data$sl65[ (data$sl65 != 0) & (data$sl65 != 1)]
unique(data$sl65)
```

```
##  [1] 0.000 0.458 1.000 0.333 0.750 0.917 0.583 0.667 0.016 0.500 0.250 0.956
## [13] 0.542 0.167 0.625 0.989 0.083 0.208 0.417 0.708 0.951 0.375 0.958
```

```r
class(data$sl65)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$sl65)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.06333333
```

```r
Variable <- c("sl 65")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)


print("sl70 values not 0 or 1")
```

```
## [1] "sl70 values not 0 or 1"
```

```r
sb<-data$sl70[ (data$sl70 != 0) & (data$sl70 != 1)]
unique(data$sl70)
```

```
##  [1] 0.000 0.667 1.000 0.083 0.417 0.984 0.750 0.500 0.833 0.375 0.792 0.583
## [13] 0.042 0.333
```

```r
class(data$sl70)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$sl70)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.01666667
```

```r
Variable <- c("sl 70")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)

print("sl75 values not 0 or 1")
```

```
## [1] "sl75 values not 0 or 1"
```

```r
sb<-data$sl75[ (data$sl75 != 0) & (data$sl75 != 1)]
unique(data$sl75)
```

```
## [1] 0.000 1.000 0.500 0.667 0.583 0.083 0.625 0.333 0.750
```

```r
class(data$sl75)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$sl75)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.0075
```

```r
Variable <- c("sl 75")
Proportion <- c(prop_sb)
```

```r
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)


print("sl70plus values not 0 or 1")
```

```
## [1] "sl70plus values not 0 or 1"
```

```r
sb<-data$sl70plus[ (data$sl70plus != 0) & (data$sl70plus!= 1) ]
unique(data$sl70plus)
```

```
##  [1] 0.000 0.667 1.000 0.083 0.417 0.984 0.500 0.750 0.833 0.375 0.792 0.583
## [13] 0.625 0.042 0.333
```

```r
class(data$sl70plus)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$sl70plus)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.02333333
```

```r
Variable <- c("sl 70 plus")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)

print("sl55 values not 0 or 1")
```

```
## [1] "sl55 values not 0 or 1"
```

```r
sb<-data$sl55[ (data$sl55 != 0) & (data$sl55 != 1)]
unique(data$sl55)
```

```
##  [1] 1.000 0.542 0.000 0.250 0.333 0.750 0.044 0.083 0.417 0.458 0.500 0.011
## [13] 0.917 0.292 0.049 0.583 0.375
```

```r
class(data$sl55)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$sl55)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.04
```

```r
Variable <- c("sl 55")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)
```

```r
print("bac10 values not 0 or 1")
```

```
## [1] "bac10 values not 0 or 1"
```

```r
sb<-data$bac10[ (data$bac10 != 0) & (data$bac10 != 1)]
unique(data$bac10)
```

```
##  [1] 1.000 0.583 0.000 0.417 0.667 0.750 0.833 0.500 0.250 0.333
```

```r
class(data$bac10)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$bac10)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.05416667
```

```r
Variable <- c("bac 10")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)

print("bac08 values not 0 or 1")
```

```
## [1] "bac08 values not 0 or 1"
```

```r
sb<-data$bac08[ (data$bac08 != 0) & (data$bac08 != 1)]
unique(data$bac08)
```

```
## [1] 0.000 0.417 1.000 0.333 0.500 0.250 0.750 0.667
```

```r
class(data$bac08)
```

```
## [1] "numeric"
```

```r
prop_sb<-length(sb)/length(data$bac08)
print("proportion:")
```

```
## [1] "proportion:"
```

```r
print(prop_sb)
```

```
## [1] 0.03333333
```

```r
Variable <- c("bac 08")
Proportion <- c(prop_sb)
df <- data.frame(Variable, Proportion)
prop_df<-rbind(prop_df, df)
```

```r
head(prop_df,20)
```

```
##          Variable Proportion
## 1 Seatbelt Values 0.00000000
## 2     Minimum Age 0.04916667
## 3  Zero Tolerance 0.03250000
## 4             gdl 0.01666667
```

```
## 5            perse 0.02250000
## 6            sl 65 0.06333333
## 7            sl 70 0.01666667
## 8            sl 75 0.00750000
## 9       sl 70 plus 0.02333333
## 10           sl 55 0.04000000
## 11          bac 10 0.05416667
## 12          bac 08 0.03333333
```

As they are not zero or one only, but have fractions to represent what portion of the year they were implemented. We will have to round some of the numbers, since the entries, that are fractions, per the table above, show the proportion to be small.

```
data %>% mutate(
  sum_bac = bac10 + bac08) %>% filter(sum_bac < 1 & sum_bac > 0)
```

```
##      year state sl55 sl65 sl70 sl75 slnone seatbelt minage zerotol gdl bac10
## 28   1982     3    1    0    0    0      0        0   18.0     0.0   0 0.417
## 54   1983     4    1    0    0    0      0        0   21.0     0.0   0 0.750
## 78   1982     5    1    0    0    0      0        0   21.0     0.0   0 0.833
## 104  1983     6    1    0    0    0      0        0   18.0     0.0   0 0.500
## 131  1985     7    1    0    0    0      0        0   20.5     0.0   0 0.250
## 154  1983     8    1    0    0    0      0        0   18.0     0.0   0 0.833
## 204  1983    11    1    0    0    0      0        0   19.0     0.0   0 0.333
## 230  1984    13    1    0    0    0      0        0   18.0     0.0   0 0.833
## 279  1983    15    1    0    0    0      0        0   21.0     0.0   0 0.333
## 331  1985    17    1    0    0    0      0        0   20.0     0.0   0 0.500
## 442  1996    21    0    1    0    0      0        1   21.0     1.0   0 0.667
## 479  1983    23    1    0    0    0      0        0   21.0     0.0   0 0.750
## 529  1983    25    1    0    0    0      0        0   18.0     0.0   0 0.500
## 579  1983    27    1    0    0    0      0        0   19.0     0.0   0 0.250
## 629  1983    29    1    0    0    0      0        0   21.0     0.0   0 0.500
## 654  1983    30    1    0    0    0      0        0   20.0     0.0   0 0.333
## 679  1983    31    1    0    0    0      0        0   21.0     0.0   0 0.750
## 705  1984    32    1    0    0    0      0        0   21.0     0.5   0 0.500
## 779  1983    35    1    0    0    0      0        0   21.0     0.0   0 0.500
## 804  1983    36    1    0    0    0      0        0   19.0     0.0   0 0.750
## 828  1982    37    1    0    0    0      0        0   18.0     0.0   0 0.500
## 904  1983    40    1    0    0    0      0        0   20.0     0.0   0 0.500
## 992  1996    43    0    1    0    0      0        2   21.0     1.0   0 0.667
## 1080 1984    47    1    0    0    0      0        0   19.0     0.0   0 0.500
## 1153 1982    50    1    0    0    0      0        0   18.0     0.0   0 0.667
##      bac08 perse totfat nghtfat wkndfat totfatpvm nghtfatpvm wkndfatpvm
## 28       0 0.000    724     337     177     3.550      1.652      0.868
## 54       0 0.000    557     232     122     3.340      1.391      0.732
## 78       0 0.000   4611    2419    1182     2.708      1.420      0.694
## 104      0 0.500    646     358     203     2.600      1.441      0.817
## 131      0 0.000    448     244     122     2.000      1.089      0.545
## 154      0 1.000    110      66      40     2.250      1.350      0.818
## 204      0 0.000   1296     608     323     2.650      1.243      0.660
## 230      0 0.000    242     114      56     3.120      1.470      0.722
## 279      0 0.333   1016     503     254     2.550      1.262      0.638
## 331      0 0.000    486     226     123     2.520      1.172      0.638
## 442      0 1.000    608     267     143     1.320      0.580      0.310
## 479      0 0.000   1314     732     413     2.160      1.203      0.679
```

```
## 529      0 0.500    715     317     180     4.020      1.782      1.012
## 579      0 0.000    286     152      68     3.980      2.115      0.946
## 629      0 0.500    253     114      52     3.680      1.658      0.756
## 654      0 0.000    191      95      47     2.660      1.323      0.655
## 679      0 0.000    932     469     227     1.780      0.896      0.434
## 705      0 0.500    497     254     147     3.850      1.968      1.139
## 779      0 0.500    116      64      31     2.160      1.192      0.577
## 804      0 0.000   1582     832     437     2.160      1.136      0.597
## 828      0 0.000   1054     547     298     3.510      1.822      0.992
## 904      0 0.000    100      54      31     1.840      0.994      0.570
## 992      0 0.000   1239     566     291     2.120      0.968      0.498
## 1080     0 0.000   1013     538     290     2.280      1.211      0.653
## 1153     0 0.000    770     413     227     2.350      1.260      0.693
##      statepop totfatrte nghtfatrte wkndfatrte vehicmiles unem perc14_24
## 28    2889868     25.05  11.660000   6.120000   20.39437  9.9      18.0
## 54    2305755     24.16  10.060000   5.290000   16.67665 10.1      17.0
## 78   24820028     18.59   9.750000   4.760000  170.29520  9.9      18.0
## 104   3133629     20.62  11.420000   6.480000   24.84615  6.6      17.4
## 131   3201131     14.00   7.620000   3.810000   22.40000  4.9      16.9
## 154    605415     18.17  10.900001   6.610000    4.88889  8.1      18.8
## 204   5728264     22.62  10.610000   5.640000   48.90566  7.5      18.1
## 230    990837     24.42  11.510000   5.650000    7.75641  7.2      16.3
## 279   5450403     18.64   9.230000   4.660000   39.84314 11.1      17.8
## 331   2427417     20.02   9.309999   5.070000   19.28571  5.0      16.2
## 442   5111986     11.89   5.220000   2.800000   46.06061  4.9      12.3
## 479   9047766     14.52   8.090000   4.560000   60.83333 14.2      18.0
## 529   2567737     27.85  12.350000   7.010000   17.78607 12.6      18.8
## 579    814027     35.13  18.670000   8.349999    7.18593  8.8      16.8
## 629    901974     28.05  12.640000   5.770000    6.87500  9.8      16.4
## 654    958123     19.93   9.920000   4.910000    7.18045  5.4      17.7
## 679   7467809     12.48   6.280000   3.040000   52.35955  7.8      16.9
## 705   1416664     35.08  17.930000  10.380000   12.90909  7.5      17.7
## 779    676685     17.14   9.460000   4.580000    5.37037  5.6      18.5
## 804  10737653     14.73   7.750000   4.070000   73.24075 12.2      17.3
## 828   3206119     32.87  17.059999   9.290000   30.02849  5.7      17.7
## 904    956374     10.46   5.650000   3.240000    5.43478  8.3      18.5
## 992   5416643     22.87  10.450000   5.370000   58.44340  5.2      13.9
## 1080  5643868     17.95   9.530000   5.140000   44.42982  5.0      17.9
## 1153  4728868     16.28   8.730000   4.800000   32.76596 10.7      18.7
##      sl70plus sbprim sbsecon d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91
## 28          0      0       0   0   0   1   0   0   0   0   0   0   0   0   0
## 54          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 78          0      0       0   0   0   1   0   0   0   0   0   0   0   0   0
## 104         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 131         0      0       0   0   0   0   0   0   1   0   0   0   0   0   0
## 154         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 204         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 230         0      0       0   0   0   0   0   1   0   0   0   0   0   0   0
## 279         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 331         0      0       0   0   0   0   0   0   1   0   0   0   0   0   0
## 442         0      1       0   0   0   0   0   0   0   0   0   0   0   0   0
## 479         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 529         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 579         0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
```

```
## 629          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 654          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 679          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 705          0      0       0   0   0   0   0   1   0   0   0   0   0   0   0
## 779          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 804          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 828          0      0       0   0   0   1   0   0   0   0   0   0   0   0   0
## 904          0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 992          0      0       1   0   0   0   0   0   0   0   0   0   0   0   0
## 1080         0      0       0   0   0   0   0   1   0   0   0   0   0   0   0
## 1153         0      0       0   0   0   1   0   0   0   0   0   0   0   0   0
##      d92 d93 d94 d95 d96 d97 d98 d99 d00 d01 d02 d03 d04 vehicmilespc sum_bac
## 28     0   0   0   0   0   0   0   0   0   0   0   0   0     7057.197   0.417
## 54     0   0   0   0   0   0   0   0   0   0   0   0   0     7232.621   0.750
## 78     0   0   0   0   0   0   0   0   0   0   0   0   0     6861.201   0.833
## 104    0   0   0   0   0   0   0   0   0   0   0   0   0     7928.874   0.500
## 131    0   0   0   0   0   0   0   0   0   0   0   0   0     6997.527   0.250
## 154    0   0   0   0   0   0   0   0   0   0   0   0   0     8075.271   0.833
## 204    0   0   0   0   0   0   0   0   0   0   0   0   0     8537.605   0.333
## 230    0   0   0   0   0   0   0   0   0   0   0   0   0     7828.139   0.833
## 279    0   0   0   0   0   0   0   0   0   0   0   0   0     7310.127   0.333
## 331    0   0   0   0   0   0   0   0   0   0   0   0   0     7944.952   0.500
## 442    0   0   0   0   1   0   0   0   0   0   0   0   0     9010.315   0.667
## 479    0   0   0   0   0   0   0   0   0   0   0   0   0     6723.575   0.750
## 529    0   0   0   0   0   0   0   0   0   0   0   0   0     6926.749   0.500
## 579    0   0   0   0   0   0   0   0   0   0   0   0   0     8827.631   0.250
## 629    0   0   0   0   0   0   0   0   0   0   0   0   0     7622.171   0.500
## 654    0   0   0   0   0   0   0   0   0   0   0   0   0     7494.288   0.333
## 679    0   0   0   0   0   0   0   0   0   0   0   0   0     7011.367   0.750
## 705    0   0   0   0   0   0   0   0   0   0   0   0   0     9112.315   0.500
## 779    0   0   0   0   0   0   0   0   0   0   0   0   0     7936.292   0.500
## 804    0   0   0   0   0   0   0   0   0   0   0   0   0     6820.927   0.750
## 828    0   0   0   0   0   0   0   0   0   0   0   0   0     9365.994   0.500
## 904    0   0   0   0   0   0   0   0   0   0   0   0   0     5682.693   0.500
## 992    0   0   0   0   1   0   0   0   0   0   0   0   0    10789.598   0.667
## 1080   0   0   0   0   0   0   0   0   0   0   0   0   0     7872.229   0.500
## 1153   0   0   0   0   0   0   0   0   0   0   0   0   0     6928.923   0.667
```

2 (30 points, total) Build and Describe the Data

1. (5 points) Load the data and produce useful features. Specifically:
   - Produce a new variable, called `speed_limit` that re-encodes the data that is in `sl55`, `sl65`, `sl70`, `sl75`, and `slnone`;
   - Produce a new variable, called `year_of_observation` that re-encodes the data that is in `d80`, `d81`, . . . , `d04`.
   - Produce a new variable for each of the other variables that are one-hot encoded (i.e. `bac*` variable series).
   - Rename these variables to sensible names that are legible to a reader of your analysis. For example, the dependent variable as provided is called, `totfatrte`. Pick something more sensible, like, `total_fatalities_rate`. There are few enough of these variables to change, that you should change them for all the variables in the data. (You will thank yourself later.)
2. (5 points) Provide a description of the basic structure of the dataset. What is this data? How, where, and when is it collected? Is the data generated through a survey or some other method? Is the data

that is presented a sample from the population, or is it a *census* that represents the entire population? Minimally, this should include:

- How is the our dependent variable of interest `total_fatalities_rate` defined?

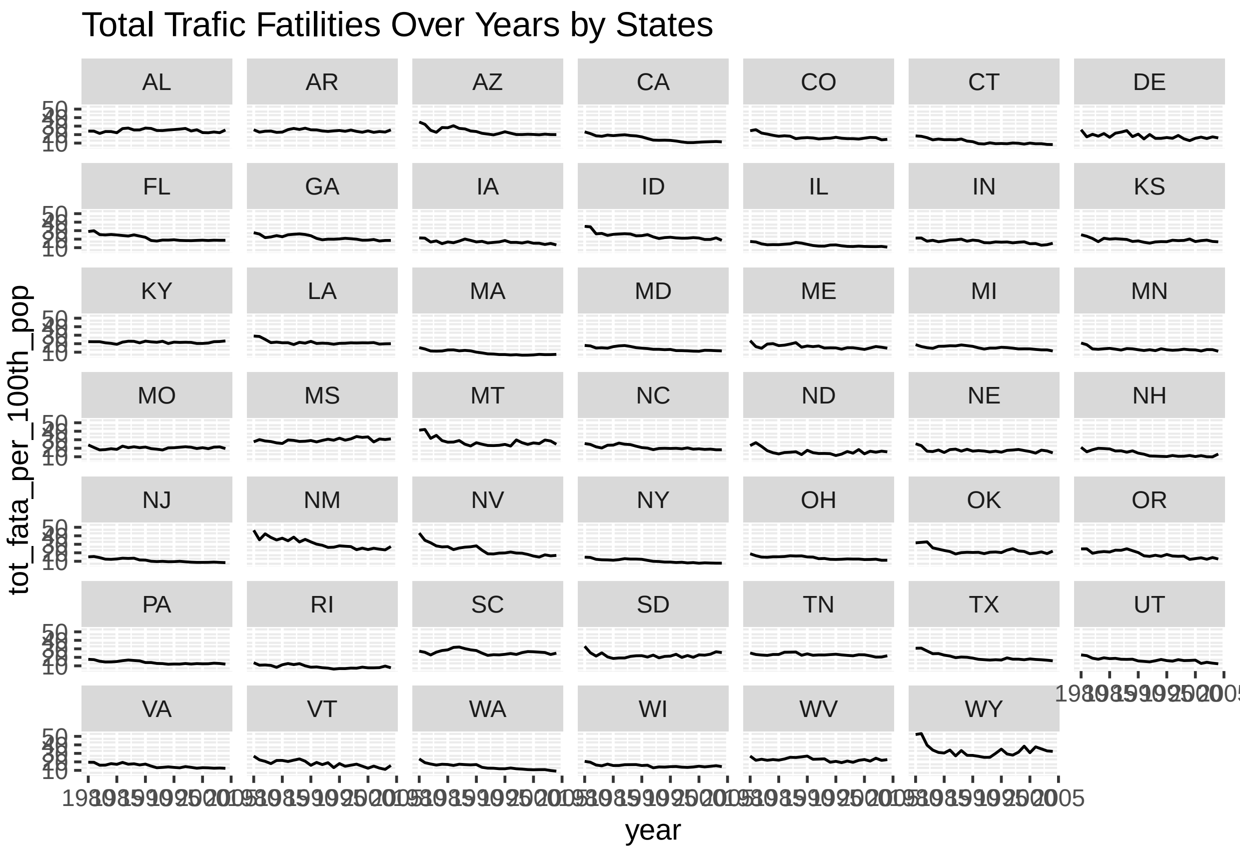
The dataset contains the total car fatalities rate from 1980 to 2004 for the 48 states in the US. The data is structured under a panel format. It is believed that the data was collected by National Highway Traffice Safety Administration (NHTSA). According to NHTSA, the traffic fatilities data is obtained from States' existing documents: police accident reports, state vehicle registration files and state driver licensing files. When dividing the total traffic fatalities by the state population, we obtained the result similar to total traffic fatality rate (fatalities per 100,000 of population), so it's likely the data represent the entire population within the 48 states.

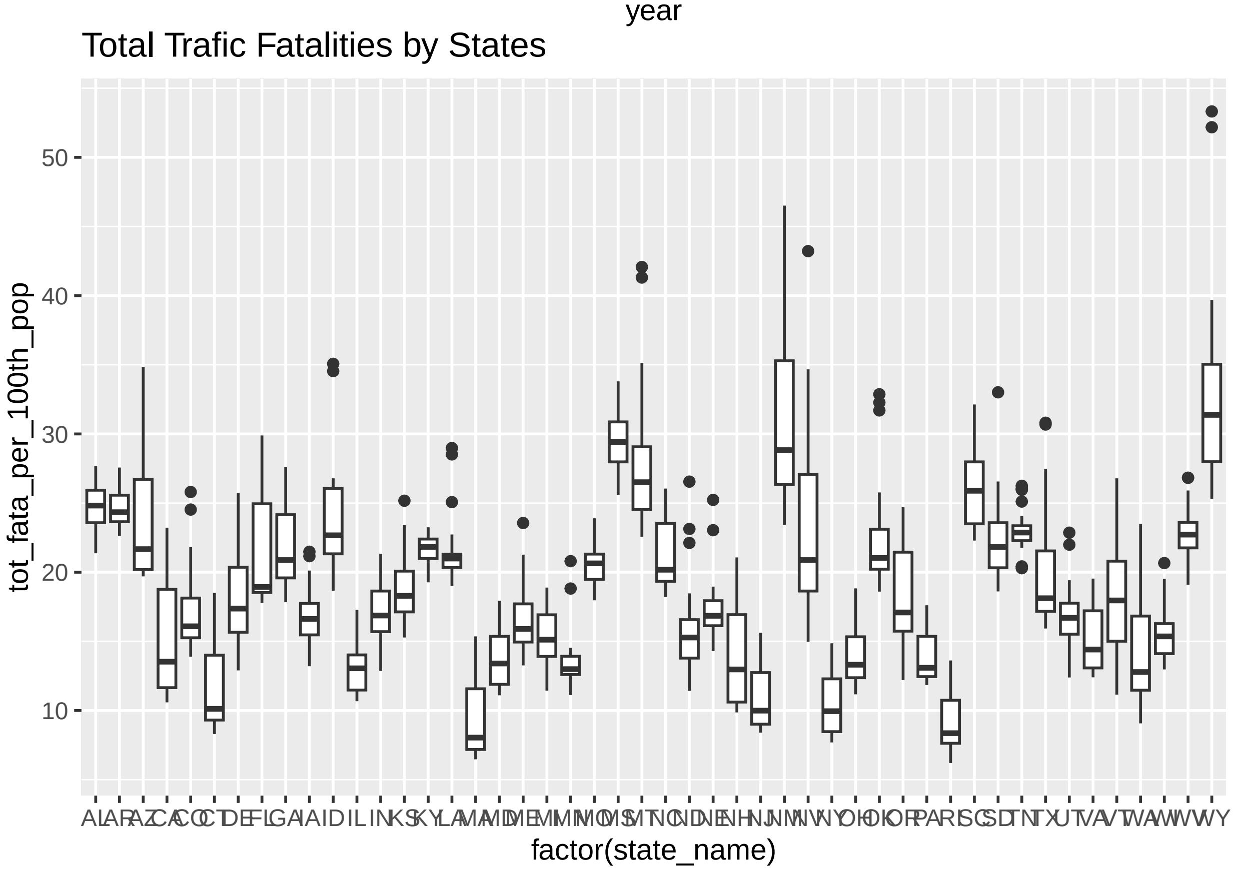
As mentioned above, the `total_fatilities_rate` is the total traffic fatalities per 100,000 of population. A `total_facilities_rate` of 24 means for every 100,000 residents within the state, there are 24 fatalities due to traffic.

Besides fatilities data, the dataset also consists of traffic law data (speed limit, blood alcohol limit, per se law, seatbelt law, zero tolerance law), and other data that may correlate with traffic fatilities (minimum drinking age, vehicle travel miles, percentage of population with age from 14 to 24 and unemployment rate).

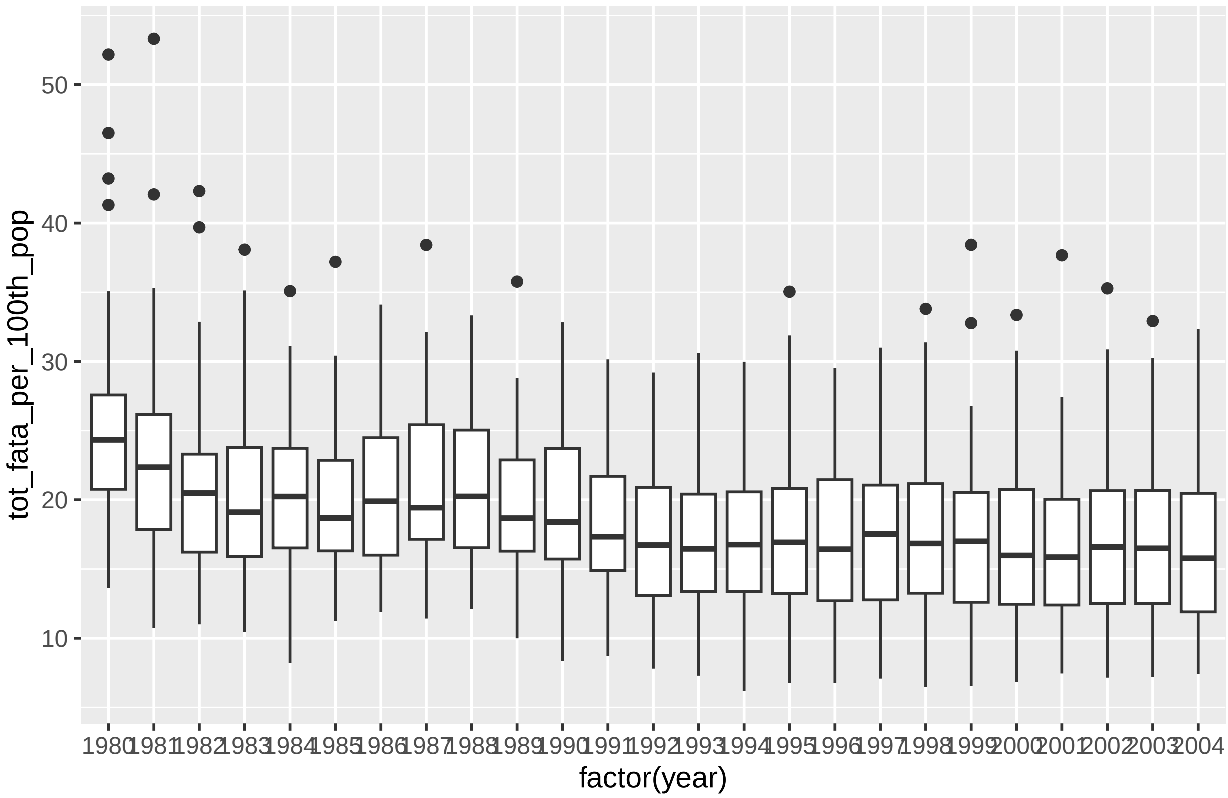
3. (20 points) Conduct a very thorough EDA, which should include both graphical and tabular techniques, on the dataset, including both the dependent variable `total_fatalities_rate` and the potential explanatory variables. Minimally, this should include:
   - How is the our dependent variable of interest `total_fatalities_rate` defined?
   - What is the average of `total_fatalities_rate` in each of the years in the time period covered in this dataset?

As with every EDA this semester, the goal of this EDA is not to document your own process of discovery – save that for an exploration notebook – but instead it is to bring a reader that is new to the data to a full understanding of the important features of your data as quickly as possible. In order to do this, your EDA should include a detailed, orderly narrative description of what you want your reader to know. Do not include any output – tables, plots, or statistics – that you do not intend to write about.

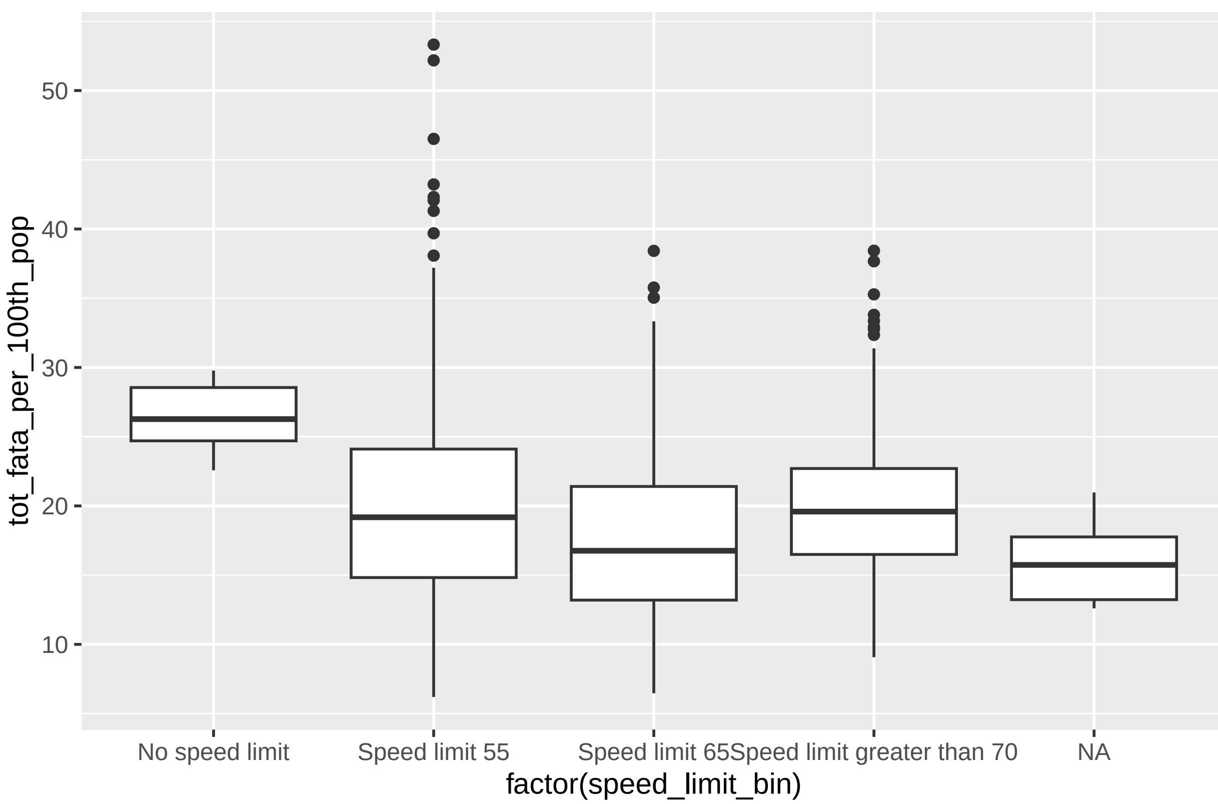

Total Trafic Fatilities Over Years by States
AL
AR
AZ
CA
CO
CT
DE
FL
GA
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY
tot_fata_per_100th_pop
year

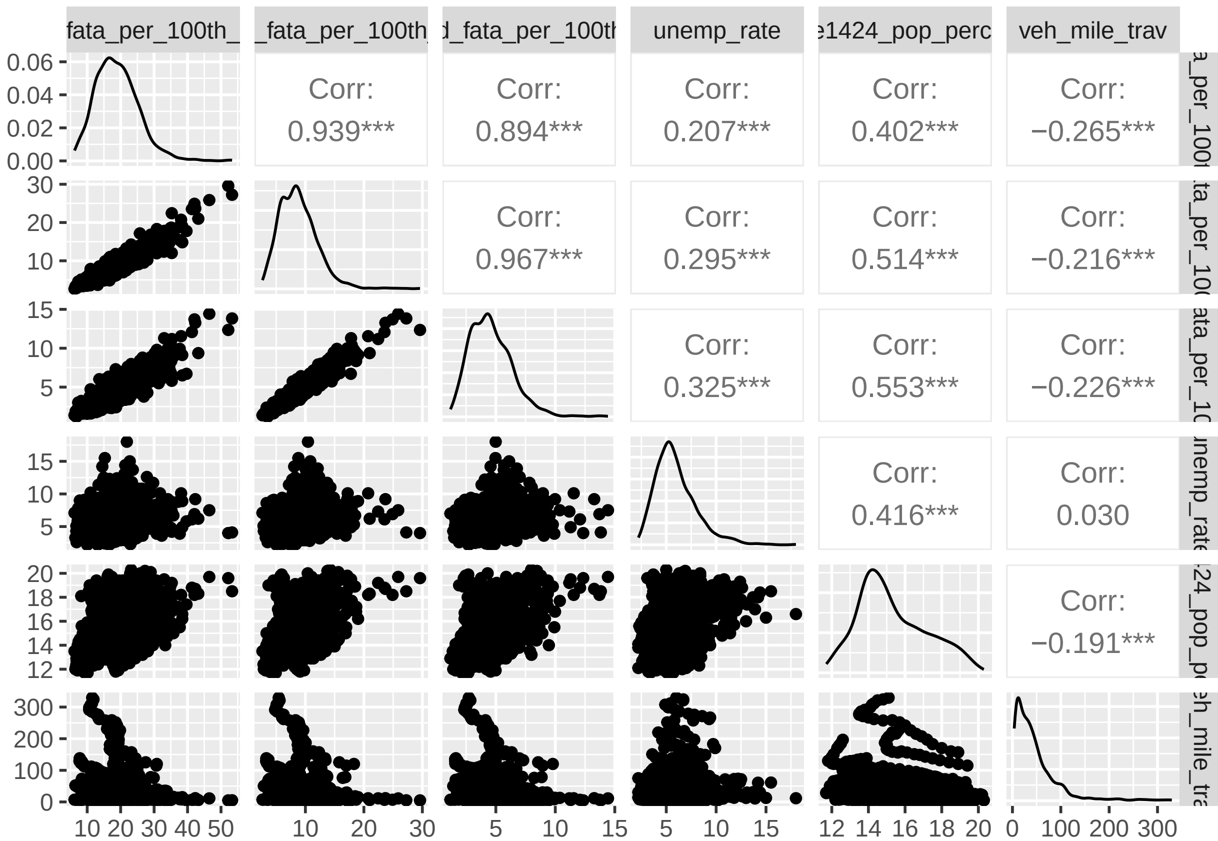

Total Trafic Fatalities by States
50
40
30
20
10
tot_fata_per_100th_pop
factor(state_name)

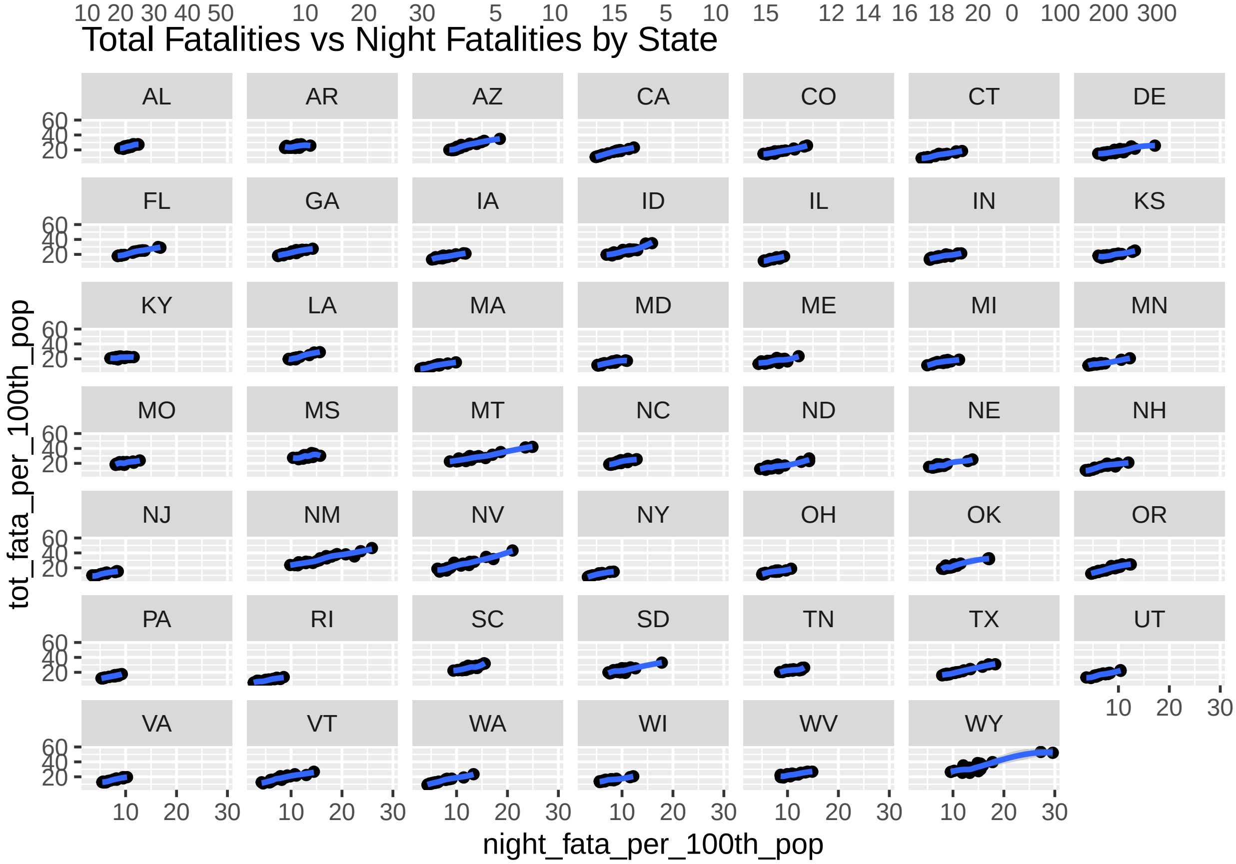
Total Trafic Fatalities by Year



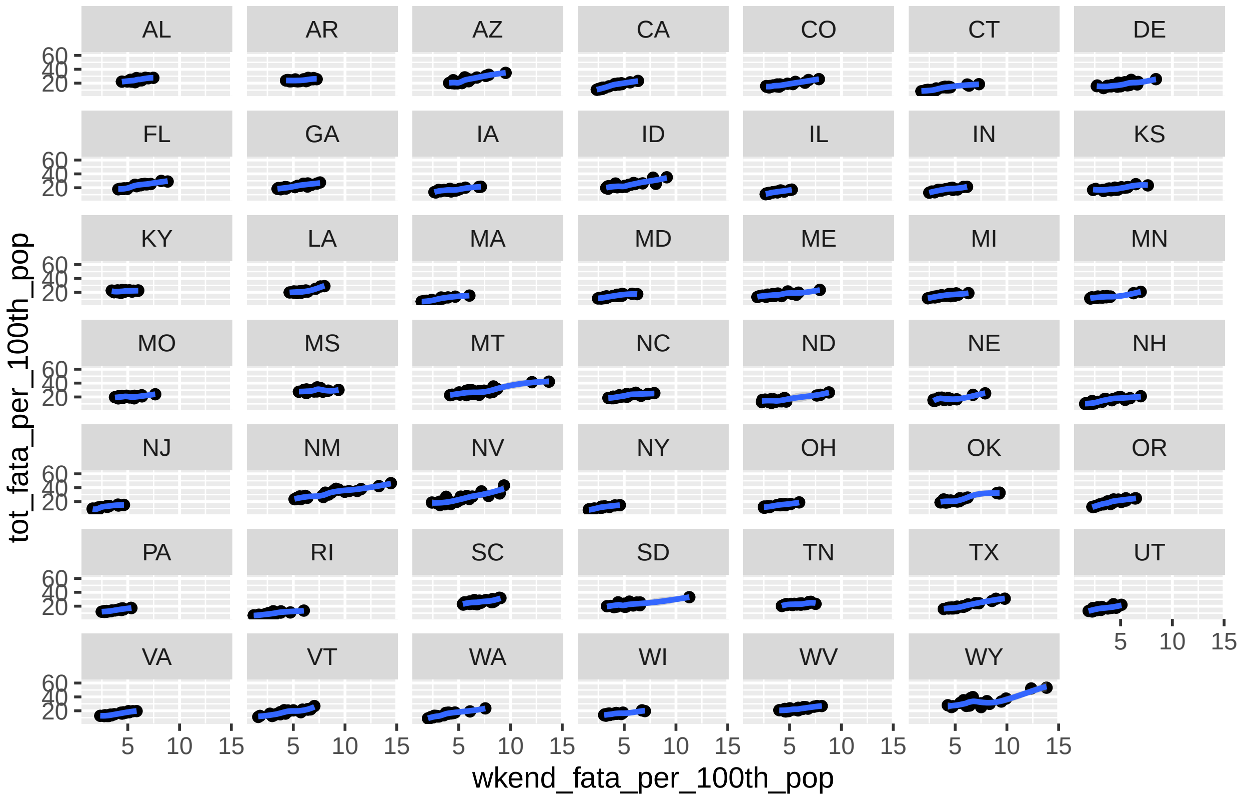
Total Trafic Fatalitites by Speed Limit Category



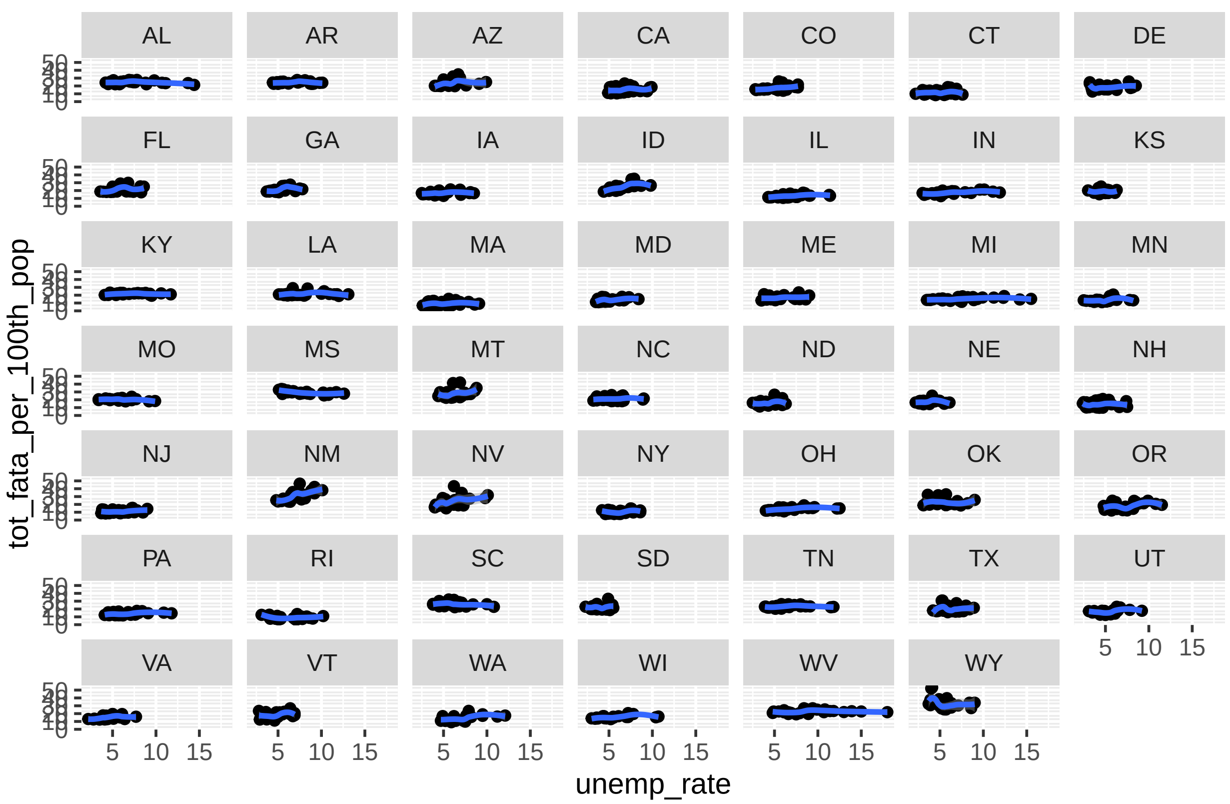

fata_per_100th_
_fata_per_100th
l_fata_per_100tl
unemp_rate
e1424_pop_perc
veh_mile_trav
Corr: 0.939***
Corr: 0.894***
Corr: 0.207***
Corr: 0.402***
Corr: −0.265***
Corr: 0.967***
Corr: 0.295***
Corr: 0.514***
Corr: −0.216***
Corr: 0.325***
Corr: 0.553***
Corr: −0.226***
Corr: 0.416***
Corr: 0.030
Corr: −0.191***

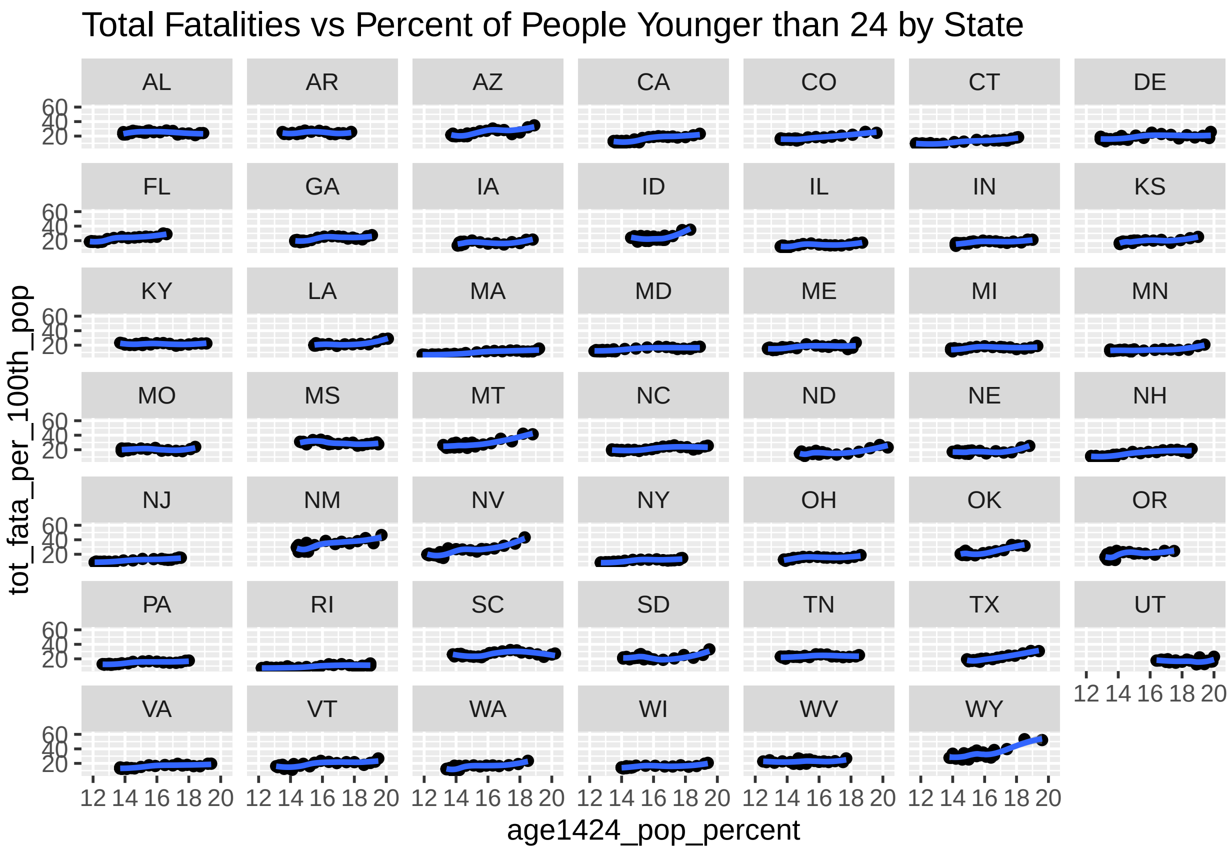

Total Fatalities vs Night Fatalities by State
AL AR AZ CA CO CT DE
FL GA IA ID IL IN KS
KY LA MA MD ME MI MN
MO MS MT NC ND NE NH
NJ NM NV NY OH OK OR
PA RI SC SD TN TX UT
VA VT WA WI WV WY
tot_fata_per_100th_pop
night_fata_per_100th_pop

Total Fatalities vs Weekend Fatalities by State

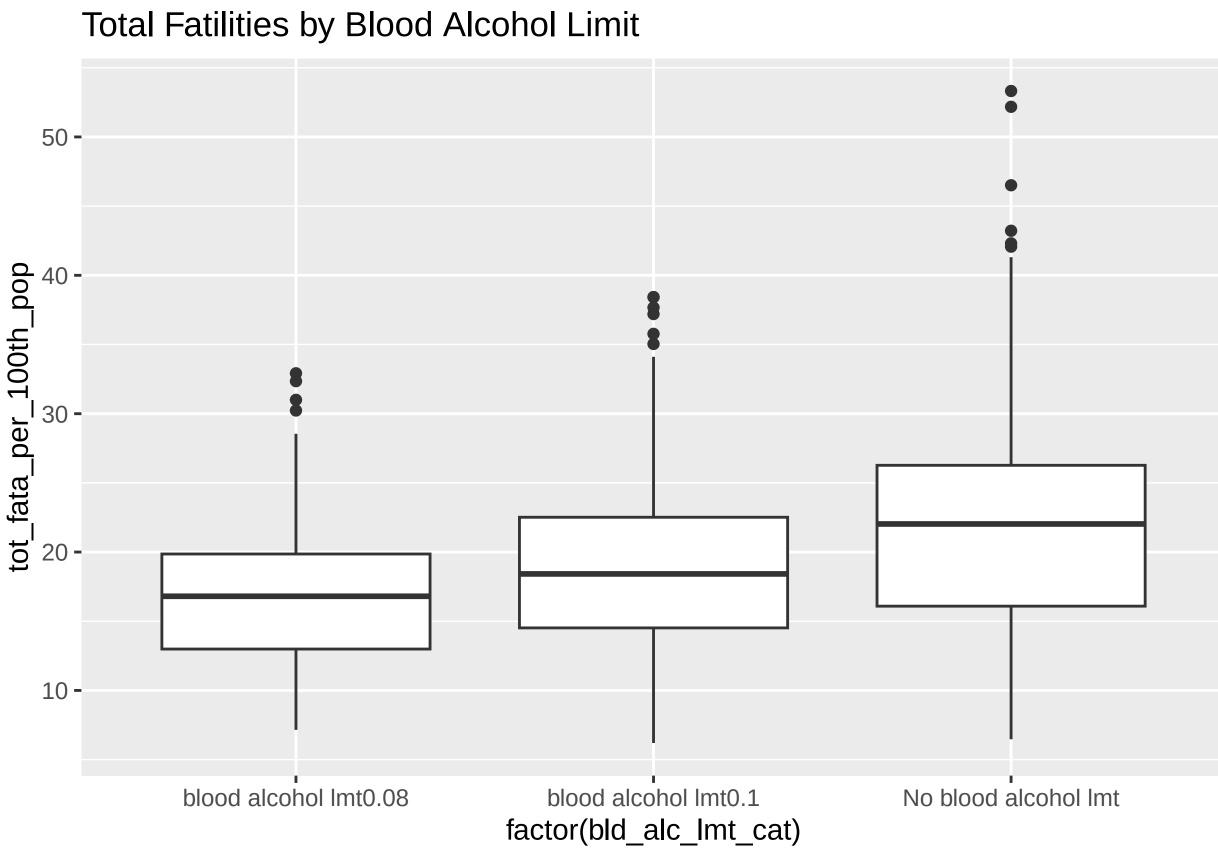


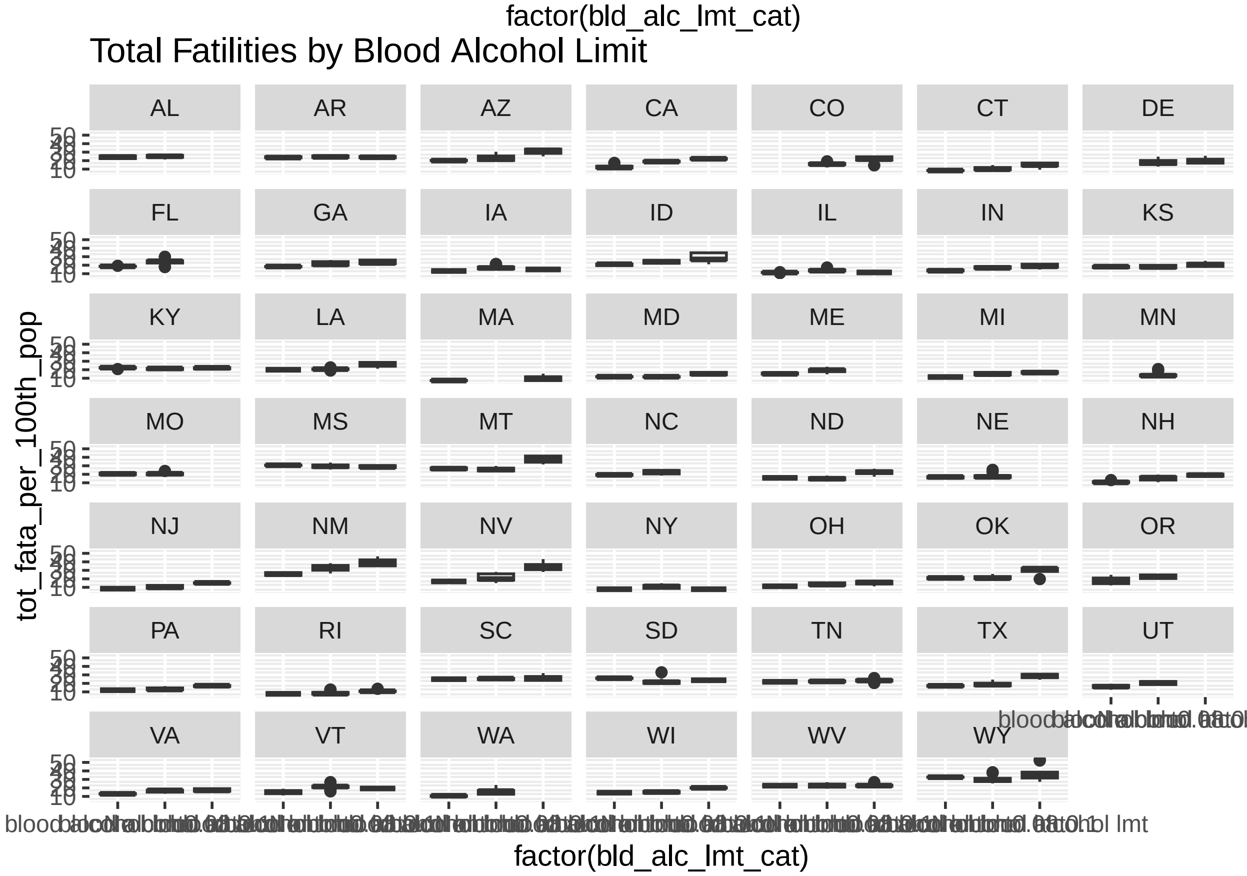
Total Fatalities vs Unemployment by State

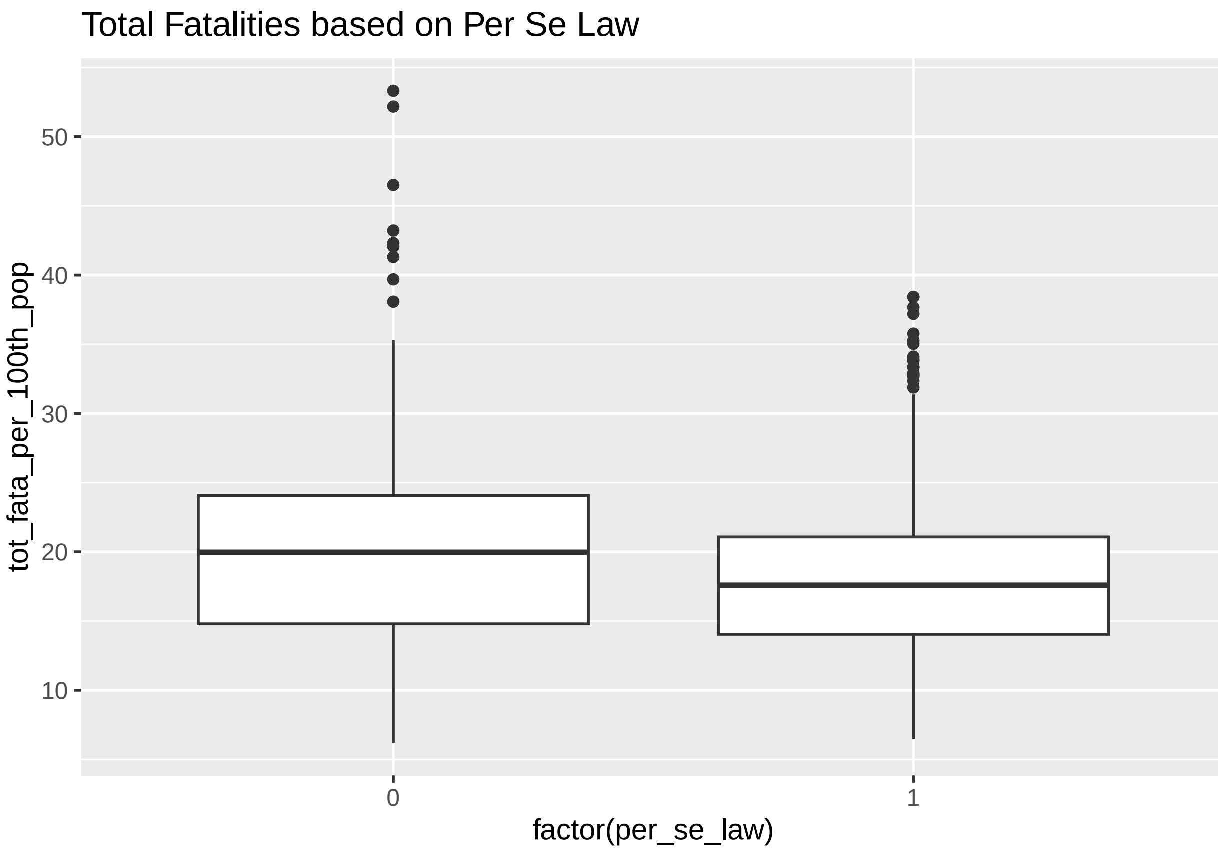


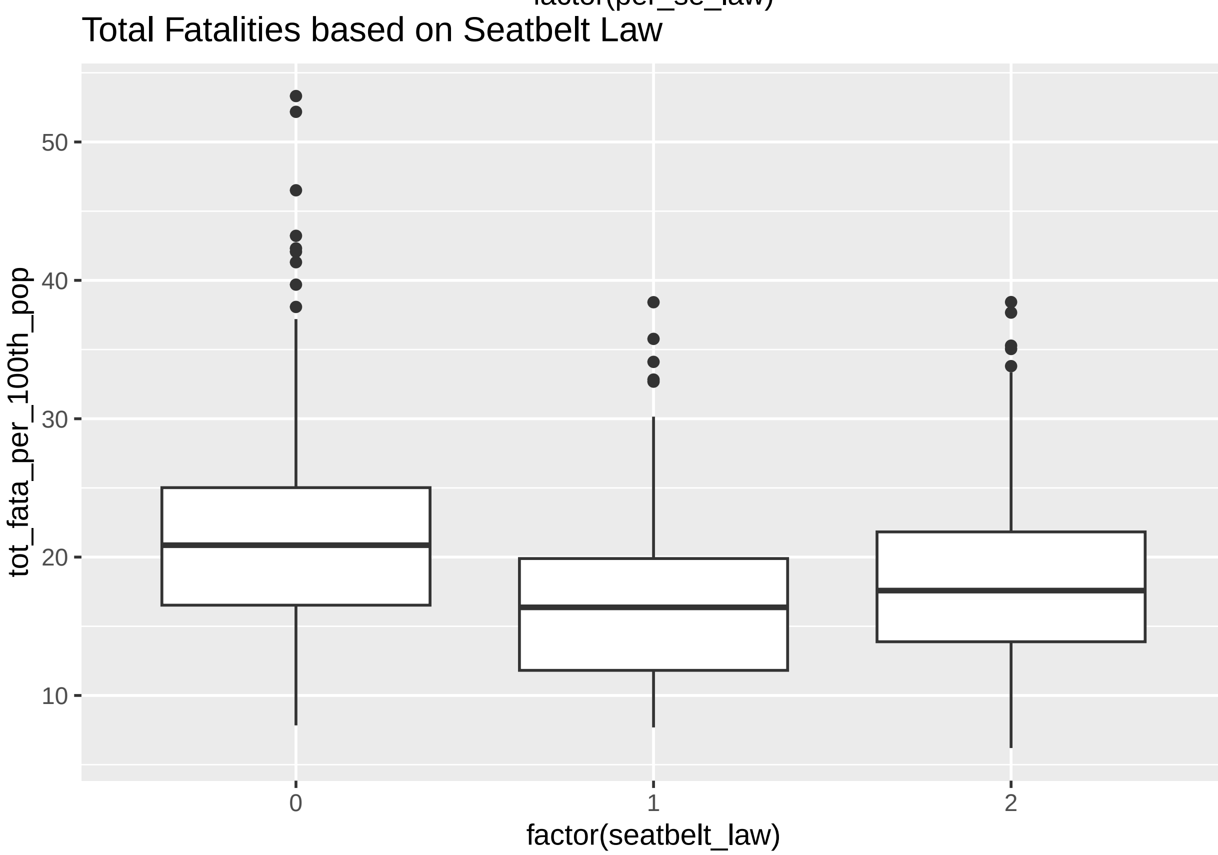

Total Fatalities vs Percent of People Younger than 24 by State
AL AR AZ CA CO CT DE
FL GA IA ID IL IN KS
KY LA MA MD ME MI MN
MO MS MT NC ND NE NH
NJ NM NV NY OH OK OR
PA RI SC SD TN TX UT
VA VT WA WI WV WY
60 40 20
12 14 16 18 20
tot_fata_per_100th_pop
age1424_pop_percent

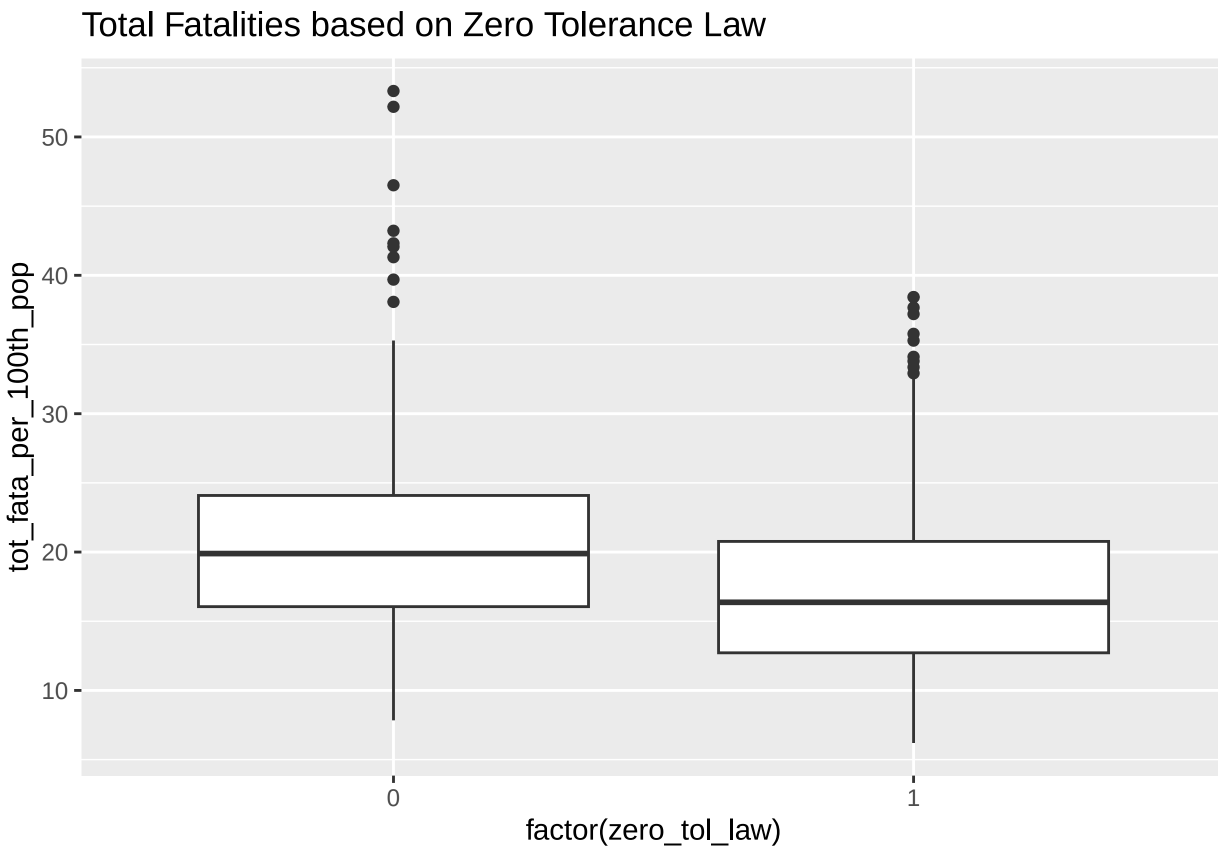

Total Fatalities vs Miles Traveling
AL AR AZ CA CO CT DE
FL GA IA ID IL IN KS
KY LA MA MD ME MI MN
MO MS MT NC ND NE NH
NJ NM NV NY OH OK OR
PA RI SC SD TN TX UT
VA VT WA WI WV WY
50 40 30 20 10
0 100 200 300
tot_fata_per_100th_pop
veh_mile_trav


Total Fatilities by Blood Alcohol Limit
tot_fata_per_100th_pop
50
40
30
20
10
blood alcohol lmt0.08
blood alcohol lmt0.1
No blood alcohol lmt
factor(bld_alc_lmt_cat)


Total Fatilities by Blood Alcohol Limit
AL
AR
AZ
CA
CO
CT
DE
FL
GA
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY
tot_fata_per_100th_pop
factor(bld_alc_lmt_cat)


Total Fatalities based on Per Se Law
tot_fata_per_100th_pop
50
40
30
20
10
0
1
factor(per_se_law)


Total Fatalities based on Seatbelt Law
tot_fata_per_100th_pop
50
40
30
20
10
0
1
2
factor(seatbelt_law)



The above plots show the following: - There is a high correlation between total traffic fatalities fatalities during night and weekend. This correlation appears to be similar among all states. - Traffic fatalities are lower when there is law enforced (applicable to all laws included in the dataset, per se law, seat belt law, zero tolerance law, blood alcohol limit). On blood alcohol limit, higher limit seems to lead to higher total fatalities. It is difficult to see the same trend at the state level. - There is a reduction in traffic fatalities from 1980 to 2004, although it could be due to the enforcement of trafic law - The downward trend of traffic fatalities seems to follow for all states - There do not seem to have high correlation between traffic fatalities with miles traveling, percent of people younger than 24 and unemployment rate for all states. There are exception but it maybe spurious correlation.

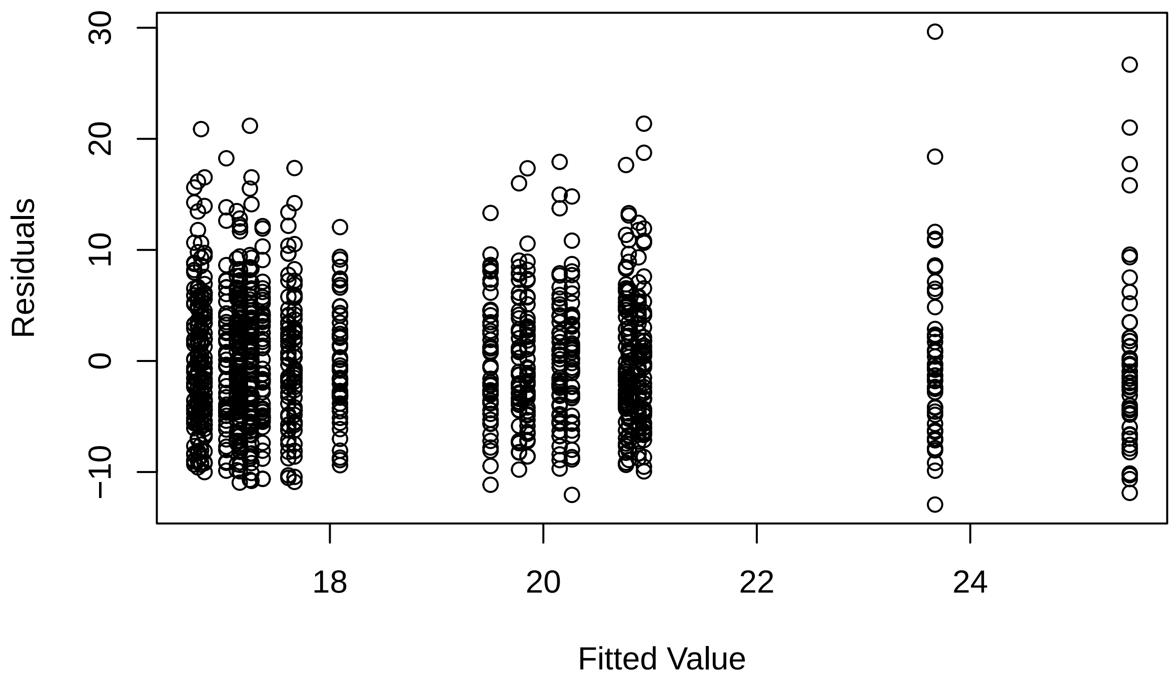
3 (15 points) Preliminary Model

Estimate a linear regression model of *totfatrte* on a set of dummy variables for the years 1981 through 2004 and interpret what you observe. In this section, you should address the following tasks:

- Why is fitting a linear model a sensible starting place?
- What does this model explain, and what do you find in this model?
- Did driving become safer over this period? Please provide a detailed explanation.
- What, if any, are the limitation of this model. In answering this, please consider **at least**:
  - Are the parameter estimates reliable, unbiased estimates of the truth? Or, are they biased due to the way that the data is structured?
  - Are the uncertainty estimate reliable, unbiased estimates of sampling based variability? Or, are they biased due to the way that the data is structured?

```
## Pooling Model
##
## Call:
```

```
## plm(formula = tot_fata_per_100th_pop ~ year_of_observation, data = data,
##     effect = "individual", model = "pooling", index = c("state_name",
##         "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -12.93021  -4.34682  -0.73052   3.74875  29.64979
##
## Coefficients:
##                         Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)             25.49458    0.86712 29.4015 < 2.2e-16 ***
## year_of_observation1981 -1.82438    1.22629 -1.4877 0.1370936
## year_of_observation1982 -4.55208    1.22629 -3.7121 0.0002152 ***
## year_of_observation1983 -5.34167    1.22629 -4.3560 1.440e-05 ***
## year_of_observation1984 -5.22708    1.22629 -4.2625 2.183e-05 ***
## year_of_observation1985 -5.64313    1.22629 -4.6018 4.644e-06 ***
## year_of_observation1986 -4.69417    1.22629 -3.8279 0.0001360 ***
## year_of_observation1987 -4.71979    1.22629 -3.8488 0.0001251 ***
## year_of_observation1988 -4.60292    1.22629 -3.7535 0.0001829 ***
## year_of_observation1989 -5.72229    1.22629 -4.6663 3.418e-06 ***
## year_of_observation1990 -5.98938    1.22629 -4.8841 1.182e-06 ***
## year_of_observation1991 -7.39979    1.22629 -6.0343 2.137e-09 ***
## year_of_observation1992 -8.33667    1.22629 -6.7983 1.681e-11 ***
## year_of_observation1993 -8.36688    1.22629 -6.8229 1.425e-11 ***
## year_of_observation1994 -8.33938    1.22629 -6.8005 1.656e-11 ***
## year_of_observation1995 -7.82604    1.22629 -6.3819 2.512e-10 ***
## year_of_observation1996 -8.12521    1.22629 -6.6258 5.246e-11 ***
## year_of_observation1997 -7.88396    1.22629 -6.4291 1.863e-10 ***
## year_of_observation1998 -8.22917    1.22629 -6.7106 3.007e-11 ***
## year_of_observation1999 -8.24417    1.22629 -6.7228 2.774e-11 ***
## year_of_observation2000 -8.66896    1.22629 -7.0692 2.666e-12 ***
## year_of_observation2001 -8.70188    1.22629 -7.0961 2.214e-12 ***
## year_of_observation2002 -8.46500    1.22629 -6.9029 8.316e-12 ***
## year_of_observation2003 -8.73104    1.22629 -7.1199 1.877e-12 ***
## year_of_observation2004 -8.76563    1.22629 -7.1481 1.542e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    48612
## Residual Sum of Squares: 42407
## R-Squared:      0.12765
## Adj. R-Squared: 0.10983
## F-statistic: 7.16387 on 24 and 1175 DF, p-value: < 2.22e-16
```



The EDA shows a series of correlation between the response variable (total traffic fatality rate) and explanatory variables. As a result, a linear model is an appropriate approach to forecast the total traffic fatality rate.

The model uses `1980` as the base year. Based on the output of the model, beside year `1981`, all other subsequent years have statistically significant impact on total traffic fatalities. The coefficient estimates of all years are negative, indicating a decrease in total fatalities from the `1980`.

It's likely that there are omitted variables within the model, the parameter estimates are not reliable and likely overestimate/underestimate the impact of time on fatalities rate. One omitted variable that we suspect is the enforcement of traffic law which happened in the later years (1980 as the reference). The enforced traffic law would likely reduce the fatalities rate (as shown in the EDA section), which would imply that the impact from years on traffic fatalities is less than what the above model shows.

Because the uncertainty of the model likely includes the effect of omitted variables, the uncertainty of the model is not reliable and may be biased.

4 (15 points) Expanded Model

Expand the **Preliminary Model** by adding variables related to the following concepts:

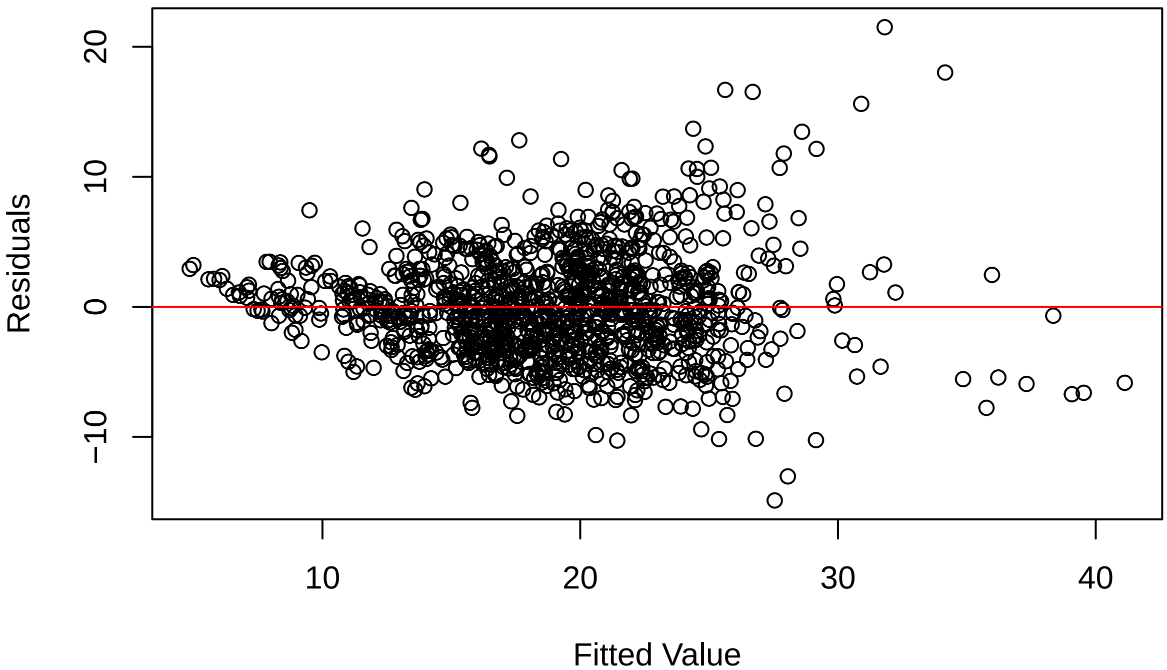
- Blood alcohol levels
- Per se laws
- Primary seat belt laws (Note that if a law was enacted sometime within a year the fraction of the year is recorded in place of the zero-one indicator.)
- Secondary seat belt laws
- Speed limits faster than 70
- Graduated drivers licenses
- Percent of the population between 14 and 24 years old
- Unemployment rate
- Vehicle miles driven per capita.

If it is appropriate, include transformations of these variables. Please carefully explain carefully your rationale, which should be based on your EDA, behind any transformation you made. If no transformation is made, explain why transformation is not needed.

- How are the blood alcohol variables defined? Interpret the coefficients that you estimate for this concept.
- Do *per se laws* have a negative effect on the fatality rate?
- Does having a primary seat belt law?

```
## Pooling Model
##
## Call:
## plm(formula = tot_fata_per_100th_pop ~ year_of_observation +
##     bld_alc_lmt_cat + factor(per_se_law) + factor(round(prim_seatbelt_law)) +
##     factor(round(second_seatbelt_law)) + factor(round(speed_lim_grter70)) +
##     factor(round(grad_driver_license_law)) + age1424_pop_percent +
##     unemp_rate + veh_mile_trav_percap, data = data, effect = "individual",
##     model = "pooling", index = c("state_name", "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -14.89623  -2.72649  -0.30325   2.33231  21.50641
##
## Coefficients:
##                                           Estimate  Std. Error  t-value
## (Intercept)                             -5.0206e+00  2.5021e+00  -2.0065
## year_of_observation1981                 -2.1840e+00  8.2903e-01  -2.6344
## year_of_observation1982                 -6.6572e+00  8.5472e-01  -7.7887
## year_of_observation1983                 -7.5890e+00  8.6714e-01  -8.7519
## year_of_observation1984                 -5.9745e+00  8.7303e-01  -6.8434
## year_of_observation1985                 -6.6031e+00  8.9149e-01  -7.4069
## year_of_observation1986                 -5.9467e+00  9.2901e-01  -6.4011
## year_of_observation1987                 -6.4588e+00  9.6555e-01  -6.6892
## year_of_observation1988                 -6.6905e+00  1.0127e+00  -6.6066
## year_of_observation1989                 -8.1588e+00  1.0518e+00  -7.7570
## year_of_observation1990                 -9.0597e+00  1.0759e+00  -8.4206
## year_of_observation1991                 -1.1206e+01  1.0992e+00 -10.1943
## year_of_observation1992                 -1.2996e+01  1.1212e+00 -11.5909
## year_of_observation1993                 -1.2882e+01  1.1342e+00 -11.3579
## year_of_observation1994                 -1.2530e+01  1.1543e+00 -10.8546
## year_of_observation1995                 -1.2033e+01  1.1825e+00 -10.1760
## year_of_observation1996                 -1.4025e+01  1.2240e+00 -11.4590
## year_of_observation1997                 -1.4304e+01  1.2420e+00 -11.5171
## year_of_observation1998                 -1.5120e+01  1.2622e+00 -11.9783
## year_of_observation1999                 -1.5185e+01  1.2760e+00 -11.9001
## year_of_observation2000                 -1.5544e+01  1.2958e+00 -11.9955
## year_of_observation2001                 -1.6449e+01  1.3159e+00 -12.5002
## year_of_observation2002                 -1.7028e+01  1.3305e+00 -12.7979
## year_of_observation2003                 -1.7418e+01  1.3364e+00 -13.0334
## year_of_observation2004                 -1.6979e+01  1.3694e+00 -12.3989
## bld_alc_lmt_catblood alcohol lmt0.1      9.5648e-01  3.7087e-01   2.5790
## bld_alc_lmt_catNo blood alcohol lmt      2.1944e+00  4.8907e-01   4.4868
## factor(per_se_law)1                     -6.4989e-01  2.9432e-01  -2.2081
## factor(round(prim_seatbelt_law))1       -9.4205e-02  4.9095e-01  -0.1919
## factor(round(second_seatbelt_law))1      6.4304e-02  4.2990e-01   0.1496
## factor(round(speed_lim_grter70))1        3.2389e+00  4.3515e-01   7.4431
## factor(round(grad_driver_license_law))1 -3.4762e-01  5.1007e-01  -0.6815
```

```
## age1424_pop_percent                       1.4010e-01  1.2292e-01   1.1398
## unemp_rate                                7.6749e-01  7.7963e-02   9.8443
## veh_mile_trav_percap                      2.9271e-03  9.4849e-05  30.8601
##                                           Pr(>|t|)
## (Intercept)                               0.045032 *
## year_of_observation1981                   0.008539 **
## year_of_observation1982                  1.489e-14 ***
## year_of_observation1983                  < 2.2e-16 ***
## year_of_observation1984                  1.247e-11 ***
## year_of_observation1985                  2.475e-13 ***
## year_of_observation1986                  2.232e-10 ***
## year_of_observation1987                  3.476e-11 ***
## year_of_observation1988                  5.967e-11 ***
## year_of_observation1989                  1.889e-14 ***
## year_of_observation1990                  < 2.2e-16 ***
## year_of_observation1991                  < 2.2e-16 ***
## year_of_observation1992                  < 2.2e-16 ***
## year_of_observation1993                  < 2.2e-16 ***
## year_of_observation1994                  < 2.2e-16 ***
## year_of_observation1995                  < 2.2e-16 ***
## year_of_observation1996                  < 2.2e-16 ***
## year_of_observation1997                  < 2.2e-16 ***
## year_of_observation1998                  < 2.2e-16 ***
## year_of_observation1999                  < 2.2e-16 ***
## year_of_observation2000                  < 2.2e-16 ***
## year_of_observation2001                  < 2.2e-16 ***
## year_of_observation2002                  < 2.2e-16 ***
## year_of_observation2003                  < 2.2e-16 ***
## year_of_observation2004                  < 2.2e-16 ***
## bld_alc_lmt_catblood alcohol lmt0.1       0.010031 *
## bld_alc_lmt_catNo blood alcohol lmt      7.945e-06 ***
## factor(per_se_law)1                       0.027433 *
## factor(round(prim_seatbelt_law))1         0.847868
## factor(round(second_seatbelt_law))1       0.881124
## factor(round(speed_lim_grter70))1        1.905e-13 ***
## factor(round(grad_driver_license_law))1  0.495681
## age1424_pop_percent                       0.254611
## unemp_rate                               < 2.2e-16 ***
## veh_mile_trav_percap                     < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    48612
## Residual Sum of Squares: 19132
## R-Squared:      0.60644
## Adj. R-Squared: 0.59495
## F-statistic: 52.799 on 34 and 1165 DF, p-value: < 2.22e-16
```



The team performed the following transformation on the raw dataset: - Generate a new variable to re-encode the blood-alcohol-limit columns (0.1 and 0.08). The values included in this variable is the sum product of variable `bld_alc_lim10` (named `bac10` in the raw data) with 0.10 and variable `bld_alc_lim08` (named `bac08` in the raw data) with 0.08. Then we convert the variable into categorical values (`Blood alcohol limit of 0.1` for rows with value of 0.1 and `Blood alcohol limit of 0.08` for rows with value of 0.08). For rows with zero values in both `bac10` and `bac08` column, we treated them as `No blood alcohol limit`. - Generate a new variable to reflect miles travel per capita (total miles divided by state population)

From the output of the expanded model, r-squared increased from 0.1098294 to 0.594955, which indicates that the new model provides a significant increase of the explanation for the variance in the response variable. The coefficients for the years dummy variables also are smaller than the ones from the preliminary model. This confirms the suspicion that the impact on traffic fatalities rate from the time variable in the preliminary model was overestimated. Overall, we observed that impacts from `blood alcohol limit`, `per se law`, `speed limit law`, `unemployment rate` and `vehicle miles travel per capita` are statistically different from zero. Interestingly, some of the enforced laws such as `primary and secondary seatbelt law` and `graduate driver license law` did not seem to have a significant impact on the traffic fatalities.

The `per se law` has a negative slope, which means that having the law enforced reduces the total traffic fatalities by -0.6498873. This confirms with the findings of our EDA.

The coefficient estimates of the `blood alcohol limt` indicates that when there is no alcohol limit law, fatality rates are higher than when blood alcohol limit is set at 0.1. When there is no blood alcohol limit or the limit is at 0.1, fatality rates are higher than when blood alcohol limit is set at 0.08.

5 (15 points) State-Level Fixed Effects

Re-estimate the **Expanded Model** using fixed effects at the state level.

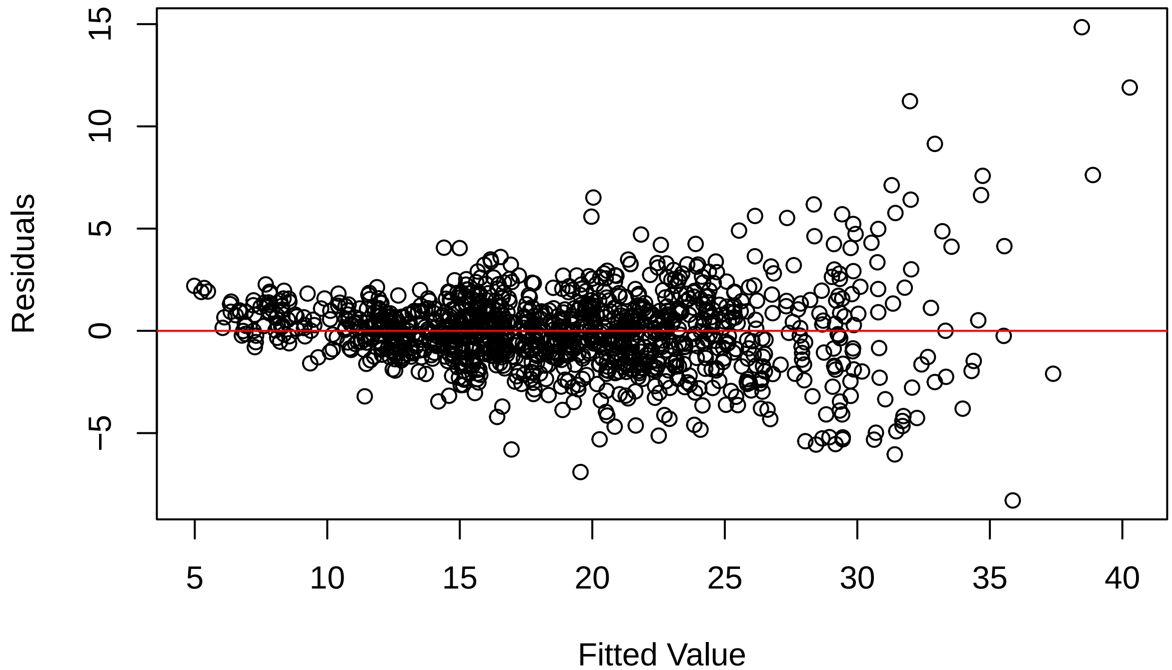
- What do you estimate for coefficients on the blood alcohol variables? How do the coefficients on the blood alcohol variables change, if at all?
- What do you estimate for coefficients on per se laws? How do the coefficients on per se laws change, if at all?
- What do you estimate for coefficients on primary seat-belt laws? How do the coefficients on primary seatbelt laws change, if at all?

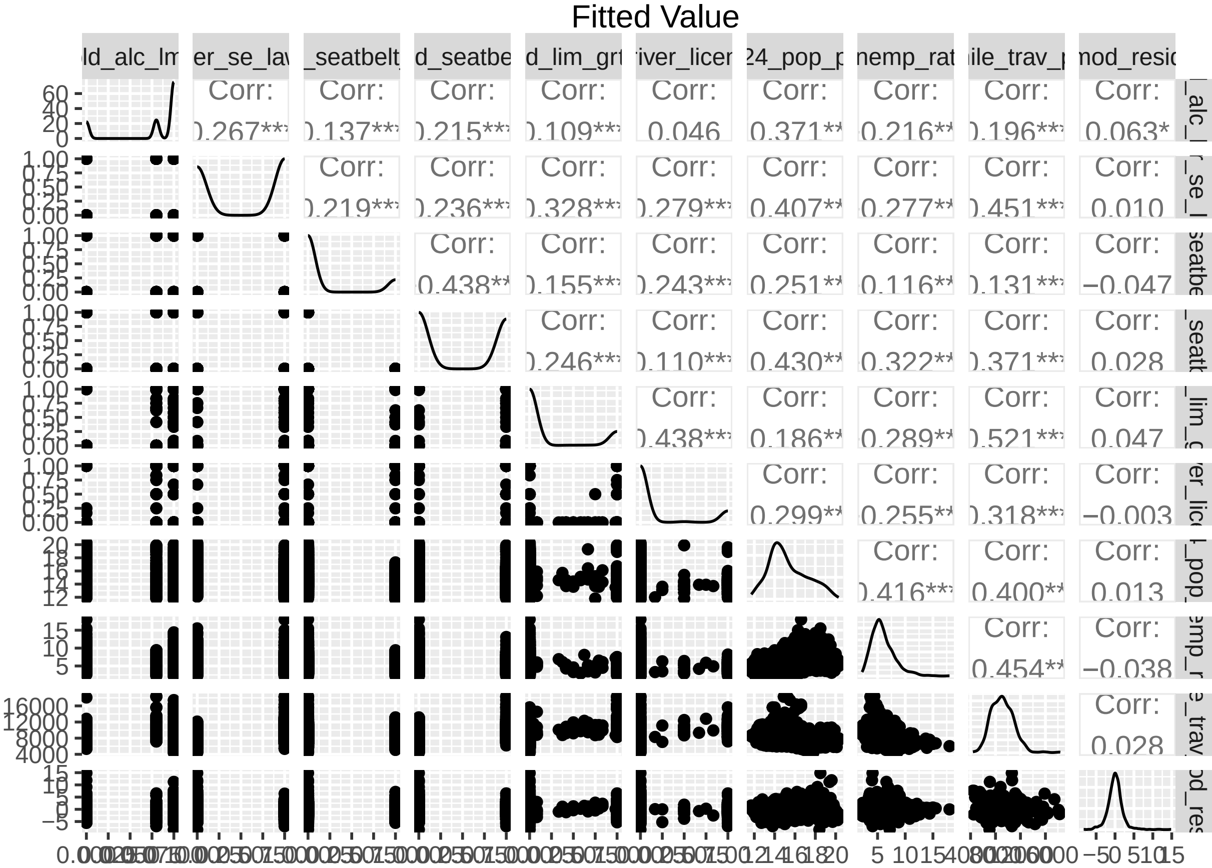
Which set of estimates do you think is more reliable? Why do you think this?

- What assumptions are needed in each of these models?
- Are these assumptions reasonable in the current context?

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = tot_fata_per_100th_pop ~ year_of_observation +
##     bld_alc_lmt_cat + factor(per_se_law) + factor(round(prim_seatbelt_law)) +
##     factor(round(second_seatbelt_law)) + factor(round(speed_lim_grter70)) +
##     factor(round(grad_driver_license_law)) + age1424_pop_percent +
##     unemp_rate + veh_mile_trav_percap, data = data, effect = "individual",
##     model = "within", index = c("state_name", "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -8.2942752 -1.0561094  0.0055576  0.9788363 14.8497790
##
## Coefficients:
##                                          Estimate  Std. Error  t-value
## year_of_observation1981                -1.5124e+00  4.1379e-01  -3.6549
## year_of_observation1982                -3.0540e+00  4.4318e-01  -6.8912
## year_of_observation1983                -3.6638e+00  4.5516e-01  -8.0495
## year_of_observation1984                -4.3998e+00  4.5966e-01  -9.5719
## year_of_observation1985                -4.8603e+00  4.8010e-01 -10.1237
## year_of_observation1986                -3.7692e+00  5.1357e-01  -7.3392
## year_of_observation1987                -4.4123e+00  5.5162e-01  -7.9989
## year_of_observation1988                -4.8877e+00  5.9837e-01  -8.1684
## year_of_observation1989                -6.2395e+00  6.3732e-01  -9.7901
## year_of_observation1990                -6.3564e+00  6.6196e-01  -9.6024
## year_of_observation1991                -7.0442e+00  6.7895e-01 -10.3752
## year_of_observation1992                -7.8905e+00  7.0039e-01 -11.2659
## year_of_observation1993                -8.2366e+00  7.1290e-01 -11.5536
## year_of_observation1994                -8.6823e+00  7.3004e-01 -11.8930
## year_of_observation1995                -8.3889e+00  7.5324e-01 -11.1370
## year_of_observation1996                -8.7648e+00  7.9400e-01 -11.0388
## year_of_observation1997                -8.9164e+00  8.1140e-01 -10.9889
## year_of_observation1998                -9.5333e+00  8.2867e-01 -11.5044
## year_of_observation1999                -9.6940e+00  8.3614e-01 -11.5938
## year_of_observation2000                -1.0223e+01  8.4713e-01 -12.0683
## year_of_observation2001                -9.9608e+00  8.5745e-01 -11.6168
## year_of_observation2002                -9.2546e+00  8.6613e-01 -10.6849
## year_of_observation2003                -9.3270e+00  8.6980e-01 -10.7232
## year_of_observation2004                -9.6676e+00  8.9310e-01 -10.8248
## bld_alc_lmt_catblood alcohol lmt0.1      3.1068e-01  2.4330e-01   1.2769
## bld_alc_lmt_catNo blood alcohol lmt      1.1805e+00  3.2987e-01   3.5786
## factor(per_se_law)1                    -1.0587e+00  2.2415e-01  -4.7230
## factor(round(prim_seatbelt_law))1      -1.2506e+00  3.4313e-01  -3.6447
## factor(round(second_seatbelt_law))1    -3.5659e-01  2.5230e-01  -1.4133
## factor(round(speed_lim_grter70))1      -3.2440e-02  2.6034e-01  -0.1246
## factor(round(grad_driver_license_law))1 -3.0503e-01  2.8029e-01  -1.0883
```

```
## age1424_pop_percent                      1.9367e-01  9.5068e-02   2.0372
## unemp_rate                              -5.7652e-01  6.0592e-02  -9.5147
## veh_mile_trav_percap                     9.2612e-04  1.1066e-04   8.3691
##                                          Pr(>|t|)
## year_of_observation1981                 0.0002692 ***
## year_of_observation1982                 9.222e-12 ***
## year_of_observation1983                 2.111e-15 ***
## year_of_observation1984                 < 2.2e-16 ***
## year_of_observation1985                 < 2.2e-16 ***
## year_of_observation1986                 4.123e-13 ***
## year_of_observation1987                 3.118e-15 ***
## year_of_observation1988                 8.379e-16 ***
## year_of_observation1989                 < 2.2e-16 ***
## year_of_observation1990                 < 2.2e-16 ***
## year_of_observation1991                 < 2.2e-16 ***
## year_of_observation1992                 < 2.2e-16 ***
## year_of_observation1993                 < 2.2e-16 ***
## year_of_observation1994                 < 2.2e-16 ***
## year_of_observation1995                 < 2.2e-16 ***
## year_of_observation1996                 < 2.2e-16 ***
## year_of_observation1997                 < 2.2e-16 ***
## year_of_observation1998                 < 2.2e-16 ***
## year_of_observation1999                 < 2.2e-16 ***
## year_of_observation2000                 < 2.2e-16 ***
## year_of_observation2001                 < 2.2e-16 ***
## year_of_observation2002                 < 2.2e-16 ***
## year_of_observation2003                 < 2.2e-16 ***
## year_of_observation2004                 < 2.2e-16 ***
## bld_alc_lmt_catblood alcohol lmt0.1     0.2018878
## bld_alc_lmt_catNo blood alcohol lmt     0.0003603 ***
## factor(per_se_law)1                     2.619e-06 ***
## factor(round(prim_seatbelt_law))1       0.0002800 ***
## factor(round(second_seatbelt_law))1     0.1578360
## factor(round(speed_lim_grter70))1       0.9008578
## factor(round(grad_driver_license_law))1 0.2767099
## age1424_pop_percent                     0.0418646 *
## unemp_rate                              < 2.2e-16 ***
## veh_mile_trav_percap                    < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12134
## Residual Sum of Squares: 4547.9
## R-Squared:      0.6252
## Adj. R-Squared: 0.59804
## F-statistic: 54.8501 on 34 and 1118 DF, p-value: < 2.22e-16
```





```
##
##  Durbin-Watson test for serial correlation in panel models
##
## data:  tot_fata_per_100th_pop ~ year_of_observation + bld_alc_lmt_cat +  ...
## DW = 1.0619, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```
##
##  Breusch-Pagan LM test for cross-sectional dependence in panels
##
## data:  tot_fata_per_100th_pop ~ year_of_observation + bld_alc_lmt_cat +     factor(per_se_law) + fac
## chisq = 3396.7, df = 1128, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

```
##
##  F test for individual effects
##
## data:  tot_fata_per_100th_pop ~ year_of_observation + bld_alc_lmt_cat +  ...
## F = 76.279, df1 = 47, df2 = 1118, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

The coefficient estimate for blood alcohol level limit at 0.1 changes from 0.9564807 in the expanded model to 0.3106834 in the expanded model with state level fixed effect. This factor level also does not seem to be statistically significant among state. The coefficient estimate for no blood alcohol limit changes from 2.1943687 in the expanded model to 1.180452 in the fixed effect model.

The coefficient estimate for per se law changes from -0.6498873 in the expanded model to -1.0586512 in the expanded model with state level fixed effect.

The coefficient estimate for primary seat-belt law changes from -0.0942048 in the expanded model to -1.2506108 in the expanded model with state level fixed effect.

Both models seem to have residuals with zero mean and a mild level of heteroskedaticity, however, the t statistics of all three estimates are higher in the state-level fixed effect model, which indicates a narrower confidence interval for these estimates. As a result, we find the estimates of the state-level fixed effect model to be more reliable.

For the estimates to be consisted, the model needs to satisfy the following assumptions: - The model is linear in parameters, the residual vs fitted value shows no pattern between the residuals and the model's fitted value, which indicate linearity. - The observations are independent across individuals but not necessarily across time. This assumption may be difficult to satisfy as individuals in close by states may share similar characteristics. - The regressors are not perfectly collinear and all regressors have non-zero variance and not too many extreme values. Based on the pair plots in the EDA section, it's unlikely that the regressors have perfect multicolinearity, or zero variance or contain too many extreme values. - The error term is uncorrelated with all explanatory variables across all time period. The pair plot above shows that the residuals is not correlated with any explanatory variables within the model. - Error term is homoskedastic and serially uncorrelated across time. The Durbin-Watson test shows a p-value of 0.988, which indicates that there is no serial correlation in the error term and satisfied this assumption. The Breusch-Pagan test indicates that the standard error is heteroskedasitc and robust standard error is a more appropriate method to determine confidence inteval for the coefficient estimates.

The F test for individual effects indicates that we reject the null hypothesis of no fixed effects and thus the state-level fixed effect model provides more reliable coefficient estimates.

6 (10 points) Consider a Random Effects Model

Instead of estimating a fixed effects model, should you have estimated a random effects model?

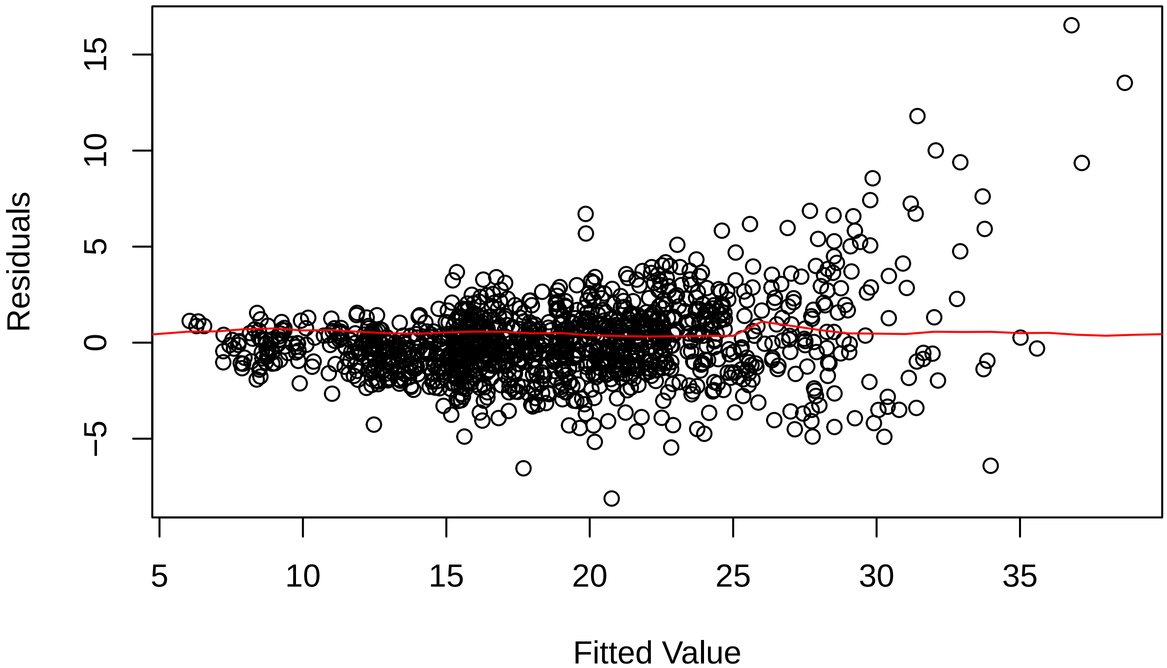
- Please state the assumptions of a random effects model, and evaluate whether these assumptions are met in the data.
- If the assumptions are, in fact, met in the data, then estimate a random effects model and interpret the coefficients of this model. Comment on how, if at all, the estimates from this model have changed compared to the fixed effects model.

- If the assumptions are **not** met, then do not estimate the data. But, also comment on what the consequences would be if you were to *inappropriately* estimate a random effects model. Would your coefficient estimates be biased or not? Would your standard error estimates be biased or not? Or, would there be some other problem that might arise?

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = tot_fata_per_100th_pop ~ year_of_observation +
##     bld_alc_lmt_cat + factor(per_se_law) + factor(round(prim_seatbelt_law)) +
##     factor(round(second_seatbelt_law)) + factor(round(speed_lim_grter70)) +
##     factor(round(grad_driver_license_law)) + age1424_pop_percent +
##     unemp_rate + veh_mile_trav_percap, data = data, effect = "individual",
##     model = "random", index = c("state_name", "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Effects:
##                 var std.dev share
## idiosyncratic 4.068   2.017 0.333
## individual    8.131   2.851 0.667
## theta: 0.8599
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -8.11399 -1.17662 -0.14744  0.92184 16.52346
##
## Coefficients:
##                                          Estimate  Std. Error  z-value
## (Intercept)                            1.5788e+01  2.1181e+00   7.4536
## year_of_observation1981               -1.5510e+00  4.2938e-01  -3.6121
## year_of_observation1982               -3.2776e+00  4.5899e-01  -7.1409
## year_of_observation1983               -3.9165e+00  4.7112e-01  -8.3131
## year_of_observation1984               -4.5252e+00  4.7579e-01  -9.5109
## year_of_observation1985               -5.0077e+00  4.9642e-01 -10.0876
## year_of_observation1986               -3.9513e+00  5.3052e-01  -7.4480
## year_of_observation1987               -4.6224e+00  5.6891e-01  -8.1249
## year_of_observation1988               -5.1162e+00  6.1621e-01  -8.3027
## year_of_observation1989               -6.4955e+00  6.5563e-01  -9.9073
## year_of_observation1990               -6.6771e+00  6.8042e-01  -9.8131
## year_of_observation1991               -7.4487e+00  6.9779e-01 -10.6747
## year_of_observation1992               -8.3746e+00  7.1928e-01 -11.6430
## year_of_observation1993               -8.7041e+00  7.3194e-01 -11.8918
## year_of_observation1994               -9.1153e+00  7.4940e-01 -12.1634
## year_of_observation1995               -8.8278e+00  7.7297e-01 -11.4206
## year_of_observation1996               -9.2735e+00  8.1441e-01 -11.3868
## year_of_observation1997               -9.4469e+00  8.3189e-01 -11.3559
## year_of_observation1998               -1.0094e+01  8.4927e-01 -11.8851
## year_of_observation1999               -1.0267e+01  8.5684e-01 -11.9827
## year_of_observation2000               -1.0797e+01  8.6816e-01 -12.4370
## year_of_observation2001               -1.0630e+01  8.7840e-01 -12.1011
## year_of_observation2002               -1.0018e+01  8.8696e-01 -11.2947
## year_of_observation2003               -1.0112e+01  8.9068e-01 -11.3532
```

```
## year_of_observation2004                -1.0419e+01  9.1468e-01 -11.3911
## bld_alc_lmt_catblood alcohol lmt0.1     3.4699e-01  2.5060e-01   1.3846
## bld_alc_lmt_catNo blood alcohol lmt     1.2664e+00  3.3941e-01   3.7311
## factor(per_se_law)1                    -1.0144e+00  2.2935e-01  -4.4230
## factor(round(prim_seatbelt_law))1      -1.1992e+00  3.5216e-01  -3.4054
## factor(round(second_seatbelt_law))1    -3.5381e-01  2.6066e-01  -1.3574
## factor(round(speed_lim_grter70))1       5.7038e-02  2.6884e-01   0.2122
## factor(round(grad_driver_license_law))1 -2.8008e-01  2.9018e-01  -0.9652
## age1424_pop_percent                     2.0316e-01  9.7274e-02   2.0886
## unemp_rate                             -4.9503e-01  6.1912e-02  -7.9957
## veh_mile_trav_percap                    1.1640e-03  1.0951e-04  10.6292
##                                         Pr(>|z|)
## (Intercept)                            9.083e-14 ***
## year_of_observation1981                0.0003037 ***
## year_of_observation1982                9.273e-13 ***
## year_of_observation1983                < 2.2e-16 ***
## year_of_observation1984                < 2.2e-16 ***
## year_of_observation1985                < 2.2e-16 ***
## year_of_observation1986                9.476e-14 ***
## year_of_observation1987                4.477e-16 ***
## year_of_observation1988                < 2.2e-16 ***
## year_of_observation1989                < 2.2e-16 ***
## year_of_observation1990                < 2.2e-16 ***
## year_of_observation1991                < 2.2e-16 ***
## year_of_observation1992                < 2.2e-16 ***
## year_of_observation1993                < 2.2e-16 ***
## year_of_observation1994                < 2.2e-16 ***
## year_of_observation1995                < 2.2e-16 ***
## year_of_observation1996                < 2.2e-16 ***
## year_of_observation1997                < 2.2e-16 ***
## year_of_observation1998                < 2.2e-16 ***
## year_of_observation1999                < 2.2e-16 ***
## year_of_observation2000                < 2.2e-16 ***
## year_of_observation2001                < 2.2e-16 ***
## year_of_observation2002                < 2.2e-16 ***
## year_of_observation2003                < 2.2e-16 ***
## year_of_observation2004                < 2.2e-16 ***
## bld_alc_lmt_catblood alcohol lmt0.1    0.1661613
## bld_alc_lmt_catNo blood alcohol lmt    0.0001906 ***
## factor(per_se_law)1                    9.735e-06 ***
## factor(round(prim_seatbelt_law))1      0.0006608 ***
## factor(round(second_seatbelt_law))1    0.1746679
## factor(round(speed_lim_grter70))1      0.8319814
## factor(round(grad_driver_license_law))1 0.3344486
## age1424_pop_percent                    0.0367456 *
## unemp_rate                             1.288e-15 ***
## veh_mile_trav_percap                   < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12850
## Residual Sum of Squares: 5104.2
## R-Squared:      0.60278
## Adj. R-Squared: 0.59118
```

```
## Chisq: 1767.85 on 34 DF, p-value: < 2.22e-16
```




```
##
##  Hausman Test
##
## data:  tot_fata_per_100th_pop ~ year_of_observation + bld_alc_lmt_cat +  ...
## chisq = 164.12, df = 34, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

The random effect model uses all of the assumptions applied to the fixed model with the addition of the following assumptions: - The unobserved effect is uncorrelated to all explanatory variables. This assumes the fixed effects at the state level are not correlated to the explanatory variable such as per se law, miles travel per, and speed limit. - The expected value of the unobserved effect given the explanatory variables is constant. - The variance of the unobserved effect given the explanatory variables is constant.

The first random effect assumption may likely be violated. The time-constant effect such as state location may have an impact on speed limit, which in turn may impact total fatalities rate. In addition, the location effect or state population may have an impact on miles travel per capita which in turn may impact total fatalities rate.

If this assumption is violated, coefficient estimates become biased and not as reliable as the fixed effect model.

For comparison purposes, we developed the random effect model to confirm our evaluation of the model assumptions.

The p-value of the Hausman test is less than 0.05, which means that we reject the null hypothese that the random effect model is appropriate. The test result also agrees with our evaluations of the random effect assumption and suggests that fixed effect model is a more appropriate approach.

7 (10 points) Model Forecasts

The COVID-19 pandemic dramatically changed patterns of driving. Find data (and include this data in your analysis, here) that includes some measure of vehicle miles driven in the US. Your data should at least cover the period from January 2018 to as current as possible. With this data, produce the following statements:

- Comparing monthly miles driven in 2018 to the same months during the pandemic:
  - What month demonstrated the largest decrease in driving? How much, in percentage terms, lower was this driving?
  - What month demonstrated the largest increase in driving? How much, in percentage terms, higher was this driving?

Now, use these changes in driving to make forecasts from your models.

- Suppose that the number of miles driven per capita, increased by as much as the COVID boom. Using the FE estimates, what would the consequences be on the number of traffic fatalities? Please interpret the estimate.
- Suppose that the number of miles driven per capita, decreased by as much as the COVID bust. Using the FE estimates, what would the consequences be on the number of traffic fatalities? Please interpret the estimate.

```
## Rows: 645
## Columns: 2
## $ DATE              <chr> "1970-01-01", "1970-02-01", "1970-03-01", "1970-04-01~
## $ TRFVOLUSM227NFWA <dbl> 80173, 77442, 90223, 89956, 97972, 100035, 106392, 10~
```

The team obtained the data from FRED Economic data. The data is the monthly vehicle miles traveled in the US from 1970. We averaged the vehicle miles traveled in 2020 and 2021 as a presentation for the travel during the pandemic. Overall, the traveling distance during the pandemic is lower than that in 2018. The team found that the smallest difference in terms of miles travel is in January with a difference of 0.0074897 percent of 2018 miles traveled and the largest difference is in April with a difference of 0.237016. The team then estimate the nation-wide miles traveled per capita from 1980 until 2004 to be 8691.2207973. Hypothetically, if the miles traveled per capita reduces as much as the highest point during the pandemic, the nation-wide miles traveled per capita would be 6631.2624414 which would result in a reduction of 2059.9583559. If the miles traveled per capita reduces as much as the lowest point during the pandemic, the nation_wide miles traveled per capita would be 8626.1261332 which would result in a reduction of 65.0946641.

The coefficient estimate of the mile travel per capita is 9.2611593×10^{-4}, which means that for every mile traveled per capita increase, the traffic fatalities rate increases by 9.2611593×10^{-4}. If the mile traveled per capita decreases, the traffic fatalities rate would decrease accordingly.

The reduction of traffic fatalities would be 1.9077603 with the highest drop in mile traveled per capita and 0.0602852 with the lowest drop.

8 (5 points) Evaluate Error

If there were serial correlation or heteroskedasticity in the idiosyncratic errors of the model, what would be the consequences on the estimators and their standard errors? Is there any serial correlation or heteroskedasticity?

```
##
##  Durbin-Watson test for serial correlation in panel models
##
## data:  tot_fata_per_100th_pop ~ year_of_observation + bld_alc_lmt_cat +  ...
## DW = 1.0619, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

```
##
##  Breusch-Pagan LM test for cross-sectional dependence in panels
##
## data:  tot_fata_per_100th_pop ~ year_of_observation + bld_alc_lmt_cat +     factor(per_se_law) + fac
## chisq = 3396.7, df = 1128, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```

Both Breusch-Pagan and Durbin-Watson test show that the idiosyncratic error is heteroskedastic and serial correlated. Serial correlation in idiosyncratic would result in model coefficient estimates to be inefficient. And heteroskedastic would make the coefficient estimates unreliable as we cannot determine the true confidence interval for the parameters. For this model, using robust standard errors would be a more appropriate approach to estimate model coefficients.