Lab 3: Panel Models

US Traffic Fatalities: 1980 - 2004

Contents

```
## Warning: package 'plm' was built under R version 4.2.3
```

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, ~
## $ slnone       <dbl> 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, 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, ~
## $ sbsecon      <int> 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, ~
## $ d81          <int> 0, 1, 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, ~
## $ d83          <int> 0, 0, 0, 1, 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, ~
## $ d85          <int> 0, 0, 0, 0, 0, 1, 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, ~
## $ d87          <int> 0, 0, 0, 0, 0, 0, 0, 1, 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, ~
## $ d89          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 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, ~
## $ d91          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 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, ~
## $ d93          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, ~
## $ d94          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, ~
## $ d95          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, ~
## $ d96          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, ~
## $ d97          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, ~
## $ 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, ~
## $ d00          <int> 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, ~
## $ d02          <int> 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, ~
## $ d04          <int> 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
```

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`. #TODO: Asked Vinod if this is stil necessary, since we didn't do this
   - 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.)

```r
df <- data %>%
  mutate(state = factor(state)) %>%
  rowwise() %>%
  # speed_limit
  mutate(
    speed_limit_70plus = factor(sl70plus),
    speed_limit = parse_number(
      colnames(
        select(data, starts_with("sl"))
      )[which.max(c_across(starts_with("sl")))],
      na = "slnone"
    ),
  ) %>%
  select(-starts_with("sl")) %>%
  mutate(year_of_observation = factor(year)) %>% # year_of_observation
  select(-starts_with("d")) %>%
  mutate(blood_alcohol_limit = parse_number(
    colnames(
      select(data, starts_with("bac"))
    )[which.max(c_across(starts_with("bac")))]
  ) / 100) %>% # blood_alcohol_limit
  select(-starts_with("bac")) %>%
  mutate(
    seatbelt = factor(seatbelt), # 'seatbelt' categorizes primary or secondary
    speed_limit_70plus = ifelse(speed_limit == 55 | speed_limit == 65, 0, 1)
  ) %>%
  select(-starts_with("sb"))


df <- df %>%
  dplyr::rename(
    "total_fatalities_rate" = "totfatrte",
    "minimum_drinking_age" = "minage",
    "zero_tolerance_law" = "zerotol",
    "graduated_drivers_license_law" = "gdl",
    "per_se_laws" = "perse",
    "total_traffic_fatalities" = "totfat",
    "total_nighttime_fatalities" = "nghtfat",
    "total_weekend_fatalities" = "wkndfat",
    "total_fatalities_per_100_million_miles" = "totfatpvm",
    "nighttime_fatalities_per_100_million_miles" = "nghtfatpvm",
    "weekend_fatalities_per_100_million_miles" = "wkndfatpvm",
    "nighttime_fatalities_rate" = "nghtfatrte",
    "weekend_fatalities_rate" = "wkndfatrte",
    "vehicle_miles" = "vehicmiles",
    "unemployment_rate" = "unem",
    "pct_population_14_to_24" = "perc14_24",
    "vehicle_miles_per_capita" = "vehicmilespc"
  ) %>%
  select(
    year_of_observation,
    state,
    year,
    # response variables
    total_fatalities_rate,
    nighttime_fatalities_rate,
    weekend_fatalities_rate,
    total_traffic_fatalities,
    total_nighttime_fatalities,
    total_weekend_fatalities,
    total_fatalities_per_100_million_miles,
    nighttime_fatalities_per_100_million_miles,
    weekend_fatalities_per_100_million_miles,
    # potential explanatory variables
    seatbelt,
    zero_tolerance_law,
    graduated_drivers_license_law,
    per_se_laws,
    minimum_drinking_age,
    speed_limit_70plus,
    speed_limit,
    blood_alcohol_limit,
    vehicle_miles,
    vehicle_miles_per_capita,
    # econ and demographic variables
    statepop,
    unemployment_rate,
    pct_population_14_to_24, vehicle_miles
  ) # keep the similar variables together

df %>% glimpse()
```

```
## Rows: 1,200
## Columns: 25
## Rowwise:
## $ year_of_observation                        <fct> 1980, 1981, 1982, 1983, 198~
## $ state                                      <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ year                                       <int> 1980, 1981, 1982, 1983, 198~
## $ total_fatalities_rate                      <dbl> 24.14, 24.07, 21.37, 23.64,~
## $ nighttime_fatalities_rate                  <dbl> 10.84, 11.08, 9.58, 10.09, ~
## $ weekend_fatalities_rate                    <dbl> 6.060000, 6.330000, 5.71000~
## $ total_traffic_fatalities                   <int> 940, 933, 839, 930, 932, 88~
## $ total_nighttime_fatalities                 <int> 422, 434, 376, 397, 421, 35~
## $ total_weekend_fatalities                   <int> 236, 248, 224, 223, 237, 22~
## $ total_fatalities_per_100_million_miles     <dbl> 3.200, 3.350, 2.810, 3.000,~
## $ nighttime_fatalities_per_100_million_miles <dbl> 1.437, 1.558, 1.259, 1.281,~
## $ weekend_fatalities_per_100_million_miles   <dbl> 0.803, 0.890, 0.750, 0.719,~
## $ seatbelt                                   <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ zero_tolerance_law                         <dbl> 0.000, 0.000, 0.000, 0.000,~
## $ graduated_drivers_license_law              <dbl> 0.00, 0.00, 0.00, 0.00, 0.0~
## $ per_se_laws                                <dbl> 0.000, 0.000, 0.000, 0.000,~
## $ minimum_drinking_age                       <dbl> 18, 18, 18, 18, 18, 20, 21,~
## $ speed_limit_70plus                         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ speed_limit                                <dbl> 55, 55, 55, 55, 55, 55, 55,~
## $ blood_alcohol_limit                        <dbl> 0.10, 0.10, 0.10, 0.10, 0.1~
## $ vehicle_miles                              <dbl> 29.37500, 27.85200, 29.8576~
## $ vehicle_miles_per_capita                   <dbl> 7543.874, 7107.785, 7606.62~
## $ statepop                                   <int> 3893888, 3918520, 3925218, ~
## $ unemployment_rate                          <dbl> 8.8, 10.7, 14.4, 13.7, 11.1~
## $ pct_population_14_to_24                    <dbl> 18.9, 18.7, 18.4, 18.0, 17.~
```

data %>% filter(sl65>0 & sl65<1) %>% select(sl55,sl65,sl70,sl75,slnone) %>% mutate(sum_sl = sum(sl55,sl65,sl70,sl75,slnone)), we are choosing the max when it's pct

speed_limit_70plus column has values that are not 0 or 1, reclassify based on speed_limit column. Also classify the states with no speed limit as 1 for this variable.

```
df$zero_tolerance_law %>%
  table(useNA = "ifany") %>%
  as.data.frame()
```

```
##                       . Freq
## 1                     0  636
## 2  0.0829999968409538    1
## 3   0.166999995708466    2
## 4                0.25    6
## 5   0.333000004291534    2
## 6   0.416999995708466    3
## 7                 0.5   17
## 8   0.583000004291534    5
## 9   0.666999995708466    2
## 10               0.75    1
## 11                  1  525
```

```
df$graduated_drivers_license_law %>%
  table(useNA = "ifany") %>%
  as.data.frame()
```

```
##                     . Freq
## 1                   0  981
## 2 0.166999995708466    1
## 3              0.25    2
## 4               0.5   14
## 5 0.670000016689301    1
## 6              0.75    1
## 7 0.833000004291534    1
## 8                   1  199
```

```
df$per_se_laws %>%
  table(useNA = "ifany") %>%
  as.data.frame()
```

```
##                     . Freq
## 1                   0  528
## 2 0.0829999968409538    1
## 3  0.166999995708466    1
## 4                0.25    4
## 5  0.333000004291534    2
## 6  0.416999995708466    2
## 7                 0.5   16
## 8                0.75    1
## 9                   1  645
```

```
df <- df %>%
  mutate(
    zero_tolerance_law = ifelse(
      zero_tolerance_law == 0 | zero_tolerance_law == 1, zero_tolerance_law, 1
    ),
    graduated_drivers_license_law = ifelse(
      graduated_drivers_license_law == 0 | graduated_drivers_license_law == 1,
      graduated_drivers_license_law,
      1
    ),
    per_se_laws = ifelse(
      per_se_laws == 0 | per_se_laws == 1, per_se_laws, 1
    )
  )
```

We observed non-binary values in the following columns: zero_tolerance_law, graduated_drivers_license_law, per_se_laws. But we expect them to be binary given the definition.

We decided to treat all non-zero values as 1 and make it a binary variable.

```
df$minimum_drinking_age %>%
  table(useNA = "ifany") %>%
  as.data.frame()
```

```
##                    . Freq
## 1                 18   98
## 2               18.5    5
## 3   18.6000003814697    1
## 4   18.7000007629395    4
## 5                 19   58
## 6               19.5    5
## 7   19.7000007629395    2
## 8   19.7999992370605    1
## 9                 20   35
## 10              20.5    2
## 11 20.7000007629395    4
## 12                21  985
```

```
df <- df %>%
  mutate(
    minimum_drinking_age = round(minimum_drinking_age, 0)
  )
```

We noticed that the minimum_drinking_age column has values that are not integers. We decided to round them to the nearest integer.

```
df %>%
  filter(is.na(speed_limit)) %>%
  select(state, year_of_observation, speed_limit, speed_limit_70plus)
```

```
## # A tibble: 9 x 4
## # Rowwise:
##   state year_of_observation speed_limit speed_limit_70plus
##   <fct> <fct>                     <dbl>              <dbl>
## 1 27    1996                         NA                 NA
## 2 27    1997                         NA                 NA
## 3 27    1998                         NA                 NA
## 4 27    1999                         NA                 NA
## 5 27    2000                         NA                 NA
## 6 27    2001                         NA                 NA
## 7 27    2002                         NA                 NA
## 8 27    2003                         NA                 NA
## 9 27    2004                         NA                 NA
```

```
# TODO: verifying with Vinod
df <- df %>%
  mutate(
    speed_limit = ifelse(
      is.na(speed_limit) & state == 27, ifelse(
        year >= 1996 & year <= 1999, 85, 75
      ), speed_limit
    ),
    speed_limit_70plus = ifelse(
      is.na(speed_limit_70plus) & state == 27, 1, speed_limit_70plus
    )
  )
```

Looking at the data, we realized the *speed_limit* is not set for State 27, between 1996 to 2004. Through some background research, this is reflecting the fact that "for three years after the 1995 repeal of the increased 65 mph limit, Montana had a non-numeric"reasonable and prudent" speed limit during the daytime on most rural roads". But it doesn't mean there was no speed limit.

We decided to set the *speed_limit* to 85 for Montana between 1996 to 1999, given the legal case of State v. Rudy Stanko (1998), who got charged for speed of 85. Effective May 28, 1999, as a result of that decision, the Montana Legislature established a speed limit of 75 mph. So we set the *speed_limit* to 75 for Montana between 2000 to 2004.

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?
      This data set is a balanced longitudinal dataset and contains traffic fatalities data for the 48 continental U.S. states from 1980 through 2004. For each year of observation, the dataset contains state-level cross sectional measurements of fatality count and rate. This data is collected and distributed by Jeffrey M. Wooldridge through this link.

After our data processing work, the clean dataset has 25 columns/fields which include:

- Index variables: year_of_observation, state
- 9 fatality variables: There are three measurements - fatality count, fatality count per 100M miles and fatality rate as defined as count per 100k population. These three measurements are provided for total, nighttime and weekend
- 10 law and vehicle variables: 8 traffic laws indicators (seatbelt, zero_tolerance_law, graduated_drivers_license_law, per_se_laws, minimum_drinking_age, speed_limit_70plus, speed_limit, blood_alcohol_limit) and 2 driving variables (vehicle_miles, vehicle_miles_per_capita)
- 3 Economics and demographic variables: statepop, unemployment_rate, pct_population_14_to_24

In this dataset, the total_fatalities_rate is defined as total fatalities per 100,000 population.

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:

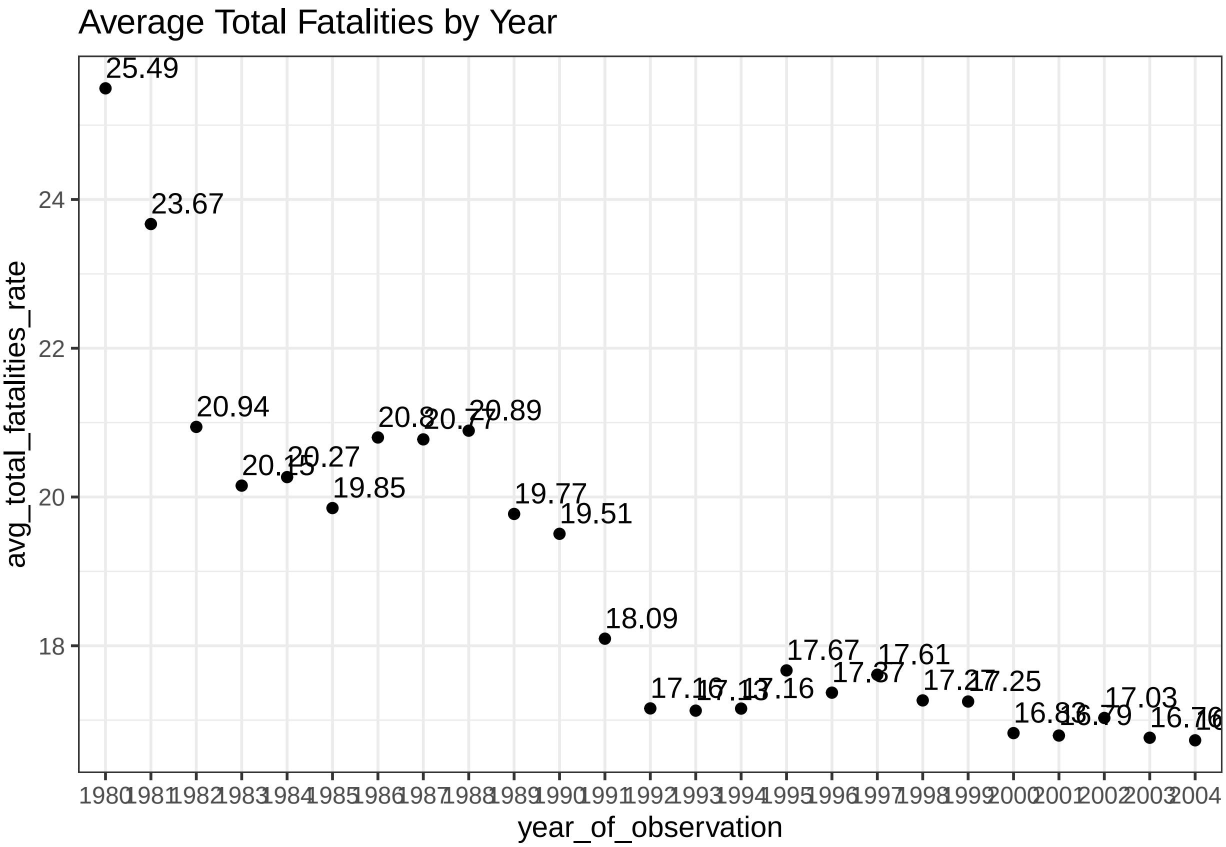
- How is the our dependent variable of interest `total_fatalities_rate` defined?

  In this dataset, the total_fatalities_rate is defined as total fatalities per 100,000 population.

- What is the average of `total_fatalities_rate` in each of the years in the time period covered in this dataset?

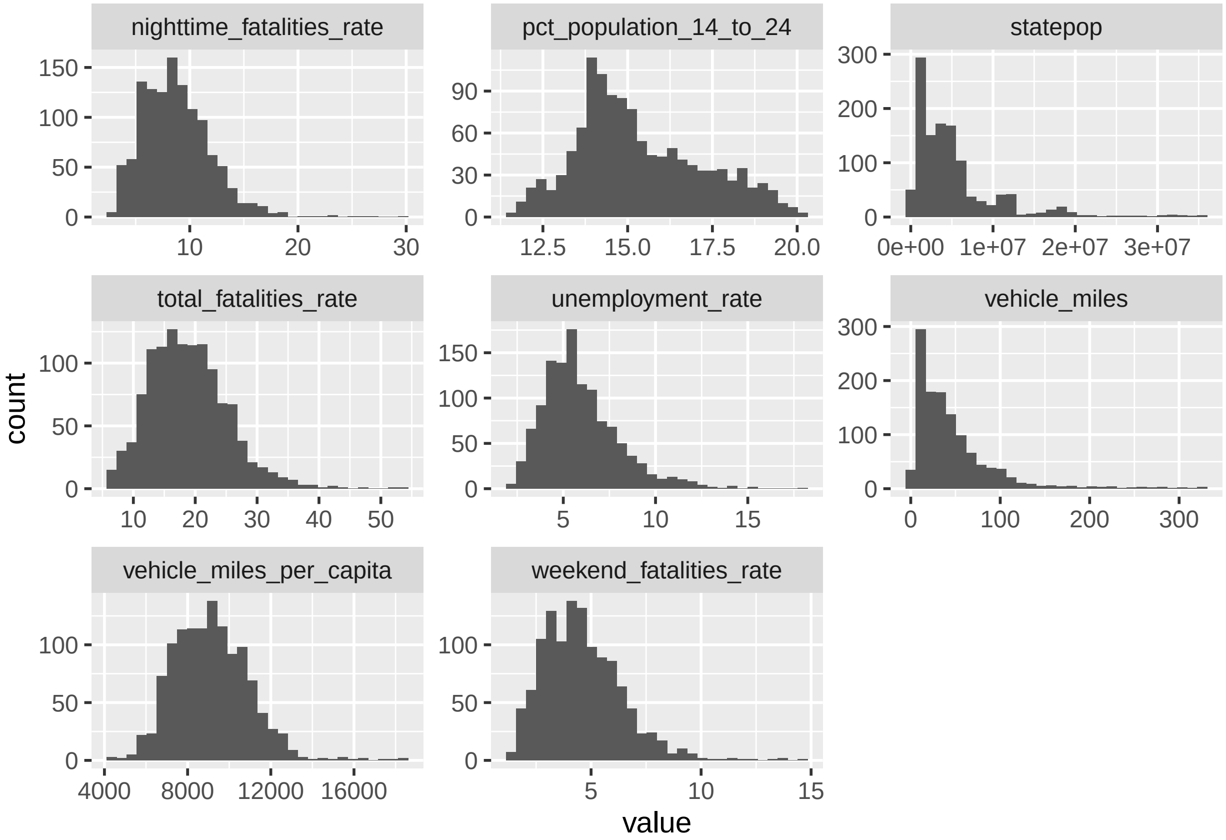
```r
df_avg <- df %>%
  group_by(year_of_observation) %>%
  summarise(avg_total_fatalities_rate = mean(total_fatalities_rate))
# average fatality by year
years <- unique(df$year_of_observation)
avg_df <- data.frame(
  year = years,
  avg_fatality = round(df_avg$avg_total_fatalities_rate, 2)
)

# plot fatality by year
df_avg %>%
  ggplot(aes(year_of_observation, avg_total_fatalities_rate,
    label = round(avg_total_fatalities_rate, 2)
  )) +
  geom_point() +
  geom_text(hjust = 0, vjust = -0.5) +
  theme_bw() +
  labs(title = "Average Total Fatalities by Year")
```



```
df %>%
  select(
    total_fatalities_rate,
    nighttime_fatalities_rate,
    weekend_fatalities_rate,
    vehicle_miles,
    vehicle_miles_per_capita,
    statepop,
    unemployment_rate,
    pct_population_14_to_24
  ) %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +
  facet_wrap(~key, scales = "free") +
  geom_histogram()
```

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

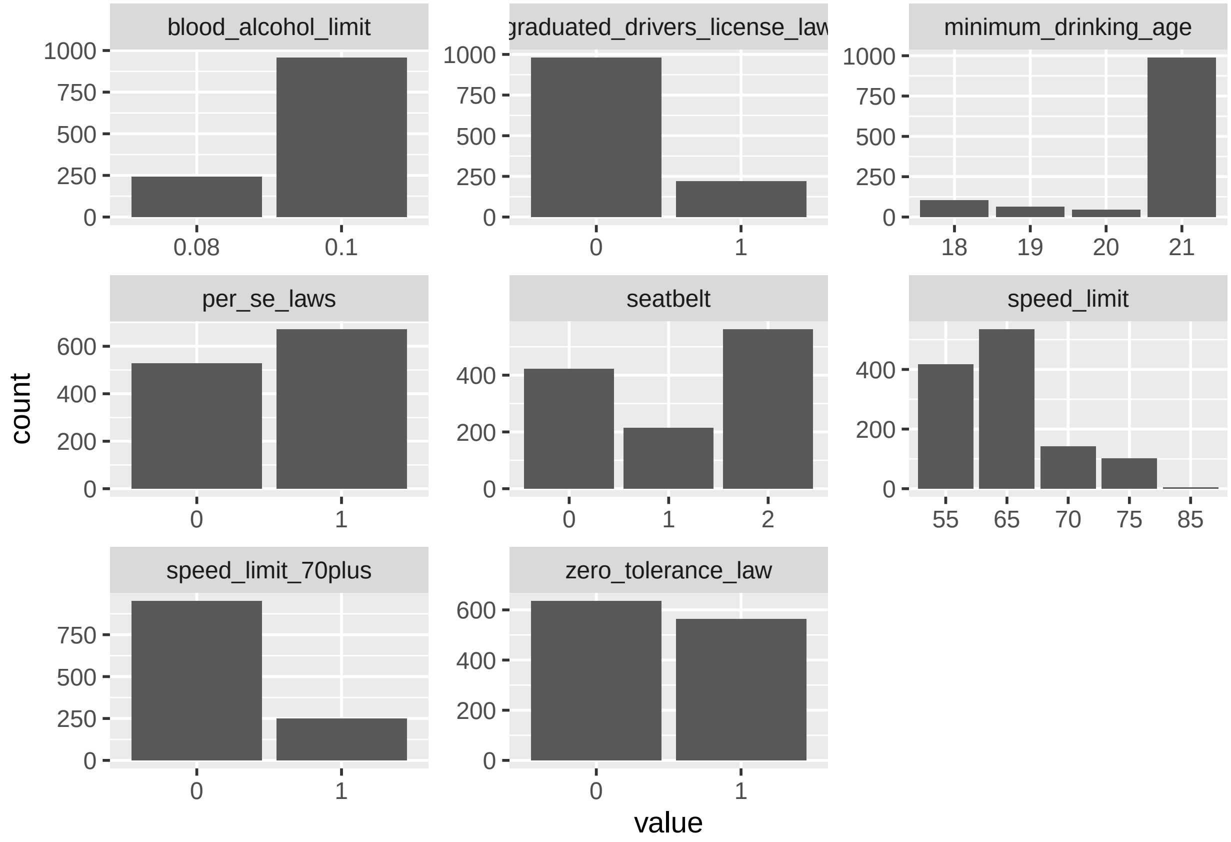


```
# hist(df$total_fatalities_rate)
# hist(log(df$total_fatalities_rate))
```

For the continuous variables, the distributions are right-skewed for most of these variables.

```
df %>%
  select(
    seatbelt,
    zero_tolerance_law,
    graduated_drivers_license_law,
    per_se_laws,
    minimum_drinking_age,
    speed_limit_70plus,
    speed_limit,
    blood_alcohol_limit
  ) %>%
  gather() %>%
  ggplot(aes(value)) +
  facet_wrap(~key, scales = "free") +
  geom_bar()
```

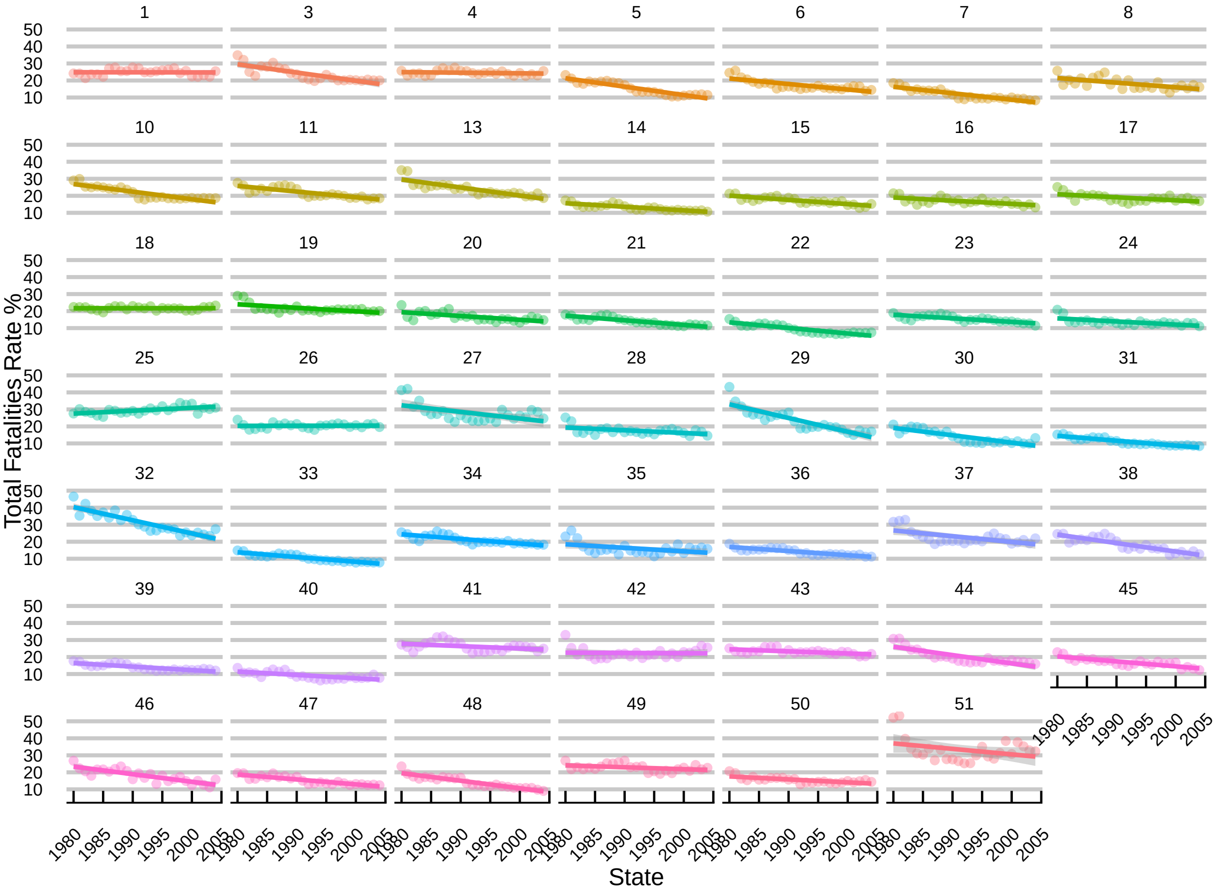
```
## Warning: attributes are not identical across measure variables; they will be
## dropped
```



We observed the following distribution for these categorical variables - Blood alcohol limit: Most states have the limit of 0.1. - Minimum drinking age: Most states have 21. - Graduated drivers license law, most states have 0 - Speed limit: Most states have less than 70 miles

```
# TODO: For Ken: why are there 51 plots, with 48 states?
df %>%
  ggplot(aes(year, total_fatalities_rate, color = state)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm") +
  facet_wrap(~state) +
  theme_economist_white(gray_bg = FALSE) +
  theme(
    legend.position = "none",
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, size = 8),
    axis.text.y = element_text(size = 8),
    strip.text = element_text(size = 8)
  ) +
  scale_y_continuous() +
  xlab("State") +
  ylab("Total Fatalities Rate %")
```

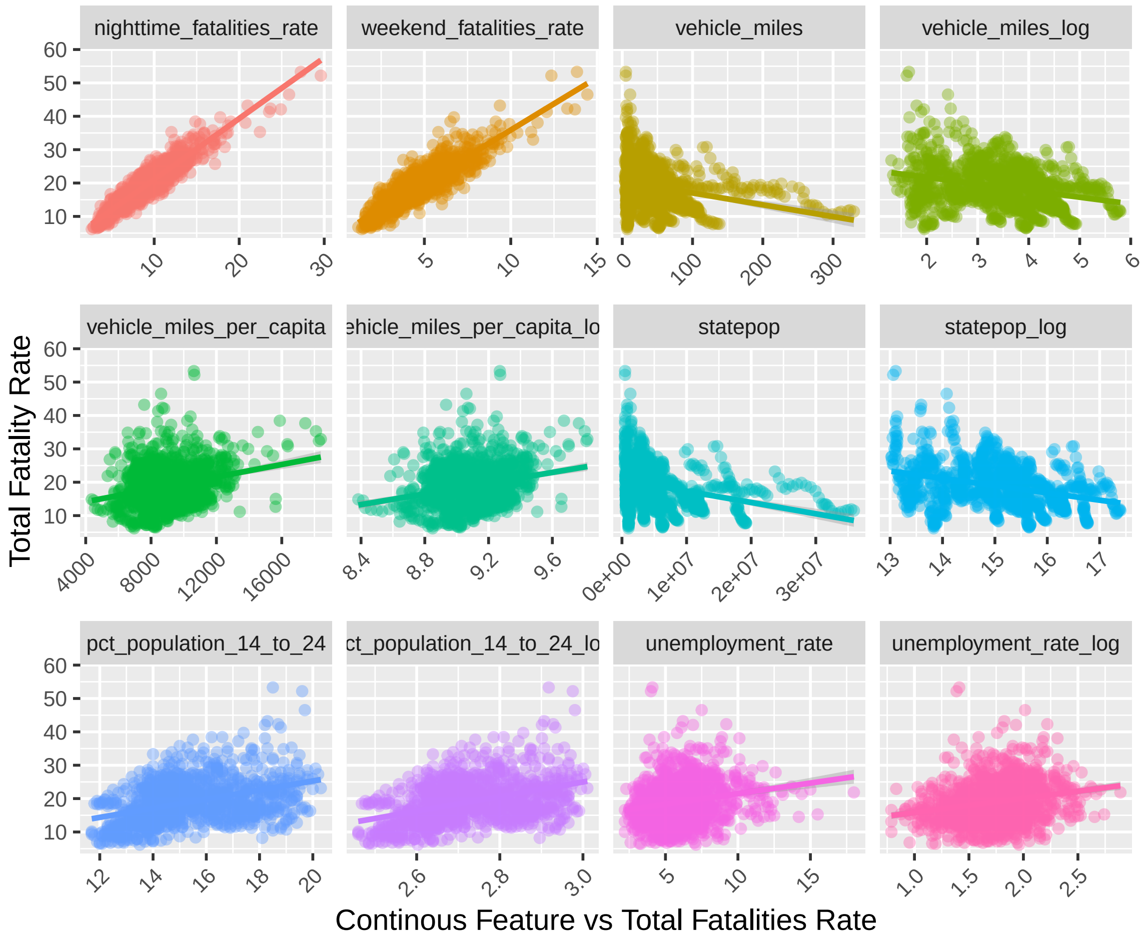
```
## `geom_smooth()` using formula = 'y ~ x'
```



Most states have downward trends in total fatalities rate.

```
df %>%
  mutate(
    vehicle_miles_log = log(vehicle_miles),
    vehicle_miles_per_capita_log = log(vehicle_miles_per_capita),
    statepop_log = log(statepop),
    pct_population_14_to_24_log = log(pct_population_14_to_24),
    unemployment_rate_log = log(unemployment_rate)
  ) %>%
  select(
    total_fatalities_rate,
    nighttime_fatalities_rate,
    weekend_fatalities_rate,
    vehicle_miles,
    vehicle_miles_log,
    vehicle_miles_per_capita,
    vehicle_miles_per_capita_log,
    statepop,
    statepop_log,
    pct_population_14_to_24,
    pct_population_14_to_24_log,
    unemployment_rate,
    unemployment_rate_log
  ) %>%
  melt(id.vars = c("total_fatalities_rate")) %>%
  ggplot(aes(value, total_fatalities_rate, color = variable)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm") +
  facet_wrap(~variable, scales = "free_x") +
  # theme_economist_white(gray_bg=F) +
  theme(
    legend.position = "none",
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1, size = 8),
    axis.text.y = element_text(size = 8),
    strip.text = element_text(size = 8)
  ) +
  # scale_y_continuous(label=percent) +
  xlab("Continous Feature vs Total Fatalities Rate") +
  ylab("Total Fatality Rate")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



We conducted log transformation on the following variables: vehicle_miles, vehicle_miles_per_capita, statepop, pct_population_14_to_24, unemployment_rate. We visualized both the original and log transformed variables, and decided to use log transformation for our interpretation for vehicle_miles, vechicle_miles_per_capita, statepop and unemployment_rate.

Total fatalities rate is positively correlated with unemployment_rate and percentage of population aged 14 through 24.

Total fatalities rate is negatively correlated with vehicle miles and state population, but positively correlated with vehicle miles per capita.

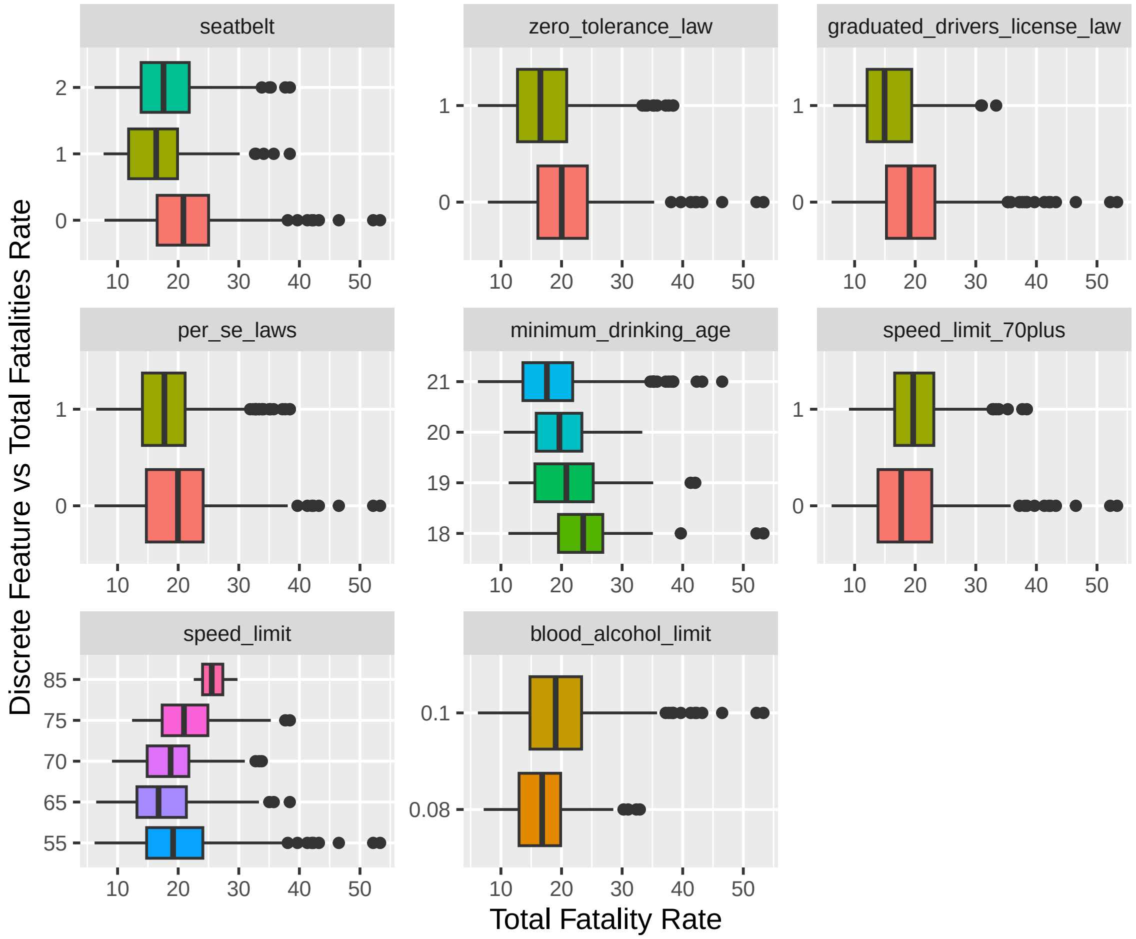
We interpret this as population and vehicle miles are both increasing with time, so is the other driving forces of fatalities rate (quality of the car, technology of the car, road conditions, etc.), so the relationship between total fatalities rate vs vehicle miles and state population is potentially spurious, and thus inconsistent with our background knowlege.

However, the relationship between total fatalities rate and vehicle miles per capita is positive, because as the density of the population increases, there is expected to be more severe traffic incidents that leads to higher fatalities rate. This is consistent with our background knowledge.

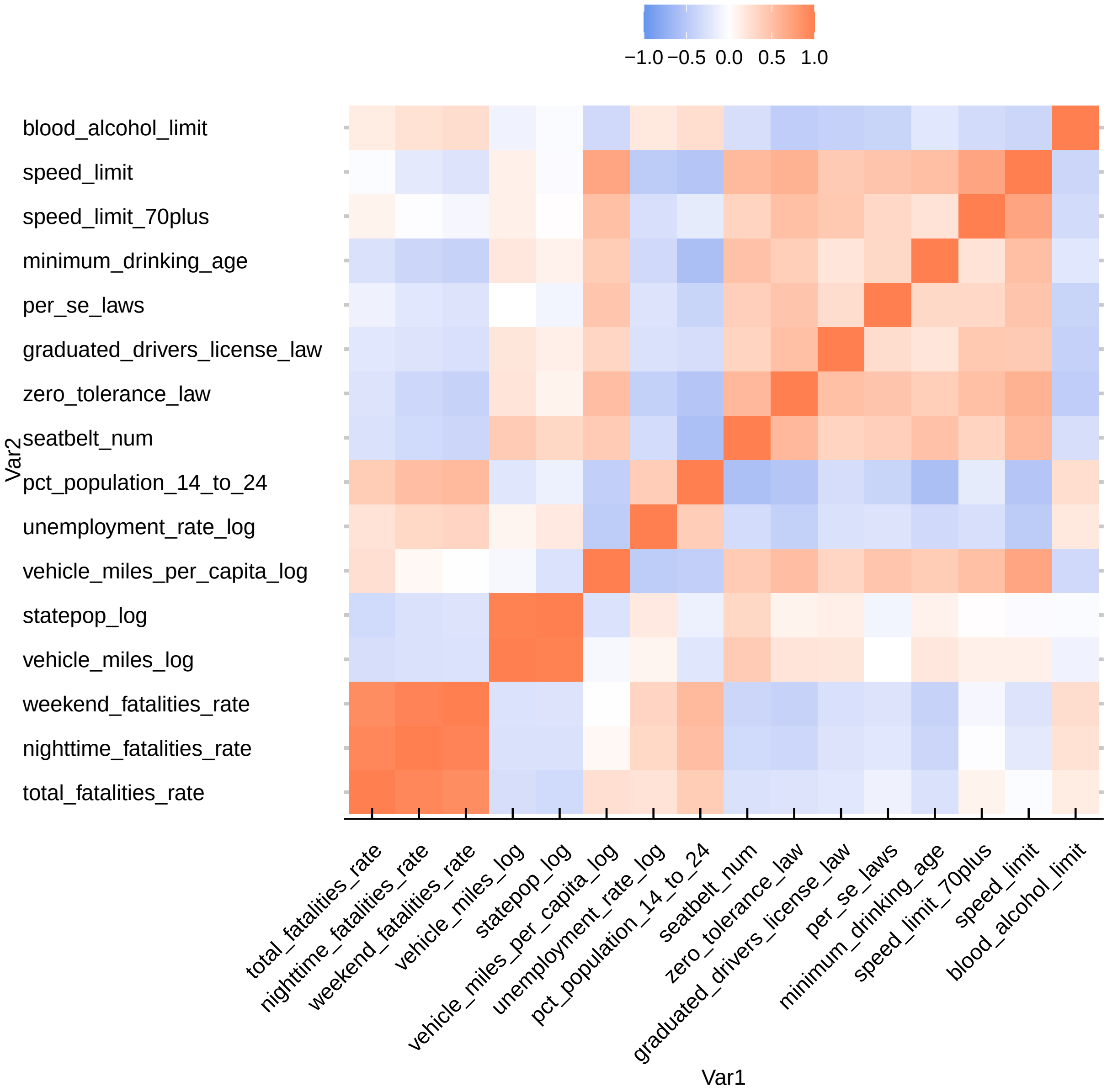
```
# TODO: talk about this?
df <- df %>%
  mutate(
    vehicle_miles_log = log(vehicle_miles),
    vehicle_miles_per_capita_log = log(vehicle_miles_per_capita),
    statepop_log = log(statepop),
    unemployment_rate_log = log(unemployment_rate)
  ) %>%
  select(
    -vehicle_miles,
    -vehicle_miles_per_capita,
    -statepop,
    -unemployment_rate
  )
```

```
df %>%
  select(
    total_fatalities_rate,
    seatbelt,
    zero_tolerance_law,
    graduated_drivers_license_law,
    per_se_laws,
    minimum_drinking_age,
    speed_limit_70plus,
    speed_limit,
    blood_alcohol_limit
  ) %>%
  melt(id.vars = c("total_fatalities_rate")) %>%
  ggplot(aes(value, total_fatalities_rate)) +
  geom_boxplot(aes(fill = factor(value))) +
  coord_flip() +
  facet_wrap(~variable, scales = "free") +
  theme(
    legend.position = "none",
    axis.text.x = element_text(size = 8),
    axis.text.y = element_text(size = 8),
    strip.text = element_text(size = 8)
  ) +
  xlab("Discrete Feature vs Total Fatalities Rate") +
  ylab("Total Fatality Rate")
```

```
## Warning: attributes are not identical across measure variables; they will be
## dropped
```



```
df %>%
  # TODO: does this even make sense to calculate corrolation on discrete variables? Don't we need to us
  mutate(seatbelt_num = as.numeric(seatbelt)) %>%
  select(
    total_fatalities_rate,
    nighttime_fatalities_rate,
    weekend_fatalities_rate,
    vehicle_miles_log,
    statepop_log,
    vehicle_miles_per_capita_log,
    unemployment_rate_log,
    pct_population_14_to_24,
    seatbelt_num,
    zero_tolerance_law,
    graduated_drivers_license_law,
    per_se_laws,
    minimum_drinking_age,
    speed_limit_70plus,
    speed_limit,
    blood_alcohol_limit
  ) %>%
  cor() %>%
  melt() %>%
  ggplot(aes(Var1, Var2, fill = value)) +
  geom_tile() +
  theme_economist_white(gray_bg = FALSE) +
  theme(
    legend.title = element_blank(),
    legend.text = element_text(size = 10),
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)
  ) +
  scale_fill_gradient2(
    low = "cornflowerblue", high = "coral", mid = "white",
    midpoint = 0, limit = c(-1, 1)
  )
```



The state population and vehicle miles are almost perfectly correlated. So to avoid the colinearity problem, we will only use vehicle_miles_per_capita in our model, which is impacted both of them.

2.0.1 Summary Of EDA

1. clean up the dataframe to keep only variables we care; #TODO: as a team, we can write some observations, and rationalize which are the variables we end up using in the next session.
2. transform some variables via log #TODO: later in the Expanded Model part: > A log transformation is applied to total_fatalities_rate and unemployment_rate because the skewed distribution needs to be normalized.

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.

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? > Fitting a linear model helps identify significant explanatory variables and evaluate how strong the linear correlation is.
- What does this model explain, and what do you find in this model? > This model explains whether a given year has a linear relationship with total fatalities rate. Based on the model, there is strong evidence that all the years except 1981 are related to total fatalities rate at the significance level of 0.

  This makes sense as we are using 1980 as the baseline year, so 1981 is the first year after 1980
  and as a result, all the time-variant effects have not kicked in.
- Did driving become safer over this period? Please provide a detailed explanation. > In 1980, the average total fatalities rate was 24% and by 2004, the average total fatalities decreased down to 16%.
- 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? > TODO: let's do this. > We can note on R^2.

```
mod.lm1 <- lm(total_fatalities_rate ~ year_of_observation, data = df)
summary(mod.lm1)
```

```
##
## Call:
## lm(formula = total_fatalities_rate ~ year_of_observation, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9302  -4.3468  -0.7305   3.7488  29.6498
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              25.4946     0.8671  29.401  < 2e-16 ***
## year_of_observation1981  -1.8244     1.2263  -1.488 0.137094
## year_of_observation1982  -4.5521     1.2263  -3.712 0.000215 ***
## year_of_observation1983  -5.3417     1.2263  -4.356 1.44e-05 ***
## year_of_observation1984  -5.2271     1.2263  -4.263 2.18e-05 ***
## year_of_observation1985  -5.6431     1.2263  -4.602 4.64e-06 ***
## year_of_observation1986  -4.6942     1.2263  -3.828 0.000136 ***
## year_of_observation1987  -4.7198     1.2263  -3.849 0.000125 ***
## year_of_observation1988  -4.6029     1.2263  -3.754 0.000183 ***
## year_of_observation1989  -5.7223     1.2263  -4.666 3.42e-06 ***
## year_of_observation1990  -5.9894     1.2263  -4.884 1.18e-06 ***
## year_of_observation1991  -7.3998     1.2263  -6.034 2.14e-09 ***
## year_of_observation1992  -8.3367     1.2263  -6.798 1.68e-11 ***
## year_of_observation1993  -8.3669     1.2263  -6.823 1.43e-11 ***
## year_of_observation1994  -8.3394     1.2263  -6.800 1.66e-11 ***
## year_of_observation1995  -7.8260     1.2263  -6.382 2.51e-10 ***
## year_of_observation1996  -8.1252     1.2263  -6.626 5.25e-11 ***
## year_of_observation1997  -7.8840     1.2263  -6.429 1.86e-10 ***
## year_of_observation1998  -8.2292     1.2263  -6.711 3.01e-11 ***
## year_of_observation1999  -8.2442     1.2263  -6.723 2.77e-11 ***
## year_of_observation2000  -8.6690     1.2263  -7.069 2.67e-12 ***
## year_of_observation2001  -8.7019     1.2263  -7.096 2.21e-12 ***
## year_of_observation2002  -8.4650     1.2263  -6.903 8.32e-12 ***
## year_of_observation2003  -8.7310     1.2263  -7.120 1.88e-12 ***
## year_of_observation2004  -8.7656     1.2263  -7.148 1.54e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.008 on 1175 degrees of freedom
## Multiple R-squared:  0.1276, Adjusted R-squared:  0.1098
## F-statistic: 7.164 on 24 and 1175 DF,  p-value: < 2.2e-16
```

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. > A log transformation is applied to total_fatalities_rate and unemployment_rate because the skewed distribution needs to be normalized.

```r
mod.lm2 <- lm(
  log(total_fatalities_rate) ~ year_of_observation
    + factor(blood_alcohol_limit)
    + per_se_laws
    + seatbelt
    + speed_limit_70plus
    + graduated_drivers_license_law
    + pct_population_14_to_24
    + unemployment_rate_log
    + vehicle_miles_per_capita_log,
  data = df
)
summary(mod.lm2)
```

```
##
## Call:
## lm(formula = log(total_fatalities_rate) ~ year_of_observation +
##     factor(blood_alcohol_limit) + per_se_laws + seatbelt + speed_limit_70plus +
##     graduated_drivers_license_law + pct_population_14_to_24 +
##     unemployment_rate_log + vehicle_miles_per_capita_log, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.58465 -0.12657 -0.00148  0.14135  0.61947
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -1.135e+01  4.023e-01 -28.208  < 2e-16 ***
## year_of_observation1981        -9.146e-02  4.117e-02  -2.221   0.0265 *
## year_of_observation1982        -2.949e-01  4.203e-02  -7.016 3.87e-12 ***
## year_of_observation1983        -3.516e-01  4.263e-02  -8.247 4.32e-16 ***
## year_of_observation1984        -3.056e-01  4.283e-02  -7.136 1.69e-12 ***
## year_of_observation1985        -3.439e-01  4.370e-02  -7.870 8.04e-15 ***
## year_of_observation1986        -3.208e-01  4.563e-02  -7.029 3.54e-12 ***
## year_of_observation1987        -3.570e-01  4.762e-02  -7.498 1.28e-13 ***
## year_of_observation1988        -3.673e-01  5.025e-02  -7.309 4.99e-13 ***
## year_of_observation1989        -4.525e-01  5.222e-02  -8.664  < 2e-16 ***
## year_of_observation1990        -5.119e-01  5.344e-02  -9.579  < 2e-16 ***
## year_of_observation1991        -6.277e-01  5.458e-02 -11.500  < 2e-16 ***
## year_of_observation1992        -7.338e-01  5.562e-02 -13.193  < 2e-16 ***
## year_of_observation1993        -7.251e-01  5.637e-02 -12.863  < 2e-16 ***
## year_of_observation1994        -7.111e-01  5.755e-02 -12.356  < 2e-16 ***
## year_of_observation1995        -6.887e-01  5.894e-02 -11.684  < 2e-16 ***
## year_of_observation1996        -8.135e-01  6.108e-02 -13.319  < 2e-16 ***
## year_of_observation1997        -8.217e-01  6.218e-02 -13.214  < 2e-16 ***
## year_of_observation1998        -8.686e-01  6.331e-02 -13.720  < 2e-16 ***
## year_of_observation1999        -8.664e-01  6.431e-02 -13.473  < 2e-16 ***
## year_of_observation2000        -8.768e-01  6.549e-02 -13.388  < 2e-16 ***
## year_of_observation2001        -9.303e-01  6.591e-02 -14.114  < 2e-16 ***
## year_of_observation2002        -9.744e-01  6.610e-02 -14.740  < 2e-16 ***
## year_of_observation2003        -9.969e-01  6.642e-02 -15.009  < 2e-16 ***
## year_of_observation2004        -9.795e-01  6.792e-02 -14.421  < 2e-16 ***
## factor(blood_alcohol_limit)0.1  4.539e-02  1.835e-02   2.473   0.0135 *
## per_se_laws                    -2.197e-02  1.437e-02  -1.529   0.1264
## seatbelt2                       1.936e-02  2.139e-02   0.905   0.3656
## seatbelt1                      -6.715e-04  2.448e-02  -0.027   0.9781
## speed_limit_70plus              2.211e-01  2.160e-02  10.238  < 2e-16 ***
## graduated_drivers_license_law  -3.434e-02  2.513e-02  -1.367   0.1720
## pct_population_14_to_24         1.780e-02  6.103e-03   2.917   0.0036 **
## unemployment_rate_log           2.673e-01  2.413e-02  11.078  < 2e-16 ***
## vehicle_miles_per_capita_log    1.541e+00  4.436e-02  34.747  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2013 on 1166 degrees of freedom
## Multiple R-squared:  0.668,  Adjusted R-squared:  0.6586
## F-statistic:  71.1 on 33 and 1166 DF,  p-value: < 2.2e-16
```

- How are the blood alcohol variables defined? Interpret the coefficients that you estimate for this concept. > #TODO: add the bac definition.
- Do *per se laws* have a negative effect on the fatality rate? > *per se laws* have a negative effect on the

fatality rate, but the variable yielded a p-value > 0.1 thus suggesting that there is weak evidence it is correlated with fatality rate.

- Does having a primary seat belt law? > *primary seatbelt laws* also has a negative effect on the fatality rate, but the variable yielded a p-value > 0.1 thus suggesting that there is weak evidence it is correlated with fatality rate.

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?

```r
# estimate the fixed effects regression with plm()
mod.fe <- plm(
  log(total_fatalities_rate) ~ year_of_observation
    + factor(blood_alcohol_limit)
    + per_se_laws
    + seatbelt
    + speed_limit_70plus
    + graduated_drivers_license_law
    + pct_population_14_to_24
    + unemployment_rate_log
    + vehicle_miles_per_capita_log,
  data = df,
  index = c("state"),
  model = "within"
)

# print summary using robust standard errors
coeftest(mod.fe, vcov. = vcovHC, type = "HC1")
```

```
##
## t test of coefficients:
##
##                                    Estimate Std. Error t value  Pr(>|t|)
## year_of_observation1981        -0.0620639  0.0168586 -3.6814  0.000243 ***
## year_of_observation1982        -0.1341301  0.0180606 -7.4267 2.203e-13 ***
## year_of_observation1983        -0.1664768  0.0216291 -7.6969 3.051e-14 ***
## year_of_observation1984        -0.2124940  0.0213503 -9.9528 < 2.2e-16 ***
## year_of_observation1985        -0.2379207  0.0262729 -9.0557 < 2.2e-16 ***
## year_of_observation1986        -0.2011135  0.0336844 -5.9705 3.172e-09 ***
## year_of_observation1987        -0.2472163  0.0383805 -6.4412 1.757e-10 ***
## year_of_observation1988        -0.2776718  0.0480131 -5.7832 9.499e-09 ***
## year_of_observation1989        -0.3510217  0.0535312 -6.5573 8.349e-11 ***
## year_of_observation1990        -0.3606396  0.0586036 -6.1539 1.051e-09 ***
## year_of_observation1991        -0.3984518  0.0627482 -6.3500 3.124e-10 ***
## year_of_observation1992        -0.4587240  0.0666755 -6.8799 9.945e-12 ***
## year_of_observation1993        -0.4755859  0.0678098 -7.0135 4.014e-12 ***
## year_of_observation1994        -0.5087505  0.0670278 -7.5901 6.711e-14 ***
## year_of_observation1995        -0.5089585  0.0725202 -7.0182 3.888e-12 ***
## year_of_observation1996        -0.5600408  0.0752837 -7.4391 2.014e-13 ***
## year_of_observation1997        -0.5828591  0.0770172 -7.5679 7.898e-14 ***
## year_of_observation1998        -0.6356498  0.0772402 -8.2295 5.183e-16 ***
## year_of_observation1999        -0.6510148  0.0789708 -8.2437 4.634e-16 ***
## year_of_observation2000        -0.6816718  0.0792438 -8.6022 < 2.2e-16 ***
## year_of_observation2001        -0.6497400  0.0831842 -7.8109 1.301e-14 ***
## year_of_observation2002        -0.6105826  0.0810339 -7.5349 1.005e-13 ***
## year_of_observation2003        -0.6127145  0.0835637 -7.3323 4.329e-13 ***
## year_of_observation2004        -0.6504450  0.0875255 -7.4315 2.127e-13 ***
## factor(blood_alcohol_limit)0.1  0.0048883  0.0176342  0.2772  0.781673    
## per_se_laws                    -0.0554823  0.0163076 -3.4022  0.000692 ***
## seatbelt2                       0.0046718  0.0162378  0.2877  0.773621    
## seatbelt1                      -0.0413988  0.0248524 -1.6658  0.096035 .  
## speed_limit_70plus              0.0727041  0.0222040  3.2744  0.001091 ** 
## graduated_drivers_license_law  -0.0311878  0.0195683 -1.5938  0.111265    
## pct_population_14_to_24         0.0192910  0.0105815  1.8231  0.068556 .  
## unemployment_rate_log          -0.1940177  0.0235478 -8.2393 4.799e-16 ***
## vehicle_miles_per_capita_log    0.6678486  0.1374350  4.8594 1.345e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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.

  Random Effect Model Assumptions: - the idiosyncratic error uit should be uncorrelated with each explanatory variable across all time period - the error uit are homoskedastic and serially uncorrelated across all time pericod - the unobservbed effect term ai is independent of all explanatory variables in all the time periods in the model

- 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.

```
# re.model <- plm(
#   log(total_fatalities_rate) ~ year_of_observation
#     + factor(blood_alcohol_limit)
#     + per_se_laws
#     + seatbelt
#     + speed_limit_70plus
#     + graduated_drivers_license_law
#     + pct_population_14_to_24
#     + log(unemployment_rate)
#     + vehicle_miles_per_capita,
#   data = df,
#   index = c("state"),
#   model = "random"
# )

# summary(re.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?

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?

  We have collected the United States Motor Vehicle Miles Traveled Total (Millions) data from the US Department of Transformation. This data is available at here [https://www.fhwa.dot.gov/policyinformation/statistics/2021/vm202.cfm].

  We have downloaded the data that is compiled by Bloomberg (under ticker: VMTDVCLE Index) and saved it in the file `data/VMTDVCLE.csv`.

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.

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?