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)
#desc
```

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

2.1 (5 points) Load the data and produce useful features.

```
# For the fractions, we are taking the majority as a speed limit
# We skipped year_of_observation since there a year column which aligns with dx
df <- data %>%
  mutate(speed_limit = ifelse(sl55 >= 0.5, '55',
                       ifelse(sl65 >= 0.5, '65',
                       ifelse(sl70 >= 0.5, '70',
                       ifelse(sl75 >= 0.5, '75',
                       ifelse(slnone >= 0.5, 'none', '0')
                       ))))) %>%
           mutate(speed_limit=factor(speed_limit,
                  levels=c('55', '65', '70', '75', 'none')),
         blood_alcohol_limit_10 = ifelse(bac10 >= 0.5, 1, 0),
         blood_alcohol_limit_08 = ifelse(bac08 >= 0.5, 1, 0)) %>%
  mutate(bac=ifelse(blood_alcohol_limit_10==1, '10',
              ifelse(blood_alcohol_limit_08==1, '8', 'none'))) %>%
  mutate(bac=factor(bac, levels=c('none', '10', '8'))) %>%
  select(!c((sl55:slnone), (d80:d04), bac10, bac08)) %>% # Excluding
  rename(minimum_drinking_age = minage, zero_tolerance_law = zerotol,
         graduated_drivers_license_law = gdl, per_se_law = perse,
         total_fatalities = totfat, nighttime_fatalities = nghtfat,
         weekend_fatalities = wkndfat, total_fatalities_per_100M_miles = totfatpvm,
         nighttime_fatalities_per_100M_miles = nghtfatpvm,
         weekend_fatalities_per_100M_miles = wkndfatpvm,
         state_population =  statepop, total_fatalities_rate = totfatrte,
```

```r
         nighttime_fatalities_rate = nghtfatrte,
         weekend_fatalities_rate = wkndfatrte,
         vehicle_miles_traveled = vehicmiles, unemployment_rate = unem,
         population_aged_14_to_24_rate = perc14_24,
         speed_limit_70_plus = sl70plus,
         seat_belt = seatbelt,
         primary_seatbelt_law = sbprim, secondary_seatbelt_law = sbsecon,
         miles_driven_per_capita = vehicmilespc) %>%
  mutate(speed_limit_70_plus =
           ifelse(speed_limit_70_plus>0.5, 1, 0)
           ) %>%
  mutate(seat_belt_law =
           ifelse(seat_belt==0, 'none',
           ifelse(seat_belt==2, 'secondary',
           ifelse(seat_belt==1, 'primary', 'na')))) %>%
  mutate(seat_belt_law=factor(seat_belt_law,
                              levels=c('none', 'secondary', 'primary')),
         ) %>%
  mutate(per_se_law=round(per_se_law, 0)) %>%
  mutate(per_se_law=factor(per_se_law, levels=c(0, 1))) %>%
  mutate(log_total_fatalities_rate = log10(total_fatalities_rate))


# Adding states to the dataframe
state_df <- data.frame("index" = 1:51,
                       "state_name" = sort(c(state.name, "District of Columbia")))
main_df <- merge(df, state_df, by.x = 'state', by.y ='index')

pdata <- pdata.frame(main_df, index=c("state", "year"))
head(main_df)
```

2.1.0.1 Status complete: TODO Check for completeness

```
##   state year seat_belt minimum_drinking_age zero_tolerance_law
## 1     1 1980         0                   18                  0
## 2     1 1981         0                   18                  0
## 3     1 1982         0                   18                  0
## 4     1 1983         0                   18                  0
## 5     1 1984         0                   18                  0
## 6     1 1985         0                   20                  0
##   graduated_drivers_license_law per_se_law total_fatalities
## 1                             0          0              940
## 2                             0          0              933
## 3                             0          0              839
## 4                             0          0              930
## 5                             0          0              932
## 6                             0          0              882
##   nighttime_fatalities weekend_fatalities total_fatalities_per_100M_miles
## 1                  422                236                            3.20
## 2                  434                248                            3.35
## 3                  376                224                            2.81
## 4                  397                223                            3.00
## 5                  421                237                            2.83
```

```
## 6                 358                224                            2.51
##   nighttime_fatalities_per_100M_miles weekend_fatalities_per_100M_miles
## 1                               1.437                             0.803
## 2                               1.558                             0.890
## 3                               1.259                             0.750
## 4                               1.281                             0.719
## 5                               1.278                             0.720
## 6                               1.019                             0.637
##   state_population total_fatalities_rate nighttime_fatalities_rate
## 1          3893888                 24.14                     10.84
## 2          3918520                 24.07                     11.08
## 3          3925218                 21.37                      9.58
## 4          3934109                 23.64                     10.09
## 5          3951834                 23.58                     10.65
## 6          3972527                 22.20                      9.01
##   weekend_fatalities_rate vehicle_miles_traveled unemployment_rate
## 1                    6.06               29.37500               8.8
## 2                    6.33               27.85200              10.7
## 3                    5.71               29.85765              14.4
## 4                    5.67               31.00000              13.7
## 5                    6.00               32.93286              11.1
## 6                    5.64               35.13944               8.9
##   population_aged_14_to_24_rate speed_limit_70_plus primary_seatbelt_law
## 1                          18.9                   0                    0
## 2                          18.7                   0                    0
## 3                          18.4                   0                    0
## 4                          18.0                   0                    0
## 5                          17.6                   0                    0
## 6                          17.3                   0                    0
##   secondary_seatbelt_law miles_driven_per_capita speed_limit
## 1                      0                7543.874          55
## 2                      0                7107.785          55
## 3                      0                7606.622          55
## 4                      0                7879.802          55
## 5                      0                8333.562          55
## 6                      0                8845.614          55
##   blood_alcohol_limit_10 blood_alcohol_limit_08 bac seat_belt_law
## 1                      1                      0  10          none
## 2                      1                      0  10          none
## 3                      1                      0  10          none
## 4                      1                      0  10          none
## 5                      1                      0  10          none
## 6                      1                      0  10          none
##   log_total_fatalities_rate state_name
## 1                  1.382737    Alabama
## 2                  1.381476    Alabama
## 3                  1.329805    Alabama
## 4                  1.373647    Alabama
## 5                  1.372544    Alabama
## 6                  1.346353    Alabama
```

2.2 (5 points) Provide a description of the basic structure of the dataset.

2.2.0.1 Status not-started

TODO TODO

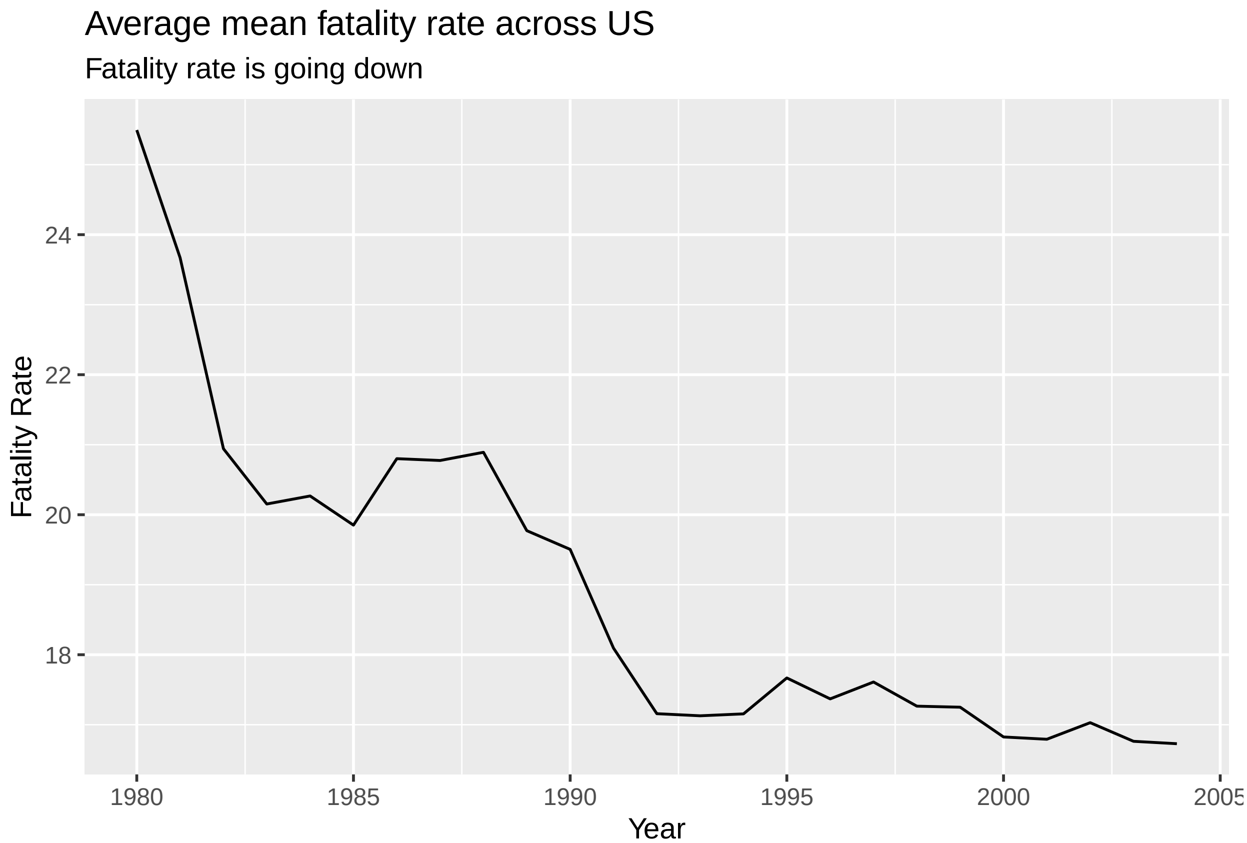
2.3 (20 points) Conduct a very thorough EDA

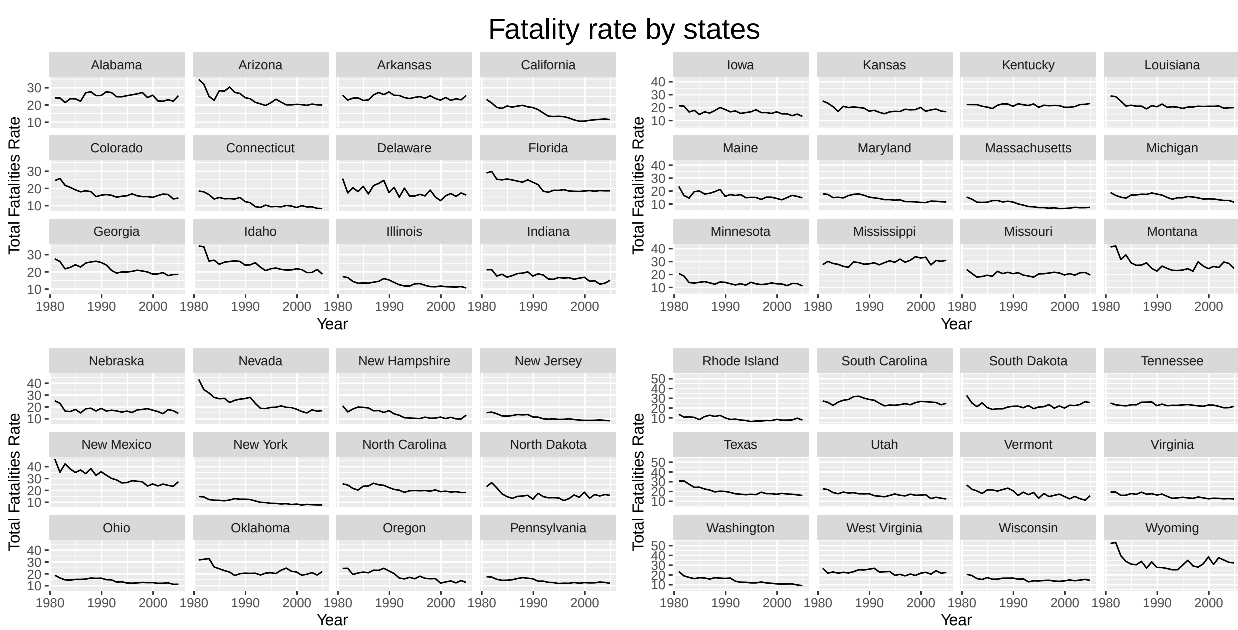
2.3.0.1 Status Complete: Check for completeness

TODO TODO

2.3.1 Average and state wide trends

2.3.1.1 Status Incomplete: Refer to graphs and add descriptive text. Does the dependent variable require a transformation?



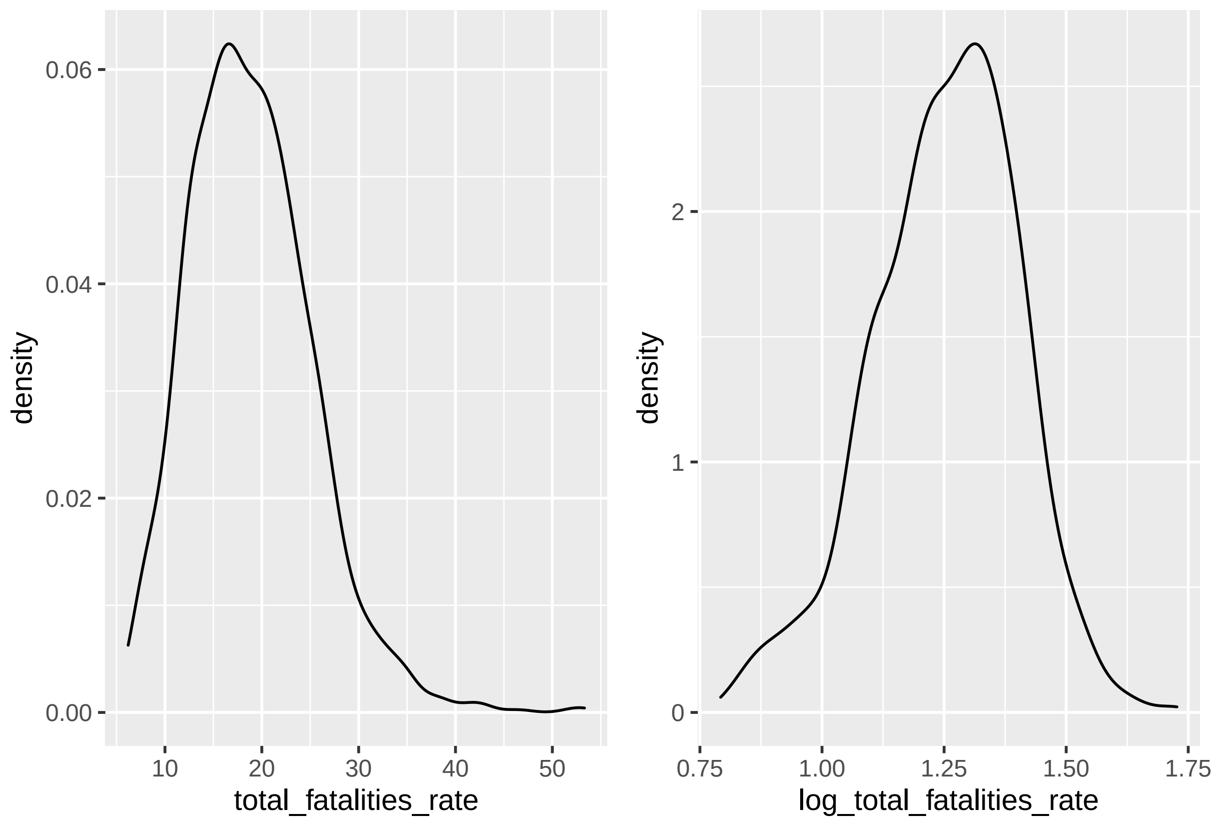


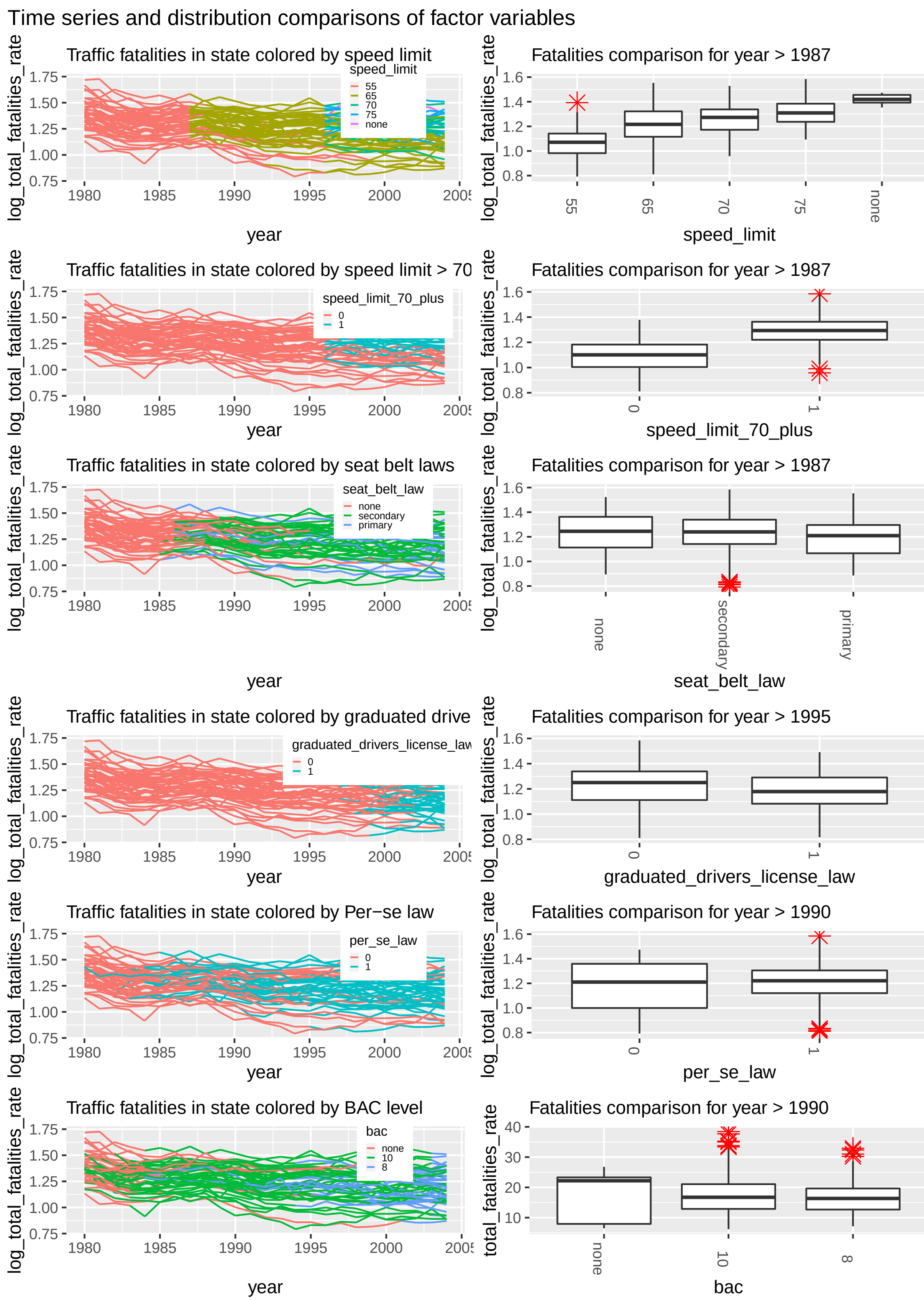
'For most states, fatality rates go down over the years, but some states like Alabama and Arkansas do not show many changes. Surprisingly, Mississippi has an increase in the fatality rate.'

```
without.tran <- main_df %>% ggplot(aes(x=total_fatalities_rate)) +
  geom_density()
```

```
with.tran <- main_df %>% ggplot(aes(x=log_total_fatalities_rate)) +
  geom_density()

without.tran | with.tran
```



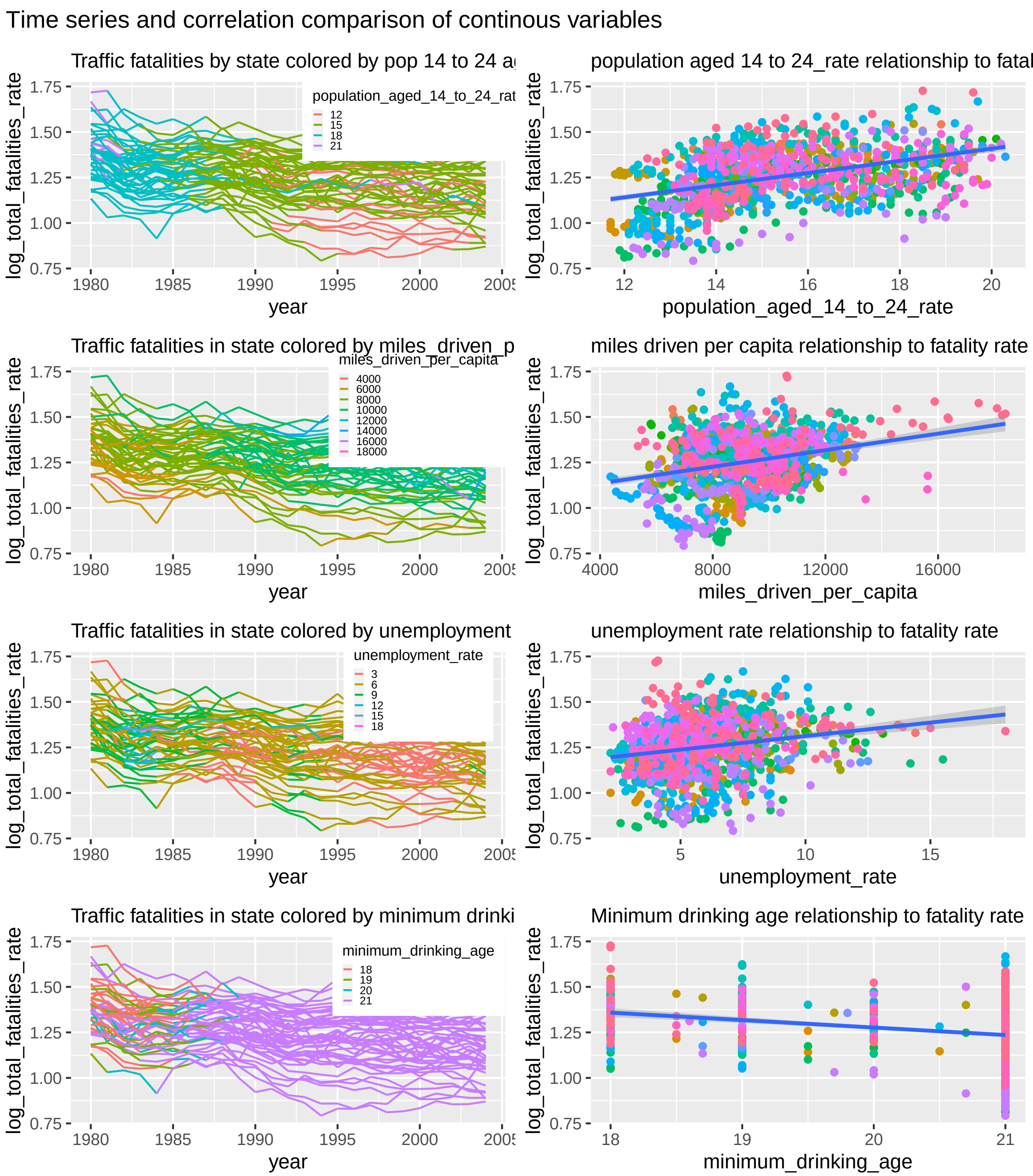

Time series and distribution comparisons of factor variables
Traffic fatalities in state colored by speed limit
speed_limit
55
65
70
75
none
Fatalities comparison for year > 1987
speed_limit
Traffic fatalities in state colored by speed limit > 70
speed_limit_70_plus
Fatalities comparison for year > 1987
speed_limit_70_plus
Traffic fatalities in state colored by seat belt laws
seat_belt_law
none
secondary
primary
Fatalities comparison for year > 1987
seat_belt_law
Traffic fatalities in state colored by graduated drive
graduated_drivers_license_law
Fatalities comparison for year > 1995
graduated_drivers_license_law
Traffic fatalities in state colored by Per−se law
per_se_law
Fatalities comparison for year > 1990
per_se_law
Traffic fatalities in state colored by BAC level
bac
none
10
8
Fatalities comparison for year > 1990
bac
log_total_fatalities_rate
total_fatalities_rate
year

2.3.2 Description of variables

2.3.2.1 Status incomplete: TODO: Add figur enumber references to each graph

2.3.2.2 Factor variables

- The highway speed limit was uniformly 55mph across all states before 1987. Since then, different states have adopted different speed limits. Especially in 1997, there was a signifcant increase in highway speeds across multiple states.The box plot compares the fatality rate across different speed limits filtered for years greater than 1987. We see that increasing speed limits are associated with increased fatality rates. As there are states with no speed limit, this variables has been treated as a factor.
- Now thresholding speed limits for greater or lower than 70mph shows a similar pattern of higher speeds associated with a higher fatality rate.
- Seat belts started to become mandatory starting mid to late 80s and today there is only one state which does not have it as mandatory. Primary laws are the strictest and allow police to ticket drivers and passengers who are not wearing a proper safety restraint, even if that is the only traffic violation they are committing. Secondary seat belt laws, on the other hand, do not grant law enforcement officials the right to ticket drivers or passengers for failing to wear a safety restraint unless another traffic violation has occurred. There are 15 states with secondary seat belt laws. Source: https://www.cooper-law-firm.com/what-is-the-difference-between-primary-and-secondary-seat-belt-laws/.
- The graduated drivers licence law was started to be introduced in the late 90's. The box plot, which has been filtered for years greater than 1995, suggests that even for that time frame, there is a reduction in fatality rate between the two groups.
- Some states had Per-Se laws before the start of the data in 1980 and some still did not have Per-Se laws in 2004. There is a gradual increase in the adoption of the law from 1980 to about the 2000s. Surorisingly, there is an increase in fatality rates in comparison of data with PerSe law as compared to without.
- Most states had adopted a BAC limit by the mid 80s with two states choosing a limit only in 2002.



2.3.2.3 Description of continous variables

- There is a decrease in the percentage of 14 to 24 year olds in the population over time. This is correlated to the decrease in the fatalities during that time period.
- Miles driven per capita also has a positive relationship with fatality rate. An increase in miles driven is associated with an increase in fatality rate
- There is an increase in failure rate with an increase in unemployment rate.
- The minimum drinking age has been 21 in most states since the late 80s. There is a general decrease

in fatality rate with an increased minimum drinking age but there also has been a general decrease in fatality rates during the time period when the age limits were changed.

```r
# traffic laws that we are exploring are seat_belt, minimum_drinking_age,
# zero_tolerance_law, graduated_drivers_license_law, per_se_law, speed_limit,
# speed_limit_70_plus, primary_seatbelt_law, secondary_seatbelt_law,
# blood_alcohol_limit_10, blood_alcohol_limit_08
```

3 (15 points) Preliminary Model

```r
#  Pooled OLS model
pooled_ols <- plm(log_total_fatalities_rate ~ year, data = pdata,
                  index = c("state", "year"),
                  effect = "individual", model = "pooling")
summary(pooled_ols)
```

3.0.0.1 Status complete: TODO Check if anything is missing

```
## Pooling Model
##
## Call:
## plm(formula = log_total_fatalities_rate ~ year, data = pdata,
##     effect = "individual", model = "pooling", index = c("state",
##         "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.4183307 -0.0961280  0.0043659  0.1008470  0.3770988
##
## Coefficients:
##              Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)  1.387907   0.020400 68.0351 < 2.2e-16 ***
## year1981    -0.034213   0.028850 -1.1859 0.2359037
## year1982    -0.086671   0.028850 -3.0042 0.0027191 **
## year1983    -0.102159   0.028850 -3.5411 0.0004141 ***
## year1984    -0.098087   0.028850 -3.3999 0.0006966 ***
## year1985    -0.105539   0.028850 -3.6582 0.0002652 ***
## year1986    -0.085474   0.028850 -2.9627 0.0031107 **
## year1987    -0.086297   0.028850 -2.9912 0.0028363 **
## year1988    -0.082018   0.028850 -2.8429 0.0045473 **
## year1989    -0.107768   0.028850 -3.7355 0.0001963 ***
## year1990    -0.116324   0.028850 -4.0320 5.886e-05 ***
## year1991    -0.149274   0.028850 -5.1742 2.690e-07 ***
## year1992    -0.174714   0.028850 -6.0560 1.875e-09 ***
## year1993    -0.174832   0.028850 -6.0601 1.830e-09 ***
## year1994    -0.177185   0.028850 -6.1416 1.116e-09 ***
## year1995    -0.167169   0.028850 -5.7945 8.794e-09 ***
## year1996    -0.173498   0.028850 -6.0138 2.416e-09 ***
```

```
## year1997    -0.167622   0.028850 -5.8102 8.028e-09 ***
## year1998    -0.177863   0.028850 -6.1651 9.666e-10 ***
## year1999    -0.180014   0.028850 -6.2397 6.108e-10 ***
## year2000    -0.189762   0.028850 -6.5776 7.181e-11 ***
## year2001    -0.189010   0.028850 -6.5515 8.500e-11 ***
## year2002    -0.185324   0.028850 -6.4238 1.927e-10 ***
## year2003    -0.190996   0.028850 -6.6204 5.437e-11 ***
## year2004    -0.194794   0.028850 -6.7520 2.286e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    26.854
## Residual Sum of Squares: 23.471
## R-Squared:      0.12598
## Adj. R-Squared: 0.10813
## F-statistic: 7.05692 on 24 and 1175 DF, p-value: < 2.22e-16
```

- 'Starting from a linear model will give us an easy and intuitively clearer overall pattern over the years via the coefficients on the year dummy variables (or instead by adding group ids as dummy variables) using controlling for any variable of interest, if any. Later on, when we perform panel data analysis, we can compare it against the linear model, justify the result and keep building the intuition on top of the linear model.'

- 'This model explains the log of fatality rate over different years compared to the baseline year, which is 1980. All coefficients except for the 1981 year are statistically significant. From this model, we learn that in all years following the baseline up to 2004, the fatality rate goes down compared to the baseline year 1980. The decline increases as the year increases but not consistently, as seen, for example, between 1986 - 1988, where the fatality rate increases compared to previous years. However, it returns to the decline track/trend in 1989. The higher, in absolute value, coefficients over the years means that driving becomes safer as the fatality rate keeps decreasing. For example, the fatality rate, which was 25.5 as of 1980, became 19.5 in 1990, 16.8 in 2000, and 16.7 in 2004, showing that driving over the years has become safer. In other words, around 8.8 fewer people are predicted to get traffic fatalities out of 100,000 people in 2004 than in 1980.'

- 'We are ignoring unobserved Heterogeneity and the group structure by taking each entry as a separate observation. Because of that, residuals generally correlate across time and have heteroskedasticity across and/or within groups. Heteroscedastic residuals are a violation of the OLS Homoscedasticity assumption, which will make it difficult to trust the standard error. As a result, the confidence interval can not be trusted as it can be too wide or narrow. Also, the independence assumption (no autocorrelation) is violated since we did not accommodate the lag/trend component, which makes the OLS estimates to be unreliable; in other words, our OLS estimator is not the Best Linear Unbiased Estimator.'

4 (15 points) Expanded Model

4.0.0.1 Status complete: TODO Check if anything is missing

4.1 Transformation(treatment) of variables

- As described in a previous section, a log transformation was performed on the response variable to make the distributionb closer to normal.

- States where highway speeds were made over 70mph during the middle of the year contained a fractional value for that year. The fraction was threshold at 0.5 to make this a binary variable. As there is no meaningful interpretation for this variable as a continuous value, this threshold was necessary to convert it into a factor.
- A state can have a primary, secondary or no seat belt laws. The seat_belt_law reflects these three factors combined into one variable.
- Graduated drivers licence law and per-se law parameters both have fractional value for years where the law was implemented mid-year. The fractions was threshold at 0.5 to make this a binary variable. As there is no meaningful interpretation for these variables as a continuous value, this threshold was necessary to convert them into a factor.
- Although Blood alcohol content (BAC) levels of 8PPM or 10PPM lends itself to a numeric interpretation, there is no numeric value associated with no BAC limit. For this reason BAC has been treated as a factor variable with levels none, 10 and 8.
- None of the continuous variables, namely, unemployment rate, miles drive per-capita, rate of 14 to 24 aged people in population required any transformation. This can be seen in the correlation plots in Figure y where their relationhip to log fatality rates appear linear.

4.1.1 Pooled OLS model creation

```
expanded.ols.data <- main_df %>% select(c(log_total_fatalities_rate, bac, per_se_law,
                   seat_belt_law, graduated_drivers_license_law,
                   population_aged_14_to_24_rate, minimum_drinking_age,
                   unemployment_rate, speed_limit_70_plus,
                   miles_driven_per_capita, year, state
                   ))

main_p <- pdata.frame(expanded.ols.data, index=c("state", "year"))
expanded.ols <- plm(log_total_fatalities_rate ~  year + bac +
                      population_aged_14_to_24_rate + miles_driven_per_capita +
                      unemployment_rate + speed_limit_70_plus + per_se_law +
                      seat_belt_law + graduated_drivers_license_law,
                    data = main_p,
                  index = c("state", "year"),
                  effect = "individual", model = "pooling")

summary(expanded.ols)
```

```
## Pooling Model
##
## Call:
## plm(formula = log_total_fatalities_rate ~ year + bac + population_aged_14_to_24_rate +
##     miles_driven_per_capita + unemployment_rate + speed_limit_70_plus +
##     per_se_law + seat_belt_law + graduated_drivers_license_law,
##     data = main_p, effect = "individual", model = "pooling",
##     index = c("state", "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.34758836 -0.05686774  0.00052907  0.06279643  0.25888532
##
```

```
## Coefficients:
##                                Estimate  Std. Error  t-value  Pr(>|t|)
## (Intercept)                   6.5776e-01  5.5569e-02  11.8367 < 2.2e-16 ***
## year1981                     -4.0821e-02  1.8599e-02  -2.1948   0.02838 *
## year1982                     -1.3080e-01  1.9183e-02  -6.8182 1.477e-11 ***
## year1983                     -1.5099e-01  1.9581e-02  -7.7113 2.659e-14 ***
## year1984                     -1.2006e-01  1.9720e-02  -6.0885 1.545e-09 ***
## year1985                     -1.3316e-01  2.0105e-02  -6.6233 5.354e-11 ***
## year1986                     -1.2207e-01  2.0907e-02  -5.8386 6.819e-09 ***
## year1987                     -1.3467e-01  2.1723e-02  -6.1997 7.839e-10 ***
## year1988                     -1.3907e-01  2.2769e-02  -6.1076 1.376e-09 ***
## year1989                     -1.7282e-01  2.3648e-02  -7.3079 5.026e-13 ***
## year1990                     -1.9557e-01  2.4190e-02  -8.0846 1.550e-15 ***
## year1991                     -2.4396e-01  2.4741e-02  -9.8605 < 2.2e-16 ***
## year1992                     -2.8991e-01  2.5218e-02 -11.4964 < 2.2e-16 ***
## year1993                     -2.8707e-01  2.5525e-02 -11.2464 < 2.2e-16 ***
## year1994                     -2.8368e-01  2.5976e-02 -10.9207 < 2.2e-16 ***
## year1995                     -2.7658e-01  2.6580e-02 -10.4057 < 2.2e-16 ***
## year1996                     -3.3148e-01  2.7527e-02 -12.0418 < 2.2e-16 ***
## year1997                     -3.4014e-01  2.7968e-02 -12.1615 < 2.2e-16 ***
## year1998                     -3.6607e-01  2.8402e-02 -12.8887 < 2.2e-16 ***
## year1999                     -3.7110e-01  2.8800e-02 -12.8855 < 2.2e-16 ***
## year2000                     -3.8060e-01  2.9269e-02 -13.0035 < 2.2e-16 ***
## year2001                     -4.0014e-01  2.9859e-02 -13.4009 < 2.2e-16 ***
## year2002                     -4.1553e-01  3.0151e-02 -13.7815 < 2.2e-16 ***
## year2003                     -4.2378e-01  3.0308e-02 -13.9825 < 2.2e-16 ***
## year2004                     -4.2116e-01  3.0978e-02 -13.5954 < 2.2e-16 ***
## bac10                        -9.9381e-03  8.7474e-03  -1.1361   0.25614
## bac8                         -2.6054e-02  1.1804e-02  -2.2073   0.02749 *
## population_aged_14_to_24_rate  6.8653e-03  2.7564e-03   2.4906   0.01289 *
## miles_driven_per_capita       6.8781e-05  2.1301e-06  32.2903 < 2.2e-16 ***
## unemployment_rate             1.6720e-02  1.7501e-03   9.5533 < 2.2e-16 ***
## speed_limit_70_plus           9.6254e-02  9.7589e-03   9.8632 < 2.2e-16 ***
## per_se_law1                  -6.8721e-03  6.6160e-03  -1.0387   0.29915
## seat_belt_lawsecondary        9.1036e-03  9.6452e-03   0.9438   0.34545
## seat_belt_lawprimary         -8.6245e-04  1.1029e-02  -0.0782   0.93768
## graduated_drivers_license_law -1.2209e-03  1.1840e-02  -0.1031   0.91789
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    26.854
## Residual Sum of Squares: 9.6296
## R-Squared:      0.64141
## Adj. R-Squared: 0.63095
## F-statistic: 61.29 on 34 and 1165 DF, p-value: < 2.22e-16
```

4.2 Interpretation of results

4.2.1 How are the blood alcohol variables defined? Interpret the coefficients that you estimate for this concept.

- In an earlier section we have defined our treatment of the BAC variable as a factor with levels 'none', '10' and '8'. In this model we note that the base level is no blood alcohol limit. We note that bac value of 0.08% is statistically significant and 0.1% is not statistically significant. The model suggests that setting a blood alcohol limit of 0.1% is associated with a `0.98` times decrease in fatality rate as compared to no BAC limit. The model suggests that a setting a blood alcohol limit of 0.08% is associated with a `0.94` times decrease in fatality rate as compared to no BAC limit.

4.2.2 Do *per se laws* have a negative effect on the fatality rate?

- We note that per-se law is not a statistically significant parameter. The model suggests that havng a per-se law is associated with a `0.98` times decrease in fatality rate as compared to not having per-se law.

4.2.3 Does having a primary seat belt law reduce fatality rates?

- We note that the seat belt law factors are not statistically significant. We also note that the implementation of the primary law leads to a `0.998` times decrease in fatality rate which is practically insignificant.

5 (15 points) State-Level Fixed Effects

5.0.0.1 Status complete: TODO Check if anything is missing

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

Model estimation for a fixed effect(within) model.

```r
expanded.within <- plm(log_total_fatalities_rate ~  bac + year  +
                      population_aged_14_to_24_rate + miles_driven_per_capita +
                      unemployment_rate + speed_limit_70_plus + per_se_law +
                      seat_belt_law + graduated_drivers_license_law,
                    data = main_p,
                  index = c("state", "year"),
                  effect = "individual", model = "within")

summary(expanded.within)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log_total_fatalities_rate ~ bac + year + population_aged_14_to_24_rate +
##     miles_driven_per_capita + unemployment_rate + speed_limit_70_plus +
##     per_se_law + seat_belt_law + graduated_drivers_license_law,
##     data = main_p, effect = "individual", model = "within", index = c("state",
##         "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.1835098 -0.0223288  0.0020551  0.0227251  0.1279559
##
```

```
## Coefficients:
##                                Estimate  Std. Error  t-value  Pr(>|t|)
## bac10                        -1.1165e-02  4.9828e-03  -2.2407  0.025241 *
## bac8                         -1.6135e-02  7.1737e-03  -2.2492  0.024694 *
## year1981                     -2.5783e-02  7.9252e-03  -3.2533  0.001175 **
## year1982                     -5.0669e-02  8.4888e-03  -5.9689 3.204e-09 ***
## year1983                     -6.1384e-02  8.7911e-03  -6.9825 4.964e-12 ***
## year1984                     -7.8231e-02  8.9205e-03  -8.7698 < 2.2e-16 ***
## year1985                     -8.6982e-02  9.2924e-03  -9.3605 < 2.2e-16 ***
## year1986                     -6.6274e-02  9.9105e-03  -6.6872 3.585e-11 ***
## year1987                     -8.1819e-02  1.0632e-02  -7.6954 3.088e-14 ***
## year1988                     -9.1423e-02  1.1516e-02  -7.9387 4.941e-15 ***
## year1989                     -1.2221e-01  1.2261e-02  -9.9670 < 2.2e-16 ***
## year1990                     -1.2726e-01  1.2730e-02  -9.9971 < 2.2e-16 ***
## year1991                     -1.4334e-01  1.3067e-02 -10.9697 < 2.2e-16 ***
## year1992                     -1.6845e-01  1.3465e-02 -12.5101 < 2.2e-16 ***
## year1993                     -1.7549e-01  1.3715e-02 -12.7957 < 2.2e-16 ***
## year1994                     -1.8929e-01  1.4049e-02 -13.4739 < 2.2e-16 ***
## year1995                     -1.8602e-01  1.4479e-02 -12.8476 < 2.2e-16 ***
## year1996                     -2.0517e-01  1.5272e-02 -13.4345 < 2.2e-16 ***
## year1997                     -2.1113e-01  1.5630e-02 -13.5084 < 2.2e-16 ***
## year1998                     -2.3185e-01  1.5945e-02 -14.5404 < 2.2e-16 ***
## year1999                     -2.3797e-01  1.6125e-02 -14.7580 < 2.2e-16 ***
## year2000                     -2.5039e-01  1.6352e-02 -15.3122 < 2.2e-16 ***
## year2001                     -2.4326e-01  1.6634e-02 -14.6247 < 2.2e-16 ***
## year2002                     -2.2996e-01  1.6803e-02 -13.6862 < 2.2e-16 ***
## year2003                     -2.3124e-01  1.6894e-02 -13.6878 < 2.2e-16 ***
## year2004                     -2.4447e-01  1.7323e-02 -14.1121 < 2.2e-16 ***
## population_aged_14_to_24_rate  9.0309e-03  1.8219e-03   4.9569 8.270e-07 ***
## miles_driven_per_capita        2.6879e-05  2.1286e-06  12.6277 < 2.2e-16 ***
## unemployment_rate            -1.2761e-02  1.1601e-03 -11.0002 < 2.2e-16 ***
## speed_limit_70_plus            2.3830e-02  4.9909e-03   4.7748 2.038e-06 ***
## per_se_law1                  -2.3217e-02  4.2933e-03  -5.4078 7.801e-08 ***
## seat_belt_lawsecondary       -6.6982e-04  4.8314e-03  -0.1386  0.889762
## seat_belt_lawprimary         -2.0458e-02  6.5736e-03  -3.1122  0.001904 **
## graduated_drivers_license_law -7.3527e-03  5.6051e-03  -1.3118  0.189860
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    6.0213
## Residual Sum of Squares: 1.6682
## R-Squared:      0.72294
## Adj. R-Squared: 0.70287
## F-statistic: 85.8017 on 34 and 1118 DF, p-value: < 2.22e-16
```

5.2 Interpretation of model

```
##
## ============================================================
##                              log_total_fatalities_rate
##                            Pooled OLS            Within
##                               (1)                  (2)
## ------------------------------------------------------------
```

```
## bac10                        -0.010            -0.011**
##                             (0.009)            (0.005)
##
## bac8                        -0.026**           -0.016**
##                             (0.012)            (0.007)
##
## per_se_law1                  -0.007           -0.023***
##                             (0.007)            (0.004)
##
## seat_belt_lawprimary         -0.001           -0.020***
##                             (0.011)            (0.007)
##
## N                             1,200             1,200
## R2                            0.641             0.723
## ==============================================================
## Notes:                   ***Significant at the 1 percent level.
##                           **Significant at the 5 percent level.
##                            *Significant at the 10 percent level.
```

5.2.1 What do you estimate for coefficients on the blood alcohol variables? How do the coefficients on the blood alcohol variables change, if at all?

We also note that bac value of 0.1% and 0.08% are both statistically significant parameters.

- This model suggests that setting a blood alcohol limit of 0.1% is associated with a `0.97` times decrease in log fatality rate as compared to no BAC limit. Compared to the OLS model which showed that setting a blood alcohol limit of 0.1% is associated with a `0.98` times decrease in fatality rate as compared to no BAC limit.
- This model suggests that a setting a blood alcohol limit of 0.08% is associated with a `0.96` times decrease in fatality rate as compared to no BAC limit. Compared to the OLS model which showed that setting a blood alcohol limit of 0.08% is associated with a `0.96` times decrease in fatality rate as compared to no BAC limit.

5.2.2 What do you estimate for coefficients on per se laws? How do the coefficients on per se laws change, if at all?

We note that per-se law is a statistically significant parameter. This model suggests that having a per-se law is associated with a `0.95` times decrease in fatality rate as compared to not having per-se law. Compared to the OLS model which showed that havng a per-se law is associated with a `0.98` times decrease in fatality rate as compared to not having per-se law

5.2.3 What do you estimate for coefficients on primary seat-belt laws? How do the coefficients on primary seatbelt laws change, if at all?

In this model we note that seat belt primary law is a significant factor. This model suggests that having a primary seat belt law is associated with a `0.95` unit decrease in fatality rate as compared to not having primary seat belt law. Compared to the OLS model for which the parameter was not statistically significant and had a close to 0 value of `1`

5.3 Model assumptions

A fundamental assumption in an pooled OLS model is that the data is IID. Let's consider a dataset where a sample of a large population was collected on different years. It is unlikely that a particular individual sample is measured twice. In such a circumstance a pooled OLS model would be applicable. However in this dataset, the individual is the state and the same state is measured multiple times across years. This violates the assumption of IID in the pooled OLS. A fixed effect model is then expected to be a better model in this scenario. We perform a F-test between the pooled and the fixed effect model to check for fixed effects. The Null hypothesis is that there are no fixed effects and the alternate hypothesis is that there are fixed effects. We test against an alpha of 0.05

```
res <- pFtest(expanded.within, expanded.ols)
```

With a p-value of `0` less than 0.05, we reject the null hypothesis of no fixed effects. This means we should include state and/or time fixed effects in our model. Hence the fixed effect model is better for this scenario.

6 (10 points) Consider a Random Effects Model

6.0.0.1 Status incomplete

6.0.1 Random effect model assumptions

Assumption for a random effects model is $cov(x_{it}, a_i) = 0$, i.e. the fixed effect is uncorrelated to the independent variables at any time period.

6.0.2 Estimating the random effect model

```
expanded.re <- plm(log(log_total_fatalities_rate) ~  bac + year  +
                      population_aged_14_to_24_rate + miles_driven_per_capita +
                      unemployment_rate + speed_limit_70_plus + per_se_law +
                      seat_belt_law + graduated_drivers_license_law,
                    data = main_p,
                  index = c("state", "year"),
                  model = "random")

summary(expanded.re)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = log(log_total_fatalities_rate) ~ bac + year + population_aged_14_to_24_rate +
##     miles_driven_per_capita + unemployment_rate + speed_limit_70_plus +
##     per_se_law + seat_belt_law + graduated_drivers_license_law,
##     data = main_p, model = "random", index = c("state", "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Effects:
```

```
##                     var  std.dev share
## idiosyncratic 0.001039 0.032234 0.224
## individual    0.003590 0.059914 0.776
## theta: 0.893
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.1660054 -0.0184966  0.0024194  0.0204280  0.0990869
##
## Coefficients:
##                                Estimate  Std. Error  z-value  Pr(>|z|)
## (Intercept)                   2.6777e-02  3.3896e-02   0.7900  0.429542
## bac10                        -6.9546e-03  4.2398e-03  -1.6403  0.100940
## bac8                         -9.2990e-03  6.0977e-03  -1.5250  0.127259
## year1981                     -1.8385e-02  6.7789e-03  -2.7121  0.006685 **
## year1982                     -3.6454e-02  7.2529e-03  -5.0261 5.005e-07 ***
## year1983                     -4.4786e-02  7.5088e-03  -5.9645 2.454e-09 ***
## year1984                     -5.8153e-02  7.6184e-03  -7.6333 2.289e-14 ***
## year1985                     -6.4079e-02  7.9313e-03  -8.0792 6.520e-16 ***
## year1986                     -4.8582e-02  8.4542e-03  -5.7466 9.107e-09 ***
## year1987                     -6.1580e-02  9.0616e-03  -6.7958 1.077e-11 ***
## year1988                     -6.9287e-02  9.8064e-03  -7.0655 1.600e-12 ***
## year1989                     -9.3947e-02  1.0435e-02  -9.0034 < 2.2e-16 ***
## year1990                     -9.9176e-02  1.0828e-02  -9.1589 < 2.2e-16 ***
## year1991                     -1.1235e-01  1.1114e-02 -10.1091 < 2.2e-16 ***
## year1992                     -1.3484e-01  1.1448e-02 -11.7785 < 2.2e-16 ***
## year1993                     -1.4047e-01  1.1659e-02 -12.0487 < 2.2e-16 ***
## year1994                     -1.5331e-01  1.1941e-02 -12.8391 < 2.2e-16 ***
## year1995                     -1.5171e-01  1.2304e-02 -12.3305 < 2.2e-16 ***
## year1996                     -1.7129e-01  1.2974e-02 -13.2019 < 2.2e-16 ***
## year1997                     -1.7747e-01  1.3275e-02 -13.3688 < 2.2e-16 ***
## year1998                     -1.9703e-01  1.3540e-02 -14.5516 < 2.2e-16 ***
## year1999                     -2.0305e-01  1.3691e-02 -14.8304 < 2.2e-16 ***
## year2000                     -2.1366e-01  1.3885e-02 -15.3877 < 2.2e-16 ***
## year2001                     -2.0870e-01  1.4122e-02 -14.7788 < 2.2e-16 ***
## year2002                     -1.9994e-01  1.4262e-02 -14.0195 < 2.2e-16 ***
## year2003                     -2.0133e-01  1.4339e-02 -14.0401 < 2.2e-16 ***
## year2004                     -2.1385e-01  1.4704e-02 -14.5436 < 2.2e-16 ***
## population_aged_14_to_24_rate  9.6276e-03  1.5455e-03   6.2293 4.684e-10 ***
## miles_driven_per_capita       2.5922e-05  1.7696e-06  14.6491 < 2.2e-16 ***
## unemployment_rate            -9.7906e-03  9.8326e-04  -9.9573 < 2.2e-16 ***
## speed_limit_70_plus           2.7523e-02  4.2568e-03   6.4658 1.008e-10 ***
## per_se_law1                  -1.7036e-02  3.6422e-03  -4.6774 2.905e-06 ***
## seat_belt_lawsecondary        1.0944e-03  4.1220e-03   0.2655  0.790617
## seat_belt_lawprimary         -1.4173e-02  5.5867e-03  -2.5369  0.011185 *
## graduated_drivers_license_law -4.0096e-03  4.7872e-03  -0.8376  0.402278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    4.3716
## Residual Sum of Squares: 1.2721
## R-Squared:      0.70902
## Adj. R-Squared: 0.70052
## Chisq: 2838.65 on 34 DF, p-value: < 2.22e-16
```

6.0.3 We conduct a Hausman test for random vs. fixed effects using phtest.

We perform this test with an $\alpha = 0.05$

```
res <- phtest(expanded.within, expanded.re)
```

With a p-value of $3.6519259 \times 10^{-51}$ much less than α, we reject the null hypothesis that random effects are appropriate, suggesting that we should use the fixed models. The random effects model is not likely to be consistent in this case.

> TODO: Comment about consequences

7 (10 points) Model Forecasts

7.0.0.1 Status not-started 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.

```
# Source
# U.S. Federal Highway Administration, Vehicle Miles Traveled [TRFVOLUSM227NFWA],
#   retrieved from FRED, Federal Reserve Bank of St. Louis;
# https://fred.stlouisfed.org/series/TRFVOLUSM227NFWA, November 29, 2022.

miles.driven <-
  readr::read_csv("./data/TRFVOLUSM227NFWA.csv", show_col_types = FALSE) %>%
  rename(miles = TRFVOLUSM227NFWA, date = DATE)

largest.decrease <- 0
largest.increase <- 0
max <- -Inf
min <- Inf

# Comparing each month miles driven
for (i in 1:12) {
```

```
  val.18 <- miles.driven %>%
    filter(date == mdy(paste(i, 1, 2018, sep="/")))

  # Comparing 2018 with 2020 and 2021
  for (j in 2020:2021) {
    val.covid <- miles.driven %>%
      filter(date == mdy(paste(i, 1, j, sep="/")))

    diff <- val.covid$miles - val.18$miles
    if(diff < min) {
      largest.decrease <- mdy(paste(i, 1, j, sep="/"))
      min = diff
    }
    if(diff > max) {
      largest.increase <- mdy(paste(i, 1, j, sep="/"))
      max = diff
    }
  }
}

# Calculating percentage
prev <- miles.driven %>%
    filter(date == largest.decrease)
lower.percentage  <- min/prev$miles

prev <- miles.driven %>%
    filter(date == largest.increase)
higher.percentage <- max/prev$miles
```

'The largest decrease in driving was on April 2020 and it was lower by -64.14% amount.'
'The largest increase in driving was on January 2020 and it was higher by 6.18% amount.'

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?

8.0.1 Analysis of residuals

8.0.1.1 Status incomplete We test the residuals for serial correlation. The Null hypothesis for the test is that the residuals are not serially correlated and the alternate hypothesis is that the residuals are serially correlated. We check against an α of 0.05.

```
#
res <- pdwtest(expanded.within)
```

With a p-value of `\ensuremath{6.5724615\times 10^{-46}}` less than 0.05, we reject the null hypothesis of no serial correlation. This suggests that we must use robust standard errors for model parameters

Hetroskedasticity test

```r
pcdtest(expanded.within, test = "lm")
```

```
##
##  Breusch-Pagan LM test for cross-sectional dependence in panels
##
## data:  log_total_fatalities_rate ~ bac + year + population_aged_14_to_24_rate +     miles_driven_per
## chisq = 2826.9, df = 1128, p-value < 2.2e-16
## alternative hypothesis: cross-sectional dependence
```