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

```
# 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,
         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)))


# 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)
```

```
##   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 state_name
## 1                      1                      0  10          none    Alabama
## 2                      1                      0  10          none    Alabama
## 3                      1                      0  10          none    Alabama
## 4                      1                      0  10          none    Alabama
## 5                      1                      0  10          none    Alabama
## 6                      1                      0  10          none    Alabama
```

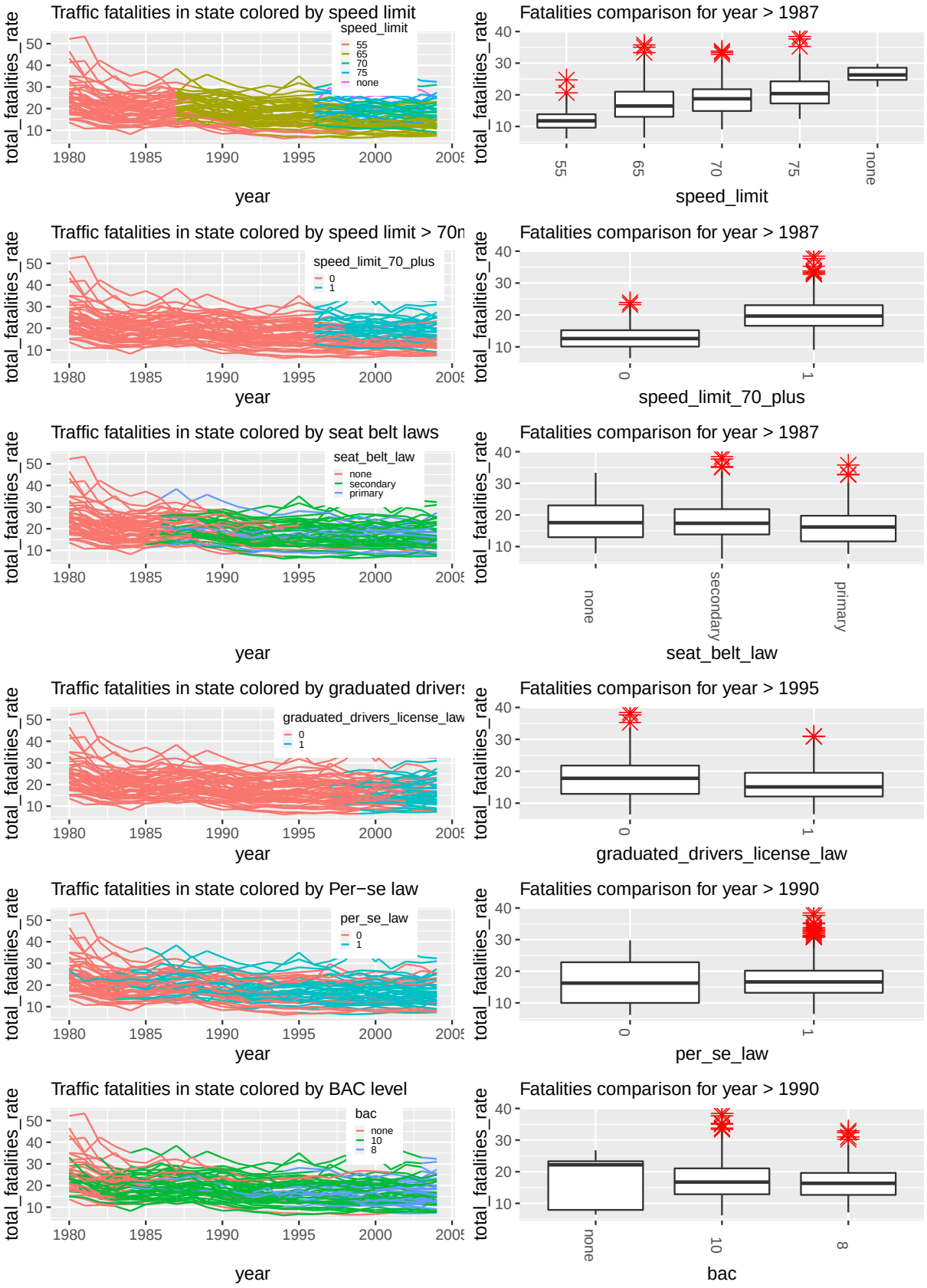
```
# Time series line plot
# Variable to plot as var_name
# Each line gets a color based on the group_name
# X-axis is always year
plots.ts.by.group <- function(df, var_name, group_name, subtitle='') {
plt <- df %>%
  mutate(!!group_name := factor(df[[group_name]])) %>%
    ggplot(aes_string(x='year', y=var_name, group='state')) +
    geom_line(aes_string(color=group_name)) +
  labs(subtitle=subtitle) +
  theme(legend.position = c(0.8, 0.8), legend.key.size = unit(0.2, "cm"),
        legend.text=element_text(size=rel(0.5)),
        legend.title=element_text(size=rel(0.7)))
  return((plt))
}


# Box plot by group
# Variable to plot as var_name
# Each group is plit by group_name
plots.box.by.group <- function(df, var_name, group_name, subtitle='')
  {
plt <- df %>% mutate(factored=factor(df[[group_name]])) %>%
    ggplot(aes_string(x='factored', y=var_name)) +
     geom_boxplot(outlier.colour="red", outlier.shape=8,  outlier.size=4) +
  theme(axis.text.x=element_text(angle=-90)) +
  labs(subtitle=subtitle) +
  xlab(group_name)
  return((plt))
}


# Time series line plot
# Variable to plot as var_name
# Each line gets a color based on the group_name
# X-axis is always year
plots.scatter.by.state <- function(df, var_name, subtitle='') {
plt <- df %>%
    ggplot(aes_string(x=var_name, y='total_fatalities_rate', group='state')) +
    theme(axis.text.x=element_text(angle=-90)) +
```

```
    geom_point(aes_string(color='state_name')) +
  labs(subtitle=subtitle) +
  theme(legend.position = 'none')
  return((plt))
}
```

2.1 Main Plots


Traffic fatalities in state colored by speed limit
speed_limit
55
65
70
75
none
total_fatalities_rate
year
Fatalities comparison for year > 1987
speed_limit
Traffic fatalities in state colored by speed limit > 70n
speed_limit_70_plus
0
1
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 drivers
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
Fatalities comparison for year > 1990
bac

2.1.1 Description of factor independent 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.

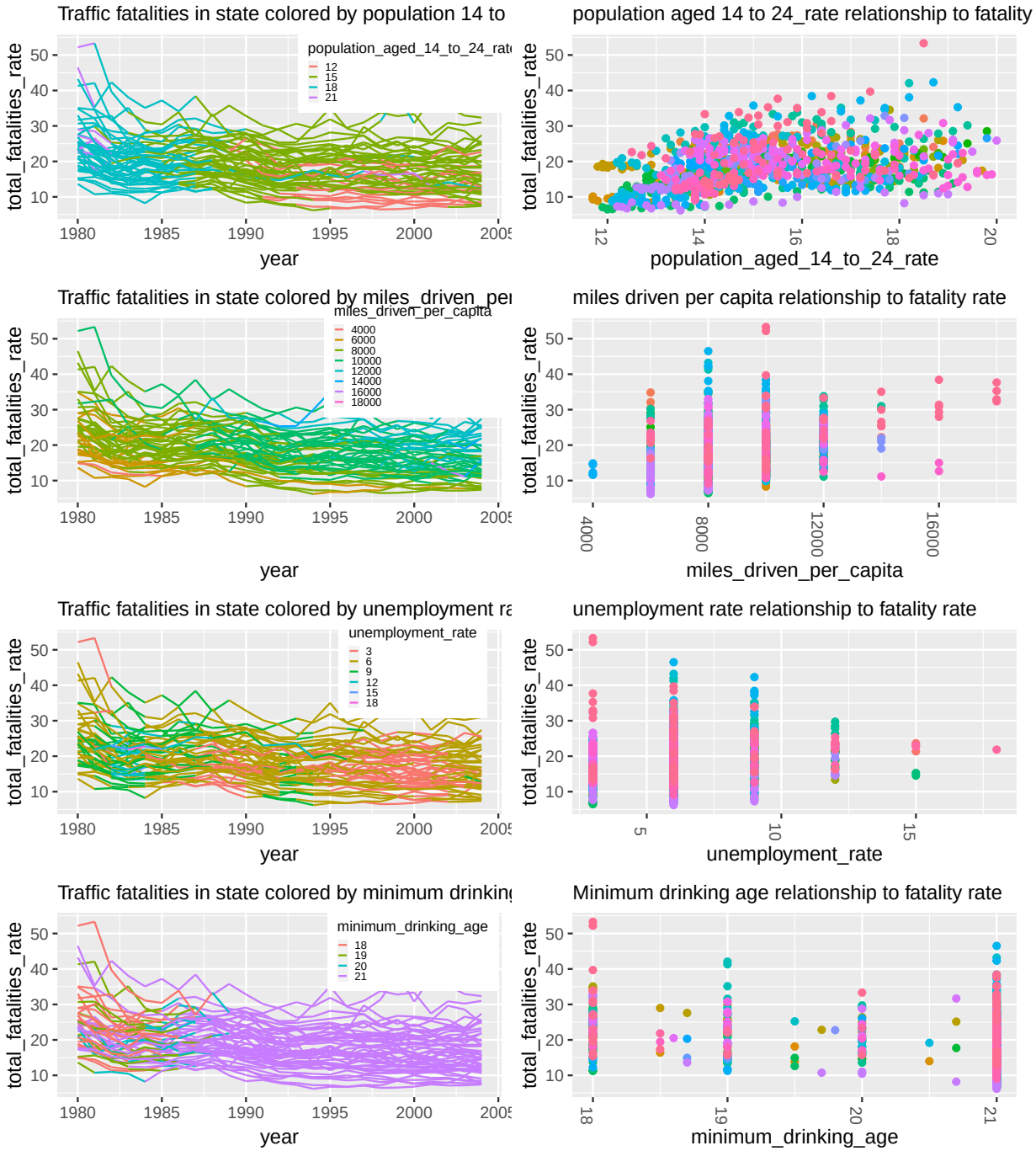
Seat belts started to become mandatory starting mid to late 80s and today there is only one state which does not have it as mandatory. Primary laws are the strictest and allow police to ticket drivers and passengers who are not wearing a proper safety restraint, even if that is the only traffic violation 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.

BAC level

2.1.2 Description of continous variables

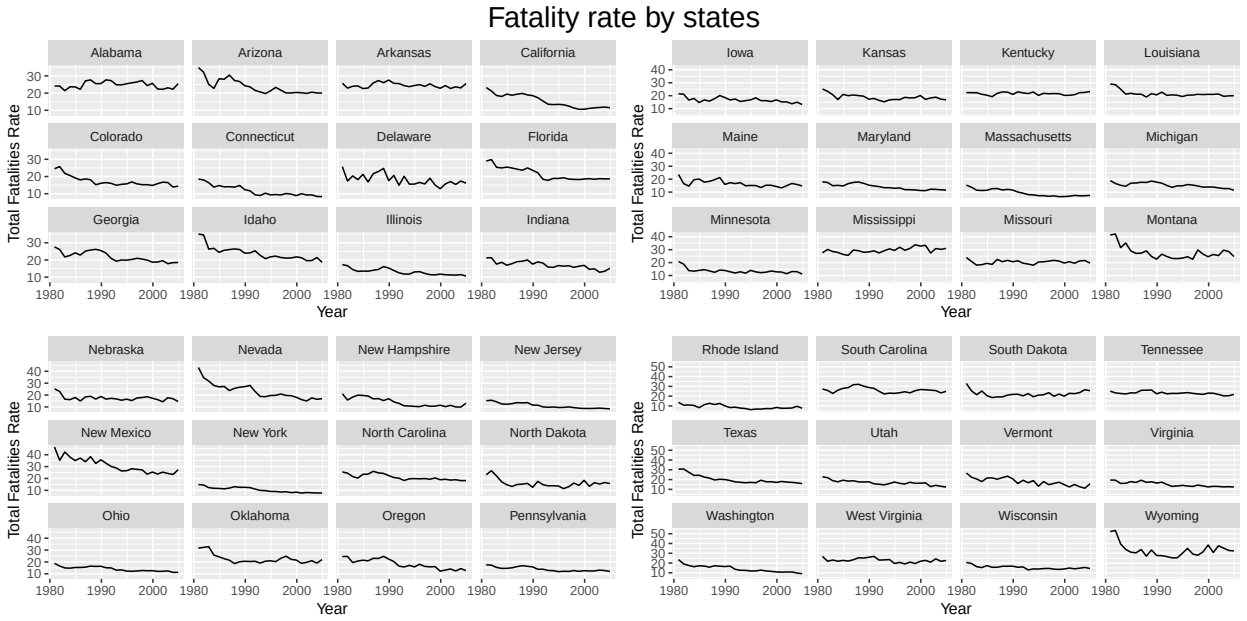


There is a decrease in the percetage of 14 to 24 year olds in the population over time. This is correlated to the decrease in the fatalities during that time period.

```
cut_point <- c(-1, 12, 24, 36, 48)
plots <- vector('list', 4)
```

```
for (i in 2:5) {
  plots[[i-1]] <- (pdata %>%
    filter(as.integer(state) > cut_point[i-1] & as.integer(state) <= cut_point[i]) %>%
    ggplot(aes(x = as.Date(year,"%Y"), y = total_fatalities_rate)) +
    geom_line() +
    facet_wrap(~ state_name, nrow = 3, ncol=4) +
    labs(x = "Year",  y = "Total Fatalities Rate") +
    theme(legend.position = "none"))
}

grid.arrange(plots[[1]], plots[[2]], plots[[3]], plots[[4]], nrow = 2, ncol = 2,
             top = textGrob("Fatality rate by states", gp=gpar(fontsize=20)))
```



> '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.'

```
# 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

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?

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?

```
expanded.ols.data <- main_df %>% select(c(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_name
                   ))

# expanded.ols <- lm(
#   total_fatalities_rate ~  factor(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 = expanded.ols.data)

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

expanded.ols <- plm(total_fatalities_rate ~  state + 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 = total_fatalities_rate ~ state + 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.
## -7.350830 -1.186176 -0.065354  1.103884 14.530220
##
## Coefficients:
##                                 Estimate  Std. Error  t-value  Pr(>|t|)
## (Intercept)                    1.4063e+01  1.9341e+00   7.2710 6.606e-13 ***
## state3                        -9.9565e-01  6.7873e-01  -1.4669 0.1426724
## state4                        -2.5705e-01  6.4682e-01  -0.3974 0.6911361
## state5                        -7.5456e+00  6.9644e-01 -10.8345 < 2.2e-16 ***
## state6                        -7.2916e+00  6.9814e-01 -10.4443 < 2.2e-16 ***
## state7                        -1.2150e+01  7.3475e-01 -16.5359 < 2.2e-16 ***
## state8                        -7.5213e+00  6.7802e-01 -11.0930 < 2.2e-16 ***
## state10                       -8.4416e-01  7.0876e-01  -1.1910 0.2338903
## state11                       -4.1616e+00  6.4065e-01  -6.4959 1.230e-10 ***
## state13                       -1.4775e+00  6.5033e-01  -2.2719 0.0232775 *
## state14                       -1.0018e+01  7.3789e-01 -13.5762 < 2.2e-16 ***
## state15                       -7.8249e+00  6.6281e-01 -11.8057 < 2.2e-16 ***
## state16                       -7.4778e+00  7.1819e-01 -10.4121 < 2.2e-16 ***
## state17                       -6.7010e+00  6.8081e-01  -9.8426 < 2.2e-16 ***
## state18                       -5.1301e+00  6.4436e-01  -7.9615 4.076e-15 ***
## state19                       -2.3429e+00  6.9604e-01  -3.3660 0.0007880 ***
## state20                       -7.4787e+00  6.7528e-01 -11.0750 < 2.2e-16 ***
## state21                       -1.1054e+01  7.1445e-01 -15.4717 < 2.2e-16 ***
## state22                       -1.6933e+01  7.4775e-01 -22.6459 < 2.2e-16 ***
## state23                       -8.8079e+00  6.7187e-01 -13.1095 < 2.2e-16 ***
## state24                       -1.1339e+01  6.9535e-01 -16.3076 < 2.2e-16 ***
## state25                        4.5302e+00  6.5638e-01   6.9019 8.493e-12 ***
## state26                       -3.7413e+00  6.5698e-01  -5.6947 1.571e-08 ***
## state27                        2.1529e+00  6.4333e-01   3.3465 0.0008452 ***
## state28                       -8.8302e+00  6.9429e-01 -12.7183 < 2.2e-16 ***
## state29                        3.4136e-01  7.1413e-01   0.4780 0.6327410
## state30                       -1.2070e+01  7.0818e-01 -17.0433 < 2.2e-16 ***
## state31                       -1.3157e+01  7.3420e-01 -17.9202 < 2.2e-16 ***
## state32                        7.1509e+00  6.4992e-01  11.0027 < 2.2e-16 ***
## state33                       -1.2886e+01  8.1765e-01 -15.7602 < 2.2e-16 ***
## state34                       -3.0253e+00  6.7844e-01  -4.4591 9.041e-06 ***
## state35                       -1.0925e+01  6.7951e-01 -16.0782 < 2.2e-16 ***
## state36                       -1.0414e+01  6.8730e-01 -15.1521 < 2.2e-16 ***
## state37                       -2.4676e+00  6.4836e-01  -3.8059 0.0001488 ***
## state38                       -3.8977e+00  6.7866e-01  -5.7432 1.190e-08 ***
## state39                       -1.0393e+01  7.3251e-01 -14.1888 < 2.2e-16 ***
## state40                       -1.6204e+01  7.5147e-01 -21.5631 < 2.2e-16 ***
```

```
## state41                       -9.6759e-01  6.7604e-01  -1.4312 0.1526324
## state42                       -5.2881e+00  6.6513e-01  -7.9505 4.435e-15 ***
## state43                       -3.2307e+00  6.6644e-01  -4.8476 1.422e-06 ***
## state44                       -4.7889e+00  6.6121e-01  -7.2427 8.070e-13 ***
## state45                       -1.0189e+01  7.3485e-01 -13.8659 < 2.2e-16 ***
## state46                       -8.1464e+00  6.5757e-01 -12.3887 < 2.2e-16 ***
## state47                       -1.1309e+01  6.6860e-01 -16.9139 < 2.2e-16 ***
## state48                       -9.0493e+00  6.7343e-01 -13.4376 < 2.2e-16 ***
## state49                        4.2030e-01  6.8695e-01   0.6118 0.5407661
## state50                       -1.0022e+01  6.7275e-01 -14.8971 < 2.2e-16 ***
## state51                        6.1474e+00  7.2314e-01   8.5010 < 2.2e-16 ***
## bac10                         -1.4242e+00  2.6001e-01  -5.4777 5.299e-08 ***
## bac8                          -1.9099e+00  3.7039e-01  -5.1563 2.967e-07 ***
## population_aged_14_to_24_rate  9.6311e-01  7.0698e-02  13.6228 < 2.2e-16 ***
## miles_driven_per_capita        2.9909e-04  1.0237e-04   2.9217 0.0035496 **
## unemployment_rate             -5.8884e-01  5.0849e-02 -11.5803 < 2.2e-16 ***
## speed_limit_70_plus           -1.1099e+00  2.3878e-01  -4.6482 3.738e-06 ***
## per_se_law1                   -1.4028e+00  2.3705e-01  -5.9175 4.317e-09 ***
## seat_belt_lawsecondary        -9.0641e-01  2.4836e-01  -3.6496 0.0002745 ***
## seat_belt_lawprimary          -1.8391e+00  3.4577e-01  -5.3187 1.257e-07 ***
## graduated_drivers_license_law -6.2026e-01  2.2720e-01  -2.7300 0.0064300 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    48612
## Residual Sum of Squares: 5540.3
## R-Squared:      0.88603
## Adj. R-Squared: 0.88034
## F-statistic: 155.757 on 57 and 1142 DF, p-value: < 2.22e-16
```

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?

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?

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.

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?