Lab 3: Panel Models

US Traffic Fatalities: 1980 - 2004

Contents


```
## Warning: package 'plm' was built under R version 4.1.3
```

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.

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

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

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`;

```r
data[data$sl55 == 0.5, 4:6] <- 0
data[data$sl65 == 0.5, 5:6] <- 0

data_cleaned <- data %>%
  pivot_longer(col = sl55:slnone, names_to="speed_limit", names_prefix="sl") %>%
  filter(value >= 0.5) %>%
  subset(select = -c(value))
```

- Produce a new variable, called `year_of_observation` that re-encodes the data that is in `d80`, `d81`, . . . , `d04`.

```r
data_cleaned <- data_cleaned[-c(which(colnames(data_cleaned)=="d80"):which(colnames(data_cleaned)=="d04

data_cleaned <- data_cleaned %>%
       rename("year_of_observation" = "year")
```

- Produce a new variable for each of the other variables that are one-hot encoded (i.e. `bac*` variable series).

```r
data_cleaned <- add_column(data_cleaned, bacnone = 0, .after = "bac08")

data_cleaned <- data_cleaned %>%
  mutate(
    bacnone = ifelse(bac10 == 0 & bac08 == 0, 1, 0),
    bac10 = ifelse(bac10 > 0 & bac08 == 0, 1, bac10)
    )

data_cleaned[data_cleaned$bac08 == 0.5, "bac10"] <- 0

data_cleaned <- data_cleaned %>%
  pivot_longer(col = bac10:bacnone, names_to="blood_alcohol_level", names_prefix="bac") %>%
  filter(value >= 0.5) %>%
  subset(select = -c(value))
```

- 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.)

```
col_name_lookup <- c(
  "min_drinking_age" = "minage",
  "zero_tol_law" = "zerotol",
  "grad_driver_law" = "gdl",
  "per_se_law" = "perse",
  "total_fatalities" = "totfat",
  "night_fatalities" = "nghtfat",
  "weekend_fatalities" = "wkndfat",
  "total_fatalities_per_100mil_miles" = "totfatpvm",
  "night_fatalities_per_100mil_miles" = "nghtfatpvm",
  "weekend_fatalities_per_100mil_miles" = "wkndfatpvm",
  "state_population" = "statepop",
  "total_fatalities_rate" = "totfatrte",
  "night_fatalities_rate" = "nghtfatrte",
  "weekend_fatalities_rate" = "wkndfatrte",
  "vehicle_miles" = "vehicmiles",
  "unemployment_rate" = "unem",
  "percent_age_14_24" = "perc14_24",
  "speed_limit_70_or_higher" = "sl70plus",
  "primary_seatbelt" = "sbprim",
  "secondary_seatbelt" = "sbsecon",
  "vehicle_miles_per_capita"="vehicmilespc"
  )

data_cleaned <- data_cleaned %>%
  rename(any_of(col_name_lookup))
```

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?

```
pdriving <- pdata.frame(
  data_cleaned,
  index=c("state", "year_of_observation")
  )

pdim(pdriving)
```

```
## Balanced Panel: n = 48, T = 25, N = 1200
```

```
# Renaming States
replacement_state <- c("AL", "AZ", "AR", "CA", "CO", "CT", "DE", "FL",
                       "GA", "ID", "IL", "IN", "IA", "KS", "KY", "LA",
                       "ME", "MD", "MA", "MI", "MN", "MS", "MO", "MT",
                       "NE", "NV", "NH", "NJ", "NM", "NY", "NC", "ND",
```

```r
                        "OH", "OK", "OR", "PA", "RI", "SC", "SD", "TN",
                        "TX", "UT", "VT", "VA", "WA", "WV", "WI", "WY")

levels(pdriving$state) <- replacement_state

replacement_region <- c(
  "south", "southwest", "south", "west", "west", "northeast", "northeast", "south", "south",
  "west", "midwest", "midwest", "midwest", "midwest", "south", "south", "northeast", "northeast",
  "northeast", "midwest", "midwest", "south", "midwest", "west", "midwest", "southwest", "northeast",
  "northeast", "southwest", "northeast", "south", "midwest", "midwest", "southwest", "west", "northeast
  "northeast", "south", "midwest", "south", "southwest", "west", "northeast", "south", "west", "south",
  "midwest", "west")

pdriving$region <- pdriving$state
levels(pdriving$region) <- replacement_region
```

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

```r
sapply(pdriving, function(x)sum(is.na(x)))
```

```
##                 year_of_observation                               state
##                                   0                                   0
##                            seatbelt                     min_drinking_age
##                                   0                                   0
##                        zero_tol_law                      grad_driver_law
##                                   0                                   0
##                          per_se_law                     total_fatalities
##                                   0                                   0
##                   night_fatalities                   weekend_fatalities
##                                   0                                   0
##   total_fatalities_per_100mil_miles   night_fatalities_per_100mil_miles
##                                   0                                   0
## weekend_fatalities_per_100mil_miles                     state_population
##                                   0                                   0
##               total_fatalities_rate                night_fatalities_rate
##                                   0                                   0
##             weekend_fatalities_rate                       vehicle_miles
##                                   0                                   0
##                   unemployment_rate                    percent_age_14_24
##                                   0                                   0
##              speed_limit_70_or_higher                    primary_seatbelt
##                                   0                                   0
##                  secondary_seatbelt              vehicle_miles_per_capita
##                                   0                                   0
##                         speed_limit                   blood_alcohol_level
##                                   0                                   0
##                              region
##                                   0
```

```
table(pdriving$year_of_observation)
```

```
##
## 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995
##   48   48   48   48   48   48   48   48   48   48   48   48   48   48   48   48
## 1996 1997 1998 1999 2000 2001 2002 2003 2004
##   48   48   48   48   48   48   48   48   48
```

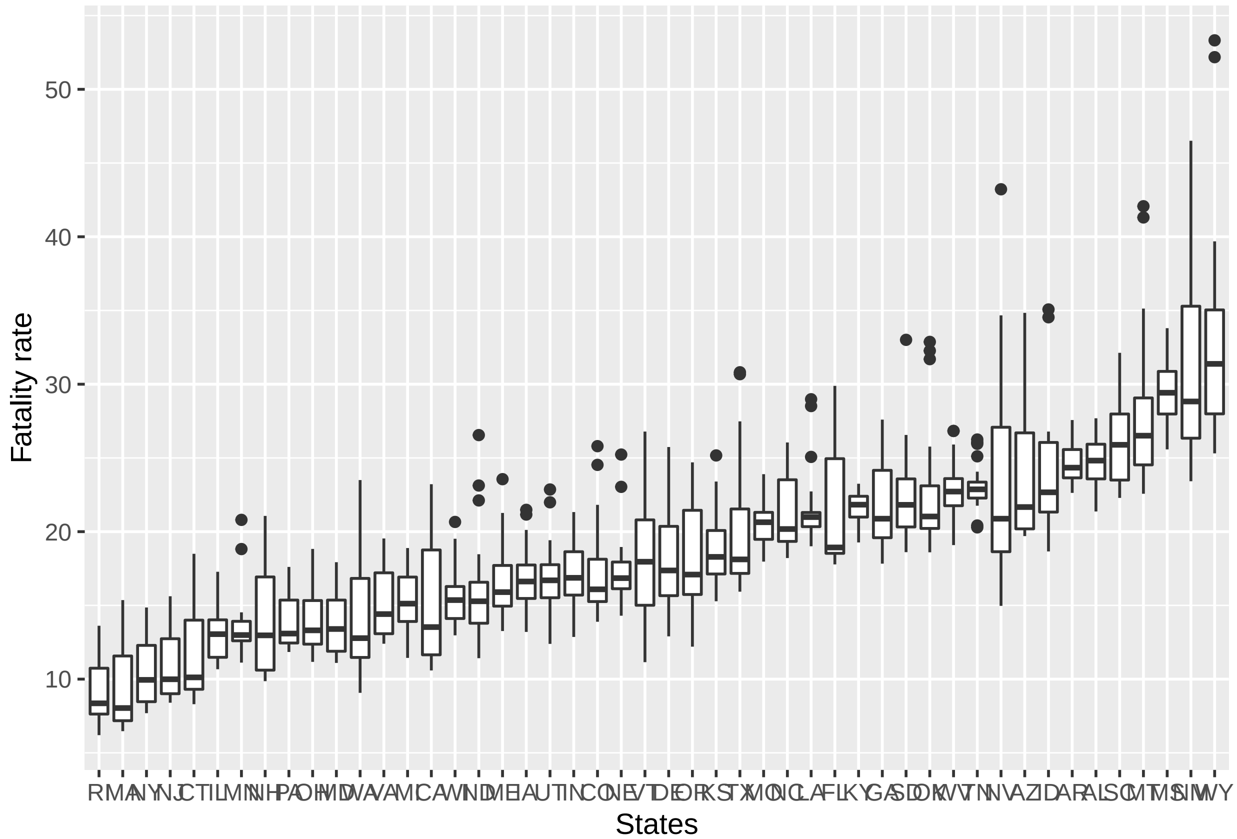
```
table(pdriving$state)
```

```
##
## AL AZ AR CA CO CT DE FL GA ID IL IN IA KS KY LA ME MD MA MI MN MS MO MT NE NV
## 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25
## NH NJ NM NY NC ND OH OK OR PA RI SC SD TN TX UT VT VA WA WV WI WY
## 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25 25
```

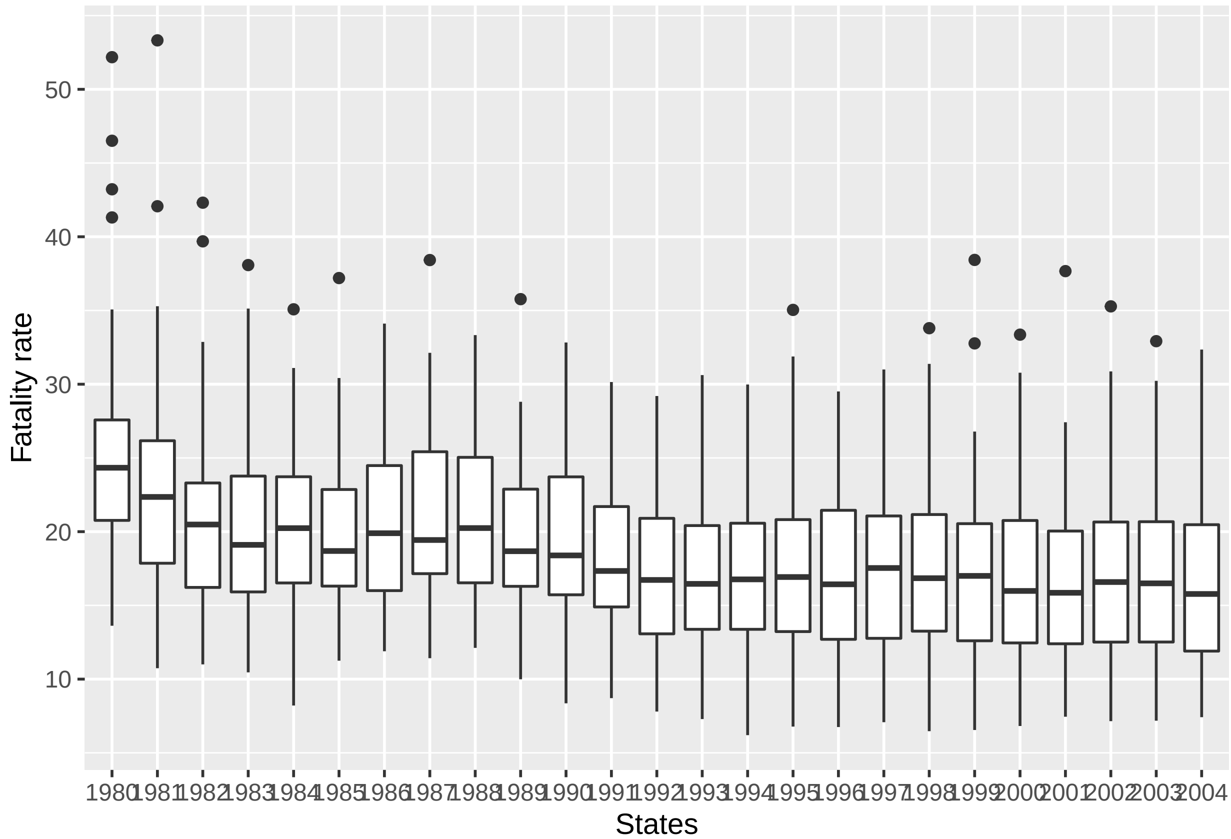
```
pdriving %>%
 is.pconsecutive()
```

```
##    1    3    4    5    6    7    8   10   11   13   14   15   16   17   18   19
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   20   21   22   23   24   25   26   27   28   29   30   31   32   33   34   35
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
##   36   37   38   39   40   41   42   43   44   45   46   47   48   49   50   51
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

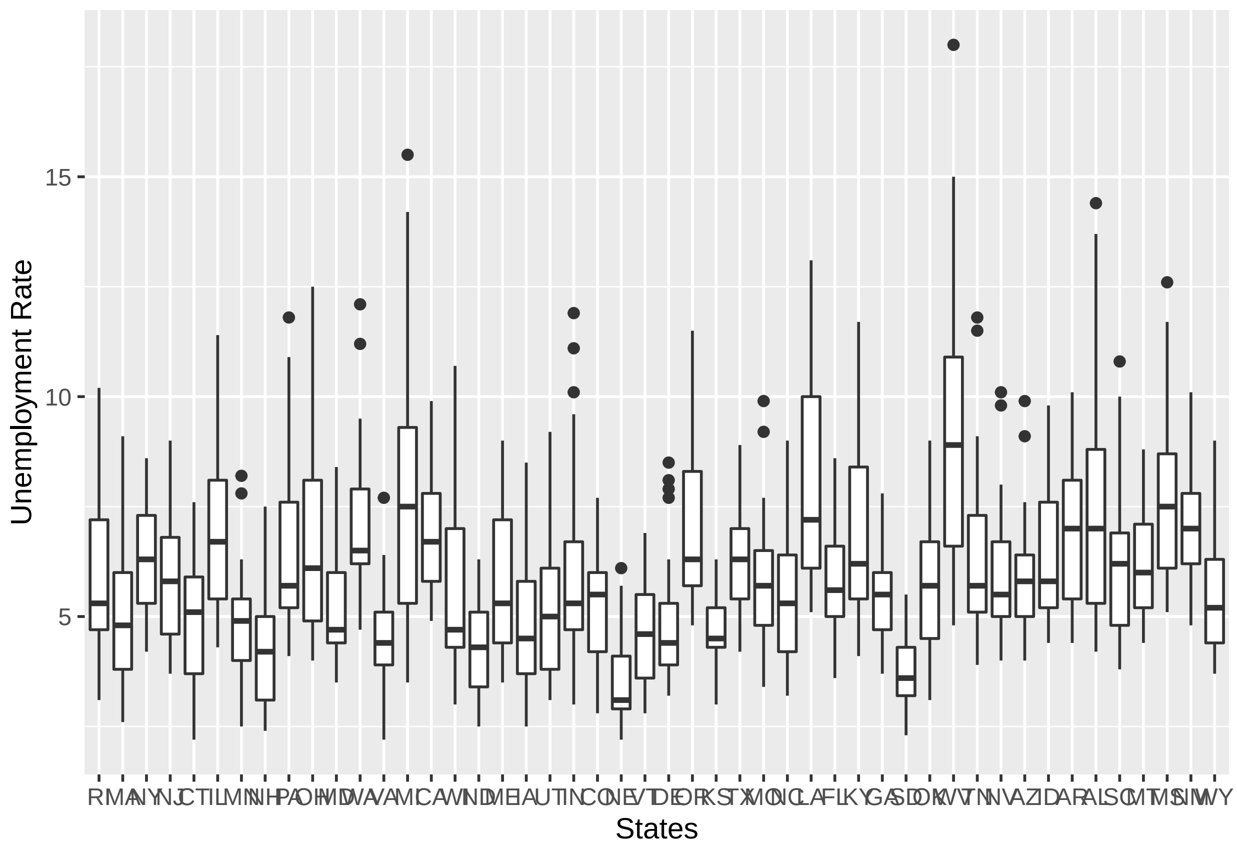
```
pdriving %>%
  group_by(state) %>%
  ggplot(
    aes(x = reorder(state,total_fatalities_rate),
        y = total_fatalities_rate)) +
  geom_boxplot() +
  labs(
    x = "States",
    y = "Fatality rate"
    )
```



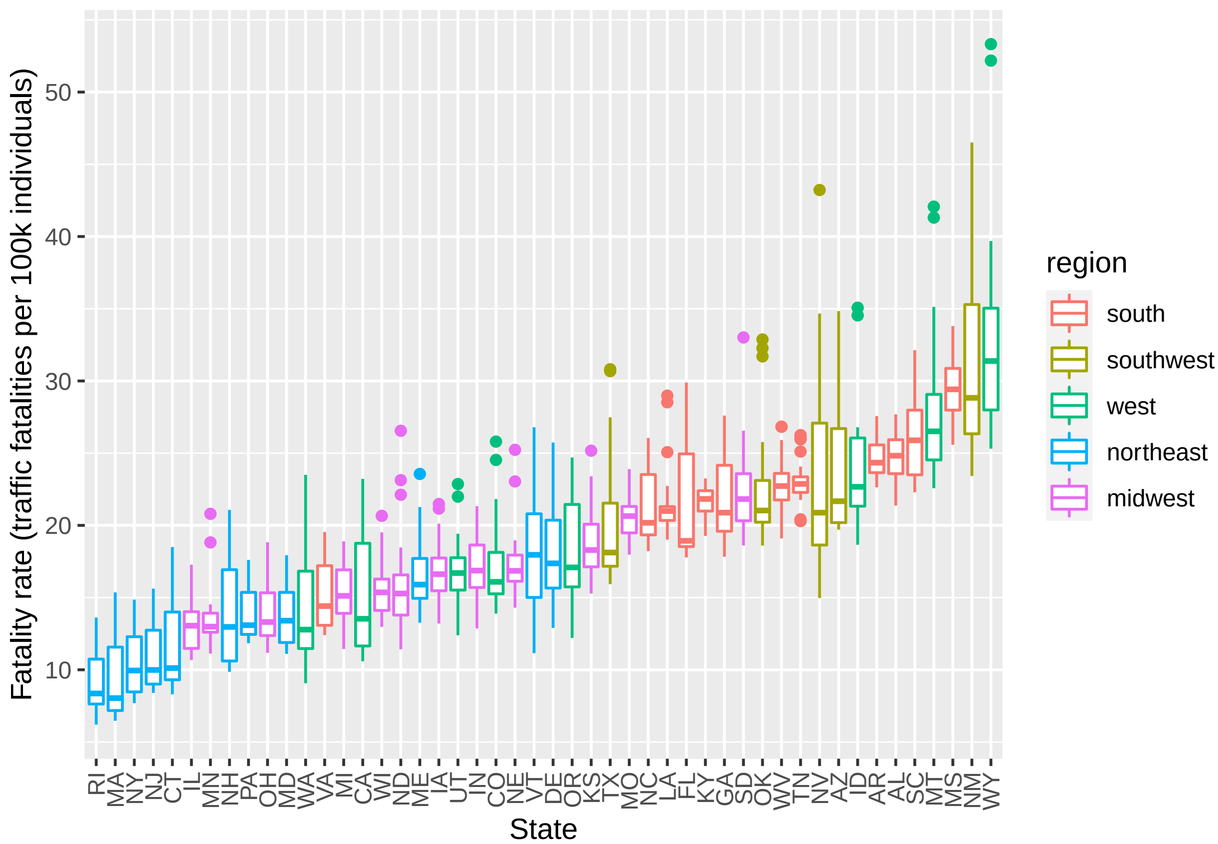

```
pdriving %>%
  group_by(year_of_observation) %>%
  ggplot(aes(x = year_of_observation, y = total_fatalities_rate)) +
  geom_boxplot() +
  labs(
    x = "States",
    y = "Fatality rate"
    )
```



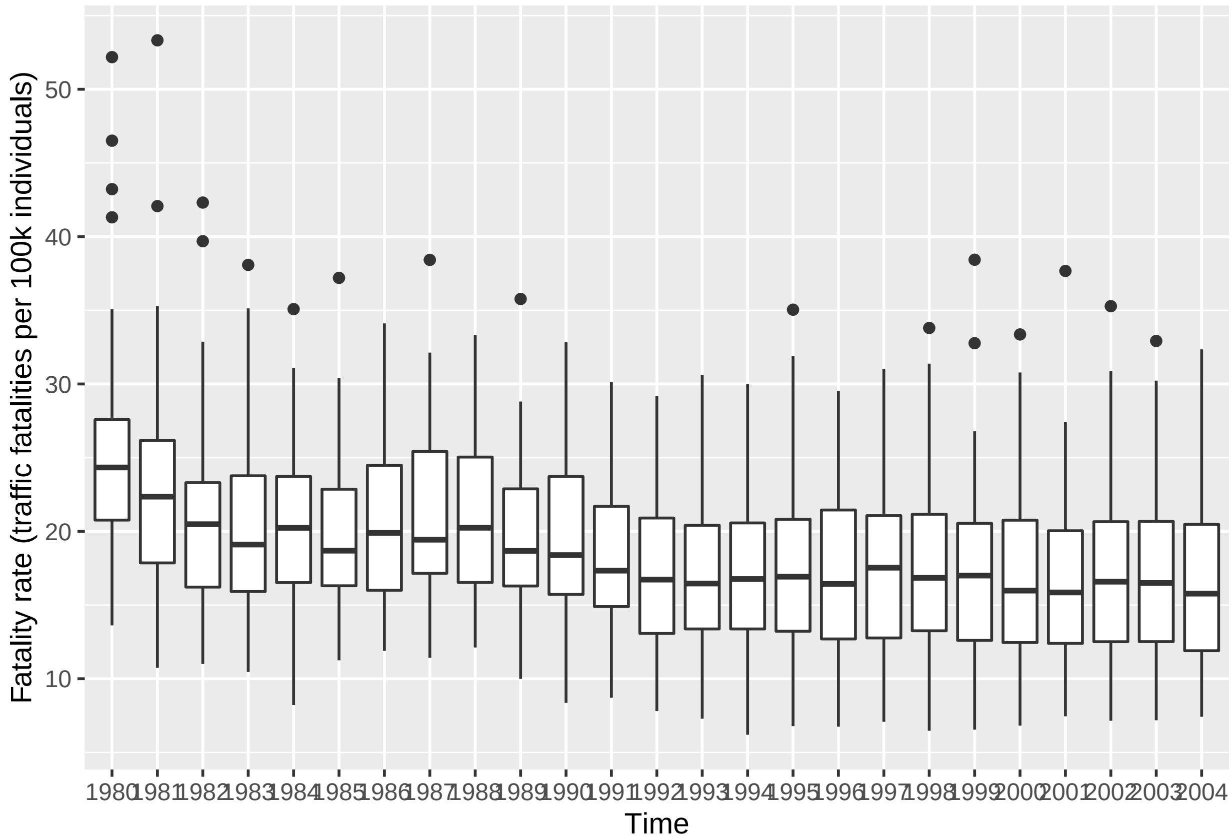

```
pdriving %>%
  group_by(state) %>%
  ggplot(aes(x = reorder(state,total_fatalities_rate), y = unemployment_rate)) +
  geom_boxplot() +
  labs(
    x = "States",
    y = "Unemployment Rate",
    )
```




```
#################################################################
pdriving %>%
  group_by(state) %>%
  ggplot(
    aes(x = reorder(state, total_fatalities_rate),
        y = total_fatalities_rate,
        colour = region)) +
  geom_boxplot() +
  labs(
    x = "State",
    y = "Fatality rate (traffic fatalities per 100k individuals)",
    ) +
  theme(
    axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)
  )
```

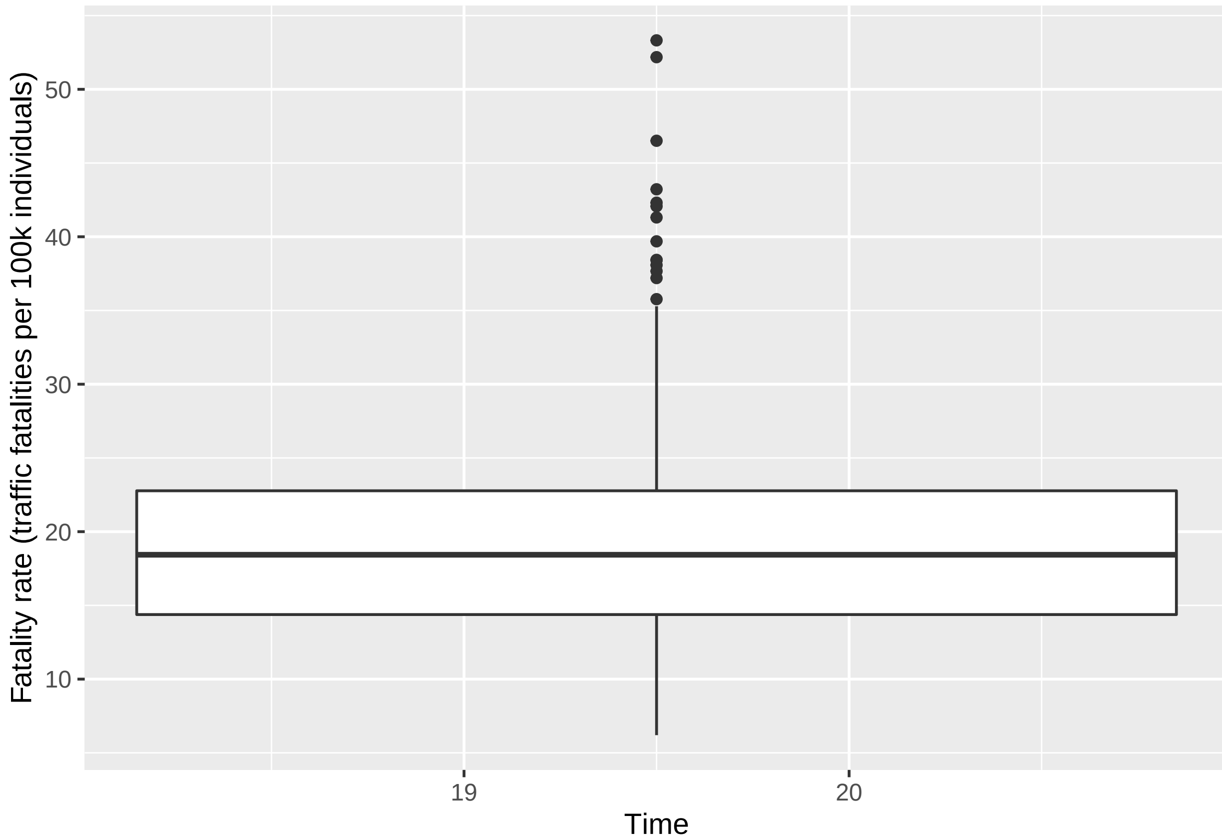



```
pdriving %>%
  group_by(year_of_observation) %>%
  ggplot(aes(x = year_of_observation,
             y = total_fatalities_rate)) +
  geom_boxplot() +
  labs(x = "Time",
       y = "Fatality rate (traffic fatalities per 100k individuals)",
       )
```

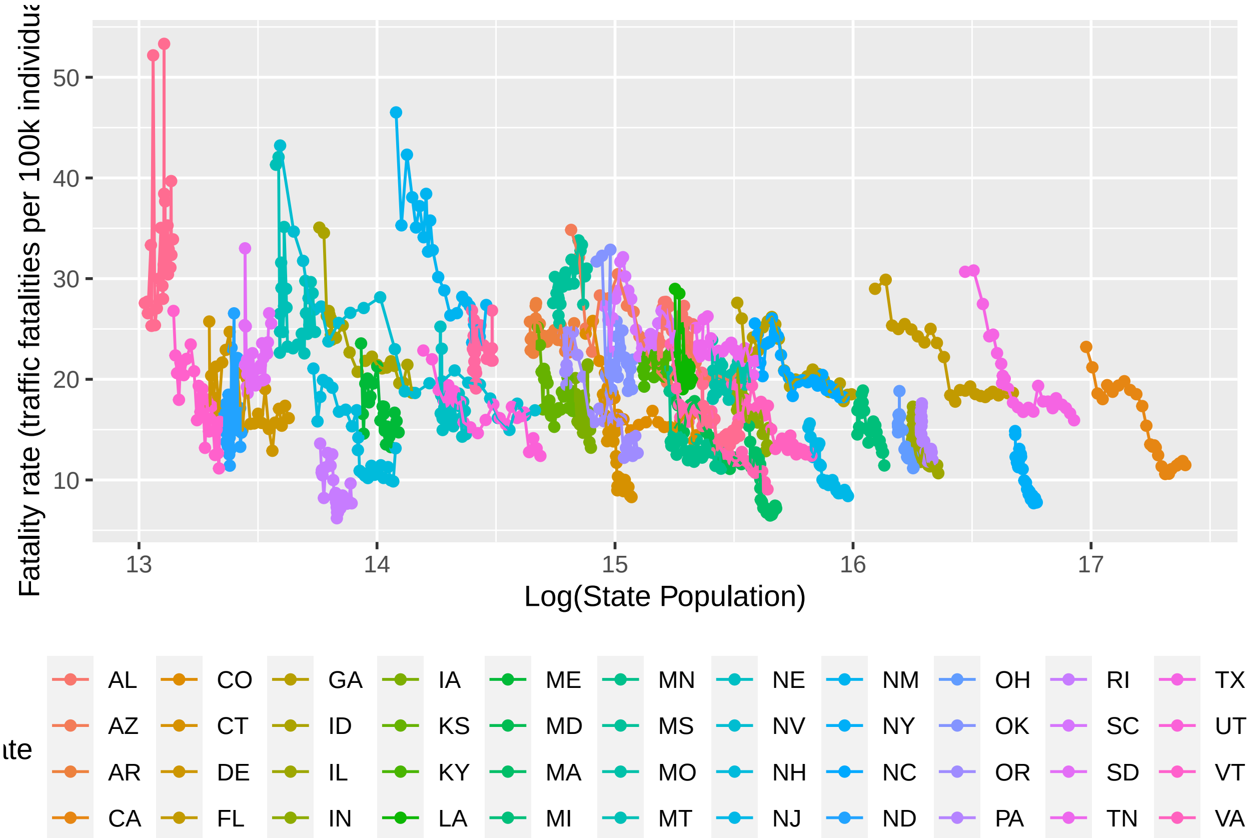



```
pdriving %>%
  group_by(year_of_observation) %>%
  ggplot(aes(x = min_drinking_age,
             y = total_fatalities_rate),
         group = min_drinking_age) +
  geom_boxplot() +
  labs(x = "Time",
       y = "Fatality rate (traffic fatalities per 100k individuals)",
       )
```

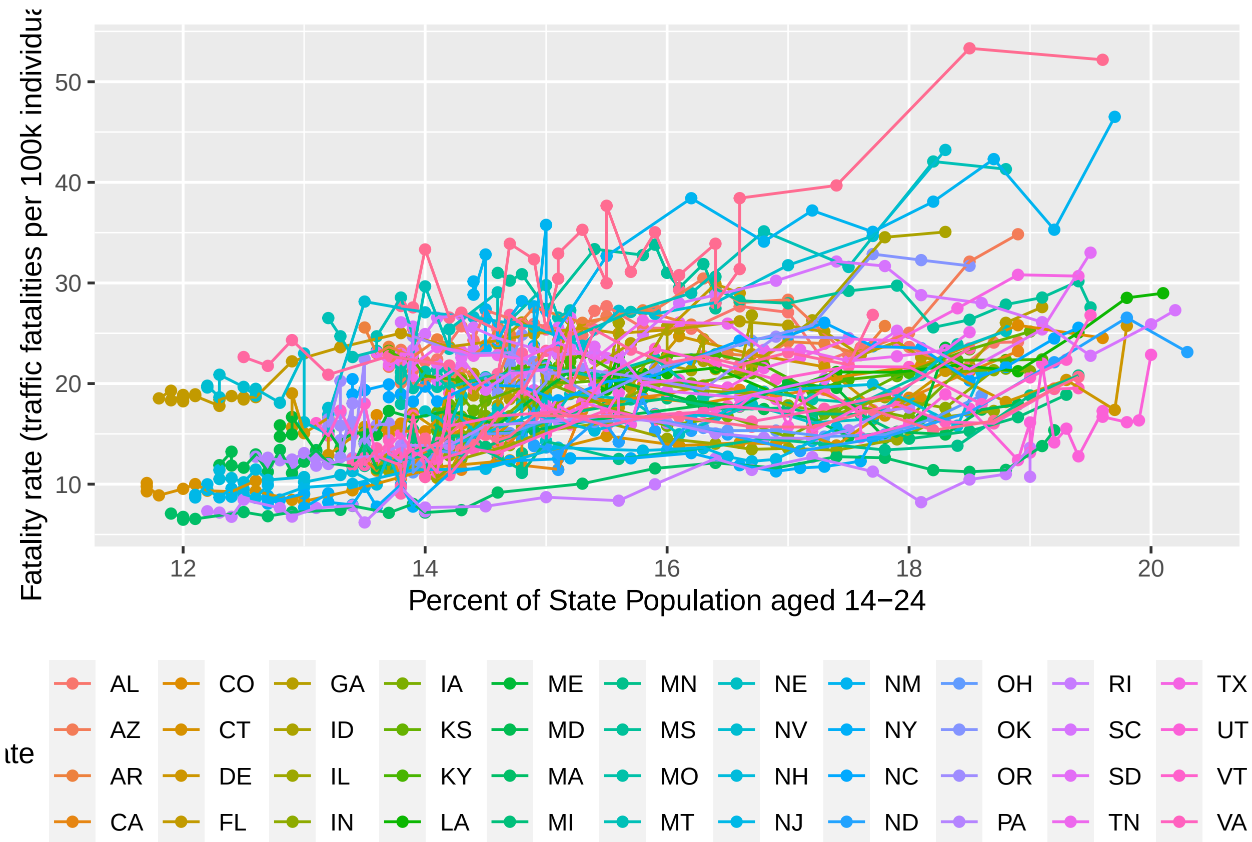
```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```




```
pdriving %>%
  ggplot(aes(x = log(state_population), y = total_fatalities_rate)) +
  geom_point(aes(colour = state)) +
  geom_line(aes(colour = state)) +
    labs(x = "Log(State Population)",
       y = "Fatality rate (traffic fatalities per 100k individuals)",
       ) +
  theme(
    legend.position = 'bottom'
  ) +
  guides(
    colour = guide_legend(nrow = 4)
  )
```

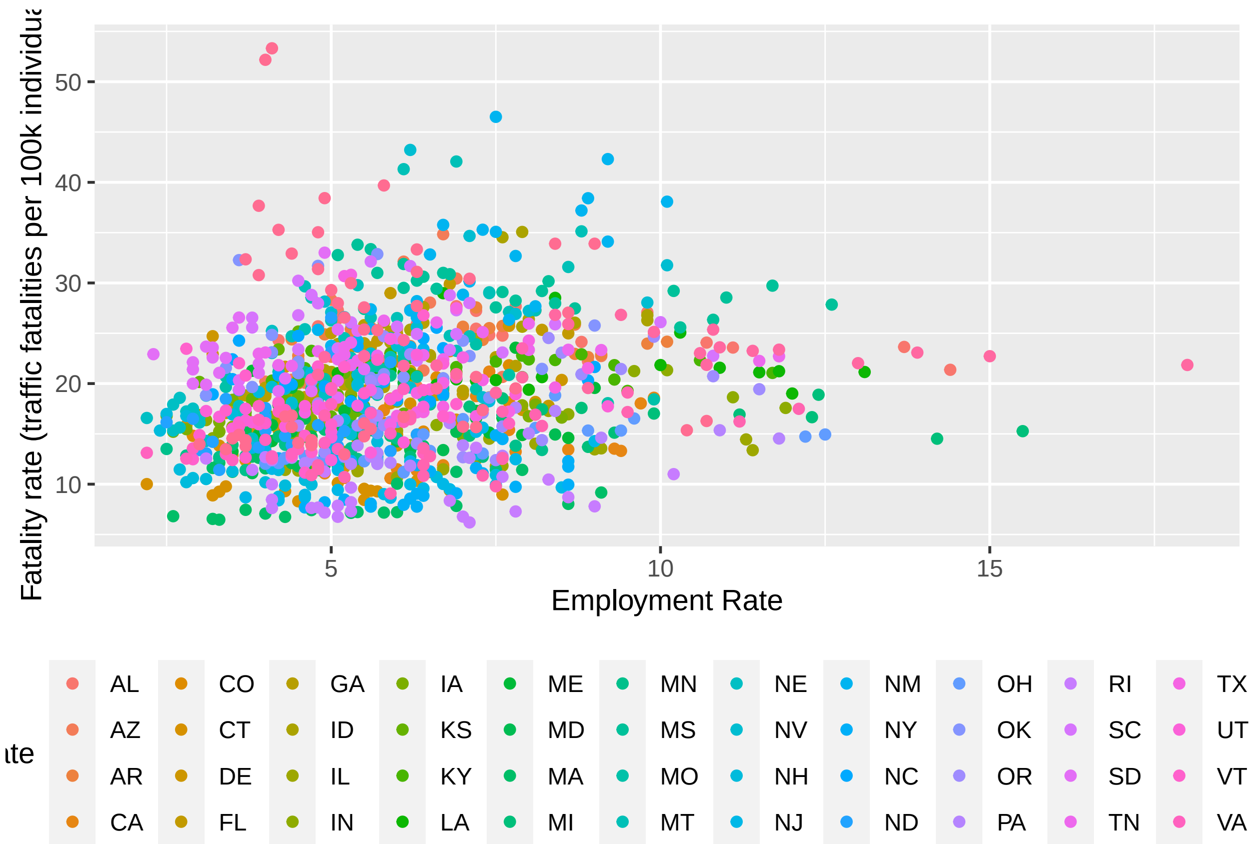



```
pdriving %>%
  ggplot(aes(x = percent_age_14_24, y = total_fatalities_rate)) +
  geom_point(aes(colour = state)) +
  geom_line(aes(colour = state)) +
    labs(x = "Percent of State Population aged 14-24",
       y = "Fatality rate (traffic fatalities per 100k individuals)",
       ) +
  theme(
    legend.position = 'bottom'
  ) +
  guides(
    colour = guide_legend(nrow = 4)
  )
```




```r
pdriving %>%
  ggplot(aes(x = total_fatalities_per_100mil_miles, y = total_fatalities_rate)) +
  geom_point(aes(colour = state)) +
  geom_line(aes(colour = state)) +
    labs(x = "Fatality rate (traffic fatalities per 100M miles driven)",
       y = "Fatality rate (traffic fatalities per 100k individuals)",
       ) +
  theme(
    legend.position = 'bottom'
  ) +
  guides(
    colour = guide_legend(nrow = 4)
  )
```


```
pdriving %>%
  ggplot(aes(x = unemployment_rate, y = total_fatalities_rate)) +
  geom_point(aes(colour = state)) +
    labs(x = "Employment Rate",
       y = "Fatality rate (traffic fatalities per 100k individuals)",
       ) +
  theme(
    legend.position = 'bottom'
  ) +
  guides(
    colour = guide_legend(nrow = 4)
  )
```



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:

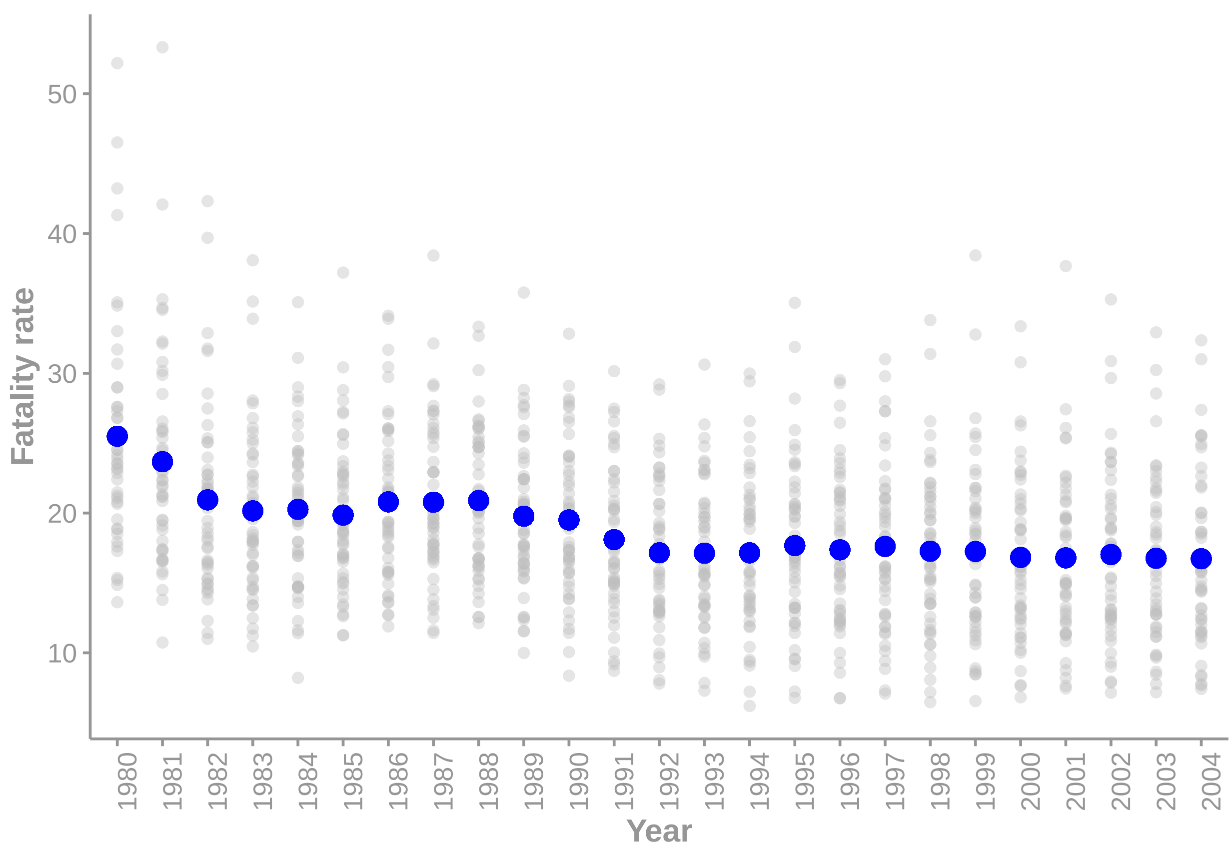
```
mod.prel <- plm(
  formula = total_fatalities_rate ~ year_of_observation,
  data = pdriving,
  model = "pooling"
  )

summary(mod.prel)
```

```
## Pooling Model
##
## Call:
## plm(formula = total_fatalities_rate ~ year_of_observation, data = pdriving,
##     model = "pooling")
```

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

```
pdriving%>%
  ggplot(
    aes(
      x = year_of_observation,
      y = total_fatalities_rate
      )
    ) +
  geom_point(
    color = "gray",
    alpha = 0.4
    ) +
```

```
  geom_point(
    data = broom::augment(mod.prel),
    aes(x = year_of_observation, y = .fitted),
    colour = "blue", size = 3) +
  labs(
    x = "Year",
    y = "Fatality rate"
    ) +
  theme_classic() +
  theme(
  plot.title = element_text(color = "#0099F8",
                            size = 14,
                            face = "bold"),
  plot.subtitle = element_text(color="#969696",
                               size = 12,
                               face = "italic"),
  axis.title = element_text(color = "#969696",
                            size = 12,
                            face = "bold"),
  axis.text = element_text(color = "#969696", size = 10),
  axis.text.x = element_text(angle = 90),
  axis.line = element_line(color = "#969696"),
  axis.ticks = element_line(color = "#969696"),
  legend.position = "none"
  )
```



- 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:

```
mod.exp <- plm(
  formula = total_fatalities_rate ~ year_of_observation +
    blood_alcohol_level + per_se_law + primary_seatbelt +
    secondary_seatbelt + speed_limit_70_or_higher + grad_driver_law +
    percent_age_14_24 + unemployment_rate + vehicle_miles_per_capita,
  data = pdriving,
  model = "pooling"
  )

summary(mod.exp)
```

```
## Pooling Model
##
## Call:
## plm(formula = total_fatalities_rate ~ year_of_observation + blood_alcohol_level +
##     per_se_law + primary_seatbelt + secondary_seatbelt + speed_limit_70_or_higher +
##     grad_driver_law + percent_age_14_24 + unemployment_rate +
##     vehicle_miles_per_capita, data = pdriving, model = "pooling")
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -15.0113  -2.6999  -0.3100   2.2578  21.4537
##
## Coefficients:
##                            Estimate  Std. Error  t-value  Pr(>|t|)
## (Intercept)              -5.1181e+00  2.5137e+00  -2.0361  0.041970 *
## year_of_observation1981  -2.1733e+00  8.2874e-01  -2.6224  0.008845 **
## year_of_observation1982  -6.5557e+00  8.5508e-01  -7.6668 3.703e-14 ***
## year_of_observation1983  -7.2840e+00  8.7811e-01  -8.2951 2.967e-16 ***
## year_of_observation1984  -5.8996e+00  8.7928e-01  -6.7096 3.039e-11 ***
## year_of_observation1985  -6.5371e+00  8.9792e-01  -7.2802 6.116e-13 ***
## year_of_observation1986  -5.9422e+00  9.3170e-01  -6.3778 2.585e-10 ***
## year_of_observation1987  -6.4635e+00  9.6788e-01  -6.6780 3.743e-11 ***
## year_of_observation1988  -6.6980e+00  1.0146e+00  -6.6020 6.150e-11 ***
## year_of_observation1989  -8.1730e+00  1.0535e+00  -7.7582 1.873e-14 ***
## year_of_observation1990  -9.0581e+00  1.0779e+00  -8.4034 < 2.2e-16 ***
```

```
## year_of_observation1991  -1.1153e+01  1.1024e+00 -10.1168 < 2.2e-16 ***
## year_of_observation1992  -1.2978e+01  1.1237e+00 -11.5488 < 2.2e-16 ***
## year_of_observation1993  -1.2825e+01  1.1375e+00 -11.2745 < 2.2e-16 ***
## year_of_observation1994  -1.2457e+01  1.1583e+00 -10.7542 < 2.2e-16 ***
## year_of_observation1995  -1.2065e+01  1.1842e+00 -10.1879 < 2.2e-16 ***
## year_of_observation1996  -1.3960e+01  1.2251e+00 -11.3953 < 2.2e-16 ***
## year_of_observation1997  -1.4340e+01  1.2514e+00 -11.4594 < 2.2e-16 ***
## year_of_observation1998  -1.5148e+01  1.2664e+00 -11.9613 < 2.2e-16 ***
## year_of_observation1999  -1.5209e+01  1.2849e+00 -11.8368 < 2.2e-16 ***
## year_of_observation2000  -1.5564e+01  1.3059e+00 -11.9183 < 2.2e-16 ***
## year_of_observation2001  -1.6300e+01  1.3349e+00 -12.2104 < 2.2e-16 ***
## year_of_observation2002  -1.6811e+01  1.3505e+00 -12.4479 < 2.2e-16 ***
## year_of_observation2003  -1.7139e+01  1.3601e+00 -12.6015 < 2.2e-16 ***
## year_of_observation2004  -1.6825e+01  1.3877e+00 -12.1241 < 2.2e-16 ***
## blood_alcohol_level10     1.0528e+00  3.7069e-01   2.8401  0.004588 **
## blood_alcohol_levelnone   2.2793e+00  5.2929e-01   4.3064 1.799e-05 ***
## per_se_law               -6.5235e-01  2.9762e-01  -2.1919  0.028585 *
## primary_seatbelt         -8.1877e-02  4.9139e-01  -0.1666  0.867695
## secondary_seatbelt        6.8084e-02  4.2943e-01   0.1585  0.874054
## speed_limit_70_or_higher  3.3281e+00  4.4588e-01   7.4641 1.636e-13 ***
## grad_driver_law          -4.1106e-01  5.2761e-01  -0.7791  0.436075
## percent_age_14_24         1.4410e-01  1.2285e-01   1.1730  0.241042
## unemployment_rate         7.5436e-01  7.8044e-02   9.6658 < 2.2e-16 ***
## vehicle_miles_per_capita  2.9303e-03  9.5117e-05  30.8073 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    48612
## Residual Sum of Squares: 19119
## R-Squared:      0.6067
## Adj. R-Squared: 0.59523
## F-statistic: 52.8573 on 34 and 1165 DF, p-value: < 2.22e-16
```

- 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?

5 (15 points) State-Level Fixed Effects

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

```
mod.fix <- plm(
  formula = total_fatalities_rate ~ state + year_of_observation +
    blood_alcohol_level + per_se_law + primary_seatbelt +
    secondary_seatbelt + speed_limit_70_or_higher + grad_driver_law +
    percent_age_14_24 + unemployment_rate + vehicle_miles_per_capita - 1,
  data = pdriving,
  effect = "individual",
  model = "pooling"
  )

summary(mod.fix)
```

```
## Pooling Model
##
## Call:
## plm(formula = total_fatalities_rate ~ state + year_of_observation +
##     blood_alcohol_level + per_se_law + primary_seatbelt + secondary_seatbelt +
##     speed_limit_70_or_higher + grad_driver_law + percent_age_14_24 +
##     unemployment_rate + vehicle_miles_per_capita - 1, data = pdriving,
##     effect = "individual", model = "pooling")
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -8.404643 -1.012656  0.021961  0.933989 14.789322
##
## Coefficients:
##                             Estimate  Std. Error  t-value  Pr(>|t|)
## stateAL                   2.3574e+01  2.2089e+00  10.6726 < 2.2e-16 ***
## stateAZ                   2.3560e+01  2.0672e+00  11.3969 < 2.2e-16 ***
## stateAR                   2.3886e+01  2.1057e+00  11.3433 < 2.2e-16 ***
## stateCA                   1.6964e+01  2.0926e+00   8.1065 1.357e-15 ***
## stateCO                   1.6953e+01  2.0634e+00   8.2159 5.772e-16 ***
## stateCT                   1.2367e+01  1.9827e+00   6.2373 6.305e-10 ***
## stateDE                   1.6797e+01  2.1532e+00   7.8009 1.403e-14 ***
## stateFL                   2.1836e+01  1.9112e+00  11.4254 < 2.2e-16 ***
## stateGA                   1.9460e+01  2.2034e+00   8.8317 < 2.2e-16 ***
## stateID                   2.3004e+01  2.1556e+00  10.6720 < 2.2e-16 ***
## stateIL                   1.5274e+01  2.0670e+00   7.3892 2.884e-13 ***
## stateIN                   1.6485e+01  2.1870e+00   7.5379 9.838e-14 ***
## stateIA                   1.7177e+01  2.1003e+00   8.1788 7.727e-16 ***
## stateKS                   1.7457e+01  2.0557e+00   8.4924 < 2.2e-16 ***
## stateKY                   1.9933e+01  2.1977e+00   9.0700 < 2.2e-16 ***
## stateLA                   2.3365e+01  2.2363e+00  10.4479 < 2.2e-16 ***
## stateME                   1.6241e+01  2.0746e+00   7.8287 1.139e-14 ***
## stateMD                   1.3616e+01  2.0601e+00   6.6092 5.969e-11 ***
## stateMA                   9.0082e+00  2.0653e+00   4.3617 1.410e-05 ***
## stateMI                   1.5848e+01  2.1348e+00   7.4236 2.254e-13 ***
```

```
## stateMN                     1.2766e+01  2.0793e+00   6.1395 1.148e-09 ***
## stateMS                     2.9356e+01  2.2863e+00  12.8400 < 2.2e-16 ***
## stateMO                     1.9495e+01  2.0989e+00   9.2884 < 2.2e-16 ***
## stateMT                     2.5554e+01  2.0992e+00  12.1733 < 2.2e-16 ***
## stateNE                     1.5429e+01  2.0436e+00   7.5501 9.002e-14 ***
## stateNV                     2.3968e+01  1.9737e+00  12.1437 < 2.2e-16 ***
## stateNH                     1.2547e+01  2.0265e+00   6.1913 8.367e-10 ***
## stateNJ                     1.1504e+01  1.9638e+00   5.8581 6.150e-09 ***
## stateNM                     3.0352e+01  2.2437e+00  13.5273 < 2.2e-16 ***
## stateNY                     1.3224e+01  1.9868e+00   6.6558 4.405e-11 ***
## stateNC                     2.0894e+01  2.1317e+00   9.8014 < 2.2e-16 ***
## stateND                     1.3839e+01  2.1696e+00   6.3787 2.610e-10 ***
## stateOH                     1.4511e+01  2.0975e+00   6.9182 7.685e-12 ***
## stateOK                     2.0539e+01  2.1750e+00   9.4432 < 2.2e-16 ***
## stateOR                     1.9592e+01  2.0491e+00   9.5612 < 2.2e-16 ***
## statePA                     1.4733e+01  2.0002e+00   7.3657 3.414e-13 ***
## stateRI                     9.8802e+00  2.0308e+00   4.8651 1.308e-06 ***
## stateSC                     2.3981e+01  2.2205e+00  10.8001 < 2.2e-16 ***
## stateSD                     1.8882e+01  2.1121e+00   8.9398 < 2.2e-16 ***
## stateTN                     2.0949e+01  2.1456e+00   9.7637 < 2.2e-16 ***
## stateTX                     1.9800e+01  2.1797e+00   9.0838 < 2.2e-16 ***
## stateUT                     1.6213e+01  2.2904e+00   7.0789 2.562e-12 ***
## stateVT                     1.5561e+01  2.1433e+00   7.2605 7.210e-13 ***
## stateVA                     1.3227e+01  2.1167e+00   6.2488 5.872e-10 ***
## stateWA                     1.4874e+01  2.0425e+00   7.2821 6.185e-13 ***
## stateWV                     2.4839e+01  2.1622e+00  11.4876 < 2.2e-16 ***
## stateWI                     1.4601e+01  2.1400e+00   6.8226 1.462e-11 ***
## stateWY                     2.7865e+01  2.3407e+00  11.9043 < 2.2e-16 ***
## year_of_observation1981    -1.5066e+00  4.1428e-01  -3.6367 0.0002888 ***
## year_of_observation1982    -2.9878e+00  4.4378e-01  -6.7325 2.660e-11 ***
## year_of_observation1983    -3.4256e+00  4.6392e-01  -7.3839 2.997e-13 ***
## year_of_observation1984    -4.3239e+00  4.6763e-01  -9.2463 < 2.2e-16 ***
## year_of_observation1985    -4.7959e+00  4.8834e-01  -9.8210 < 2.2e-16 ***
## year_of_observation1986    -3.7554e+00  5.1886e-01  -7.2378 8.460e-13 ***
## year_of_observation1987    -4.4044e+00  5.5646e-01  -7.9149 5.925e-15 ***
## year_of_observation1988    -4.8675e+00  6.0273e-01  -8.0757 1.724e-15 ***
## year_of_observation1989    -6.2316e+00  6.4147e-01  -9.7146 < 2.2e-16 ***
## year_of_observation1990    -6.3264e+00  6.6632e-01  -9.4944 < 2.2e-16 ***
## year_of_observation1991    -7.0038e+00  6.8373e-01 -10.2435 < 2.2e-16 ***
## year_of_observation1992    -7.8644e+00  7.0474e-01 -11.1592 < 2.2e-16 ***
## year_of_observation1993    -8.1845e+00  7.1779e-01 -11.4023 < 2.2e-16 ***
## year_of_observation1994    -8.6050e+00  7.3592e-01 -11.6929 < 2.2e-16 ***
## year_of_observation1995    -8.3621e+00  7.5782e-01 -11.0345 < 2.2e-16 ***
## year_of_observation1996    -8.7057e+00  7.9844e-01 -10.9034 < 2.2e-16 ***
## year_of_observation1997    -8.8207e+00  8.2188e-01 -10.7324 < 2.2e-16 ***
## year_of_observation1998    -9.4735e+00  8.3546e-01 -11.3393 < 2.2e-16 ***
## year_of_observation1999    -9.6111e+00  8.4537e-01 -11.3691 < 2.2e-16 ***
## year_of_observation2000    -1.0134e+01  8.5731e-01 -11.8212 < 2.2e-16 ***
## year_of_observation2001    -9.7897e+00  8.7377e-01 -11.2040 < 2.2e-16 ***
## year_of_observation2002    -9.0664e+00  8.8401e-01 -10.2560 < 2.2e-16 ***
## year_of_observation2003    -9.1175e+00  8.9026e-01 -10.2415 < 2.2e-16 ***
## year_of_observation2004    -9.5422e+00  9.1056e-01 -10.4795 < 2.2e-16 ***
## blood_alcohol_level10       2.7426e-01  2.4371e-01   1.1254 0.2606818
## blood_alcohol_levelnone     1.1165e+00  3.7740e-01   2.9584 0.0031570 **
```

```
## per_se_law              -1.1874e+00  2.3299e-01  -5.0963 4.065e-07 ***
## primary_seatbelt        -1.2254e+00  3.4351e-01  -3.5674 0.0003758 ***
## secondary_seatbelt      -3.4882e-01  2.5269e-01  -1.3804 0.1677369
## speed_limit_70_or_higher -6.0765e-02  2.7006e-01  -0.2250 0.8220186
## grad_driver_law         -3.8227e-01  2.9304e-01  -1.3045 0.1923309
## percent_age_14_24        1.9075e-01  9.5331e-02   2.0010 0.0456390 *
## unemployment_rate       -5.7812e-01  6.0689e-02  -9.5258 < 2.2e-16 ***
## vehicle_miles_per_capita  9.3198e-04  1.1138e-04   8.3675 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    48612
## Residual Sum of Squares: 4558.8
## R-Squared:      0.90622
## Adj. R-Squared: 0.89943
## F-statistic: 1416.25 on 82 and 1118 DF, p-value: < 2.22e-16
```

- 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?