Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Lab 4

Grace Lin, Neil Mukerje, Ankitkumar Patel

Contents

Description of the Lab

In this lab, you are asked to answer the question **"Do changes in traffic laws affect traffic fatalities?"** To do so, you will conduct the tasks specified below using the data set *driving.Rdata*, which includes 25 years of data that cover changes in various state drunk driving, seat belt, and speed limit laws.

```r
load("~/git/main-2019-summer/labs/lab4/driving.RData")
head(data,5)
```

```
##   year state sl55 sl65 sl70 sl75 slnone seatbelt minage zerotol gdl bac10
## 1 1980     1    1    0    0    0      0        0     18       0   0     1
## 2 1981     1    1    0    0    0      0        0     18       0   0     1
## 3 1982     1    1    0    0    0      0        0     18       0   0     1
## 4 1983     1    1    0    0    0      0        0     18       0   0     1
## 5 1984     1    1    0    0    0      0        0     18       0   0     1
##   bac08 perse totfat nghtfat wkndfat totfatpvm nghtfatpvm wkndfatpvm
## 1     0     0    940     422     236      3.20      1.437      0.803
## 2     0     0    933     434     248      3.35      1.558      0.890
## 3     0     0    839     376     224      2.81      1.259      0.750
## 4     0     0    930     397     223      3.00      1.281      0.719
## 5     0     0    932     421     237      2.83      1.278      0.720
##   statepop totfatrte nghtfatrte wkndfatrte vehicmiles unem perc14_24
## 1  3893888     24.14      10.84       6.06   29.37500  8.8      18.9
## 2  3918520     24.07      11.08       6.33   27.85200 10.7      18.7
## 3  3925218     21.37       9.58       5.71   29.85765 14.4      18.4
## 4  3934109     23.64      10.09       5.67   31.00000 13.7      18.0
## 5  3951834     23.58      10.65       6.00   32.93286 11.1      17.6
##   sl70plus sbprim sbsecon d80 d81 d82 d83 d84 d85 d86 d87 d88 d89 d90 d91
## 1        0      0       0   1   0   0   0   0   0   0   0   0   0   0   0
## 2        0      0       0   0   1   0   0   0   0   0   0   0   0   0   0
## 3        0      0       0   0   0   1   0   0   0   0   0   0   0   0   0
## 4        0      0       0   0   0   0   1   0   0   0   0   0   0   0   0
## 5        0      0       0   0   0   0   0   1   0   0   0   0   0   0   0
##   d92 d93 d94 d95 d96 d97 d98 d99 d00 d01 d02 d03 d04 vehicmilespc
## 1   0   0   0   0   0   0   0   0   0   0   0   0   0     7543.874
## 2   0   0   0   0   0   0   0   0   0   0   0   0   0     7107.785
## 3   0   0   0   0   0   0   0   0   0   0   0   0   0     7606.622
## 4   0   0   0   0   0   0   0   0   0   0   0   0   0     7879.802
## 5   0   0   0   0   0   0   0   0   0   0   0   0   0     8333.562
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(plm)
```

```
##
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```r
str(data)
```

```
## 'data.frame':    1200 obs. of  56 variables:
##  $ year        : int  1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 ...
##  $ state       : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ sl55        : num  1 1 1 1 1 ...
##  $ sl65        : num  0 0 0 0 0 ...
##  $ sl70        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sl75        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ slnone      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ seatbelt    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ minage      : num  18 18 18 18 18 20 21 21 21 21 ...
##  $ zerotol     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gdl         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ bac10       : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ bac08       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ perse       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ totfat      : int  940 933 839 930 932 882 1080 1111 1024 1029 ...
##  $ nghtfat     : int  422 434 376 397 421 358 500 499 423 418 ...
##  $ wkndfat     : int  236 248 224 223 237 224 279 300 226 247 ...
##  $ totfatpvm   : num  3.2 3.35 2.81 3 2.83 ...
##  $ nghtfatpvm  : num  1.44 1.56 1.26 1.28 1.28 ...
##  $ wkndfatpvm  : num  0.803 0.89 0.75 0.719 0.72 ...
##  $ statepop    : int  3893888 3918520 3925218 3934109 3951834 3972527 3991569 4015261 40238
##  $ totfatrte   : num  24.1 24.1 21.4 23.6 23.6 ...
##  $ nghtfatrte  : num  10.84 11.08 9.58 10.09 10.65 ...
##  $ wkndfatrte  : num  6.06 6.33 5.71 5.67 6 ...
##  $ vehicmiles  : num  29.4 27.9 29.9 31 32.9 ...
##  $ unem        : num  8.8 10.7 14.4 13.7 11.1 ...
##  $ perc14_24   : num  18.9 18.7 18.4 18 17.6 ...
##  $ sl70plus    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbprim      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbsecon     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d80         : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ d81         : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ d82         : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ d83         : int  0 0 0 1 0 0 0 0 0 0 ...
```

```
##  $ d84         : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ d85         : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ d86         : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ d87         : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ d88         : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ d89         : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ d90         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d91         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d92         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d93         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d94         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d95         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d96         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d97         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d98         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d99         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d00         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d01         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d02         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d03         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ d04         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ vehicmilespc: num  7544 7108 7607 7880 8334 ...
##  - attr(*, "datalabel")= chr ""
##  - attr(*, "time.stamp")= chr "22 Jan 2013 14:09"
##  - attr(*, "formats")= chr  "%8.0g" "%8.0g" "%9.0g" "%9.0g" ...
##  - attr(*, "types")= int  252 251 254 254 254 254 254 251 254 254 ...
##  - attr(*, "val.labels")= chr  "" "" "" "" ...
##  - attr(*, "var.labels")= chr  "1980 through 2004" "48 continental states, alphabetical" "s
##  - attr(*, "version")= int 12
```

```r
summary(data)
```

```
##       year          state            sl55             sl65       
##  Min.   :1980   Min.   : 1.00   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:1986   1st Qu.:15.75   1st Qu.:0.0000   1st Qu.:0.0000  
##  Median :1992   Median :27.50   Median :0.0000   Median :0.0000  
##  Mean   :1992   Mean   :27.15   Mean   :0.3533   Mean   :0.4399  
##  3rd Qu.:1998   3rd Qu.:39.25   3rd Qu.:1.0000   3rd Qu.:1.0000  
##  Max.   :2004   Max.   :51.00   Max.   :1.0000   Max.   :1.0000  
##       sl70            sl75             slnone            seatbelt    
##  Min.   :0.000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000  
##  1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000  
##  Median :0.000   Median :0.00000   Median :0.000000   Median :1.000  
##  Mean   :0.119   Mean   :0.08024   Mean   :0.007569   Mean   :1.116  
##  3rd Qu.:0.000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:2.000  
##  Max.   :1.000   Max.   :1.00000   Max.   :1.000000   Max.   :2.000  
##      minage        zerotol             gdl              bac10       
##  Min.   :18.0   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
```

```
##  1st Qu.:21.0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :21.0   Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :20.6   Mean   :0.4519   Mean   :0.1741   Mean   :0.6231
##  3rd Qu.:21.0   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :21.0   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      bac08            perse            totfat          nghtfat
##  Min.   :0.0000   Min.   :0.0000   Min.   :  63.0   Min.   :  26.0
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 310.0   1st Qu.: 139.8
##  Median :0.0000   Median :1.0000   Median : 676.0   Median : 316.0
##  Mean   :0.2135   Mean   :0.5471   Mean   : 900.7   Mean   : 427.3
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1099.5   3rd Qu.: 518.2
##  Max.   :1.0000   Max.   :1.0000   Max.   :5504.0   Max.   :2918.0
##     wkndfat         totfatpvm       nghtfatpvm       wkndfatpvm
##  Min.   :  10.0   Min.   :0.780   Min.   :0.2700   Min.   :0.1140
##  1st Qu.:  70.0   1st Qu.:1.577   1st Qu.:0.6847   1st Qu.:0.3410
##  Median : 163.0   Median :2.020   Median :0.9130   Median :0.4770
##  Mean   : 222.3   Mean   :2.122   Mean   :0.9990   Mean   :0.5255
##  3rd Qu.: 277.0   3rd Qu.:2.500   3rd Qu.:1.2110   3rd Qu.:0.6420
##  Max.   :1499.0   Max.   :5.700   Max.   :3.0030   Max.   :1.6750
##     statepop          totfatrte       nghtfatrte       wkndfatrte
##  Min.   :  453401   Min.   : 6.20   Min.   : 2.660   Min.   : 1.180
##  1st Qu.: 1641938   1st Qu.:14.38   1st Qu.: 6.338   1st Qu.: 3.240
##  Median : 3700425   Median :18.43   Median : 8.420   Median : 4.390
##  Mean   : 5329896   Mean   :18.92   Mean   : 8.796   Mean   : 4.606
##  3rd Qu.: 6069563   3rd Qu.:22.77   3rd Qu.:10.650   3rd Qu.: 5.680
##  Max.   :35894000   Max.   :53.32   Max.   :29.600   Max.   :14.430
##    vehicmiles           unem          perc14_24        sl70plus
##  Min.   :  3.703   Min.   : 2.200   Min.   :11.70   Min.   :0.0000
##  1st Qu.: 14.574   1st Qu.: 4.500   1st Qu.:13.90   1st Qu.:0.0000
##  Median : 33.863   Median : 5.600   Median :14.90   Median :0.0000
##  Mean   : 46.323   Mean   : 5.951   Mean   :15.33   Mean   :0.2068
##  3rd Qu.: 58.639   3rd Qu.: 7.000   3rd Qu.:16.60   3rd Qu.:0.0000
##  Max.   :329.600   Max.   :18.000   Max.   :20.30   Max.   :1.0000
##      sbprim           sbsecon            d80            d81
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00   Min.   :0.00
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:0.00
##  Median :0.0000   Median :0.0000   Median :0.00   Median :0.00
##  Mean   :0.1792   Mean   :0.4683   Mean   :0.04   Mean   :0.04
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00   3rd Qu.:0.00
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00   Max.   :1.00
##       d82            d83            d84            d85
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
##  Median :0.00   Median :0.00   Median :0.00   Median :0.00
##  Mean   :0.04   Mean   :0.04   Mean   :0.04   Mean   :0.04
##  3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00
##       d86            d87            d88            d89
```

```
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00  
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00  
##  Median :0.00   Median :0.00   Median :0.00   Median :0.00  
##  Mean   :0.04   Mean   :0.04   Mean   :0.04   Mean   :0.04  
##  3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00  
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00  
##       d90            d91            d92            d93      
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00  
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00  
##  Median :0.00   Median :0.00   Median :0.00   Median :0.00  
##  Mean   :0.04   Mean   :0.04   Mean   :0.04   Mean   :0.04  
##  3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00  
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00  
##       d94            d95            d96            d97      
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00  
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00  
##  Median :0.00   Median :0.00   Median :0.00   Median :0.00  
##  Mean   :0.04   Mean   :0.04   Mean   :0.04   Mean   :0.04  
##  3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00  
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00  
##       d98            d99            d00            d01      
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.00  
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00  
##  Median :0.00   Median :0.00   Median :0.00   Median :0.00  
##  Mean   :0.04   Mean   :0.04   Mean   :0.04   Mean   :0.04  
##  3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00  
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.00  
##       d02            d03            d04         vehicmilespc  
##  Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   : 4372  
##  1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.: 7788  
##  Median :0.00   Median :0.00   Median :0.00   Median : 9013  
##  Mean   :0.04   Mean   :0.04   Mean   :0.04   Mean   : 9129  
##  3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.:10327  
##  Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :18390
```

```r
#hist(data$year,col='lightblue',breaks=40)
#hist(data$state,col='lightblue',breaks=40)

#ggplot(gather(subset(data, select=c("year","state"))),
#    aes(value)) +
#    geom_histogram(bins = 100) +
#    facet_wrap(~ key, scales = 'free')

E <- pdata.frame(data, index=c("state","year"))

table(E$state)
```

```
##
```

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

```
table(E$year)
```

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

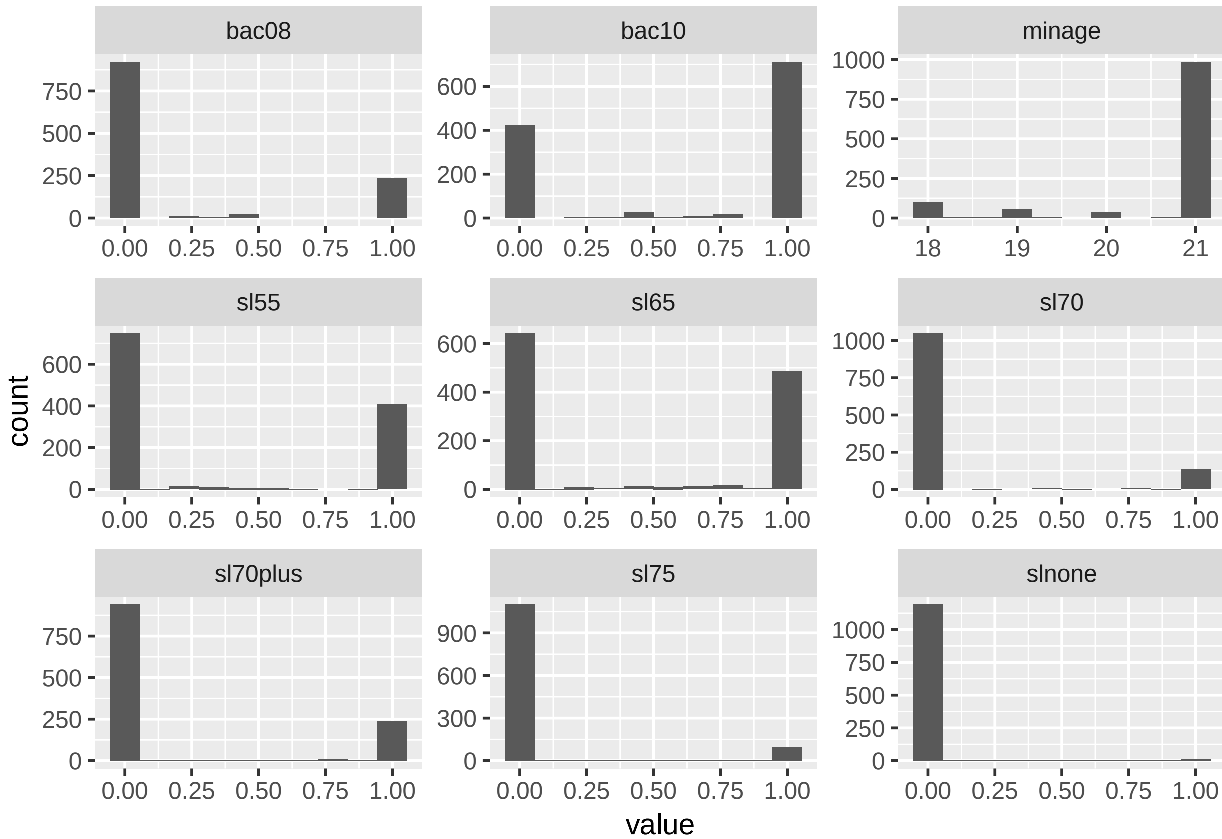
```
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

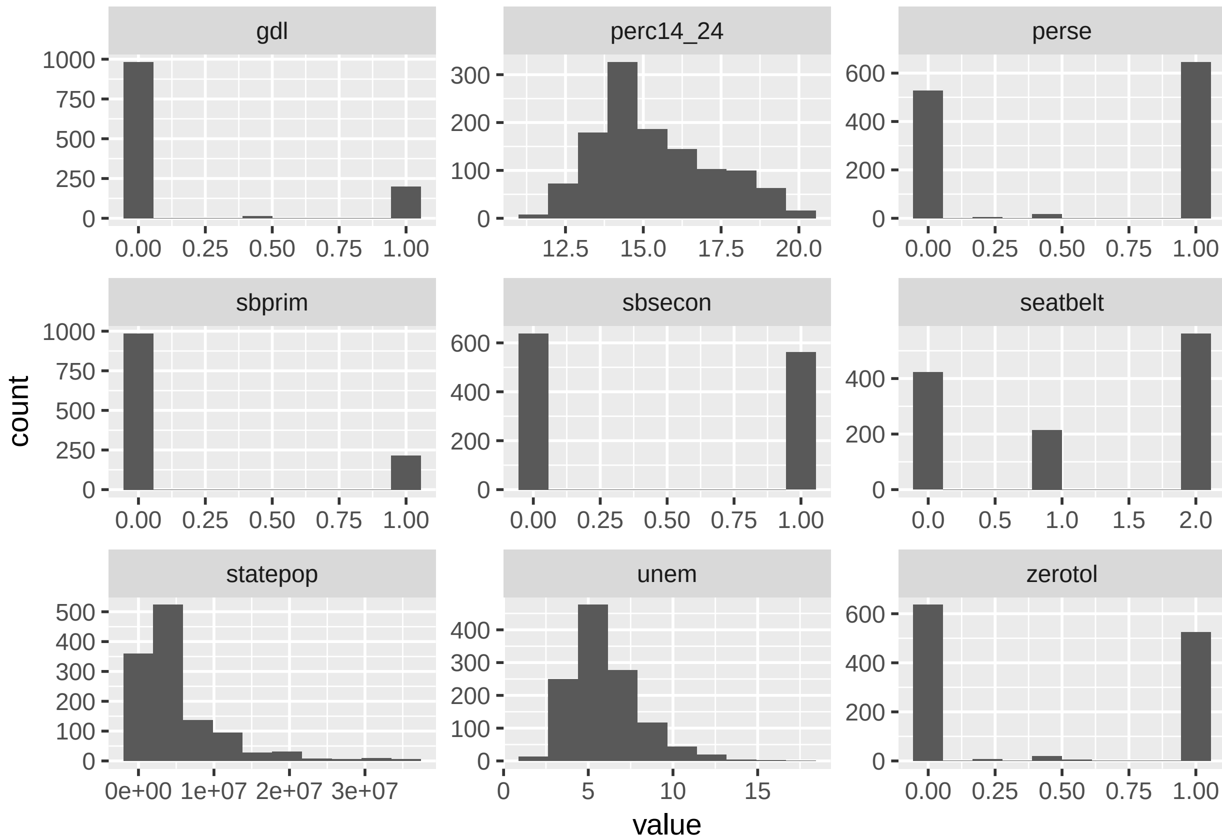
```
library(tidyr)
library(lattice)

# Let's organize the 29 terms
speed_limit_terms = c("sl55","sl65","sl70","sl75","slnone","sl70plus")
drunk_driving_terms = c("minage","bac10","bac08")
admin_law_terms = c("seatbelt","perse","zerotol","gdl","sbprim","sbsecon")
pop_metric_terms = c("unem","perc14_24","statepop")
vehicle_miles_terms = c("vehicmilespc","vehicmiles")
fatality_terms = c("totfat","nghtfat","wkndfat","totfatpvm","nghtfatpvm",
                   "wkndfatpvm","totfatrte","nghtfatrte","wkndfatrte")

date_terms = c("d80","d81","d82","d83","d84","d85","d86","d87","d88","d89",
               "d90","d91","d92","d93","d94","d95","d96","d97","d98","d99",
               "d00","d01","d02","d03","d04")

all_terms = cbind (speed_limit_terms,drunk_driving_terms,admin_law_terms,vehicle_miles_terms,
                   pop_metric_terms,fatality_terms)
```

```
## Warning in cbind(speed_limit_terms, drunk_driving_terms, admin_law_terms, :
## number of rows of result is not a multiple of vector length (arg 1)
```

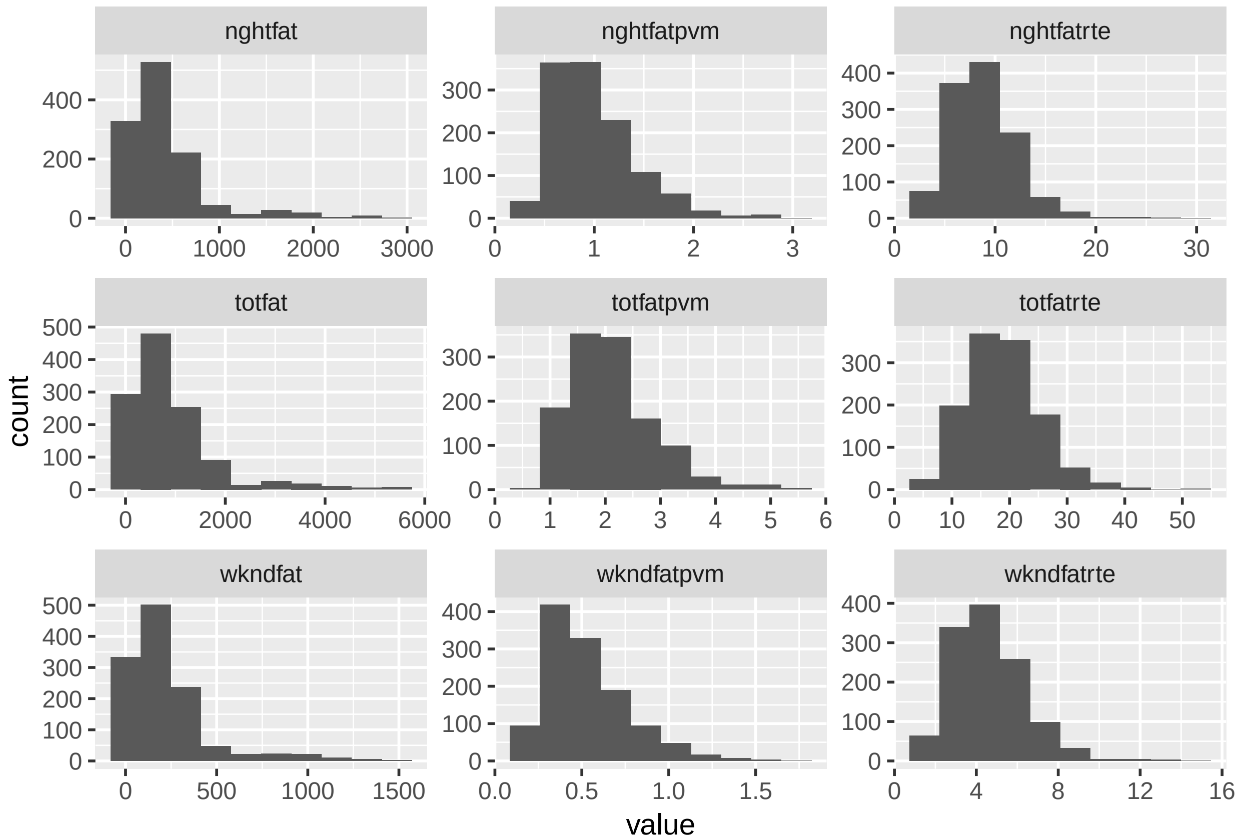
```
# speed_limit_terms + drunk_driving_terms
ggplot(gather(subset(data, select=c(speed_limit_terms,drunk_driving_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 3)
```



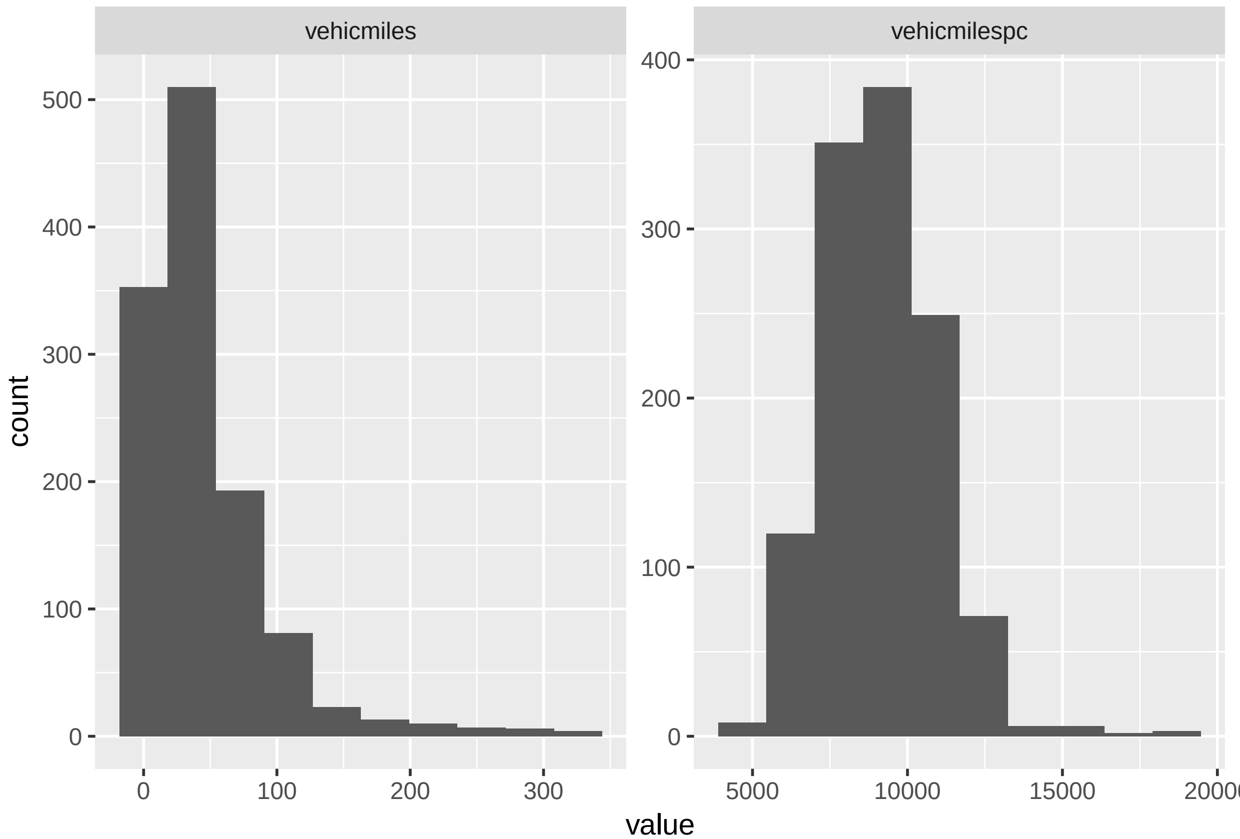

```r
# admin_law_terms + pop_metrics_terms
ggplot(gather(subset(data, select=c(admin_law_terms,pop_metric_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 3)
```



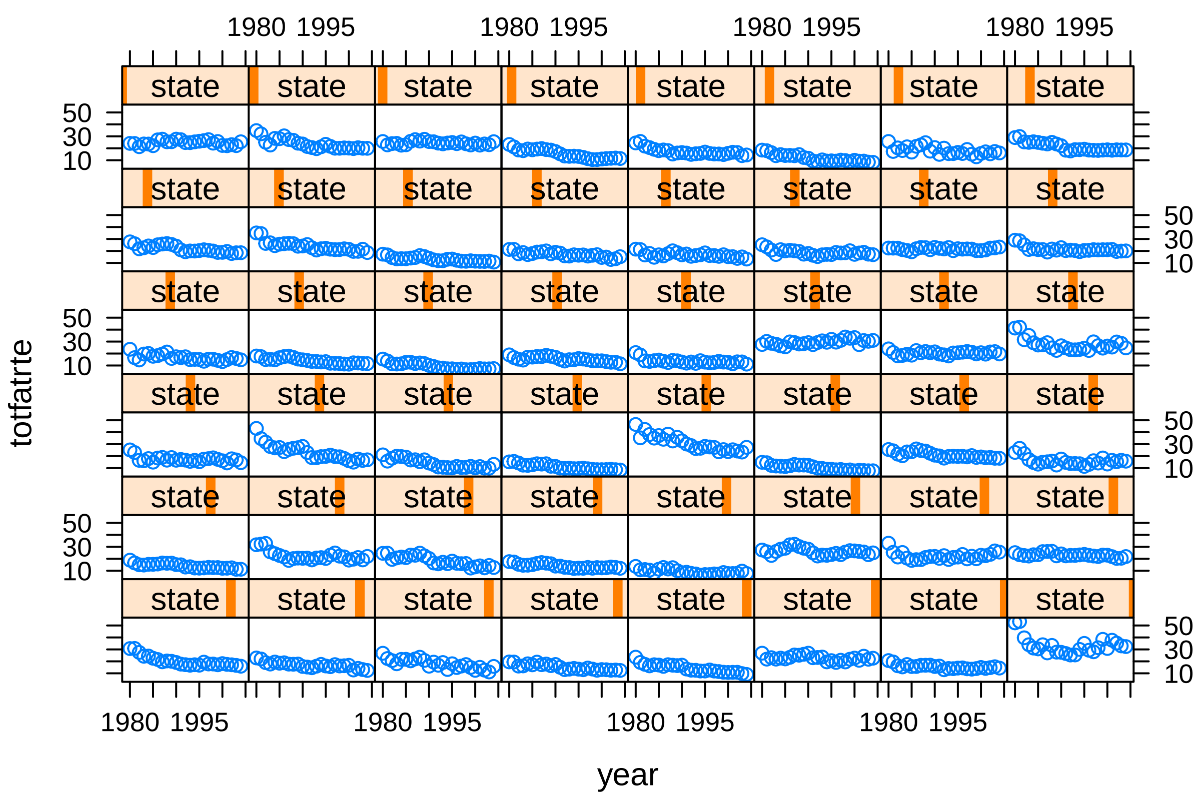

```r
# fatality_terms
ggplot(gather(subset(data, select=c(fatality_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 3)
```



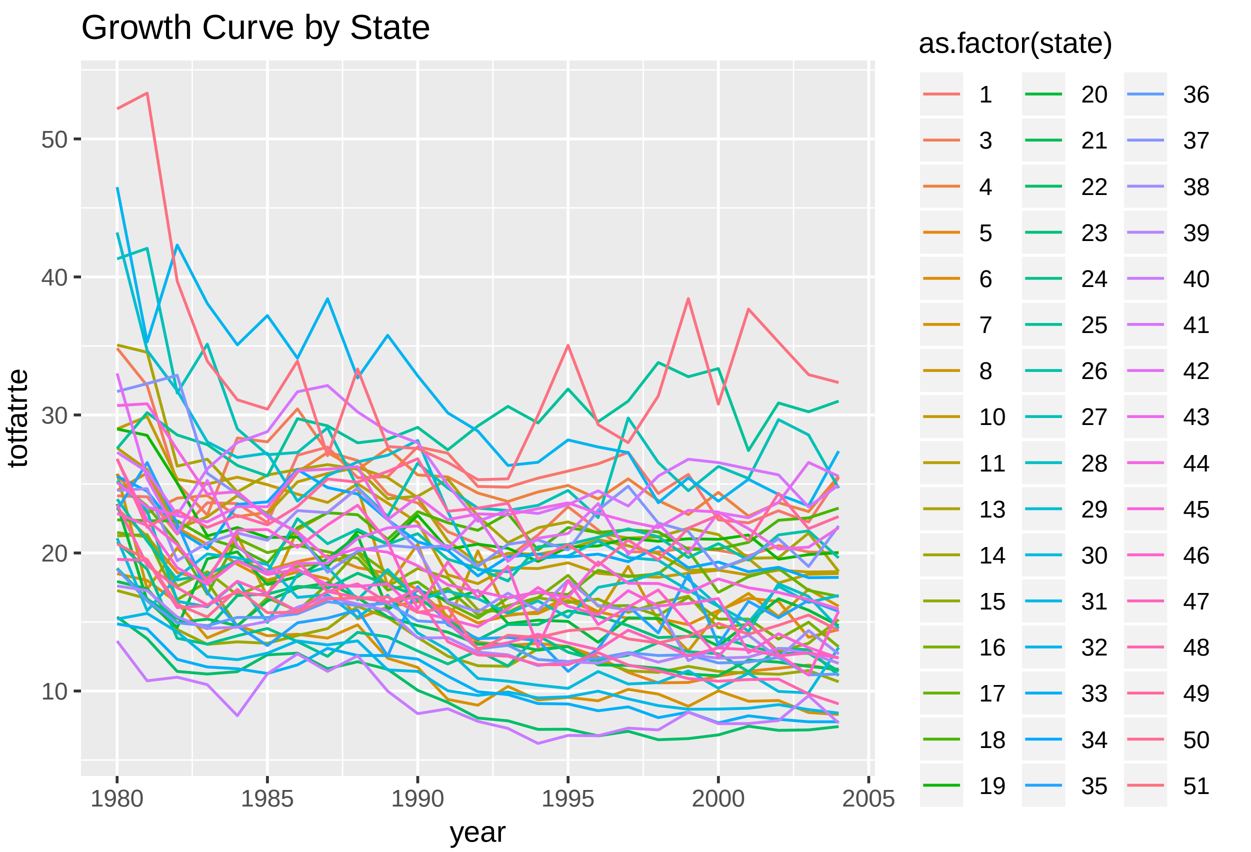

```r
# vehicle_miles_terms
ggplot(gather(subset(data, select=c(vehicle_miles_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 2)
```




```r
xyplot(totfatrte ~ year | state, data=data, as.table=T)
```



```r
ggplot(data, aes(year, totfatrte, colour = as.factor(state))) +
  geom_line() + ggtitle("Growth Curve by State")
```



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 come with the dataste.

Exercises:

Question 1

1. Load the data. Provide a description of the basic structure of the dataset, as we have done throughout the semester. Conduct a very thorough EDA, which should include both graphical and tabular techniques, on the dataset, including both the dependent variable *totfatrte* and the potential explanatory variables. You need to write a detailed narrative of your observations of your EDA. *Reminder: giving an "output dump" (i.e. providing a bunch of graphs and tables without description and hoping your audience will interpret them) will receive a zero in this exercise.*

Question 2

2. How is the our dependent variable of interest *totfatrte* defined? What is the average of this variable in each of the years in the time period covered in this dataset? Estimate a linear regression model of *totfatrte* on a set of dummy variables for the years 1981 through 2004. What does this model explain? Describe what you find in this model. Did driving become safer over this period? Please provide a detailed explanation.

totfatrte is defined as $totfatrte = totfat * 100,000/statepop$

```
data %>%
    group_by(year) %>%
    summarise(mean_totfatrte = mean(totfatrte))
```

```
## # A tibble: 25 x 2
##     year mean_totfatrte
##    <int>          <dbl>
##  1  1980           25.5
##  2  1981           23.7
##  3  1982           20.9
##  4  1983           20.2
##  5  1984           20.3
##  6  1985           19.9
##  7  1986           20.8
##  8  1987           20.8
##  9  1988           20.9
## 10  1989           19.8
## # ... with 15 more rows
```

```
data1 <- data %>% select(c("totfatrte",date_terms))
lm_fit1 = lm(totfatrte ~ ., data=data1)
summary(lm_fit1)
```

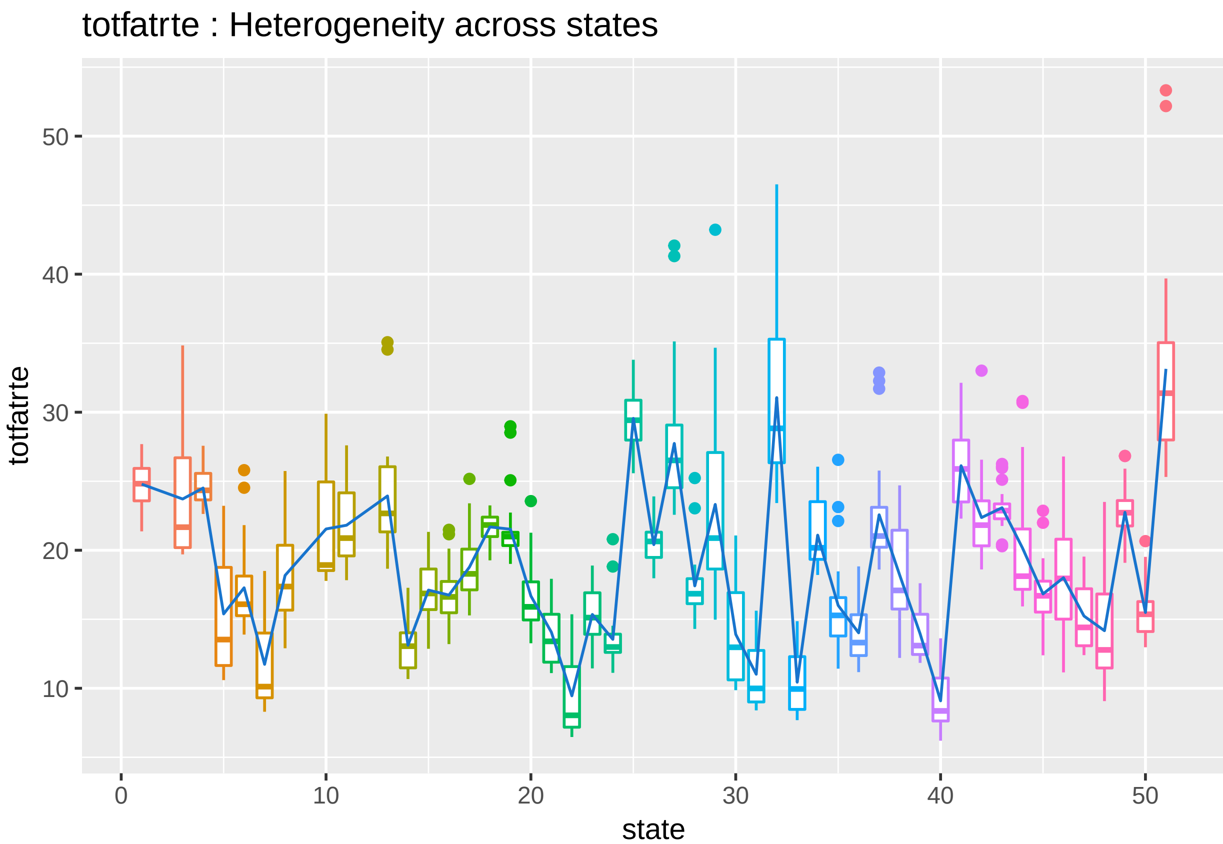
```
##
## Call:
## lm(formula = totfatrte ~ ., data = data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9302  -4.3468  -0.7305   3.7488  29.6498
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.72896    0.86712  19.293  < 2e-16 ***
## d80          8.76563    1.22629   7.148 1.54e-12 ***
## d81          6.94125    1.22629   5.660 1.90e-08 ***
## d82          4.21354    1.22629   3.436 0.000611 ***
## d83          3.42396    1.22629   2.792 0.005321 **
## d84          3.53854    1.22629   2.886 0.003979 **
## d85          3.12250    1.22629   2.546 0.011014 *
```

```
## d86          4.07146    1.22629   3.320 0.000927 ***
## d87          4.04583    1.22629   3.299 0.000999 ***
## d88          4.16271    1.22629   3.395 0.000710 ***
## d89          3.04333    1.22629   2.482 0.013213 *
## d90          2.77625    1.22629   2.264 0.023759 *
## d91          1.36583    1.22629   1.114 0.265596
## d92          0.42896    1.22629   0.350 0.726550
## d93          0.39875    1.22629   0.325 0.745112
## d94          0.42625    1.22629   0.348 0.728208
## d95          0.93958    1.22629   0.766 0.443712
## d96          0.64042    1.22629   0.522 0.601603
## d97          0.88167    1.22629   0.719 0.472302
## d98          0.53646    1.22629   0.437 0.661855
## d99          0.52146    1.22629   0.425 0.670745
## d00          0.09667    1.22629   0.079 0.937183
## d01          0.06375    1.22629   0.052 0.958549
## d02          0.30062    1.22629   0.245 0.806383
## d03          0.03458    1.22629   0.028 0.977506
## d04               NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.008 on 1175 degrees of freedom
## Multiple R-squared:  0.1276, Adjusted R-squared:  0.1098
## F-statistic: 7.164 on 24 and 1175 DF,  p-value: < 2.2e-16
```

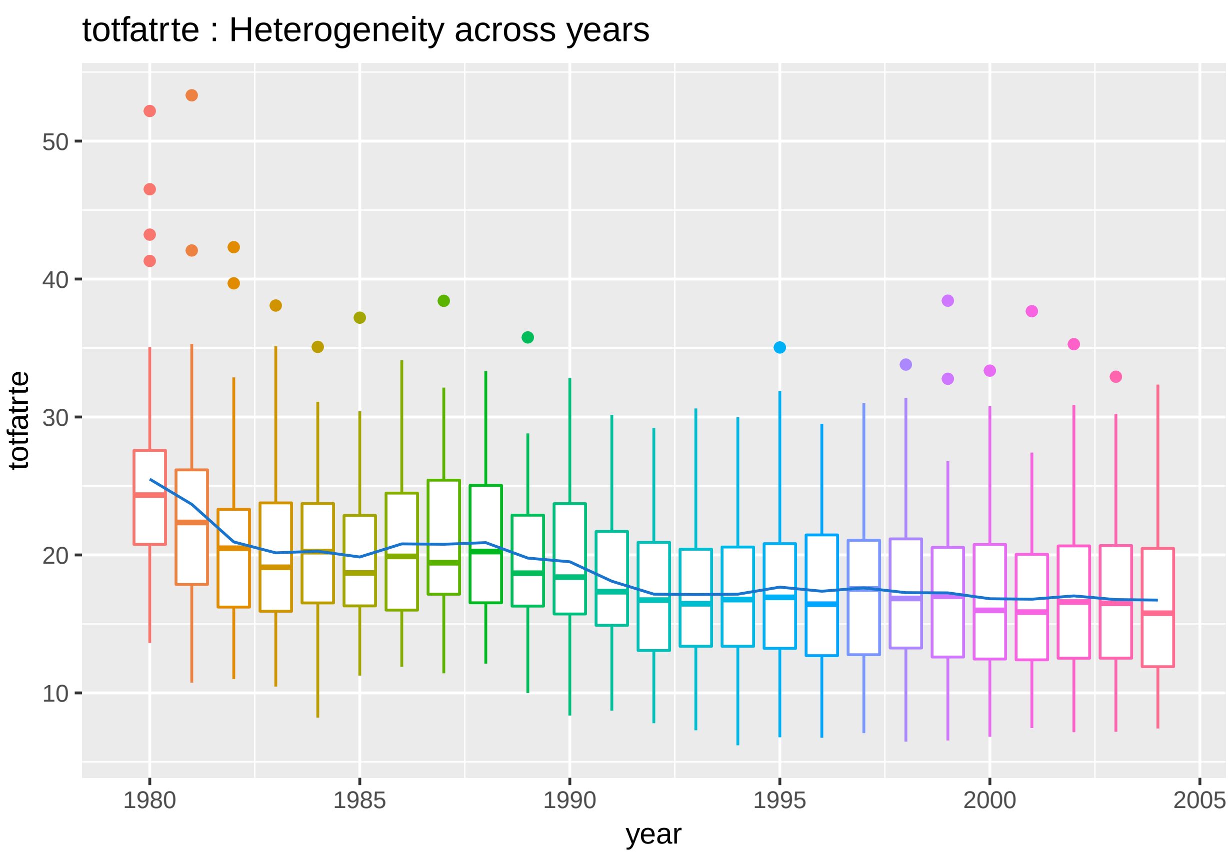
```r
library(grid)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
data1 <- data %>% group_by(state) %>% summarise(totfatrte_mean = mean(totfatrte))

# Plot boxplot of totfatrte for each state along
# with mean values.
ggplot(data, aes(y = totfatrte, x = state,
               group = state,color=as.factor(state))) +
    geom_boxplot() +
    ggtitle("totfatrte : Heterogeneity across states") +
    geom_line(color='dodgerblue3',data = data1, aes(group = 1, x=state, y=totfatrte_mean)) +
    theme(legend.position = "none")
```

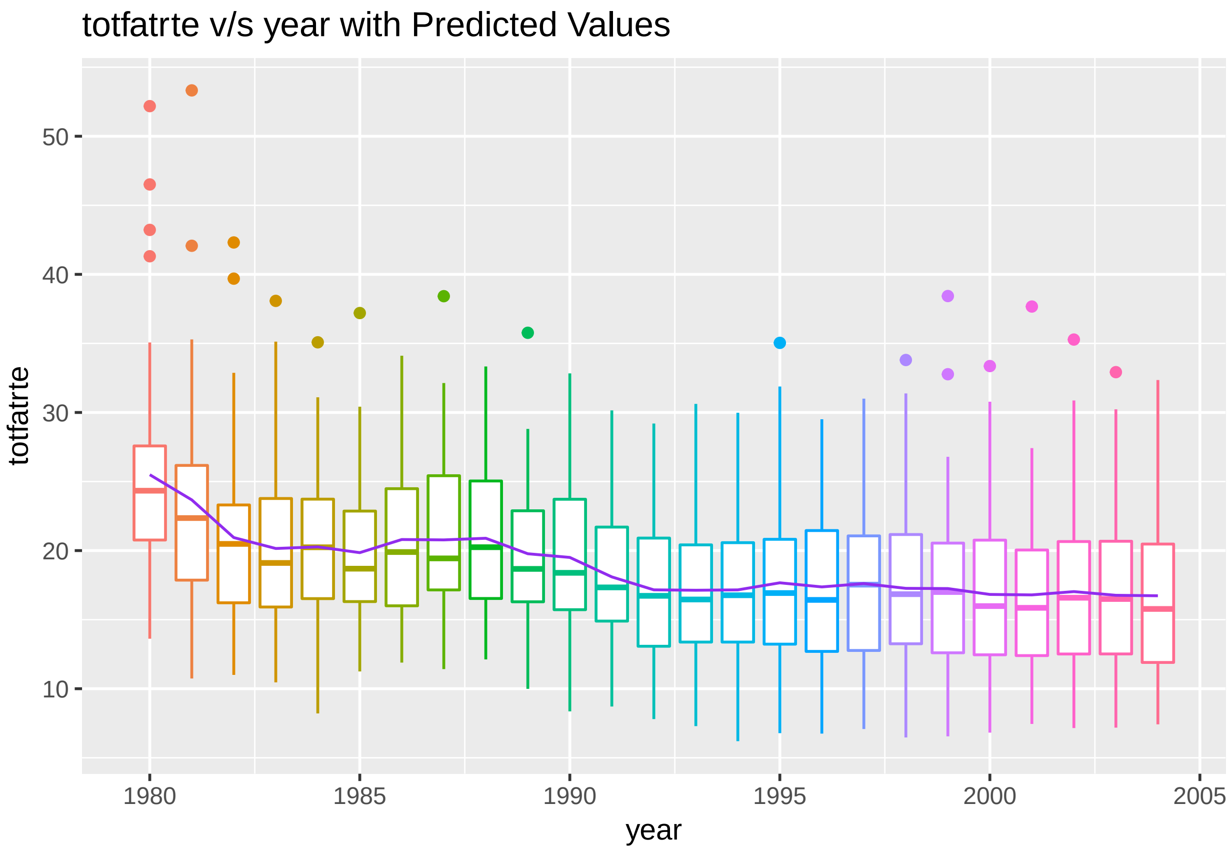



```r
data2 <- data %>% group_by(year) %>% summarise(totfatrte_mean = mean(totfatrte))

# Plot boxplot of totfatrte for each year along
# with mean values.
ggplot(data, aes(y = totfatrte, x = year,
               group = year,color=as.factor(year))) +
    geom_boxplot() +
    ggtitle("totfatrte : Heterogeneity across years") +
    geom_line(color='dodgerblue3',data = data2, aes(group = 1, x=year, y=totfatrte_mean)) +
    theme(legend.position = "none")
```




```
#grid.arrange(p1, p2, nrow = 1, ncol = 1)
```

```
# get predicted values
data$predicted_value=predict(lm_fit1)

# Plot boxplot of totfatrte for each year along
# with predicted values.
ggplot(data, aes(y = totfatrte, x = year,
               group = year,color=as.factor(year))) +
    geom_boxplot() +
    ggtitle("totfatrte v/s year with Predicted Values") +
    geom_line(color='purple2',data = data, aes(group = 1, x=year, y=predicted_value)) +
    #geom_line(color='dodgerblue3',data = data2, aes(group = 1, x=year, y=totfatrte_mean)) +
    theme(legend.position = "none")
```

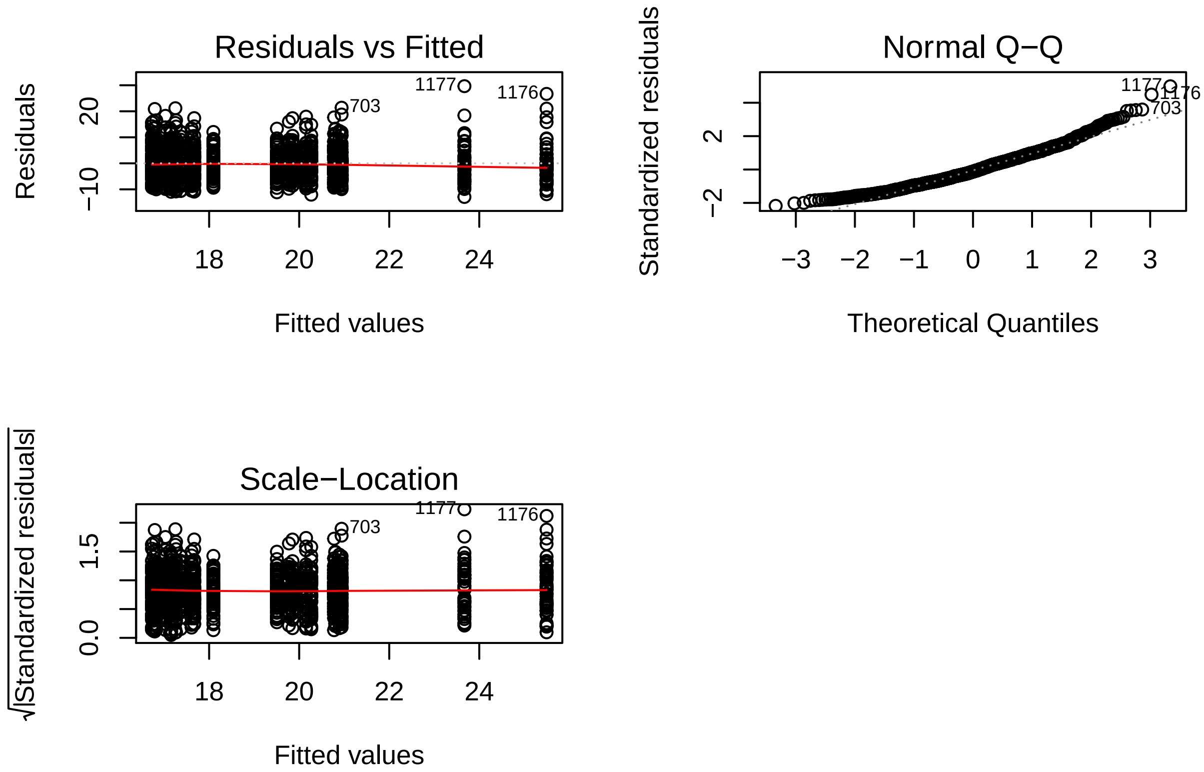


Looking at the significance of the date terms to *totfatrte* in the linear regression model, we observe that the early years 1980 - 1990 have the most significance. The later years after 1990 have the least contributon. From this we can conclude that driving did become safer over the years.

The predicted values from the regression model overlaid on the box plot of the *totfatrte* values over the years confirms the same i.e. totfatrte decreased over the years.

```
par(mfrow=c(2,2))
plot(lm_fit1)
```

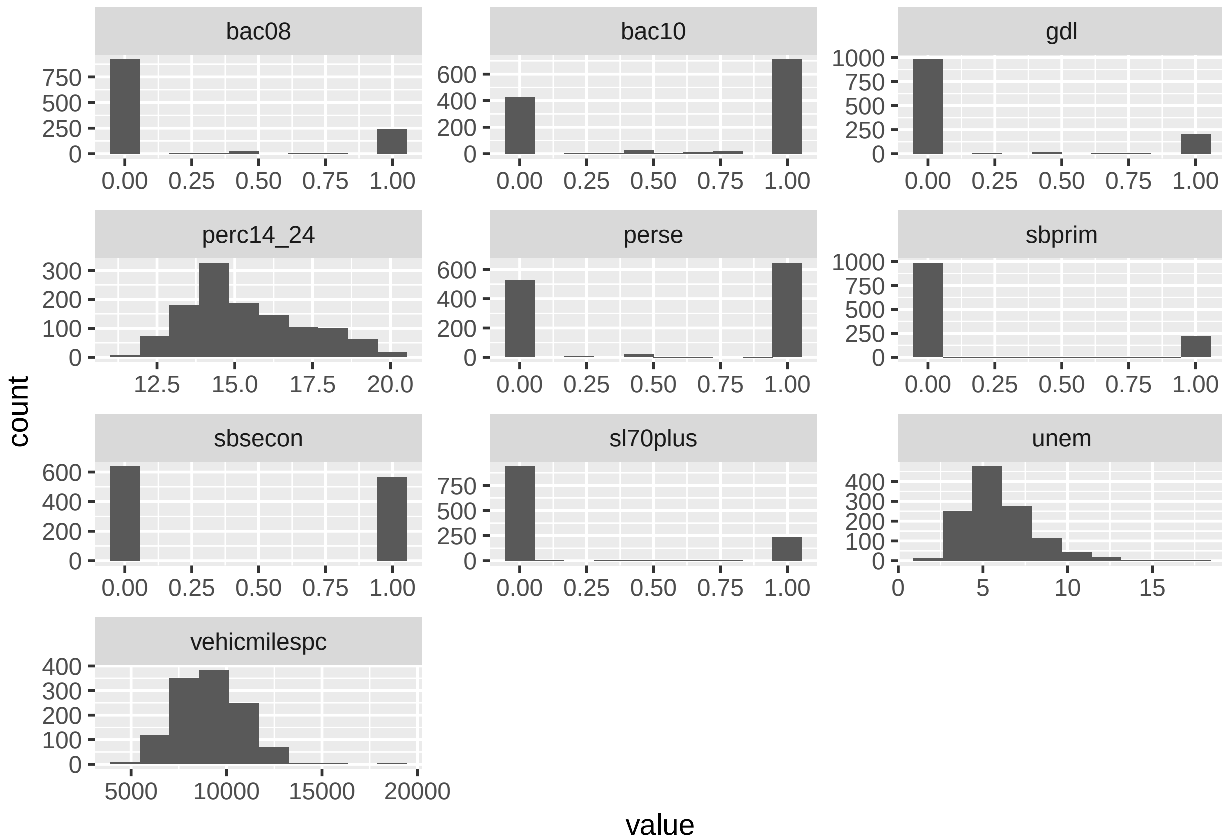
```
## hat values (leverages) are all = 0.02083333
##  and there are no factor predictors; no plot no. 5
```



Question 3

3. Expand your model in *Exercise 2* by adding variables *bac08, bac10, perse, sbprim, sbsecon, sl70plus, gdl, perc14_24, unem, vehicmilespc*, and perhaps *transformations of some or all of these variables*. Please 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 variables *bac8* and *bac10* defined? Interpret the coefficients on *bac8* and *bac10*. Do *per se laws* have a negative effect on the fatality rate? What about having a primary seat belt law? (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.)

```
regression_terms = c("bac08", "bac10", "perse", "sbprim", "sbsecon",
                     "sl70plus", "gdl", "perc14_24", "unem", "vehicmilespc")

ggplot(gather(subset(data, select=c(regression_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 3)
```

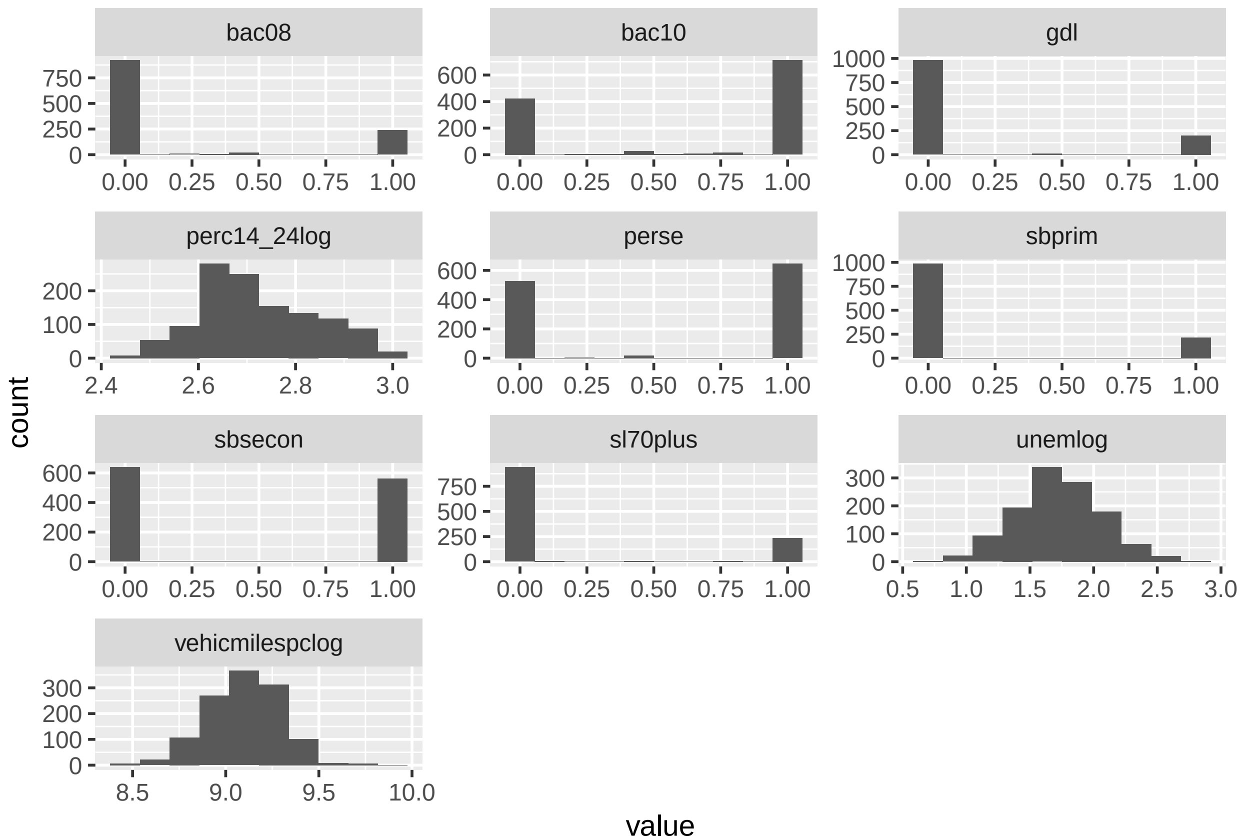



```
data %>% select(c(regression_terms))  %>% head()
```

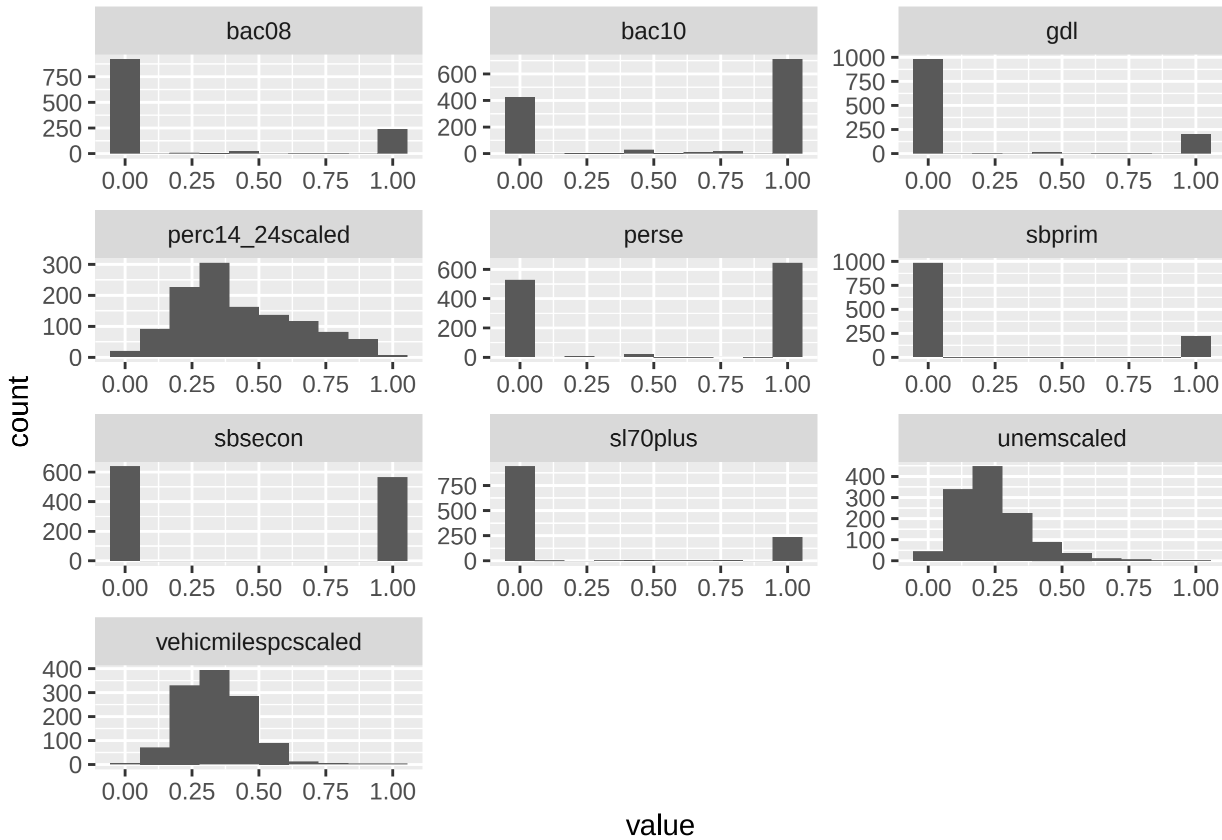
```
##   bac08 bac10 perse sbprim sbsecon sl70plus gdl perc14_24 unem
## 1     0     1     0      0       0        0   0      18.9  8.8
## 2     0     1     0      0       0        0   0      18.7 10.7
## 3     0     1     0      0       0        0   0      18.4 14.4
## 4     0     1     0      0       0        0   0      18.0 13.7
## 5     0     1     0      0       0        0   0      17.6 11.1
## 6     0     1     0      0       0        0   0      17.3  8.9
##   vehicmilespc
## 1     7543.874
## 2     7107.785
## 3     7606.622
## 4     7879.802
## 5     8333.562
## 6     8845.614
```

```
minmax <- function(x){(x-min(x))/(max(x)-min(x))}

data$vehicmilespclog <- log(data$vehicmilespc)
data$unemlog <- log(data$unem)
data$perc14_24log <- log(data$perc14_24)

data$vehicmilespcscaled <- minmax(data$vehicmilespc)
data$unemscaled <- minmax(data$unem)
```

```
data$perc14_24scaled <- minmax(data$perc14_24)

regression_terms = c("bac08", "bac10", "perse", "sbprim", "sbsecon",
                     "sl70plus", "gdl", "perc14_24log", "unemlog", "vehicmilespclog")

ggplot(gather(subset(data, select=c(regression_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 3)
```




```
minmax <- function(x){(x-min(x))/(max(x)-min(x))}

data$vehicmilespcscaled <- minmax(data$vehicmilespc)
data$unemscaled <- minmax(data$unem)
data$perc14_24scaled <- minmax(data$perc14_24)

regression_terms = c("bac08", "bac10", "perse", "sbprim", "sbsecon",
                     "sl70plus", "gdl", "perc14_24scaled", "unemscaled", "vehicmilespcscaled")

ggplot(gather(subset(data, select=c(regression_terms))),
    aes(value)) +
    geom_histogram(bins = 10) +
    facet_wrap(~ key, scales = 'free',ncol = 3)
```




```r
data3 <- data %>% select(c("totfatrte",c(regression_terms,date_terms)))
lm_fit3 = lm(totfatrte ~ ., data=data3)
summary(lm_fit3)
```

```
##
## Call:
## lm(formula = totfatrte ~ ., data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9160  -2.7384  -0.2778   2.2859  21.4203
##
## Coefficients: (1 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -3.31516    1.21135  -2.737 0.006299 **
## bac08               -2.49848    0.53751  -4.648 3.73e-06 ***
## bac10               -1.41757    0.39633  -3.577 0.000362 ***
## perse               -0.62011    0.29820  -2.079 0.037791 *
## sbprim              -0.07533    0.49078  -0.153 0.878032
## sbsecon              0.06728    0.42930   0.157 0.875492
## sl70plus             3.34791    0.44517   7.521 1.09e-13 ***
## gdl                 -0.42691    0.52691  -0.810 0.417978
## perc14_24scaled      1.21768    1.05504   1.154 0.248675
## unemscaled          11.96144    1.23091   9.718  < 2e-16 ***
```

```
## vehicmilespcscaled 41.00853    1.33126  30.804  < 2e-16 ***
## d80                16.71127    1.38697  12.049  < 2e-16 ***
## d81                14.53579    1.36963  10.613  < 2e-16 ***
## d82                10.11530    1.36766   7.396 2.67e-13 ***
## d83                 9.31458    1.32894   7.009 4.05e-12 ***
## d84                10.86088    1.26403   8.592  < 2e-16 ***
## d85                10.22802    1.22683   8.337  < 2e-16 ***
## d86                10.85848    1.16023   9.359  < 2e-16 ***
## d87                10.34388    1.12027   9.233  < 2e-16 ***
## d88                10.11969    1.08537   9.324  < 2e-16 ***
## d89                 8.64031    1.06449   8.117 1.21e-15 ***
## d90                 7.75260    1.05368   7.358 3.52e-13 ***
## d91                 5.64272    1.04916   5.378 9.08e-08 ***
## d92                 3.83288    1.04551   3.666 0.000257 ***
## d93                 3.98055    1.03708   3.838 0.000131 ***
## d94                 4.34644    1.01967   4.263 2.18e-05 ***
## d95                 4.75872    1.01399   4.693 3.01e-06 ***
## d96                 2.83490    0.98023   2.892 0.003898 **
## d97                 2.45289    0.96149   2.551 0.010864 *
## d98                 1.66960    0.93502   1.786 0.074420 .
## d99                 1.62073    0.89626   1.808 0.070813 .
## d00                 1.26733    0.87920   1.441 0.149724
## d01                 0.52756    0.85108   0.620 0.535463
## d02                -0.01308    0.83829  -0.016 0.987557
## d03                -0.31004    0.83130  -0.373 0.709252
## d04                      NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.046 on 1165 degrees of freedom
## Multiple R-squared:  0.6078, Adjusted R-squared:  0.5963
## F-statistic:  53.1 on 34 and 1165 DF,  p-value: < 2.2e-16
```

```r
data4 <- data %>% select(c("totfatrte",c(regression_terms,"year","state")))
plm_data <-  pdata.frame(data4, index=c("state","year"))

plm_fit1 <- plm(totfatrte~year+bac08+bac10+perse+sbprim+sbsecon+
                     sl70plus+gdl+perc14_24scaled+unemscaled+vehicmilespcscaled, data = plm_da
summary(plm_fit1)
```

```
## Pooling Model
##
## Call:
## plm(formula = totfatrte ~ year + bac08 + bac10 + perse + sbprim +
##     sbsecon + sl70plus + gdl + perc14_24scaled + unemscaled +
##     vehicmilespcscaled, data = plm_data, effect = "individual",
##     model = "pooling")
##
```

```
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -14.91602  -2.73839  -0.27779   2.28591  21.42027
##
## Coefficients:
##                      Estimate Std. Error  t-value  Pr(>|t|)
## (Intercept)         13.396109   1.145957  11.6899 < 2.2e-16 ***
## year1981            -2.175479   0.827611  -2.6286 0.0086859 **
## year1982            -6.595970   0.853403  -7.7290 2.330e-14 ***
## year1983            -7.396690   0.869019  -8.5115 < 2.2e-16 ***
## year1984            -5.850394   0.876336  -6.6760 3.792e-11 ***
## year1985            -6.483252   0.894798  -7.2455 7.820e-13 ***
## year1986            -5.852796   0.930669  -6.2888 4.516e-10 ***
## year1987            -6.367393   0.966960  -6.5850 6.869e-11 ***
## year1988            -6.591578   1.013708  -6.5024 1.170e-10 ***
## year1989            -8.070967   1.052622  -7.6675 3.684e-14 ***
## year1990            -8.958670   1.076953  -8.3185 2.463e-16 ***
## year1991           -11.068552   1.101162 -10.0517 < 2.2e-16 ***
## year1992           -12.878398   1.122518 -11.4728 < 2.2e-16 ***
## year1993           -12.730718   1.136283 -11.2038 < 2.2e-16 ***
## year1994           -12.364833   1.157223 -10.6849 < 2.2e-16 ***
## year1995           -11.952549   1.183602 -10.0985 < 2.2e-16 ***
## year1996           -13.876377   1.223339 -11.3430 < 2.2e-16 ***
## year1997           -14.258378   1.249804 -11.4085 < 2.2e-16 ***
## year1998           -15.041676   1.265488 -11.8861 < 2.2e-16 ***
## year1999           -15.090547   1.284308 -11.7499 < 2.2e-16 ***
## year2000           -15.443946   1.305337 -11.8314 < 2.2e-16 ***
## year2001           -16.183715   1.334040 -12.1314 < 2.2e-16 ***
## year2002           -16.724350   1.348033 -12.4065 < 2.2e-16 ***
## year2003           -17.021308   1.359468 -12.5206 < 2.2e-16 ***
## year2004           -16.711273   1.386970 -12.0488 < 2.2e-16 ***
## bac08               -2.498483   0.537505  -4.6483 3.729e-06 ***
## bac10               -1.417565   0.396328  -3.5768 0.0003622 ***
## perse               -0.620108   0.298202  -2.0795 0.0377907 *
## sbprim              -0.075335   0.490785  -0.1535 0.8780318
## sbsecon              0.067280   0.429300   0.1567 0.8754918
## sl70plus             3.347914   0.445170   7.5205 1.086e-13 ***
## gdl                 -0.426911   0.526906  -0.8102 0.4179781
## perc14_24scaled      1.217677   1.055042   1.1542 0.2486752
## unemscaled          11.961436   1.230911   9.7176 < 2.2e-16 ***
## vehicmilespcscaled  41.008531   1.331262  30.8042 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    48612
## Residual Sum of Squares: 19067
```

```
## R-Squared:      0.60778
## Adj. R-Squared: 0.59633
## F-statistic: 53.0957 on 34 and 1165 DF, p-value: < 2.22e-16
```

```r
data2 <- data %>% select(c("totfatrte",c(regression_terms,date_terms)))
lm_fit2 = lm(totfatrte ~ ., data=data2)
summary(lm_fit2)
```

```
##
## Call:
## lm(formula = totfatrte ~ ., data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9160  -2.7384  -0.2778   2.2859  21.4203
##
## Coefficients: (1 not defined because of singularities)
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -3.31516    1.21135  -2.737 0.006299 **
## bac08              -2.49848    0.53751  -4.648 3.73e-06 ***
## bac10              -1.41757    0.39633  -3.577 0.000362 ***
## perse              -0.62011    0.29820  -2.079 0.037791 *
## sbprim             -0.07533    0.49078  -0.153 0.878032
## sbsecon             0.06728    0.42930   0.157 0.875492
## sl70plus            3.34791    0.44517   7.521 1.09e-13 ***
## gdl                -0.42691    0.52691  -0.810 0.417978
## perc14_24scaled     1.21768    1.05504   1.154 0.248675
## unemscaled         11.96144    1.23091   9.718  < 2e-16 ***
## vehicmilespcscaled 41.00853    1.33126  30.804  < 2e-16 ***
## d80                16.71127    1.38697  12.049  < 2e-16 ***
## d81                14.53579    1.36963  10.613  < 2e-16 ***
## d82                10.11530    1.36766   7.396 2.67e-13 ***
## d83                 9.31458    1.32894   7.009 4.05e-12 ***
## d84                10.86088    1.26403   8.592  < 2e-16 ***
## d85                10.22802    1.22683   8.337  < 2e-16 ***
## d86                10.85848    1.16023   9.359  < 2e-16 ***
## d87                10.34388    1.12027   9.233  < 2e-16 ***
## d88                10.11969    1.08537   9.324  < 2e-16 ***
## d89                 8.64031    1.06449   8.117 1.21e-15 ***
## d90                 7.75260    1.05368   7.358 3.52e-13 ***
## d91                 5.64272    1.04916   5.378 9.08e-08 ***
## d92                 3.83288    1.04551   3.666 0.000257 ***
## d93                 3.98055    1.03708   3.838 0.000131 ***
## d94                 4.34644    1.01967   4.263 2.18e-05 ***
## d95                 4.75872    1.01399   4.693 3.01e-06 ***
## d96                 2.83490    0.98023   2.892 0.003898 **
## d97                 2.45289    0.96149   2.551 0.010864 *
## d98                 1.66960    0.93502   1.786 0.074420 .
```

```
## d99                 1.62073    0.89626   1.808 0.070813 .
## d00                 1.26733    0.87920   1.441 0.149724
## d01                 0.52756    0.85108   0.620 0.535463
## d02                -0.01308    0.83829  -0.016 0.987557
## d03                -0.31004    0.83130  -0.373 0.709252
## d04                      NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.046 on 1165 degrees of freedom
## Multiple R-squared:  0.6078, Adjusted R-squared:  0.5963
## F-statistic:  53.1 on 34 and 1165 DF,  p-value: < 2.2e-16
```

Question 4

4. Reestimate the model from *Exercise 3* using a fixed effects (at the state level) model. How do the coefficients on *bac08, bac10, perse, and sbprim* compare with the pooled OLS estimates? Which set of estimates do you think is more reliable? What assumptions are needed in each of these models? Are these assumptions reasonable in the current context?

```
plm_fit1 <- plm(totfatrte~year+bac08+bac10+perse+sbprim+sbsecon+
                     sl70plus+gdl+perc14_24scaled+unemscaled+vehicmilespcscaled, data = plm_da
summary(plm_fit1)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = totfatrte ~ year + bac08 + bac10 + perse + sbprim +
##     sbsecon + sl70plus + gdl + perc14_24scaled + unemscaled +
##     vehicmilespcscaled, data = plm_data, model = "within", index = c("year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -8.4273592 -1.0258600 -0.0029547  0.9572345 14.8109310
##
## Coefficients:
##                     Estimate Std. Error  t-value  Pr(>|t|)
## year1981           -1.511071   0.413215  -3.6569 0.0002672 ***
## year1982           -3.025496   0.442431  -6.8383 1.316e-11 ***
## year1983           -3.503601   0.456577  -7.6736 3.628e-14 ***
## year1984           -4.259361   0.464943  -9.1610 < 2.2e-16 ***
## year1985           -4.726793   0.485470  -9.7365 < 2.2e-16 ***
## year1986           -3.661185   0.517698  -7.0721 2.686e-12 ***
## year1987           -4.305788   0.555329  -7.7536 2.001e-14 ***
## year1988           -4.767121   0.601556  -7.9246 5.501e-15 ***
## year1989           -6.129973   0.640191  -9.5752 < 2.2e-16 ***
```

```
## year1990           -6.229738   0.664851  -9.3701 < 2.2e-16 ***
## year1991           -6.917140   0.681954 -10.1431 < 2.2e-16 ***
## year1992           -7.774172   0.702886 -11.0604 < 2.2e-16 ***
## year1993           -8.094109   0.715947 -11.3055 < 2.2e-16 ***
## year1994           -8.504217   0.734109 -11.5844 < 2.2e-16 ***
## year1995           -8.255402   0.756236 -10.9164 < 2.2e-16 ***
## year1996           -8.606619   0.795950 -10.8130 < 2.2e-16 ***
## year1997           -8.707817   0.819757 -10.6224 < 2.2e-16 ***
## year1998           -9.349240   0.833735 -11.2137 < 2.2e-16 ***
## year1999           -9.474891   0.843991 -11.2263 < 2.2e-16 ***
## year2000           -9.991860   0.856064 -11.6719 < 2.2e-16 ***
## year2001           -9.631217   0.872554 -11.0380 < 2.2e-16 ***
## year2002           -8.906730   0.882053 -10.0977 < 2.2e-16 ***
## year2003           -8.936503   0.889947 -10.0416 < 2.2e-16 ***
## year2004           -9.339361   0.911070 -10.2510 < 2.2e-16 ***
## bac08              -1.437221   0.394212  -3.6458 0.0002788 ***
## bac10              -1.062668   0.268838  -3.9528 8.208e-05 ***
## perse              -1.151617   0.233987  -4.9217 9.867e-07 ***
## sbprim             -1.227400   0.342715  -3.5814 0.0003564 ***
## sbsecon            -0.349708   0.252171  -1.3868 0.1657826
## sl70plus           -0.062533   0.269311  -0.2322 0.8164283
## gdl                -0.411776   0.292574  -1.4074 0.1595790
## perc14_24scaled     1.609246   0.817857   1.9676 0.0493567 *
## unemscaled         -9.035071   0.957141  -9.4397 < 2.2e-16 ***
## vehicmilespcscaled 13.177679   1.556613   8.4656 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12134
## Residual Sum of Squares: 4535.3
## R-Squared:      0.62624
## Adj. R-Squared: 0.59916
## F-statistic: 55.0943 on 34 and 1118 DF, p-value: < 2.22e-16
```

- FE removes the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable.
- Those time-invariant characteristics are unique to the individual and should not be correlated with other individual characteristics.

Question 5

5. Would you prefer to use a random effects model instead of the fixed effects model you built in *Exercise 4*? Please explain.

```
plm_fit2 <- plm(totfatrte~year+bac08+bac10+perse+sbprim+sbsecon+
                     sl70plus+gdl+perc14_24scaled+unemscaled+vehicmilespcscaled, data = plm_da
summary(plm_fit2)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = totfatrte ~ year + bac08 + bac10 + perse + sbprim +
##     sbsecon + sl70plus + gdl + perc14_24scaled + unemscaled +
##     vehicmilespcscaled, data = plm_data, model = "random", index = c("state",
##     "year"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Effects:
##                 var std.dev share
## idiosyncratic 4.057   2.014 0.328
## individual    8.294   2.880 0.672
## theta: 0.8615
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -8.25582 -1.15221 -0.15787  0.93086 16.45691
##
## Coefficients:
##                      Estimate Std. Error  z-value  Pr(>|z|)
## (Intercept)         23.505008   0.982620  23.9207 < 2.2e-16 ***
## year1981            -1.548874   0.428297  -3.6164 0.0002988 ***
## year1982            -3.243321   0.457724  -7.0858 1.383e-12 ***
## year1983            -3.744747   0.472120  -7.9318 2.161e-15 ***
## year1984            -4.372875   0.480638  -9.0981 < 2.2e-16 ***
## year1985            -4.860931   0.501363  -9.6954 < 2.2e-16 ***
## year1986            -3.829544   0.534160  -7.1693 7.539e-13 ***
## year1987            -4.501403   0.572131  -7.8678 3.610e-15 ***
## year1988            -4.981938   0.618874  -8.0500 8.279e-16 ***
## year1989            -6.371318   0.657971  -9.6833 < 2.2e-16 ***
## year1990            -6.535657   0.682788  -9.5720 < 2.2e-16 ***
## year1991            -7.302698   0.700302 -10.4279 < 2.2e-16 ***
## year1992            -8.238956   0.721261 -11.4230 < 2.2e-16 ***
## year1993            -8.541762   0.734487 -11.6296 < 2.2e-16 ***
## year1994            -8.918308   0.752970 -11.8442 < 2.2e-16 ***
## year1995            -8.676932   0.775406 -11.1902 < 2.2e-16 ***
## year1996            -9.096863   0.815728 -11.1518 < 2.2e-16 ***
## year1997            -9.220296   0.839844 -10.9786 < 2.2e-16 ***
## year1998            -9.892173   0.853804 -11.5860 < 2.2e-16 ***
## year1999           -10.031557   0.864259 -11.6071 < 2.2e-16 ***
## year2000           -10.549021   0.876674 -12.0330 < 2.2e-16 ***
## year2001           -10.273651   0.893359 -11.5000 < 2.2e-16 ***
## year2002            -9.637586   0.902780 -10.6755 < 2.2e-16 ***
## year2003            -9.682820   0.910895 -10.6300 < 2.2e-16 ***
## year2004           -10.054258   0.932541 -10.7816 < 2.2e-16 ***
```

```
## bac08               -1.569317   0.403835  -3.8860 0.0001019 ***
## bac10               -1.138042   0.276041  -4.1227 3.744e-05 ***
## perse               -1.093281   0.238852  -4.5772 4.712e-06 ***
## sbprim              -1.176078   0.351436  -3.3465 0.0008184 ***
## sbsecon             -0.347578   0.260245  -1.3356 0.1816862
## sl70plus             0.029969   0.277721   0.1079 0.9140655
## gdl                 -0.385244   0.302487  -1.2736 0.2028095
## perc14_24scaled      1.693737   0.836030   2.0259 0.0427722 *
## unemscaled          -7.779559   0.977059  -7.9622 1.690e-15 ***
## vehicmilespcscaled  16.463335   1.539593  10.6933 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12834
## Residual Sum of Squares: 5078.6
## R-Squared:      0.60429
## Adj. R-Squared: 0.59274
## Chisq: 1779.05 on 34 DF, p-value: < 2.22e-16
```

```r
plm_fit2 <- plm(totfatrte~year+bac08+bac10+perse+sbprim+sbsecon+
                     sl70plus+gdl+perc14_24scaled+unemscaled+vehicmilespcscaled, data = plm_da
summary(plm_fit2)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = totfatrte ~ year + bac08 + bac10 + perse + sbprim +
##     sbsecon + sl70plus + gdl + perc14_24scaled + unemscaled +
##     vehicmilespcscaled, data = plm_data, model = "random", index = c("state"))
##
## Balanced Panel: n = 48, T = 25, N = 1200
##
## Effects:
##                 var std.dev share
## idiosyncratic 4.057   2.014 0.328
## individual    8.294   2.880 0.672
## theta: 0.8615
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -8.25582 -1.15221 -0.15787  0.93086 16.45691
##
## Coefficients:
##                      Estimate Std. Error  z-value  Pr(>|z|)
## (Intercept)         23.505008   0.982620  23.9207 < 2.2e-16 ***
## year1981            -1.548874   0.428297  -3.6164 0.0002988 ***
## year1982            -3.243321   0.457724  -7.0858 1.383e-12 ***
```

```
## year1983           -3.744747   0.472120  -7.9318 2.161e-15 ***
## year1984           -4.372875   0.480638  -9.0981 < 2.2e-16 ***
## year1985           -4.860931   0.501363  -9.6954 < 2.2e-16 ***
## year1986           -3.829544   0.534160  -7.1693 7.539e-13 ***
## year1987           -4.501403   0.572131  -7.8678 3.610e-15 ***
## year1988           -4.981938   0.618874  -8.0500 8.279e-16 ***
## year1989           -6.371318   0.657971  -9.6833 < 2.2e-16 ***
## year1990           -6.535657   0.682788  -9.5720 < 2.2e-16 ***
## year1991           -7.302698   0.700302 -10.4279 < 2.2e-16 ***
## year1992           -8.238956   0.721261 -11.4230 < 2.2e-16 ***
## year1993           -8.541762   0.734487 -11.6296 < 2.2e-16 ***
## year1994           -8.918308   0.752970 -11.8442 < 2.2e-16 ***
## year1995           -8.676932   0.775406 -11.1902 < 2.2e-16 ***
## year1996           -9.096863   0.815728 -11.1518 < 2.2e-16 ***
## year1997           -9.220296   0.839844 -10.9786 < 2.2e-16 ***
## year1998           -9.892173   0.853804 -11.5860 < 2.2e-16 ***
## year1999          -10.031557   0.864259 -11.6071 < 2.2e-16 ***
## year2000          -10.549021   0.876674 -12.0330 < 2.2e-16 ***
## year2001          -10.273651   0.893359 -11.5000 < 2.2e-16 ***
## year2002           -9.637586   0.902780 -10.6755 < 2.2e-16 ***
## year2003           -9.682820   0.910895 -10.6300 < 2.2e-16 ***
## year2004          -10.054258   0.932541 -10.7816 < 2.2e-16 ***
## bac08              -1.569317   0.403835  -3.8860 0.0001019 ***
## bac10              -1.138042   0.276041  -4.1227 3.744e-05 ***
## perse              -1.093281   0.238852  -4.5772 4.712e-06 ***
## sbprim             -1.176078   0.351436  -3.3465 0.0008184 ***
## sbsecon            -0.347578   0.260245  -1.3356 0.1816862
## sl70plus            0.029969   0.277721   0.1079 0.9140655
## gdl                -0.385244   0.302487  -1.2736 0.2028095
## perc14_24scaled     1.693737   0.836030   2.0259 0.0427722 *
## unemscaled         -7.779559   0.977059  -7.9622 1.690e-15 ***
## vehicmilespcscaled  16.463335   1.539593  10.6933 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12834
## Residual Sum of Squares: 5078.6
## R-Squared:      0.60429
## Adj. R-Squared: 0.59274
## Chisq: 1779.05 on 34 DF, p-value: < 2.22e-16
```

```r
phtest(plm_fit1,plm_fit2)
```

```
##
##  Hausman Test
##
## data:  totfatrte ~ year + bac08 + bac10 + perse + sbprim + sbsecon +  ...
## chisq = 148.69, df = 34, p-value = 2.727e-16
```

```
## alternative hypothesis: one model is inconsistent
```

Let's run the Hausman test where the null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects. A small p-value (typically <= 0.05) indicates strong evidence against the null hypothesis hence we can conclude that the fixed effects model is a better fit.

Question 6

6. Suppose that *vehicmilespc*, the number of miles driven per capita, increases by $1,000$. Using the FE estimates, what is the estimated effect on *totfatrte*? Please interpret the estimate.

```r
min_vehicmilespc <- min(data$vehicmilespc)
max_vehicmilespc <- max(data$vehicmilespc)

totfatrte_change = (1000*13.177679)/(max_vehicmilespc-min_vehicmilespc)
totfatrte_change
```

```
## [1] 0.9400519
```

```r
total_pop <- data %>% summarise(sum_state_pop = sum(statepop[year==2004])) %>% pull(sum_state_
total_pop * totfatrte_change / 100000
```

```
## [1] 2737.487
```

Thus, a $1,000$ increase in *vehicmilespc* would increase *totfatrte* by 0.94. A more intuitive way to express this would be that for the next year i.e. 2005, if the *vehicmilespc* driven would increase by $1,000$, then the total fatalities is projected to increase by 2737.487.

Question 7

7. If there is serial correlation or heteroskedasticity in the idiosyncratic errors of the model, what would be the consequences on the estimators and their standard errors?

```r
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
bptest(plm_fit1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  plm_fit1
## BP = 150.07, df = 34, p-value < 2.2e-16
```

The null hypothesis of the Breusch-Pagan test is homoscedasticity i.e. variance does not depend on auxiliary regressors. In this case, the null hypothesis is rejected as the p-value is small.