# ARE 213 PS 2a

S. Sung, H. Husain, T. Woolley, A. Watt

2021-11-08

## Contents



## Packages

```
library(tidyverse)
library(haven)
library(plm)
library(lmtest)
library(stargazer)
library(ggplot2)
library(gridExtra)
library(grid)
library(gtable)
```

# Problem 1

**Question 10.3 from Wooldridge: For $T = 2$ consider the standard unobserved effects model:**

$$y_{it} = \alpha + x_{it}\beta + c_i + u_{it} \tag{1}$$

**Let $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ represent the fixed effects and first differences estimators respectively.**

## Part (a)

**Show that $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical. Hint: it may be easier to write $\hat{\beta}_{FE}$ as the "within estimator" rather than the fixed effects estimator.**

Writing $\hat{\beta}_{FE}$ as the within estimator, $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are given by

$$\hat{\beta}_{FD} = (\Delta X'\Delta X)^{-1}(\Delta X'\Delta y) \qquad \text{and} \qquad \hat{\beta}_{FE} = \left(\ddot{X}'\ddot{X}\right)^{-1}\left(\ddot{X}'\ddot{y}\right)$$

Expanding the inner products, we have

$$\hat{\beta}_{FD} = \left(\sum_i \sum_t \Delta X_{it}'\Delta X_{it}\right)^{-1}\left(\sum_i \sum_t \Delta X_{it}'\Delta y_{it}\right)$$

and

$$\hat{\beta}_{FE} = \left(\sum_i \sum_t \ddot{X}_{it}'\ddot{X}_{it}\right)^{-1}\left(\sum_i \sum_t \ddot{X}_{it}'\ddot{y}_{it}\right)$$

Since there are only two periods, $\hat{\beta}_{FD}$ simplifies to

$$\hat{\beta}_{FD} = \left(\sum_i \Delta X_i'\Delta X_i\right)^{-1}\left(\sum_i \Delta X_i'\Delta y_i\right)$$

where

$$\Delta X_i \equiv X_{i2} - X_{i1} \qquad \text{and} \qquad \Delta y_i \equiv y_{i2} - y_{i1}$$

Now we note that

$$\ddot{X}_{i1} = X_{i1} - \frac{1}{2}(X_{i1} + X_{i2}) = \frac{1}{2}(X_{i1} - X_{i2}) = -\frac{1}{2}\Delta X_i$$

and similiarly

$$\ddot{X}_{i2} = \frac{1}{2}\Delta X_i, \quad \ddot{y}_{i1} = -\frac{1}{2}\Delta y_i \quad \ddot{y}_{i2} = \frac{1}{2}\Delta y_i$$

Then, $\hat{\beta}_{FE}$ becomes

$$\begin{aligned}
\hat{\beta}_{FE} &= \left(\sum_i \sum_t \ddot{X}_{it}' \ddot{X}_{it}\right)^{-1} \left(\sum_i \sum_t \ddot{X}_{it}' \ddot{y}_{it}\right) \\
&= \left(\sum_i \frac{1}{4} \Delta X_i' \Delta X_i + \frac{1}{4} \Delta X_i' \Delta X_i\right)^{-1} \left(\sum_i \frac{1}{4} \Delta X_i' \Delta y_i + \frac{1}{4} \Delta X_i' \Delta y_i\right) \\
&= \left(\frac{1}{2} \sum_i \Delta X_i' \Delta X_i\right)^{-1} \left(\frac{1}{2} \sum_i \Delta X_i' \Delta y_i\right) \\
&= \left(\sum_i \Delta X_i' \Delta X_i\right)^{-1} \left(\sum_i \Delta X_i' \Delta y_i\right) \\
&= \hat{\beta}_{FD}
\end{aligned}$$

So $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical.

## Part (b)

**Show that the standard errors of $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical. If you wish, you may assume that $x_{it}$ is a scalar (i.e. there is only one regressor) and ignore any degree of freedom corrections. You are not clustering the standard errors in this problem.**

The standard errors are estimates of the square root of the asymptotic variances of our estimators, so WLOG, we can compare the asymptotic variances. The asymptotic variances of our estimators are

$$\widehat{Avar(\hat{\beta}_{FE})} = \hat{\sigma}^2_{u,FE} \left(\ddot{X}'\ddot{X}\right)^{-1} \qquad \text{and} \qquad \widehat{Avar(\hat{\beta}_{FE})} = \hat{\sigma}^2_{u,FD} \left(\Delta X' \Delta X\right)^{-1}$$

where $\hat{\sigma}^2_{u,FE}$ and $\hat{\sigma}^2_{u,FD}$ are estimated from the residuals of the corresponding regressions and using the correct degrees of freedom:

$$\hat{\sigma}^2_{u,FE} = \frac{\sum_i \sum_t \widehat{\ddot{u}}^2_{it}}{N(T-1)-K} = \frac{\sum_i \sum_t \widehat{\ddot{u}}^2_{it}}{N-K} \qquad \text{and} \qquad \hat{\sigma}^2_{u,FD} = \frac{\sum_i \widehat{\Delta u_i}^2}{N(T-1)-K} = \frac{\sum_i \widehat{\Delta u_i}^2}{N-K}$$

Let $\hat{\beta} := \hat{\beta}_{FD} = \hat{\beta}_{FE}$. Then, from part (a), we can find the relationship between $\widehat{\Delta u_i}$ and $\widehat{\ddot{u}}_{it}$:

$$\begin{aligned}
\widehat{\ddot{u}}^2_{it} &= (\ddot{y}_{it} - \ddot{X}_{it}\hat{\beta})^2 \\
&= \left((-1)^t \left(\frac{1}{2}\Delta y_i - \frac{1}{2}\Delta X_i \hat{\beta}\right)\right)^2 \\
&= \frac{1}{4}\left(\Delta y_i - \Delta X_i \hat{\beta}\right)^2 \\
&= \frac{1}{4}\widehat{\Delta u_i}^2
\end{aligned}$$

So the estimated error variances are related by

$$\begin{aligned}
\hat{\sigma}^2_{u,FE} &= \frac{\sum_i \sum_t \widehat{\ddot{u}}^2_{it}}{N-K} \\
&= \frac{\sum_i \sum_t \frac{1}{4}\widehat{\Delta u_i}^2}{N-K} \\
&= \frac{\sum_i \frac{1}{2}\widehat{\Delta u_i}^2}{N-K} \\
&= \frac{1}{2}\frac{\sum_i \widehat{\Delta u_i}^2}{N-K} \\
&= \frac{1}{2}\hat{\sigma}^2_{u,FD}
\end{aligned}$$

We know from part (a) that

$$\begin{aligned}
\left(\sum_i \sum_t \ddot{X}'_{it}\ddot{X}_{it}\right)^{-1} &= \left(\frac{1}{2}\sum_i \Delta X'_i \Delta X_i\right)^{-1} \\
&= 2\left(\sum_i \Delta X'_i \Delta X_i\right)^{-1}
\end{aligned}$$

And putting all these together, we have

$$
\begin{aligned}
\widehat{Avar(\hat{\beta}_{FE})} &= \hat{\sigma}^2_{u,FE}\left(\ddot{X}'\ddot{X}\right)^{-1} \\
&= \frac{1}{2}\hat{\sigma}^2_{u,FD}2\left(\sum_i \Delta X_i' \Delta X_i\right)^{-1} \\
&= \hat{\sigma}^2_{u,FD}\left(\sum_i \Delta X_i' \Delta X_i\right)^{-1} \\
&= \widehat{Avar(\hat{\beta}_{FE})}
\end{aligned}
$$

Because the estimates of the asymptotic variances are equal, the standard errors (the square roots) will be equal.

# Problem 2

**Question 21-3 from Cameron-Trivedi (enhanced): Consider the fixed effects, two-way error component panel data model:**

$$y_{it} = \alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \tag{2}$$

## Part (a)

**Show that the fixed effects estimator of $\beta$ can be obtained by applying two within (one-way) transformations on this model. The first is the within transformation ignoring the time effects followed by the within transformation ignoring the individual effects. Assume the panel is balanced. (Hint: it may be easier to analyze the fixed effects regression using partitioned regression.)**

We want to show that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

Ignoring time effects, we get

$$\begin{aligned} \ddot{y}_{it} = y_{it} - \overline{y_i} = y_{it} - \frac{1}{T}\sum_t y_{it} \\ \implies \ddot{y}_{it} = y_{it} - \frac{1}{T}\sum_t (\alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it}) \end{aligned}$$

Applying the second within transformation, we get that

$$\begin{aligned} \dddot{y}_{it} = \ddot{y}_{it} - \overline{\ddot{y}_{it}} \\ = \ddot{y}_{it} - \frac{1}{N}\sum_i \ddot{y}_{it} \\ = y_{it} - \overline{y_i} - \overline{y_t} + \overline{y} \\ = \alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \\ - \left(\frac{1}{T}\sum_t(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\ - \left(\frac{1}{N}\sum_i(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\ + \left(\frac{1}{NT}\sum_i\sum_t(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\ = \beta\left(x_{it} - \overline{x_i} - \overline{x_t} + \overline{x}\right) + \epsilon_{it} - \overline{\epsilon_i} - \overline{\epsilon_t} + \overline{\epsilon} \end{aligned}$$

Since

$$\dddot{x}_{it} = x_{it} - \overline{x_i} - \overline{x_t} + \overline{x}$$

and

$$\dddot{e}_{it} = e_{it} - \overline{e_i} - \overline{e_t} + \overline{e}$$

We get that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

## Part (b)

**Show that the order of the transformations is unimportant. Give an intuitive explanation for why.**

Reversing the order, we can show that we get the same result

Again, we want to show that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

Ignoring individual effects, we get

$$\begin{aligned} \ddot{y}_{it} &= y_{it} - \overline{y_t} = y_{it} - \frac{1}{N}\sum_i y_{it} \\ \implies \ddot{y}_{it} &= y_{it} - \frac{1}{N}\sum_i (\alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it}) \end{aligned}$$

Applying the second within transformation, we get that

$$\begin{aligned} \dddot{y}_{it} &= \ddot{y}_{it} - \overline{\ddot{y}_{it}} \\ &= \ddot{y}_{it} - \frac{1}{T}\sum_i \ddot{y}_{it} \\ &= y_{it} - \overline{y_t} - \overline{y_i} + \overline{y} \\ &= \alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \\ &- \left(\frac{1}{N}\sum_i(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\ &- \left(\frac{1}{T}\sum_t(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\ &+ \left(\frac{1}{NT}\sum_i\sum_t(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\ &= \beta\left(x_{it} - \overline{x_t} - \overline{x_i} + \overline{x}\right) + \epsilon_{it} - \overline{\epsilon_t} - \overline{\epsilon_i} + \overline{\epsilon} \end{aligned}$$

Since

$$\dddot{x}_{it} = x_{it} - \overline{x_t} - \overline{x_i} + \overline{x}$$

and

$$\dddot{e}_{it} = e_{it} - \overline{e_t} - \overline{e_i} + \overline{e}$$

We get that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

## Part (c)

**Does your answer to part (a) change if the panel becomes unbalanced (i.e., contains different numbers of observations for each individual $i$). Why or why not?**

# Problem 3

**We now begin with an actual analysis of the data. The goal here is to determine what effect, if any, primary belt laws have on the log of traffic fatalities per capita (we log the LHS variable because we believe the effect of safety belt laws should be proportional to the overall level of fatalities per capita).**

```
data = read_dta('traffic_safety2.dta') %>%
    filter(state != 99) %>%
    mutate(fatal_per_cap = fatalities / population,
           vmt_per_cap = totalvmt/population)
```

## Part (a)

**Run pooled bivariate OLS. Interpret. Add year fixed effects. Interpret. Add all covariates that you believe are appropriate. Think carefully about which covariates should be log transformed and which should enter in levels. What happens when you add these covariates? Why?**

total vehicle miles traveled per capita: logged because, after using per capita variables, we think the increase in fatalities per capita would be proportional to a percentage point increase in vehicle miles traveled per capita, not absolute levels, since the percentage point increase is a better measure of deviation from the norm, and drivers are more likely to adjust poorly to a deviation from the norm than an absolute increase in vehicle miles traveled (since an absolute increase may not be very differnt from the norm in states that have relatively large average miles traveled.)

Not inlcuding population because we are using per-capita variables

Log precip because we car about percentage-point deviation from the norm. Don't log snow because it has zeros.

When we add the covariates, the estimated effect on primary switches from significantly negative to insignificantly positive. Since we haven't yet controlled for state-level fixed effects, this could be from trying to compare states that have similar observable covariates that adopt primary safety belt laws at different times, but may have significant reasons for adopting when they did, and thus may have significant unobserved factors that are correlated with the primary safety belt law being in place.

```
# Bivariate with lm
reg_a_bivariate_lm = lm(log(fatal_per_cap) ~ primary , data = data)
# Bivariate with plm
reg_a_bivariate_plm = plm(log(fatal_per_cap) ~ primary,
                       data = data,
                       model = "pooling")
# FE with lm
reg_a_yfe_lm = lm(log(fatal_per_cap) ~ primary + factor(year), data = data)
# FE with plm
reg_a_yfe_plm = plm(log(fatal_per_cap) ~ primary,
                 data = data,
                 index = c("year"),
                 model = "within",
                 effect = "individual")
# FE + covars with lm
reg_a_full_lm = lm(log(fatal_per_cap) ~ primary + factor(year) + college + beer
                   + secondary + unemploy + log(vmt_per_cap) + log(precip) + snow32
                   + log(rural_speed) + log(urban_speed), data = data)
# FE + covars with plm
reg_a_full_plm = plm(log(fatal_per_cap) ~ primary + factor(year) + college + beer
                     + secondary + unemploy + log(vmt_per_cap) + log(precip) + snow32
```

```r
                      + log(rural_speed) + log(urban_speed),
                      data = data,
                      index = c("state", "year"),  # order matters: unit-var, time-var
                      model = "within",
                      effect = "time")  # only do time


# Checking in-state variation in rural_speed
# d = data %>% group_by(state) %>% mutate(rural_speed_dev = abs(rural_speed - mean(rural_speed)), rural_
# plot(d$state, d$rural_speed_dev_mean)
```

```r
stargazer(reg_a_bivariate_plm, reg_a_yfe_plm, reg_a_full_plm, type='text')
```

===========================================================================================
Dependent variable:

————————————————————————————————————————————- log(fatal_per_cap)

(1) (2) (3)

——————————————————————————————————————————————————————— primary -0.155*** -0.091*** 0.031

(0.027) (0.028) (0.020)

college -2.387***
(0.146)

beer 0.176***
(0.025)

secondary 0.038**
(0.017)

unemploy 0.017***
(0.004)

log(vmt_per_cap) 1.154***
(0.041)

log(precip) -0.051***
(0.012)

snow32 -0.171***
(0.014)

log(rural_speed) 0.554***
(0.127)

log(urban_speed) 0.182**
(0.091)

Constant -1.705***
(0.011)

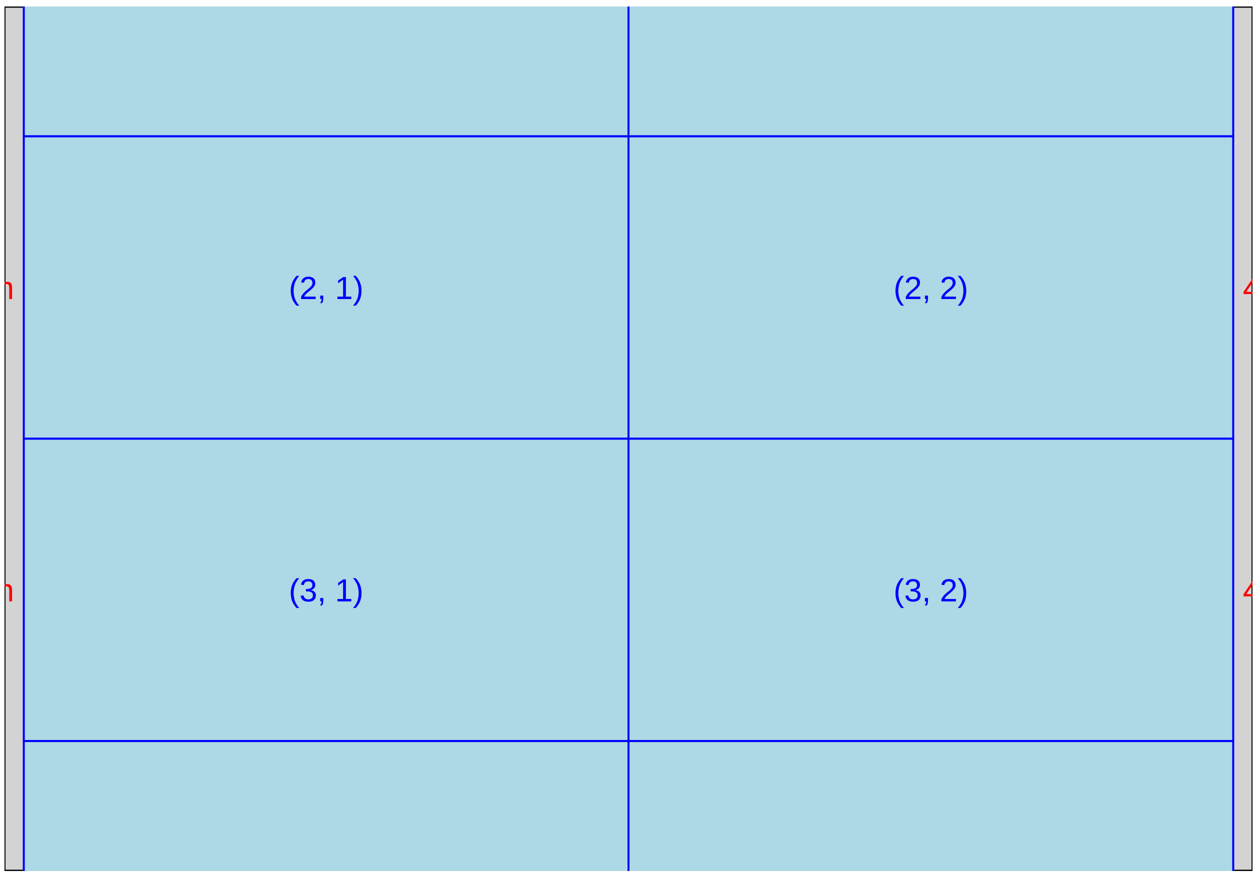
——————————————————————————————————————————

Observations 1,104 1,104 1,104
R2 0.029 0.010 0.744
Adjusted R2 0.028 -0.011 0.736
F Statistic 32.799*** (df = 1; 1102) 10.848*** (df = 1; 1080) 311.366*** (df = 10; 1071)
===========================================================================================
Note: *p<0.1;* ***p<0.05;*** p<0.01

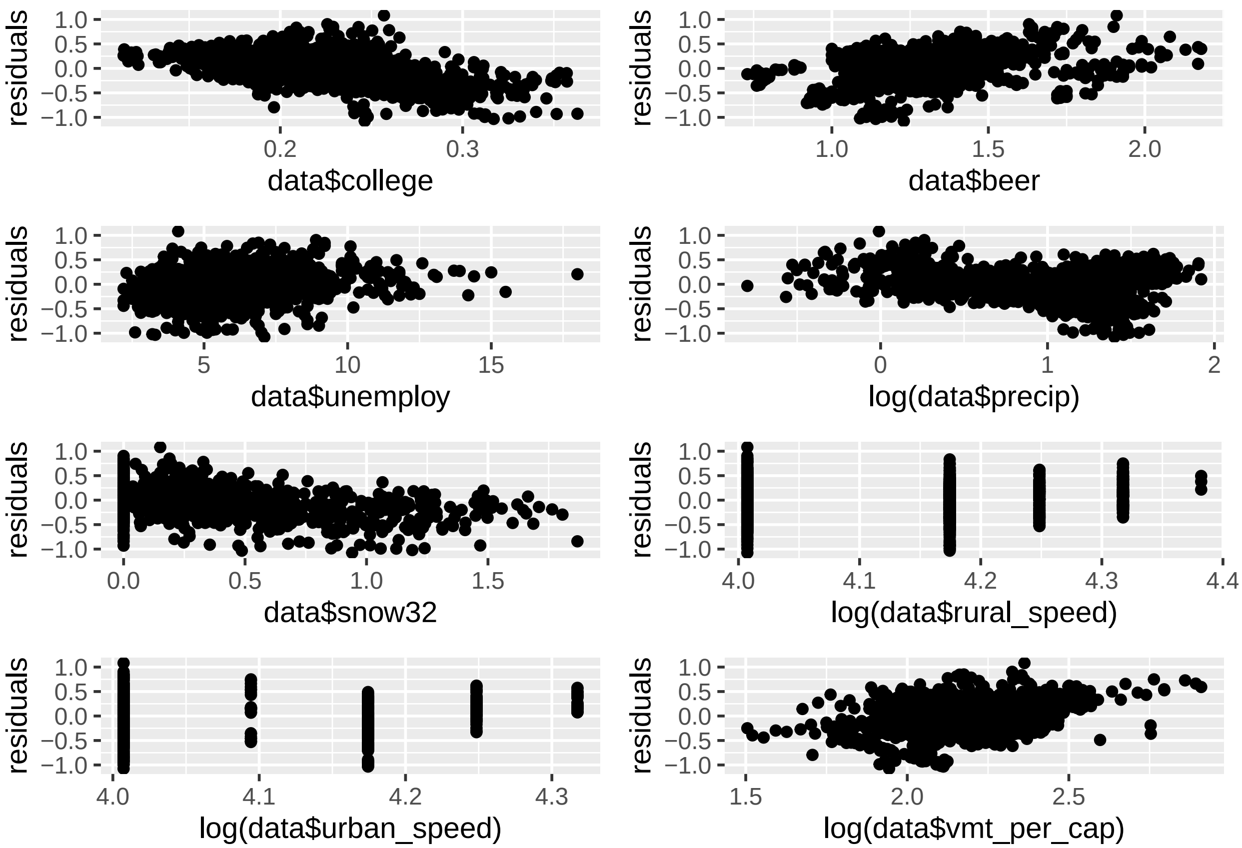
### Part (b)

**Ignore omitted variables bias issues for the moment. Do you think the standard errors from above are right? Compute the Huber-White heteroskedasticity robust standard errors (e.g., ", robust"). Do they change much? Compute the clustered standard errors that are robust to within-state correlation (e.g., ", cluster(state)"). Do this using both the "canned" command and manually using the formulas we learned in class. Do the standard errors change much? Are you surprised? Interpret.**

```
a <- gtable(unit(rep(8,2), c("cm")), unit(rep(4,4), "cm"))
gtable_show_layout(a)
```

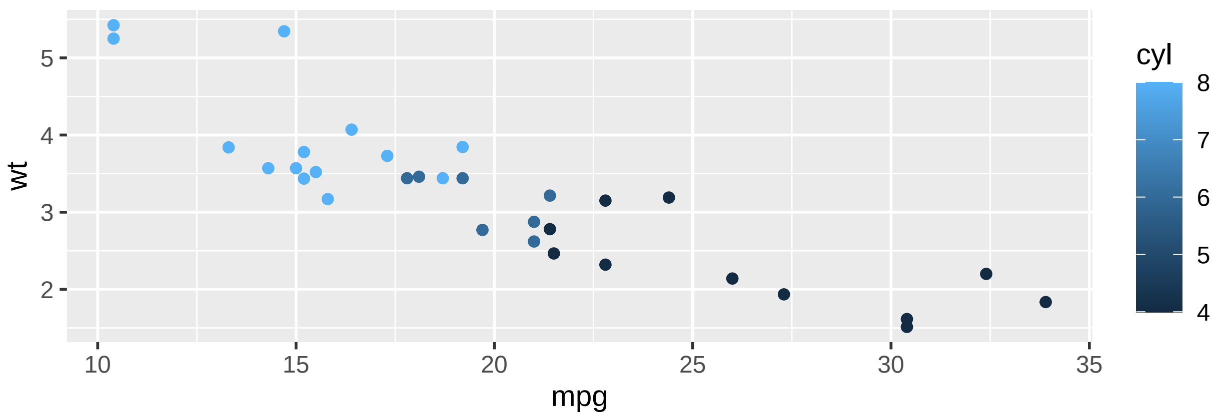


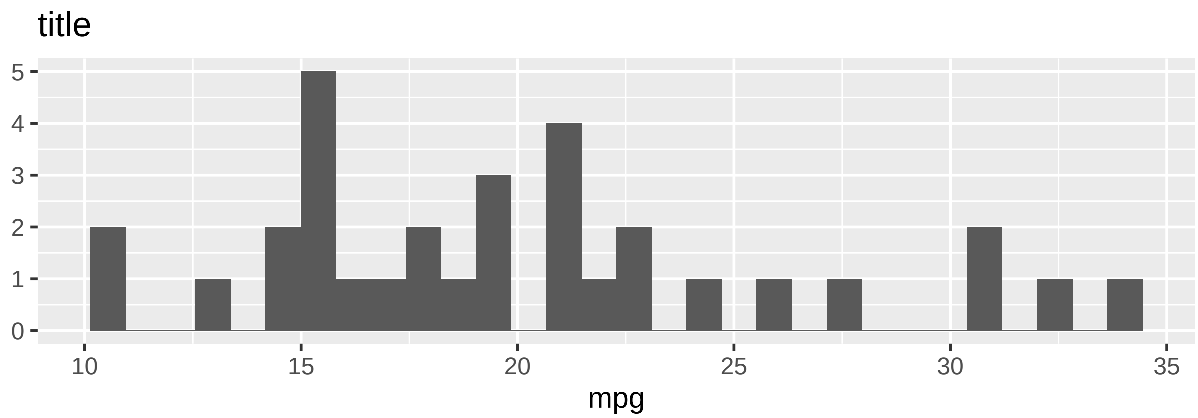

```
grid.arrange(
qplot(data$college, resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(data$beer, resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(data$unemploy, resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(log(data$precip), resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(data$snow32, resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(log(data$rural_speed), resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(log(data$urban_speed), resid(reg_a_bivariate_plm), ylab="residuals"),
qplot(log(data$vmt_per_cap), resid(reg_a_bivariate_plm), ylab="residuals"),
nrow = 4, heights=rep(6,4))
```




```
p1 <- qplot(mpg, wt, data = mtcars, colour = cyl)
p2 <- qplot(mpg, data = mtcars) + ggtitle("title")
p3 <- qplot(mpg, data = mtcars, geom = "dotplot")
p4 <-
  p1 + facet_wrap( ~ carb, nrow = 1) + theme(legend.position = "none") +
  ggtitle("facetted plot")
grid.arrange(p1,p2)
```

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




### Part (c)

**Compute the between estimator, both with and without covariates. Under what conditions will this give an unbiased estimate of the effect of primary seat belt laws on fatalities per capita? Do you believe those conditions are met? Are you concerned about the standard errors in this case?**

### Part (d)

**Compute the Random Effects estimator (including covariates). Under what conditions will this give an unbiased estimate of the effect of primary seat belt laws on fatalities per capita? What are its advantages or disadvantages as compared to pooled OLS?**

### Part (e)

**Do you think the standard errors from RE are right? Compute the clustered standard errors. Are they substantially different? If so, why? (i.e., what assumption(s) are being violated?)**

### Part (f)

**Compute the FE estimator using only primary and year fixed effects as the covariates. Compute the normal standard errors and the clustered standard errors. If they are different, why?**

## Part (g)

**Add the same range of covariates to the FE estimator that you did to the OLS estimator. Are the FE estimates more or less stable than the OLS estimates? Why?**

## Part (h)

**Estimate a first-differences estimator, a 5-year differences estimator, and a long differences estimator, including year fixed effects (when feasible) and the appropriate covariates in each case. Briefly describe the pattern that emerges from the three differencing estimates. Where does the FE estimate fall in this pattern? Are you surprised?**

## Part (i)

**Make the case that the first-differences estimate is superior to the 5-year or long differences estimates.**

## Part (j)

**Make the case that the 5-year or long differences estimates are superior to the first-differences estimate.**