ECON 334 Week 8 Assignment

Name: TAK SHING POON

Student Number: 301389544

```
library(tidyverse)
library(modelr)
library(broom)
```

```
LSFdata <-read.csv("pub0922.csv")
```

Part 1

Creating a data object that includes "employed, at work" and who work full-time in their main job.

```
employnFULL <-  LSFdata %>%
  filter(LFSSTAT == 1 & FTPTMAIN == 1)
#data set now only contains subjects that are employed and working full-time
```

Part 2

Transform "EDUC" to EduCat

```
employnFULL <-  employnFULL %>%
  mutate(EduCat = if_else(EDUC == 0, "0 to 8 years", ""),
        EduCat = if_else(EDUC == 1, "Some high school", EduCat),
        EduCat = if_else(EDUC == 2, "High school graduate", EduCat),
        EduCat = if_else(EDUC == 3, "Some postsecondary", EduCat),
        EduCat = if_else(EDUC == 4, "Postsecondary certificate or diploma", EduCat),
        EduCat = if_else(EDUC == 5, "Bachelor's degree", EduCat),
        EduCat = if_else(EDUC == 6, "Above bachelor's degree", EduCat),
        Wage = HRLYEARN/100)
```

Part 3

Regress Wage on job tenure with the current employer and

EduCat

```
mod1 <- lm(Wage ~ TENURE + EduCat, data = employnFULL)
summary(mod1)
```

```
##
## Call:
## lm(formula = Wage ~ TENURE + EduCat, data = employnFULL)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -45.847  -8.993  -2.584   6.611  83.009
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             18.828209   0.607357  31.000
## TENURE                                   0.041650   0.000795  52.392
## EduCatAbove bachelor's degree           21.542447   0.634535  33.950
## EduCatBachelor's degree                 15.571656   0.618033  25.196
## EduCatHigh school graduate               4.215650   0.622873   6.768
## EduCatPostsecondary certificate or diploma  8.608980   0.611595  14.076
## EduCatSome high school                   1.970906   0.677058   2.911
## EduCatSome postsecondary                 5.374532   0.685926   7.835
##                                          Pr(>|t|)
## (Intercept)                              < 2e-16 ***
## TENURE                                   < 2e-16 ***
## EduCatAbove bachelor's degree            < 2e-16 ***
## EduCatBachelor's degree                  < 2e-16 ***
## EduCatHigh school graduate              1.32e-11 ***
## EduCatPostsecondary certificate or diploma  < 2e-16 ***
## EduCatSome high school                     0.0036 **
## EduCatSome postsecondary                4.78e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.85 on 43013 degrees of freedom
##   (6290 observations deleted due to missingness)
## Multiple R-squared:  0.1931, Adjusted R-squared:  0.193
## F-statistic:  1471 on 7 and 43013 DF,  p-value: < 2.2e-16
```

The reference category is the education category of "0 to 8 years" of education.

Observations and interpretations:

According to the regression summary and one month increase in job tenure increases wage by $0.04 ($0.0416) dollars, holding education variables constant. Holding job tenure constant:

- On average, observations with "some high school" education earn $1.97 more dollars than the reference category.
- On average, participants who are "High school graduate" earn $4.22 more dollars than the reference category.

- On average, observations who are "Some post secondary" education earn $5.38 more dollars than the reference category.
- On average, people who have a "post-secondary certificate or diploma" earn $8.61 more dollars than the reference category.
- On average, observations with a Bachelor's degree earn $15.57 more dollars than the reference category.
- Finally on average, observations with a Bachelor's degree and above earn $18.83 more dollars than the reference category.

The data seems constant with what I was thinking where increase in education has a positive correlation with increase in wage. The data suggest a direct positive relationship between education and wage increase.

Part 4) Use broom's tidy() to plot the estimated coefficients (and their 95% confidence intervals)

```
p <- tidy(mod1, conf.int = TRUE) %>%
  filter(term != "TENURE" ,term != "(Intercept)") %>%
  mutate(term =gsub("EduCat", "", term)) %>%
  ggplot(aes(x=reorder(term,estimate), y =estimate, ymin=conf.low,ymax=conf.high))
```

```
p + geom_pointrange()+coord_flip()+
  labs(
  x= "Education Level",
  y= "Coefficients Estimation",
  title = "Estimated coefficients and 95% CI of education level on Wage",
  caption = "Source: Statistic Canada")
```

Estimated coefficients and 95% CI of education level on Wage
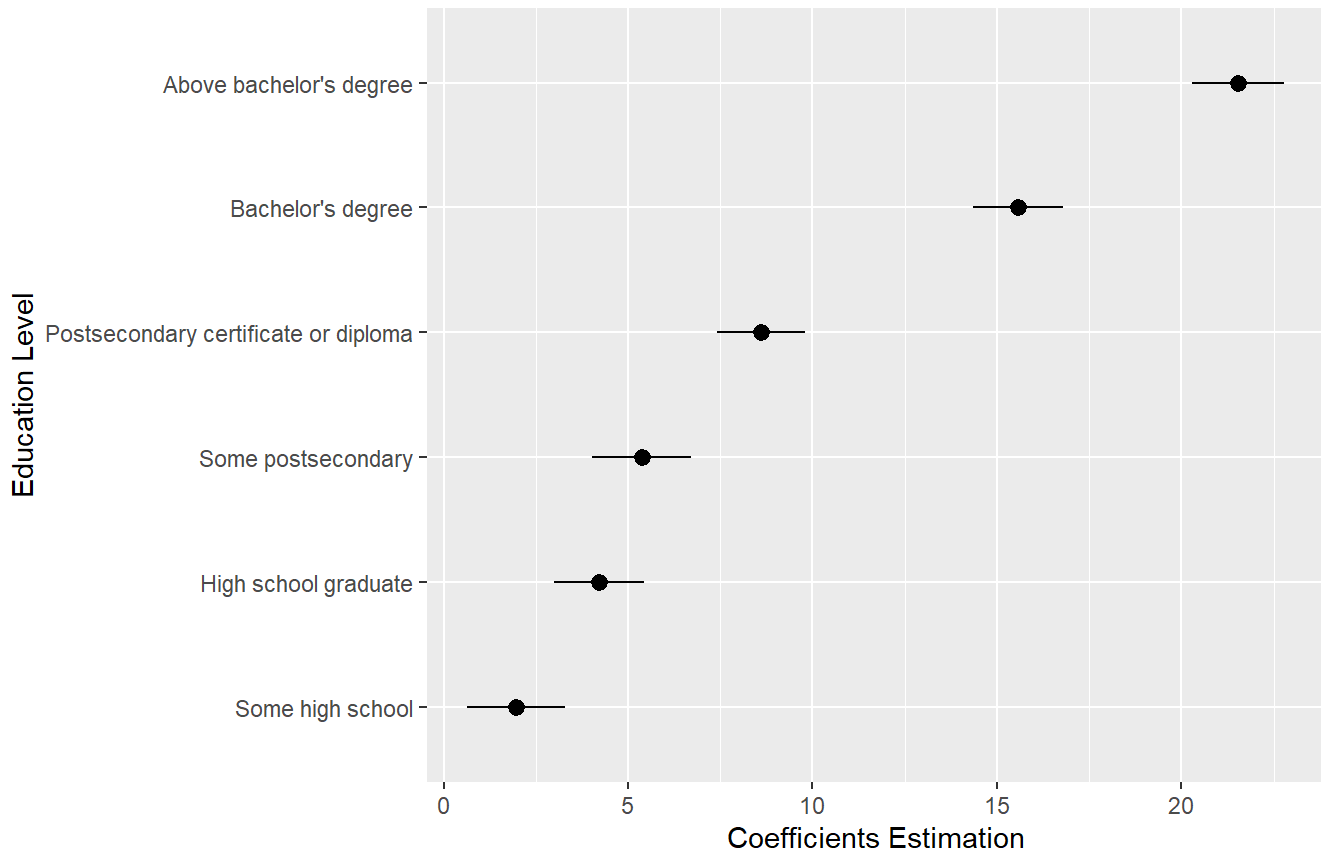
Source: Statistic Canada

Observations:

If a confidence interval contains zero then we would say there is strong evidence (95% confidence in this case) that there is not a 'significant' difference between the two population means. None of the education coefficients includes zero, this suggests that they are significant/useful in predicting the effect of education level on wage. In other words, we are 95% confident that each education level have their each respective effect on wage in respect to the reference category.

Part 5 Further Regression

```
emploγnFULL <- emploγnFULL %>%
  mutate(ImmigCat = if_else(IMMIG == 1, "Immigrant, landed 10 or less years earlier",""),
         ImmigCat = if_else(IMMIG == 2, "Immigrant, landed more than 10 years earlier", ImmigCa
t),
         ImmigCat = if_else(IMMIG == 3, "Non-immigrant", ImmigCat))
```

```
mod2 <- lm(Wage ~ TENURE+ EduCat+ImmigCat,data = emploγnFULL)
summary(mod2)
```

```
##
## Call:
## lm(formula = Wage ~ TENURE + EduCat + ImmigCat, data = employnFULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.585  -8.940  -2.596   6.566  82.528
##
## Coefficients:
##                                                      Estimate Std. Error
## (Intercept)                                         1.580e+01  6.455e-01
## TENURE                                              3.986e-02  8.043e-04
## EduCatAbove bachelor's degree                       2.188e+01  6.331e-01
## EduCatBachelor's degree                             1.561e+01  6.163e-01
## EduCatHigh school graduate                          3.938e+00  6.215e-01
## EduCatPostsecondary certificate or diploma          8.327e+00  6.102e-01
## EduCatSome high school                              1.627e+00  6.756e-01
## EduCatSome postsecondary                            5.038e+00  6.844e-01
## ImmigCatImmigrant, landed more than 10 years earlier 2.388e+00  3.003e-01
## ImmigCatNon-immigrant                               3.855e+00  2.602e-01
##                                                      t value Pr(>|t|)
## (Intercept)                                           24.478  < 2e-16 ***
## TENURE                                                49.560  < 2e-16 ***
## EduCatAbove bachelor's degree                         34.565  < 2e-16 ***
## EduCatBachelor's degree                               25.325  < 2e-16 ***
## EduCatHigh school graduate                             6.336 2.38e-10 ***
## EduCatPostsecondary certificate or diploma            13.645  < 2e-16 ***
## EduCatSome high school                                 2.409    0.016 *
## EduCatSome postsecondary                               7.361 1.86e-13 ***
## ImmigCatImmigrant, landed more than 10 years earlier   7.955 1.84e-15 ***
## ImmigCatNon-immigrant                                 14.818  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.81 on 43011 degrees of freedom
##   (6290 observations deleted due to missingness)
## Multiple R-squared:  0.1978, Adjusted R-squared:  0.1977
## F-statistic:  1179 on 9 and 43011 DF,  p-value: < 2.2e-16
```

Observations:

The variable on the job tenure variable suggest that one month increase in job tenure increases wage by $0.0398 dollars. Campare to the previous regression of $0.0416, this regression equation causes the job tenure value to decrease. The estimation coefficient of TENURE decreased because we added new variables in to the regression. Since the estimate coefficient suggests the relationship of the variables (e.g. TENURE's relationship with wage) by introducing new variables in to the regression, we are basically suggesting additional relationship on the model. If the the new variables have any casual effect on the model, then the original estimation would decrease. Furthermore, this suggest that the previous model's estimation overestimates the variables to account for the unknown/error/unobserved effects.

Part 6 Visualizing regression model's predictions

```
#Wage(continuous),TENURE(continuous),EduCat(7),ImmigCat(3)
#mod2 <- lm(Wage ~ TENURE+ EduCat + ImmigCat,data = employnFULL)

grid <- employnFULL %>%
  data_grid(TENURE = seq_range(TENURE,100),
            EduCat =c("0 to 8 years","Some postsecondary","Some high school","Postsecondary cert
ificate or diploma","High school graduate","Bachelor's degree", "Above bachelor's degree"),
            ImmigCat =c("Immigrant, landed 10 or less years earlier","Immigrant, landed more tha
n 10 years earlier","Non-immigrant")) %>%
  gather_predictions(mod2, .pred = "Wage")

predictions <- cbind(grid, predict(object = mod2, newdata = grid, interval = "prediction"))
```
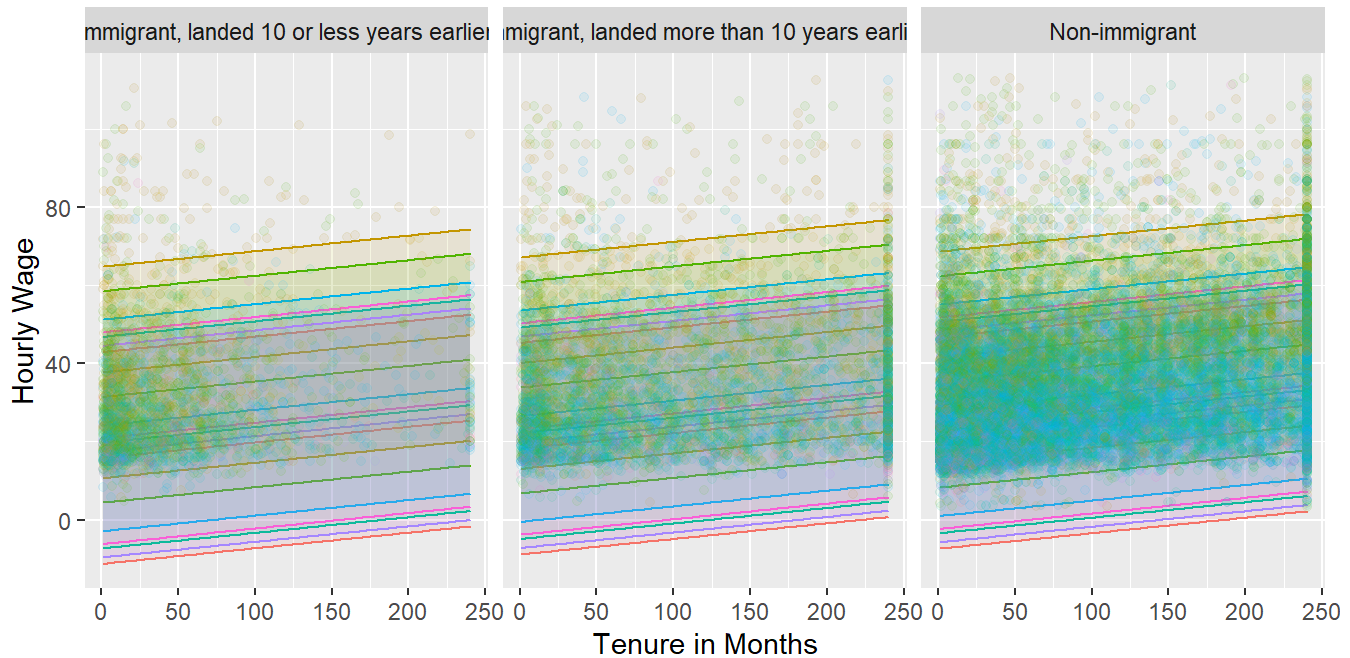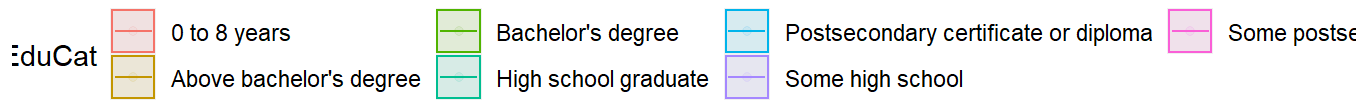
```
p1 <- predictions %>%
  ggplot(mapping= aes(x= TENURE, y = Wage, ymin = lwr, ymax = upr, color = EduCat, fill = EduCa
t)) +
  geom_line() +
  geom_ribbon(alpha = 0.1) +
  geom_point(employnFULL, mapping = aes(x = TENURE, y = Wage, color=EduCat), alpha = 0.07, inher
it.aes = FALSE) +
  facet_wrap(~ImmigCat) +
  theme(legend.position = "top") +
  labs(title = "Effects of immigration status, education and tenure on wage",
       x = "Tenure in Months",
       y = "Hourly Wage",
       subtitle = "Model : Wage ~ TENURE + EduCat + ImmigCat",
       caption = "Source: Statistic Canada"
       )

p1
```

```
## Warning: Removed 6290 rows containing missing values (geom_point).
```

Effects of immigration status, education and tenure on wage

Model : Wage ~ TENURE + EduCat + ImmigCat

EduCat

▢ 0 to 8 years	▢ Bachelor's degree	▢ Postsecondary certificate or diploma	▢ Some postse
▢ Above bachelor's degree	▢ High school graduate	▢ Some high school	

Source: Statistic Canada

Obervation :

We can tell that immigration status as the most observations, followed by immigrant landed more than 10 years and then less than 10 years. We can also see that "above bachelor's degree" education yields the highest hourly wage and 0-8 years of education yields the least.

MADE BY_TAK SHING POON(301389544)