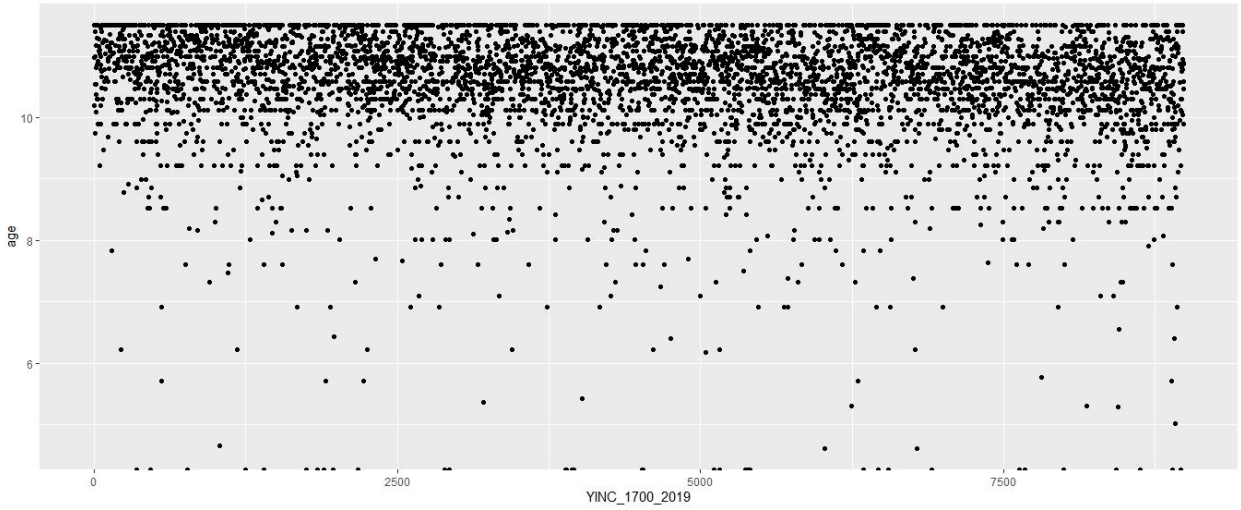
Homework 4

Tianpei Zhu

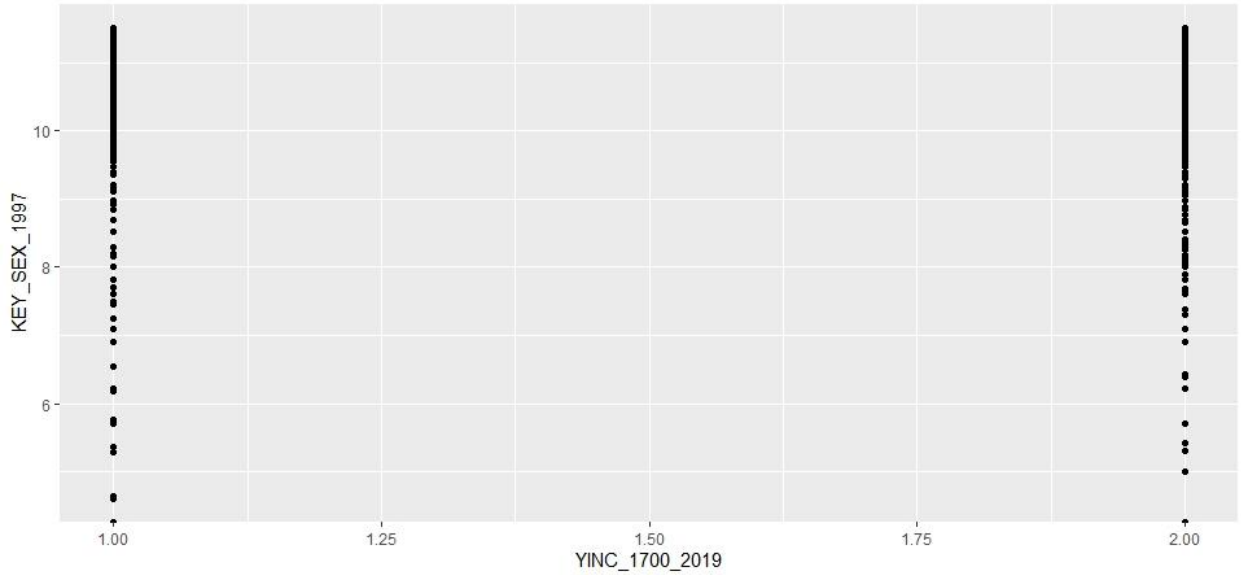
Exercise 1 Preparing the Data:

Plot of income data and age groups:



The graph depict that more age group and income has positive association ship. The increase in income lead to increase the age of the worker. When the two variables "age" and "salary" are considered together, a substantial positive connection is expected: as people get older, they tend to earn more money. As a result, the relationship between age and wage most likely yields a value.

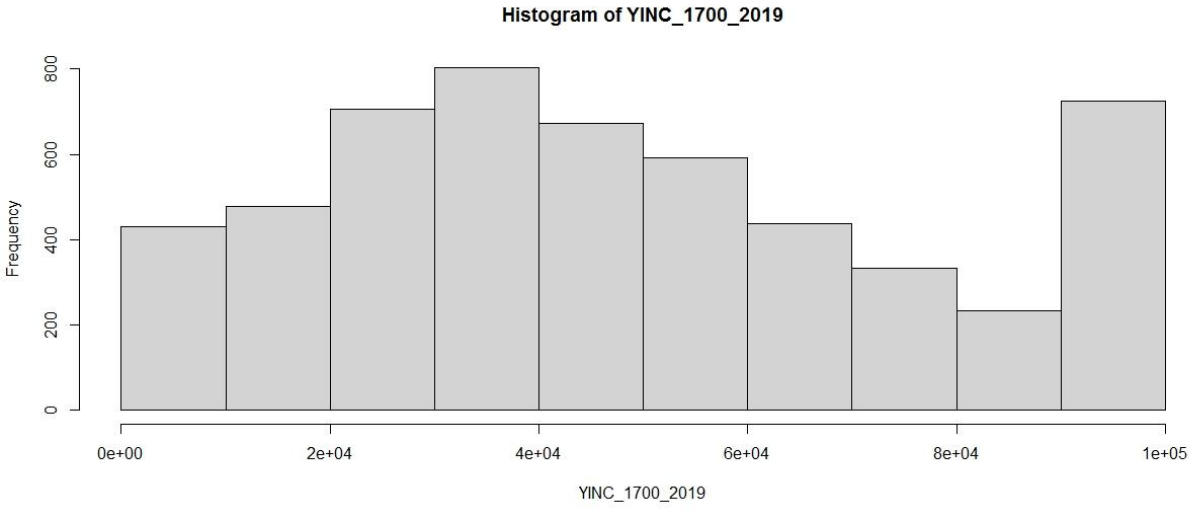
Plot of income data and Gender Group:



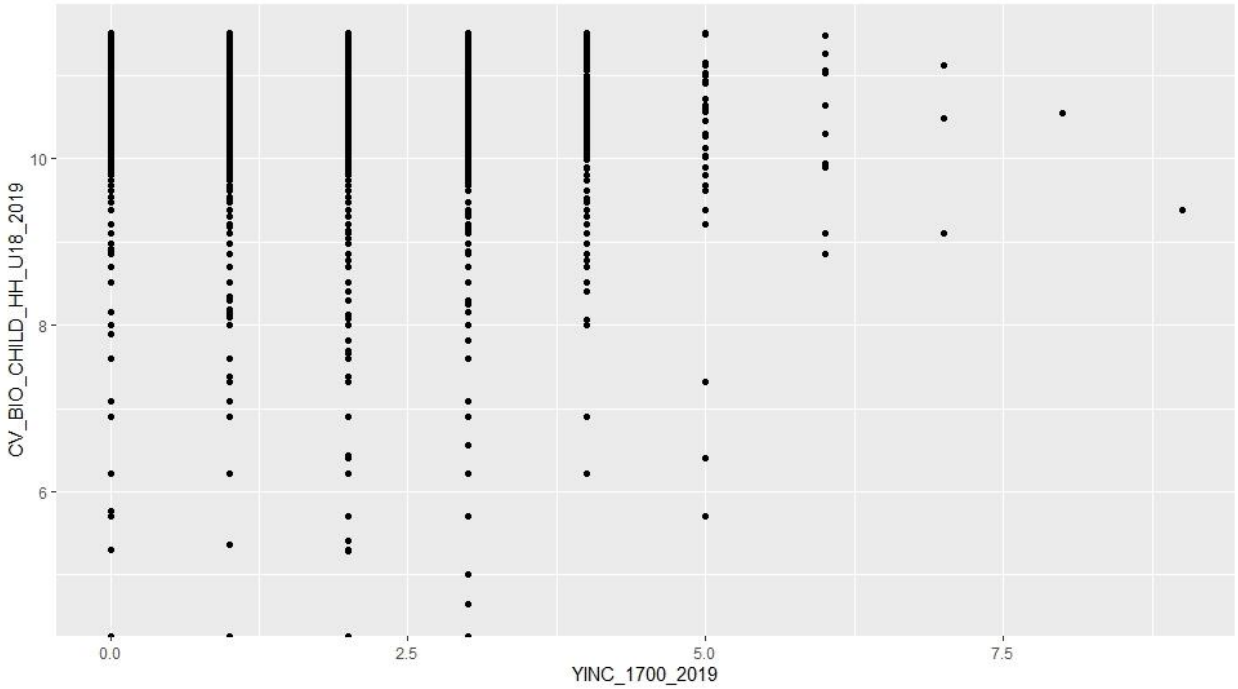
Gender identity, specifically an aversion to women earning more than men, has an impact on gender development, female labour force involvement, female income conditional on working, family permission and satisfaction, chance of opportunity, and home production division.

Plot of income data and number of children:

Within and across families, there is an inverse relationship between income and total fertility rate. In any industrialized country, the larger the number of children and working opportunities of a human population, subpopulation, or socioeconomic stratum, the fewer children are born.

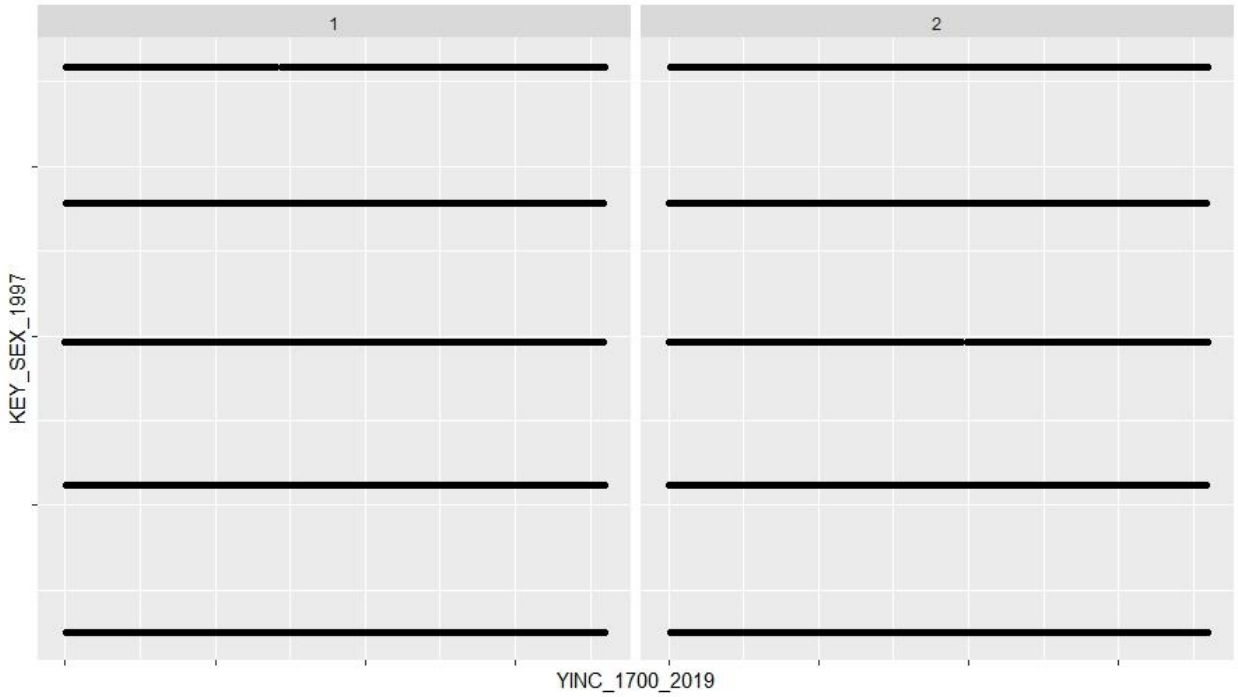


Plot of income data and age group:



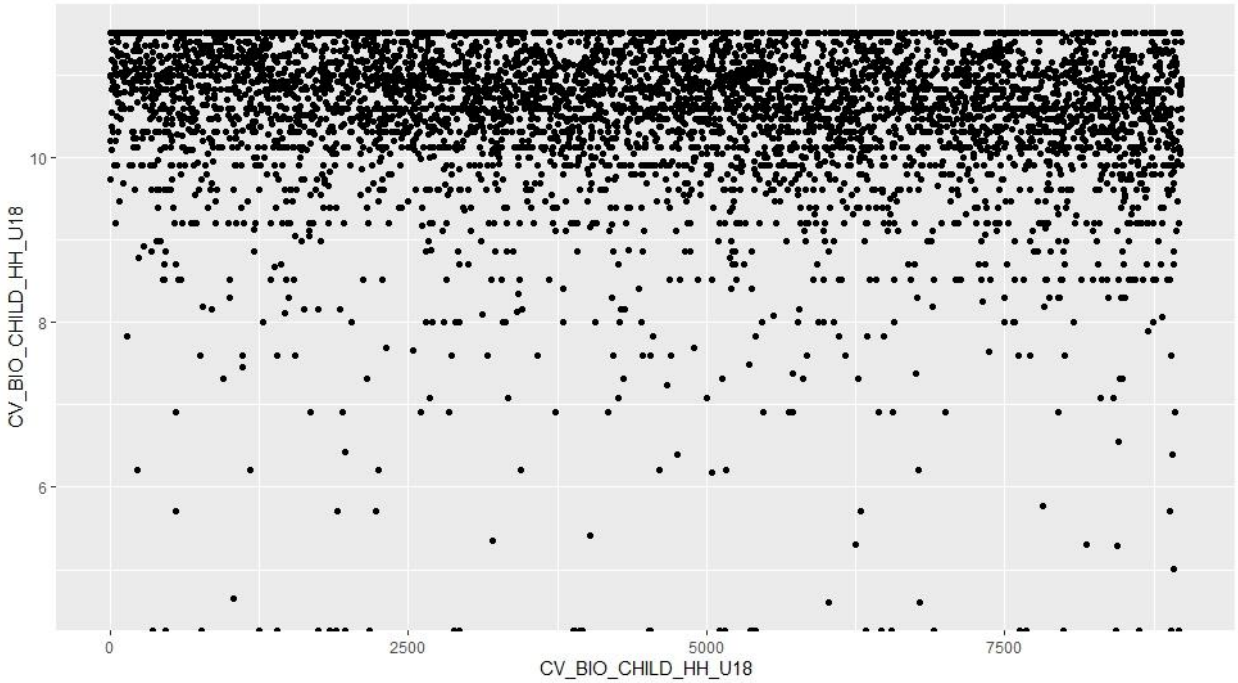
Graph depicting the association between the average number of children under the age of eight and income level for different age groups of household heads. The higher the working-age population, the higher the income, and the smaller the working-age population, the lower the income.

Plot of income data and Gender group:



Plot showing that both male and female has almost same earning pattern

Plot of income data and Number of children:



The more number of children tend to more income, as family number of income rises the income also goes to increase.

Exercise 2

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   12480.57   14442.23   0.864 0.38756
age            1032.04     358.12   2.882  0.00398 **
education       407.53      36.65  11.121  < 2e-16 ***
KEY_SEX_1997 -13909.01    1002.87 -13.869 < 2e-16 ***


Residual standard error: 27700 on 3060 degrees of freedom
  (5920 observations deleted due to missingness)
Multiple R-squared:  0.09804,  Adjusted R-squared:  0.09716
F-statistic: 110.9 on 3 and 3060 DF,  p-value: < 2.2e-16
```

The model is overall seem too good but coefficient of determinant is very small. The lower value of the coefficient of determinant suggest selection problem when estimating an OLS. Age and education has positive impact on the income level of the worker. As education increase it tend to add 407 percent in the income of the worker and age coefficient shows that as age goes to increase it tend to 1032 percent increase in the income of the consumer.

- Heckman's correction is used when one wishes to determine whether a subsample (or subgroup) of some larger sample is biased. For example, suppose I have a sample of 200 individuals I've collected data from (150 who use Facebook, 50 who don't). If I want to conduct a test only with the Facebook users, I can only do so with a subset of the data (i.e., the 150 Facebook users). Thus, it's possible my estimate will be biased since these 150 participants were not selected at random.

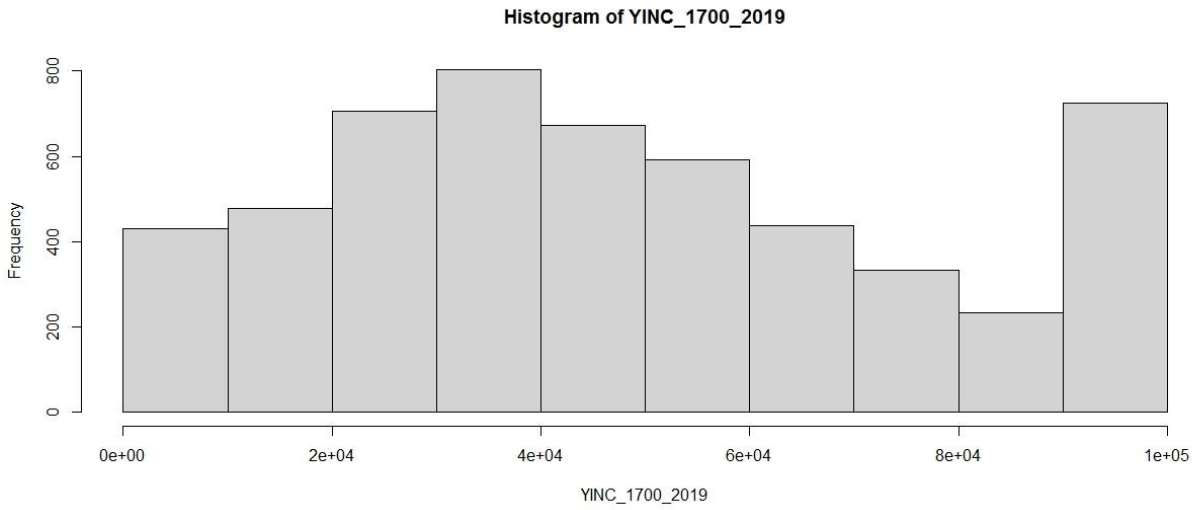
Hechman model:

```
   coeftest(vm1, vcov = sandwich)

t test of coefficients:

              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  2.4839331  0.0123068 201.8349   <2e-16 ***
KEY_SEX_1997 0.0049110  0.0076699   0.6403    0.522
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

•

Likelihood and optimize the two-stage Heckman model shows that gender does not positively influence the earning of the individual.

Exercise 3 Censoring:



The histogram shows that there is outlier in the model data set. the problem of outlier in the data set can be resolve through censoring optimization.

The model of censoring problem can be done through

- To reduce the amount of censored data, cut off the end of the sample period earlier.
- Use up-to-date data, including censored observations, but calculate a stand-in measurement or otherwise weight them differently.

Residuals:

Min	1Q	Median	3Q	Max
-2.48884	-0.00885	-0.00394	0.15021	2.06503

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.483933	0.012045	206.217	<2e-16	***
KEY_SEX_1997	0.004911	0.007690	0.639	0.523	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3245 on 7126 degrees of freedom

(1856 observations deleted due to missingness)

Multiple R-squared: 5.723e-05, Adjusted R-squared: -8.31e-05

F-statistic: 0.4078 on 1 and 7126 DF, p-value: 0.5231

The results of the analysis shows that gender of the workers had not significance influence on the earning.

Exercise 4 Panel Data:

The notion that abled individuals are the norm in society, and that people with disabilities must either aspire to achieve that norm or keep their distance from them, is known as ability bias. As a result, a disability is a "bad" thing that must be overcome.

Ability bias – A bias in the returns to education that might arise from the fact that persons with more education may have higher intrinsic skills that would allow them to earn more money even if they didn't go to school. Similar wage determinants show that higher competence leads to higher salaries, and higher experience leads to higher pay.

```
Call:
lm(formula = CV_WKSWK_JOB_DLI.01_2019 ~ age + education + KEY_SEX_1997,
    data = dat_A4)

Residuals:
    Min      1Q  Median      3Q     Max
-370.68 -216.95  -79.22  179.90 1195.30

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -697.21238  133.12652  -5.237 1.73e-07 ***
age            25.98757    3.30293   7.868 4.84e-15 ***
education       0.05903    0.33447   0.176  0.85993
KEY_SEX_1997  -25.41565    9.26723  -2.743  0.00613 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 266.2 on 3302 degrees of freedom
  (5678 observations deleted due to missingness)
Multiple R-squared:  0.02051,  Adjusted R-squared:  0.01962
F-statistic: 23.05 on 3 and 3302 DF,  p-value: 9.148e-15

> summary(spec1)

Call:
lm(formula = log(CV_WKSWK_JOB_DLI.01_2019) ~ log(age) + log(education) +
```

```
    log(KEY_SEX_1997), data = dat_A4)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3923 -0.7651  0.2495  1.0193  2.0603

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -9.23459    2.33820  -3.949 8.00e-05 ***
log(age)           3.76495    0.63040   5.972 2.59e-09 ***
log(education)     0.14330    0.07770   1.844   0.0652 .
log(KEY_SEX_1997) -0.14892    0.06393  -2.329   0.0199 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.273 on 3302 degrees of freedom
  (5678 observations deleted due to missingness)

Multiple R-squared:  0.01337,  Adjusted R-squared:  0.01247
F-statistic: 14.91 on 3 and 3302 DF,  p-value: 1.213e-09
```

By regressing the individual age averages influence on the individual averages of earning and a constant using OLS, the between estimator takes advantage of the data's cross-sectional dimension (differences between units). Fixed effects estimator (also known as the within estimator) is a term used in panel data analysis to refer to an estimator for the coefficients in the regression model that include those fixed effects (one time-invariant intercept for each subject). An ordinary least squares estimator of the vector of parameters in a model with panel data, computed using deviations from the data's temporal averages for each cross-section unit (deviations from group means). The results of the within estimators and between estimator and fixed effect adjusted results which provide unbiased estimators.

Age coefficient shows around 25 percent point increase in earning of the individual by within estimator and 0.5 percent positive impact on the earning of the individual. The between estimator shows that 3.76 percent point positive impact on the earning of the individual. The female earning is 0.14 time less as compare to male participant.