ECON1000 - Using Big Data to Solve Economic and Social Problems
Economics and Watson
Brown University
Prof. John Friedman

BROWN

ECON 1000

Empirical Exercise #5 Submission Template

Due TUESDAY 11/14/2023, by 10am EDT on Gradescope

Gradescope Course Link: https://www.gradescope.com/courses/565275

Name: Sonya Rashkovan

Group members with whom you worked[1]:

1.

2.

3.

4.

[1] In this class we encourage working in groups because you will learn a great deal from your peers. At the end of the day, however, it is important that you write up your own analysis. Your exploration in Q7 and your narrative should be your own. Please then list all group members with whom you worked. If you have any questions, please ask.

1. Calculate the weighted average earnings for men and women.

average earnings for men: 47570.33

average earnings for women: 24782.5

2. Calculate the weighted average earnings for men and women with and without children.

average earnings for men without children: 39904.62

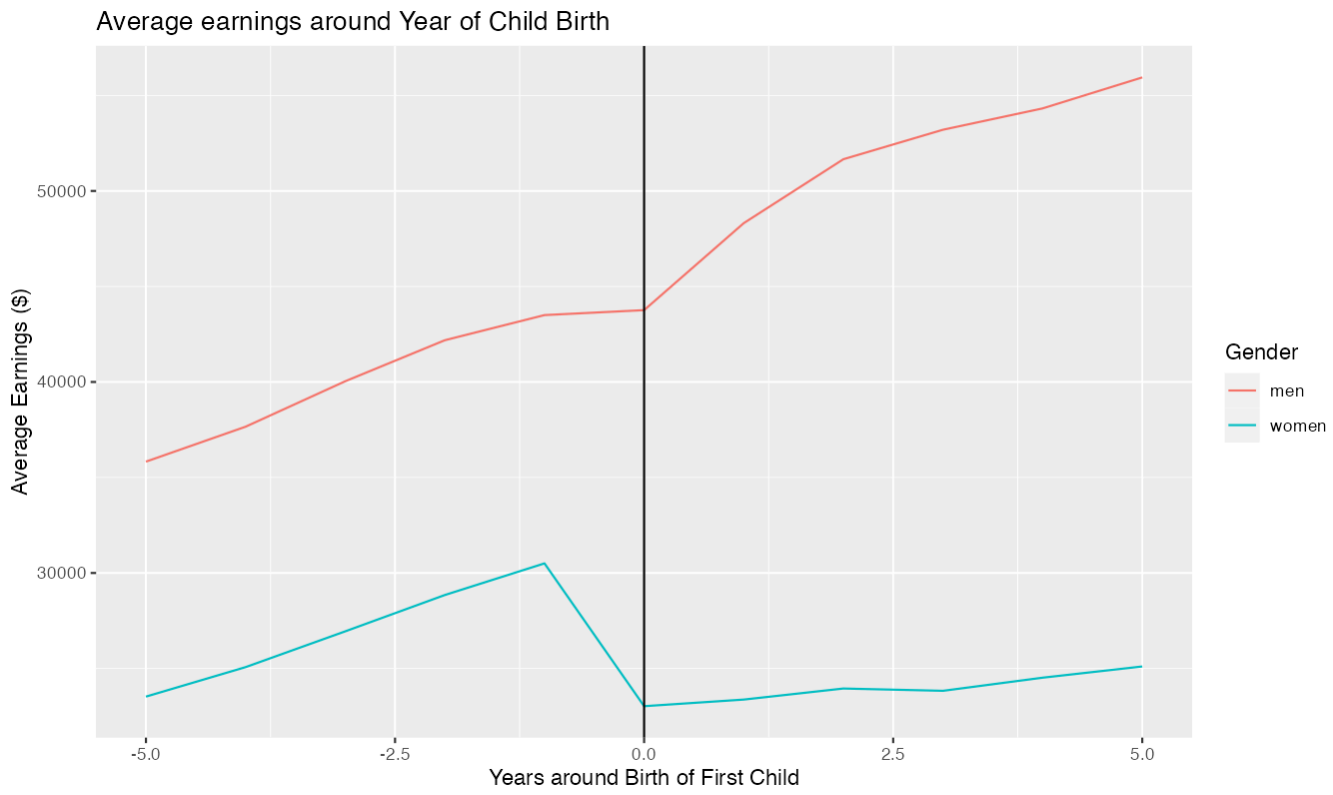
average earnings for women without children: 26977.97

average earnings for men with children: 50619.33

average earnings for women with children: 23882.06

When it comes to men and women without children, the gap between them is a bit less as childless men earn on average $39,904.62 and women $26,977.97. When it comes to parents, the largest gap between sexes appears: women with children earn on average $23,882.06 while men earn $50,619.33. A possible explanation for this is that it's more socially acceptable for women to not come back to work after having children while men are expected to start earning more as now they have to provide for more people in the family.

3. Plot earnings for men and women separately around the year of the birth of their first child.

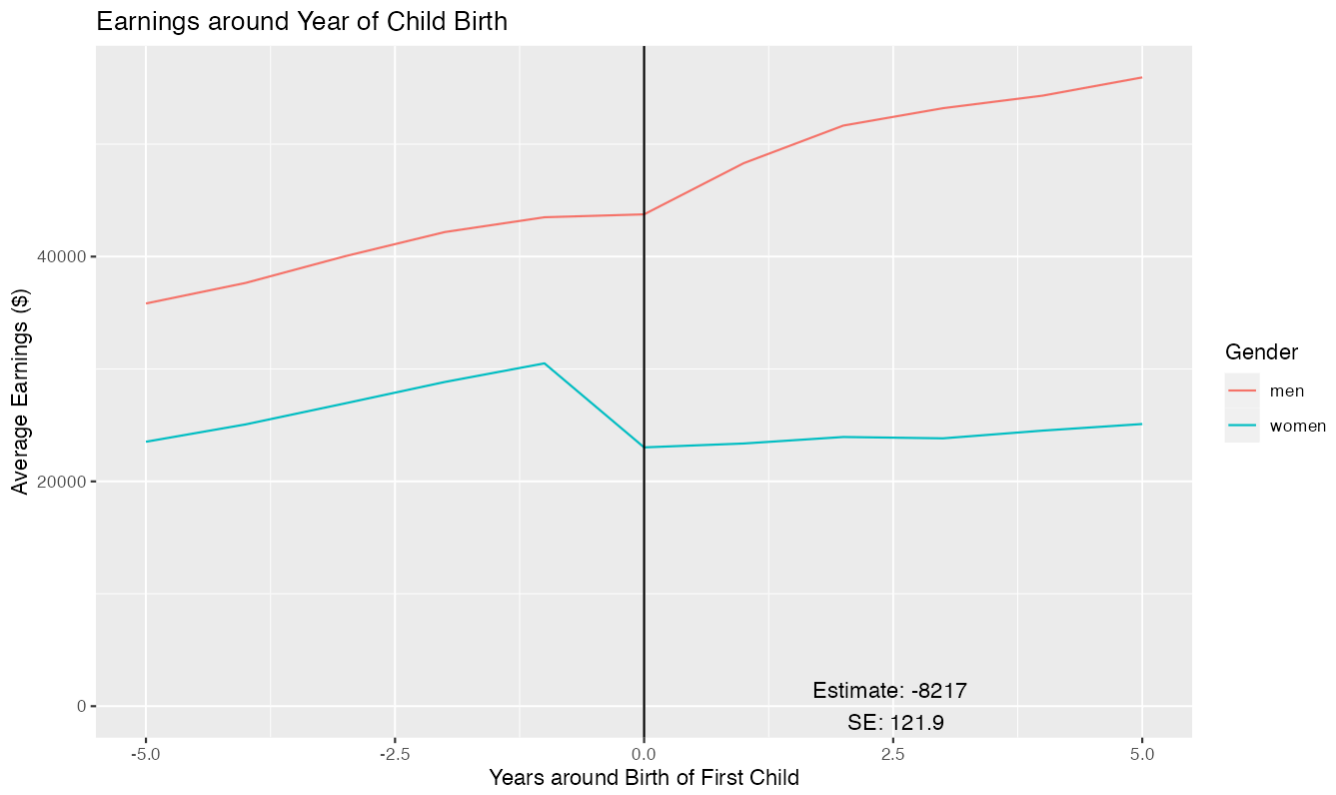


4. Run a difference in difference regression for men's and women's earnings around the birth of their first child.

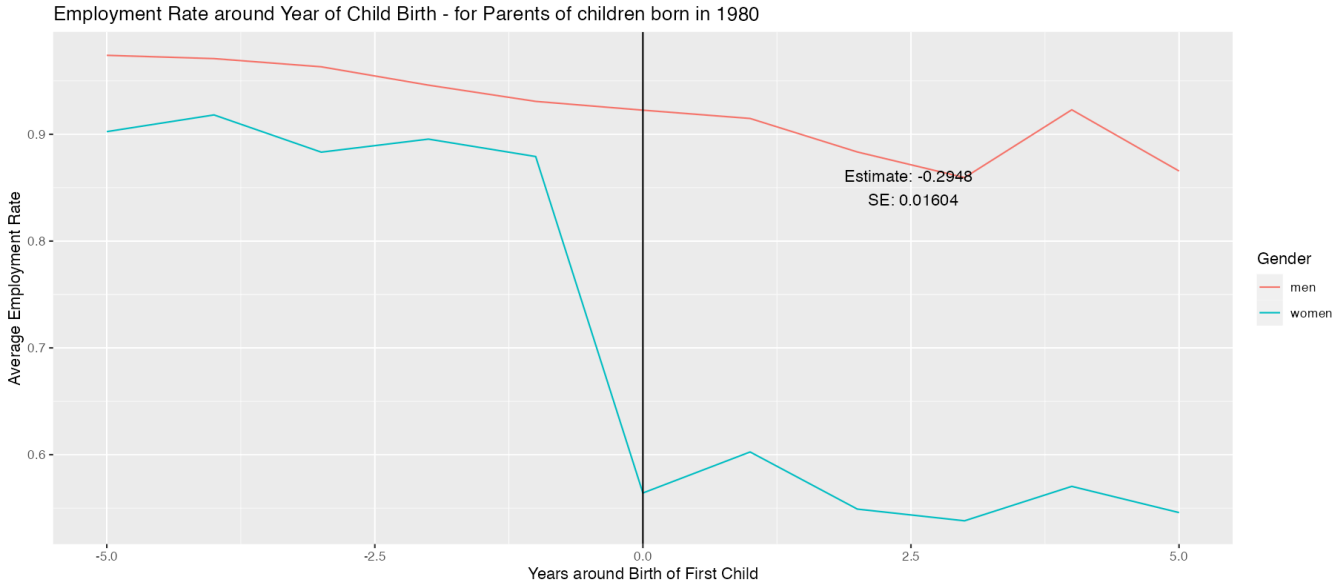
Coefficient: -8217

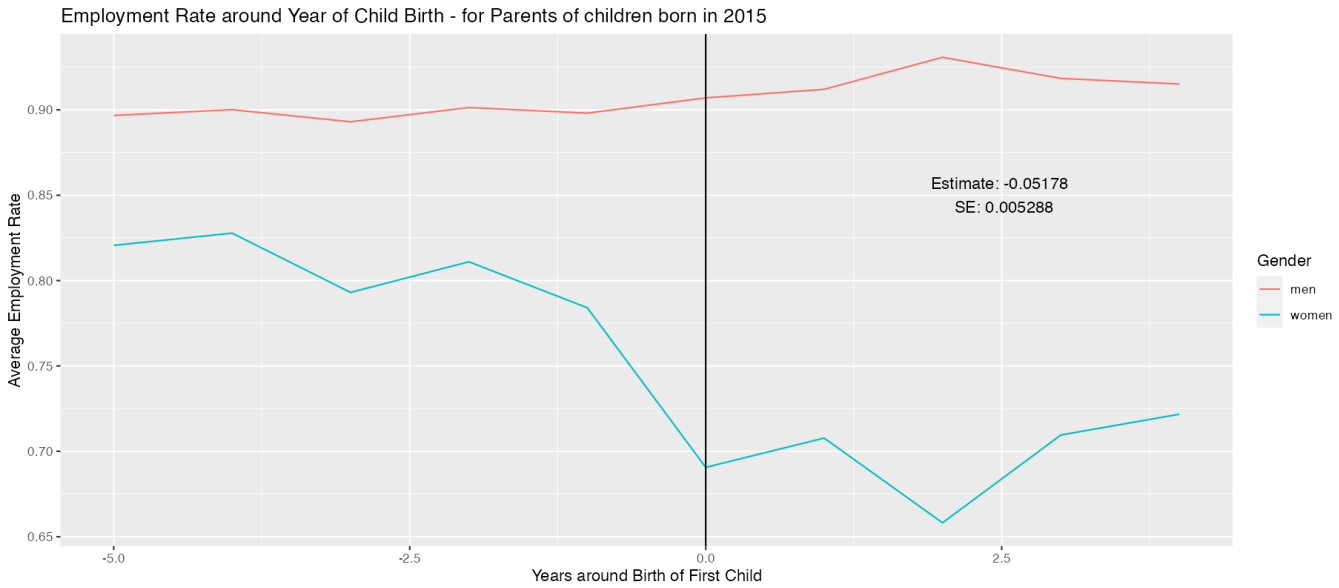
Standard error: 121.0

This regression is statistically significant as the P-value is < 2.2e-16 which is below 5% or .05 which means that is statistically significant.



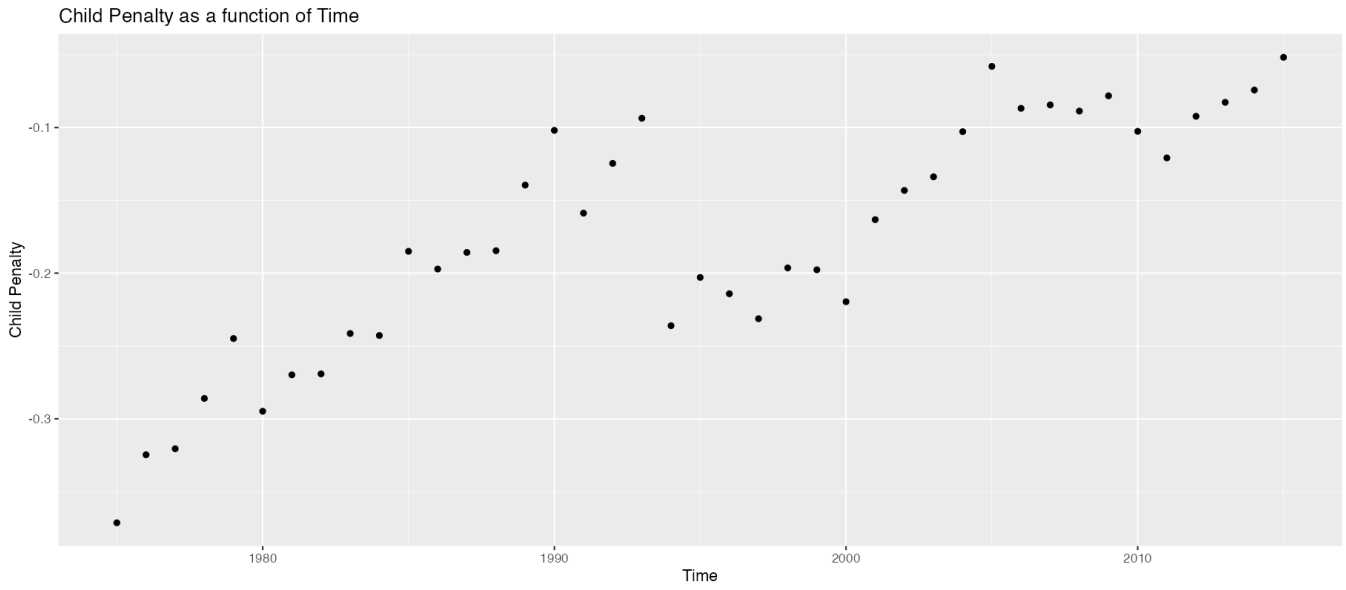
5. Plot the difference in difference regression in 1980 and 2015.





There has been progress in gender equality after childbirth between 1980 and 2015. In 1980 women's employment fell and didn't come back up after the first child's birth, but in 2015 around 2 years after childbirth, women's employment rate went back up but not even close to the pre-childbirth employment rate. Men's employment rate in 2015 steadily increased with the birth of the child not being a significant event. In 1980, there was a slight negative slope of the employment rate throughout the years but with the child's birth not being a decisive factor.

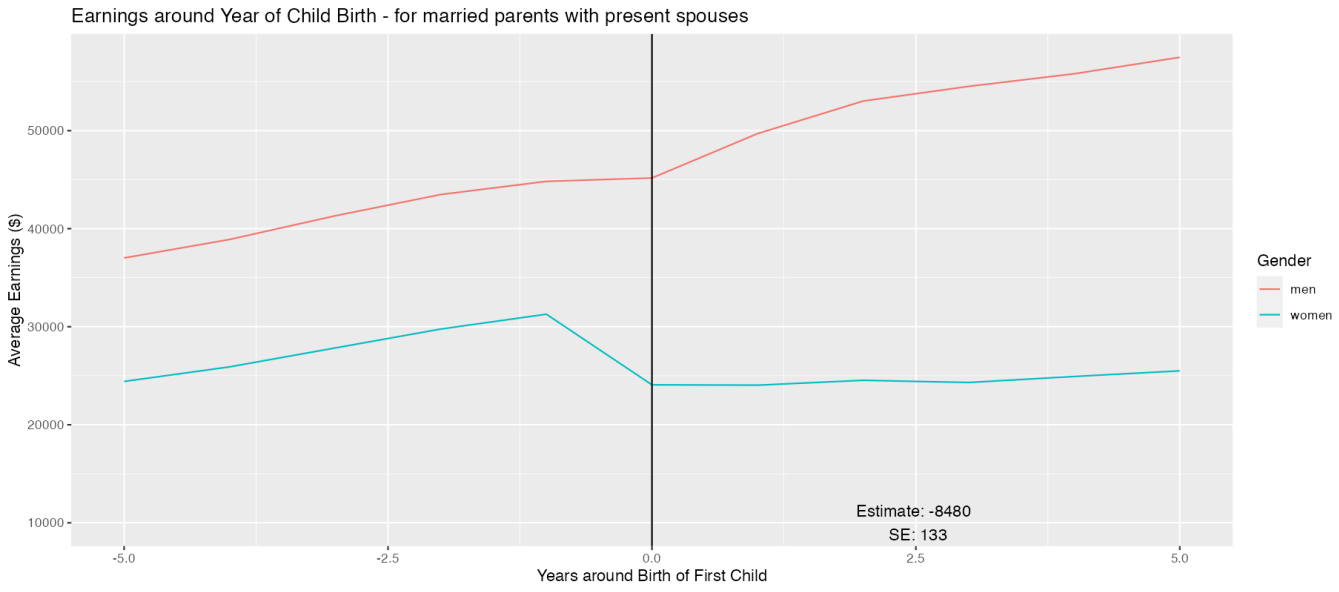
6. Plot child penalty over time using the child_penalty_df dataframe.

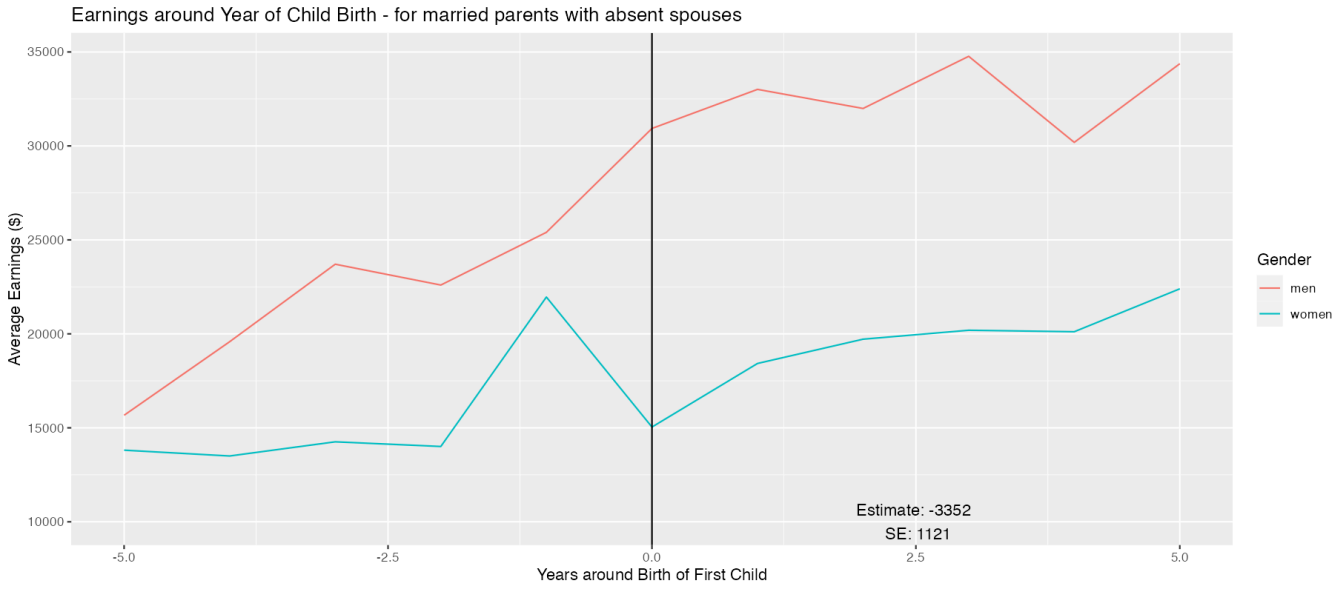


7. Explore the differences in the child penalty along at least two other dimensions.

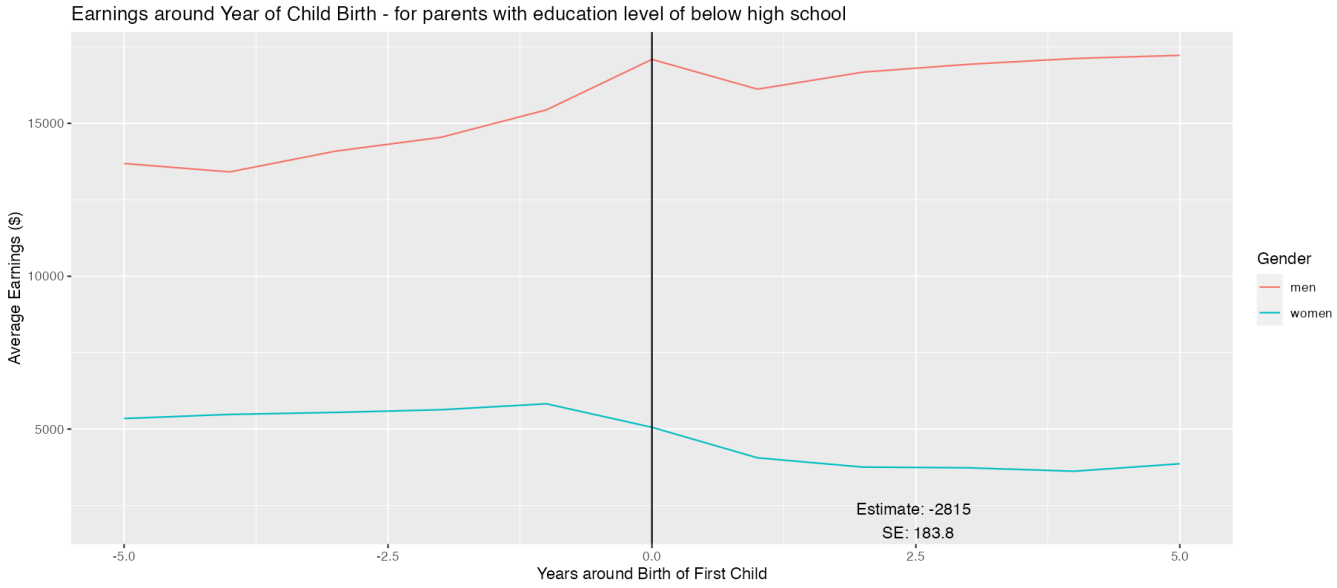
**note: no writing required for 7, but you will need to write about Q7 results/takeaways in the narrative*

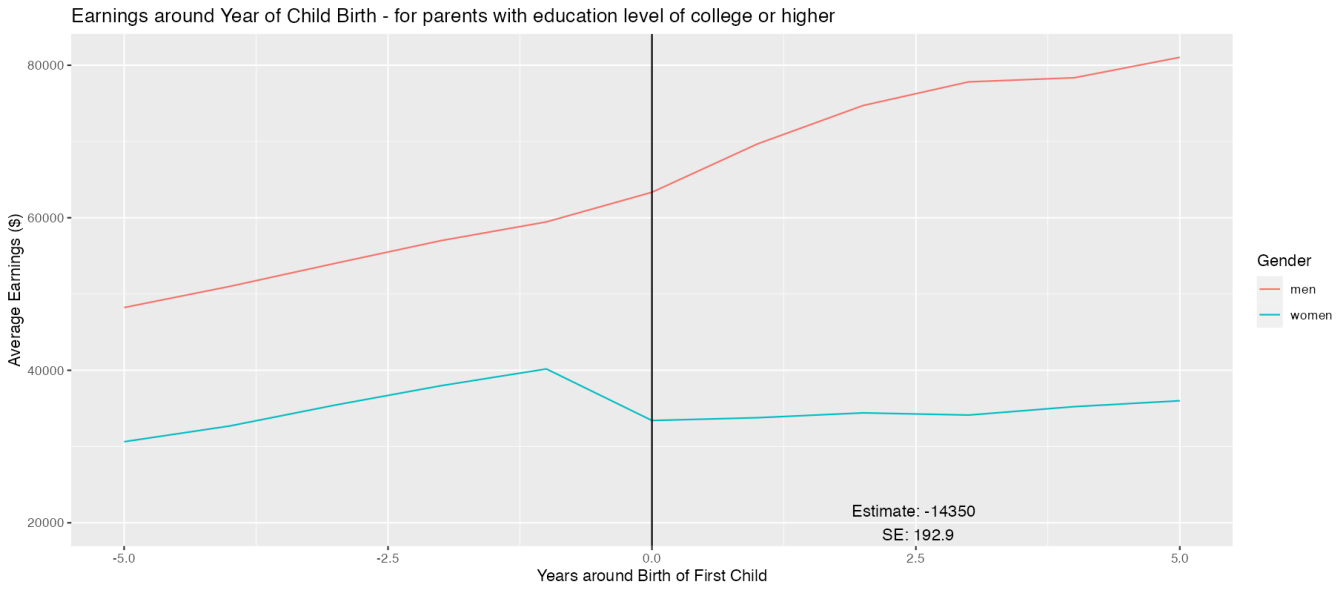
[Include plot #1 here.]



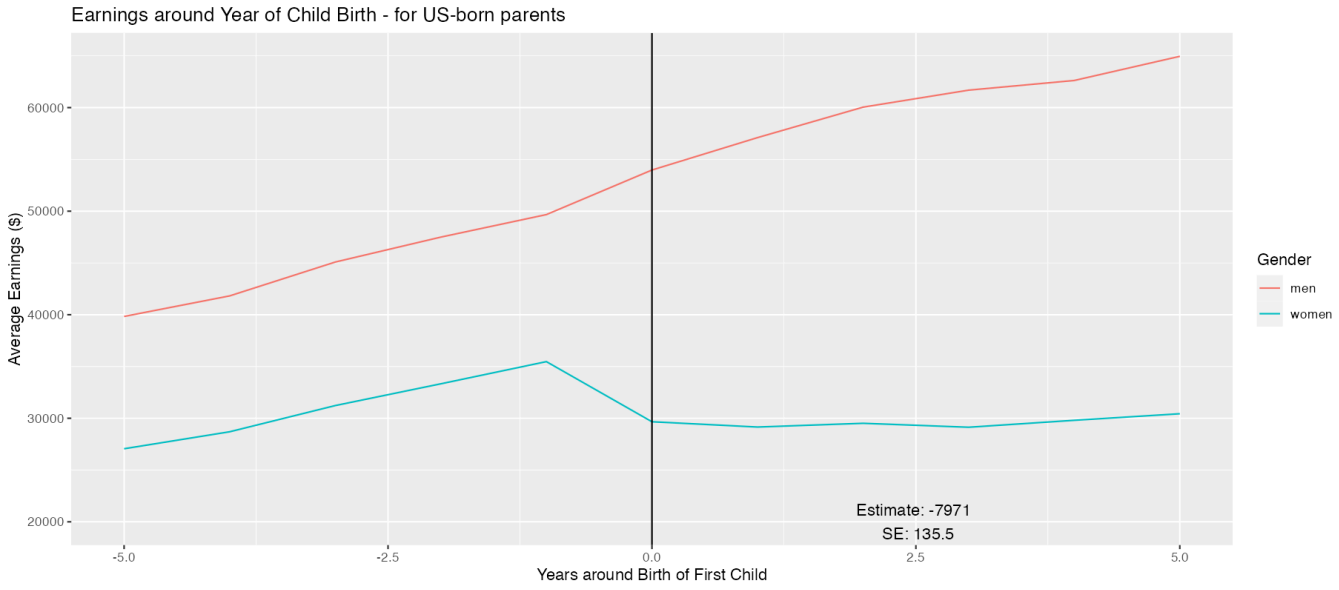


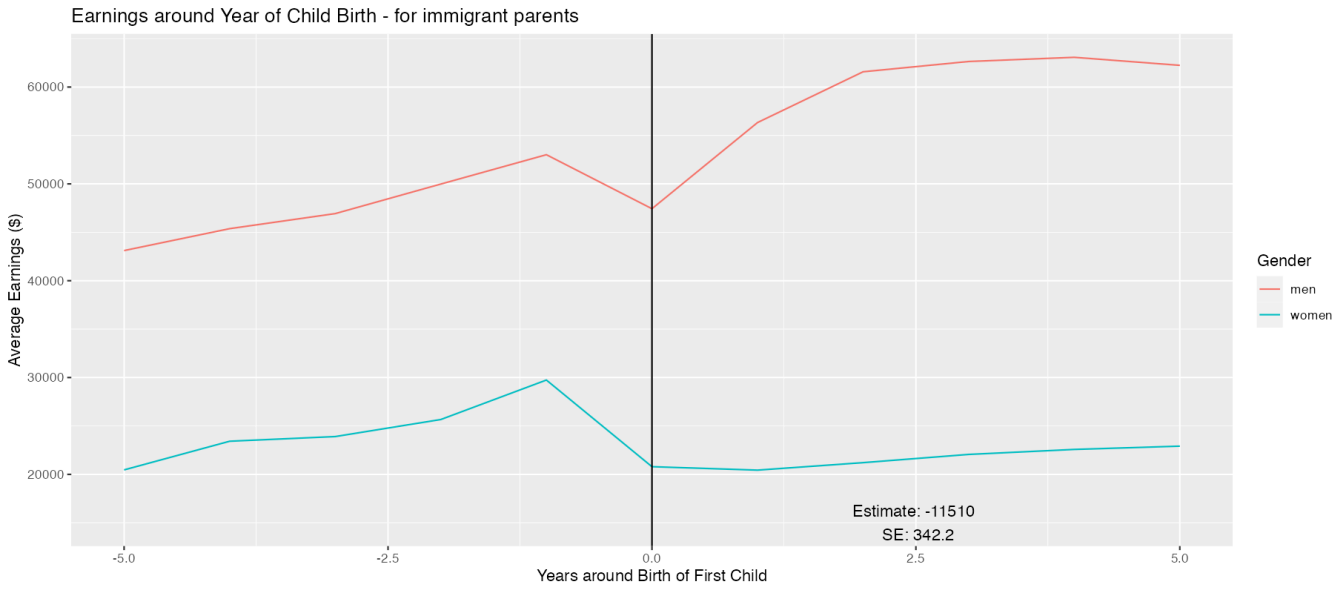
[Include plot #2 here.]





[Include plot #3 here.]




Earnings around Year of Child Birth - for immigrant parents
Average Earnings ($)
60000
50000
40000
30000
20000
-5.0
-2.5
0.0
2.5
5.0
Years around Birth of First Child
Estimate: -11510
SE: 342.2
Gender
men
women

Please include your code report here. The simplest way of doing this is to open the html file (generated from the Download Code button at the end of the HW - see below) and copy all of that text here.

Report - Exercise 5: HW

Name: Sonya Rashkovan

Explore the Data

Box 1:

Question 1

Box 1:

```
gss_data %>% filter(gender == "men") %>% summarise(wtd.mean = wtd.mean(earnings,
weights = wgt, na.rm = TRUE))
```

Box 2:

```
gss_data %>% filter(gender == "women") %>% summarise(wtd.mean = wtd.mean(earnings,
weights = wgt, na.rm = TRUE))
```

Question 2

Box 1:

```
gss_data %>% filter(gender == "men", child == 0) %>% summarise(wtd.mean =
wtd.mean(earnings, weights = wgt, na.rm = TRUE))
```

Box 2:

```
gss_data %>% filter(gender == "women", child == 0) %>% summarise(wtd.mean =
wtd.mean(earnings, weights = wgt, na.rm = TRUE))
```

Box 3:

```
gss_data %>% filter(gender == "men", child == 1) %>% summarise(wtd.mean =
wtd.mean(earnings, weights = wgt, na.rm = TRUE))
```

Box 4:

```
gss_data %>% filter(gender == "women", child == 1) %>% summarise(wtd.mean =
wtd.mean(earnings, weights = wgt, na.rm = TRUE))
```

Question 3

Box 1:

```
gss_data %>%
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE)) %>%
  ggplot(aes(year_around_birth, earnings, color = gender)) +
  geom_line() +
  labs(title = "Earnings around Year of Child Birth",
       x = "Years around Birth of First Child",
       y = "Average Earnings",
       color = "Gender") +
  geom_vline(xintercept = 0)
```

Question 4

Box 1:

```
reg <- lm(formula = earnings ~ child*gender, data = gss_data)
summary(reg)
```

Box 2:

```
reg <- lm(formula = earnings ~ child*gender, data = gss_data)



gss_data %>%
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE)) %>%
  ggplot(aes(year_around_birth, earnings, color = gender)) +
  geom_line() +
```

```
  labs(title = "Earnings around Year of Child Birth",

       x = "Years around Birth of First Child",

       y = "Average Earnings ($)",

       color = "Gender") +

  geom_vline(xintercept = 0) +

  annotate(geom="text", x = 2.5, y = 0.85, label =

             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",

                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Question 5

Box 1:

```
reg <- gss_data %>%

  filter(child_birth_year == 1980) %>%

  lm(formula = employed ~ child*gender)



gss_data %>%

  filter(child_birth_year == 1980) %>%

  group_by(year_around_birth, gender) %>%

  dplyr::summarize(avg_employed = wtd.mean(employed, weights = wgt, na.rm = TRUE))
%>%

  ggplot(aes(year_around_birth, avg_employed, color = gender)) +

  geom_line() +

  labs(title = "Employment Rate around Year of Child Birth - for Parents of children
born in 1980",

       x = "Years around Birth of First Child",
```

```
       y = "Average Employment Rate",

       color = "Gender") +

  geom_vline(xintercept = 0) +

  annotate(geom="text", x = 2.5, y = 0.85, label =

             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",

                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Box 2:

```
reg <- gss_data %>%

  filter(child_birth_year == 2015) %>%

  lm(formula = employed ~ child*gender)



gss_data %>%

  filter(child_birth_year == 2015) %>%

  group_by(year_around_birth, gender) %>%

  dplyr::summarize(avg_employed = wtd.mean(employed, weights = wgt, na.rm = TRUE))
%>%

  ggplot(aes(year_around_birth, avg_employed, color = gender)) +

  geom_line() +

  labs(title = "Employment Rate around Year of Child Birth - for Parents of children
born in 2015",

       x = "Years around Birth of First Child",

       y = "Average Employment Rate",
```

```
      color = "Gender") +
  geom_vline(xintercept = 0) +
  annotate(geom="text", x = 2.5, y = 0.85, label =
             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",
                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Question 6

Box 1:

```
child_penalty_df %>%
  ggplot(aes(year, child_penalty)) +
  geom_point() +
  labs(title = "Child Penalty as a function of Time",
       x = "Time",
       y = "Child Penalty")
```

Question 7

Box 1:

```
reg <- gss_data %>%
  filter(marital_status == 3) %>%
  lm(formula = earnings ~ child*gender)


gss_data %>%
  filter(marital_status == 3) %>%
  group_by(year_around_birth, gender) %>%
```

```
  dplyr::summarize(avg_earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE))
%>%

  ggplot(aes(year_around_birth, avg_earnings, color = gender)) +

  geom_line() +

  labs(title = "Earnings around Year of Child Birth - for married parents with
present spouses ",

       x = "Years around Birth of First Child",

       y = "Average Earnings ($)",

       color = "Gender") +

  geom_vline(xintercept = 0) +

  annotate(geom="text", x = 2.5, y = 10000, label =

             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",

                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Box 1:

```
reg <- gss_data %>%
  filter(marital_status == 2) %>%
  lm(formula = earnings ~ child*gender)


gss_data %>%
  filter(marital_status == 2) %>%
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(avg_earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE)) %>%
  ggplot(aes(year_around_birth, avg_earnings, color = gender)) +
  geom_line() +
  labs(title = "Earnings around Year of Child Birth - for married parents with absent
spouses ",
       x = "Years around Birth of First Child",
       y = "Average Earnings ($)",
       color = "Gender") +
```

```
  geom_vline(xintercept = 0) +
  annotate(geom="text", x = 2.5, y = 10000, label =
             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",
                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Box 2:

```
reg <- gss_data %>%
  filter(edlevel == 1) %>%
  lm(formula = earnings ~ child*gender)



gss_data %>%
  filter(edlevel == 1) %>%
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(avg_earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE))
%>%
  ggplot(aes(year_around_birth, avg_earnings, color = gender)) +
  geom_line() +
  labs(title = "Earnings around Year of Child Birth - for parents with education
level of below high school  ",
       x = "Years around Birth of First Child",
       y = "Average Earnings ($)",
       color = "Gender") +
  geom_vline(xintercept = 0) +
  annotate(geom="text", x = 2.5, y = 2000, label =
             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",
```

```
                "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Box 2:

```
reg <- gss_data %>%
  filter(edlevel == 4) %>%
  lm(formula = earnings ~ child*gender)


gss_data %>%
  filter(edlevel == 4) %>%
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(avg_earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE)) %>%
  ggplot(aes(year_around_birth, avg_earnings, color = gender)) +
  geom_line() +
  labs(title = "Earnings around Year of Child Birth - for parents with education level
of college or higher ",
       x = "Years around Birth of First Child",
       y = "Average Earnings ($)",
       color = "Gender") +
  geom_vline(xintercept = 0) +
  annotate(geom="text", x = 2.5, y = 20000, label =
             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",
                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Box 3:

```
reg <- gss_data %>%
  filter(immigrant == 0) %>%
  lm(formula = earnings ~ child*gender)


gss_data %>%
  filter(immigrant == 0) %>%
```

```
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(avg_earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE))
%>%
  ggplot(aes(year_around_birth, avg_earnings, color = gender)) +
  geom_line() +
  labs(title = "Earnings around Year of Child Birth - for US-born parents ",
       x = "Years around Birth of First Child",
       y = "Average Earnings ($)",
       color = "Gender") +
  geom_vline(xintercept = 0) +
  annotate(geom="text", x = 2.5, y = 20000, label =
             paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",
                   "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Box 3:

```
reg <- gss_data %>%
  filter(immigrant == 1) %>%
  lm(formula = earnings ~ child*gender)



gss_data %>%
  filter(immigrant == 1) %>%
  group_by(year_around_birth, gender) %>%
  dplyr::summarize(avg_earnings = wtd.mean(earnings, weights = wgt, na.rm = TRUE))
%>%
  ggplot(aes(year_around_birth, avg_earnings, color = gender)) +
```

```
geom_line() +

labs(title = "Earnings around Year of Child Birth - for immigrant parents ",

     x = "Years around Birth of First Child",

     y = "Average Earnings ($)",

     color = "Gender") +

geom_vline(xintercept = 0) +

annotate(geom="text", x = 2.5, y = 15000, label =

           paste("Estimate:", signif(summary(reg)[["coefficients"]][4], 4), "\n",

                 "SE:",  signif(summary(reg)[["coefficients"]][8], 4)))
```

Please submit this entire document to Gradescope. **You must select the corresponding pages for each answer when uploading this document to Gradescope.** Failure to do so makes it difficult to locate your answers.