HW6_Markdown

2024-03-07

Manay Divatia

md46245

1

a

```r
justices = read.csv("~/Downloads/justices.csv")
library(dplyr)
```

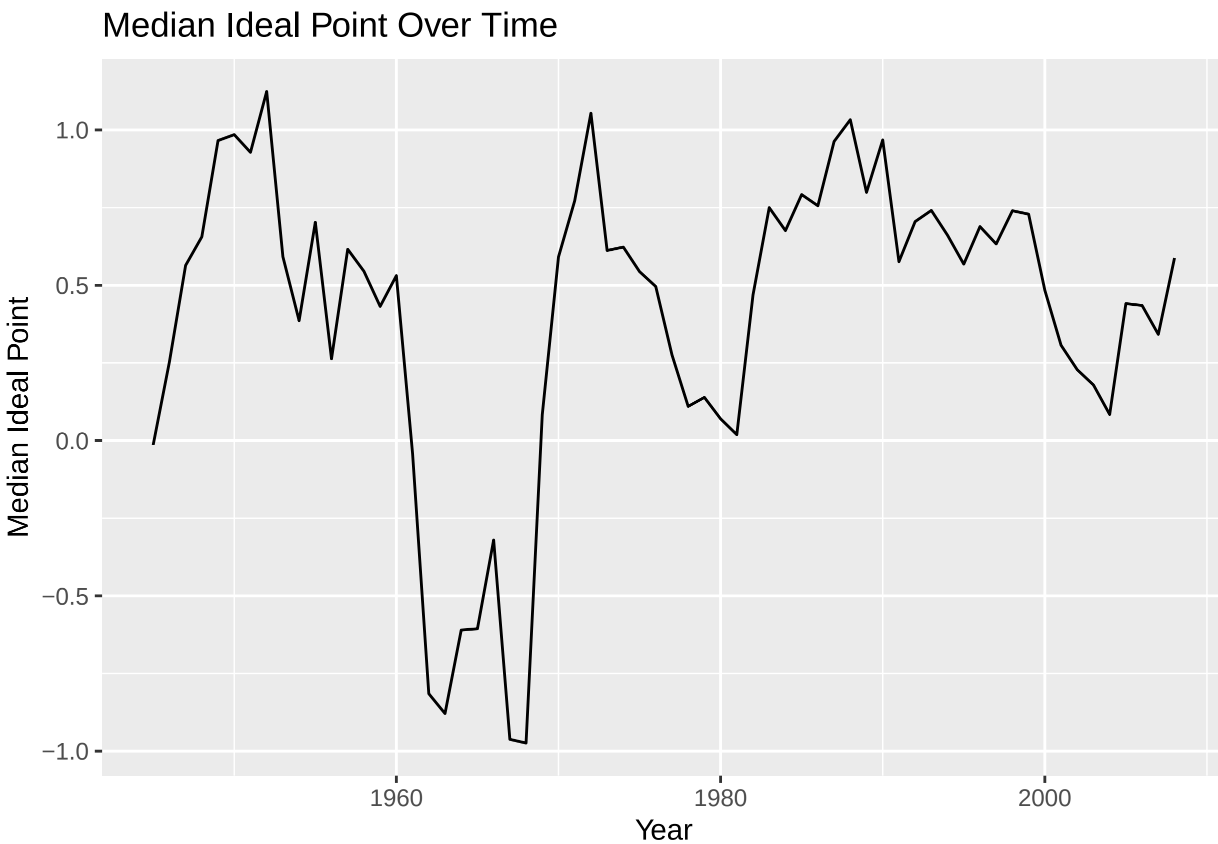
```
##
## Attaching package: 'dplyr'

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

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

```r
median_ideal_by_year <- justices %>%
  group_by(term) %>%
  summarize(median_ideal_point = median(idealpt, na.rm = TRUE))
median_ideal_by_year
```

```
## # A tibble: 64 x 2
##     term median_ideal_point
##    <int>              <dbl>
##  1  1945             -0.014
##  2  1946              0.253
##  3  1947              0.564
##  4  1948              0.656
##  5  1949              0.966
##  6  1950              0.985
##  7  1951              0.928
##  8  1952              1.12
##  9  1953              0.592
## 10  1954              0.386
## # i 54 more rows
```

```r
library(ggplot2)
ggplot(median_ideal_by_year, aes(x = term, y = median_ideal_point)) +
  geom_line() +
  labs(title = "Median Ideal Point Over Time",
       x = "Year",
```

```
       y = "Median Ideal Point")
```



b

```
merged_data <- merge(justices, median_ideal_by_year, by = "term")

justice_with_median <- merged_data %>%
  filter(idealpt == median_ideal_point) %>%
  group_by(term) %>%
  slice(which.min(abs(idealpt - median_ideal_point))) %>%
  ungroup()

justice_counts <- table(justice_with_median$justice)

most_common_justice <- names(justice_counts)[which.max(justice_counts)]

print(justice_with_median)
```

```
## # A tibble: 64 x 6
##     term justice     idealpt pparty pres       median_ideal_point
##    <int> <chr>         <dbl> <chr>  <chr>                   <dbl>
##  1  1945 Reed         -0.014 D      Truman                 -0.014
##  2  1946 Reed          0.253 D      Truman                  0.253
##  3  1947 Reed          0.564 D      Truman                  0.564
##  4  1948 Frankfurter   0.656 D      Truman                  0.656
##  5  1949 Burton        0.966 D      Truman                  0.966
```

```
##   6  1950 Burton        0.985 D      Truman                   0.985
##   7  1951 Burton        0.928 D      Truman                   0.928
##   8  1952 Clark         1.12  D      Truman                   1.12
##   9  1953 Clark         0.592 R      Eisenhower               0.592
##  10  1954 Frankfurter   0.386 R      Eisenhower               0.386
## # i 54 more rows
```

```r
print(justice_counts)
```

```
##
##       Black    Blackmun     Brennan      Burton       Clark Frankfurter
##           3           2           1           3           6           3
##    Goldberg      Harlan     Kennedy    Marshall    O'Connor      Powell
##           2           1           9           2           8           3
##        Reed      Souter     Stewart       White
##           3           2           3          13
```

```r
print(paste("Justice with the most appearances:", most_common_justice))
```

```
## [1] "Justice with the most appearances: White"
```

White was the justice that had the median ideal point the most. This justice served on the Court for 13 terms. His average ideal point over his entire tenure was .772.

c

```r
democratic_presidents <- unique(justices$pres[justices$pparty == "D"])
republican_presidents <- unique(justices$pres[justices$pparty == "R"])
democratic_ideology_shifts <- numeric(0)
republican_ideology_shifts <- numeric(0)
```

d

```r
for(president in democratic_presidents) {
  president_data <- justice_with_median %>% filter(pres == president)
  ideology_shift <- last(president_data$median_ideal_point) - first(president_data$median_ideal_point)
  democratic_ideology_shifts <- c(democratic_ideology_shifts, ideology_shift)
}
for(president in republican_presidents) {
  president_data <- justice_with_median %>% filter(pres == president)
  ideology_shift <- last(president_data$median_ideal_point) - first(president_data$median_ideal_point)
  republican_ideology_shifts <- c(republican_ideology_shifts, ideology_shift)
}

print("Democratic Ideology Shifts:")
```

```
## [1] "Democratic Ideology Shifts:"
```

```r
print(democratic_ideology_shifts)
```

```
## [1]  1.138 -0.837 -0.364 -0.206 -0.257
```

```r
print("Republican Ideology Shifts:")
```

```
## [1] "Republican Ideology Shifts:"
```

```r
print(republican_ideology_shifts)
```

```
## [1] -0.06099999  0.54000003 -0.04800004  1.01399999 -0.09400004  0.28099999
```

e

```r
mean(democratic_ideology_shifts)
```

```
## [1] -0.1052
```

```r
sd(democratic_ideology_shifts)
```

```
## [1] 0.7384542
```

```r
mean(republican_ideology_shifts)
```

```
## [1] 0.272
```

```r
sd(republican_ideology_shifts)
```

```
## [1] 0.4403894
```

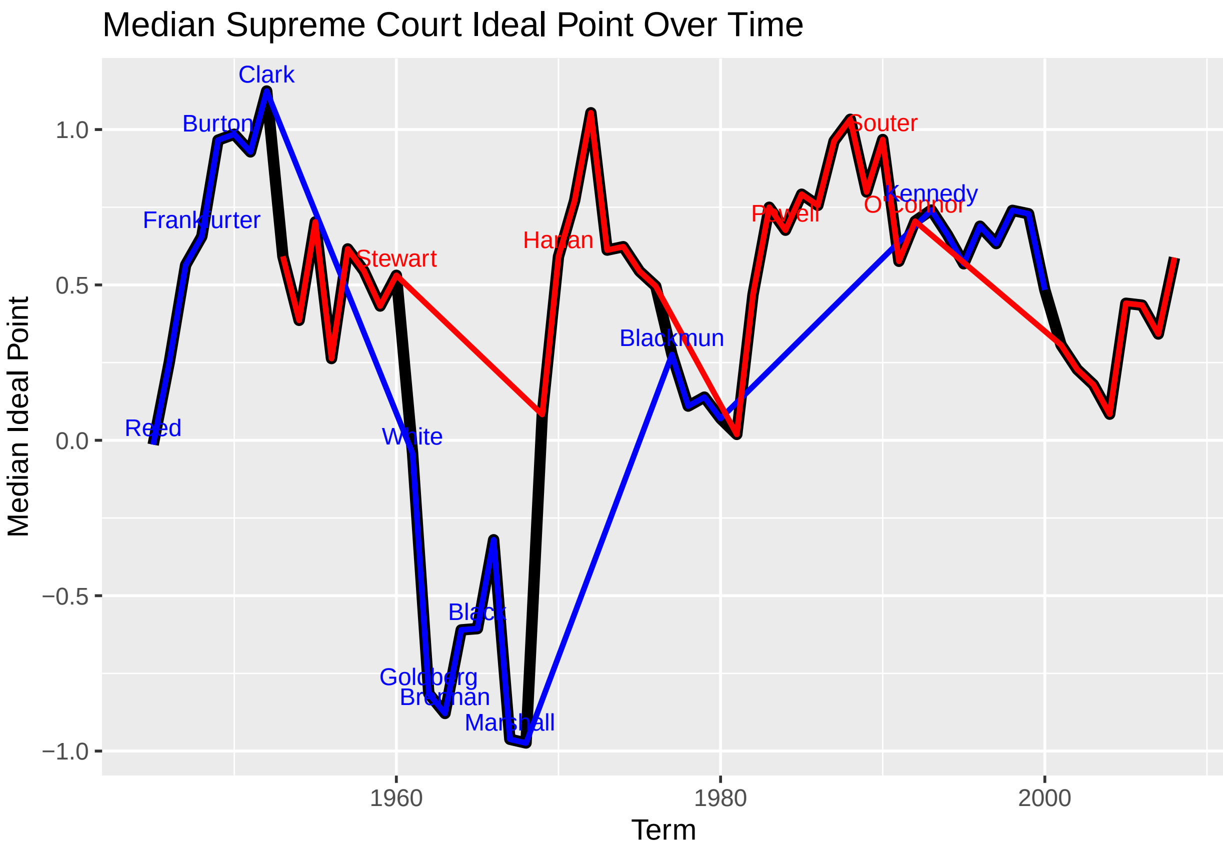
Reagan had the largest conservative shift on the Court and Kennedy had the largest liberal shift on the Court. # f

```r
plot <- ggplot(median_ideal_by_year, aes(x = term, y = median_ideal_point)) +
  geom_line(color = "black", linewidth = 2) +  # Black line for overall median ideal point
  labs(title = "Median Supreme Court Ideal Point Over Time",
       x = "Term",
       y = "Median Ideal Point")

plot <- plot +
  geom_line(data = justice_with_median, aes(x = term, y = idealpt, color = pparty), size = 1) +
  geom_text(data = justice_with_median %>% distinct(justice, .keep_all = TRUE),
            aes(x = term, y = idealpt, label = justice, color = pparty),
            vjust = -0.5, hjust = 0.5, size = 3) +
  scale_color_manual(values = c("blue", "red"), guide = FALSE)  # Blue for Democratic, Red for Republic
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
# Show the plot
print(plot)
```

```
## Warning: The `guide` argument in `scale_*()` cannot be `FALSE`. This was deprecated in
## ggplot2 3.3.4.
## i Please use "none" instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



The plot has a lot of data within it. What we can see though is how the median ideal point follows one party or the other. We can see how vast the difference is between democratic and republican justices. ## 2

a

```r
mother_df = read.csv("~/Downloads/yu2017sample.csv")
length(unique(mother_df$PUBID))
```

```
## [1] 1484
```

```r
summary(mother_df)
```

```
##        X             PUBID           year          wage          
##  Min.   :    3   Min.   :  15   Min.   :1998   Min.   :     5.0  
##  1st Qu.: 4477   1st Qu.:2347   1st Qu.:2003   1st Qu.:   702.0  
##  Median : 8963   Median :4427   Median :2006   Median :   972.5  
##  Mean   : 9011   Mean   :4487   Mean   :2006   Mean   :  1250.6  
##  3rd Qu.:13505   3rd Qu.:6626   3rd Qu.:2009   3rd Qu.:  1400.0  
##  Max.   :18282   Max.   :9022   Max.   :2013   Max.   :220000.0  
##       age         numChildren         educ              school       
##  Min.   :16.00   Min.   :0.0000   Length:11572       Mode :logical  
##  1st Qu.:21.00   1st Qu.:0.0000   Class :character   FALSE:7979     
##  Median :24.00   Median :0.0000   Mode  :character   TRUE :3593     
##  Mean   :23.97   Mean   :0.6251                                     
##  3rd Qu.:27.00   3rd Qu.:1.0000                                     
##  Max.   :34.00   Max.   :6.0000                                     
##    experience        tenure            tenure2            fullTime      
##  Min.   :  10.0   Min.   : 0.1346   Min.   :  0.01812   Mode :logical  
```

```
## 1st Qu.: 166.0   1st Qu.: 0.6923   1st Qu.:  0.47929   FALSE:4487
## Median : 309.0   Median : 1.4808   Median :  2.19268   TRUE :7085
## Mean   : 329.8   Mean   : 2.2720   Mean   : 10.05661
## 3rd Qu.: 471.0   3rd Qu.: 3.0769   3rd Qu.:  9.46746
## Max.   :1024.0   Max.   :17.6154   Max.   :310.30178
## multipleLocations   unionized          marstat              region
## Min.   :0.0000    Min.   :0.00000   Length:11572       Min.   :1.000
## 1st Qu.:0.0000    1st Qu.:0.00000   Class :character   1st Qu.:2.000
## Median :1.0000    Median :0.00000   Mode  :character   Median :3.000
## Mean   :0.6732    Mean   :0.09722                      Mean   :2.664
## 3rd Qu.:1.0000    3rd Qu.:0.00000                      3rd Qu.:3.000
## Max.   :1.0000    Max.   :1.00000                      Max.   :4.000
##     urban           industry            autonomy     competitiveness
## Min.   :0.0000   Length:11572       Min.   :2.694   Min.   :1.499
## 1st Qu.:1.0000   Class :character   1st Qu.:3.620   1st Qu.:2.529
## Median :1.0000   Mode  :character   Median :3.978   Median :2.745
## Mean   :0.8053                      Mean   :3.879   Mean   :2.841
## 3rd Qu.:1.0000                      3rd Qu.:4.137   3rd Qu.:3.231
## Max.   :1.0000                      Max.   :4.784   Max.   :4.640
##   hazardous       regularility       teamwork
## Min.   :1.000   Min.   :1.000   Min.   :2.467
## 1st Qu.:1.097   1st Qu.:1.076   1st Qu.:4.162
## Median :1.242   Median :1.204   Median :4.293
## Mean   :1.421   Mean   :1.219   Mean   :4.246
## 3rd Qu.:1.558   3rd Qu.:1.280   3rd Qu.:4.456
## Max.   :4.146   Max.   :2.400   Max.   :4.934
```

```r
length(unique(mother_df$PUBID)) / 15
```

```
## [1] 98.93333
```

One advantage of using a contemporary cohort of women rather than an older cohort is that we can make better claims about our current time rather than a previous time. One disadvantage, though, is that we are limited in the data that we have since the cohort is contemporary. # b

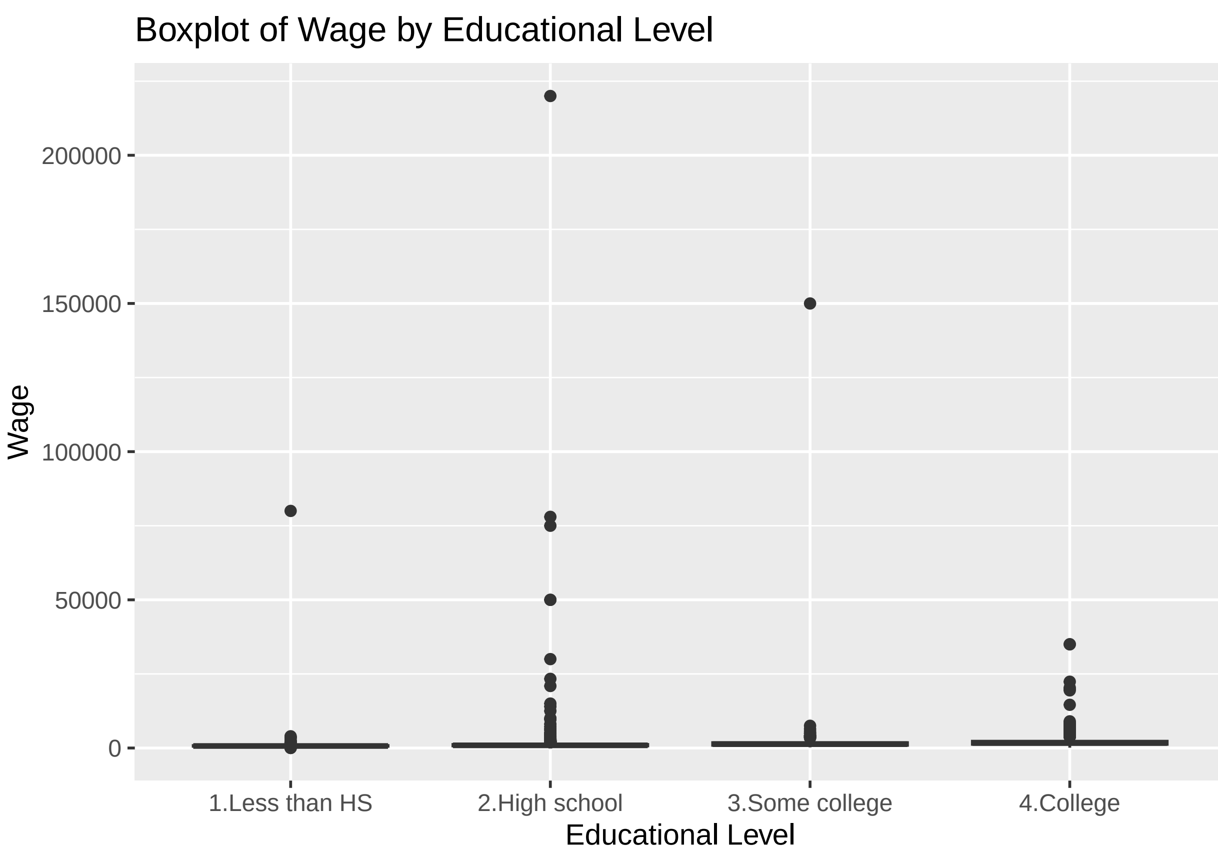
```r
children_counts <- table(mother_df$numChildren)
```

We see from the table that the majority of women have no children. From there, there is less and less observations as the number of children increase. This could be some bias in our data of who are who aren't mothers. # c

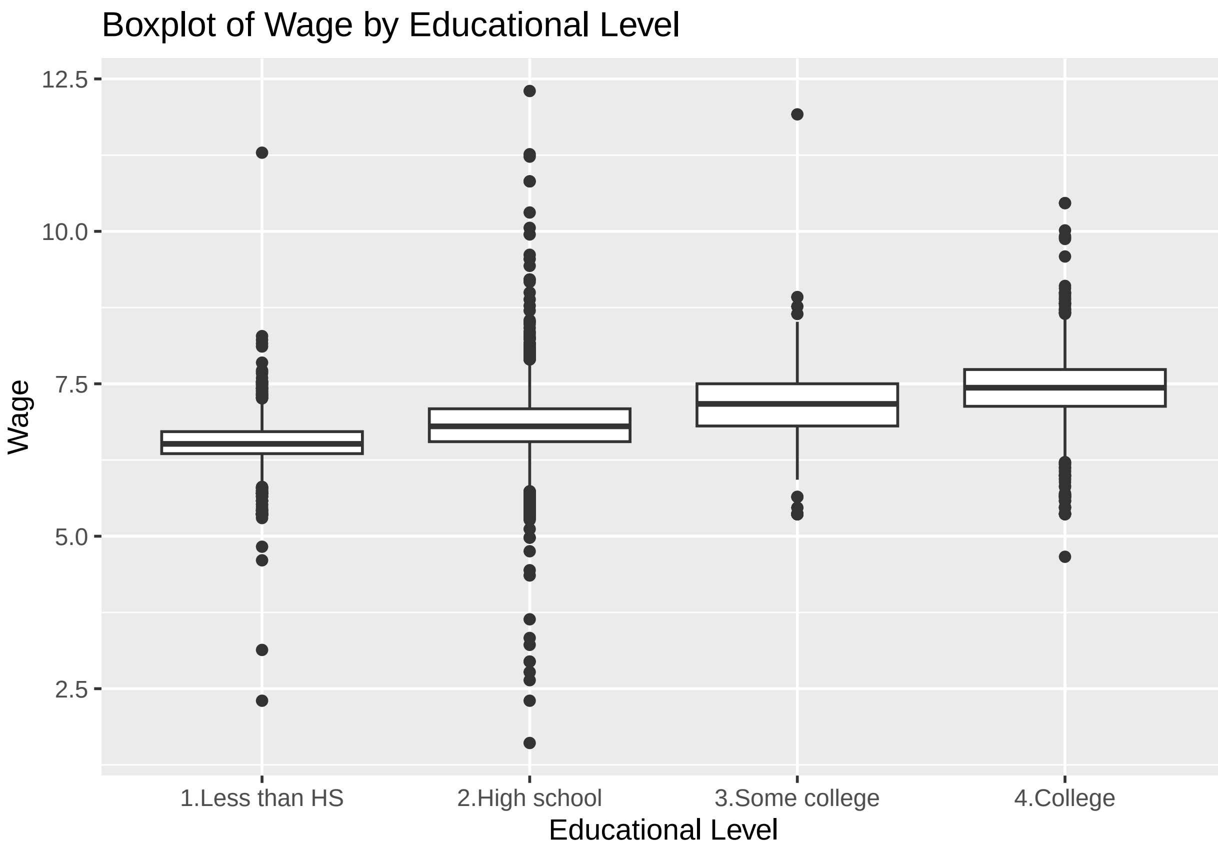
```r
mother_df$isMother <- ifelse(mother_df$numChildren > 0, 1, 0)
has_children_count <- table(mother_df$isMother)
```

Again we can see that there are far more women who aren't mothers than those who are which could be something to consider. # d

```r
mother_df$logwage <- log(mother_df$wage)
wage_plot <- ggplot(mother_df, aes(x = factor(educ), y = wage)) +
  geom_boxplot() +
  labs(title = "Boxplot of Wage by Educational Level",
       x = "Educational Level",
       y = "Wage")
wage_plot
```




```
logwage_plot <- ggplot(mother_df, aes(x = factor(educ), y = logwage)) +
  geom_boxplot() +
  labs(title = "Boxplot of Wage by Educational Level",
       x = "Educational Level",
       y = "Wage")
logwage_plot
```



From the plots, we can see how as the women gain more and more education, their wage increases with the median seemingly linearly increasing. # e

```
logwage_plot <- ggplot(mother_df, aes(x = year, y = logwage, color = factor(isMother))) +
  geom_line(stat = "summary", fun = "mean", size = 1, linetype = "solid") +
  labs(title = "Mean Logwage Over Time for Mothers and Non-Mothers",
       x = "Year",
       y = "Mean Logwage",
       color = "Mother Status") +
  scale_color_manual(values = c("blue", "red"), labels = c("Non-Mothers", "Mothers")) +
  theme_minimal()
```

Until 2005, the mean logwage seemed to be somewhat equal between non-mothers and mothers. However, after that the data shows that non-mothers make more money than mothers. In 2012, non-mothers had a mean logwage of 7.5 while mothers had 7.25. # f

```
woman_fe <- as.factor(mother_df$PUBID)
year_fe <- as.factor(mother_df$year)

fixed_effects_model <- lm(logwage ~ numChildren + fullTime + multipleLocations + unionized + industry,
#summary(fixed_effects_model)
```

The coefficient of numChildren is 0.0157. I think this is small but it makes sense considering there are so many variables that could play a role in logwage.