

# Final Class Project: Analyzing the Gender Wage Gap

*Ananthset, Iris*

*December 8, 2016*

Women earn less than men. This gender wage gap has been prevalent throughout history, and it is still an issue today. However, is this due to differences in their characteristics, such as education, age, and productivity? Or, are the employers actually paying employees differently because of their gender? This paper will provide a better understanding of the gender wage gap. The data set contains 23,144 observations - 15,058 males and 8,086 females. Below is a table that summarizes the data set.

**Table 1**

##	All	Female	Male	T-stat
## Hourly Wage	1.567162	1.438668	1.636162	-29.854182
## Age	36.404727	36.315113	36.452849	-1.199075
## Education	8.723989	9.151496	8.494422	12.548892
## Productivity	3.071540	2.944748	3.139626	-24.022307

As shown from this table, men do earn more than women on average. The average change in real hourly wages for both men and women is \$1.57. The average change in real hourly wages for men (\$1.64) is higher than the overall average change while the average for women (\$1.44) is lower than it. The difference between the change in real hourly wages for men and women is significant (t-stat of -29.85), meaning that the hypothesis of men and women earn the same amount can be rejected.

However, there are indeed differences in the characteristics of men and women, specifically education, age, and productivity. The employees in the data set, on average, are approximately 36 years old. The average age for male employees and female employees is also about 36 years old. Age is associated with experience; older employees tend to have more experience. Therefore, the gender wage gap may not be due to difference in experience between men and women. On the other hand, the differences in the other characteristics between men and women are significant, so they might be the reason for the gender wage gap. The employees in the data set have a mean of 8.72 years of education. Women, on average, have higher education (9.15 years) than both the overall average and average education of men (8.49 years). Typically, there is a positive correlation between education and wages. If women, on average, have higher education than men, why are men paid more? The answer to this is that men tend to work at more productive firms than women. The data measures productivity using the change in output per worker at the firm. The firms men work in have higher change in output (3.14) than both the overall average firms (3.07) and the firms of women (2.94).

Figure 1a: Histogram of Log Hourly Wages

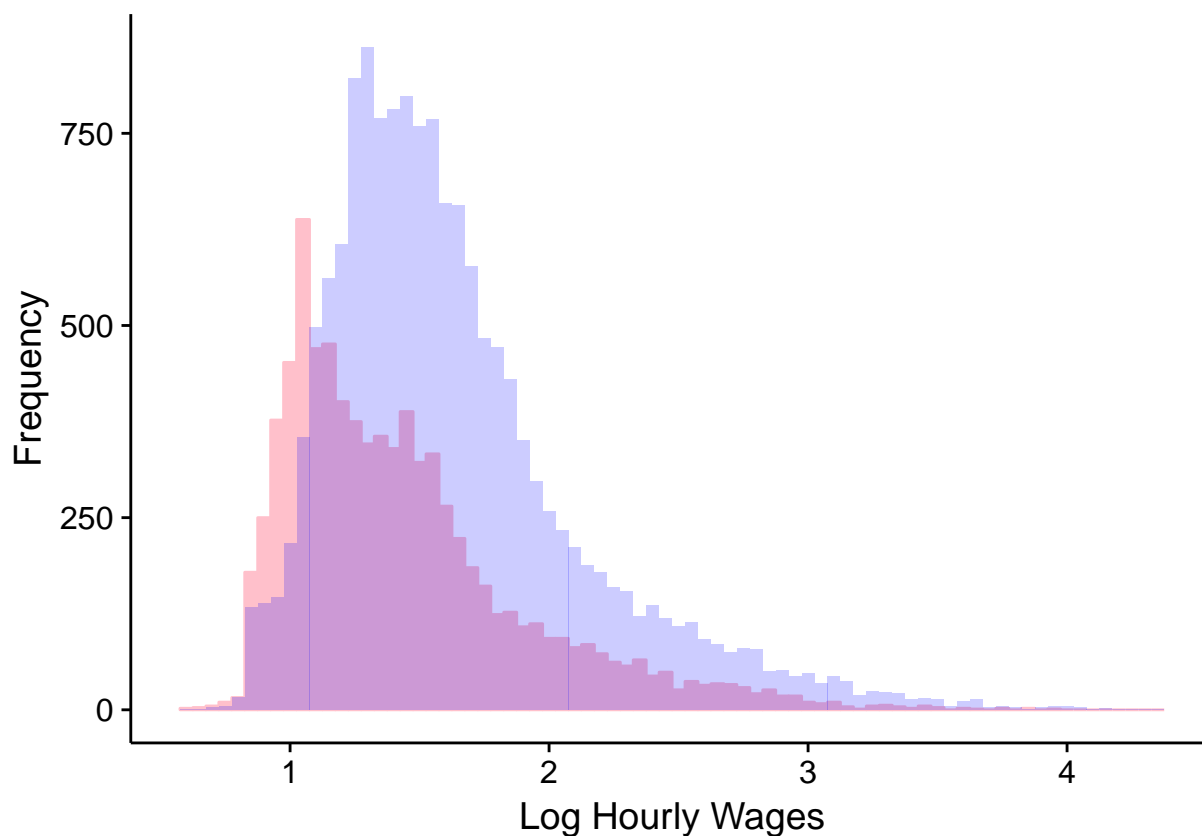


Figure 1a illustrates the distribution of log hourly wages for both male (blue) and female (pink) employees. For both men and women, the graph is right-skewed. This means that there are more people earning less than the average. The data set has twice as many observations for men than women, hence there is more blue on the graph. Despite this bias, there is a portion of the histogram for women that does not align with the histogram for men (purely pink part). This portion is towards the lower values of the log hourly wages. This means there are a lot more women earning lower hourly wages compared to men. There is also a huge portion of the histogram for men that does not align with the histogram for women (purely blue part). This portion is towards the higher values of log hourly wages. This means that there are many male employees earning higher hourly wages than female employees.

Figure 1b: Histogram of Productivity



Figure 1b illustrates the distribution of the productivity of the firms men (blue) and women (pink) work in. In contrast to Figure 1a, the graph is normally distributed. This means that people, on average, work at firms that produce the average output. Again, since data set has twice as many observations for men than women, there is more blue on the graph. Despite this bias, there is a portion of the histogram for women that does not align with the histogram for men (purely pink part). Again, this is observed at the lower values of the firm's log of output per worker. This means more women are working at firms that are considered less productive. There is a significant portion of the histogram for men that does not align with the histogram for women (purely blue part). This means that there are a lot more men working at firms considered as more productive.

To summarize, the following are the main findings from the tables and figures above:

- 1) The gender wage gap is real. On average, men earn higher wages than women.
- 2) A possible explanation for the gender wage gap is the difference in the firms men and women tend to work in. Women tend to work at firms that are less productive compared to the firms men work in.

What is the relationship between wage, education, and age? The following is are hypothesized models that show the relationship. They are simple OLS regressions with education and cubic age as the covariates.

Table 2: Series of Standard Wage Models

```
##
## =====
##                               Dependent variable:
```

```

##          -----
##                  Log Hourly Wages
##          (1)      (2)      (3)      (4)
## -----
## educ          0.083*** 0.081*** 0.086***
##                (0.001)  (0.001)  (0.001)
##
## age            0.095*** 0.110*** 0.064***
##                (0.015)  (0.020)  (0.024)
##
## age2           -0.002*** -0.002*** -0.001
##                (0.0004)  (0.001)  (0.001)
##
## age3           0.00001** 0.00001** 0.00000
##                (0.00000) (0.00000) (0.00001)
##
## female        -0.197*** -0.251***
##                (0.007)  (0.005)
##
## Constant      1.636*** -0.859*** -1.065*** -0.675**
##                (0.004)  (0.189)  (0.244)  (0.295)
##
## -----
## Observations  23,144    23,144    15,058    8,086
## R2             0.036     0.400     0.349     0.442
## Adjusted R2   0.036     0.400     0.349     0.441
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01

```

Both covariates have a positive relationship with wages. An additional year in education increases change in wages by 8.3%, holding all other variables constant. For men, the returns to education is less than the returns to education for women. An additional year in age also has a positive effect on wages, holding all the other variables constant.

In the first two columns, the regression models fit the data for both men and women. The only difference is that the first column (1) only uses the female dummy as the control variable while the second column (2) uses the female dummy as well as the other characteristic variables (education and age). The coefficient of females in (1) is smaller in magnitude than the coefficient for females in (2),  $|-0.197| > |-0.251|$ . This is because wages is not only explained by the gender of the worker; there are other explanatory variables such as education level and age which (2) considers.

In the third column (3), the regression model fits the data for only men while the fourth column (4) fits the data for only women. Both models include are the explanatory variables that were used in (2). The coefficient of education in (3) is smaller than the coefficient of education in (4),  $0.081 < 0.086$ . This means that the returns for each year of education is smaller for men than women. If this is the case, shouldn't women earn more than men on average since women tend to have more education than men? The answer to this question is the difference in returns to age. The overall coefficient of age in (3) is bigger than the overall coefficient of age in (4),  $0.108 > 0.063$ . This means that the older men get, they not only earn more than younger men, but they also earn more than women of the same age. This poses the possibility that the the gender wage gap might be due to differences in work experience of men and women. Men have lower education levels, which means they may start working at an earlier age than women. As a result, they obtain more work experience than women which might be reflected in wages.

So, how much of the gender wage gap is explained by differences in characteristics between women and men, and how much of the gender wage gap is explained by the differences in returns for the different explanatory variables between women and men? The following Oaxaca style decomposition equation answers this question:

##	Male	Female	diffA	fcoeff
## education	8.494422	9.151496	0.6570748	8.575271e-02
## age	36.452849	36.315113	-0.1377364	6.409267e-02
## age2	1399.605725	1387.453624	-12.1521010	-8.588519e-04
## age3	56518.374485	55695.500618	-822.8738670	3.471890e-06
##	mcoeff	dotA	dotB	
## education	8.133944e-02	0.056345949	0.03748814	
## age	1.100451e-01	-0.008827897	-1.67509634	
## age2	-1.855026e-03	0.010436855	1.39425050	
## age3	1.052407e-05	-0.002856927	-0.39857762	

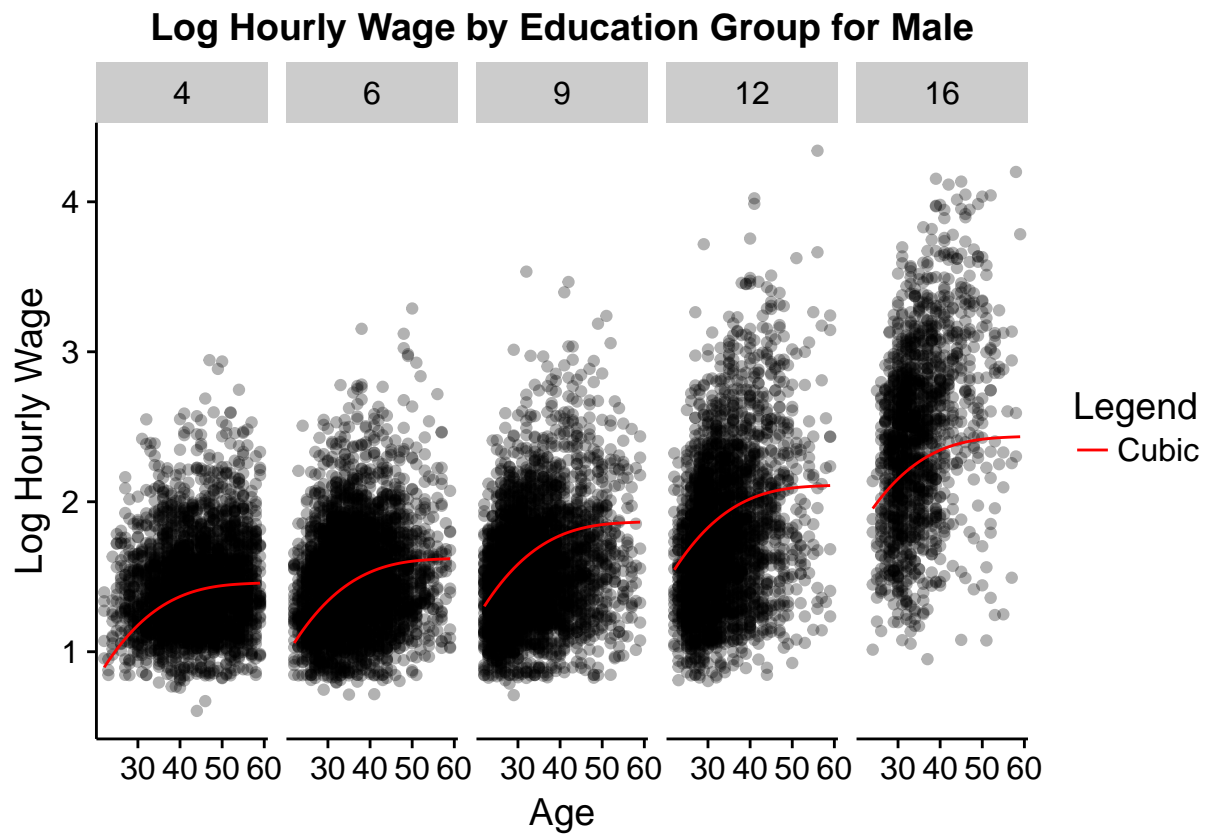
DotA is a dot product between the vector of differences in characteristics of men and women and the vector of coefficients of the explanatory variables for females is the part of the gender wage gap due solely to the differences in characteristics. DotB is a dot product between the vector of differences in coefficients of the explanatory variables for men and women and the vector of average characteristics of men is the part of the gender wage gap due solely to the differences in returns of the characteristics. The gender wage gap (-0.197) is comprised of DotA, DotB, and the difference between the constant coefficient for females and males.

The contribution of the difference in characteristics on the gender wage gap is about 0.055 out of the total -0.197. Specifically, the contribution of the difference in just education is 0.054 out of the total -0.197 while the difference in age is -0.00126 out of the total -0.197. The supposed contribution of the difference in returns of the covariates is -0.641, which is greater than the gender wage gap. This is because the difference in returns of the covariates “over-explains” the wage gap by a lot. However, the portion of the gender wage gap explained by differences in the returns to education is 0.0375 out of the -0.197 while that part explained by the differences in the return to age is -0.679 out of the -0.197.

Since women tend to have a higher education level than men and their returns are greater, the contribution of the difference in education and the difference in the returns to education between men and women are opposite to the sign of the wage gap. This means that these differences should decrease the wage gap rather than contribute to it. Therefore, although education is related to wages, it does not explain the wage gap. So far, only difference in age and the difference in the returns of age explains the wage gap.

Since difference in age profile is such an important aspect of the difference in men and women and our hypothesis is that age is one of the contributing factors of the wage gap, it must be further analyzed.

Figure 2: Relationship Between Wages and Age Between Men and Women



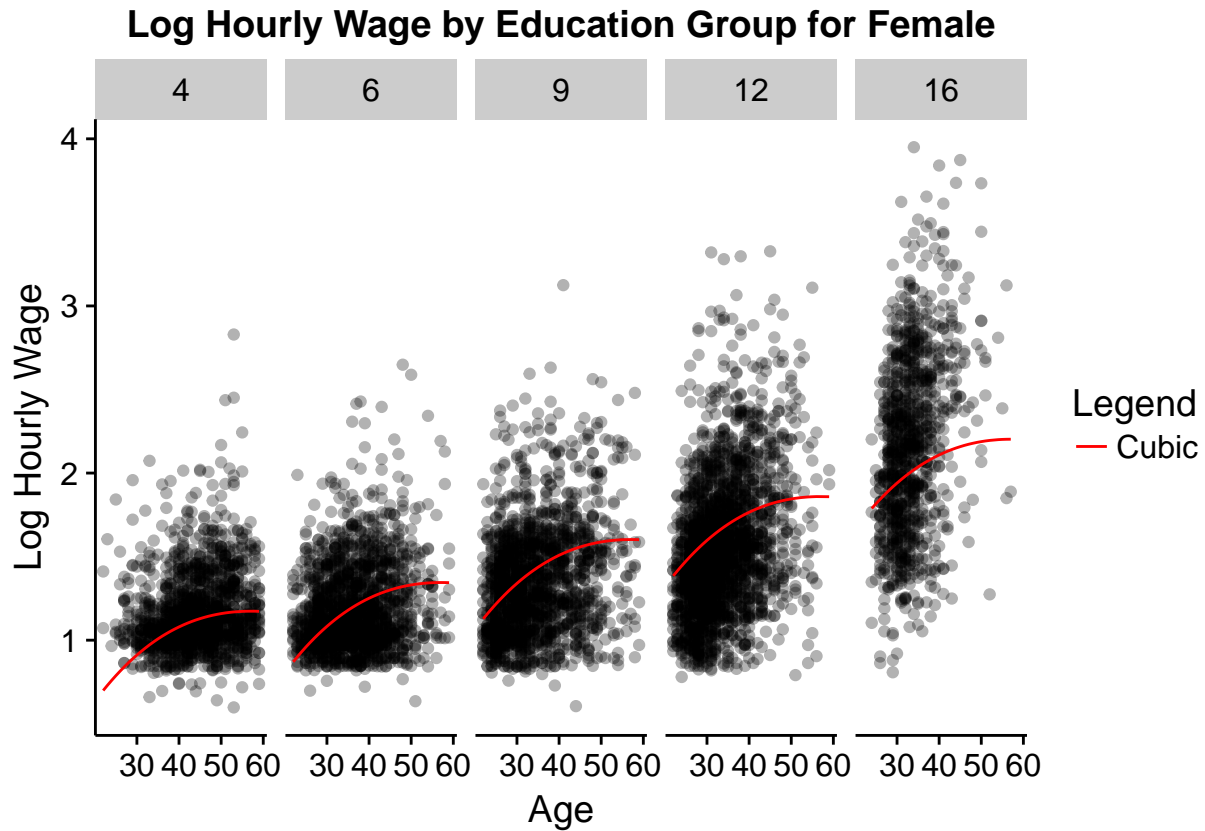
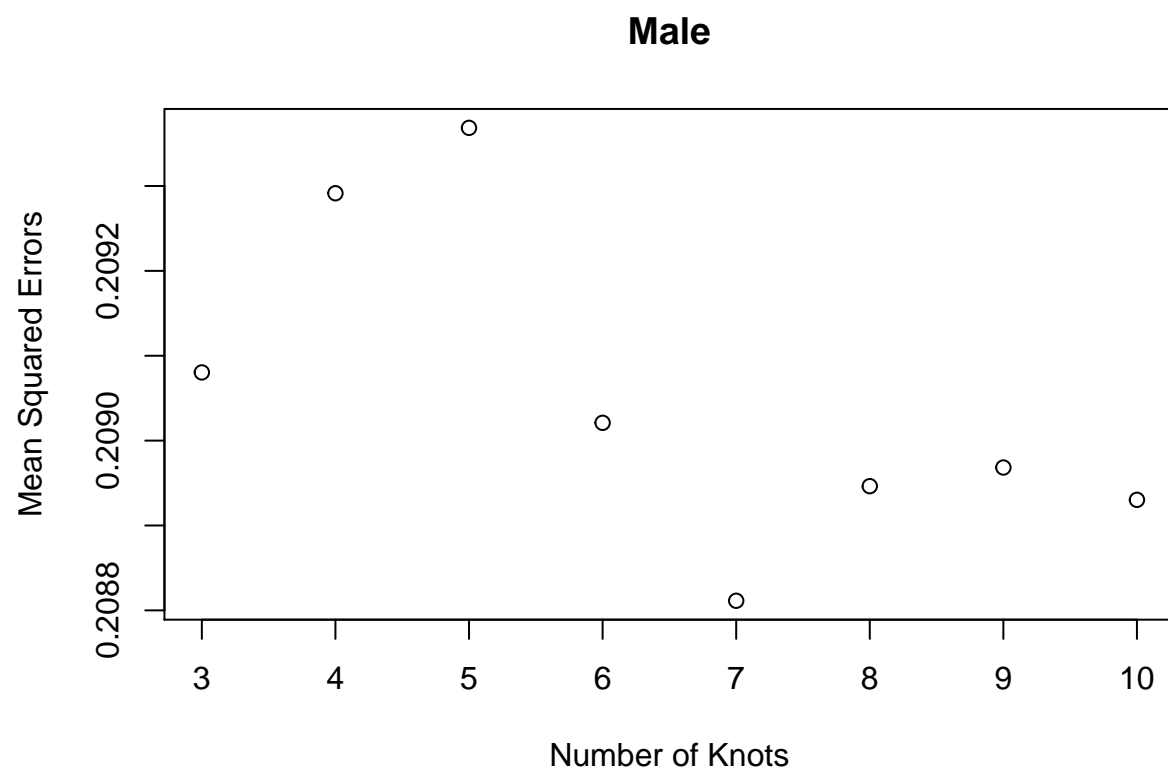


Figure 2 shows the relationship between wages and age for men and women, separated by their education level. There is a general trend that those with higher education level earn higher wages. For those with 4 years of education, there seems to be no relationship between wages and age. The hypothesized model (simple OLS with education and cubic age, column 3 of Table 2) does a poor job in estimating the relationship. The model states that age has an increasing marginal return on wages, but the marginal returns diminishes as people get older. As education level increases, the better fit the hypothesized model becomes.

The hypothesized model is not the ideal model to use because a cubic function is not always the best way to describe the age profiles for men and women. A better model to use is one that includes a cubic polynomial with a cubic spline function. The problem with this model is finding the best number of knots to use.

Figure 3: Finding the Best Number of Knots for Cubic Spline Model





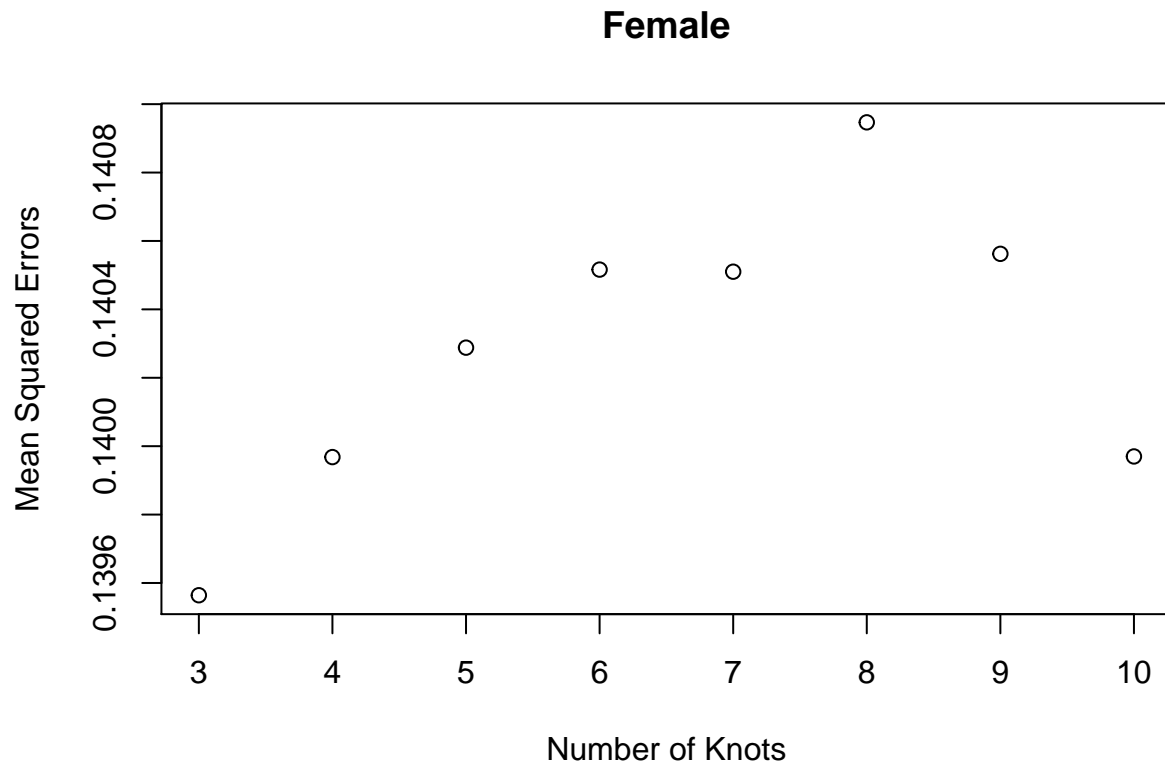
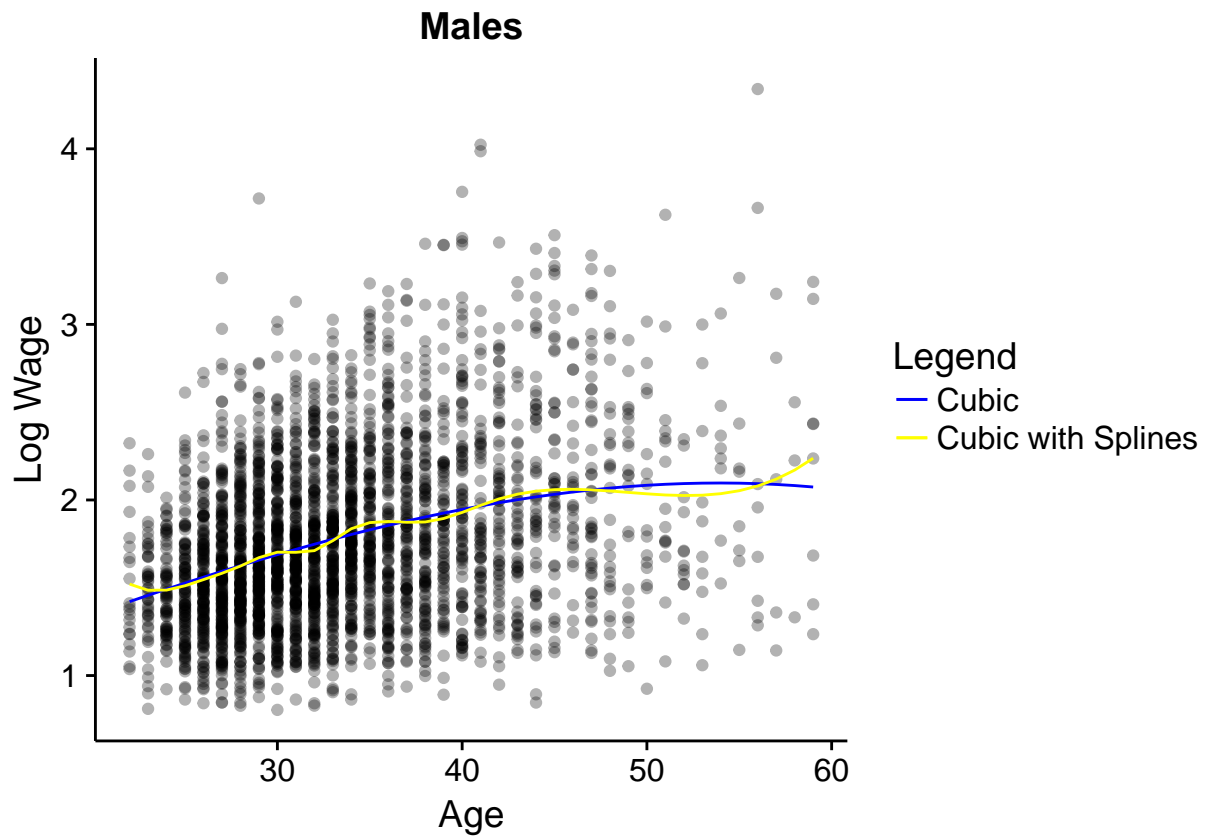


Figure 3 shows the best number of knots to use for men and women with exactly 12 years of education. For men, the model should include 7 knots since it is the one that minimizes the mean squared errors. For women, a model with 3 knots minimizes the mean squared errors.

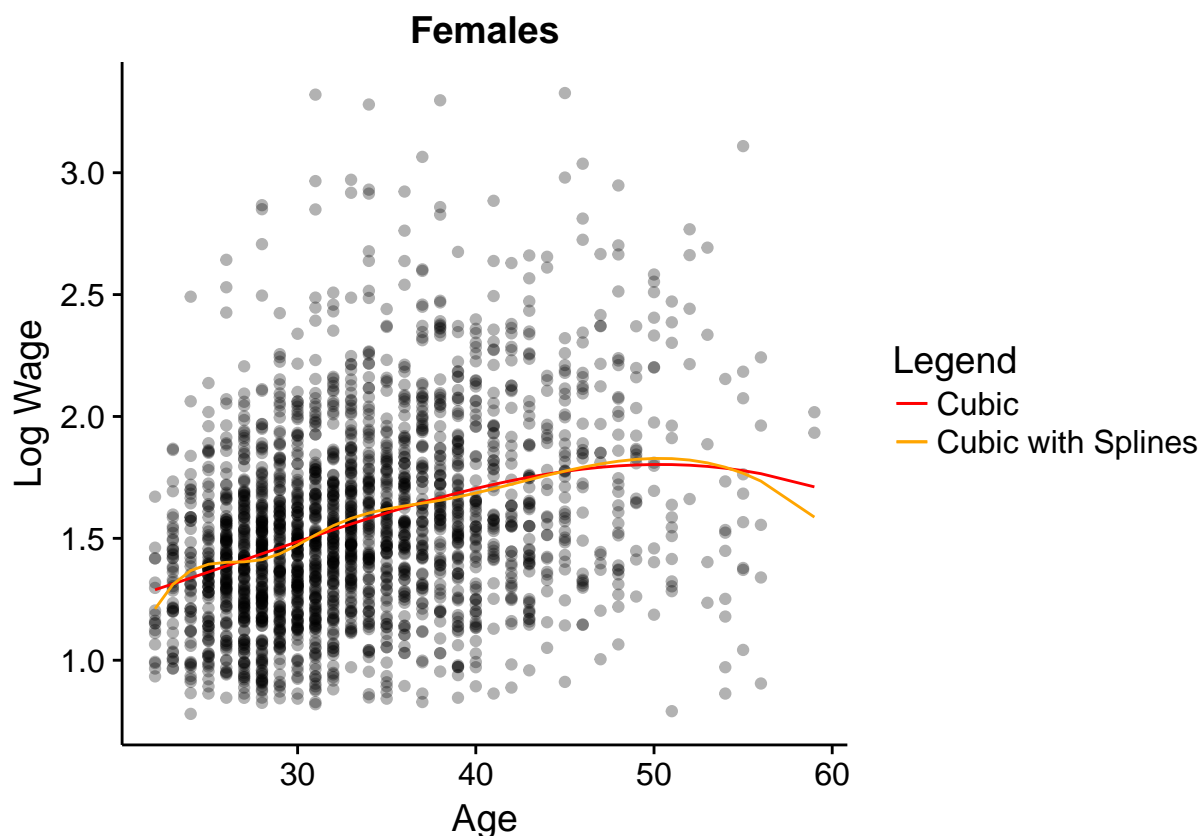
The cubic spline has smaller mean squared errors for both men and women; therefore, it is a better model for relating age to wages. Using data for only men, the model with only the cubic function gives a mean squared error of 0.213 while the model with the cubic spline and 7 knots, 0.212. Using data for only women, the model with only the cubic function gives a mean squared error of 0.1383 while the model with the cubic spline and 3 knots, 0.1379.

Figure 4: Log Wages vs. Age Plot



As illustrated, the data for men that have an education level of 12 years has many peaks and troughs. The cubic model, unlike the cubic with spline model, is not able to capture them. The 7 knots in the cubic with splines model follows the peaks and troughs of the data, hence it shows the relationship between age and wages better.

Figure 4b: Log Wages vs. Age Plot for Females



The data for women with education level of 12 years also have peaks and troughs, but it does not have as many as the data for men. Again, the cubic model is unable to capture them. The cubic spline model with 3 knots follows peaks and troughs of data, which is why the cubic spline with 3 knots is a better model.

Employer productivity is highly related to wages men and women earn. As mentioned earlier, women tend to work at less productive firms. Therefore, how much of the gender wage gap is “explained” by this fact?

Table 3: Series of Standard Wage Model (Including Firm Productivity)

##	=====			
##	Dependent variable:			
##	-----			
##	y			
##	(1)	(2)	(3)	(4)
##	-----			
## female	-0.251***	-0.197***		
##	(0.005)	(0.005)		
##				
## educ	0.083***	0.071***	0.071***	0.072***
##	(0.001)	(0.001)	(0.001)	(0.001)
##				
## age	0.095***	0.093***	0.099***	0.076***
##	(0.015)	(0.014)	(0.019)	(0.023)

```

##
## age2          -0.002*** -0.002*** -0.002*** -0.001**
##              (0.0004)  (0.0004)  (0.0005)  (0.001)
##
## age3          0.00001** 0.00001*** 0.00001** 0.00001
##              (0.00000) (0.00000) (0.00000) (0.00000)
##
## va              0.233*** 0.254*** 0.198***
##              (0.004)  (0.006)  (0.007)
##
## Constant      -0.859*** -1.427*** -1.616*** -1.254***
##              (0.189)  (0.179)  (0.230)  (0.281)
##
## -----
## Observations  23,144    23,144    15,058    8,086
## R2             0.400    0.464    0.421    0.495
## Adjusted R2   0.400    0.463    0.421    0.495
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01

```

According to Table 3, the effect of employer productivity on the wage gap is -0.45 (product of coefficient of employer productivity and the difference in means for women and men). This means that the difference in employer productivity explains 23% of the gap.

The estimated effect of employer productivity on wages is that a 1% increase in change of output increases the change in wages by 23%. However, in this model, the part of wages not explained by the observed characteristics (female dummy variable, age, and employer productivity) consists of two parts:

- (1) unobserved factors that makes the worker more productive that may be correlated with the observed characteristics
- (2) pure noise and uncorrelated with the observed characteristics

Hence, the model suffers from omitted variable bias, meaning the estimated effect of employer productivity is biased. People with high productivity are more likely to work for more productive employers. The more productive worker these firms hire, the higher the productive the firms become. Since the individual productivity is not included in the model, it overestimates the true effect of employer productivity on wages.

Another interesting observation is that working at a more productive firm has a smaller effect for women (0.198) than men (0.254). Therefore, there is a component of the gender wage gap due to the difference in returns to employer productivity.

Extending on the previous Oaxaca style decomposition equation used before to distinguish the part of the gender gap explained by differences in the education and age and the part explained by the differences in the returns of those characteristics, the following equation includes employer productivity.

```

##              Female          Male          diffA          fcoeff
## education    9.151496      8.494422      0.6570748  7.182704e-02
## age          36.315113     36.452849     -0.1377364  7.628905e-02
## age2         1387.453624    1399.605725    -12.1521010 -1.220050e-03
## age3         55695.500618    56518.374485   -822.8738670 6.728032e-06
## va           2.944748       3.139626       -0.1948772  1.980385e-01
##              dotA          mcoeff          diffB          dotB
## education    0.047195739    7.099986e-02    8.271754e-04  0.007026377
## age          -0.010507782    9.932189e-02    -2.303284e-02 -0.839612702
## age2         0.014826171    -1.631677e-03    4.116269e-04  0.576115309
## age3         -0.005536321    9.094830e-06    -2.366799e-06 -0.133767608
## va          -0.038593199     2.542883e-01    -5.624981e-02 -0.176603364

```

The contribution of the difference in characteristics on the gender wage gap decreased to about 0.007 out of the total -0.197. Additionally, contribution of the difference in just education changed to 0.047 out of the total -0.197 while the difference in age is now -0.00129 out of the total -0.197. The supposed contribution of the difference in returns of the covariates is -0.567, which is still greater than the gender wage gap. This means that the difference in returns of the covariates still “over-explains” the wage gap by a lot. The difference in returns to education is 0.007, which again is opposite of the sign of the wage gap. On the other hand, the part of the gender wage gap explained by differences in returns to employer productivity between men and women is -0.177 out of -0.197. This Oaxaca decomposition also supports the hypothesis that differences in education and gender inequality in returns to education does not contribute to the gender wage gap. Both age and employer productivity, however, are possible factors for the wage gap.

The numbers associated with employer productivity are way too big. They’re overestimations because individual productivity is not included in the model and the estimated effect of employer productivity is much bigger than the true effect. To get a better estimation, employer productivity must be renormalized. Renormalizing the employer productivity to reflect this changes the portion of the wage gap explained by the difference in the returns.

##	Female	Male	diffA	fcoeff
## education	9.151496	8.494422	0.6570748	7.182704e-02
## age	36.315113	36.452849	-0.1377364	7.628905e-02
## age2	1387.453624	1399.605725	-12.1521010	-1.220050e-03
## age3	55695.500618	56518.374485	-822.8738670	6.728032e-06
## va	1.241117	1.435995	-0.1948772	1.980385e-01
##	dotA	mcoeff	diffB	dotB
## education	0.047195739	7.099986e-02	8.271754e-04	0.007026377
## age	-0.010507782	9.932189e-02	-2.303284e-02	-0.839612702
## age2	0.014826171	-1.631677e-03	4.116269e-04	0.576115309
## age3	-0.005536321	9.094830e-06	-2.366799e-06	-0.133767608
## va	-0.038593199	2.542883e-01	-5.624981e-02	-0.080774435

The part of the gender wage gap explained by differences in returns to employer productivity between men and women decreases to -0.081 out of -0.197, which is a much reasonable number.

**Table 4: Series of Standard Change in Wage Model (Considering Job Changes)**

##	=====			
##	Dependent variable:			
##	-----			
##	dy			
##	(1)	(2)	(3)	(4)
##	-----			
## age	-0.010***	-0.010***	-0.010***	-0.010***
##	(0.002)	(0.002)	(0.003)	(0.002)
##				
## age2	0.0001***	0.0001***	0.0001***	0.0001***
##	(0.00002)	(0.00002)	(0.00003)	(0.00003)
##				
## female		-0.013***		
##		(0.004)		
##				
## dva	0.076***	0.076***	0.062***	0.083***
##	(0.003)	(0.003)	(0.004)	(0.004)
##				
## Constant	0.258***	0.262***	0.227***	0.276***

```

##                (0.034)   (0.034)   (0.050)   (0.044)
##
## -----
## Observations   23,144     23,144     8,086     15,058
## R2              0.035     0.035     0.031     0.037
## Adjusted R2    0.035     0.035     0.030     0.037
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01

```

To find the true coefficient of employer productivity, the model must take individual productivity into account. Therefore, you need to fix individual specific characteristics. Using change in employer productivity instead of current employer productivity and change in wages subtracts out the error that is constant for individuals throughout time, which includes individual productivity. The true effect of employer productivity on wages is 0.076. This is much less than the estimated effect stated in Table 3 (0.233). As found in Table 3, Table 5 columns 3 and 4 also show that the returns on working at a productive firm is less for women compared to men. This gap may be due to the fact that men have higher positions in firms or that women work less hours because they also have to juggle being a caregiver. The gender inequality in this may be one of the contributions to the gender wage gap.

**Table 5:**

```

##   Quartile Male (va_p) Female (va_p) Male (va) Female (va)
## 1         1  0.2156993    0.3138758 0.1979679  0.3468959
## 2         2  0.2633152    0.2252041 0.2759331  0.2908731
## 3         3  0.2510958    0.2480831 0.2500996  0.1606480
## 4         4  0.2698898    0.2128370 0.2759995  0.2015830

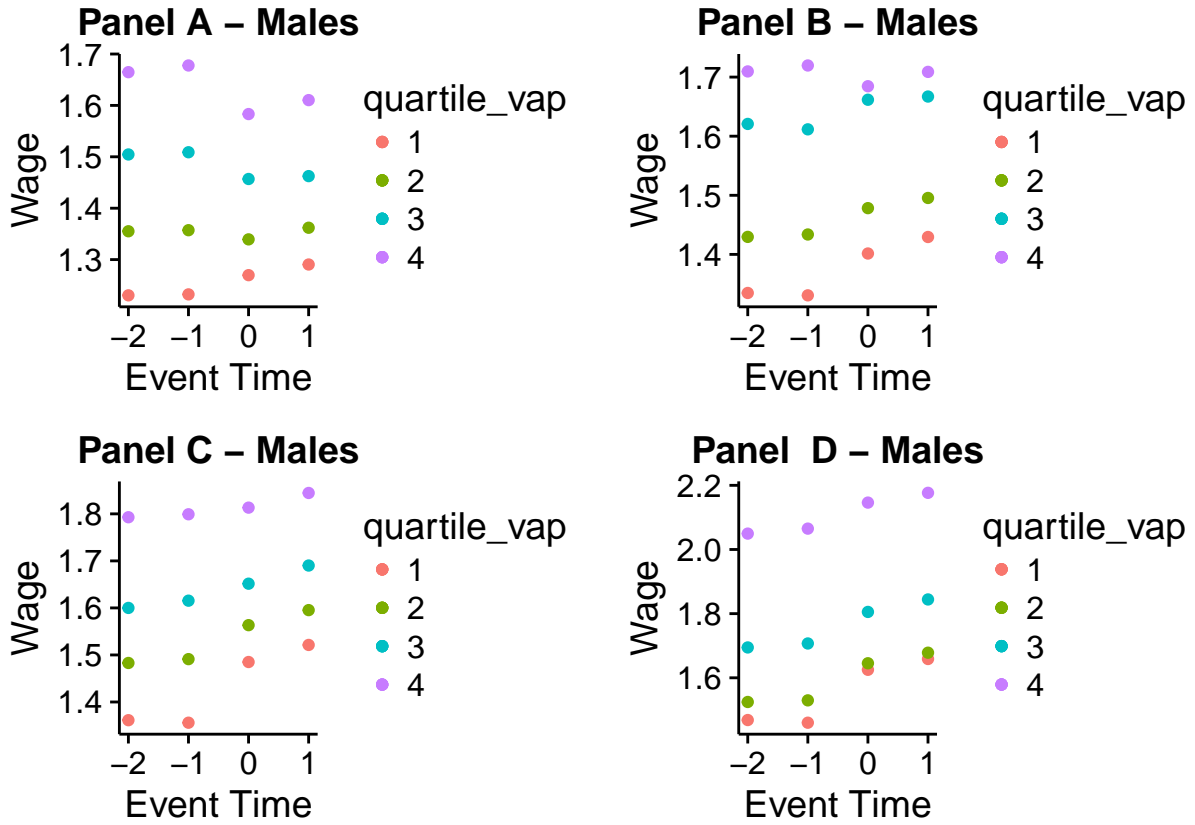
```

To establish whether or not working at a productive firm causes a change in wages, an event study must be conducted. In Table 5, workers are classified by whether their initial employer (va\_p) and current employer (va) are in quartile 1,2,3, or 4.

For men, there is a trend that illustrates that they move to better, more productive firms. The percentage of men in quartile 1 decreased between the previous period and the current period. The percentage of men in quartile 2,3,4 more or less increased in the current period.

Women, on the other hand, have an opposite trend. The percentage of women employed at firms in quartile 1 increased between the previous period and current period. There are much more women working in firms in quartile 2 during the current period; however, the percentage of women in firms categorized in quartile 3 and 4 plummeted. Women, in general, are transferring to less productive firms.

Figure 5



If men are moving to more productive firms and women are moving to less productive firms, how does this reflected in the average wages?

Figure 5 comprises of 4 graphs. Panel A plots the means of wages for male workers who started in different quartiles and end up in quartile 1. Panel B plots it for those who started in different quartiles and end up in quartile 2. Panel C does this for those who end up in quartile 3, and Panel D does this for those who end up in quartile 4.

According to Figure 5, people who move to a firm with the same or higher productivity level receive a raise in their wages. Those who move to a firm with a lower productivity level receive lower wages (males originally in quartile 2,3,4 and end up in quartile 1). If employer productivity has a positive causal effect on wages, this is exactly what is expected. Male workers who start at a firm in quartile 1 and end up at a firm in quartile 4 earns about the same amount as those who start off at a firm in quartile 4 and end up at a firm in quartile 1 because those in the first scenario get a raise when they move firms while those in the second scenario receive less wages when they moved.

Figure 6

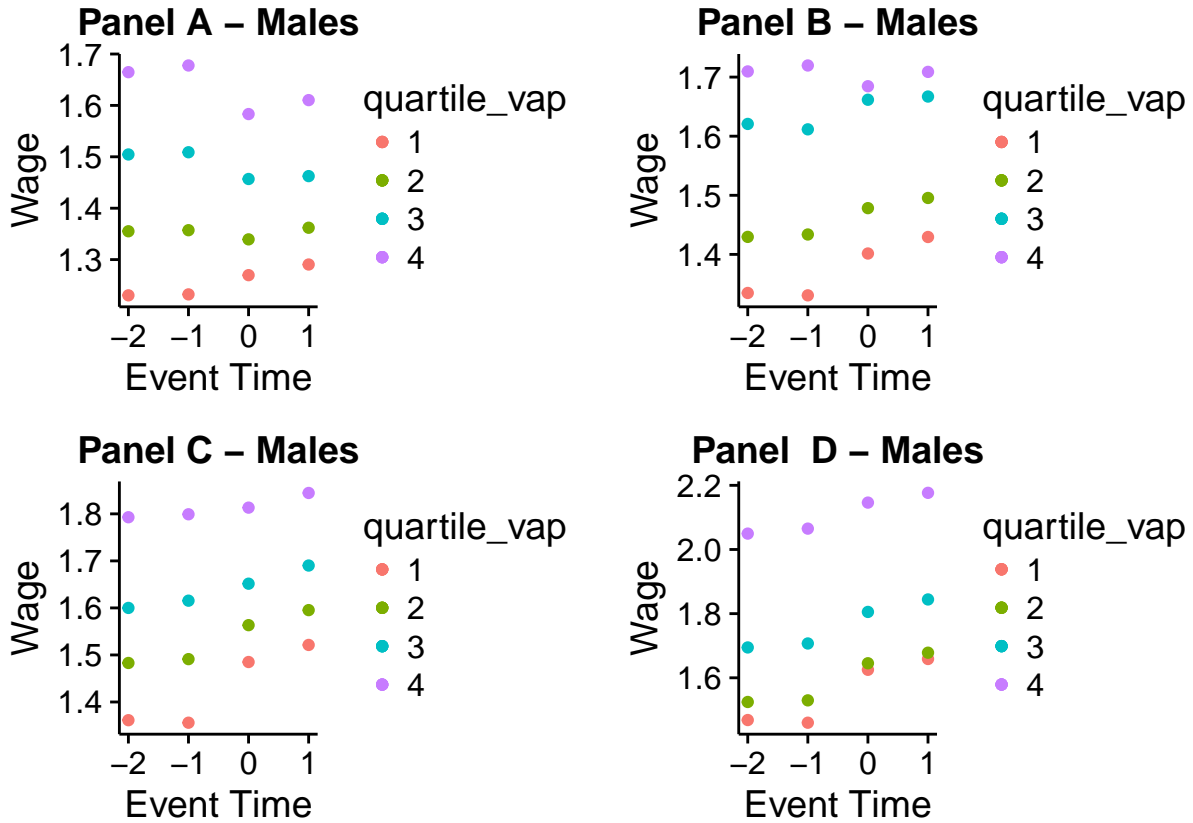


Figure 6 is comprised of the same graphs as Figure 5 but for women instead.

According to Figure 6, females who move to a firm with the same or higher productivity level also receive a raise in their wages. Those who move to a firm with a lower productivity level receive lower wages. These are the same findings as the previous figure for men. Figure 5 and Figure 6 are basically the same. Women just generally have lower reference points for wages.

There seems to be a positive relationship between employer productivity and wages. As workers, both male and female, move to firms that are more productive, they receive a raise in their wages. However, this positive trend happens to wages in general. Those who remain at firms with the same production level also experience a raise in wages. When people transfer to firms with a lower level of production, their wages decrease for that period, but they experience a raise in wages again the following period. This means that there are other things causing an increase in wages other than moving to a firm with higher productivity level.

##	Female	Male	diffA	fcoeff	dotA
## 1	9.151496	8.494422	0.6570748	8.041239e-02	0.052836956
## 2	36.315113	36.452849	-0.1377364	6.876983e-02	-0.009472112
## 3	1387.453624	1399.605725	-12.1521010	-9.973669e-04	0.012120104
## 4	55695.500618	56518.374485	-822.8738670	4.720580e-06	-0.003884442
## 5	2.944748	3.139626	-0.1948772	6.209729e-02	-0.012101349
##	mcoeff	diffB	dotB		
## 1	7.825144e-02	2.160944e-03	0.01835597		
## 2	1.068425e-01	-3.807268e-02	-1.38785768		
## 3	-1.788321e-03	7.909538e-04	1.10702341		
## 4	1.009721e-05	-5.376634e-06	-0.30387860		
## 5	8.311883e-02	-2.102153e-02	-0.06599974		



Building off the idea that the true coefficient for employer productivity is in the regressions used in Table 4, the true coefficients for the other covariates can be found by subtracting the part of the wages attributed to employer productivity from the wages. These new true coefficients changes the Oaxaca decomposition of the wage gap once again.

The contribution of the differences in characteristics on the wage gap is now 0.039 out of the -0.197. This means that women in general have higher levels of the characteristics. Since these characteristics would typically be attributed to higher wages, the differences in the characteristics cannot be the reason for the wage gap. However, the part of the wage gap due to the differences in the mean value of employer productivity is -0.0162 out of the -0.197. Therefore, the fact that men tend to work at more productive firms than women explains 8% of the wage gap. The contribution of the differences in the returns of the characteristics, on the other hand, is -0.632 out of the -0.197. The number is greater than the wage gap, meaning that the differences in returns still over explains the wage gap but still contributes to it. The contribution of the differences in returns of working at productive firms on the wage gap is -0.066, meaning that the inequality in returns explains 33.5% of the wage gap. However, the majority of the wage gap is explained by the difference in returns of additional year of age for men and women (-0.585/0.197).

In conclusion, the wage gap can be explained by: (1) differences in types of firms men and women work in (8% of wage gap) (2) differences in age between men and women (0.6% of wage gap) (3) differences in returns to working at productive firms (33.5% of wage gap) (4) differences in the returns to age (270% of wage gap)

There are a few reasons that men and women have different returns for an additional year in age. First one, as mentioned earlier in the report, men, on average, have finish school earlier than women, meaning they can enter the workforce before women. More work experience translates to more promotions and higher pay; hence, the wage gap appears. Another reason may be outright discrimination against older women. Older men may be seen as wiser and taken more seriously while older women appear to be weaker and less efficient workers. These judgements may be reflected in their pay; as a result, there's a wage gap. Another possibility is that as women grow older, the more time they take off of work to become the caregiver. Since they're working less, their change in wages as they grow older is less than the wages of men. The gender wage gap then becomes an issue.