Jadyn Bell

Econ 494

Executive Summary

**Data Collection Process**

According to a 2016 Glassdoor survey, more than two-thirds (67 percent) of U.S. employees say they would not apply for jobs at employers where they believe a gender pay gap exists.[[1]](#footnote-0) Today, the gender pay gap is more than a social or legal issue. It’s an issue that can affect the ability of employers to attract and retain talent. My data was created by [Glassdoor](https://www.kaggle.com/nilimajauhari/glassdoor-analyze-gender-pay-gap), and I chose this dataset in hopes of assessing whether there are significant differences between gender and wage in jobs of the same title.

**Variables of Interest**

The variables included in my dataset were as follows:

* Job Title
* Gender
* Age
* Education
* Performance Evaluation
* Department
* Seniority
* Base Pay
* Bonus Pay

The job title labeled the jobs of every individual surveyed and ranged from lower level jobs such as Warehouse Associate, to higher level jobs such as Software Engineer. Gender broke down whether the individual was male or female as that is the main variable I will be analyzing. The age column broke down the different ages ranging from 18 being the youngest age to 65 being the oldest age. Education is the highest level of education that an individual received before entering the workforce, ranging from High School diploma to a PhD. Performance evaluation is the last score the individuals received rated on a scale of 1-5. Department breaks down each department the job is categorized as such as, management, sales, operations, etc. Seniority, according to Glassdoor, is the number of years worked ranging on a scale of 1-5. Base Pay is the annual basic wage in dollars that each individual receives and bonus pay is the annual bonus pay each individual receives, also in dollars.

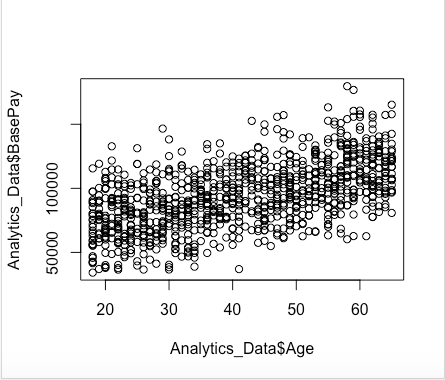
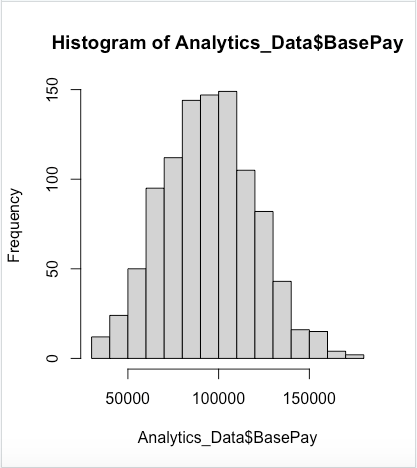
**Overall Structure of Dataset**

The dataset is set up in 1,000 observations of 10 variables, meaning there are 10 columns and 1,000 rows. There is both categorical and numerical data involved. There is no missing or questionable data at first glance.

**Analysis**

For this analysis, I chose to investigate the gender wage gap in the workforce according to Glassdoor. From this data, I am trying to understand the validity behind the difference in earnings between women and men. As there have been many studies showcasing that women are paid less than men for the same job titles, I believe this data set will be helpful in identifying the depth of the gender-based pay gap.

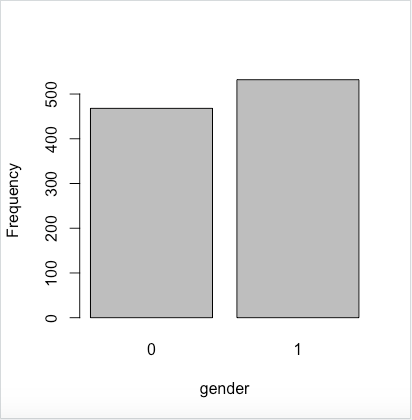
To begin, I first did an initial exploratory analysis. I looked at a summary of the data and immediately noticed a large difference between the minimum base pay value of $34,208 and the maximum base pay value of $179,726. Seeing as there are many factors potentially involved in creating this gap this difference doesn’t really tell us much. To further my initial analysis, I decided to plot the data to see if I could pick up on any trends.



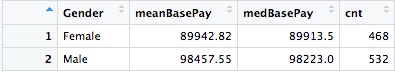
Looking at the histogram I noticed that the majority of base pay is around $100,000 a year and the scatter plot showed a slightly positive relationship between age and base pay, not taking into consideration gender. Meaning, the older the individual is, the higher their base pay seems to be. This relationship makes sense if we consider our own assumptions such as experience leading to higher pay, and over time being able to work your way up in your job.

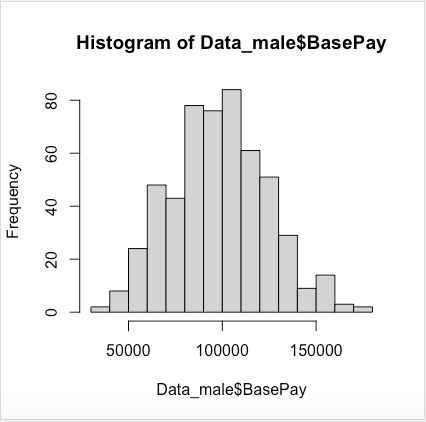
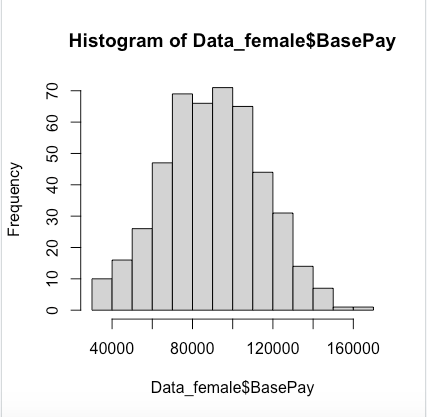
Continuing with my exploratory analysis, I took notice that there wasn’t much cleaning of the data that needed to happen. There were no missing or repeated values. I deleted the “Department” column as I didn’t find it useful in my analysis of the gender based pay gap. I also deleted the “Bonus Pay” column as I believed it would be hard to control for the other extenuating factors that could be affecting how much annual bonus each individual receives.

Once I cleaned up my data, I wanted to look further into the difference between gender and base pay specifically.

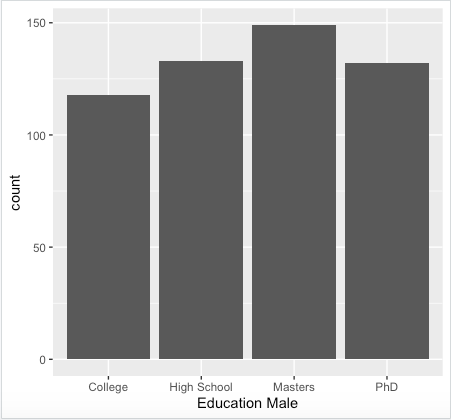
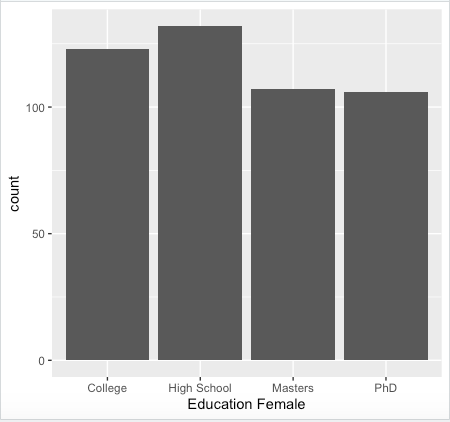


In order to plot the data, I changed the gender variable into a numeric variable so that female = 0 and male = 1. Keeping this in mind, you immediately see that the males are represented more in this dataset with 532 male responses and 468 female responses. I then created two new datasets, one for the female data and one for the male data in order to compare the differences. The resulting table and histograms are shown below. Men on average are paid $98,458 per year while women on average earn $89,943 per year — an overall or “unadjusted” pay gap of $8,515 or about 8.6 percent of male pay. Using a similar command, I looked at whether men and women are clustered more into certain job titles. After running my code, I noticed that men are overrepresented in management and software engineering roles, which are more highly paying, while women are overrepresented among marketing associate roles. This is a common pattern we see in real-world labor market data. Although this is not a big issue since it is quite possible there are more men than women who are working in these jobs, this does help further our belief that STEM jobs tend to be more male dominated.



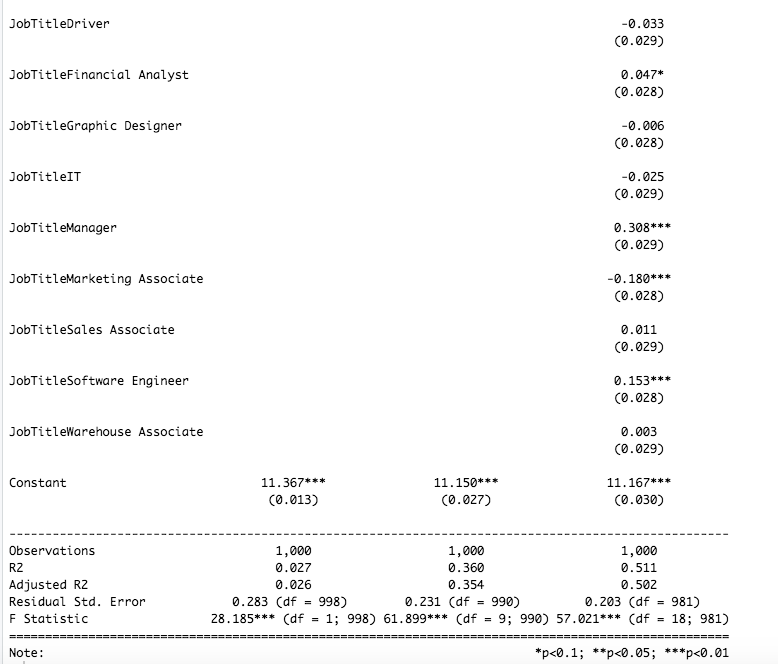
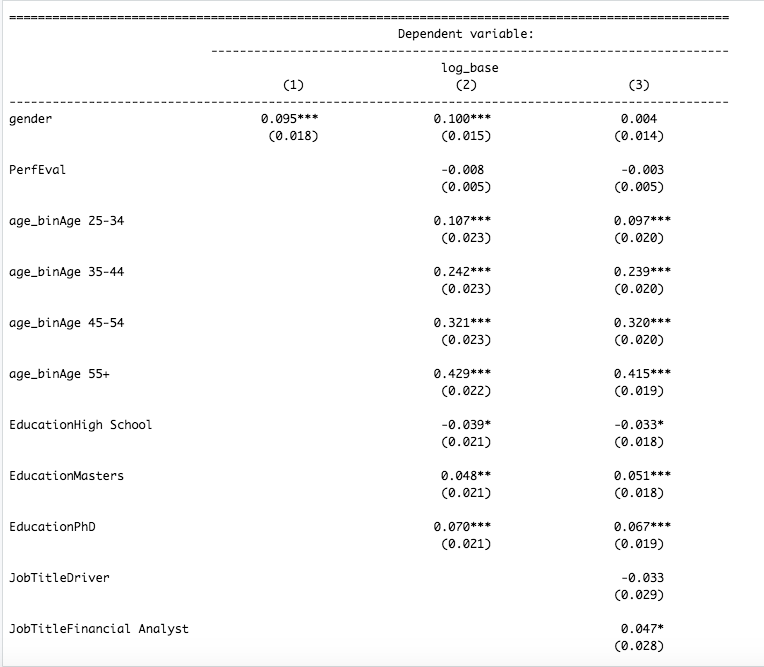


I realize that without taking into consideration the other variables that might be skewing the data, such as education, that alone is not enough to conclude the gender pay gap is prevalent in the modern day workforce. Since we have all been taught to believe that education is an avenue to success, I wanted to look at the education levels of the individuals and how that affected annual pay.



After looking at the difference between levels of education of females and males, I noticed that the females seem to be hired with lower levels of education such as high school and college. Meanwhile, the males seem to be hired with higher levels of education with a masters degree being the highest for males. This helps further my belief that females are hired into lower level jobs such as associate level positions, while the males are hired into higher level positions such as software engineering.

In order to take a closer look at the relationships between each variable, I decided to run regressions.



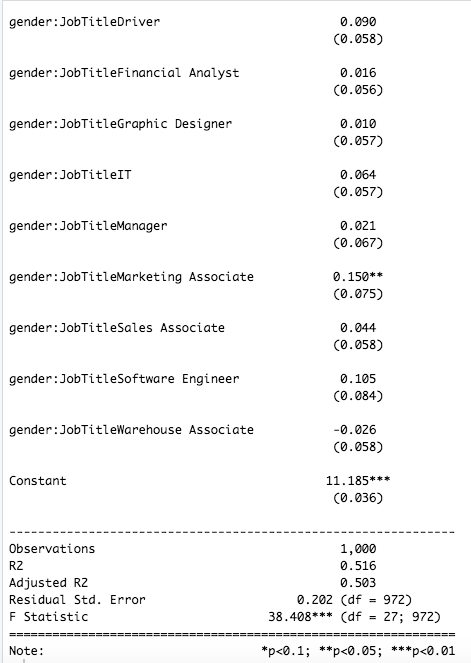
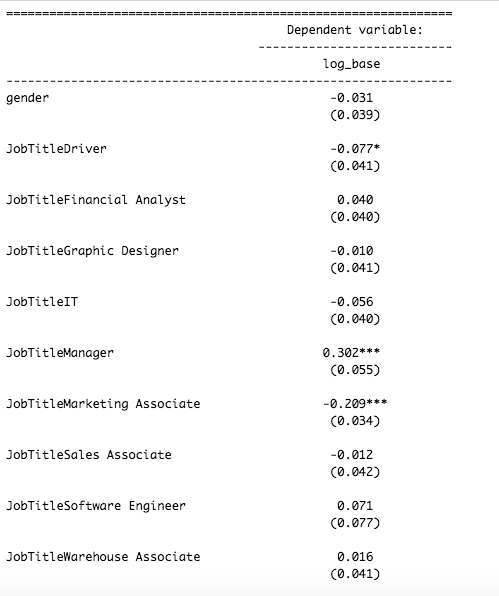
The first row of the table shows our estimates of the “unadjusted” and “adjusted” gender pay gap. That’s the coefficient on our male-female dummy. In Column 1, a coefficient of 0.095 means there is approximately 9.5 percent “unadjusted” gender pay gap in our hypothetical company. Put differently, men on average earn about 9.5 percent more than women on average in this hypothetical company. We also see this estimate is highly statistically significant.

In Column 2, we add individual controls like age, performance evaluations and education. Here we see the gender pay gap hasn’t changed much. It’s still a 10.1 percent pay gap, which is still highly statistically significant. Whatever is causing the pay gap in our hypothetical employer isn’t due to differences in education, age or performance evaluations of men and women.

In Column 3 we add all of the controls we have in our data. This is the “adjusted” gender pay gap. In this case, we see the gender pay gap shrink to 0.4 percent after controlling for job title. More importantly, this estimate is no longer statistically significant — so we can’t conclude it’s really different from zero. In this case, we say there’s no evidence of a systematic gender pay gap on an “adjusted” basis, after controlling for observable differences between male and female workers.

In our hypothetical data, the gender pay gap shrinks to near zero once we control for differences in job titles between men and women. I believe this to be due to our previous conclusion in our analysis about men being over-represented in higher-paying software engineer and manager roles, while they are underrepresented in lower-paying marketing roles. Once we control for the fact that men and women work in different roles in this company, the remaining pay difference that’s due to differences in gender turns out to be very small — in this case, close to zero.

Even if there is no overall pay gap between men and women, it’s still possible that gender pay gaps are hidden only within certain job titles. To test for differences in the gender pay gap among job titles, I made a slight adjustment to the models used above. By looking at whether the coefficients on these interaction terms are statistically significant or not, we can test whether the gender pay gap within certain job titles or departments differs from the overall company average.



After looking at these results we see that there are only a few instances where the coefficients seem to be statistically significant, meaning that the particular job title is different from the overall average of the company. In the cases where the coefficients are not statistically significant, there is no evidence to support the idea of pay gaps by job titles.

**Conclusion**

After statistically comparing workers with similar job titles, with comparable education and experience, we found little to no statistical significance between male and female pay. Before any statistical controls, men earned on average 9 percent more than women. This amounts to women earning on average 81 cents per dollar earned by men. Although those gaps are smaller than appear from a simple comparison of average male and female pay, they are a large and statistically significant difference between male and female earnings.

An often-overlooked point in gender pay gap studies is that just because the gap in male-female wages declines when we statistically control for worker characteristics, doesn’t mean the gap is not real or caused by unfair barriers women face in the workplace. For example, if women are systematically excluded from certain occupations, or encouraged to work only in certain industries, or discouraged from pursuing particular college majors, these factors can statistically “explain” the gender pay gap but still represent social biases against women that most observers would consider unfair and worthy of criticism. One of the most important findings of this study is that differences in education, age and experience—what economists call “human capital” of workers—explains a trivially small part of the gender pay gap between men and women. As women have closed the gap in rates of college education and labor force participation in recent decades, less and less of the pay difference between men and women can be explained by gender differences in skills and education. Instead, the vast majority of the gender gap today is caused by one important factor: sorting of men and women into systematically different occupations and industries throughout the economy. In order to continue to decrease this pay gap I offer up a couple suggestions.

**Improving Free Occupational Choice**

An intelligent way to design policies aimed at closing the gender pay gap is to focus on the biggest factors causing it. Although overt workplace discrimination almost certainly still occurs, intentional employer bias can only account for a small fraction of the overall gender pay gap. Though policies that address workplace discrimination are important, focusing on policies that target the dramatic occupation and industry sorting of men and women into separate and financially unequal types of work throughout their careers can help alleviate the gender pay gap by addressing its most important causes. For example, research has shown that women disproportionately tend to be primary caregivers for children and the elderly in families, and that this can lead women to sort into lower-paying occupations offering more flexibility. This is an example of occupational choice that can be statistically controlled for, but is heavily influenced by social norms largely beyond the control of individual female workers. Similarly, occupation and industry sorting is heavily affected by institutions that lead men and women down different paths through the education system, pressures that divert men and women into different college majors and career tracks, and a variety of social norms regarding family responsibilities. In these cases, societal and public policy encouraging more equal access to science, technology and health care training, targeted initiatives to encourage career growth and advancement among female business leadership, and policies that support child care and assistance for the elderly may help address the root causes of these types of occupational and industry differences between male and female workers.

**Expanding Pay Transparency**

A second example of a policy not explicitly aimed at gender pay but which can help narrow the gap over time is greater workplace transparency. For example, a 2012 study by economists Andreas Leibbrandt and John List found that a major contributor to the gender pay gap is a negotiation gap. Women are less likely to negotiate over salary than men. But when researchers explicitly told job seekers that pay was negotiable, the gender gap disappeared. Recognizing this power of transparency to eliminate pay gaps, many legal scholars have called for mandatory pay transparency among U.S. employers—a proposal endorsed by the Obama administration in 2016. Although equity in gender pay has improved dramatically in recent decades, Glassdoor data suggest there remains much to be done.

1. Gender Pay Gap Survey (February 2016), Glassdoor. https://www.glassdoor.com/blog/the-gender-pay-gap-is-it-real-new-survey/ [↑](#footnote-ref-0)