What's Behind the Gender Pay Gap?

Ariel Silbert
Final Project, EMSE 6992
December 3, 2019

Abstract

Women continue to be paid less than their male counterparts. As data from the US Census has shown, this is true regardless of whether pay for men and women generally, men and women in the same field, or even men and women in the same position is considered. This project examines a selection of hypotheses that would explain why women's pay lags behind men's to see if the data supports them.

Introduction

Research has consistently shown that diverse workplaces tend to be more innovative and productive.[1] However, pay disparity between genders could prove to be a barrier to a diverse national workforce. Currently, in the United States, women's pay lags at 78 cents to every dollar that a man makes.[2] There are many explanations given for this disparity, but oftentimes, these explanations are thrown about without any substantial proof. Without understanding the basis of a problem, one cannot apply a solution.

Research Questions

This project tested if some hypotheses for the existence of the pay gap hold up under scrutiny. Specifically, this analysis tested the following reasons given for gendered pay discrepancies:

- The pay for men is not actually that different from the pay for women, it's just that men tend to be in more prestigious positions that pay more. The gender pay gap is actually due to a few outliers and not an overall trend.[3]
- Female-dominated jobs pay less. When there are enough women in an occupation, one can be assured that it is lower paying.[4]
- When an occupation has parity in numbers between women and men, it also has parity in pay. Equality on one front influences equality on the other.[5]
- STEM occupations follow a different trend than other occupations. They tend to have more men and pay more, and this explains why the overall pay for men versus women appears to be different.[6]

The following questions were used to test these hypotheses:

- Does the distribution of pay for men differ from that for women? Does it have a different center or a longer tail on the high-pay end?
- Is the number of women and/or men in an occupation correlated to each gender's pay? In other words, does a large number of women in an occupation indicate that the pay will be more evenly distributed?
- Does a larger difference between the number of men and the number of women in an occupation correlate to a larger difference in pay between the two? In other words, if an occupation shows more gender parity in the number of workers, does it also show more gender parity in pay?
- Does an occupation being a STEM occupation affect whether there is a greater pay or number disparity between the genders?

Information on the Dataset

The majority of this data was gathered from the publicly available US Census Bureau Microdata service. It is the "ACS 5-Year Estimates - Public Use Microdata Sample" for 2017.[7] The values included are as follows:

- Occupation (2010 OCC codes) -- The occupation codes from the BLS's 2010 SOC codes, some occupations abbreviated for space considerations. Designated as "SOCP" on the site.
- Wages (past 12 months) -- The average wages or salary income over the past 12 months for the occupation as a whole. Designated as "WAGP" on the site.
- Wages-Male -- The average wages or salary income for male workers in the occupation over the past 12 months. Designated with "SEX" and WAGP" on the site.
- Wages-Female -- The average wages or salary income for female workers in the occupation over the past 12 months. Designated with "SEX" and WAGP" on the site.
- Number-Male -- The number of male workers in the occupation. Designated with "SEX" on the site.
- Number-Female -- The number of female workers in the occupation. Designated with "SEX" on the site.

In addition, part of the dataset came from Occupational Employment Statistics from the Bureau of Labor Statistics.[8]

- STEM -- The occupations listed in the file are those occupations that the Bureau of Labor Statistics considers to be STEM occupations. This information was represented with a 1 if the occupation was a STEM occupation and a 0 otherwise.

Methodology

The following methods were used to assess the hypotheses:

- Basic feature analysis to assess count, mean, standard deviation, minimum, 25th percentile, median, 75th percentile and maximum values for each part of the dataset.
- A scatterplot matrix (SPLOM) of wages, male wages, female wages, male numbers and female numbers.
- Correlation analysis on the same features plus STEM values.
- A histogram showing pay by gender across occupations.
- A SPLOM of the normalized pay and number differences across occupations.
- A histogram of those normalized differences.
- 3D k-means clustering of wages, male wages, and female wages.
- 2D k-means clustering of male and female wages.
- 3D k-means clustering of wages, male numbers, and female numbers.
- 2D k-means clustering of male and female numbers.

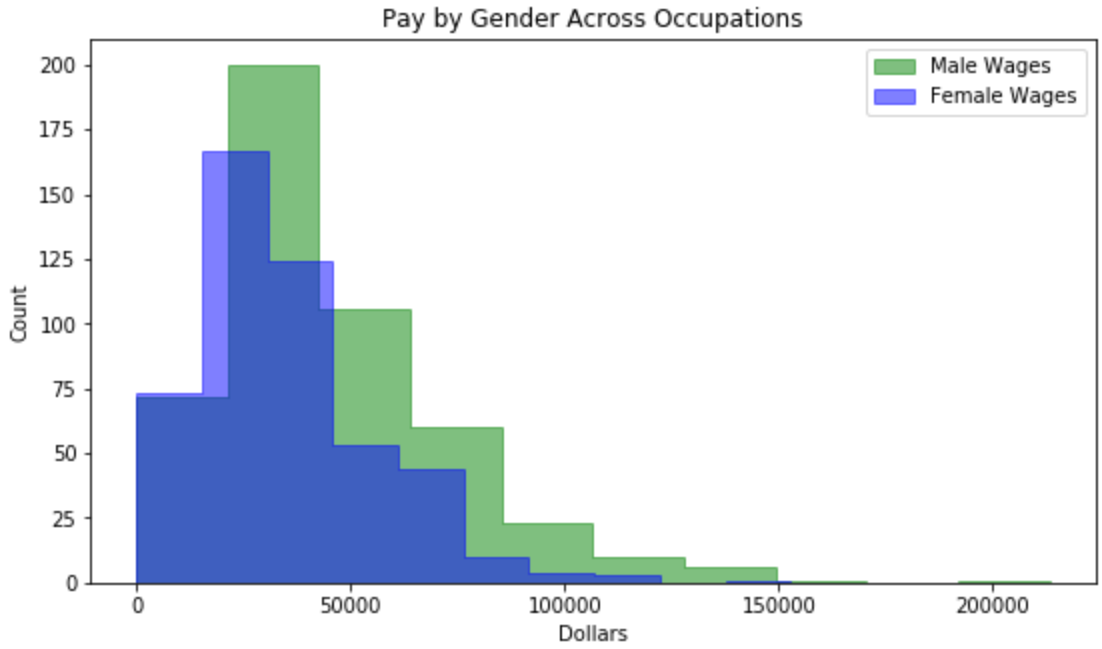
Analysis

First, a basic feature analysis was run on the dataset to ensure that nothing out of the ordinary was happening that would be worth exploring:

	Wages (past 12 months)	Wages- Male	Wages- Female	Number- Male	Number- Female
count	479	479	479	479	479
mean	41625.642456	45910.165293	35072.283098	2.048505e+05	1.911033e+05
std	24863.779334	27332.778927	21034.593307	3.995294e+05	4.612860e+05
min	0.000000	0.000000	0.000000	2.549000e+03	1.900000e+02
25%	23167.550060	26549.268710	19352.601050	2.630950e+04	8.644500e+03
50%	36301.328950	39482.129500	30489.704560	6.275600e+04	4.117300e+04
75%	52013.546290	57960.589875	45255.683625	1.801175e+05	1.497175e+05
max	191740.35520	213222.07340	152777.89180	3.968151e+06	3.987371e+06

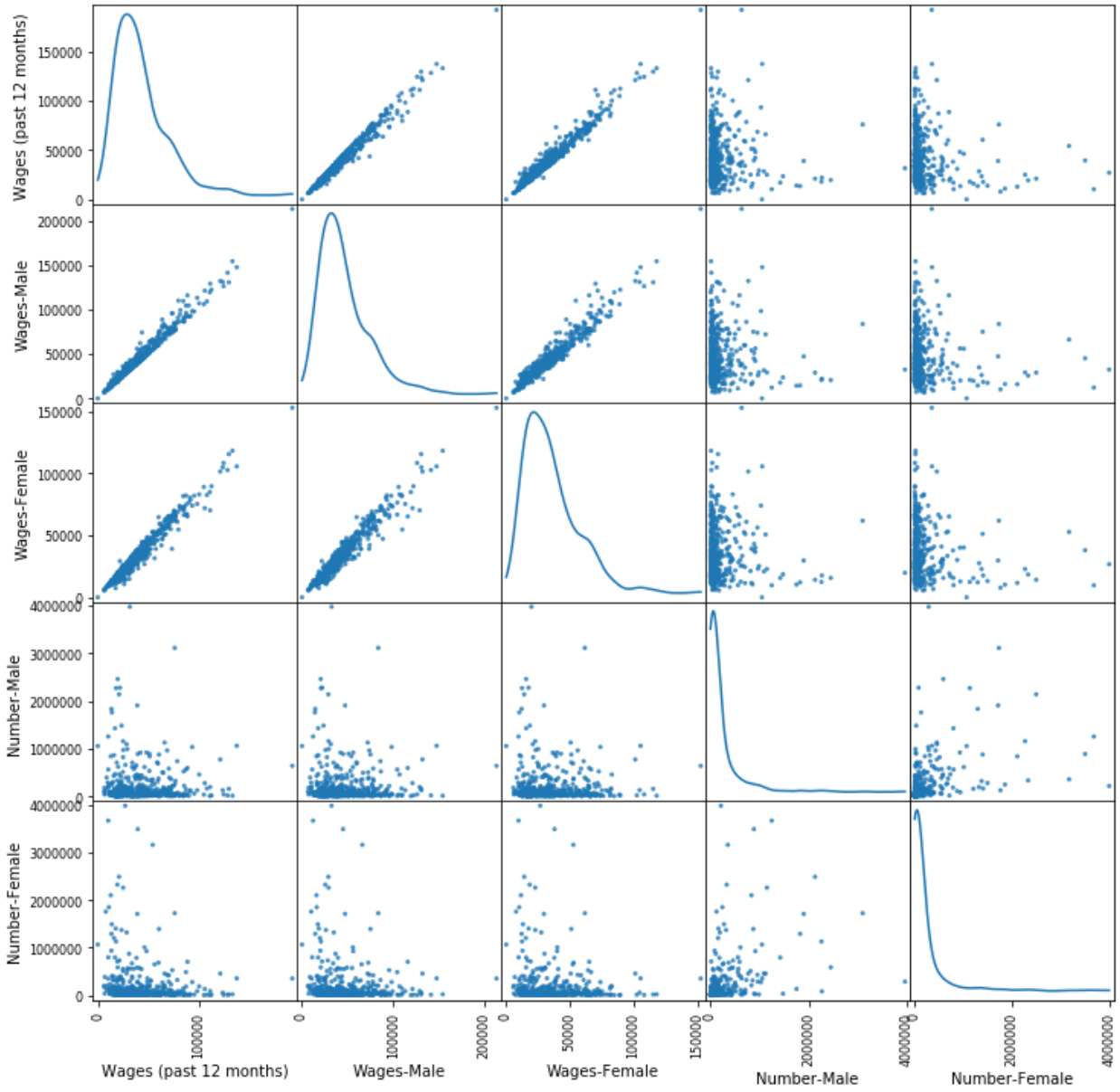
The mean wages for men were higher than for women, as were the median wages. The discrepancy between them was fairly large -- close to $10,000 for each. The standard deviation for men's wages was also larger than for women's, suggesting that men's wages have a larger spread than women's. While the minimum wages for each group was the same (since unemployed people make $0 regardless of gender), the maximum wages for men were higher than for women.

All of this information was in line with this research's baseline assumption: men make more than women on average. However, the new information about the difference in standard deviation and maxima suggests that part of that discrepancy could be due to men's wages having a longer tail on the high-pay side. This appears to be borne out by the following histogram showing that exact relationship:



However, it is also clear from this graph and the information about mean and median wages that the center for men's wages is also higher than for women's, so the longer tail does not fully explain the discrepancy.

Next, a SPLOM of the data was generated:



From this SPLOM, a few things are clear. First and most obvious, each element is highly correlated to itself. Moreover, each of the wages subsets are highly correlated, though it is difficult to determine how highly. This is not surprising because women and men both make up a significant portion of the workforce, and their wages are both used to calculate the total wages. Also, if women's and men's wages were not highly correlated, it would be obvious to even the most casual observer that women and men were being paid differently and that rampant discrimination based on gender exists. The remainder of the elements do not seem that highly correlated. This suggests that the hypothesis that a greater number of women being in an occupation influences women's pay may not hold water.

To verify the visual cues evident in the SPLOM, a correlation analysis on the features was run, including the STEM designation as well.

	Wages (past 12 months)	Wages-Male	Wages-Female	Number-Male	Number-Female	STEM
Wages (past 12 months)	1.000000	0.986729	0.979212	-0.005841	-0.120729	0.431259
Wages-Male	0.986729	1.000000	0.966439	-0.010526	-0.085219	0.388524
Wages-Female	0.979212	0.966439	1.000000	-0.033101	-0.098145	0.439224
Number-Male	-0.005841	-0.010526	-0.033101	1.000000	0.428073	-0.045650
Number-Female	-0.120729	-0.085219	-0.098145	0.428073	1.000000	-0.104114
STEM	0.431259	0.388524	0.439224	-0.045650	-0.104114	1.000000

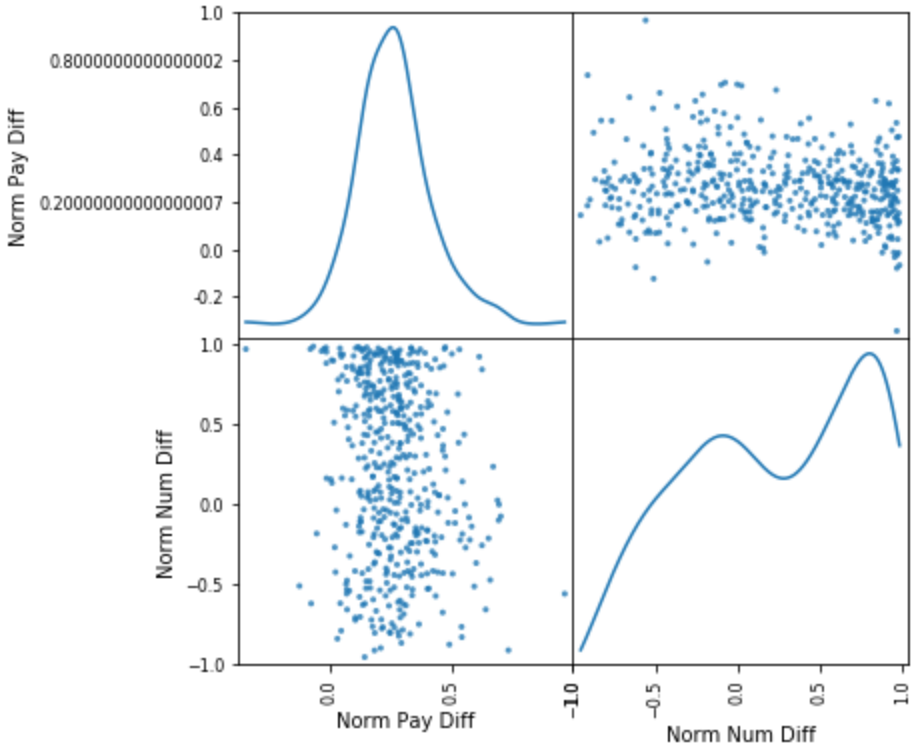
This analysis verifies the high correlation between each of the wages elements that the SPLOM suggested. It also verified that there is very little correlation between the number of male or female workers in an occupation and the pay for either gender. There is a fairly strong positive correlation between a STEM designation and pay, regardless of gender. This is not surprising given the fact that STEM occupations are widely regarded as high paying given the level of expertise and skill they require.[9]

Given this correlation between each subset of pay, it was worth looking at this data in a different way that removes this inherent correlation. While it seems that pure numbers of women in an occupation does not make it more pay equitable, the differential in numbers between the genders may still have influence on the differential in pay. To test this, the data was manipulated to create two new variables:

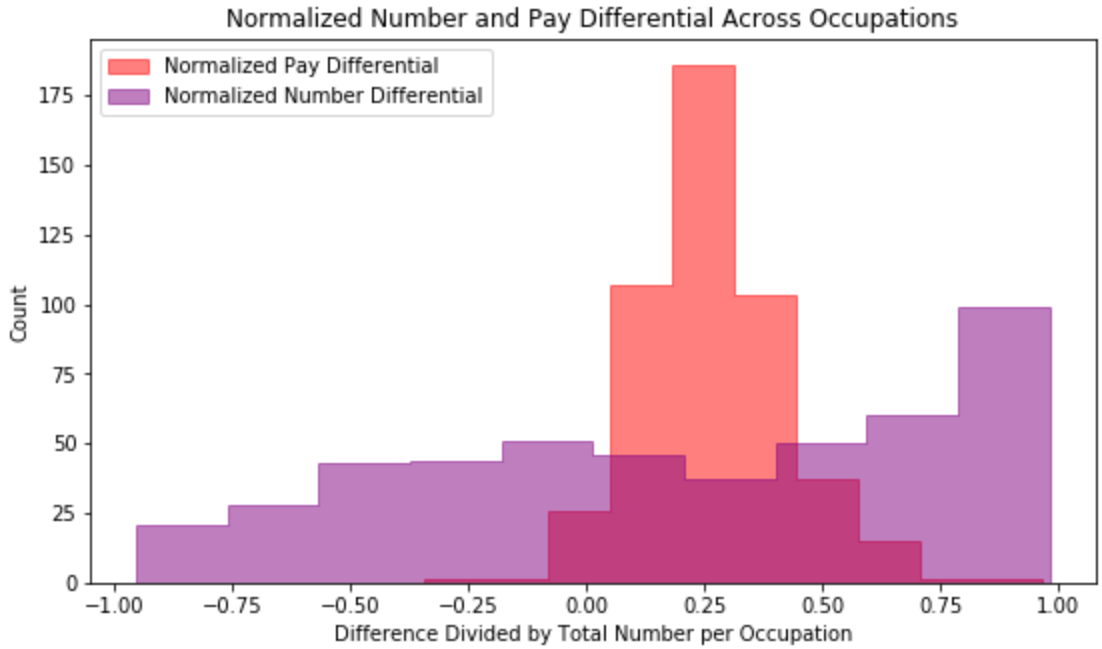
$$Normalized\ Pay\ Differential\ =\ \frac{Male\ Wages - Female\ Wages}{Wages}$$

$$Normalized\ Number\ Differential\ =\ \frac{Male\ Number - Female\ Number}{Male\ Number + Female\ Number}$$

Next, a SPLOM was generated to look for correlation between these new elements:



Neither element seems highly correlated with the other, a suspicion that was verified by their correlation coefficient of -0.156724. Moreover, the following histogram shows that the two elements do not even share a shape. The pay differential is a bell curve, while the number differential is not. The number differential does have a notable lean towards more occupations being populated by men than women -- interesting, but not for the current analysis.

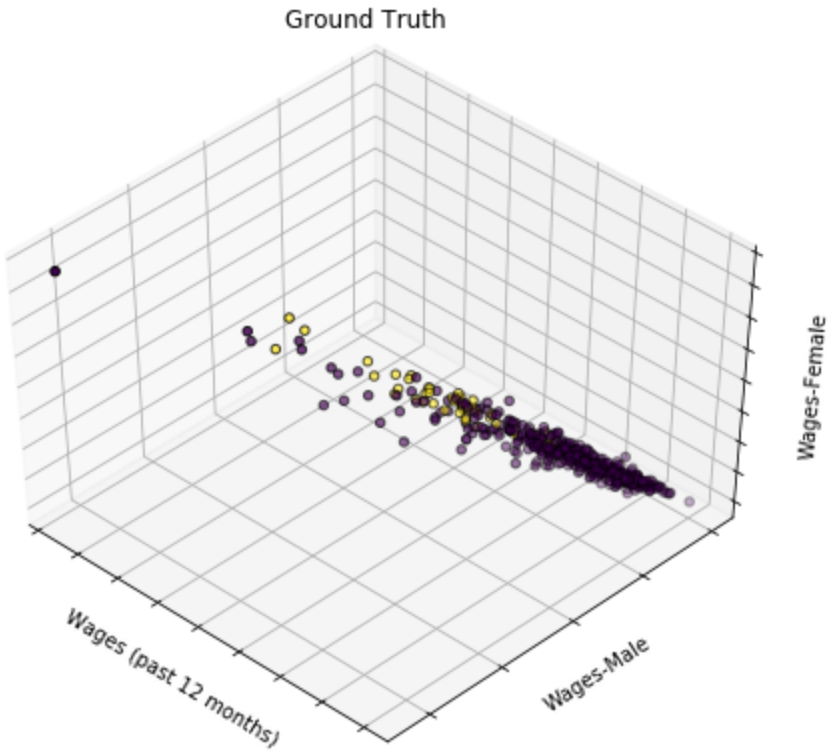


The correlation between these new factors and the previous elements was low:

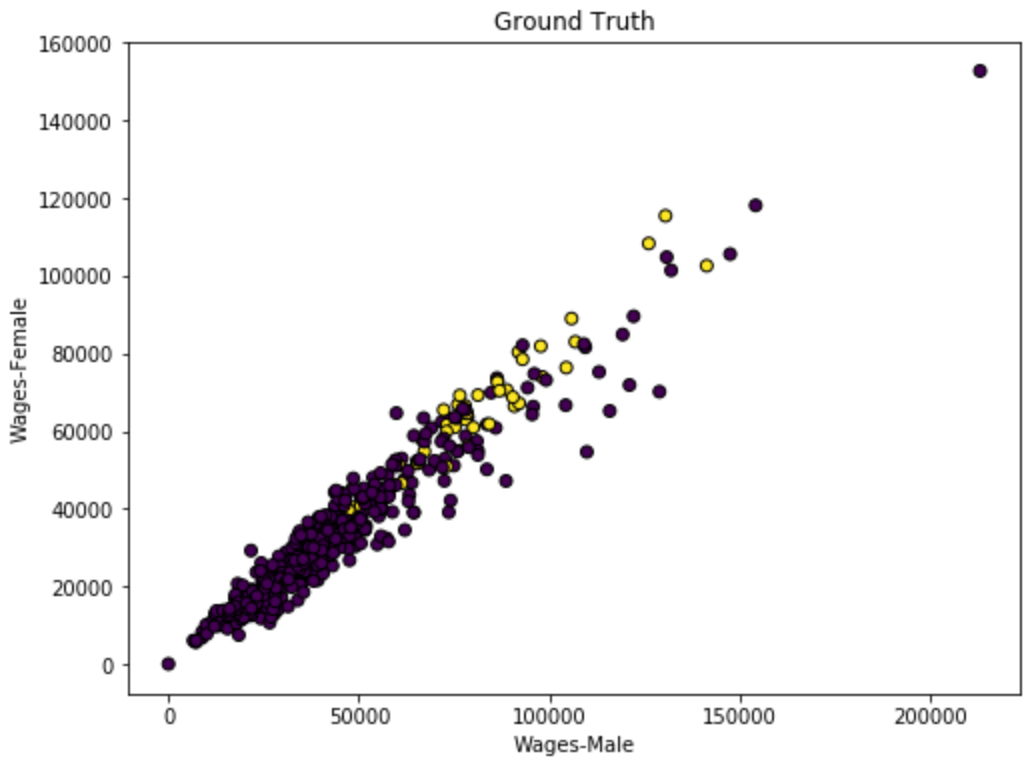
	Wages (past 12 months)	Wages-Male	Wages-Female	Number-Male	Number-Female	STEM	Norm Pay Diff	Norm Num Diff
Wages (past 12 months)	1.000000	0.986729	0.979212	-0.005841	-0.120729	0.431259	-0.048376	0.174642
Wages-Male	0.986729	1.000000	0.966439	-0.010526	-0.085219	0.388524	0.049680	0.067147
Wages-Female	0.979212	0.966439	1.000000	-0.033101	-0.098145	0.439224	-0.169228	0.087462
Number-Male	-0.005841	-0.010526	-0.033101	1.000000	0.428073	-0.045650	0.062070	0.126207
Number-Female	-0.120729	-0.085219	-0.098145	0.428073	1.000000	-0.104114	0.078129	-0.381682
STEM	0.431259	0.388524	0.439224	-0.045650	-0.104114	1.000000	-0.133800	0.166206
Norm Pay Diff	-0.048376	0.049680	-0.169228	0.062070	0.078129	-0.133800	1.000000	-0.156724
Norm Num Diff	0.174642	0.067147	0.087462	0.126207	-0.381682	0.166206	-0.156724	1.000000

For the time being, it is safe to set aside the hypotheses about pay equity having to do with the number of women or relative number of women in an occupation.

K-means clustering was used to assess if STEM occupations cluster in a way to subvert or heavily influence the trend of men being paid more than women. Below, the STEM occupations are in yellow and the non-STEM purple:

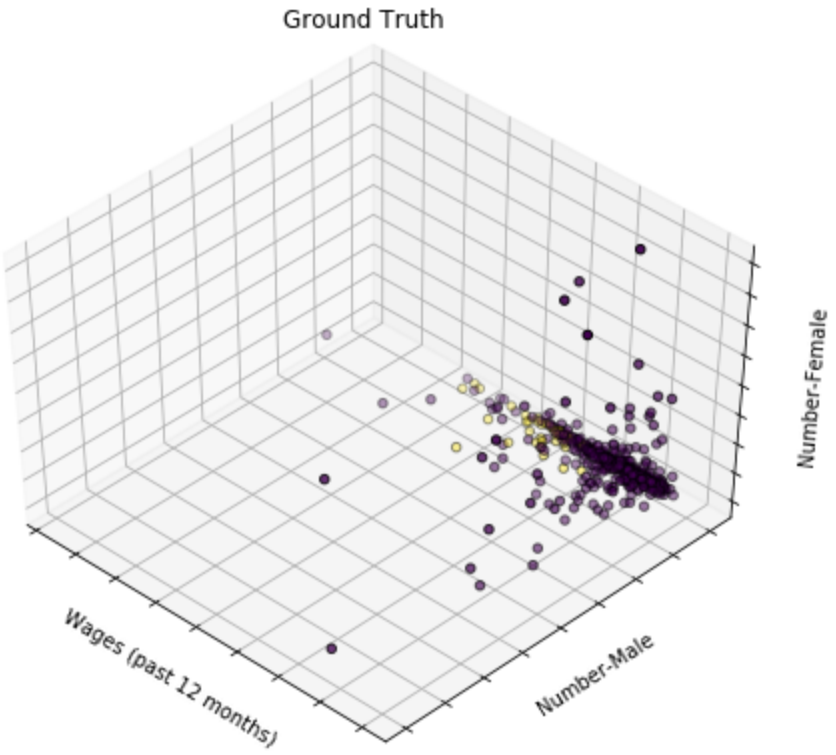


However, from this three-dimensional graph, it is hard to see how much the STEM occupation data clusters. Since it is already well-known that STEM occupations are higher paying than non-STEM overall, the k-means clustering for just the women's and men's pay was examined more closely:

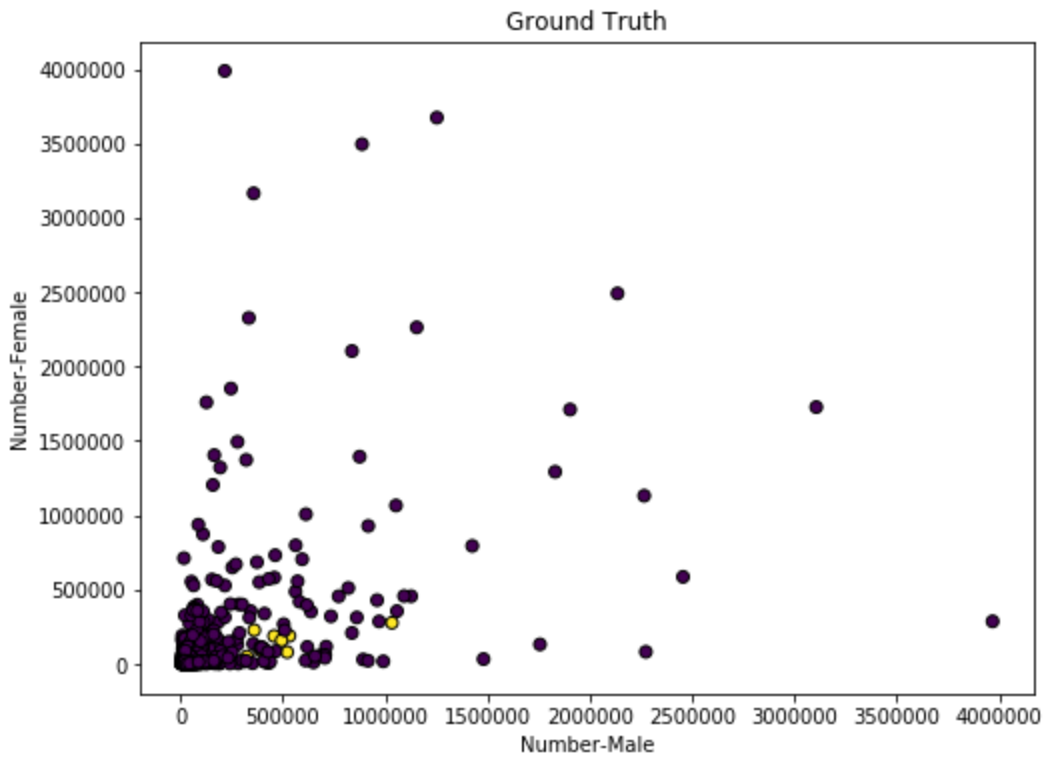


While it is apparent that the STEM positions cluster somewhat, they do not appear to have any sort of abnormal behavior in relation to the remainder of the data. They do not have a different trendline.

The process was repeated for wages and numbers of men and women in each occupation:



Again, it is difficult to assess the relationship of the STEM occupations (again in yellow) to the others (purple), so again, a two dimensional plot was created to look more closely at the numbers of men and women:



Here, the only thing that is apparent is that there is a greater number of men than women in the STEM occupations.

Results

To return to my original questions, it seems that:

- The distribution of pay for men does differ from that for women. It has a higher median and a longer tail on the high-pay side.
- The number of women in an occupation is not correlated to women's pay. In other words, a large number of women in an occupation does not indicate that the pay will be more evenly distributed.
- A larger difference between the number of men and the number of women in an occupation does not correlate to a larger difference in pay between the two. In other words, if an occupation shows more gender parity in the number of workers, it does not necessarily show more gender parity in pay.
- An occupation being a STEM occupation does affect whether there is a greater pay, but it does not clearly indicate that there is greater or lesser pay or number disparity between the genders.

This suggests the following about the justifications discussed before:

- The pay for men is not actually that different from the pay for women, it's just that men tend to be in more prestigious positions that pay more. The gender pay gap is actually due to a few outliers and not an overall trend.
    - **PARTIALLY TRUE, PARTIALLY FALSE**: Men's pay does have a longer tail on the high-paying end, but it also has a higher mean and median than women's pay.
- Female-dominated jobs pay less. When there are enough women in an occupation, one can be assured that it is lower paying
    - **FALSE**: The number of women in an occupation does not correlate with pay parity.
- When an occupation has parity in numbers between women and men, it also has parity in pay. Equality on one front influences equality on the other.
    - **FALSE**: The two do not appear to correlate.
- STEM occupations follow a different trend than other occupations. They tend to have more men and pay more, and this explains why the overall pay for men versus women appears to be different.
    - **INCONCLUSIVE**: Further analysis would be necessary to show this to be true.

Conclusions

This research has verified some hypotheses and found inconclusive results for others. Going forward, it would be interesting to see how other factors than STEM or gender affect pay. It would be interesting, for example, to examine how geography or race affect pay and how highly correlated these factors are to the ones studied here.

It would also be interesting to push this research forward through the use of other analytic methods to assess these hypotheses. For example, given certain biased training data, does a machine learning algorithm that mimics our pay per occupation turn out data that looks similar to our dataset. Such a result would go a long way toward proving that bias affects pay disparity.

References

[1] "Several New Studies Tie Diversity to Innovation, Profitability." Gartner Staff. *Talent Daily*. February 16, 2018. https://www.cebglobal.com/talentdaily/several-new-studies-tie-diversity-to-innovation-profitability/

[2] "Questions & Answers on Pay Equity." *National Committee on Pay Equity*. https://www.pay-equity.org/info-Q&A.html

[3] "Men Want Powerful Jobs More Than Women Do." HBS Working Knowledge. *Forbes*. September 23, 2015.

https://www.forbes.com/sites/hbsworkingknowledge/2015/09/23/men-want-powerful-jobs-more-than-women-do/#702526bd3e94

[4] "Explaining the Gender Wage Gap." Sarah Jane Glynn. *Center for American Progress*. May 19, 2014. https://www.americanprogress.org/issues/economy/reports/2014/05/19/90039/explaining-the-gender-wage-gap/

[5] "Separate and Not Equal? Gender Segregation in the Labor Market and the Gender Wage Gap." Ariane Hegewisch, M.Phil., Hannah Liepmann, Jeff Hayes, Ph.D., Heidi Hartmann, Ph.D. *Institute for Women's Policy Research*. August 31, 2010. https://iwpr.org/publications/separate-and-not-equal-gender-segregation-in-the-labor-market-and-the-gender-wage-gap/

[6] "How the gender pay gap permeates science and engineering."Nic Fleming. *New Scientist*. February 28, 2018. https://www.newscientist.com/article/mg23731670-100-how-the-gender-pay-gap-permeates-science-and-engineering/

[7] "ACS 5-Year Estimates - Public Use Microdata Sample." US Census Bureau. 2017. https://data.census.gov/mdat/#/

[8] "www.bls.gov/oes/stem_list.xlsx." Occupational Employment Statistics. *Bureau of Labor Statistics*. October 29, 2019. https://www.bls.gov/oes/topics.htm#stem

[9] "Most Americans believe STEM jobs pay better, but few see them as offering more flexibility for family time." Cary Funk, Kim Parker. *Pew Research Center*. January 9, 2018. https://www.pewsocialtrends.org/2018/01/09/most-americans-believe-stem-jobs-pay-better-but-few-see-them-as-offering-more-flexibility-for-family-time/