What's Behind the Gender Pay Gap?

Ariel Silbert
December 3, 2019

Introduction

- Diverse workplaces tend to be more innovative and productive.
- Pay disparity between genders could prove to be a barrier.
- Women make 78¢ to every $1 that a man makes.

* Note: For citations, please see the references section of the paper "What's Behind the Gender Pay Gap?"

Research Questions

Reason	Hypothesis
Female-dominated jobs pay less.	Is the number of women and/or men in an occupation correlated to each gender's pay?
Men tend to be in more prestigious positions that pay more.	Does the distribution of pay for men differ from that for women?
When an occupation has parity in numbers between women and men, it also has parity in pay.	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?
STEM occupations tend to have more men and pay more.	Does an occupation being a STEM occupation affect whether there is a greater pay or number disparity between the genders?

Information on the Dataset

- US Census Bureau Microdata, "ACS 5-Year Estimates - Public Use Microdata Sample" for 2017
- Occupational Employment Statistics, Bureau of Labor Statistics, "stem_list.xlsx"

Methodology

- 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.

Feature Analysis

	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	191740.355200	213222.073400	152777.891800	3.968151e+06	3.987371e+06

Distribution of Pay

- Mean wages for men were higher than for women, as were the median wages.
  - However, the discrepancy between them was larger than expected.
- The standard deviation for men's wages was also larger than for women's.
  - This suggests 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.
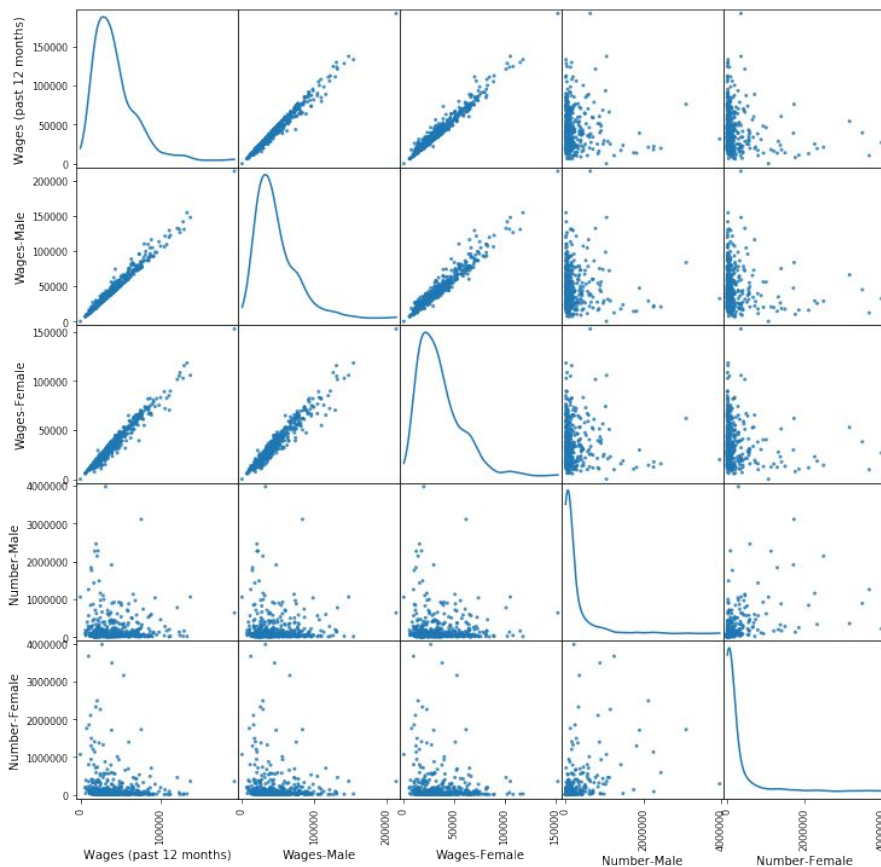
Distribution of Pay

Men's pay does have a longer tail on the high-pay end. However, it's center is also higher.

Pay by Gender Across Occupations
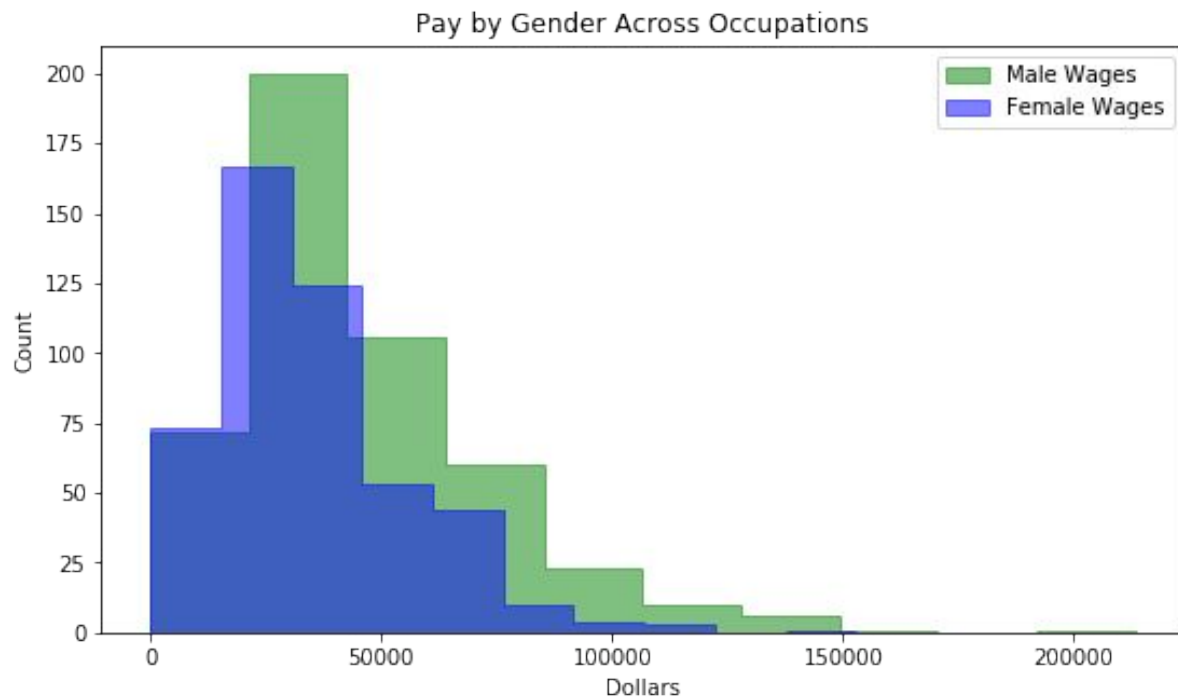
SPLOM

- Each of the wages subsets are highly correlated.
- The remainder of the elements do not seem that highly correlated.

Correlation Analysis

|	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

Female Dominated Roles Pay Less

Both the SPLOM and the correlation analysis show that there is very little correlation between the number of male or female workers in an occupation and the pay for either gender.

A Different Perspective

Normalized Pay Differential = (Male Wages - Female Wages) / Wages

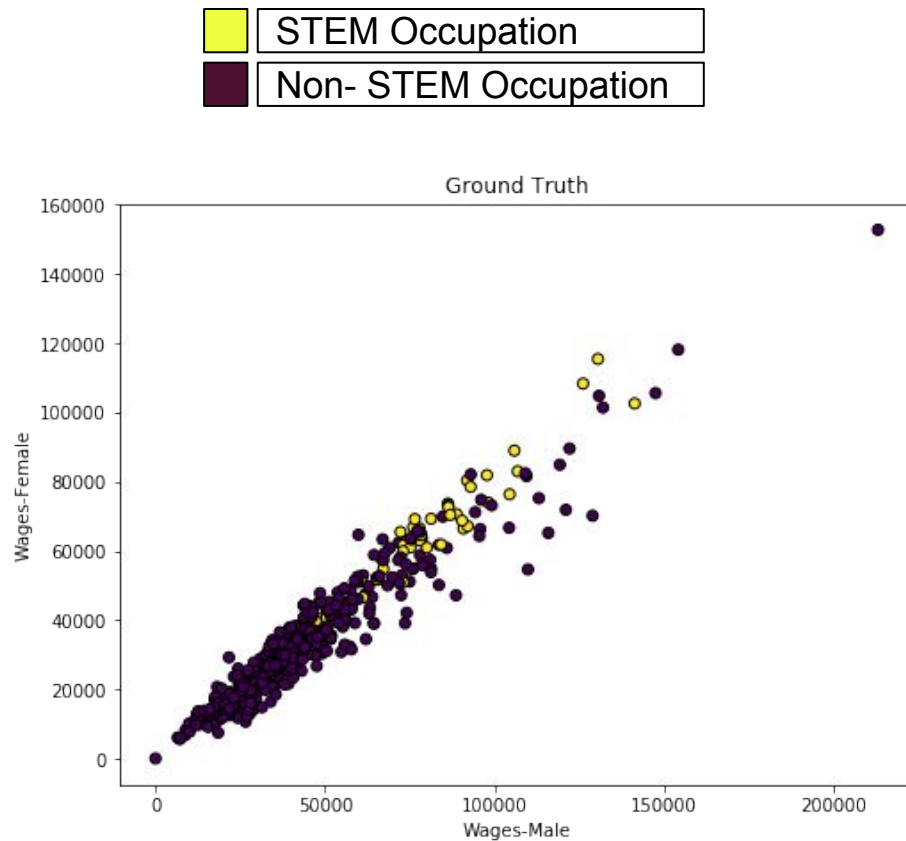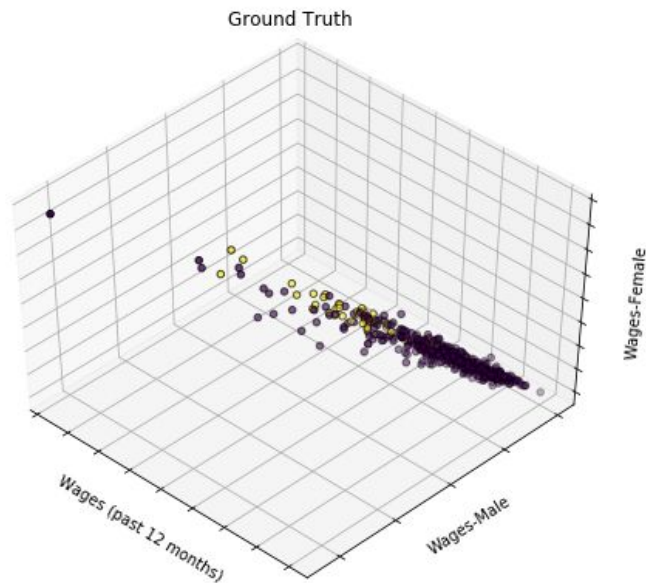Normalized Number Differential = (Male Number - Female Number) / (Male Number + Female Number)
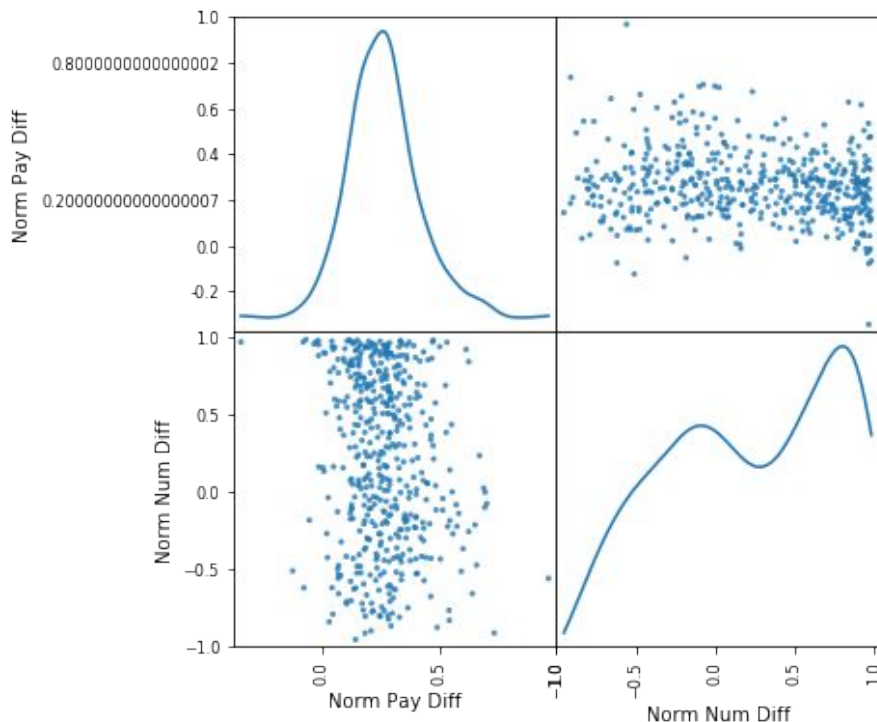
Parity in Numbers Leads to Pay Parity

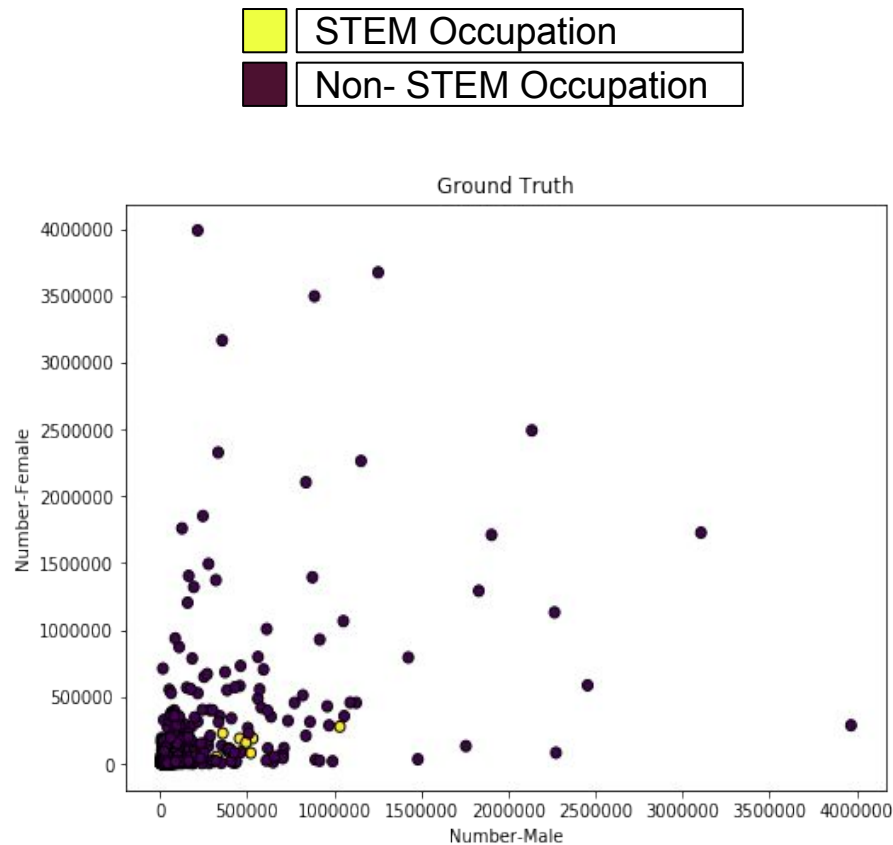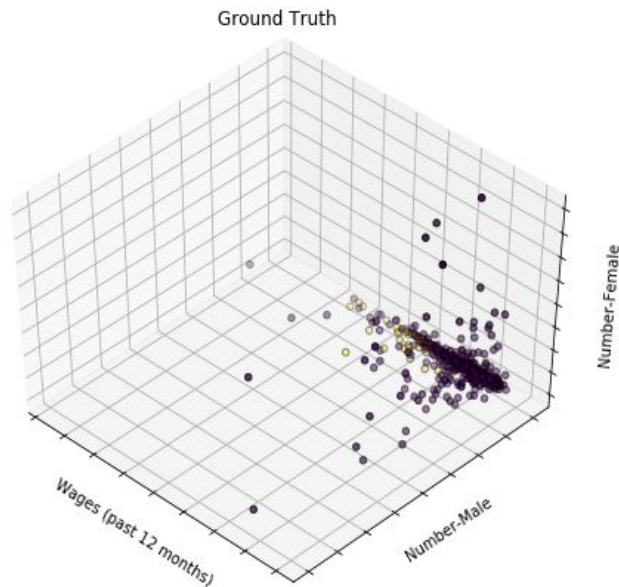Neither element seems highly correlated with the other, a suspicion that was verified by their correlation coefficient of -0.156724.

Normalized Number and Pay Differential Across Occupations

K-Means Clustering

Ground Truth

Ground Truth
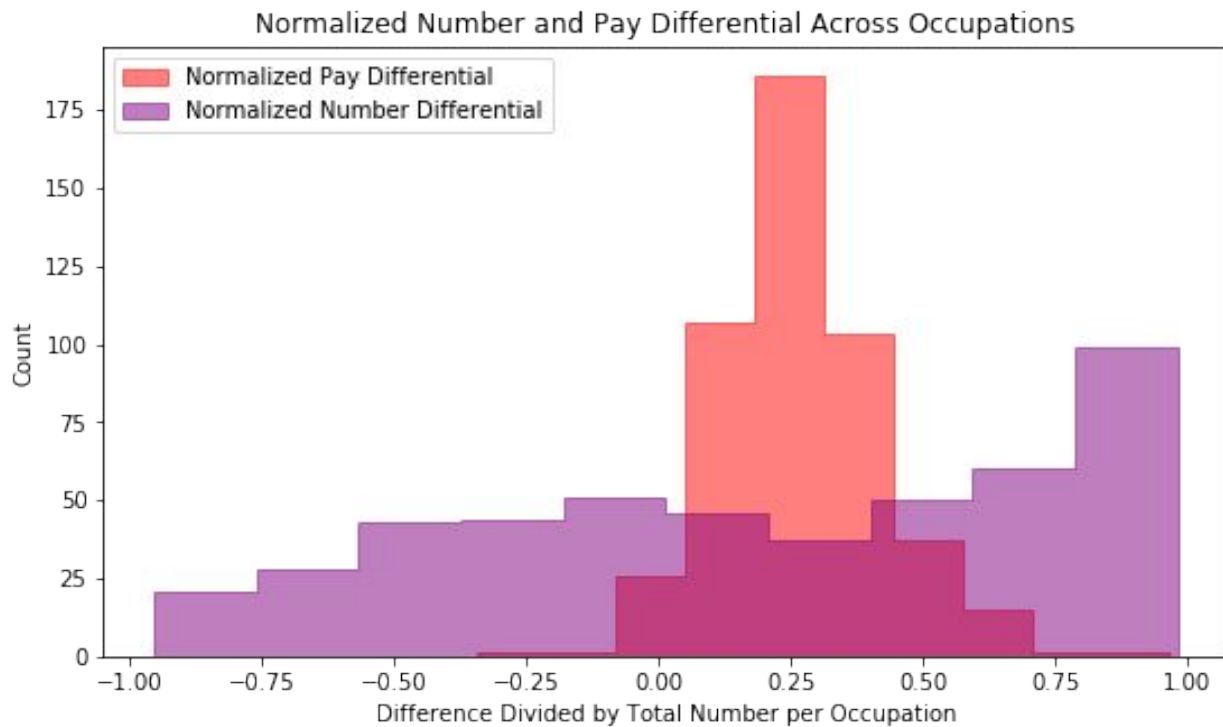
- STEM Occupation
- Non- STEM Occupation

K-Means Clustering

STEM Occupations Provide Undue Influence on Pay

INCONCLUSIVE: Further analysis would be necessary to show this to be true.

Thank you.