

Statistical Analysis of Gender Pay Differences Across College Majors in US Census Data

Philip Xu (jx24), Henry Laseter (hlaseter)

Introduction

Gender pay gap has increasingly entered the national conversation, and is an important topic of discussion for gender equality. We want to explore the numbers behind the pay gap and analyze how the US census data reflect the national discussion.

Data

We obtained the data from the U.S. census bearau, from the Public Use Microdata Series (PUMS) From this dataset, we selected the relevant fields (sex, wage, degree type, hours worked) and cleaned the data to construct our male and female datasets.

Data Cleaning

After visualising initially, we noticed random null values in our FOD1P (degree type) field. We trimmed our full dataset of approximately 3 million down to approximately 800,000 by removing entries where the FOD1P field was null.

Hypothesis

We believe that there's a statistically significant connection between gender and wage, especially for the highest earning majors.

Methodology

We decided to use linear regression to model the relationship between sex and wage, as well as a few other variables (age, education, year of entry) just to explore their impact. A t-test was a clear choice to test for statistical significance; as it test for relationships,

Results and Visualisation

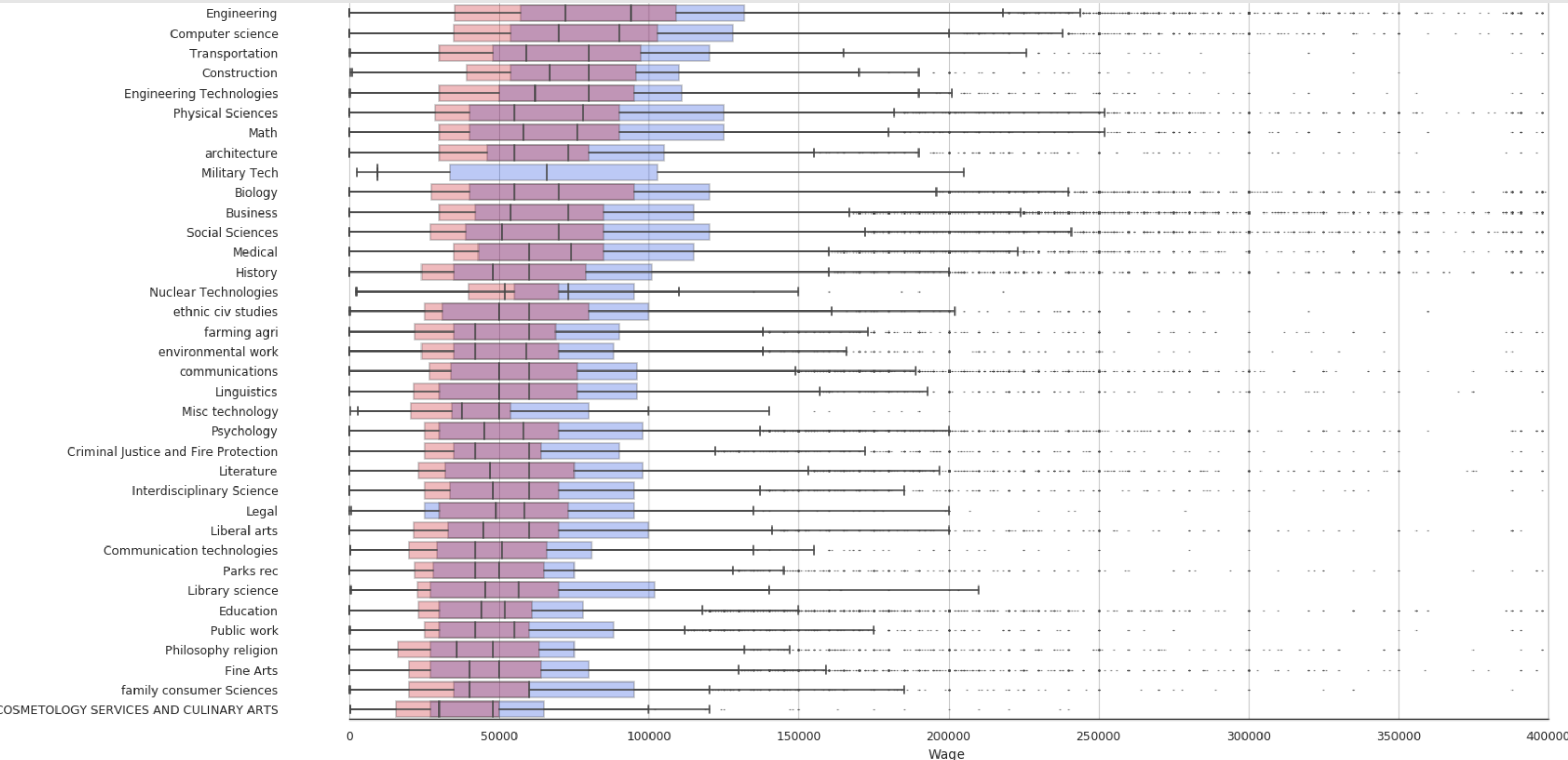


Figure 1: Income distribution across all majors, male = blue, female = red. As we can see there is a significant difference in wage distributions between male and female across all majors.

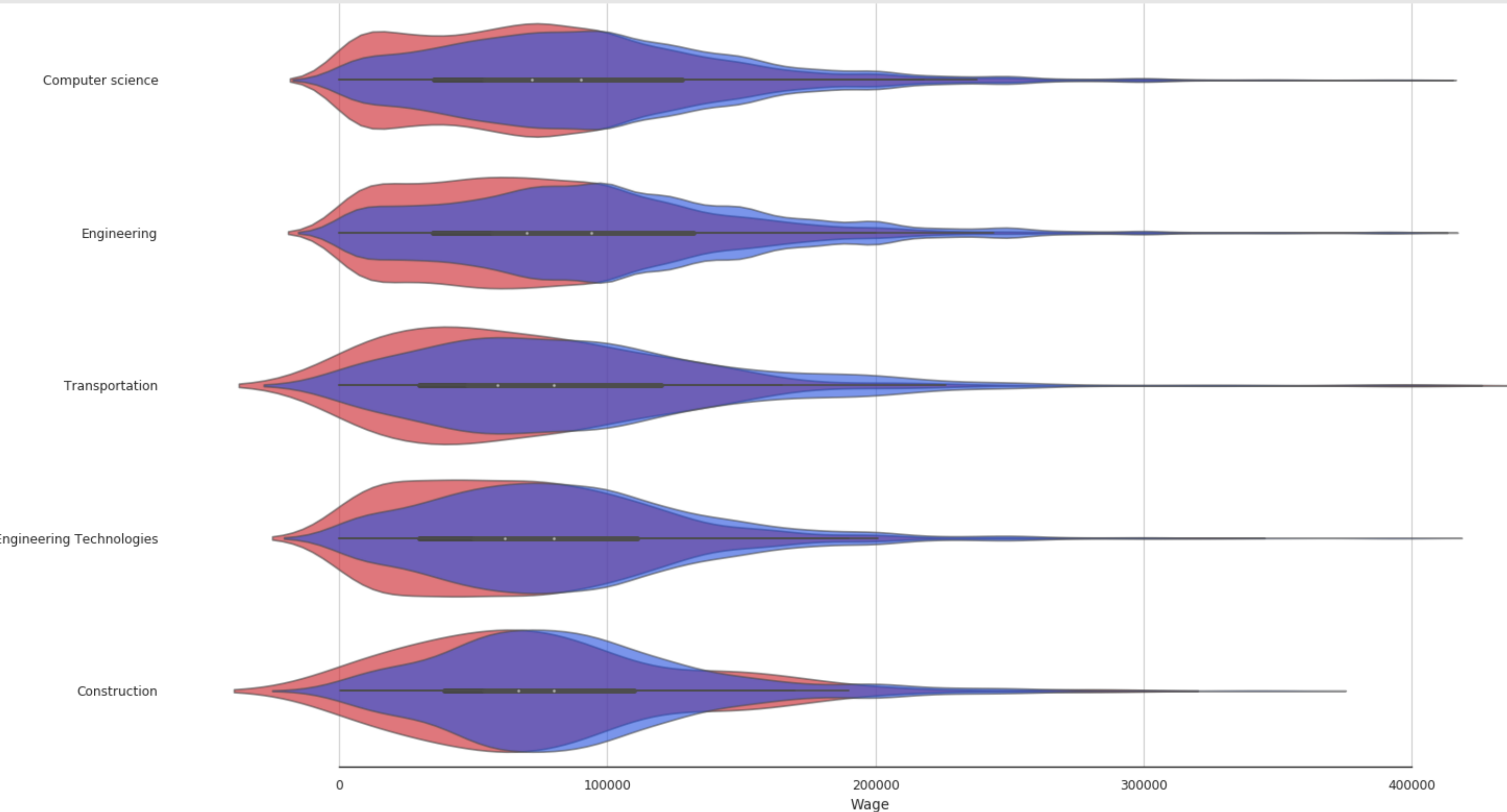


Figure 2: Income distribution across top 5 highest paying majors, male = blue, female = red.

	coef	std err	t	P> t
const	-140.6687	1.929	-72.931	0.000
SEX	-13.7775	0.144	-95.653	0.000
Y0EP	0.0956	0.007	13.767	0.000
SCHL	8.4288	0.090	94.125	0.000
AGEP	0.3984	0.005	74.855	0.000

	coef	std err	t	P> t
const	60.7539	0.233	261.113	0.000
SEX	-14.9712	0.146	-102.666	0.000

Figure 3: Results for multiple and single variable linear regression with respect to wage.

The Effect of Age

We found clear evidence that older age relates with higher income.

We also found an interesting relationship. We found statistically significant evidence that the wage gap grows with age for younger (age 20-45) workers, but not for older (45-70) workers.



T-Test Result

With an alpha level of 5%, we clearly showed the impact of sex on wages, both in the single variable case and in multiple regression involving age, experience, education level, and average wage within a person's degree.

Conclusion

We have enough evidence to reject the null hypothesis and conclude that there is a statistically significant relationship between sex and wage.