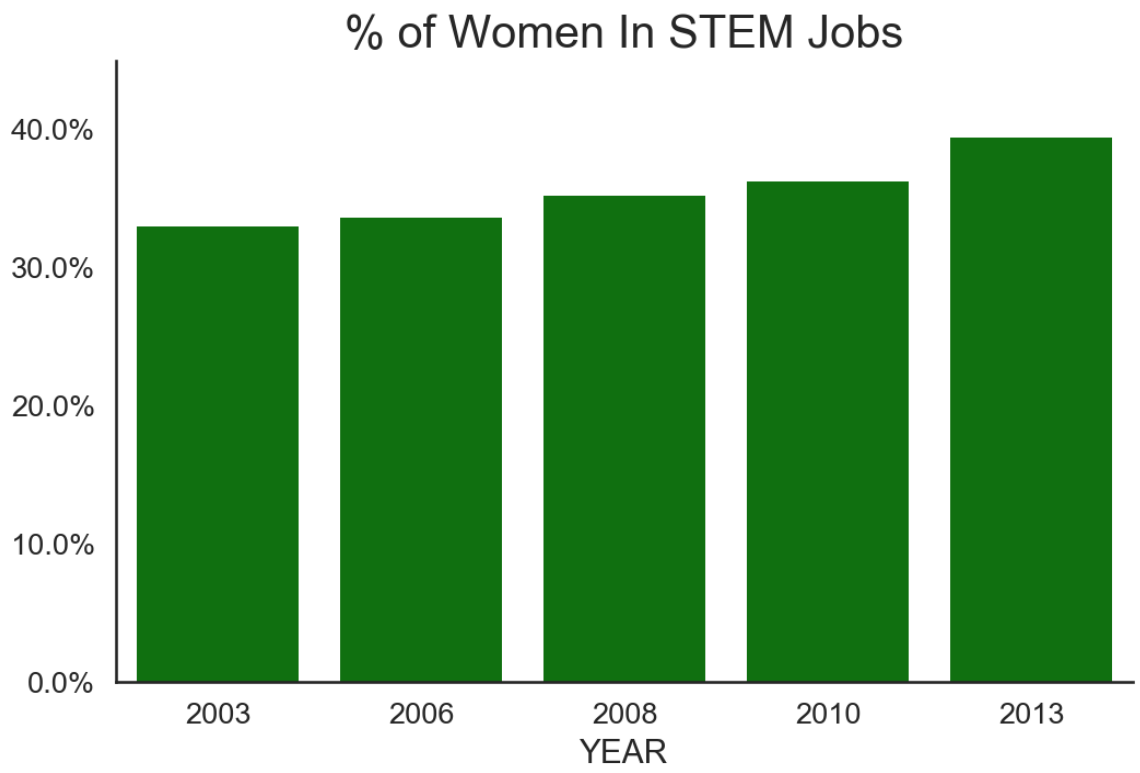


Women In STEM - Midterm Report

Prepared by: Shuojia Shi
Contact: shishuojia1987@gmail.com
March 17, 2018

DEFINE THE PROBLEM

STEM, which stands for science, technology, engineering and math produces many job opportunities in US. However, it is reported by U.S. Department of Commerce that women comprise 48 percent of the U.S. workforce but just 24 percent of STEM workers. This is means half as many women are working in STEM jobs. Why is this happening? What people majored in during higher education plays a very important role in the types of work force they enter after they graduate. So, does this mean less female students are majored in STEM topics as well? There might be many factors contributing to the unbalanced situation, including personal interest, cultures, and plainly not seeing of the job opportunity after graduation. In this project is going to look into the job outcome between men and women. I would like to find the facts about what types of jobs the females STEM students can get into, what kind of salaries would they expect to receive, and the possibility to optimize their job searching strategy.

IDENTIFY THE CLIENT¶

1. Women majoring in STEM or preparing themselves into STEM jobs.
2. Career building websites such as Linkedin.com and CareerBuilder.com.
3. Universities and higher education institutes.
4. Media and Journalists

DATA WRANGLING¶

My data was found on the Higher Ed website (https://highered.ipums.org/highered/). The relevant variables such as demographic, education and employment were carefully chosen with the available samples between 2003 and 2013. It comes in csv format. The detailed information about the labels can be found in the pdf file ‘code_label'. With Python's pandas package, the data can be read into the system. There are 478747 rows and 33 columns in the data file, with a size of 120+ MB.
The various null values in the data was filtered and replaced with python friendly identifiers. Several groups of data were stored in different data frames for easy manipulation. For example, the comparison of the education levels of graduate and undergraduate can be easily called from two data frames.The data comes in a clean form. Each column represents a variable and each row is an observation. The columns containing majority of null values are dropped and won't be considered for this work. The columns missing about 10% values turn out to be the exact columns giving details about the

employment. The number of the missing values are the same as the number of the unemployment count. The NAN values in these cases actually have real life meanings and were left where they are. They will be taken into considerations during analysis.

OTHER POTENTIAL DATA SETS¶

1. General US labor force data for employment rate and salary gap comparison.
2. Major/job fields detailed data

KEY FINDINGS

% Women in STEM Jobs

The figure on the front page of this report shows the percentage of women workers in STEM jobs between the year 2003 and 2013. It is clear that in this Higher Ed survey, less than 40% of the STEM workers are women . There is a steady increasing tread in the female proportion over the years (from 30% to 40%). It could be a good sign: more women are getting into the STEM field. Next let's take a look at how does the salary differ between the genders.

Pay Discrepancy Between Genders

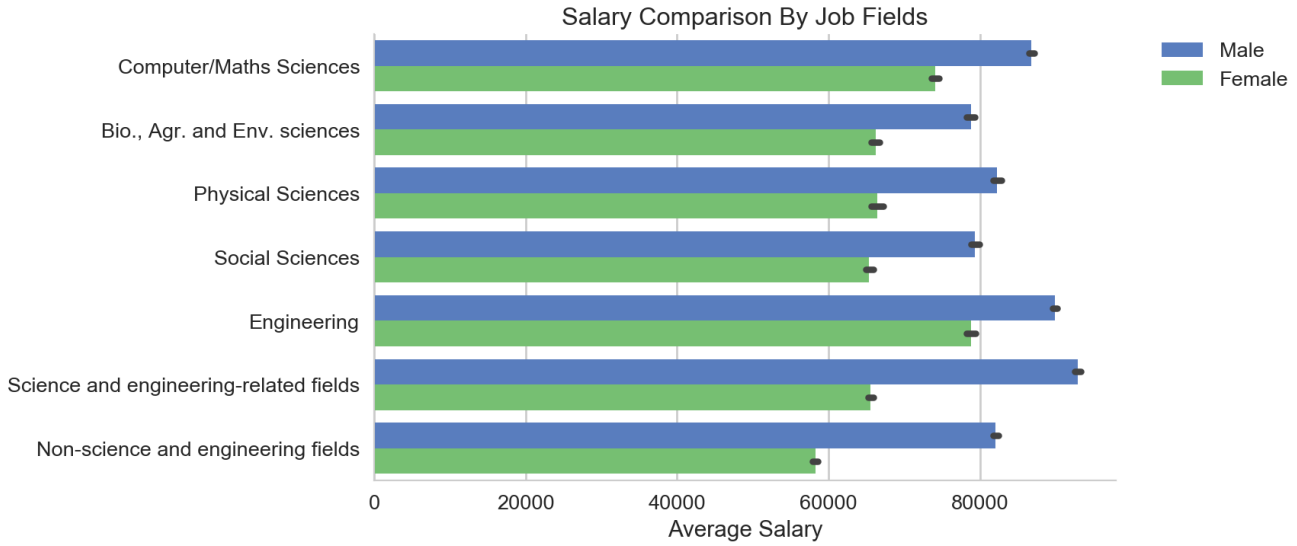
The box plot in the following page shows the median and IQR for both men and women workers in STEM between the year 2003 and 2013. Several things were shown here:

1. The median, and the quatiles wages for men is about 30% higher than women, for all the studied years. The statistical analysis and central limit theorem tell us the difference in this data set is significant.
2. There is NO shrinking trend in the gender discrepancies during these years.
3. The mean wages increased from year 2003 to 2008. Then we see a small decrease from 2008 to 2013. This makes sense considering the financial crisis in 2008. It is interesting to see the magnitude of it affecting the job market.

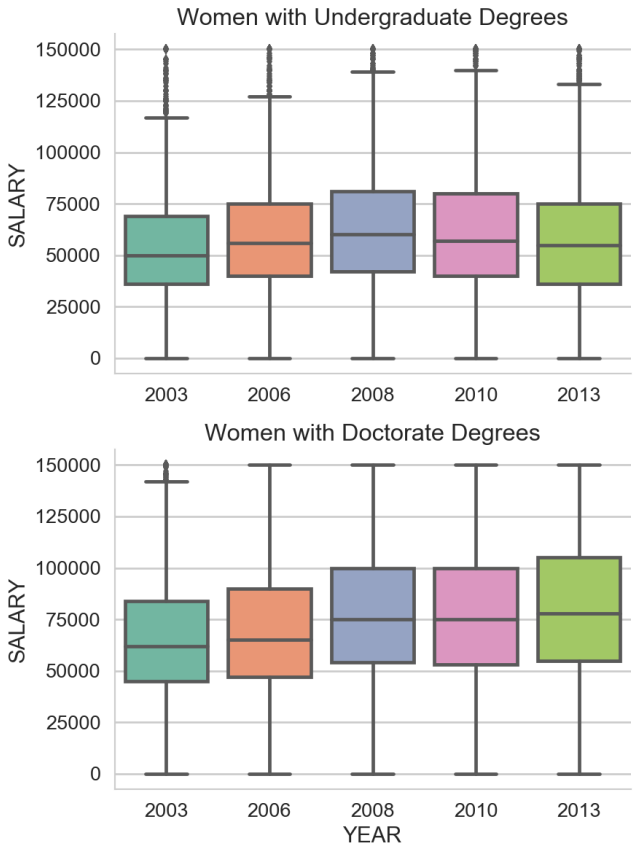
Gender Disparity By Job Fields

The next plot shows the salary comparison between men and women in several STEM fields. We see that the jobs that have the least pay gap between men and women are: engineering and computer/ maths sciences. These two fields also held two highest pay positions for women workers in STEM.

On the other hand, the largest pay gap goes to non-science and engineering fields and science and engineering related fields. These are the two fields containing management positions, STEM or not. This could be indicating the even severe gender disparity in the salaries of management positions.



Highest Degree of Women STEM Workers



How does the highest educational degree of a woman affect her career in STEM?

The box plot on the left presents several points: First of all, it is easy to see a steady increase in salary from year 2003 to 2008. But things get interesting after 2008.

The upper subplot shows the salary for women with undergraduate highest degree. It clearly indicates the trending down after year 2008.

The lower subplot shows the salary boxplot with doctorate degrees over the years. We see a difference! It did not show a down trend after 2008. Instead, the mean salary for year 2010 kept about the same as 2008 and an even an increased in 2013.

In a financial crisis like what happened in 2008, the doctorate women group in average was influenced least and recovered fastest. It gives some insights to the need for highly specialized professionals from our society.

SUMMARY

In this mid-term report of my capstone project: Women In STEM, I have re-iterate the problem (motivation) and suggested the clients who could be interested in this project. The steps taken for data wrangling in this study was briefly explained. Some keys findings were listed above: the increase of women workers in STEM seems to indicate the shrinking gender disparity between genders in STEM. But then we found the unchanged pay gap between men and women (as well as among different job fields). It is also found that the salary increased more over the years for women with higher educational degrees.

There are still aspects left to study: comparison of this analysis with US labor analysis, look into the details of the fields of STEM, machine learning and solving problems and etc. They will be included in the final report of this project.

Find the data and the EDA/statistical analysis code on my Github:
https://github.com/shuojiashi/capstone_project_1/tree/master/notebooks