Women In STEM

Shuojia Shi
Springboard Data Science Career Track

Outlines

- Motivation. Potential Clients.
- Data Source and Wrangling
- Explorative Data Analysis
  - Women participation in STEM
  - Unemployment and gender pay gap
  - Women in STEM in job fields.
- Statistical Analysis— is it statistically significant?
- Classification
  - Feature selection
  - Predict salary levels
  - Predict principal job fields
- Summary

Motivation & Potential Clients

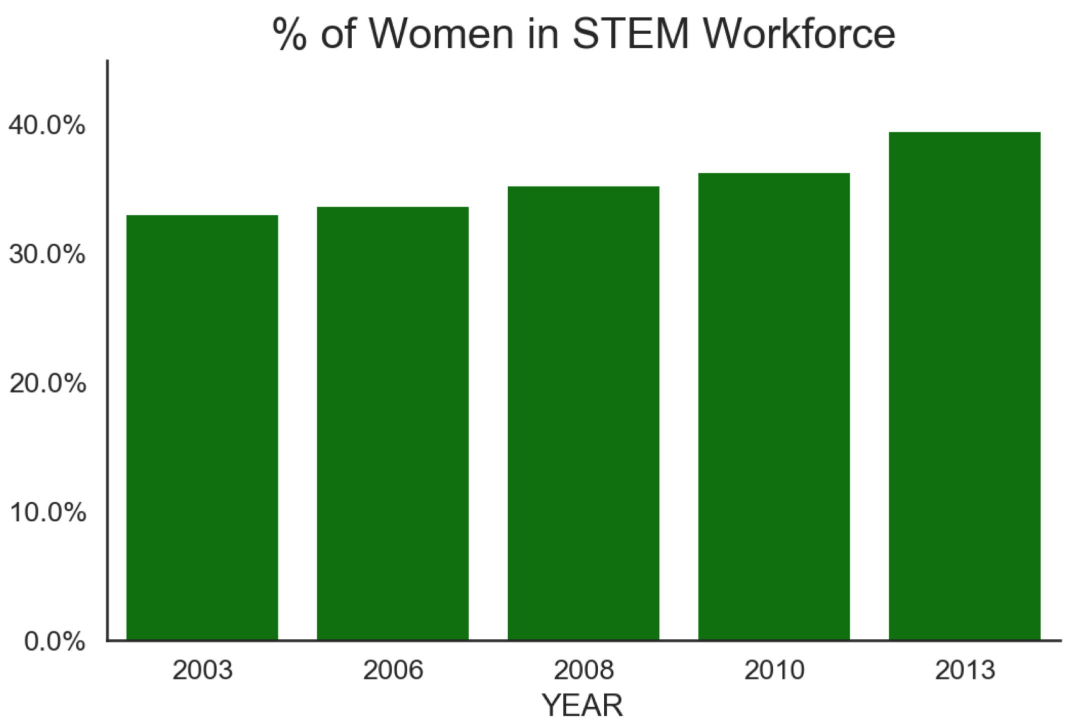
- It is reported that half as many women are working in STEM jobs[1]. It is also reported that women make 78 cents for every dollar men make[2].
- Questions I would like to answer:
    - How many women are actually in STEM workforce?
    - Does the gender pay gap change over the years?
    - How do race/job fields affect the gender gap?
    - Can we predict the salary level for women STEM workers?
    - Can we recommend a job field for a certain individual? …
- As a woman in STEM field myself, I am intrinsically interested in these problems.
- Potential clients:
    - Women majoring in STEM or preparing themselves into STEM jobs.
    - Career building websites such as Linkedin.com and CareerBuilder.com.
    - Universities and higher education institutes.
    - Media and Journalists

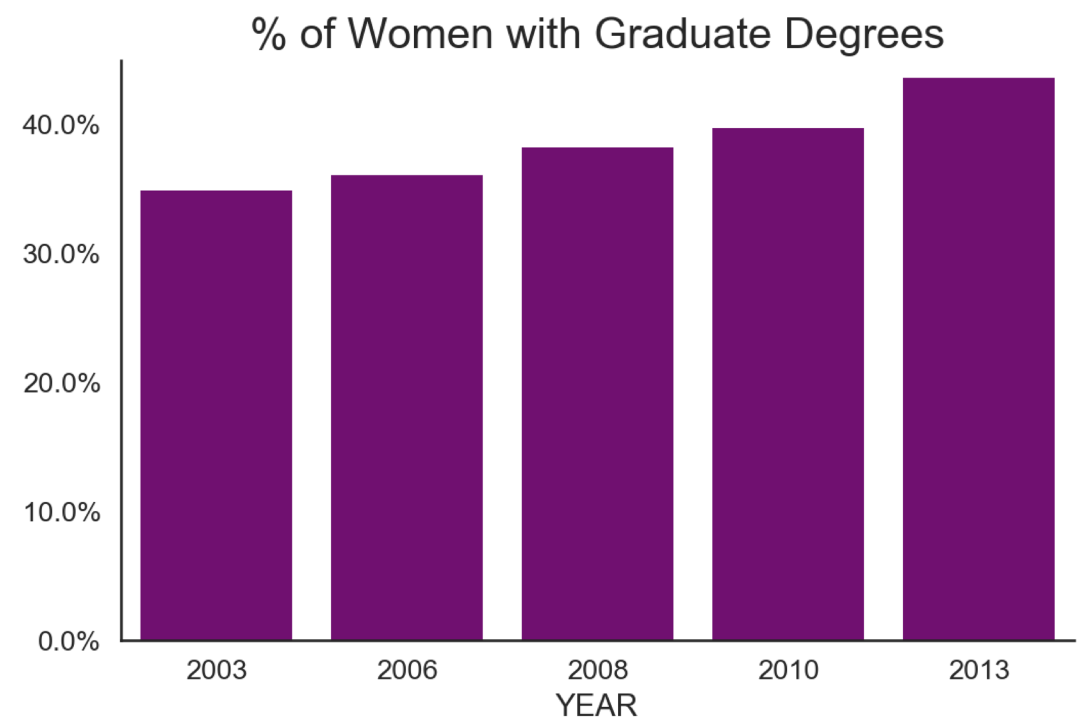
1. **http://www.esa.doc.gov/sites/default/files/womeninstemagaptoinnovation8311.pdf**
2. **http://money.cnn.com/2016/04/12/pf/gender-pay-gap-equal-pay-day/index.html**

Data Source and Wrangling

- Survey data from IPUMS Higher Ed https://highered.ipums.org/highered/
- Leading surveys for studying the science and engineering (STEM) workforce in the United States.
- Data include National Surveys of College Graduates (NSCG) and Doctorate Recipients (SDR)
- The relevant variables such as demographic, education and employment were chosen with the available samples between 2003 and 2013.
- Filter columns with mostly missing data.
- Final data size: 478747 entries and 47 columns (175.3+ MB).

Explorative Data Analysis
Women participation in STEM

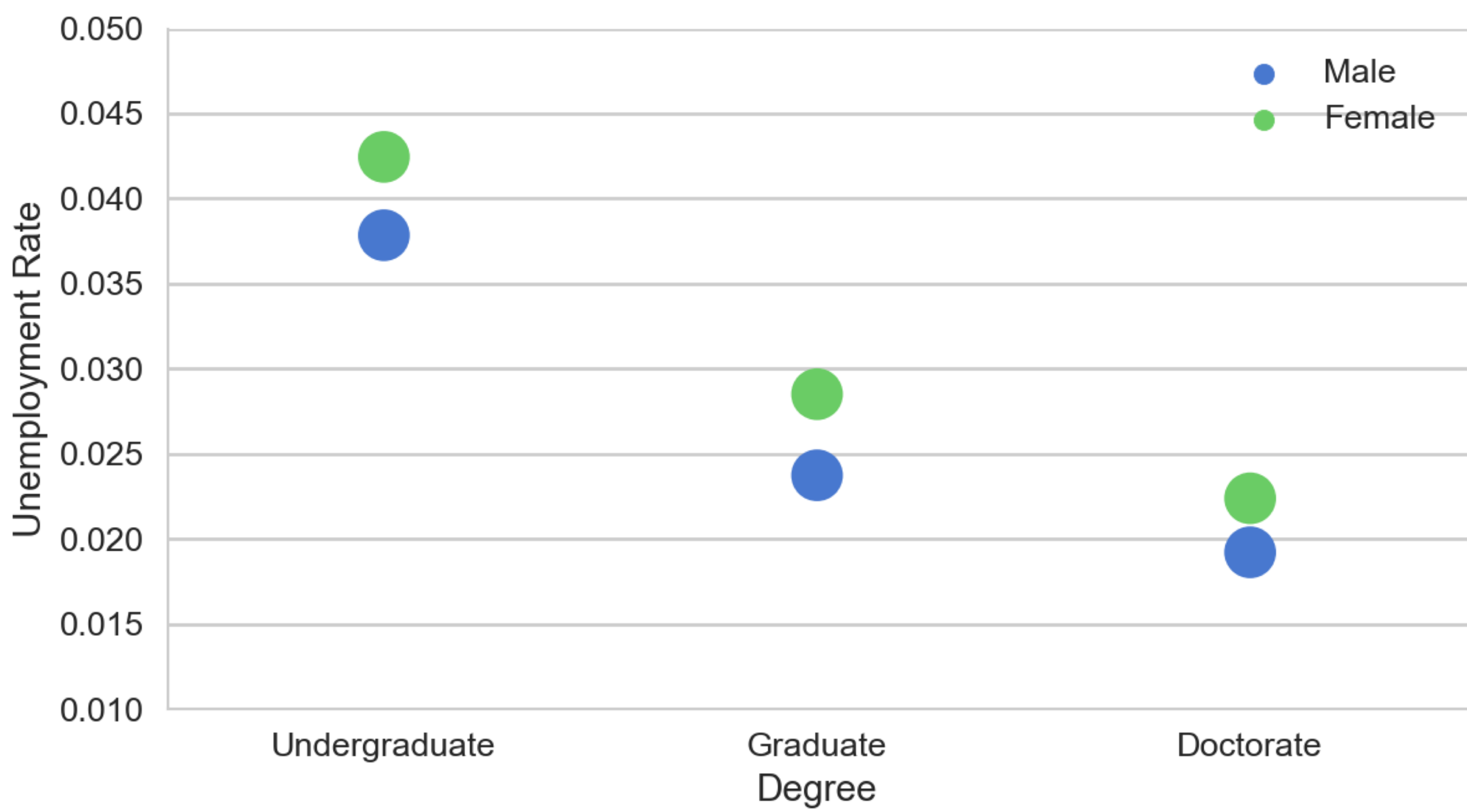




- Bar plot on the left: women made up 33% of the STEM workforce in 2003.
- That number rises steadily over the next ten years with a jump from 2010 to 2013, as the percentage rose to almost 40%!
- We see a similar trend in the % of women with graduate degrees in the bar plot on the right.

Explorative Data Analysis

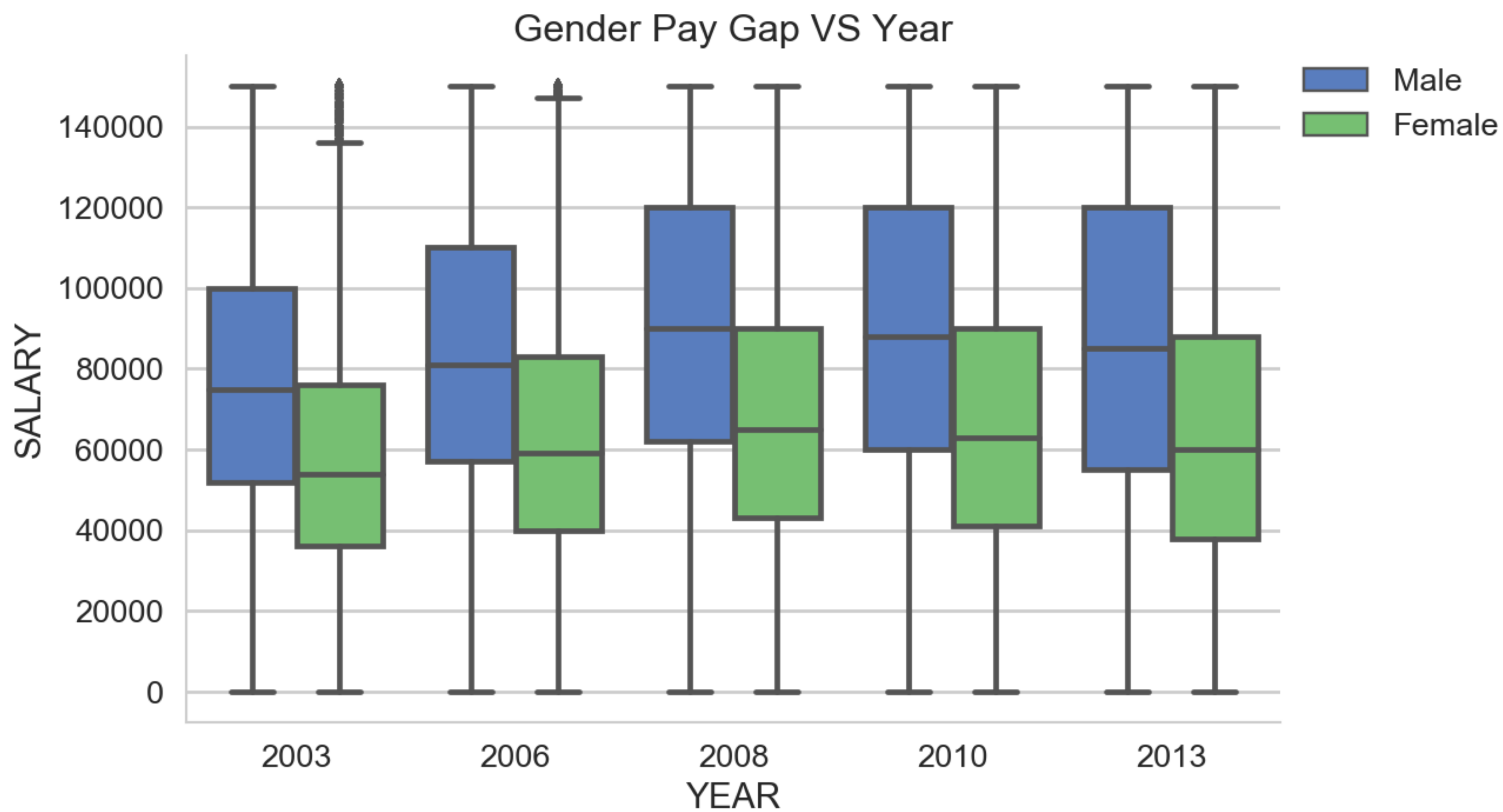
Gender unemployment gap



- Unemployment rate trending down with higher education degrees.
- Men has slightly lower (1%) unemployment rate than women.
- The gender unemployment gap shrinks with higher education degrees.
- Note the rate here only consider individuals that are in labor force.

Explorative Data Analysis

Gender pay gap (2003 to 2013)

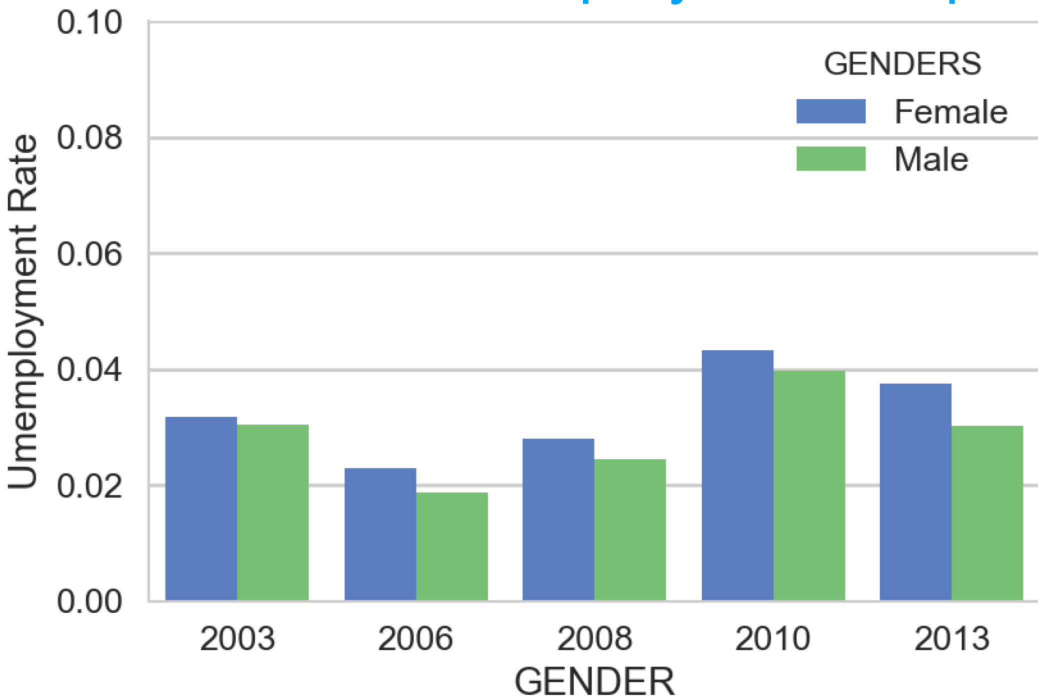


- Women have come a long way but are still not at parity. We have seen women's increased education and workforce participation. But women makes consistently 30% less than men!
- There is NO shrinking trend in the gender pay gap over the years.
- The mean wages increased from year 2003 to 2008. Then we see a small decrease from 2008 to 2013. This makes sense considering the financial crisis in 2008.

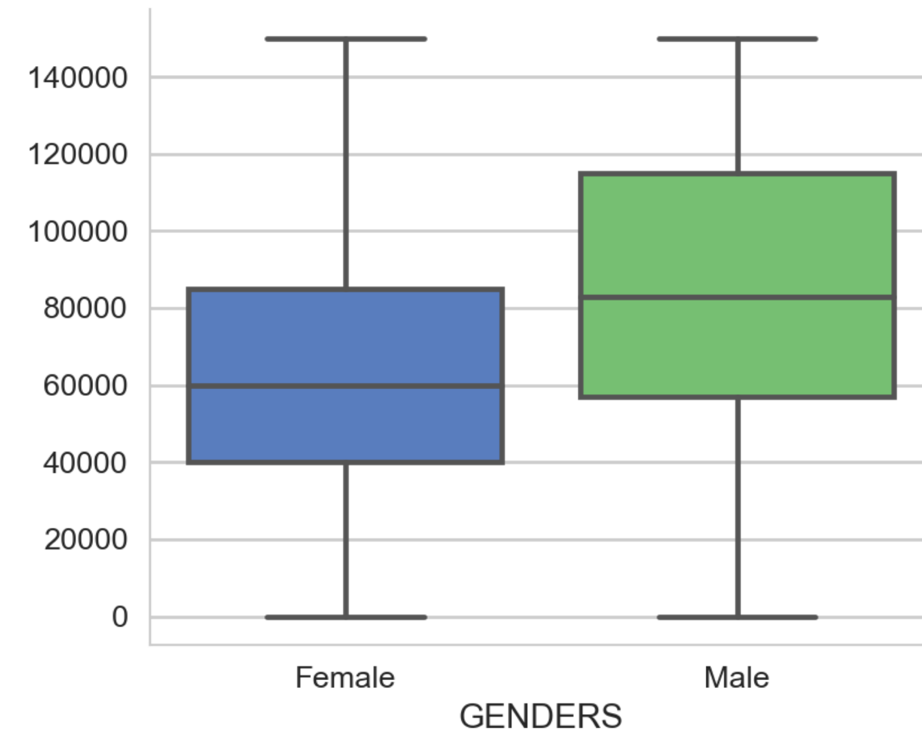
Statistical Analysis

Gender makes a difference in STEM workforce, statistically.

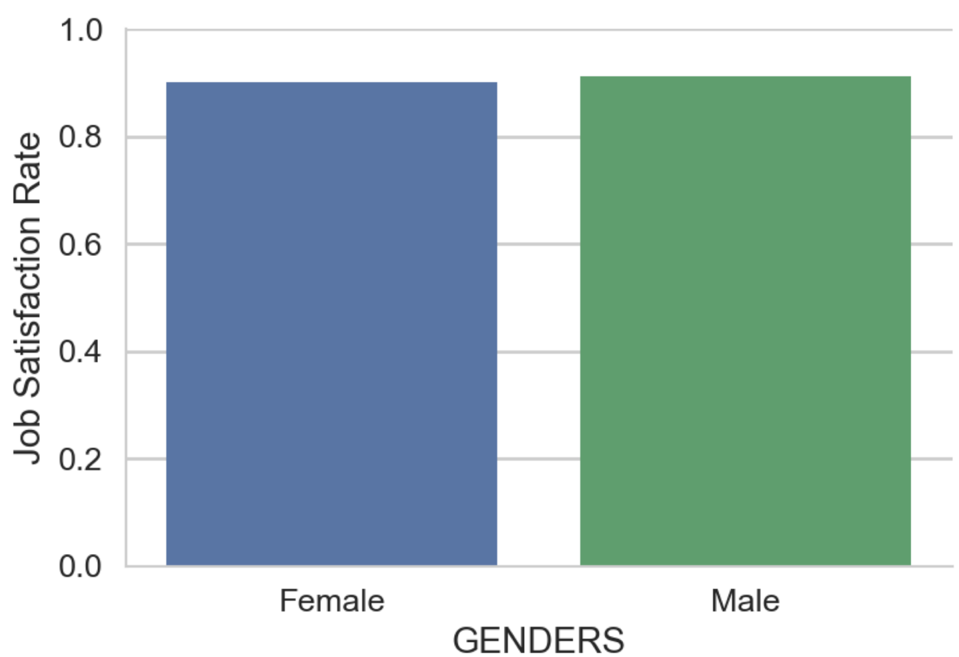
Gender Unemployment Gap



Gender Pay Gap

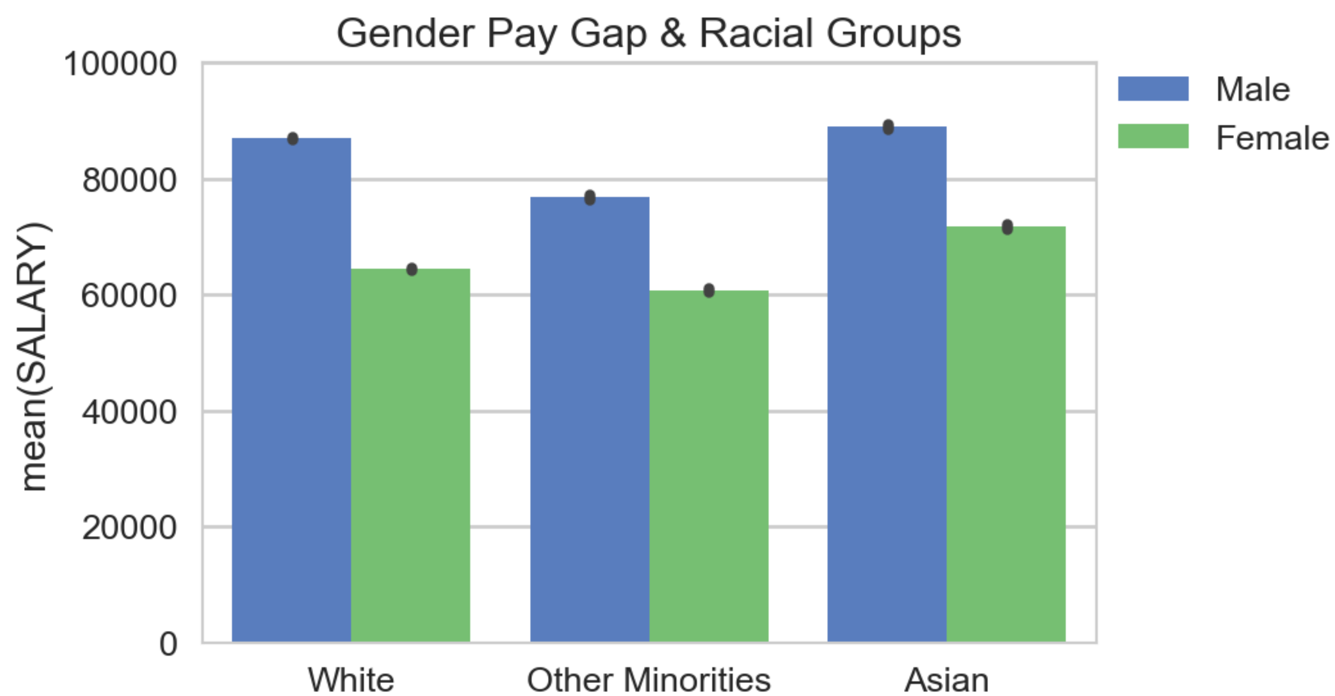


Gender Job Satisfaction Gap



Feature Examined with Gender	Difference	P-value from χ^2 test	Conclusion
Gender Unemployment Gap	0.5%	7.8e-17	Significant
Gender Pay Gap	30%	0	Significant
Gender Satisfaction Gap	1%	3.3e-24	Significant

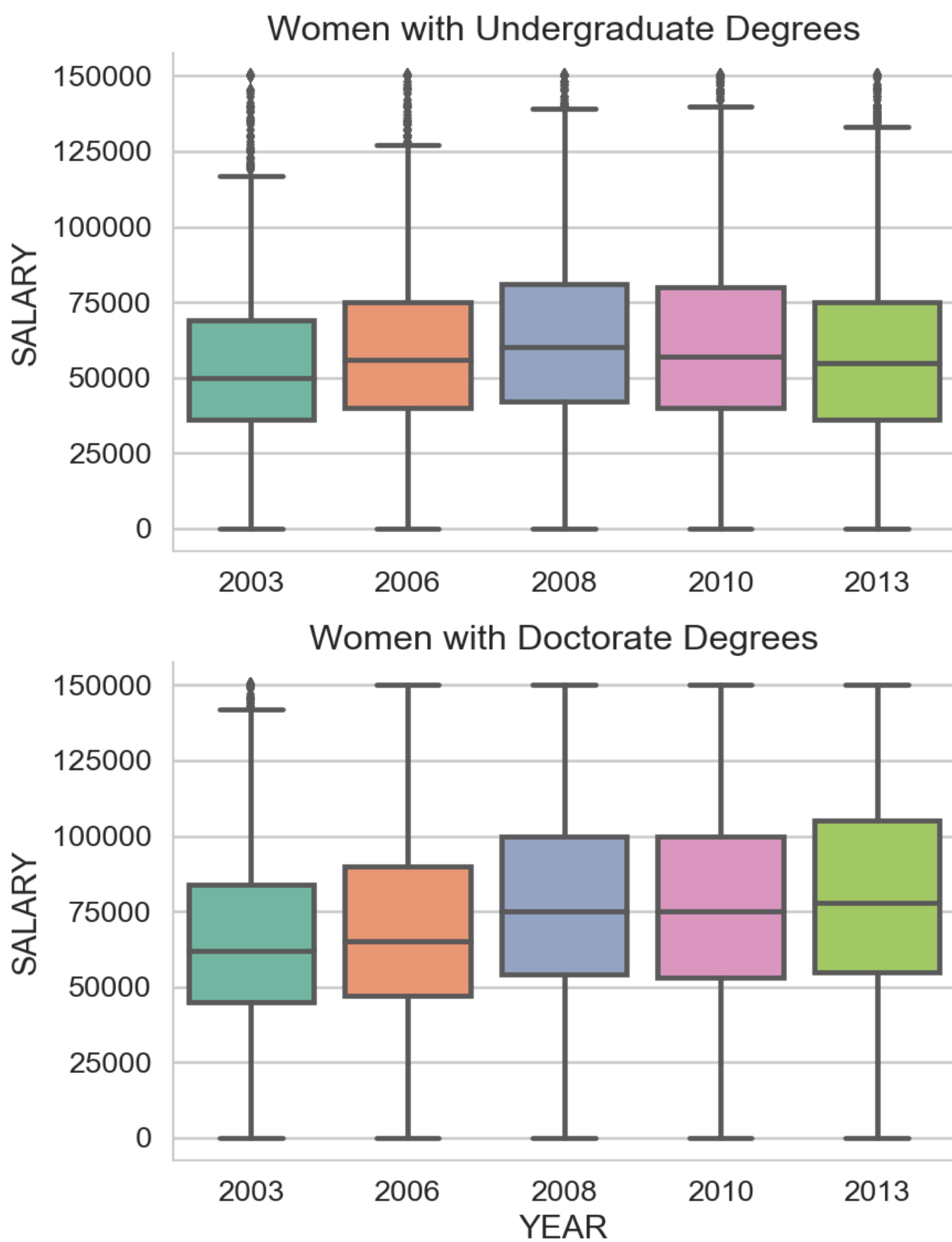
Explorative Data Analysis
Gender/race pay gap



- The pay gap between men and women exist in all race groups.
- For every dollar Asian men make, **Asian** women make **81 cents**.
  For every dollar white men make, **white** women make **74 cents**.
  For every dollar **other minorities** (include black, Hispanic, and etc.) men make, women in the same group make **79 cents**.

Explorative Data Analysis

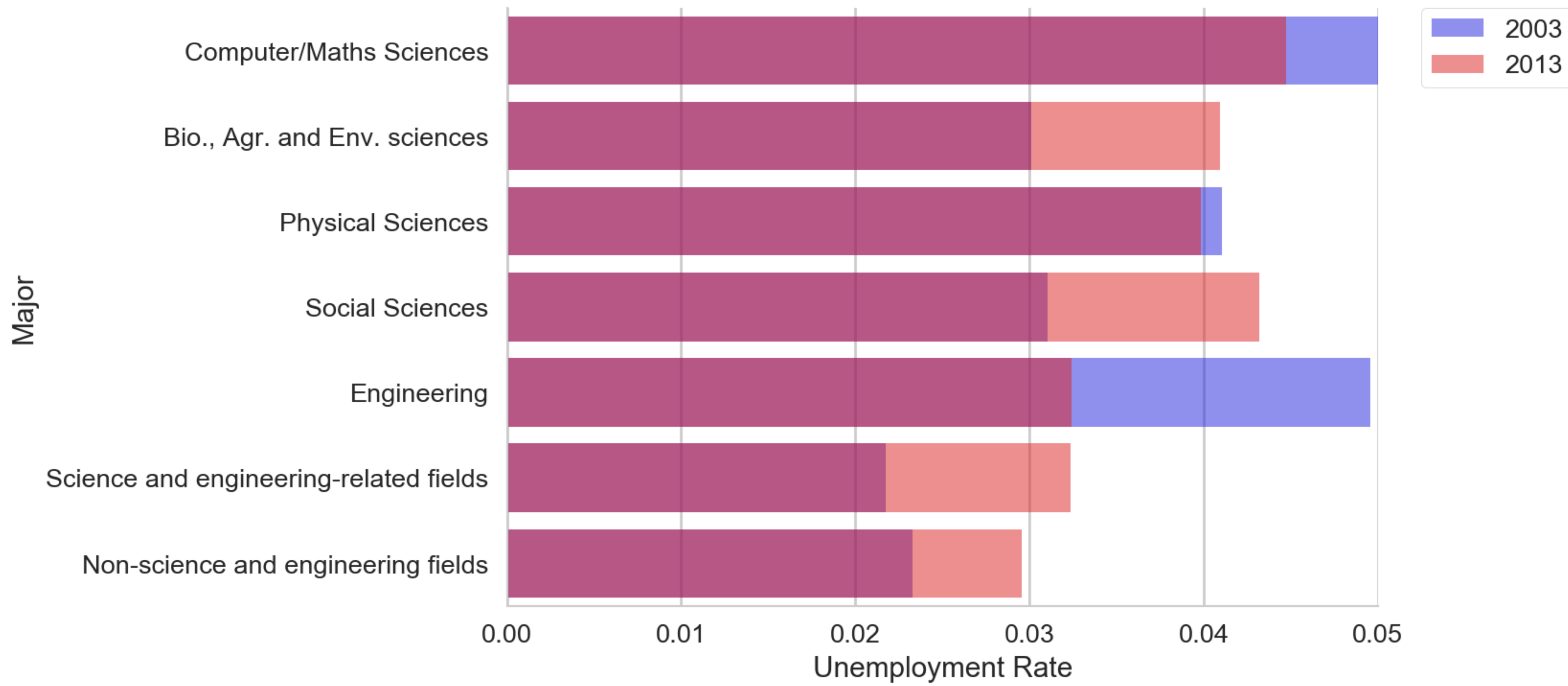
Women In Education Level Groups



- The plot on top presents the quantiles of the salaries for women with undergraduate degrees. It clearly indicates the trending down after year 2008.
- The plot on the bottom shows salaries for women from the doctorate group. Big difference! This median wage and quantiles kept increasing after 2008, even though the increase from 2008 to 2010 is minimal!
- Median wage for women doctors is 20% higher than that of women with undergraduate degrees. This difference is even larger after 2008.

Explorative Data Analysis

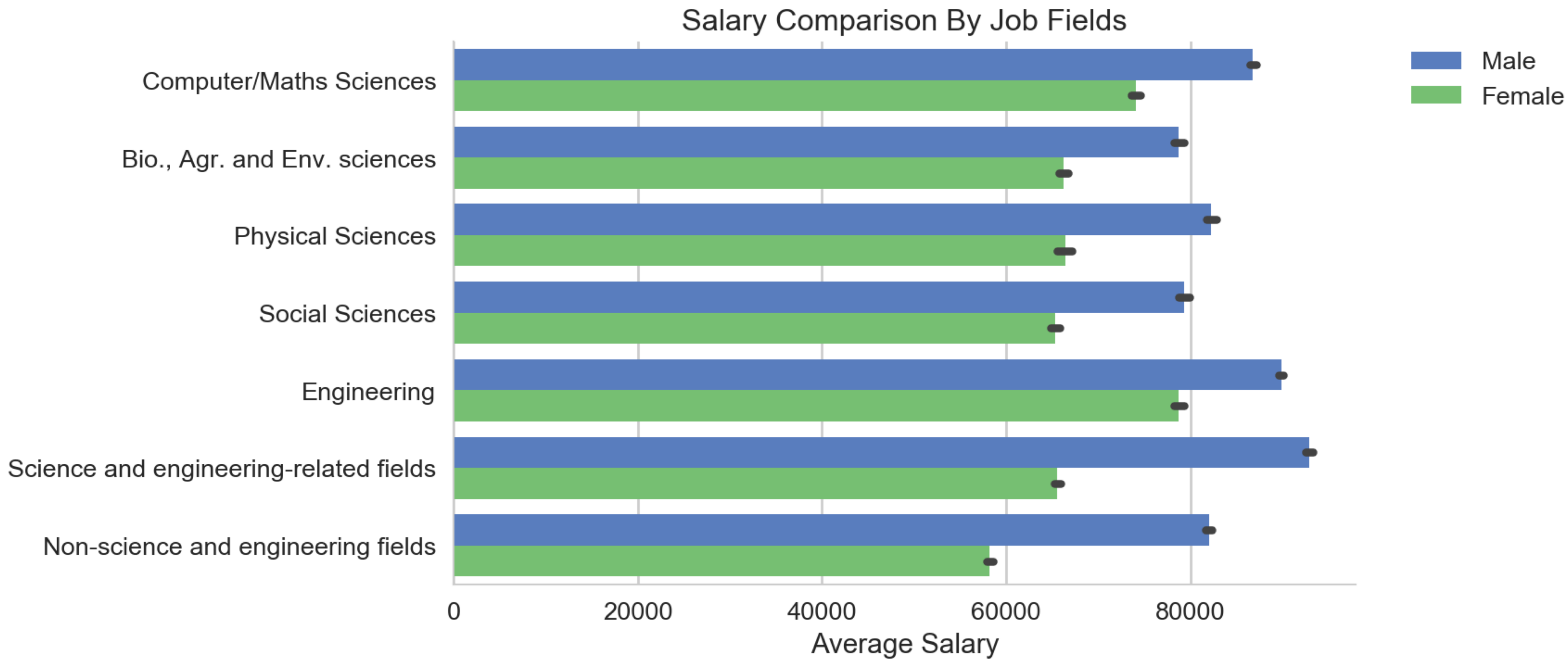
Fields of major– women unemployment rate



- The plot of female unemployment rate for field of major groups in 2003 and 2013 can give us some insight what majors are having a better performance in job placement.
- There is a significant decrease in the unemployment rate for the female engineering group (from 5% to 3%).
- Fields saw a larger increase of unemployment rate: social sciences, bio and life sciences, and other science and engineering related fields.

Explorative Data Analysis

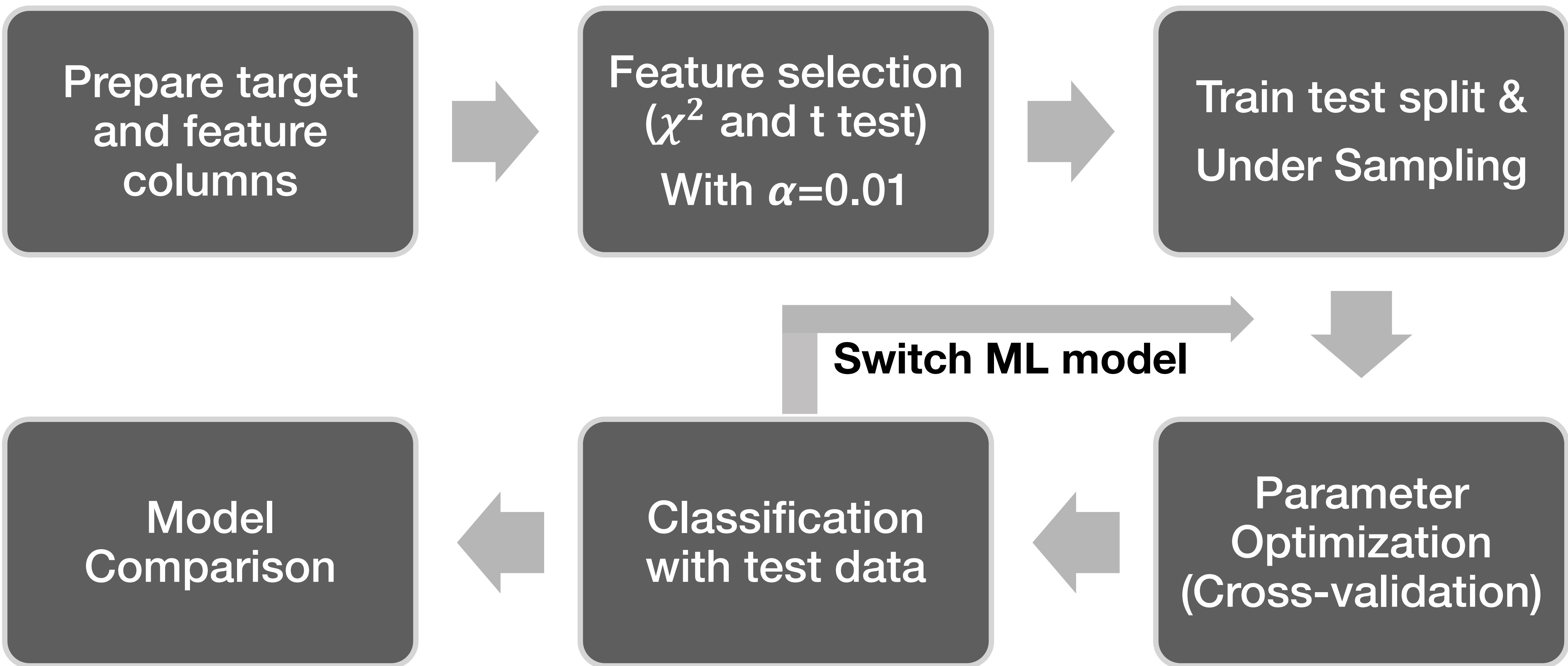
Principal job fields - gender pay gap



- The job fields that have the least pay gap between men and women are: engineering, computer/Maths sciences and bio/life sciences.
- Engineering field gives the best compensation to their female employees compared with other jobs. And women in non-science and engineering field reports lowest income averagely.
- Non-science and engineering fields has the largest gender pay gap. The field group contains job title such as 'Management' and 'Administration'. This could be indicating the even more severe gender disparity in these positions.

Classification

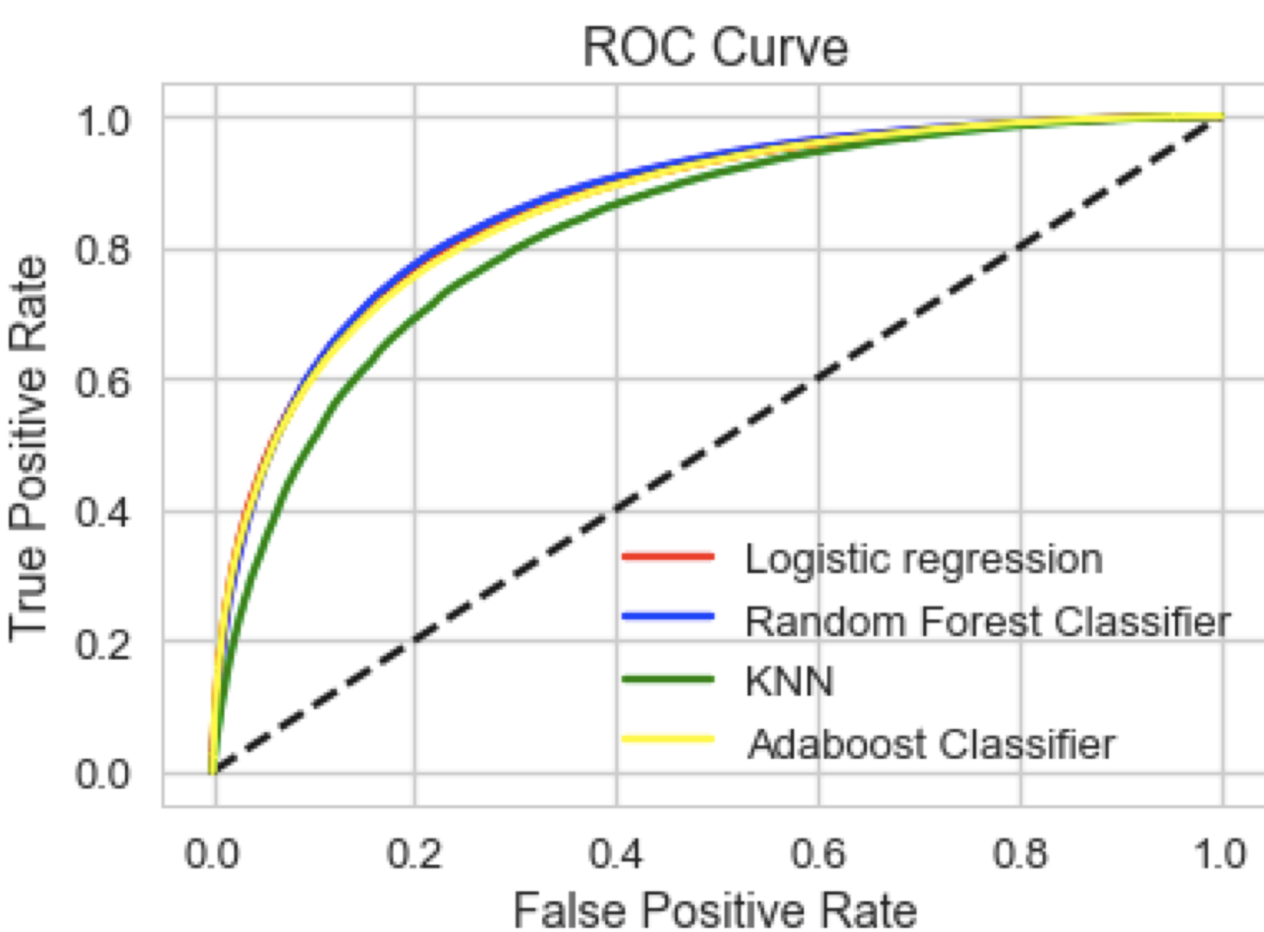
Predict Salary Level (>=60k or <60k)



Classification

Predict Salary Level

Model	Accuracy score on train	Accuracy score on test
Logistic Regression	0.782	0.815
Random Forest Classification	0.784	0.825
KNN	0.745	0.800
Adaboost	0.779	0.813



- **Random forest** ranks first among the classification models. It gives 0.78 accuracy score in predicting the salary level for full time STEM workers. Considering this survey data is missing location information for privacy reasons, this is a very encouraging score.
- **ROC curve** indicates similar result. KNN gives the worst performance among models tested.
- From the coefficients of attributes by logistic regression, we the top 5 attributes positively contributing the high salary level are **size of the employers** (>5000 employees), **professional or doctorate degrees.**
- If a recommendation system is to be built, the **job location** could be associated with the cost of living and added to the model for a more tailored result.

Classification

Predict Principal Job Field

Use only pre-employment portion of the survey data.

Similar data process as the classification of salary level.

Binarize the target column so there are *n* classification given *n* principal job fields.

Choose random forest and loop through the n principal job fields.

Field	Accuracy score on train	Accuracy score on test
Social and related scientists	0.908	0.905
Physical and related scientists	0.887	0.889
Engineers	0.878	0.878
Biological, agricultural and other life scientists	0.872	0.876
Computer and mathematical scientists	0.797	0.798
Science and engineering related occupations	0.744	0.746
Non-science and engineering occupations	0.721	0.722

Classification

Predict Principal Job Field

- Accuracy score with two tiers:
  - Tier I: Accuracy score > 0.85, social scientists, physicist, engineers, and bio/life scientists.
  - Tier II: Accuracy score < 0.8, computer and maths scientists, other STEM or non STEM related occupations.
- Top coefficients of logistic regression shows field of major and level of degree play important roles. Makes sense!
- The classification used pre-employment data. It could give reference for **women** majoring in STEM to **prepare** themselves into **STEM jobs.**
- Together with the salary level prediction, the classification could be a part of the **recommendation system** for **career building websites**.

Summary

- Women show **increased education and workforce participation** in STEM.
- **Gender DOES make a difference** in STEM workforce. We looked at the gender unemployment gap, gender pay gap. Women expect higher unemployment rate and consistently make 30% less than men!
- The gender gaps are statistically **significant**.
- Using random forest classifier, the salary level prediction report an accuracy score of 0.785 even though the survey data is missing location information. The prediction of principal job fields give an average of 0.85 accuracy score!
- Machine learning models give reference for **women majoring in STEM** to **prepare** themselves into STEM jobs. The classifications produced in this project could be a part of **the recommendation system** for **career building websites**.