PPOL670 Final Report

Team: LSX (Qianying Liu, Terry Sang, Yuchen Xiang)

Problem Statement and Background

In the past few decades, China has experienced rapid economic growth. Women's education level and social status has been improving a lot as well. As a result, more and more women have become career-oriented and entered job market. However, the issue of employment competition has also become increasingly prominent, and the current labor market has already oversupplied. Under this circumstance, gender discrimination in workplace has become a common phenomenon in Chinese job market. For example, during job interviews, Chinese women are often asked explicitly by the employers about their marriage status and fertility decisions during interviews.

Thus, for our project, we try to examine whether there is a gender gap in job market in terms of income and labor force participation, and then which factors may contribute to labor force participation in China. Since safeguarding the right of women to equal employment is of great significance in eliminating gender discrimination, balancing the status of the two sexes, promoting economic development, and building a harmonious society, our data-based study may help policymakers advocate for promoting gender equality, and provide data-based support for policymakers to make targeted policy decisions to protect and ensure women's rights and gender equality in job market.

Research Design/Approach

To explore our questions, we adopted survey data from China Nutrition and Health Survey. This survey provided us several separate datasets for more than 20,000 observer demographic information which including their, gender, education level, income, employment status, job type, marriage status, living areas and household size etc. Based on real Chinese employment situations, we considered several potential indicators of gender gap in workplace and merged these variables into one master dataset. Our goals of analysis are to examine whether there is a gender gap in Chinese job market using logistic regressions and what are the strong indicators of Chinese people's employment status through supervised learning. Based on the analysis results, policy implications will be made.

Empirical Methods

To examine our questions, we use data from the China Nutrition and Health Survey. This dataset is a representative sample of Chinese individuals, and includes information on each individual's gender, household size, fertility, education levels, and living areas. Due to the cross-sectional and complex nature of the data, the survey team divided them into different data sets. We merged several elements of the survey to arrive at a comprehensive dataset. We first visualize the descriptive statistics to explore if there are any patterns in employment status. Then, we use logistic regressions using a dummy indicating whether a person is working as dependent variable to explore if gender will be a significant indicator of employment status. Later, we use classification and regression trees to explore other potential significant indicators of employment status. Lastly, we visualize the results using Shiny.

Results and Discussion

1. Preliminary Findings

First, we used data visualization to have general understandings about the gender gap in terms of income and working status in labor market and how some variables including education level, marital status and living area influence income of men and women.

- Figure1 Average Income by Gender and Education Level

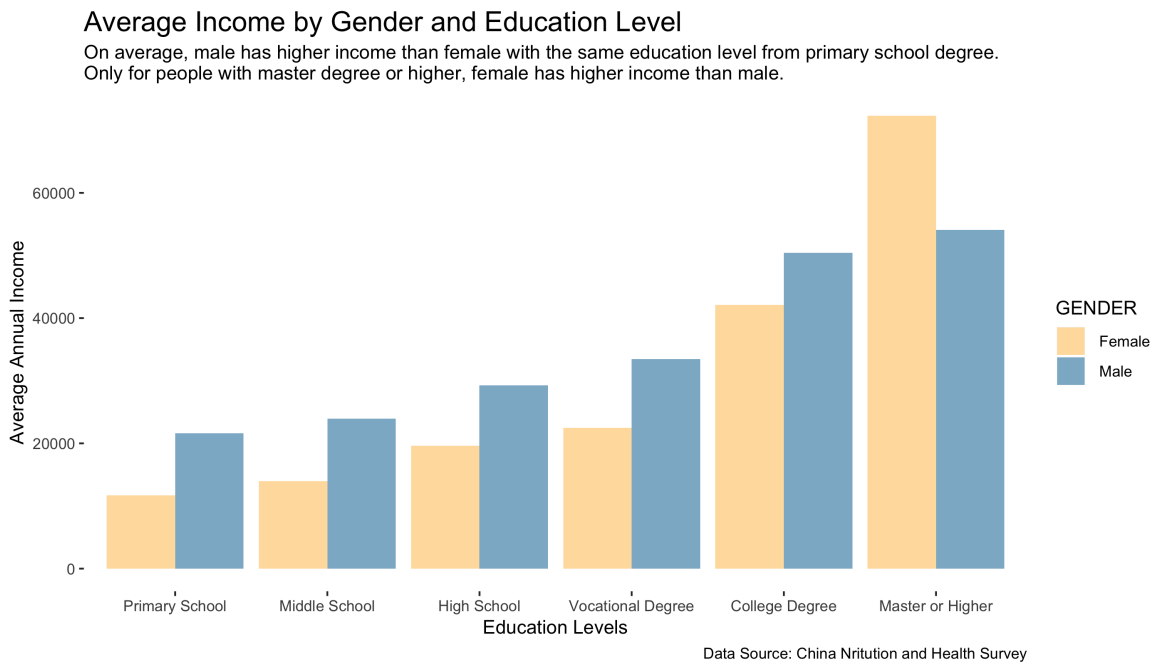


Figure 1 shows that for degrees from primary school to college, male income was obviously higher than female income in average. However, with master or higher degrees, female average income was much above male average income. Also, higher education level are closely connected with higher income for both male and female.

Also, for all the education level we considered, the percentage of employed men of total is higher than employed women of total. As a result, it may imply that male and female isn't treated equally in job seeking process. However, for both male and female, ratio of

working people with master degree or higher is lower than that with college degree.(Figure in Appendix)

- Figure2 Average Income by Gender and Marital Status

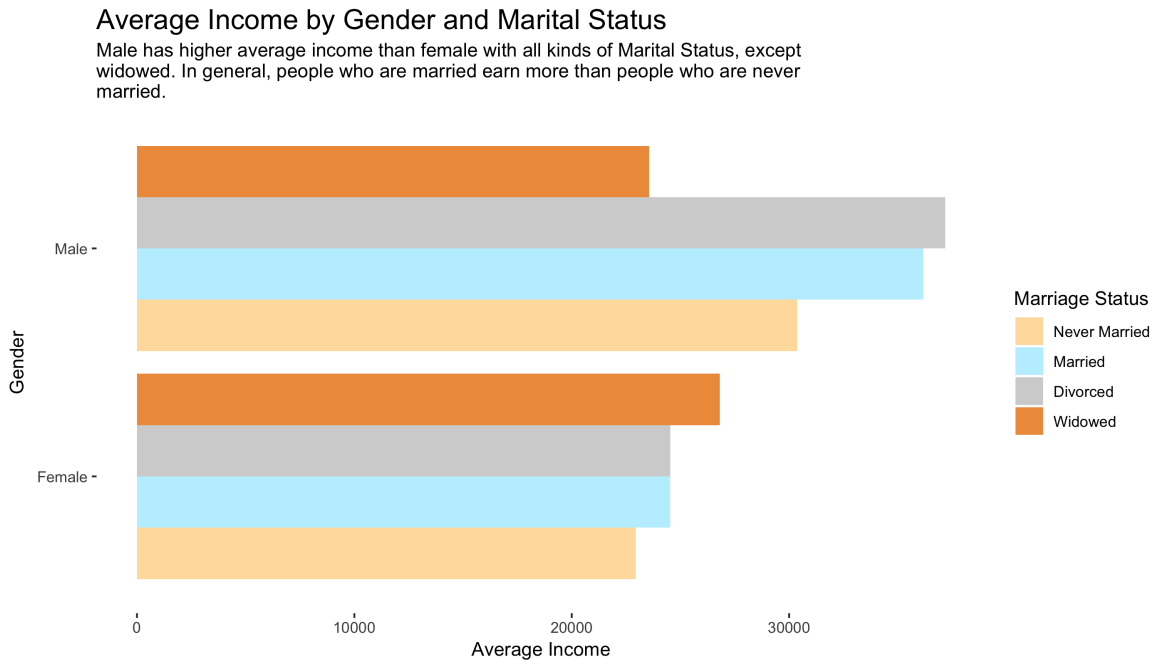


Figure 2 shows that Male has higher average income than female with all kinds of Marital Status, except widowed. In general, people who are married earn more than people who are never married. The difference of income between married men and never married men is bigger than the difference between married women and never married women.

- Figure3 Income by area Faceted by Gender

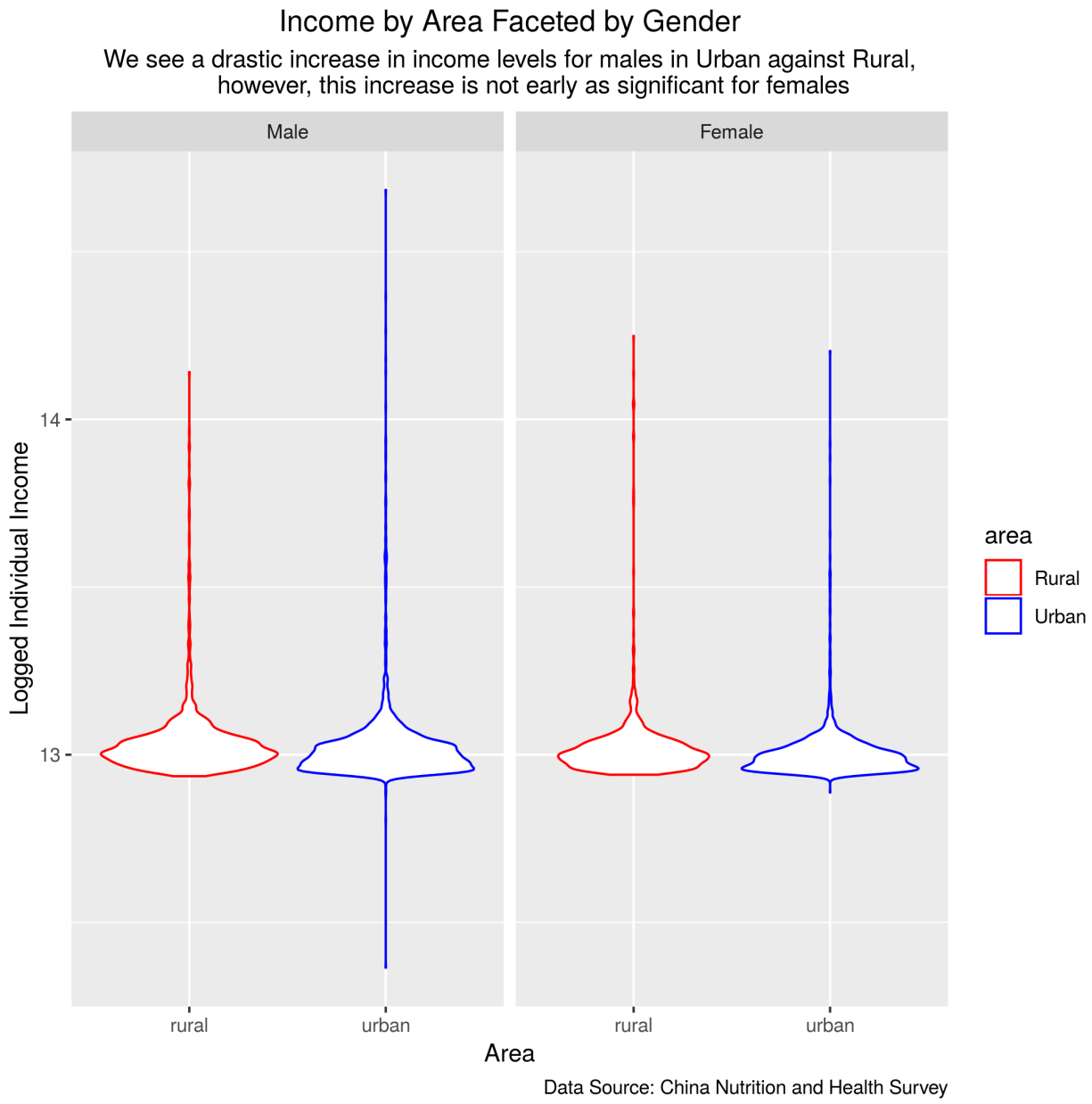


Figure 3 shows the logged income of both men and women from urban and rural areas. There is a drastic increase in income of men in urban area against urual area.

Also, we looked at income of both men and women after childbirth.There is a slight decrease in income after childbirth. However, how it impacts each gender is slightly different.(Figure in appendix)

2. Empirical Results

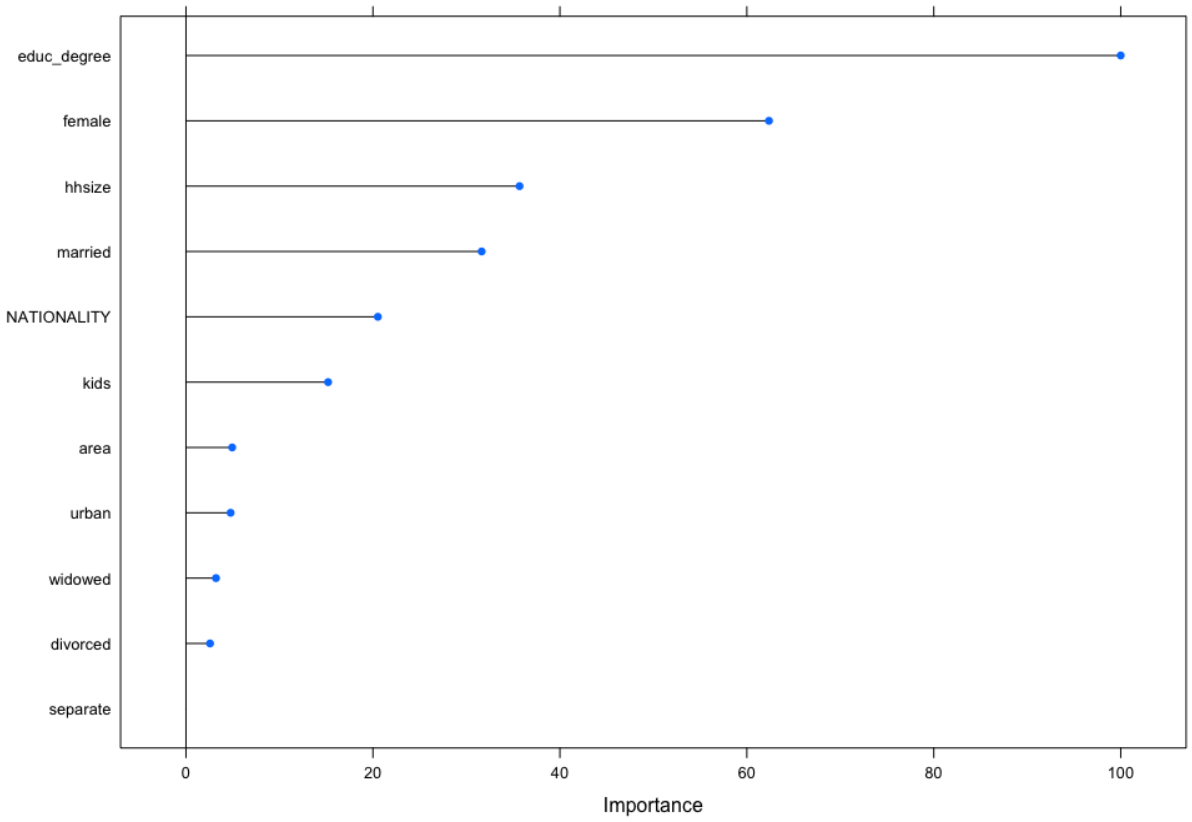
In order to examine whether there is a gender gap in labor force participation in China, we use Logistic Regression to explore whether gender will be a significant indicator of a person's employment status. In table 1, we can see that if a person is female, she is 41% less likely to get employed than male, holding all else constant. Besides, education is another significant indicator of a person's employment status. If a person's education level increases by one degree, his/her chance of getting employed will be 18.6% higher. Thus, it is clear that Chinese women, on average, experience more difficulties in labor market. Below is logit Regressions and Odds Ratio.

Table 1. Logit Regressions and Odds Ratio

	Dependent variable:	*Currently Working?*
	Coef.	Odds Ratio
Female	-0.522***	0.5930892
	(0.033)	
Education Degree	0.170***	1.1858870
	(0.013)	
Urban	-0.136***	0.8725755
	(0.036)	
Married	0.212***	1.2365713
	(0.035)	
Divorced	0.219	1.2443945
	(0.158)	
Widowed	0.052	1.0538148
	(0.186)	
Separate	-0.231	0.7933630
	(0.523)	
Household Size	-0.009	0.9912036
	(0.010)	
Children	0.134***	1.1428660
	(0.039)	
Constant	0.623***	1.8650587
	(0.061)	
Observations	18,284	18,284
Log Likelihood	-10,971.800	-10,971.800
Akaike Inf. Crit.	21,963.600	21,963.600
Note:		*p<0.1; **p<0.05; ***p<0.01

3. Machine Learning

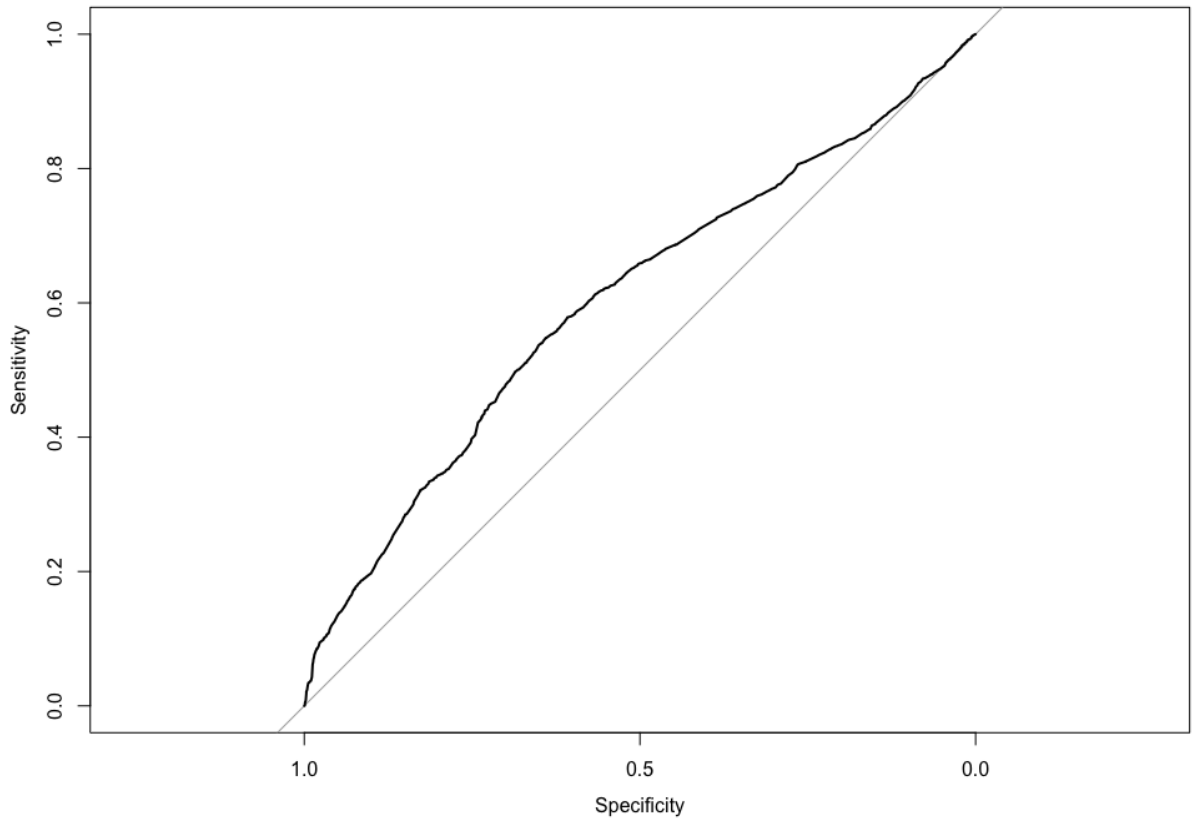
Once we determine that there is a gender gap in labor force participation in China, we use supervised machine learning to explore what are the important indicators of employment status. The methods include K-Nearest, Classification and Regression Trees, and Random Forest. After we compare the output of all three models, Random Forest stands out as the model has the best predict power. Figure 3 shows the importance of variables indicating employment status in China. It is clear that education level and gender are two important indicators of whether a person is working, which is consistent with our findings from using Logistic Regression.



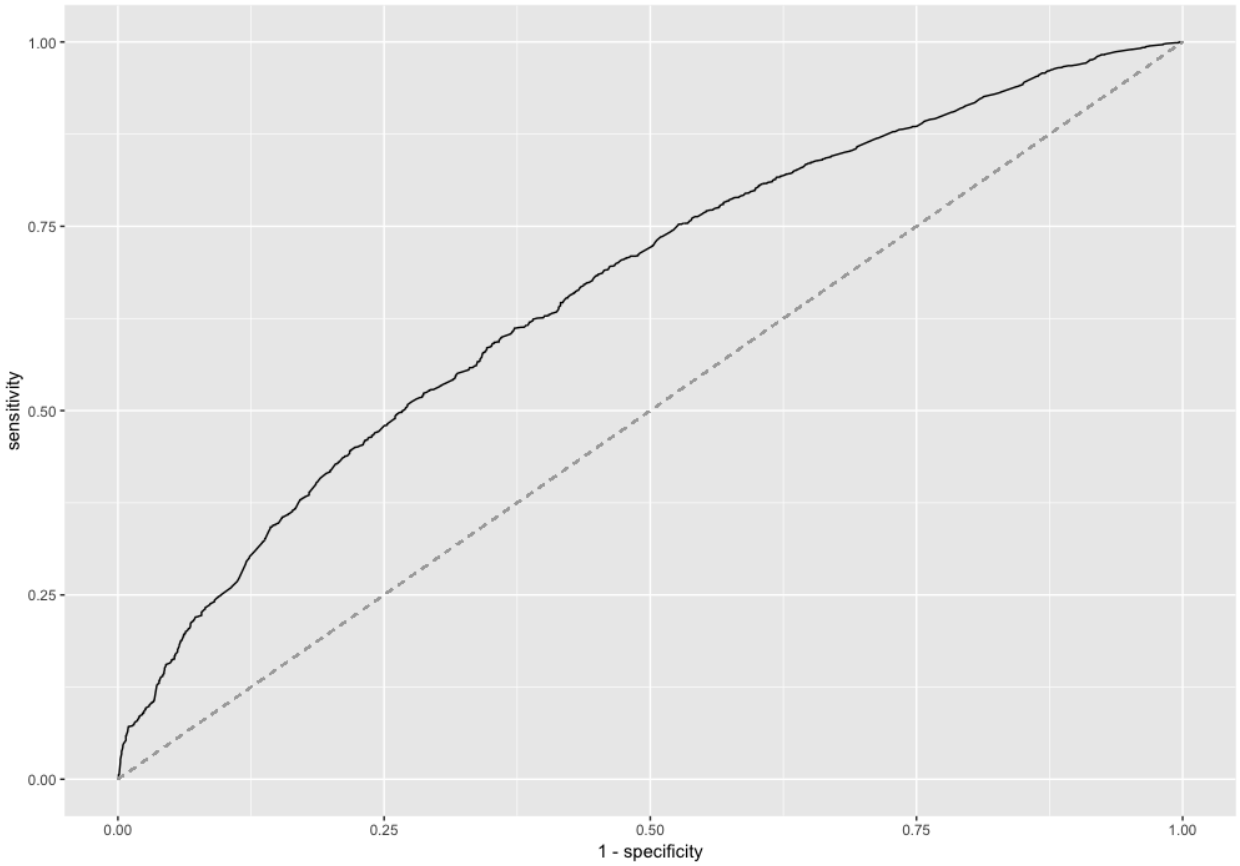
Since we have logistic model and random forest model, we are wondering which model has better performance. Therefore, we used ROC curve to assess the accuracy of measurements for predicting a binary outcome. Below are ROC curves of two models. As we can see from two curves, Random forest model has better performance. Thus, we could say that Random

forest is a better model for our study.

Logistic Model



Random Forest



Conclusions and Policy Implications

Given the results, we find that education level and gender are two significant indicators of a person's employment status in China. As a person's education level increases, his/her chance of getting education is higher. Moreover, being a female in China serves as a disadvantage in labor market.

In the light of our findings, we recommend: Introducing a parental leave policy that equally applies to men. In modern China, gender identity is still a statistically significant factor that negatively influences Chinese women's labor force participation and earnings, because once women make the fertility decision, they have to take parental leave, thus incurring a higher operational cost on their employers. Introducing a parental leave policy that equally applies to men can equalize the operational costs of hiring males and females.

Introducing an education policy reform that can reduce cost to women of receiving education. Our empirical results show that education is a strong indicator of whether a person can get employed. If we can reduce the cost of women receiving education, women can get access to education more easily and correspondingly find jobs more easily.

As a society develops to a stage where farming is no longer a major source of income for most families, and jobs increasingly require mental labor instead of physical labor, women will become as competitive as men in labor market. Thus, reforms in parental leave policy and education policy are necessary, not just to address the issue of gender gap in workplace, but also to promote the rights that women deserve.

Appendix

Figure 1

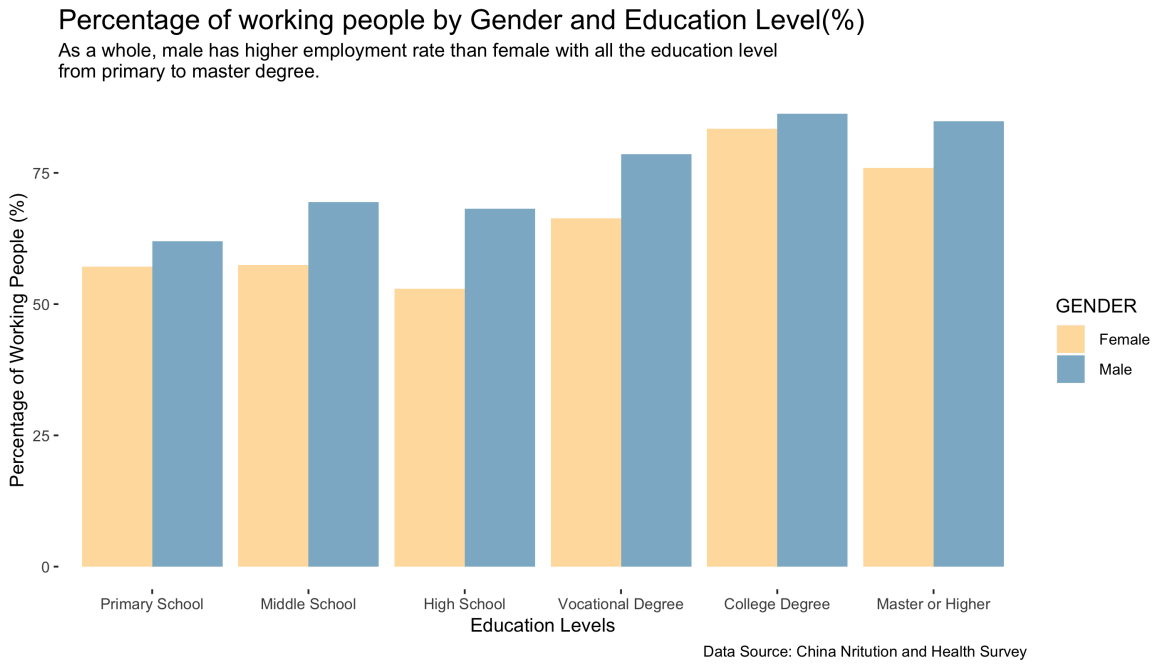
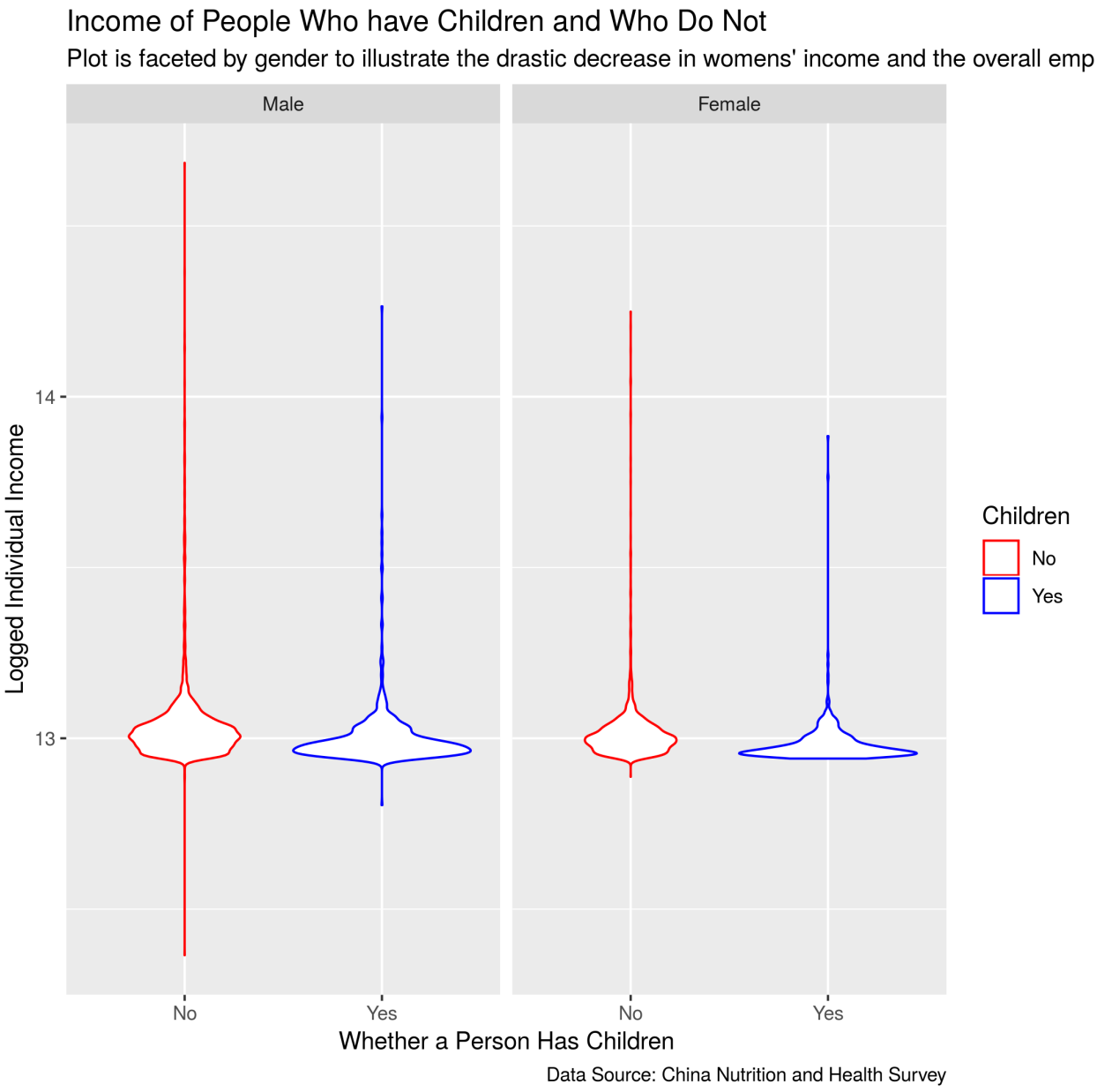


Figure 2


Income of People Who have Children and Who Do Not
Plot is faceted by gender to illustrate the drastic decrease in womens' income and the overall emp
Male
Female
Logged Individual Income
14
13
No
Yes
No
Yes
Whether a Person Has Children
Children
No
Yes
Data Source: China Nutrition and Health Survey