Research on the Main Factors Affecting Income and Its Implications

Xiaoqing Zheng, Yuang Fan, Qimin Li, Kairui Wang

Abstract—**Inflation in the United States continues to soar. On the one hand, the wealth of the top 1% of high-income people has increased significantly; on the other hand, the problem of homeless people has become increasingly rampant, and low-income families bear serious financial pressure.The problem of income disparity is growing in the richest country in the world [1]. Based on the comprehensive analysis of more than 40000 pieces of data and 15 attributes, this paper systematically investigated the influencing factors and mechanisms behind the income gap. Education promotes the value of human capital and makes income premium; Marriage enhances people's sense of responsibility and has a positive impact on income; Together with the continuation of inequality effects of traditional variables such as race and gender, they have distorted the pattern of income and wealth distribution in the United States.**

I. Introduction

THE issue of income disparity and its influencing factors are not only a hot topic of widespread concern in all walks of life, but also a key topic of research for policy makers and scholars, and it has an important impact on economic development and social stability. Since the 1970s, the widening gap between the rich and the poor and the increasing polarization in the United States have brought a series of social problems, especially the subprime mortgage crisis in 2007, which made people pay more attention to the household debts of low-income people and the factors affecting their income [2]. A scientific measurement of the income distribution gap and an analysis of the causes of this gap can provide a scientific basis for the government to formulate reasonable income distribution policies, rectify the order of distribution and reform the financial and taxation system, and at the same time provide reference suggestions for individuals to make career choices, improve income and quality of life, and enhance personal happiness, which is of great theoretical and practical significance to promote social harmony and stability and sustainable economic development.

At present, the research on factors affecting income is relatively scattered, and most generally mention that the increase of income gap is the result of the joint action of many factors, among which there is probably no "key factor" [3]. There is no quantitative conclusion on the impact of the increase of income gap.

On the basis of reading a lot of theories that affect income, this paper starts with analyzing the personal factors that affect income gap, focusing on the analysis of four key factors that have a greater impact on income gap: education, marriage, race and gender, and expounds their impact on income gap, trying to enlighten the social level in exploring effective tax systems and welfare policies.

A. Data Set

This income data set comes from the Kaggle platform and contains more than 40000 pieces of data which have 15 attributes, such as the income of each person, people's marriage status, people's gender, people's race etc. In summary, this data set contains many aspects of people and their income.

B. Focused research questions

The focused research question is the relationship between different factors in this data set and income, mainly including education, gender, race,and marital status, as well as the deep reason behind the relationship.

II. Background/Related Work

In fact, many other researchers have done research on the impact of various factors on income.Some predecessors also put forward many widely accepted theories.

A. The effect of education on income

There are many theories that examine the relationship between education and income. For example, the signal model of higher education believes that education can effectively reflect people's ability, and workers with more education are more capable of working than those without much education [4]. Going to college is just one way to showcase their talents and ability to work. In other words, going to college is an important way to gain recognition from others. As long as they are highly talented, they can do well in their jobs in the future. The level of education of a worker can indicate his ability to work to a certain extent, so most workers with higher levels of education earn more than those with lower levels of education [4]. As another example, the human capital theory holds that investment in education is the core of human capital, and it is a productive investment that improves the ability hidden in the human body [5]. Education can improve people's knowledge and skills, increase productivity, and thus increase personal income [5]. Another theory is the labor market segmentation theory. It considers that the labor market is divided into labor market and non-labor market. In the main labor market, the education level of the labor force is relatively high, and the proportional relationship between education level and wage level is basically established [6]. But in the secondary labor market, education and wage levels are not necessarily positively correlated [6]. People with higher education have a good chance of entering the main labor market, and the higher the education level in the main labor market, the higher the wage level [6].

B. The effect of marriage on income

What many people don't know is that marriage also has a big impact on income. A research project once investigated many large enterprises and found that after removing the influence of age, the average salary of married employees is a part higher. For the same number of years of service, married employees are more likely to be promoted than unmarried employees. Researchers have proposed several different explanations for this phenomenon. First of all, being married means greater responsibility and financial pressure, and employees will work harder.In other words, married people tend to climb the job ladder faster[7]. At the same time, married people are more likely to give people the impression of being stable and reliable, and they will have more opportunities in the workplace. In addition, maybe being married makes men more effective, since women still do most of the housework so men can focus more on their careers [7].

C. The effect of race on income

Although the United States abolished slavery very early, and the United States of this era is also very pursuing equality and political correctness, income inequality between races still exists.

Racial income inequality has a lot to do with history. Although there is no racial segregation system now, most blacks in the United States are still concentrated in poor areas, whether rural or urban. In the United States, most funding for KM schools comes from state governments. State governments distribute school funding unequally, but according to local government taxes. The wealthier the area, the higher the taxes, the more public schools, the more funding, and the better the quality of education. Taxes are low in poverty-stricken areas, public schools have little funding, and schools without funding have a poor environment and cannot hire good teachers. No wonder so many students are not getting a good education. If you don't study hard, you won't find a good job and you won't make money. Residents, local governments and schools remain impoverished, forming a vicious circle.

D. The effect of gender on income

Right now, regardless of the survey, the median income of men is slightly higher than that of women.

Part of the gender pay gap is that women are more likely to take breaks in their careers to have children or to seek lower-paying positions that offer greater flexibility and make household management easier. Factors that can affect the gender pay gap include unconscious bias and discrimination against women. Many women said they had experienced sexism at work, while not many men said they had experienced sexism at work. Secondly, the survey found that the income of men and women is not equal. Very few women say they earn less than men in the same job, and very few men say they earn less than women in the same job. Women with children also earn less than men with children or women without children.
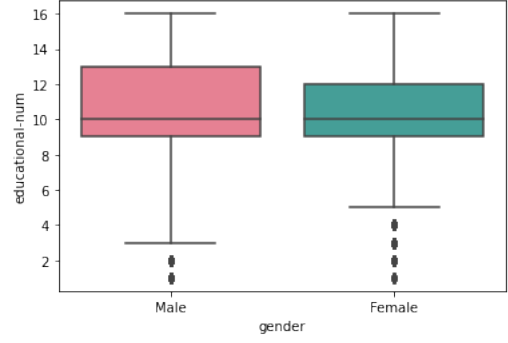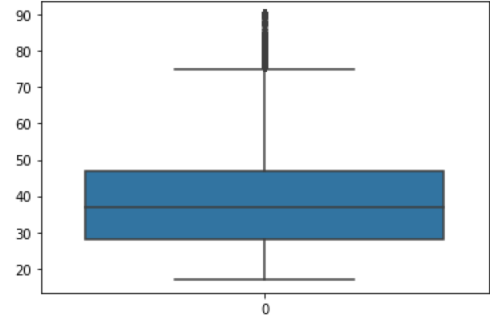

Fig. 1. Education vs Gender

Fig. 2. Age distribution

III. DATA CLEANING

Now that we have the data, this is the stage where we need to examine it to determine if it is a usable data set and what type of work we need to put into it. The condition of the data set will most likely make or break the visualization. The main tasks during the process of data examination are to check the data types, missing values, size of the data set, ranges of values, trustworthiness and integrity, and the completeness of the data set.

There are 43957 rows and 15 columns in the raw data set but there are also some missing values. After examination, as shown in Table 1, there are 2498 missing values (5.6%) in the "workclass" column, 2506 missing values (5.7%) in the "occupation" column, and 763 missing values (1.7%) in the "native-country" column. Because the missing values account for only a small portion in the data set, we can simply drop them.

Then we check the distribution of some important numerical variables in the data set, such as age(Fig. 2.), educational-num(Fig. 3.), and hours-per-week(Fig. 4.). The overall distributions of these variables are reasonable and no noticeable exceptions and outliers are identified. As for the completeness and integrity of the data set, after deleting the rows that contain missing values, all the values are in reasonable ranges. Besides, there are 15 factors, like different workclass, educational background and genders, in the data set that may influence income, which is comprehensive.

After the examination and data cleaning process, the final size of the data set is 40727 rows, 15 columns.

Columns	total missing	ratio of missing
workclass	2498	0.056828
occupation	2506	0.057010
native-country	763	0.017358

TABLE I
MISSING VALUES
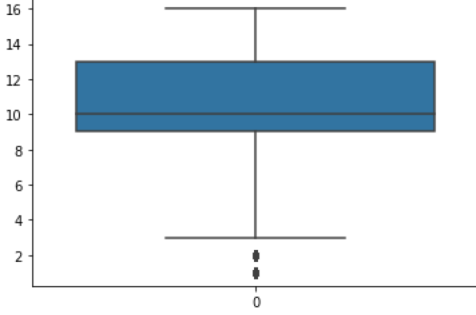
Fig. 3. Education distribution

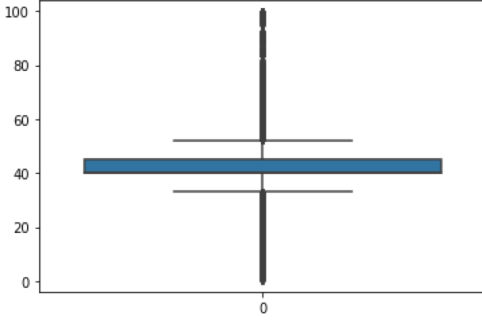Fig. 5. Education vs Gender

Fig. 4. Working hours distribution

Fig. 6. Education vs Gender Race

IV. DESCRIPTIVE STATISTICS

First, we should get an overview of the data set. We calculate the mean and median of some numerical variables. The median age of the interviewees is 37, which means that most of the interviewees are at a stage where they are married or older single people. This is an age when most of the people live a stable life so the result of these data would be more general and of greater reference value. Besides, the mean and median of the educational-num are 10.1 and 10. Combined with the distribution of the educational-num, this data set is mainly about people with a medium and high education level.

Then we calculate a correlation matrix of all the numerical variables in the whole data set to find out if there are any related variables. The correlation matrix is a useful tool to check the relationship between variables in the data set, which is a foundation to explore the data later. From the matrix, it can be seen that there is a relatively high correlation coefficient between "income>50k" and "educational-num", which indicates that a high education level may create high income and we will explore this further. From the correlation matrix, we know that whether income is higher than 50k is related to people's educational-num. From this point of view, we may expect a relatively high average income from the data set.
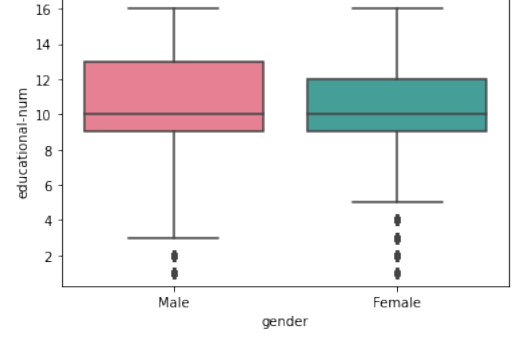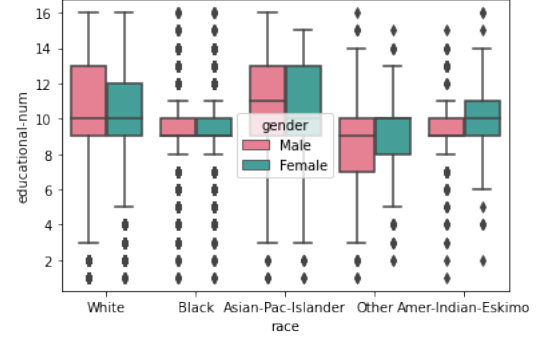
V. DATA VISUALIZATION

From Fig. 5., the educational level of male and female interviewees are overall the same. To be specific, the differences are that the minimum bound of the female is higher than the male's. The Q3 (75th percentile) of the male is higher than the female's. From the boxplot above, it can be concluded that there are fewer female interviewees whose educational level is very low. However, the ratio of the highly educated male interviewees is larger than the female ones. According to the correlation matrix, maybe we can expect a higher income among male interviewees.

Now, let's take race and gender together into consideration. From Fig. 6., it is easy to find out that the White's and Asian-Pac-Islander's general educational levels are higher than the other races. However, it may be resulted from the relatively small amounts of the data of these other races. If the dataset is too small, the biases are almost inevitable. Actually, in the United States, White and Asian-Pac-Islander have a larger chance to receive a high level of education, as displayed in the result of the boxplot. According to the sociologist of education and anti-racist activist Barry Troyna, society and the role of schooling perpetuates a system of racial inequalities in
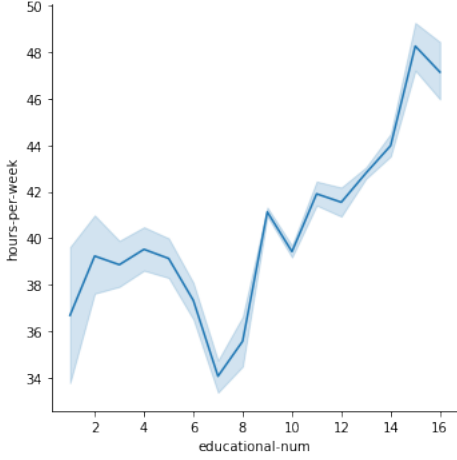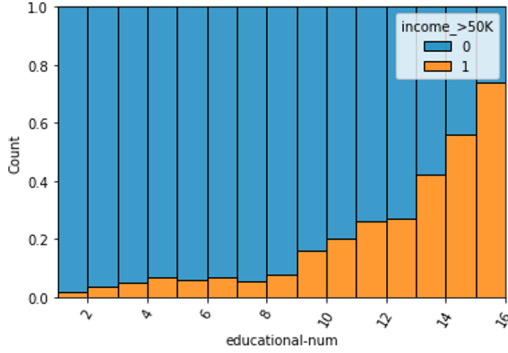
Fig. 7. Working hours vs Education

Fig. 8. Income vs Education

Fig. 9. Working hours vs Relationship

Fig. 10. Hours Income vs Relationship

education and beyond, which is a noteworthy problem so the complex and pervasive influence of racism in education which the contributors draw attention to necessitates both short and long term antiracist strategies [8].

As seen in the line chart Fig. 7., from educational-num 7, there is a strong positive correlation between hours-per-week and educational-num. This phenomenon indicates that the high income of highly educated people not only simply resulted from their advanced degree, but also their hard work. It is reasonable that along with the advancement of education, people will be more and more aware of the principle that labor creates value. Besides, as shown in the histogram Fig. 8., there is a strong positive correlation between educational-num. However, as we stated above, there is pervasive educational inequality in society. It is always harder for some races to get better education. However, it is also common sense that students from richer families have easier access to better education, which will create a vicious circle.

From the bar chart Fig. 9., it is interesting that when people get married, their working hours even increase. Besides, when people have children, their working hours dramatically decrease. The probable reason is that when they get married, people have more responsibility to support the family, so they work harder to relieve the financial burden. Then we check the income from 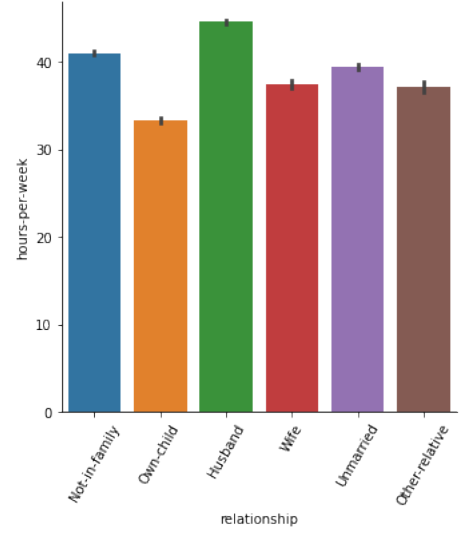Fig. 10., as stated above, the avera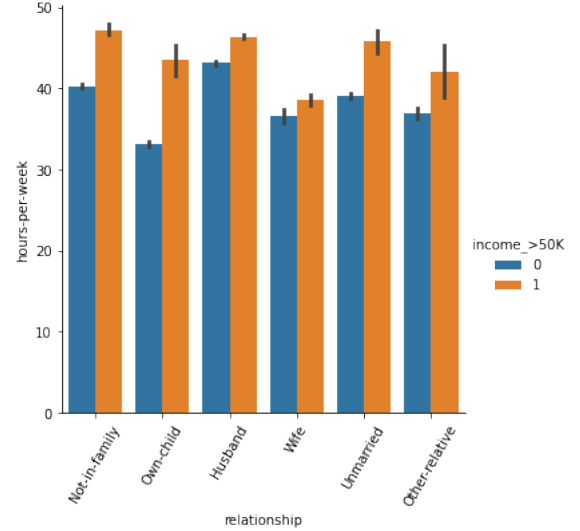ge salary of married employees was 12% higher. Married employees are 34% more likely to be promoted than unmarried employees for the same length of service. However, after giving birth to children, people should spend time with their children, which makes the working hours decrease.

This is another interesting phenomenon in Fig. 11. that couples without children have a far higher ratio among the people whose income is more than 50k. Combined with the former chart about the working hours per week, it is reasonable that couples work harder and stimulate each other to earn a higher salary.

VI. Initial Inferences

Through the data visualization, we have some interesting discoveries for the relationship between income and social factors. The first important thing is that the sample range of this income dataset did not pick randomly. In this dataset, 85%
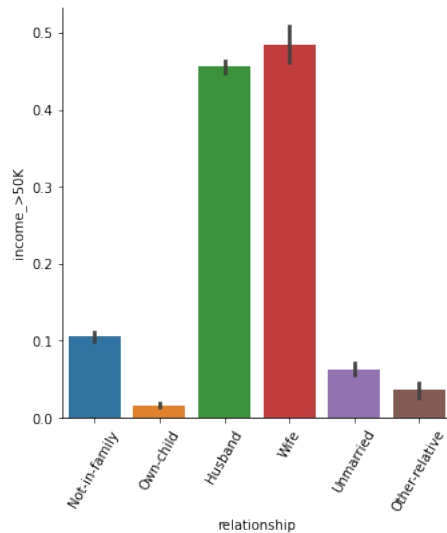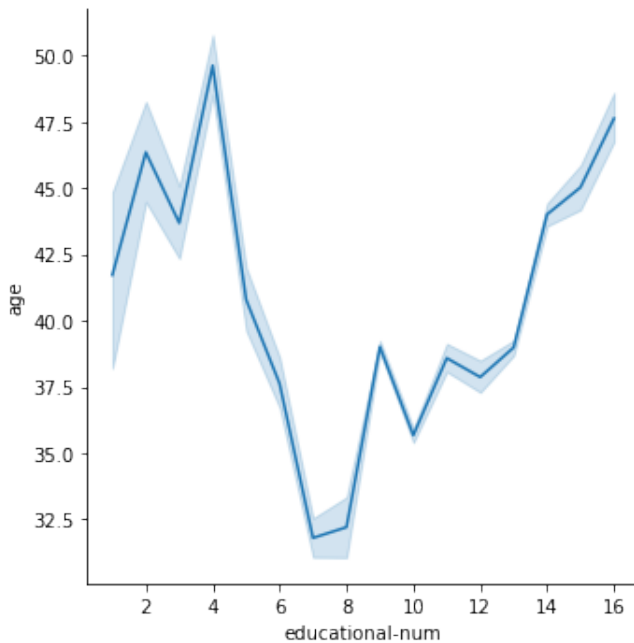
Fig. 11.   Income vs Relationship

Fig. 12.   Age vs Education level

from the Age vs Education level graph, the average age of people taken 1-8th years of school are ten more years higher than people in 11th and 12th grade. As the minimum age in the dataset is 17, that means there are fresh graduates of high school that counted in this dataset, and many elders who only received primary education. However, from looking through the average income of each educational level, it seems that the age of the low education people does not have much influence on their incomes. People who have the highest income are in educational levels of 1. Doctorate, 2. Professional school, 3. Masters. People in preschool level of education have the least ratio of earning more than 50k. As in common sense, income has a positive correlation with the educational level. People who receive higher education have more opportunities to get better jobs with higher salaries. That's why people are struggling to study at better schools. There are some people who believe education is not important. They joke that elementary school dropouts have become bosses after several decades of social life, while college students have to work for them after graduation. But through the analytic of the income dataset, only a few cases of people with low education have good jobs, the others do not get high wages.

From the graph Working Hours vs Income, we found that in most instances, people who work more hours per week have higher income, as people always said hard work is rewarded. However, combined with what we had before, although people with education levels of 11th and 12th grade work the relatively lower hours per week, their income is not so low compared to people with less education. That means even many people with low education have to work hard every day, they don't make a lot of money. Most of them can only work for the service occupation, like driver, waiter, and cleaner, which are tough but not profitable. People engaged in management and professional specialty are more likely to have higher income. One thing that surprises me is that the occupations in protection areas like armed-forces and protective service also have relatively high incomes.

Income is also related to gender and marital status as mentioned in the previous studies. Male have relatively higher salaries than female, not only because some gender discrimination and also females' life status after they have children. People work more hours after they get married, but the time decreases when they have children. This could happen because people have more pressure for living and preparation for raising children after marriage, but it will take a lot of time for parents to spend in the family when they have children. From the previous studies, we found that income is related to race, and our analysis from this dataset also shows white, Asian and pacific islander have relatively higher income, but as the selection of data is not comprehensive, and the classification of race is unspecific, we cannot get an accurate relationship between income and race from our analysis.

What's more, there is one thing worth mentioning: According to this dataset, males are having relatively higher educational levels than females, while white people and Asian-Pac-Islander have higher educational levels than the other races. This seems back to talk about the relationship between education and income, but it also comes to a more complicated

of the samples are white people, 10% black, and only 5% of others; The number of male is almost twice as many as the number of females; The range of sample's age is 17-90, and most are under 45. The non-random sample can make our analysis result inaccurate and even biased. According to this dataset, males are having relatively higher educational levels than females, while white people and Asian-Pac-Islander have relatively higher educational levels than other races.

According to the working hours vs education graph, it seems people with higher educational level work more hours per week as universally known, but people with educational level of 11th and 12th year of school work the lowest time every week. This is a surprise finding that people with the lowest educational level work more hours. However, as we can see

question: How does the gender and race discrimination affect on income? It seems that females and blacks' income are affected by their education level, but are the discrimination directly related to income? As June Ellenoff indicated that the income of blacks is surely lowered by discrimination, but "Jewish and Japanese Americans, for instance, have had incomes substantially above those of white non-Jewish groups, despite evidence of discrimination against them" [9]. Race discrimination may have different influences on different race. This is the question we need to think about in the later analysis.

Overall, income is related to many social factors like educational level, working hours, gender, races, and martial states. These conclusions are considered working for the most white people as the samples are not representative. There are still several aspects we need to study about income in the future work, like the influence of discrimination on income. Besides the environmental forces and normal human behaviors that we mentioned, Dale stated "public policy may exert a positive influence on reducing income inequality through economic policy that promotes economic growth" [10]. Policy and taxation also need to be taken into account when we study about income.

REFERENCES

[1] "The hilltop review," 2012. [Online]. Available: scholarworks.wmich.edu

[2] "International review of financial analysis," 2016. [Online]. Available: https://scholar.google.com/citations?user=I9Qmm9oAAAAJ&hl=zh-CN&oi=sra

[3] A. Fan, "The british journal of psychiatry," 2001. [Online]. Available: https://scholar.google.com/citations?user=DD3XK0EAAAAJ&hl=zh-CN&oi=sra

[4] M. Spence, "Job market signaling," 1973.

[5] S. Ross, "What is the human capital theory and how is it used?" March 2021. [Online]. Available: https://www.investopedia.com/ask/answers/032715/what-human-capital-and-how-it-used.asp

[6] D. G. M Reich and R. Edwards, "A Theory of Labor Market Segmentation," *The American Economic Review*, 1973.

[7] S. SKILLS, "Married people earn more than single individuals. here's why." January 2022. [Online]. Available: https://bigthink.com/smart-skills/married-wage-gap/

[8] "Racial inequality in education," 2012. [Online]. Available: https://books.google.com/books?id=DSD6ltdoKkAC&printsec=frontcover&hl=zh-CN#v=onepage&q&f=false

[9] J. E. O'Neill, "Discrimination and income inequality: Social philosophy and policy." January 2009. [Online]. Available: https://www.cambridge.org/core/journals/social-philosophy-and-policy/article/abs/discrimination-and-income-inequality/D9FF0F6142053E717001B2E8DAB85DD9#access-block.

[10] D. O. Cloninger, "What factors influence income inequality?" July 2016. [Online]. Available: https://theconversation.com/what-factors-influence-income-inequality-61837.