U.S Wages

(Data Visualization)

(Ibrahim Khan Jabarkhail)

2024-03-09

1 Downloading and storing the data

In this project, I have looked at the data related to wages from US for years 1976-1982. Firstly, I installed the Ecdat package and looked at the data for wages and then checked the dataset. It includes 4165 observations and 12 variables.

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

2 Task 1

Here, I will describe the data or the data set that I am analyzing.

```
## Warning: package 'knitr' was built under R version 4.3.2
```

Table 1: Summary Statistics of Wages Dataset

Total Observations	Total Variables	Average Wage	Median Wage
4165	15	882.9307	799.9986

Table 2: List of Variables in wages_formatted dataset

Variable	Description
exp	Experience of work in full time
wks	Weeks of work per year
bluecol	blue collar
ind	If an individual works in a manufacturing industry
south	If an individual lives in a metropolitan area
smsa	If an individual is living in south
married	Is the individual married

Variable	Description
sex	Male or female
union	Is the salary set by union
ed	The number of years of education
black	If an individual is black
lwage	log of monthly wage
id	Unique identifier for each individual
year	Which year it is
wage	wage of workers

Looking at the the dimensions, we can see that there are a total of 4165 observations and 15 are variables. The dataset is a pane of 595 observations from 1976 to 1982. The variables also include three new variables that we formed i.e., ID, Year and wages. Initially, the data set had 12 variables.

Next, I will draw the chart to look at the patterns of wages of individuals over the years 1976-1982.

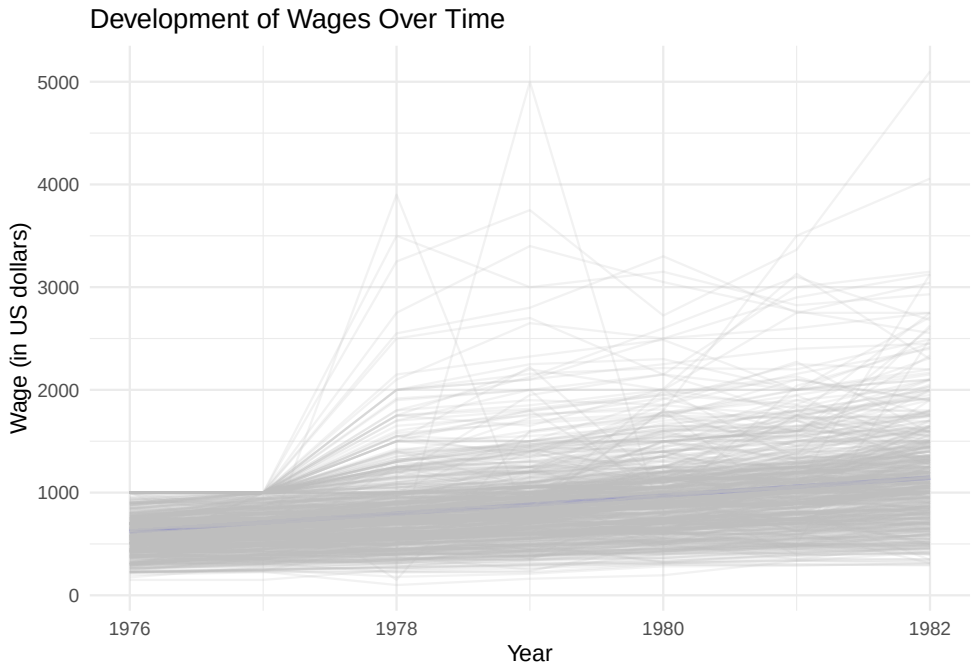

Figure 1: Distribution of wages over the years studied in the data set

The plot above demonstrates that the wages have been mostly flat and there are also few individuals whose where the wages have increased to their high levels for example the wages have reached up till USD 5000 over the course of the years for some individuals. From 1976 to around 1978, wages for the majority of individuals/entities appear relatively stable and clustered around the lower end of the wage spectrum. After 1978, there's increased variability, with some entities experiencing a sharp increase in wages, peaking around 1980. However, most remain at the lower wage range.

The observed fluctuations in the wages are very few compared to the amount of wages

that have remained flattened. However, there appears to be significant variability between different individuals, with some experiencing steep wage increases while others exhibit more stagnant wage trajectories.

2.1 Task 2

Here, I will filter the data first to evaluate the relationship between the two variables only for the year 1982. I can then check if the relationship between these two variables is positive or negative.

Now, I will plot the figure to see the relationship.

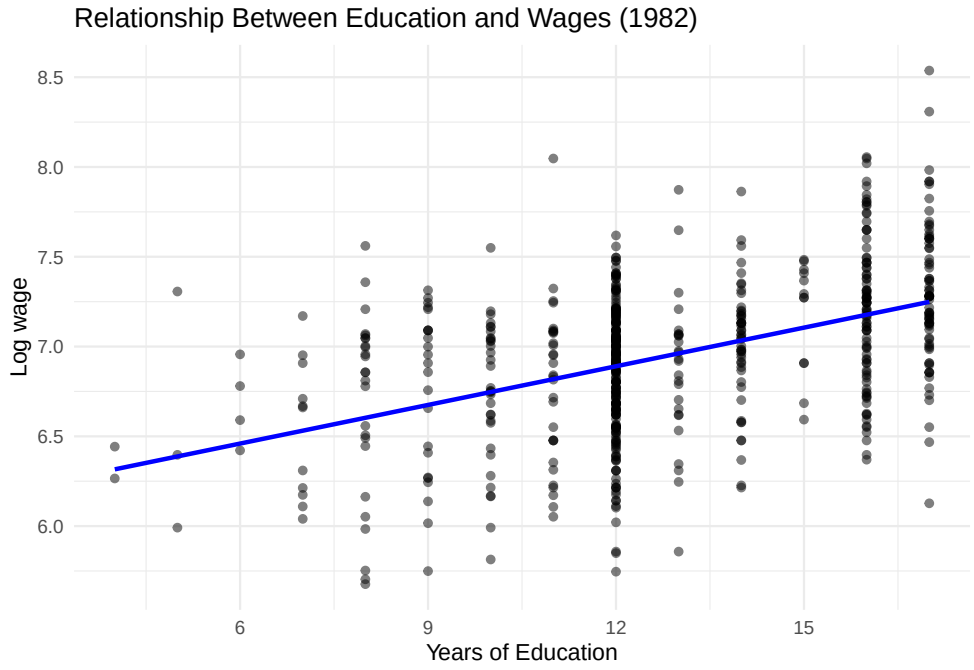

Figure 2: Relationship between wages and year of education

It is now evident from the graph that there is positive relationship between years of education and higher wages. In other words, looking at the graph we can say that with the increase in education levels, the wage of the workers also increased. There is linear relationship between the two variables for the year 1982. The stragiht line in the chart also indicate the positive relationship between years of education and the wage levels. One can easily that with for example 12 years of education results in a higher wage than with 6 years of education. It could also be the case that with the higher education people secure jobs in good industries or secure white collar jobs which make their wages higher.

3 Task 3

Here, I evaluate the relationship between wage and some of the other variables in the data set. Firstly, I created the new categorical variabele wage_levels which would

highlight if the wage for particular variables or category is low, average or high. Later, I comapred it with different categorical variables.

Now, here I will assess the relationship between wage levels and one of the categorical variables in the data set. I have checked the wage_levels against different categorical variables like married/non-married.

Figure 3: Difference in wage levels among male and female

Among women, the graph reveals that mostly females have low level of wages, whereas only a margin of men have higher wages. The middle wage is also higher for men than women.

Figure 4 also shows that the wage levels of the married ones are higher than the ones who are not married in all categories of the wages. This could also mean that the number of individuals who are married is higher than the ones who are not married.

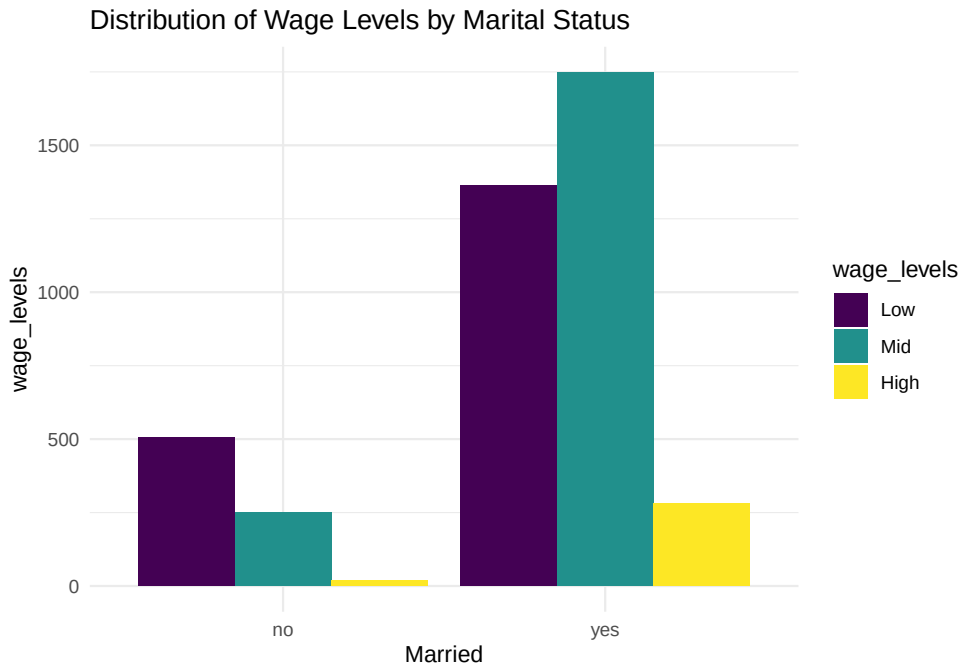

Figure 4: Difference in wage levels among the individuals who are married/unmarried