# Assignment III

Data Visualization

Benjámin Olivér Maday

2023-10-12

## 1 Task 1

In this assignment I will be working with the Wages dataset from the Ecdat package. The panel data consists of 4165 observations, meaning it contains socio-economic information about 595 different individuals from the United States from 1976 up until 1982. There are 15 different variables, which I would like to describe in the following table:

Table 1: Description of variables in the mutated Wages dataset

| Variable description | Data type |
|---|---|
| Years of full-time work experience | integer |
| Weeks worked | integer |
| Whether a person is a blue collar worker | categorical |
| Whether they work in the manufacturing industry | categorical |
| Whether they reside in the south | categorical |
| Whether they reside in a metropolitan area | categorical |
| The individuals marital status | categorical |
| The individuals sex | categorical |
| Is their wage determined by union contract | categorical |
| How many years did they spend in education | integer |
| Whether the individual is black or not | categorical |
| The logarithm of the wage variable | continuous |
| The identification number of a given individual | integer |
| The year of the observation | integer |
| The monthly wage of a person in USD | continuous |

We are interested in the connections between the different variables and their effect on a given person's income. The first thing we should consider is time. If we take a look at Figure 1, we can see that even though there is a great variance on the level of the individual (sometimes they have a huge income, but the next year the FED realizes that aforementioned income came from money laundering), the overall tendency suggests that, at least nominally, wages steadily increased in this period. There are a few individual increases which stuck for the remainder of the time period, potentially meaning a promotion or an industry change.

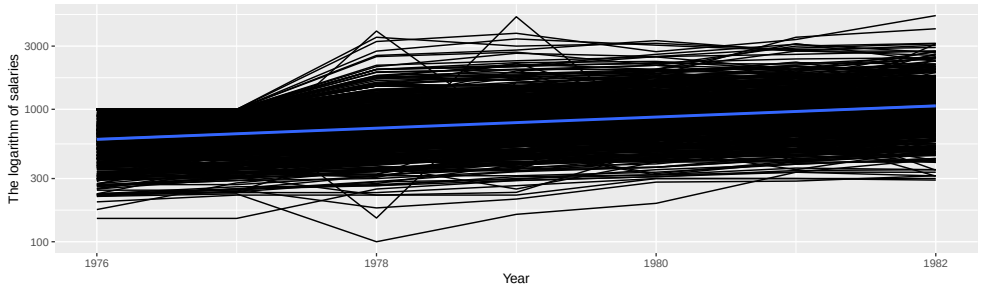


Figure 1: The evolution of wages in the US between 1976 and 1982

## 2 Task 2

Next we should take a look at the connection between education and income. Looking at Figure 2 we can determine that there is a positive correlation between the two variables, however I would advise the reader to take this information with a pinch of salt. More years spent in education does not always mean a higher level of education, nor does it mean a better paying profession (we can take the example of teachers and social workers and compare them to masons or carpenters, especially in the US). The fact that an overwhelming number of observed individuals had the same amount of schooling, and the lack of data from people with less than 12 years of education should also make us cautious when drawing conclusions.

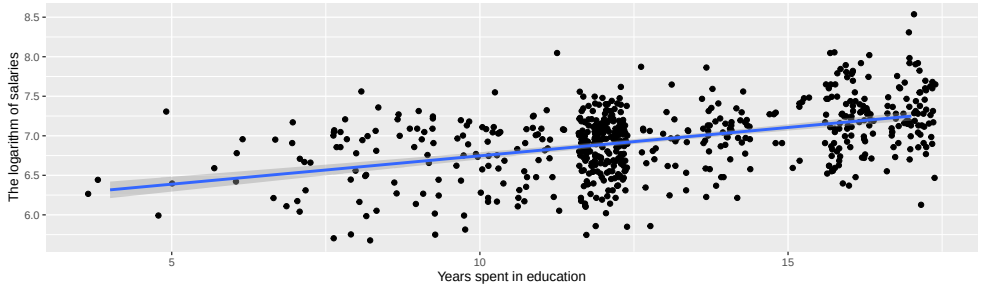


Figure 2: Figure 2. Salaries mapped to the years of education

## 3 Task 1

Lastly, I would like to investigate the potential relationship between the geographical location and wages. As we have way less southerners in our dataset I decided to create a mosaic plot mapping the relationship between the region and the income category of individuals. It is clearly visible, that my prejudgments where just, the southern parts labour composition does rely more heavily on low income workers.

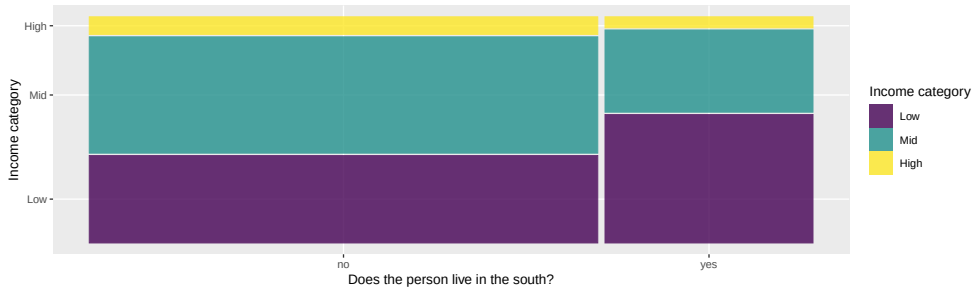


Figure 3: Figure 3. Mosaic