

# Assignment1-Report

## 1. Table Introduction:

My research data is from table1. My goal is to find the most pivotal property for education attainment distribution. Based on this. I checked the table's structure: Both of them are separated by gender inside. Furthermore, based on the table's name, they can be separated by married persons or single persons. In this way, there are three important attributes, gender, race, marital status.

## 2. Technology

In order to make the data more conclusive, I used the data group technology.

The education is grouped into three parts: under high school(UN), between high school and bachelor(BE), up to the bachelor(UP).

The age is grouped into four parts. 18-25, 25-34, 35-55, over 55

Another technology is pandas. I used the pandas library to process the excel data.

## 3. Summary

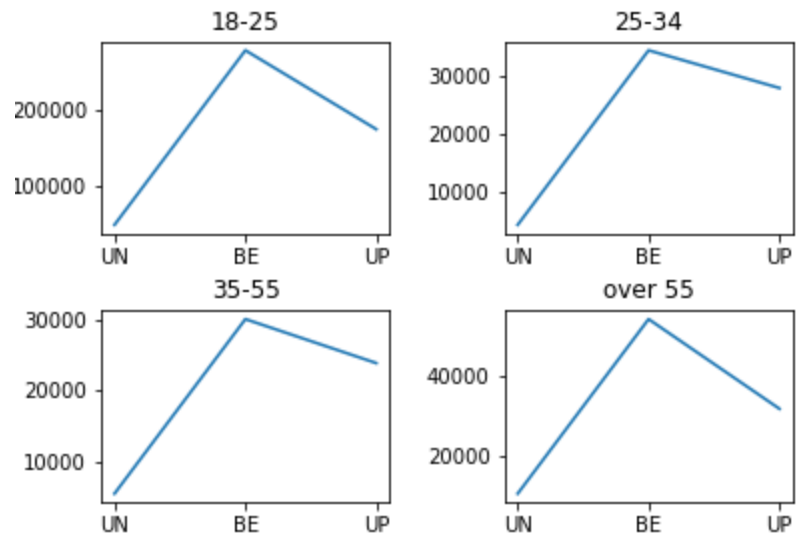
I mainly focused on AllRaces, White related, Asian related, Black related tables.

I first make the gender as a variable, combination and race as constant, then find the Inside the same table, male and female have the same line chart structure.

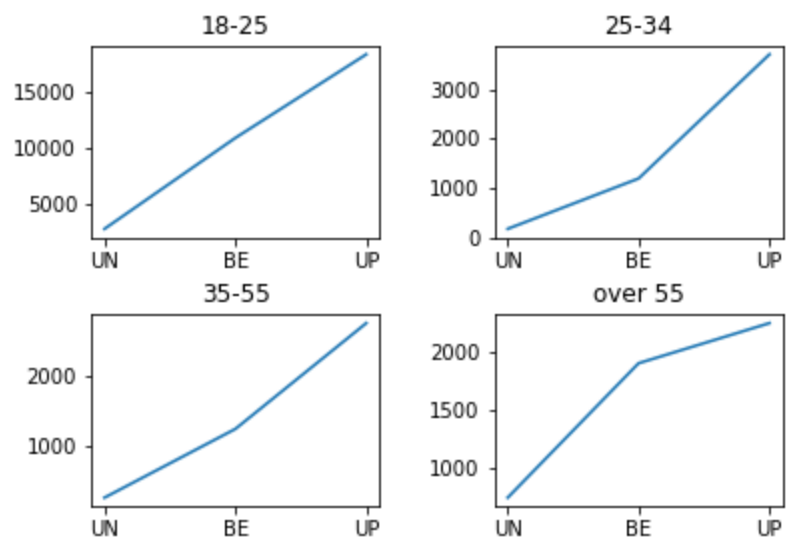
And then I make the combination as a variable, gender and race as the constant. Also, line chart under same combinations are the same.

Finally, I make the race as a variable, gender, and combination constant. I found the line chart looks very different under different races.

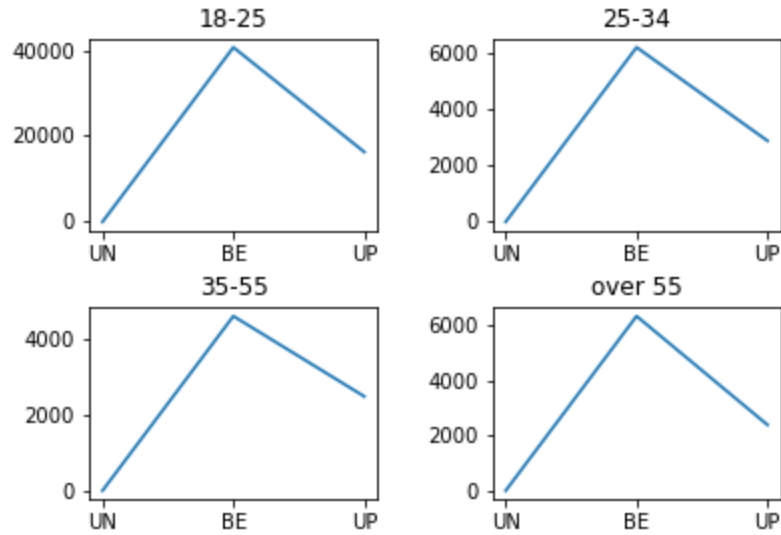
### All Race:



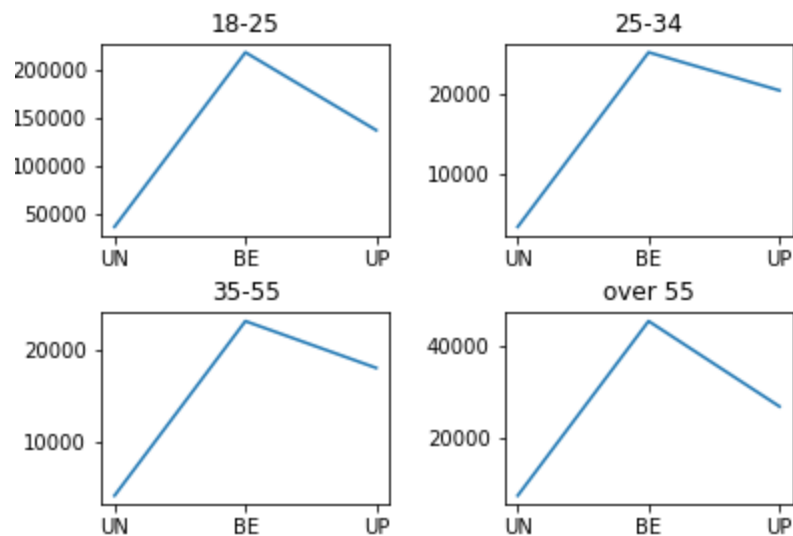
## Asian:



## Black



## White



As we can see, the distribution of asian is very different from white and black. Under every age range, Bachelor degree and above accounted for the highest proportion; For the white, In the age range of 25-34 and 35-55, "BE" and "UP" have almost the same amount. For black, "BE" always has the highest amount of people under different age ranges. In other words, the distribution of black is closer to Allraces people distribution.

As a result, compared to combination and gender, race has a greater impact on education.

#### **4. what I studied:**

My first study is that if there are many rows and columns, we need to group them to make it easy to check the line chart towards.

Another study is that if there are many uncertain properties and hard to find which one is more important, trying to only make one as variable, others as constants. In this way, we can easily find the most important property.