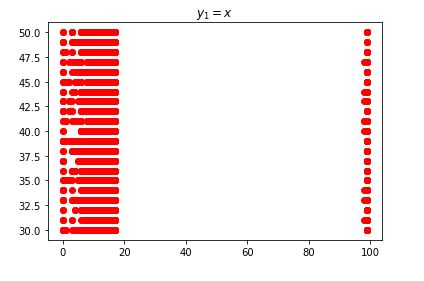
**Detail Report**

The Panel Study of Income Dynamics (PSID) dataset contain data on 4856 individuals. Information on these individuals represent data covering age, education, earnings, hours, no of kids they have and their marital status. The question we are trying to answer is the impact of an individual’s education level to his/her earning capacity.

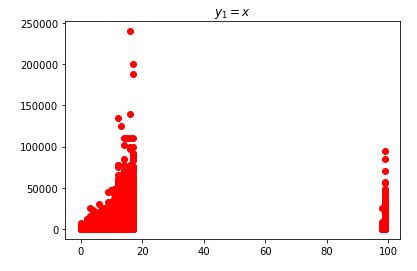
1. **Experiments**

As the first step, we analyzed the PSID dataset and painted them on different graphs.

* X axis – education level , Y –axis – Age

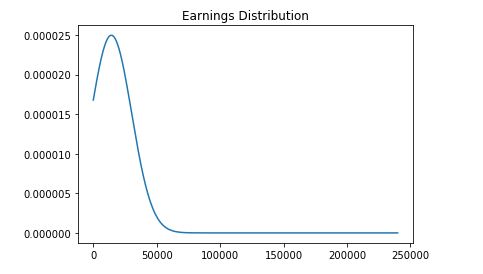


* X axis – education level , Y –axis – earnings

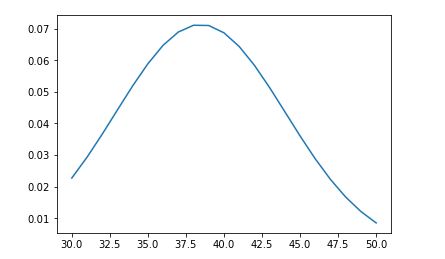


We can see that the second graph shows a linear regression if we omit the outliers.

We plotted the linear regression for Earnings distribution and age distribution.



This is left skewed because of outliers. Below is the age distribution.



1. **Outputs**

Age distribution shows a clear normal distribution for the whole population while earnings distribution is left skewed more due to the outliers. Since age shows a clear normal distribution, we took two samples (sample size = 100) and plotted.