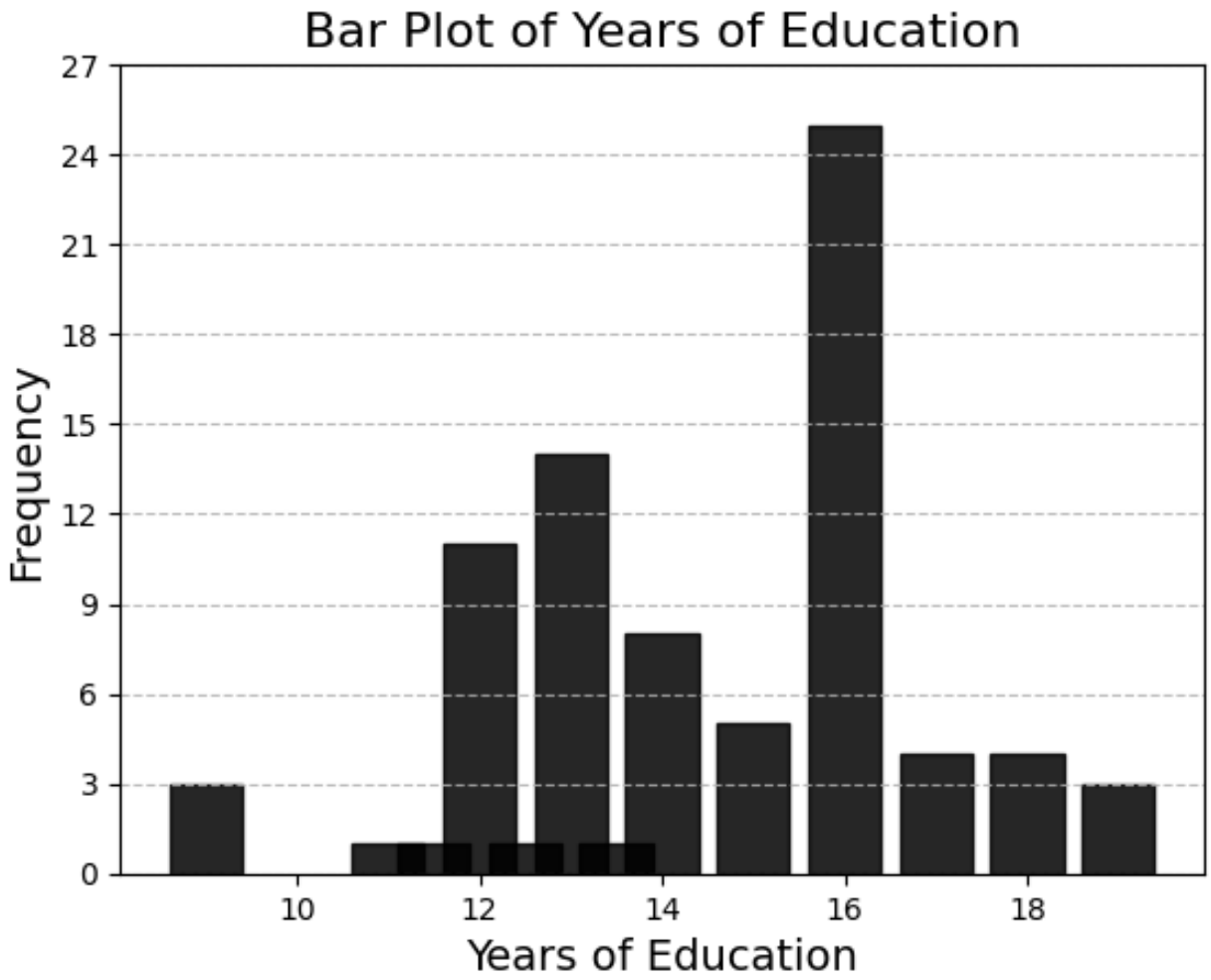
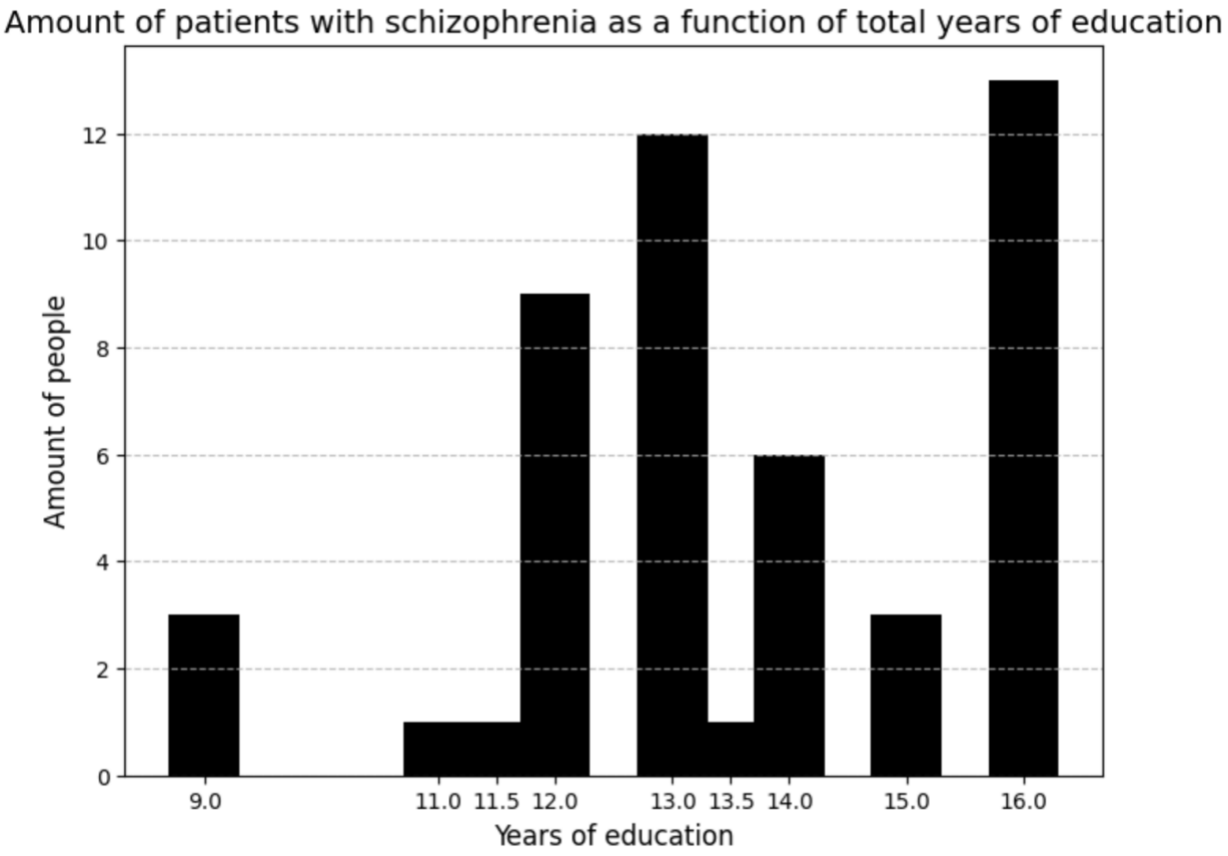
Miya Estis 343478575

Carmel Mandelblit 212373096

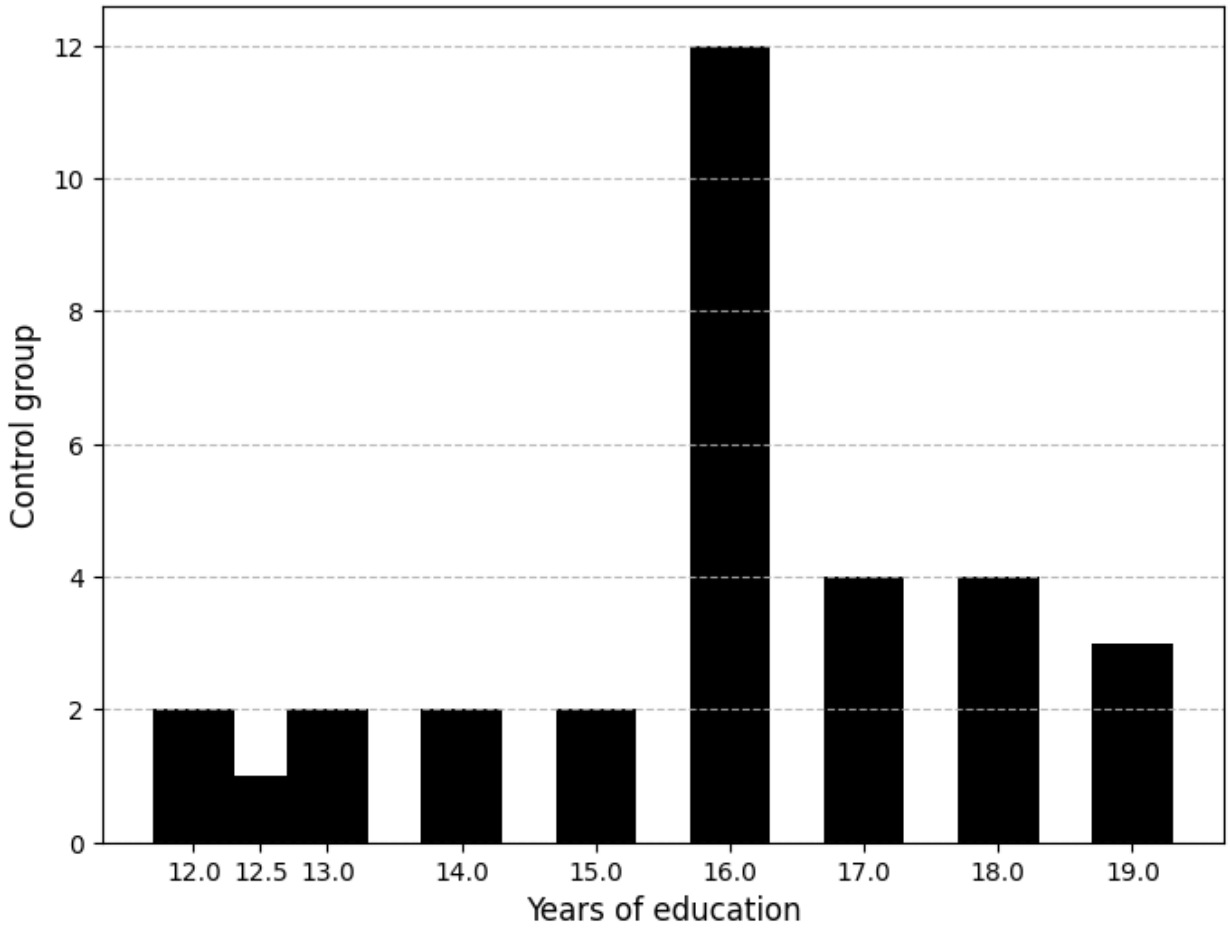
The dataset we found https://www.kaggle.com/datasets/broach/button-tone-sz/data?select=demographic.csv was about EEG data findings took from basic sensory task in people who have schizophrenia (total 81 subjects, 32 control group, 49 had schizophrenia). The data set was well structured and very big (20.8 GB). After making a quick search we found a csv file called 'demographic' that included information such as gender, age, education etc. Suddenly we came up with a hypothesis about the negative correlation between the years of education and having the disease.



The first plot was created to illustrate the distibution of education years and the total amount of participants.



The second plot provides you information about a subgroup of participants with shcizophrenia.

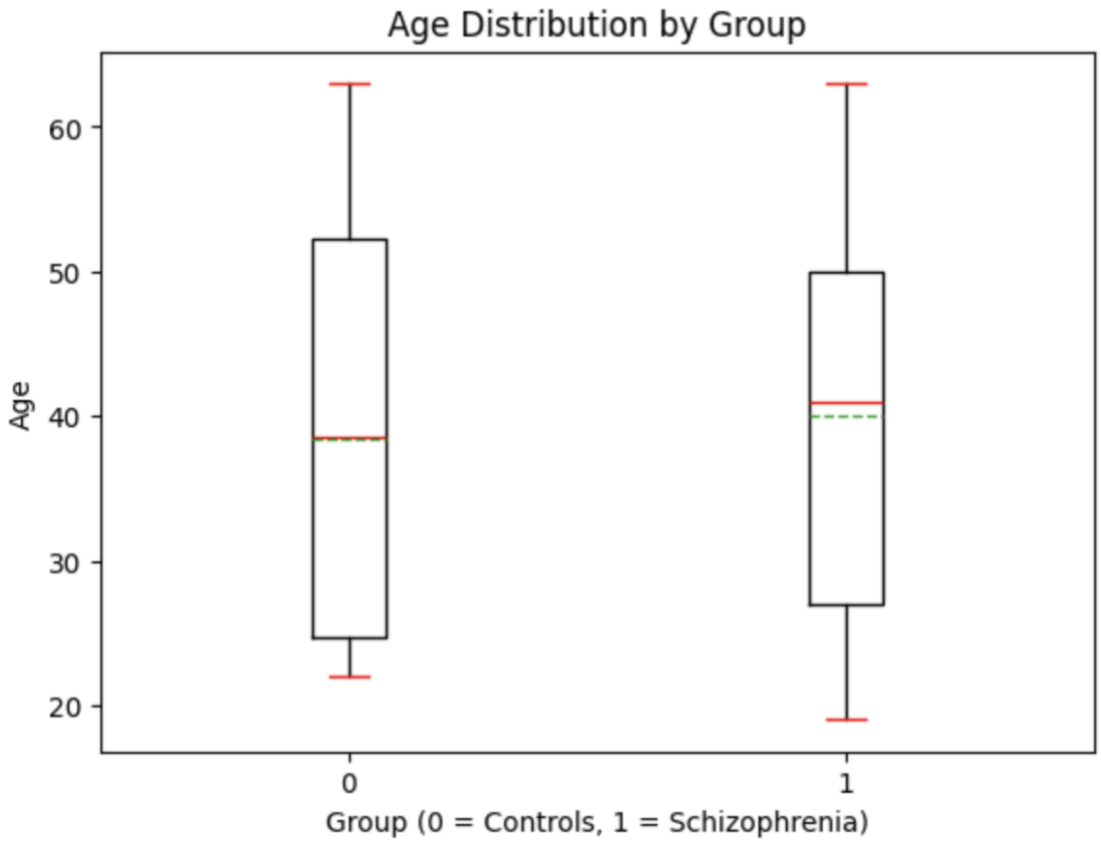


The third plot was provided to show the difference in the years of education in control group. The comparison shows us a clear answer that there is no any correlation between the education level and the disease. Moreover, in our case we are able to notice that people with schizophrenia more educated than people from the control group.

Conclusion – the hypothesis is disprooved.

Right after we refuted our hypothesis we came up with the other one. What if there will be a correlation between the age and patients?

```
Correlation between age and schizophrenia: 0.06
<Figure size 800x600 with 0 Axes>
```



We made a boxplot that represents the interquartile range, which contains the middle 50% of the data. It checks the age distibution between groups. The visual comparison and correlation coefficient show us a very weak relationship between controls and patients.