This report focuses on the data collected in the 1994 US Census. The specific data in the dataframe were used by researchers as a part of an experiment to try and create a model which predicts whether certain people would earn more or less than $50,000 USD based on characteristics such as their race, sex, highest level of education attained, and hours worked per week.

One of the variables in the dataset is called 'education_num,' and it is a numerical variable which takes an integer value from 1 to 16, which corresponds to the highest level of education which the person has completed. In this data, a higher education_num corresponds to a higher level of education. For instance, an education_num of 16 occurs when the person has a doctorate, whereas a 1 represents completion of preschool. One possible research question which we could study through this data would be "are older people more likely to be highly educated?"

My hypothesis for this research question was that older people would tend to be more highly educated, because people who are older would simply have more time to be able to seek out higher education. For example, someone who is 40 years old is more likely to have a higher level of education for multiple reasons. First, they have simply had more time available to seek out education, in case they want to pursue further education later in life. Second, it would be very unlikely for anyone with a Masters Degree or doctorate to be young, since these degrees take time to earn.

To find results for our research question, we first can compute the correlation between the variables age and education_num, to see if there was any significant correlation between these variables. There potentially could be some correlation between these variables, The results show that the correlation between these two variables is approximately 0.0365. Correlation is measured on a scale from -1 to 1, but the strength of the correlation between the variables does not look at the 'sign,' and is measured from 0 to 1, where a lower value means the variables are less correlated. Thus, there is a very weak positive correlation between these variables, but it is far from convincing.

We can split these observations by sex, and separately compute the correlation between education_num and age for males and females, to see if this pattern is consistent across the sexes. We find that the correlation between education_num and age for males is 0.06, and the correlation coefficient is -0.018 for females. The stronger positive correlation for males is to be expected based on our hypothesis that age would be positively correlated with education_num, but the result for females is counterintuitive.

A possible explanation for this phenomenon could be societal expectations of women. Since the data is from 1994, many of the older women in the dataset would have grown up or been adults in the 1940s through 1960s, a time period where it was less common in the United States for women to seek higher education, as women would tend to be relegated to familial duties such as housework, cooking, and cleaning, instead of having a job. These societal norms prevented many women from seeking out higher education, and as such, we find that older women had lower values of education_num, because the societal pressure on women to have families and stay at home instead of getting a degree and a job outweighs the potential which they had to seek out education as they grew older.

One thing to note is that there are often other barriers to education which could have affected people in the data, such as the price of going to school, and the opportunity cost of not working or having reduced work hours as a student. Because of these barriers, education_num does not necessarily reflect the highest level of education which these people wanted, as there could be some people who graduated high school but were unable to study in a post-secondary institution due to external factors such as cost, familial affairs, or even societal pressure to do something else.

In this exploration of the data, one message rings clear, which is that women typically were not given the same opportunities as men to go to school. The fact that age is positively correlated with education for men and negatively correlated with education for women implies that older generations have a massive gender gap with regards to education, and that women have historically lacked the educational opportunities which men have always had available to them.