The exploratory data analysis of the NSFG dataset provided some useful correlations to several variables. Some of the variables I used was agepreg which shows the age at the time of pregancy, oldwantr which shows the readiness of the mother during her pregancy, ager was the age at the time of the interview, hieduc was the highest achieved education at the time of pregancy, and birthwgt_lb which is the weight of the baby at the time of birth.

The initial hypothesis I was focusing on is the correlation between hieduc and agepreg and poverty and birthwgt_lb. The dataset was normally distributed as confirmed by the data points in close proximity with the Q Q line. The Shapiro Wilk test was another test conducted to see if the dataset was normally distributed.

The outcome of the exploratory data analysis shows that it has a moderate to strong linear relationship on both sets of variables that were tested.

Some of the variables that could have helped in the analysis was comparing readiness of the father which we could have compared with the readiness of the mother to see if there is a relationship with the two variables.

An assumption that I thought was true was the more educated the mother is the older they would have the pregnancy. That is not always the case as there are several other factors that could have affected the age in which the mother was pregnant.

Some of the challenges I faced was modifying the dataset to filter out some variables to have two datasets with different variables. An example would be filtering out the dataset into two using the poverty level variable to see if the poverty level influenced the age of conception.