DATA WORK:

I decided to work with data on engineers (specifically combining mechanical and civil engineers, codes 1460 and 1360 respectively). I created the dataframe, new variables, and inspected the data.

DISTRIBUTION OF WAGES FOR ENGINEERS (IRRESPECTIVE OF SEX)

Firstly, I wanted to see the distribution of wages for these engineers using bar charts(not focusing on the sexes). I found that the average wages per hour were about $35.

WAGES FOR ENGINEERS (FOCUSING ON SEX)- Regressions and Scatter Plots

-I then changed the data to find which of the engineers were female. I found that about 12.5% of the engineers were female.

-I then looked at the gender gap between male and female wages for the occupations. I took a level-level regression and discovered it didn't fit the data too well (R-squared was 0.002 or the regression line fit 0.2% of the data). I then took a look at the data by modeling the regression via a scatter plot. From the slope I saw that female mechanical/civil engineers earn 2% less, on average, than male engineers. We can be 95% confident that the average difference between hourly earnings of female mechanical/civil engineers versus male was -5% to 1%. The CI includes 0 which means we can't say with 95% confidence that their average earnings (between males and females) are the not same. The t-stat is 1.2 which is less than 1.96 so we can't reject the H0 hypothesis. The p-value is also 0.2 which is much larger than 0.05. The coefficient can't be considered statistically significant at 5%.

-I then took a log-level regression and got a really weird reading/summary with mostly 'nan' values. I then visualized the regression with another scatter plot to see why. I found that the data looked somewhat odd and tended not to fit the regression line at all for the log-level regression.

WAGES FOR ENGINEERS v. HIGHEST LEVEL OF EDUCATION (FOCUSING ON SEX as well)- Regression and Scatter Plot

-I then looked at the wage gender gap regarding level of education. I visualized a scatter plot with a regression line. I then looked at the summary of my regression and found that an increase in grade level meant 3.3 more dollars in wages; however, if you're a woman, it also means 2.7 dollars less in wages per education level increase. (Could perhaps look at this like for a woman, an increase in education level means slope(grade92)-slope(female) or 0.53 dollar increase in wages for every increase in education level. The p value for sex in regards to wages and education level is 0.090 which is more than 0.05 so we most likely fail to reject the null that there is a difference in wages for education level/sex for these engineers. We can be 95% confident that the average difference between hourly earnings of female mechanical/civil engineers based on education level versus male is -5.9% to 0.4%. The CI includes 0 which means we can't say with 95% confidence that their average earnings are not the same. The R-squared value is 0.088 meaning that the regression line counts for 8.8% of the data- which isn't bad.