There has been much written about how income increases at a greater rate with increasing education. There also has been much written about how females earn less than males even when they have the same level of education. For this project, we will use data from the General Social Survey consisting of a random sample of 570 cases of those who graduated high school and who are within the working age demographic of between 22 and 55 years old. The data may be found in the file **gss_projectV3.dta**. In addition to investigating how income increases with education overall, we also will investigate whether the increase in income with education is different for males and females.

The data are from the General Social Survey, which, since 1972, has been a reliable source of information to help monitor and explain trends in American behaviors, demographics, and opinions.

***Question #1**
Generate a new variable, weekly_wage, from income by dividing income by 52
gen weekly_wage = income/52

***Question #2**
Compute the mean and standard deviation of female. The following command is useful:
tabstat (female), stats (n, mean, min, max) What does the mean of female represent?

Using the same command **tabstat** (weekly_wage educ), find in addition to the mean and sd, the p25, p50, p75, skew, min, max for weekly_wage and educ.

***Question #3**
Compute the mean, sd, min, and max for educ and weekly_wage separately for males and females, and interpret results. Use the command **tabstat** on weekly_wage and educ with the appropriate **if** added to each analysis.

***Question #4**
Use boxplots to compare the female and male distributions of educ and weekly_wage. In addition to shape, make comparisons in terms of the five summary statistics depicted in the boxplot. For example, for educ, you may use the command
graph box educ, over(female)

***Question #5**
*Compute all pairwise correlations between the three variables in this analysis

***Question #6**
*Interpret these correlations.

***Question #7**
*Construct a scatterplot of educ on female with the regression line superimposed
graph twoway (scatter educ female) (lfit educ female), ytitle (Education) ///
xtitle (Gender) title(Education vs. Gender)

***Question #8**
Why does it make sense for the slope of this regression line to be reasonably flat?

***Question #9**
Construct a scatterplot of weekly_wage on education, superimposing the regression line on the plot and assigning a title and labels on y- and x-axes analogous to what was done in #7

***Question #10**
Obtain the equation of the regression line depicted in #9 and write it out.

***Question #11**
Interpret the equation obtained in #10

***Question #12**
Is the relationship between weekly wages & educ different for males and females? Regress weekly_wage on education separately for females and males to find out.

Write out the two equations, one for females and the other for males, and interpret each.

***Question #13**
Obtain a graphic of these regressions on the same set of coordinate axes to visualize the difference in the relationship between income and education for males and females. You may use:

```
graph twoway (scatter weekly_wage educ if female == 0, ms(+)) ///
             (lfit weekly_wage educ if female == 0) ///
             (scatter weekly_wage educ if female == 1, ms(X)) ///
             (lfit weekly_wage educ if female == 1), ///
              legend(row(1) order(2 4) label(2 "male +") label(4 "female X")) ///
              ytitle(Weekly Wage)
```

***Question #14**
Use these regression lines to estimate the mean weekly-wage for females & also for males for educ = 12 (hs degree), 16 (college degree), & 20 (grad schl).

Are the differences in estimated weekly-wage for males and females what you would have expected given the two regression lines and the graphic describing them? Explain.

***Question #15**
Generate a new dichotomous variable, college_degree, which equals 1 if educ >=16 and equals 0 if educ < 16

```
generate college_degree = 0
replace college_degree = 1 if educ >= 16 & educ != .
```

***Question #16**
Use the command tabulate to construct a frequency distribution of this new variable

***Question #17**
Test whether males and females who have a college degree differ in terms of their mean weekly_wage. Carry out Levene's test, but note why, in this case, the results of it are not a concern. (Hint: are the sample sizes close to equal?) Interpret your results of the following t-test, ignoring the results of Levene's test.

```
ttest weekly_wage if college_degree == 1, by (female)
```

***Question #18**
Test whether males and females who do NOT have a college degree differ in terms of their mean weekly_wage and interpret your results. Use the same approach as in #17.

***Question #19**
Write up a paragraph to summarize your main results as they relate to the motivating reasons given in the first paragraph for carrying out these analyses.

All figures and tables should be placed at the end of your write-up and referred to in the main body of the write-up (e.g., as shown in Table 1 or Figure 1)as you answer question by question. Be sure to include your do-file for this project in a second appendix.