**Independent T-test**

This file is designed to act as a walkthrough in how to manage, analyze, and visualize a dataset for comparison between two groups. The dataframe is a publically available dataset from AddHealth's longitudinal study in Adolescent Development. For this exercise we will be using Wave 4's data to test the hypothesis that males weigh more than females. The data has already been subset to include only the key variables, and it is included in this repository.

When conducting a t-test there are a few statistical assumptions that must be considered.

1. The Dependent Variable is Continuous and the Independent Variable is Categorical
2. There is no relationship or overlap between the Independent Variable categories
3. Dependent Variable is normally distributed
4. Variances across groups are roughly equivalent

In our dataset, the first statistical assumption is met because we are looking at how changes in biological sex (a factor with two recorded levels) leads to changes in weight. Since biological sex was recorded as either "Male" or "Female" without the possibility of multiple classification we can also say that the second statistical assumption has been met. Though it could be argued that sex preferences between genders lead to weight not truly being independent in this context, we do not have data to support that claim and therefor can neither argue for it or against it.

The third requirement can be checked using the descriptives function. Liberal cut-offs for Skew and kurtosis are -3.00 to 3.00 and -10.00 to 10.00 respectively. If you'd like something a little more conservative, Curran West and Finch (1996) recommend -2.00 to 2.00 and -7.00 to 7.00. You can also visually verify this by looking at the histograms and seeing if they look like they are roughly centered at the highest point with symmetrical decreases from there out (like a bell). If you find that skew or kurtosis exceed these values, you'll need to remove univariate outliers to make sure they aren't pulling your data.

For comparison of variance between two groups, we will use Levene's test of equality of variance (Levene, 1960). If the p-value (pr>F) is less than 0.05, this assumption is violated and you'll need to add Welch's correction for more valid results.

It is also important to verify that your data is clean at this point. In this study that means removing the NA's, removing outliers (which you probably already did), and ensuring that you have relatively equal groups for comparison. If the groups are unequal, you can use random sampling with replacement to equalize them, or consider using a Mann-Whitney U test for non-parametric samples.

Both the parametric and non-parametric test demonstrate that there is a real difference between males and females in weight (p < .001) with males weighing more than females (M2-M1 = 28 or 29 depending on which one you're looking at).

To aide in explaining this, the code for converting this data to a simple bar graph is included using standard error to demonstrate the confidence interval on the means.