The data set was provided in a nearly cleaned manner. However, multiple data wrangling techniques were performed on the data frame to transom it into a document that can be analyzed. These techniques include removing unnecessary variables, changing the names of some columns, adding variables, and including functions for calculation.

**Removing Unnecessary Variables:** The original data frame included 19 variables. Some of the variables were not necessary to perform propensity score matching as they did not provide relevant information for the matching algorithm. Removing the variables reduced the data frame to 10 variables. Some of the variables are still irrelevant but it is necessary to map the data back to specific observations. Therefore, the data frame does include more variables that will be used in the analysis.

**Changing Columns Names:** There are ten variables included in the data frame, and all variables underwent a name change. When the data was imported into R Studio, some variables names were too long, and others did not accurately describe the particular data. Therefore, the names of the variables were changed to shorter, concise names that accurately describe the data in the column.

**Adding Variables and Data:** Two variables were necessary for building the model. A treatment column and a column that displayed standardized z scores for the final grade earned. In the treatment column, an if-else statement was written to place a 1 for the observation that was treated and a 0 otherwise. This information helped to accurately and easily determine the control and treatment groups.

**Standardized z-score:** In the standardized z-score column, a function was written to include a standardized score for each observation in the “Score” column of the data frame. This standardized z-score normalizes the data and has a mean of 0 and a standard deviation of 1. It represents the signed fractional number of standard deviations by which the value of an observation or data point lies above or below the mean value of the data set that is measured. Values above the mean have positive standard scores, while values below the mean have negative standard scores. Adding this column increased the data frame to 11 variables.