The original data frame consisted of census data and consequently included numerous variables. I was able to utilize several data wrangling techniques to transform it into a document that can be easily analyzed. These techniques include removing unnecessary variables, changing “NA” to “0”, removing “NA” for numerical data, and changing the names of some columns.

Removing Unnecessary Variables: The original data frame included 46 variables. Many of these variables were not necessary for the classification study that I endeavor to undertake. Removing the variables reduced the data frame to 10 variables. Presently, all of the variables included should be relevant to the capstone project.

Changing “NA” to “0”, Removing “NA” for Numerical Data: In order to have the data more actively reflect the characteristics of the specific variables, I opted to change the “NA” assigned to several variables to “0”. This made the data more easily understandable as variables such as birthplace could more easily be identified, established. The designation of “NA” was also removed for all numerical data but was maintained for categorical data.

Changing Columns Names: Of the ten variables that are included in the data frame, several of them 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 more accurately describe the data in the column.