Elections in the United States of America are determined by a number of factors. Voters are primarily swayed by their identities, beliefs, and contexts. Since these issues are associated to their individual demographic classifications, one can use a regression model to predict the outcome of the 2020 federal elections. For this purpose, we came up with a logistic regression model using the Nationscape dataset as a sample population. This dataset describes a number of demographic factors in the United States which affect the political affiliations of voters. The dataset enabled us to develop an accurate model to predict the outcome. The logistic model was preferred since we aimed at getting a binary output. The output would simply state whether President Trump would win the election or lose, presumably to Joe Biden.

## Model

After downloading and unzipping the files, the dataset was cleaned to eliminate such rows as the voters who were not registered since they would not be allowed to vote. The cleaning process also assigned numeric values to the educational level, the political ideals, to facilitate the fitting of the model. The cleaning process was conducted in order to reduce the dataset to only include the essential information. This would facilitate the creation of a model from the data since the factors being used were converted into numbers. The independent variables chosen for the model included age, household income, educational level, political ideals, and ethnicity. The formula below was used in the regression.

Given the current state of affairs, the matters of race and ethnicity are bound to reappear once again like they have for a number of elections in the recent past. While political ideals and the racial or ethnic heritage of a voter have an obvious relationship with their choice at the elections, age and educational level are less obvious but their effects are greatly pronounced in the 2016 election and in the feedback obtained from the survey. The household income determines a person’s choice in voting since it reflects a person’s satisfaction in the status quo. The model can be summarised as follows.

The model was then fitted to the census data using the following code in order to predict the outcome of the elections.

##