Data Wrangling Summary

Upon initial inspection of the data, it is seen that the last 32 observations have missing values for the variables of county name and the percentage share of the Republican and Democratic vote share in both the 2012 and 2016 presidential elections. Likely, this data is missing since these were counties with low populations where reporting the vote share may constitute an undue invasion of privacy. This hypothesis is feasible given that 29 of these 32 missing observations are in the state of Alaska, a very low population density state. In fact, the state of Alaska does not have a single complete observation in the data set. Given the fact that none of the observations for Alaska had the dependent variable of interest and one of the key explanatory variables, the decision was made to delete these 32 missing observations from the data set. Otherwise all the column names and values were organized and sensible, and no further data wrangling was needed.