# Predictive Algorithm for Congressional Districts Voting

I will be attempting to form a predictive algorithm that will predict the likelihood of a congressional district voting either Republican or Democrat. This could be used by political parties to gauge how much money or attention to focus on an a particular district, or by political watchdog groups to assist in discovering what factors lead to how people vote and whether current districts allow for equal and fair voting.

I will be using data from the United States Census Bureau website as variables in my algorithm:

<https://www.census.gov/mycd/>

I will also use the square miles of each district from Wikipedia:

<https://en.wikipedia.org/wiki/List_of_United_States_congressional_districts>

I will use this information to divide the districts by region:

<https://en.wikipedia.org/wiki/List_of_regions_of_the_United_States>

And, lastly, voting information by district from Daily Kos:

<https://docs.google.com/spreadsheets/d/1VfkHtzBTP5gf4jAu8tcVQgsBJ1IDvXEHjuMqYlOgYbA/edit#gid=1474862967>

My planned approach is to look for connections between demographics (race, age, gender, etc), household income, square miles per person (rural or suburban), and local industries (categorized by groups such as Labor, Professional, Education and the Arts). I will also break the congressional districts into geographic regions to see if variables in one region weigh more heavily than in another. I will also use the voting results versus the population of people of voting age to see if percent of turnout has an effect on results. My hopes are that I find connections between these variables that I may build a predictive algorithm on.

My final results would include graphs showing the connections between these variables and explain how they were used to build the predictive algorithm, and results of the algorithm compared to the results of the last 3 Presidential elections.