Feature Summary

Everett Sussman, Miruna Cristus, Peter Lee

Our ethnicity composition data was downloaded from the United Census Bureau website (<https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>), by selecting the “Congressional District” filter from the “Geographies” category. For each Congress from the 115th to 110th (2016-2008), we used the “White” filter from the “Race and Ethnic Groups” category. We then downloaded the “RACE” (B02001) tables for the relevant years, which contain the yearly American Community Survey (ACS) 1-year demographic estimates. The relevant columns from these tables were the congressional district name, the total population, and the total white population. We then wrote a script to merge all our datasets into one data frame, registering the congressional district columns, year (by parsing the file name), and the minority percentage as 1 minus the white population over the total population.

Since this feature would be the same for both the Democratic and Republican candidate, we compiled that data separately by scraping the Wikipedia pages corresponding to the election for each of our years (e.g. <https://en.wikipedia.org/wiki/United_States_House_of_Representatives_elections,_2016>).

The data features can be described as follows:

congressional\_district – string representing the congression district number and state.

minority\_percentage – percentage of minorities in a given district rounded to 5 decimal places (as defined by the census bureau).

year – integer representing the year of the election (2008, 2010, 2012, 2014, 2016).

candidates – list of candidates that ran for congressional office in order of vote percentage. For example, all first candidate names are those candidates that won the election.