The feature "dem\_percent" gives an approximation of the number of votes for the democratic candidate in the most recent presidential election, as a percentage of the approximate total number of votes for the democrats and republicans (we ignore third party candidates). The number of votes is approximate since presidential voting data is only publicly available by county, and not district. A county may be split among multiple districts. Therefore, we approximate the percentage of democrat votes in a district by summing the votes across all counties which are at least partially contained in that district, and doing the same for the republican vote. Though there is some inaccuracy involved, this gives an indicator of the district's preference in a presidential vote, which may be predictive of party preference in the next house vote. 2016 presidential county data was downloaded from MIT Election Data and Science Lab (https://electionlab.mit.edu/data), 2012 data was from  https://www.theguardian.com/news/datablog/2012/nov/07/us-2012-election-county-results-download#data, and 2008 data was from https://www.theguardian.com/news/datablog/2009/mar/02/us-elections-2008. Counties were matched to districts using data from the relevant US census data (https://www.census.gov/geo/maps-data/data/cd\_national.html). County assignments to districts changed during the relevant time period only in 2013 (https://en.wikipedia.org/wiki/United\_States\_congressional\_apportionment), so we use the 115th congressional district data for the 2016 election and the 110th congressional district data for presidential elections before 2016. Missing data can be explained due to some mismatch between county names between 2008 and 2012 which some districts could not be constructed.  "State\_CD" is the state and congressional district name. "year" is the year of the election.