US votes data set: description file

April 7, 2025

The US votes data set has been adapted from three publicly available datasets, whose records cover, among other things, at the county level: voting records in the 2024 US presidential election; crude estimates of depression levels; and estimates of self-reported racial identity. The data set consists of $n = 3107$ counties from all fifty US states, totalling some 152,519,316 votes. The data set includes the following variables:

STNAME The state name to which the county belongs;

CTYNAME The county name;

TOT_POP The total population of the county;

TOT_MALE The total number of males in the county;

TOT_FEMALE The total number of females in the county;

total_votes The total number of votes tallied in the county;

per_gop Percentage of votes (out of 100) in favour of Republican party;

Crude.Prevalence.Estimate Estimate of percentage of county (out of 100) self-reporting as being depressed;

race Proportion of county that describe themselves as white and not as any other race.

The assignment problem and the way that the data set in question was constructed was heavily motivated by the following posts: https://freerangestats.info/blog/2024/12/23/depression-and-vote and https://freerangestats.info/blog/2025/01/03/depression-and-vote-a See here for further information about the data sets used for analysis, and an alternative perspective on how to model these data.