There are a total of 3,111 counties or county equivalents in the data. 75% - 2331 rows are in the training data. The remaining 780 are in the test data.

The goal is to predict the percentage of voters in a county that voted for Biden in the 2020 US Presidential Election. Predictor columns are demographic and education information for each county. The demographic information are estimates made by the US Census Bureau.

Notes: Alaska has been removed from the data set. Votes in Alaska are tabulated in 41 districts. However, census information for Alaska is recorded in 30 boroughs or census areas. I could not match the demographic information from the census to the voting districts.

Goal: The goal is to predict the percentage of voters
in a county that voted for Biden in the 2020 US Presidential Election. Predictor columns are demographic and education information for each county.

- **Introduction**: context and background information
    - External sources citation
    - Mention what variables are believed to be associated with the response variable based on the background information
    - Approximately 100 words
- **Exploratory Data Analysis**
    - Explore potential relationships between the variables
    - Provide graphs and visualization showing relationships with descriptions (around 20 words for each)
    - Make transformations of some variables
    - Discover some possible interactions between variables
- **Preprocessing / Recipes**
    - Create different recipes, and explain reasons behinds the steps
    - Perform preprocessing of variables
    - Approximately 100 words

- **Candidate Models**
    - Construct various candidate models
    - Describe each candidate model briefly
    - Include a table listing of all candidate models attempted with
        - Model Identifier
        - Type of Model
        - Engine
        - Recipe used or listing of variables in the model
        - Hyperparameters
- **Model Evaluation and Tuning**
    - Discuss the evaluation and comparison of the candidate models that were attempted
    - Construct V-fold cross validation to measure the performance of the candidate models
    - Tuning the hyperparameters
    - Summarize the performance of each model with a table, including:
        - Model identifier
        - Metric score: RMSE
    - Include autoplot comparing the performance of the different models
- **Appendix:**
    - The final script used to produce results with annotated comments