# Voting Frequencies

Ben Lee (Team Lead, bele@ucdavis.edu), Ander Duong (andduong@ucdavis.edu),
Namita Hegde (nahegde@ucdavis.edu), Eric Daniel Tuason (ertuason@ucdavis.edu),
Rui Zhao (ruizhou@ucdavis.edu)

## 1 Problem Statement

How do education, race, gender, and income bracket affect voting frequency?

## 2 Dataset Description

This data was collected from surveys and polls from different countries and was updated as of March 2023. The data contains four independent variables (education, race, gender, and income) and one dependent variable (voting habits). The education category describes the voter's highest education level: college, high school or less, or some college. The race category describes the races of voters: White, Black, Hispanic, or other/mixed communities. The gender category describes voters of two genders: male or female. The income category describes voters in four different income brackets: less than $40K, $40-75K, $75-125K, or $125K or more. The voting habit category describes the frequency with which the individual tends to vote: sporadic, always, or rarely/never.
This dataset also has a quality rating of 10.00, which was one of the primary reasons we chose to work with it. It has a 100% rating in completeness, credibility, and compatibility which means it should be effortless to analyze.

## 3 Goals

We would like to know which voter description category or categories can be best used to predict voting frequency. Additionally, we would like to be able to create accurate models that form an ideal fit to map our independent variables (education, race, gender, income level) to our dependent variable (voter frequency).

## 4 Timeline

**Week 3**: Background study
**Week 4**: Use a correlation matrix/pairplot to understand and analyze the dataset
**Week 5**: Use supervised machine learning to develop an accurate model
**Week 6**: Use SSE and variance to evaluate the models and test performance
**Week 7-8**: Develop a basic web-based front-end to invoke and run the model(s) on input data and display the prediction output
**Week 9-10**: Create Final Report