Mikhail Kouzminov

CMPS 161 Programming Assignment 2 Technical Document

The Problem:

Proposition 3, the California Water Infrastructure and Watershed Conservation Bond Initiative, was one of the closest votes this last election. According to ballotpedia, it was a "measure to authorize $8.877 billion in general obligation bonds for water infrastructure, groundwater supplies and storage, surface water storage and dam repairs, watershed and fisheries improvements, and habitat protection and restoration. It would likely generate $8.4 billion in interest over 40 years, costing the state about 17.3 billion". However, water conservation has been a major issue in California, with a lot of draughts over the past few years, and this bill might be important for California's survival.

Proposition 3 was one of the most closely contested propositions this election, and ultimately failed 49.3 to 50.7 %. I wanted to see if there was any way to determine what social and economic characteristics could be best used to predict the results of this and future legislature like it. Here is a map of distribution of votes (approval is red), with the official map from the government website displayed to the right:
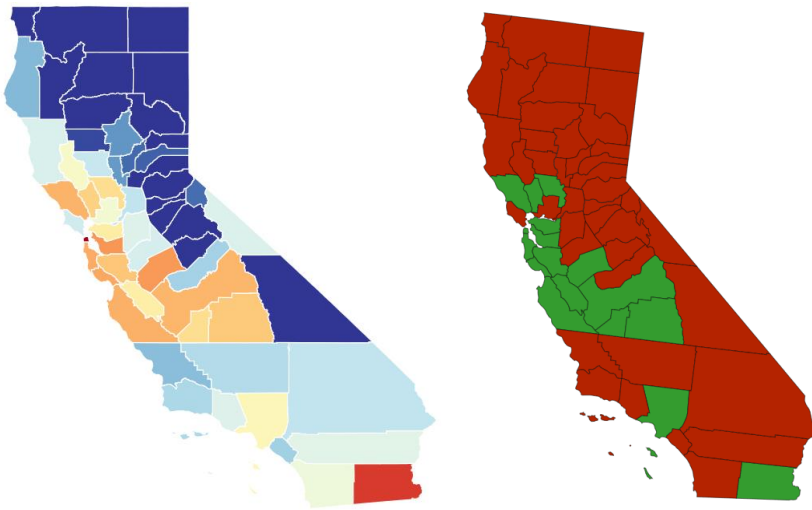

This left image will be displayed next to (to the right of) all future images for comparison. Here, for purposes of explanation, is the image of Majority approval vs disapproval, copied from the government vote registry website, with red symbolizing disapproval.

My Approach:

Note that there is no obvious pattern, except majority approval in Central California and Los Angeles, and preety much nowhere else. Because of this, I decided to limit my problem, and look specifically at counties where over 55% of people disapproved of it. Many of the counties where the majority voted against, are actually fairly even on further examination. As such, this was the map of counties where a noticeably large majority (55%) disapproved of the legislation (disapproval painted in red):
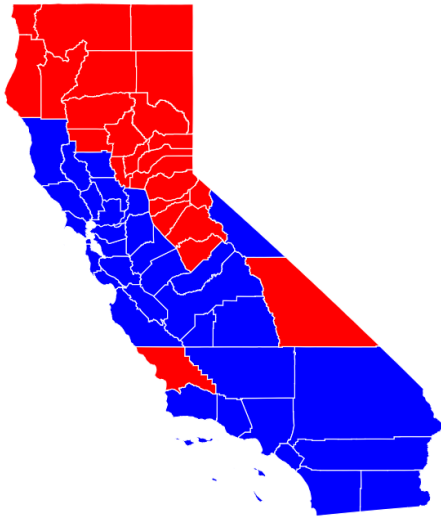

As you can see, there are far fewer counties where this many people disapproved of it. Since I decided to specifically look at economic and cultural indicators, one of the first probable indicators I looked at was median income. This was the result of that mapping. Here, red represents a high median income:
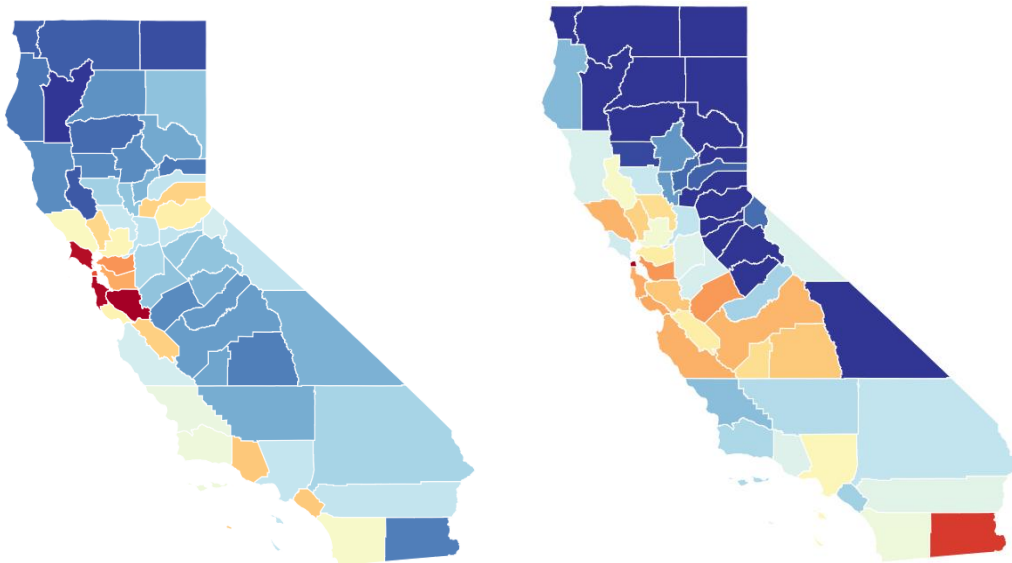
Figure 1 Map of Median Income vs Prop 3

As shown, there is obviously some slight correlation – the far Northern counties tend to be the poorest in California. However, it isn't very clear, since some north-eastern counties have relatively high median income and still voted against the legislation. Further, some of the poorer counties clearly voted for the proposition: in particular, the counties in Central California, and Imperial County to the far South of California approved the legislation. There are too many outlying cases for this to be a good vector for analyzing the problem.

Next, I looked at examining whether there was any correlation between farming jobs – or lack thereof, and approval of legislature. I theorized that counties with higher amounts of farmers would be more likely to accept the legislature, since water is important for farming. Here red represents counties where a larger proportion of people work in farming:
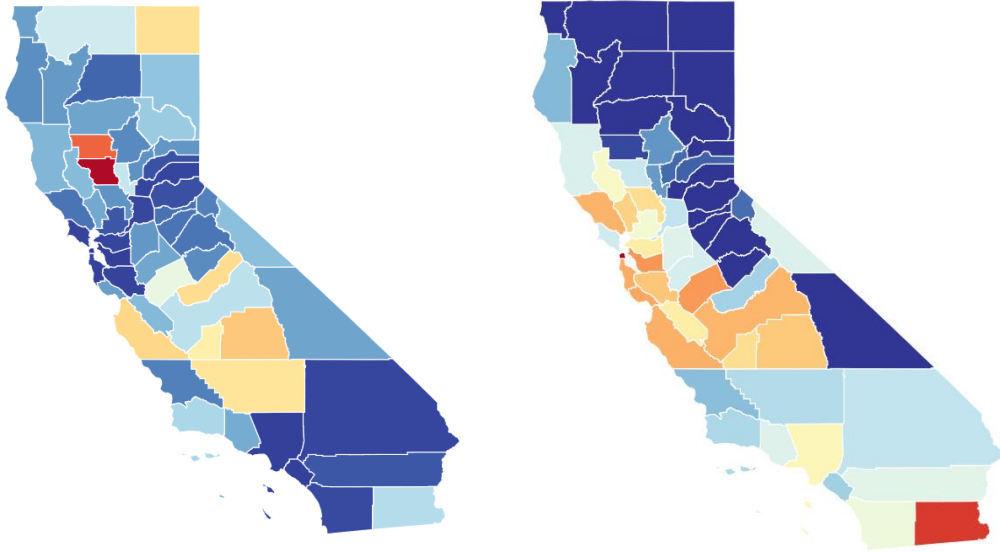
Figure 2 Map of Farming occupations vs Prop 3

As shown, while there is there is the logical possibility for a correlation, there is none in reality. Some of the areas with the highest amount of farming definitively rejected the proposition, while some of the ones with the lowest did not.

I then tried analyzing population density (red is higher), to see if perhaps that correlated with approval:
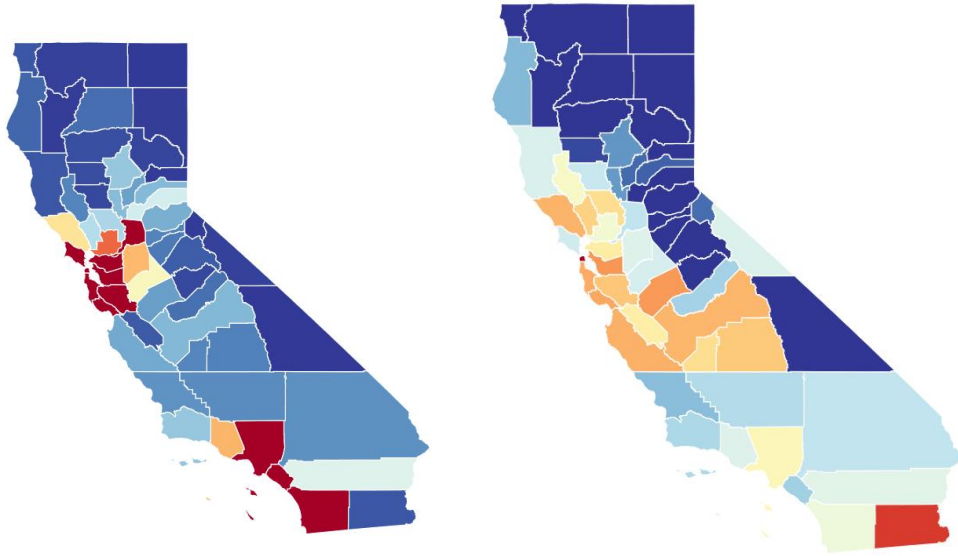
Figure 3 Population Density vs Prop 3 aproval

This is far closer to a direct correlation, and might explain why certain Northern areas of California were against Prop 3. Also, areas with higher population are more for it. However, some areas with slightly higher population density in central-eastern California still disapproved of Prop 3, while some areas with higher population in Northwestern and Central California were not as disapproving.

There is one more correlation to approval of Prop 3, that I analyzed and found interesting – counties with a higher median age had a higher likelihood of disapproving Prop 3, as did, specifically counties with a higher proportion of people over the age of 65.

Here is a heatmap showing median age of the population of each county, with red being older:
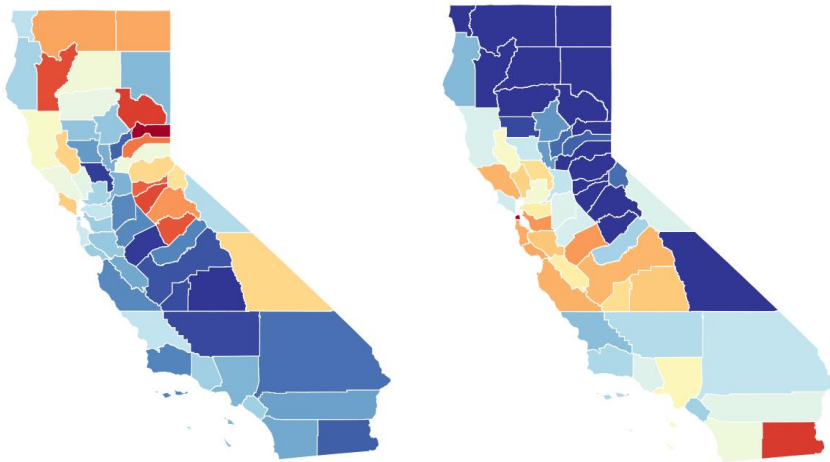
Figure 4 Median Age vs Prop 3 Aproval

And a map of proportion of people over 65, with red meaning a higher proportion of elderly:
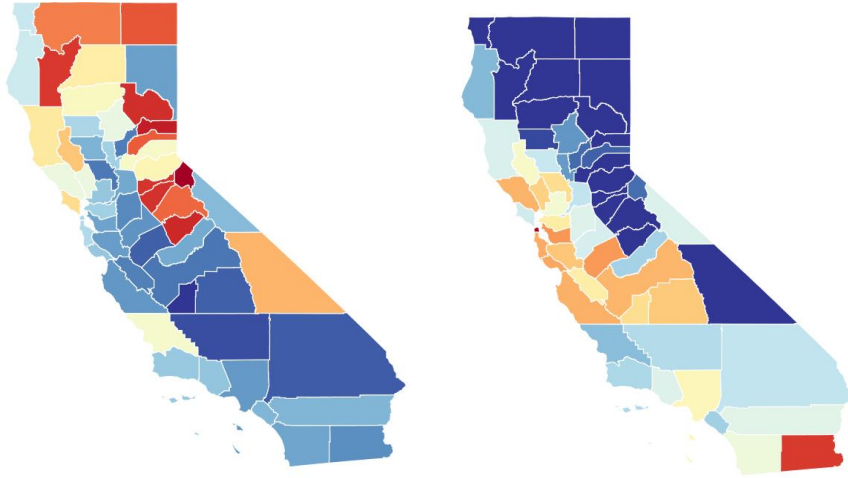
Figure 5 Proportion of Retiree Aged people vs Prop. 3 Approval

As shown, there is a clear correlation for both for counties in Northern California. There are exceptions, but these places in Northern and Eastern California appear to likely have a high concentration of retirees. While the same does not hold true of Western California, there is a positive correlation between counties with a high median age specifically disapproving of Prop 3. There does not seem to be a direct logical reason for this, except that perhaps these areas are used for retirement homes, and already have good infrastructure. Retirees might also be less interested in planning for the future.

Final Analysis:

These results point out specific social and economic demographics that would be useful to target if trying to pass a bond to help improve California's water infrastructure through a similar proposition:

Most importantly, it would help to target Northern California – Central and Southern California tend to approve of improving water infrastructure.

Secondly: elderly retirees. Most areas who heavily disapproved Prop.3 had a high amount of elderly retirees. This might be because these areas are potentially used as retirement spots, and may have better local water infrastructure.

Alternate options for visualization (not used):

I was considering using Parallel Coordinates, however, this would not work as well in this case as in some others: Specific geographic location very much seems to affect the amount of correlation between an analyzable dimension, and the resulting approval or disapproval. There does not seem to be a single dimension of information with which to predict how people vote, and several factors contributing in various ways contribute to predicting a decision.

Local geographic conditions are difficult to analyze for how they affect the people living there when the vote, but this is such a situation. Moreover, there does not appear to be a single overall rating for water infrastructure for a given county, and conditions in neighboring counties can possibly affect each other. It would be difficult to analyze the need and quality of water infrastructure for each county, so an analysis of likelihood of approval using such a rating would be potentially useful, but not necessarily relevant in places with low populations, such as these counties.