Money in Politics : An Analysis of 2016 California Ballot Propositions Contributions

Serigne Mourtallah M'backe Faye

17 April 2021

In November 2016 Californians voted on 17 different propositions, including the most in over a decade. This paper is an exploratory analysis of the political contributions donated in support or in opposition of these propositions, including investigations of anomalous, paradoxical observations, a two sample t-test comparing the means of the sums of contributions donated to propositions that passed to those that failed, and lastly, a exploration of efficacy of modeling the probability of a proposition being passed with the amount of money contributed to that propositions using a logistic regression model.

Overview

In California, ballot propositions are initiatives or referendums posed to it's electorate during presidential elections to which they may vote "Yes" or "No" in a direct[1] vote. If passed, these propositions becomes state law. It is by this process that Proposition 9 of 1974, as known as the Political Reform Act, was passed requiring every dollar raised and spent on a political campaign be disclosed and that the committees associated with the propositions register with the secretary of state and disclose on their contributors including their names, occupations and places of residence.

While every state has a database to manage these records California's CAL-ACCESS is notorious for being engineered especially poorly, earning the ire of journalists, academics and transparency advocates with Secretary of State Alex Padilla calling emphatically for it's "complete overhaul"[2] in 2015.

Fortunately by that time, a group of journalists and developers had already been working on this problem for over a year, forming the California Civic Data Coalition to address this very problem. Although the project is still in development, their mission is to make the dirty, jumbled CAL-ACCESS databases accessible so that novices can analyze and understand it's contents. It has been well received: experimental versions of data mined and cleaned by CCDC has powered several investigations into money in politics conducted by the Los Angeles Times[3]. One of its co-founders, Cheryl Phillips, and one of its former lead developers, James Gordon, in collaboration with Ben Welsh of The Los Angeles Times and Andrea Suozzo of the Seven Days newspaper authored an introduction to Python course that provides data on the committees and contributions for the 2016 election and inspired this paper.

[1] i.e. majority rules.

[2] White, Jeremy. "Secretary of state hits California's 'Frankenstein monster' campaign finance database." *Sacramento Bee* 15 October 2015.

[3] The team also earned the 2015 Knight News Challenge Award for their efforts.

Data

Summary

The 2016 CalProps data is comprised of two dataframes (contributions and committees), which we combine to create a 90,264 x 18 contribution level dataframe. Each observation contains the amount of the contribution, the committee the contribution is associated with, the proposition that the committee is associated with, and stances assumed by the committee on said proposition, and some basic information about the contributor, including their name, occupation and state or residence.

Table 1: Structure of merged contributions and committees data.

	datatype
calaccess_committee_id	numeric
ocd_prop_id	character
calaccess_prop_id	numeric
ccdc_prop_id	numeric
prop_name	character
ccdc_committee_id	numeric
committee_name.x	character
committee_position	character
committee_name.y	character
calaccess_filing_id	numeric
date_received	Date
contributor_city	character
contributor_state	character
contributor_zip	character
contributor_employer	character
contributor_occupation	character
contributor_is_self_employed	logical
amount	numeric

Propositions

Odd Observations : Negative Support

It was presumed that negative values of "amount" corresponded to opposing a proposition but upon closer inspection at the subset below, we observe that there are contributions of sizable amounts for which the committee position is SUPPORT, yet it's amount is negative.

Table 2: Sample of irreconcilable observations.

	amount	committee_position
67317	-27000	SUPPORT
67318	-9000	SUPPORT
67636	-150000	SUPPORT
77730	-2500	SUPPORT
81650	-5000	SUPPORT
83742	-2450	OPPOSE

We may expect the amount of contributions that have this paradoxical attribute to be distributed uniformly across all propositions. The barplot in Figure 1 shows that this is not the case: that the frequency of these types of records are much higher among contributions to Proposition 63: "Background Checks for Ammunition Purchases and Large-Capacity Ammunition Magazine Ban Initiative."

This proposition, when passed, banned the possession of magazines of capacity of more than ten rounds, and require background checks for the purchase of ammunition.

We deal with these ambiguous observations by applying an absolute value transformation to the amount.

Missing Data

This dataset would be completely dense if not for 16 missing observations in amount coming from contributions on exactly two propositions, namely, Prop 67, The Plastic Bag Referendum, and, once again, Prop 63. It is not clear why proposition 63 is over-represented in corrupted data but it is clear that these corrupted and missing observations within Prop 63 come from a singular source: The Coalition for Civil Liberties, a project of the California Rifle and Pistol Association. These observations are dropped in subsequent aggregations.

Aggregations

An aggregate of ~$4.7M was donated across all propositions with an average amount donated per proposition of $5,177.06, and an average number of contributions per proposition of 5309.64. Mixing these two ratios, we observe that the average dollar to contributor ratio is .97, approximately a dollar per contributor.

In examining aggregations on contributions across proposition in Figure 3, we observe that number of contributions for Prop 62 and 66 are much higher than the other at over 30,000. Prop 63 and 67 have
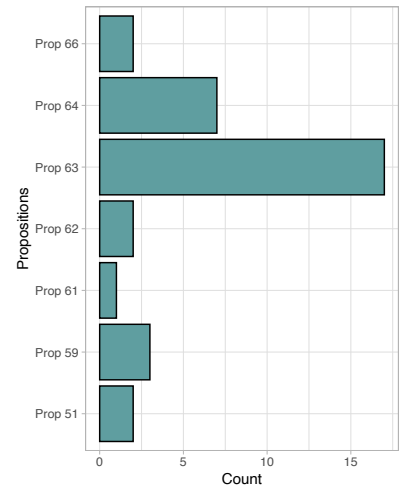

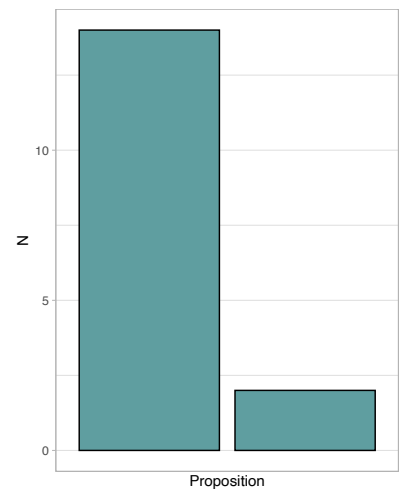Figure 1: Frequency of Corrupted Records by Proposition

Figure 2: Frequency of Missing Data grouped by Proposition

around 11000 and 6000 contributors respectively, and the number of contributors for the remainder of propositions were negligible in comparison.
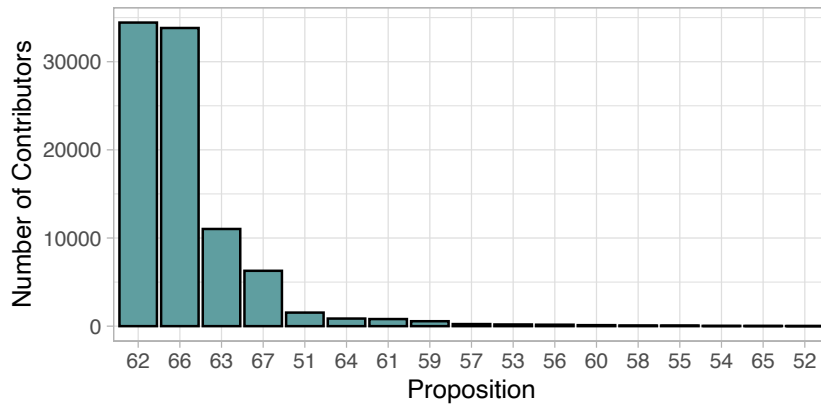

Figure 3: Unique Contributors by Proposition

In Figure 4 we aggregate dollars contributed across propositions and find that propositions 55, 56 and 61.
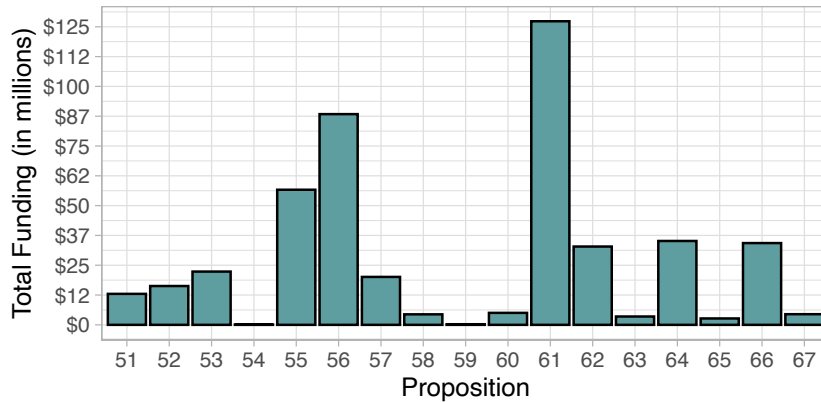

Figure 4: Total Funding Per Proposition

The relationship between the number of contributors per proposition is further explored in the scatterplot and the violin-dot plot in figures 5 and 6. We observe that there are at least two distinct profiles for propositions: propositions like 55, 56, and 61 that are extremely well funded by albeit by a small handful of very large contributions, and propositions like 66 and 62 which do not have large sums, in terms of contributions, but garnered at least three times as many contributions as the other propositions.
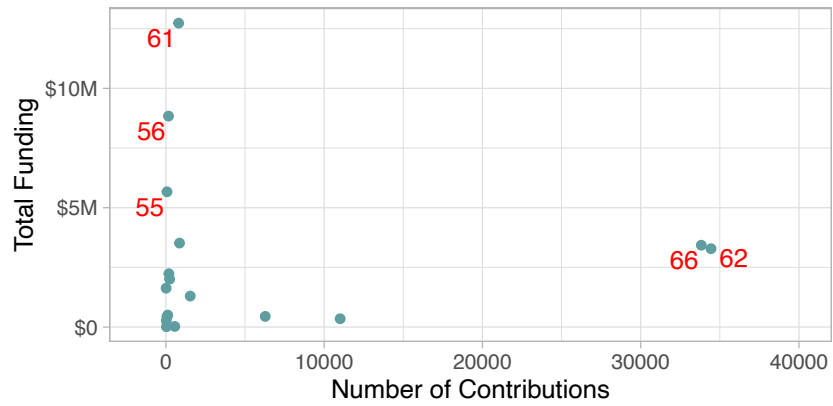
Figure 5: Proposition Total Funding by Number of Contributions

Figure 6: "Violin-Dot Plot of contributions by Propositions by Stance"

Bivariate Relationships

A closer look at figure 6 gives us an some insight into the distributions of the contribution amounts for each proposition. Well funded propositions like 56 and 61 were hotly contested by large contributions in support of and in opposition to their respective propositions, while propositions like 60 and 61 seem to have garnered a sizable amount of small contributions.[4]

Models

Welch's Two Sample T-Test

A two sample t-test was conducted to investigate the difference in the sums contributed to propositions that were passed and propositions that failed. The t-test was not significant $t = .6158, df = 4.93$, $p = .5652$, thus we fail to reject the null hypothesis that the difference in total contributions to losing and failing propositions is not statistically different from zero.

Logistic Regression

A binary logistic regression was conducted to investigate if funds raised for a proposition or if popular support for a proposition predicts if a proposition will be passed. The outcome of interest was the success or failure of the proposition. Funds raised for the proposition was measure by the average per proposition contribution size, and popular support was measured by the proportion of contributors supported the proposition.

The Hosmer-Lemeshow goodness of fit test was not significant $(p > 0.5)$ indicating that the model is correctly specified. The model resulted that the mean contribution size was not significant $(p = .6118)$, however, the proportion of supports among contributors was found to be significant at the $\alpha = .1$ level. The full model had a $\chi^2 = 5.41, p = 0.07$ and a Cragg-Ulher Pseudo-R^2 of .39. Thus, after controlling for the average size of the contribution for a proposition, the proportion of supporters was found to contribute to the model. The un-standardized coefficient, $\beta = 3.607$ 90% CI = $.49, 6.72, SE = 2.992 * 10^-6, p = .0569$ corresponding to a approximate predicted 15 probability for a proposition being passed with no contributors in support of it and an approximate .05% increase in the probability that a proposition will be passed for every 1% increase in the proportion of supporting contributors to a contribution.

[4] Welsh et al. warns in their tutorial that contributions less than $100 are absent from this dataset, citing the fact these contributions are not required to be reported. Nevertheless many small contributions appear, contradicting this. What proportion of the total amount of contributions in these amounts appear here are unclear.
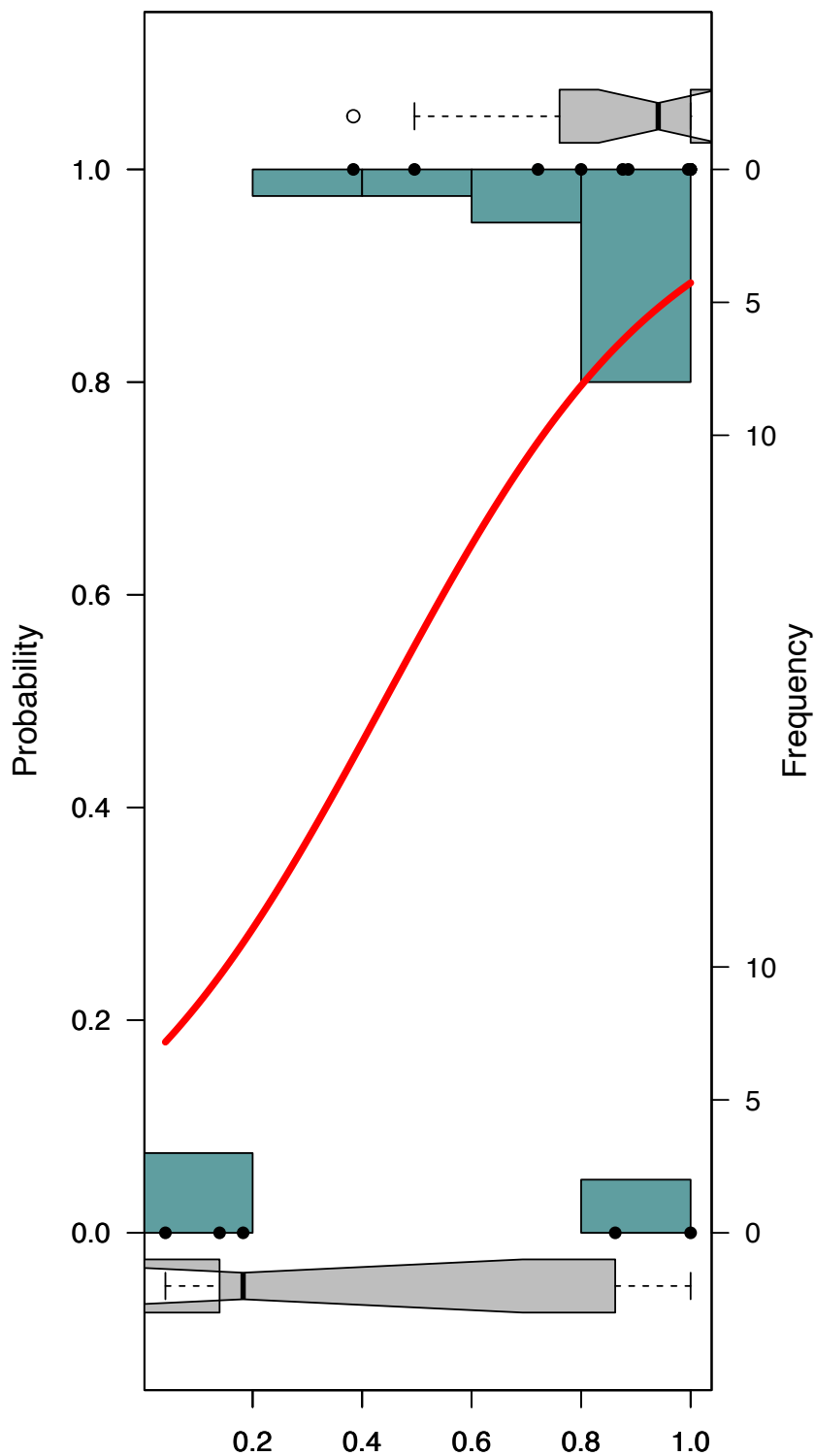
Figure 7: Logit model of Log-odds of Proposition Passing by proportion of pontributors that support.

Conclusion

From the results of a welch's two sample t-test we concluded that there was insufficient evidence to conclude that the total amount contributed to a proposition is different between the groups of propositions that failed, or succeeded. In analyzing the outcomes of the 2016 California propositional ballot elections, we observe that the proportion of contributors that were supporters of a proposition is associated with an increase in the probability of the proposition being passed. While the p-value this coefficient, is significant at the $\alpha = .1$, indicating that the proportion of contributors that support a proposition descriptive value, the confidence intervals are too wide to justify using this model predicatively. This limitation is no doubt due to the small sample size, but these results serve as impetus for subsequent research in how proportions of financial contributors to a proposition may effect the probability that a proposition will succeed, irrespective of what the size or amount of those contributions are.