Introduction

In California, a special election to recall the governor can be triggered by a petition with support from 12% of the voters who participated in the most recent gubernatorial election. California has had two recall elections in the past, one in 2003 for the recall of Governor Gray Davis which succeeded and replaced him with Governor Arnold Schwarzenegger, and another in 2021 for the recall of Governor Gavin Newsom which failed to pass. Due to the unique factors presented in the California 2021 Special Election to recall Governor Gavin Newsom, this election is a special case study of voter turnout. My research question of interest is, "Did age have a unique effect upon voter turnout by political party in the 2021 California Special Election to recall Governor Gavin Newsom?"

Age is considered a primary motivator of voter turnout. Hence, this study will consider voter turnout by age group across political parties to analyze if voter motivation was heightened in non-incumbent parties due to these unique factors, and if this effect is applied evenly across age groups. I hypothesize that older voters will retain motivation to vote regardless of predicted election outcome, but in the Democratic party (the incumbent party in this recall election) there will be less motivation in younger voters. I predict this discrepancy in younger voters across political parties results from lacking turnout urgency for Democrats due to incumbency advantage and the strength of the Democratic party in California. After Gavin Newsom won his election for Governor of California in 2018 with 61.9% of the vote, the likelihood of a successful recall without

a popular Democratic candidate in 2021 in replacement was perceivably low. I predict this to have created a relatively low turnout for Democrats, as perceived low turnout urgency would fail to overcome the individual time costs involved in voting. California Republicans, however, had already demonstrated their willingness to turnout by triggering the recall election via petition signature. With this heightened urgency, I predicted this to have created a relatively high turnout for Democrats. In older voters, however, voter turnout is traditionally much higher than in younger age groups due to the negligibility of the individual time costs to vote for older voters. This relationship will be studied through multivariate linear regression, with the implementation of an interaction term to capture the unique effect of age upon voter turnout between Democrats and Republicans in the 2021 Special Election to recall Governor Gavin Newsom.

First, I will discuss the existing literature surrounding the California's 2003 and 2021 Special Elections to demonstrate the unique voting results which are presented by recall elections. This will highlight the nonuniform effect that California recall elections have upon voter turnout and voting outcomes by demographic. I will demonstrate that the existing research thoroughly analyzes demographics such as turnout by racial demographics but does not analyze age to the same depth. Next, I will establish that this recall election breaks the theory that recall elections are primarily a tool of accountability and occur when voter urgency is at its highest. As not many recalls happen in conditions of low voter urgency and with large victory margins, this research uniquely analyzes the California 2021 Special Election for its contribution to explain voter urgency in recall elections. I will then employ three linear ordinary least squares

regressions to investigate the causal impact of age upon voter turnout between political parties. My findings indicate that there is a significant effect, but that the direction of the relationship is contrary to my predictions. Democrats had higher turnout at ages under 50, and non-Democrats had higher turnout at ages over 50. I will then investigate potential causes of concern for internal validity and discuss whether the implications of my findings suggest the reassessment of the minimum threshold to initiate a recall election in California.

Literature Review

The Uniqueness of Special Elections

Recall elections have been frequently studied in the literature. This includes many papers which study the 2003 California Special Election to recall Governor Gray Davis. However, the 2021 Special Election in California to recall Governor Newsom has not yet been studied for its contribution to the existing literature per my cursory findings. Hence, continuing in examining California's recall elections, this study greatly contributes to the existing literature by being an early paper to quantitatively analyze the results of voter turnout for this election.

Within racial and ethnic groups, noteworthy and unique voting patterns emerge from the two-step recall vote. In the 2003 California Special Election, African American voters, Latinos voters, and Asian American voters responded "Yes" on the recall 37.9%, 14.8%, and 12.5% less than Anglo voters respectively (Segura & Fraga, 2008). This is likely due to political affiliation being unevenly distributed across ethnicity. However,

beyond the vote to recall, there were significant differences in replacement candidates between racial groups. Within those who voted "No" upon the recall, Latino voters were nearly twice as likely to vote for Bustamante (a leading Democrat candidate) as their replacement candidate than Anglo voters (Segura & Fraga, 2008). These types of significant effects are well documented in recall elections. The existing literature thoroughly demonstrates the importance of ethnic factors upon voter outcomes but does not as thoroughly examine the effect of age upon voting outcomes. As age is proven to be a primary motivator of voting activity due to lowered costs of voting and more time to fulfill voting costs (Harder & Krosnick, 2008), this study will fill this hole in the existing literature by analyzing age groups by party for their turnout in this recall election.

What Drives Recall Elections and makes them Successful?

Michael Haman's paper *Recall Elections: A Tool of Accountability? Evidence From Peru* examines recall elections through the lens of political accountability, observing over one thousand local level elections in Peru (primarily mayoral elections) for the factors which lead to the initiation of recalls and the likelihood that they succeed (Haman, 2021). Contrary to the existing theory that recall elections are a tool of governmental accountability, Haman concludes that recall elections are held in municipalities which hold high rates of party affiliation and recently had close margin victories for the position (Haman, 2021). Haman cites this outcome because of heightened political urgency which arises from these conditions (Haman, 2021). Through this lens, recalls are triggered when they have a reasonable chance of success. Haman finds that the largest factor of success in recall elections is the political

experience of the organizer (Haman, 2021). This implies that it is the opposition to the incumbent which determines the result of the recall more than the competency of the incumbent themselves. These findings are held in contrast given the circumstances of the 2021 California Special Election. Governor Newsom did not win by a narrow margin in 2018; he won the race with 61.9% of the vote, a 23.8% margin over that of his opponent, the largest landslide for a non-incumbent since 1930 (Williams, 2021). This contrasts Haman's theory that recalls are driven by contentious elections. This discrepancy may be due to a variety of factors, including the relatively low threshold to begin a California Gubernatorial recall of 12% of the voters of the last election compared to the threshold in Peruvian municipalities. It was always unlikely that the recall would pass given Governor Newsom's 2018 support, and the results of the recall election were 61.9% in Newsom's favor (Briz et al., 2018): the exact same percentage who supported his 2018 campaign. This suggests that previous Newsom voters were steadfast in their support, and that the recall election was triggered primarily by a vocal minority in opposition of Governor Newsom. Hence, this study will examine voter turnout by age group across parties to analyze the urgency felt by different political demographics to participate in the recall vote given the relatively safe vote in favor of the incumbent. This type of analysis is uncommon in the existing literature, as most studies focus upon those elections with close margins, or those with successful recalls. This study uniquely will analyze which voters choose to mobilize despite an absence of typical factors which enable recalls per existing research.

Data and Methods

All data used in my research is from the "Statewide Database", a non-profit public access online database for all California districting. The database hosts official government voter registration and voting result data for all California elections, as provided by the State of California (2021 Special Election Precinct Data, 2021). Because this data is authoritative, it is extremely reliable. It accurately portrays the target population of California registered voters at the census level, as every registered voter is represented in this comprehensive data. All data is aggregated at the precinct level to exclude any personally identifiable information from the data. Hence, each observation is representative of one precinct in the state of California. I use the "SR Precinct", a consolidated precinct of voting precincts and registration precincts used for statistical merging. There are 15,525 SR Precincts in the data, representing all precincts in California at the time of the 2021 Special Election. Notably, this is the first implementation of the 2020 Census Block redistricting boundaries in an election. The average number of voters per precinct is 1,328, with a total of 20,619,582 registered voters in total, which represents 52.55% of the population of California in 2021.

The dependent variable (Y) is *TotalRegTurnout%*, a continuous numerical variable ranging from 0-100% with a mean of 58.91%, representing the turnout of registered voters in the precinct as a percentage. There are two ways to calculate voter turnout: votes cast as a fraction of all eligible voters, or votes cast as a fraction of only registered voters. For this research, voter turnout will be measured as a percentage of turnout for registered voters. From the Statewide Database, I downloaded two files: one containing data for all registered voters by precinct and one containing data for number of votes cast by precinct. These were imported in R and saved as data-frame objects,

where I operationalized the variable accordingly: $\frac{Voted\ Total}{Registered\ Total} * 100 = TotalRegisteredTurnout\%$.

Voter registration data allows me to compare registered party identities to voter turnout, which is not possible for turnout as a percentage of all eligible voters. Additionally, registration data is publicly accessible, where eligible voters data is not.

The independent variable of interest (X_1) is *AvgRegAge*, a continuous numerical variable ranging from 21-72 years of age with a mean of 48.37 years, representing the average age of registered voters in the precinct. The data includes a breakdown the total of registered voters by age bracket (18-24, 25-34, 35-44, 55-64, 65+) across gender (Male or Female) and party affiliation (Democrat, Republican, Decline to Answer, Other). For example, one column of the original data would be "DEMM1824" (representing the total number of Democrat registered men aged 18-24). I tabulated all gender and party affiliation permutations by age bracket and applied a weighted average of the center value of the age bracket ($Total_{18-24} = \frac{24-18}{2} * \sum DEMM1824 + DEMF1824 + REPM1824 + REPF1824 + DCLM1824 + DCLF1824 + OTHM1824 + OTHF1824$). For the "65+" age bracket, I chose the upper age limit to be the 2021 life expectancy in California, 79 years old. I replicated the weighted average for each age bracket, tabulated them and divided by the total number of registered voters for each precinct ($\frac{\sum AgeRanges}{Registered\ Total} = AvgRegAge$). The finished operationalized variable is an average age of registered voters in each precinct.

The first control variable (X_2) is *NonDemReg%*, a continuous numerical variable ranging from 0-100% with a mean of 55.83%, representing the percentage of registered voters in the precinct who are not registered as Democrats. This serves as a control for voter urgency, as belonging to the non-incumbent party is predicted to increase urgency to recall Governor Newsom, hence increasing turnout. Party affiliation is also predicted

to correlate with registered age, as Republicans are typically older than Democrats. Hence, I expect a positive coefficient on this control variable due to its positive relationship with both *Y* and *X₁*. This variable is operationalized as the complement of the proportion of registered voters who are registered Democrats ($\left[1 - \left(\frac{Registered\ Democrat}{Registered\ Total}\right)\right] * 100 = NonDemReg\%$).

The second control variable (*X₃*) is *MaleReg%*, a continuous numerical variable ranging from 0-100% with a mean of 51.28%, representing the percentage of registered voters in the precinct who are male. This serves as a control for gender, as women tend to turnout in higher rates than men. Gender is also predicted to correlate with average age of the precinct, as women tend to live longer. Hence, I expect a positive coefficient on this control variable due to its negative relationship with both *Y* and *X₁*. This variable is operationalized as the proportion of registered voters who are male. ($\left[1 - \left(\frac{Registered\ Male}{Registered\ Total}\right)\right] * 100 = MaleReg\%$).

	(*Y*) TotalRegTurnout%,	(*X₁*) AvgRegAge	(*X₂*) NonDemReg%	(*X₃*) MaleReg%
Minimum	0.00	21.00	0.00	0.00
Median	61.15	48.13	55.03	50.78
Mean	58.91	48.37	55.83	51.28
Maximum	100.00	79.00	100.00	100.00
Standard Deviation	16.22	5.58	15.46	7.51

Each column represents a separate variable. The first 5 rows are summary statistics, rounded to two decimal places.

Model 1 is a naive regression of *TotalRegTurnout%* on *AvgRegAge* using linear ordinary least squares regression. The regression uses the following form: $TotalRegTurnout\% = \alpha + \beta_1 AvgRegAge + \epsilon$. The coefficient of interest β_1 is predicted to be a positive value, representing that as average age of registered voters in a precinct increases, so will registered voter turnout. Model 2 is an interaction regression of *TotalRegTurnout%* on *AvgRegAge*, *NonDemReg%, and NonDemReg% * AvgRegAge* using multivariate linear ordinary least squares regression. The regression uses the following form: $TotalRegTurnout\% = \alpha + \beta_1 AvgRegAge + \beta_2 NonDemReg\% + \beta_3 AvgRegAge * NonDemReg\% + \epsilon$. The coefficient β_2 is predicted to be a positive value, representing that as non-Democrat registration in a precinct increases, so will registered voter turnout due to heightened urgency. The coefficient β_3 represents the additional effect of *AvgRegAge* on *TotalRegTurnout%* at greater levels of *NonDemReg%*. In other words, this is the additional effect of age upon turnout rate caused by increased proportions of registered voters of non-incumbent parties. This coefficient is predicted to be negative, as I expect a diminished effect of being registered as non-Democrat in older voters who are not as affected by the individual time costs of voting. Model 3 is a variation of Model 2 with the inclusion of *MaleReg%* to control for the effect of gender of registered voters by precinct. The regression uses the following form: $TotalRegTurnout\% = \alpha + \beta_1 AvgRegAge + \beta_2 NonDemReg\% + \beta_3 AvgRegAge * NonDemReg\% + \beta_4 MaleReg\% + \epsilon$. I expect the coefficient β_4 to be negative, as men typically have lower voter turnout than women.

Results

	Dependent Variable (Y): TotalRegTurnout%		
	(1) TotalRegTurnout%	(2) TotalRegTurnout%	(3) TotalRegTurnout%
(X_1) AvgRegAge	1.69954*** (0.01889)	2.269*** (0.05026)	2.283*** (0.05039)
(X_2) NonDemReg%		0.5578*** (0.04088)	0.5661*** (0.04093)
($X_1 * X_2$) AvgRegAge * NonDemReg%		-0.01051*** (8.155e-04)	-0.01070*** (8.169e-04)
(X_3) MaleReg%			0.05044*** (0.01401)
Constant/Intercept	-23.29125*** (0.91961)	-53.35*** (2.451)	-56.55*** (2.606)
Confidence Interval Lower Bound of X_1	1.662517	2.170613	2.184069
Confidence Interval Upper Bound of X_1	1.736556	2.367642	2.381586
(n) Sample Size	15525	15525	15525
Adjusted R-squared	0.3427	0.3509	0.3514

Each column represents a separate regression. The entries in the first 4 rows are estimated regression coefficients, with standard errors below them in parenthesis. The asterisks indicate whether the coefficient is statistically significant at the 10% level (*), 5% level (**) or the 1% level (***).

In Model 1, it is predicted that increasing the average age of registered voters in the precinct by one year will increase the total turnout of registered voters by 1.70%. This effect is statistically significant at the 10%, 5%, and 1% significance levels. The adjusted R-squared is .3427, which is meaningfully large. This strength of the

relationship is quite strong, and clearly demonstrates that age is a prominent predictor of voter turnout among those registered, with a positive relationship as predicted. However, with no control variables in the regression, the analysis is not sufficient to claim this relationship causal. Having demonstrated the importance of age in predicting voter turnout, I will move to testing the interaction term of age and party affiliation to analyze how age affects voter turnout uniquely by the party identity of the precinct.
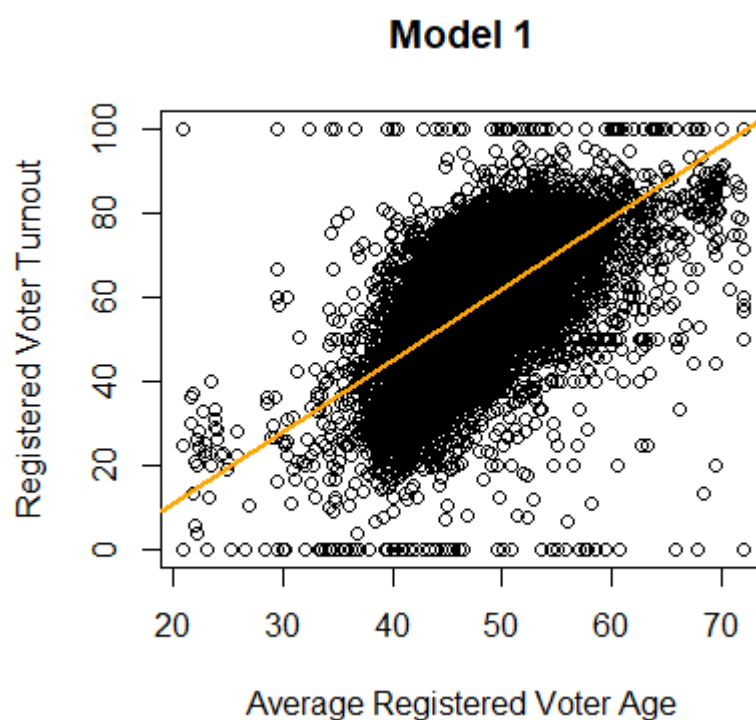
Model 1

Model 2 includes a control for the percentage of non-Democrats registered in the precinct, which is statistically significant at the 10%, 5%, and 1% significance levels. The inclusion of this control captured some omitted variable bias that was uncontrolled in Model 1, as the effect is statistically significant and meaningfully large. However, with only a 2.39% increase in adjusted R-squared, it did not account for much omitted

variable bias. The effect of the interaction term, β_3 is statistically significant at the 10%, 5%, and 1% significance levels, and is negative as predicted. Because the effect is statistically significant, we can conclude that the additional effect of *AvgRegAge* on *TotalRegTurnout%* at greater levels of *NonDemReg%* is meaningful. In other words, the additional effect of age upon registered voter turnout rate caused by increasing proportions of registered voters of non-incumbent parties is statistically significant.

Model 2
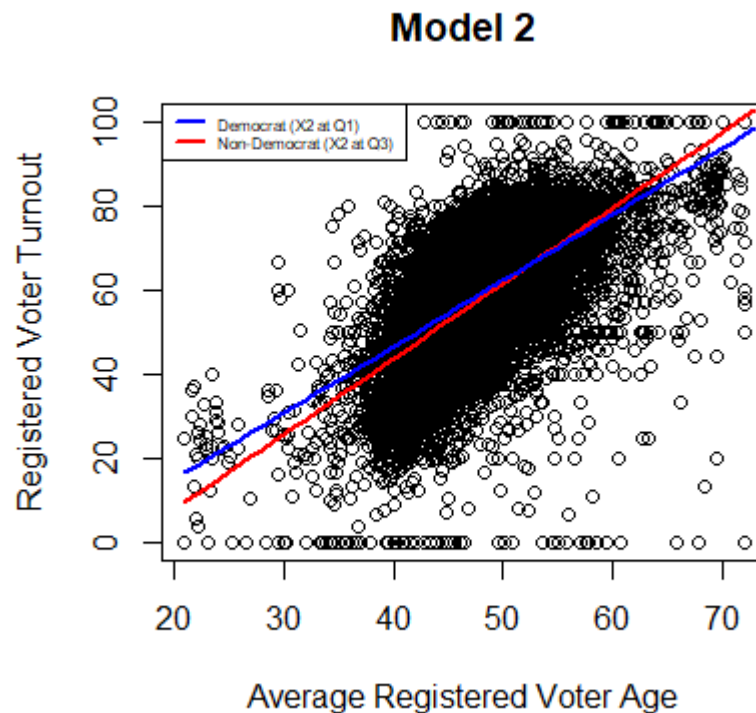
The graph demonstrates that at high values of *NonDemReg%*, where most voters are registered non-Democrats, turnout is lower in the younger registered voter precincts than in low values of *NonDemReg%*, where most voters are registered Democrats. For older precincts, the opposite effect proved true, where turnout was greater in high values of *NonDemReg%* than in low values of *NonDemReg%*.

Model 3 includes an additional control for the percentage of men registered in the precinct, which is statistically significant at the 10%, 5%, and 1% significance levels. The inclusion of this control captured some omitted variable bias that was uncontrolled in Model 1 and 2, as the effect is statistically significant but not meaningfully large. With only a .14% increase in adjusted R-squared, it barely accounted for much omitted variable bias, and all other coefficients did not meaningfully change with the inclusion of this control variable.
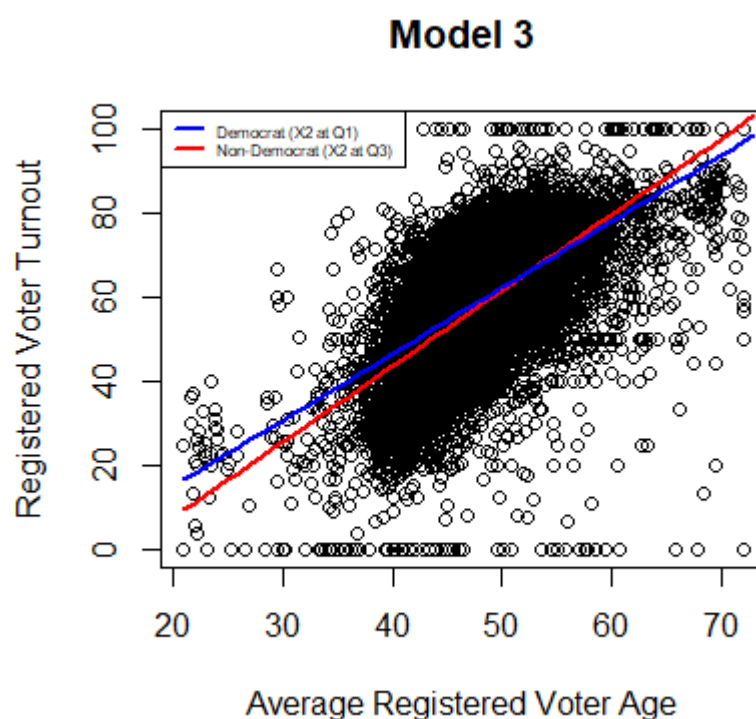
Discussion

My predictions were correct in hypothesizing that there would be a different effect of increasing average precinct age upon voter turnout by party affiliation proportion, but I was incorrect as to the direction of the effect. The results indicate that Democrats

voted in higher proportion than Republicans up until the age of roughly 50, and that non-Democrats voted in higher proportion at ages above 50.

There may be additional omitted variables contributing bias not yet controlled for that create this effect. One issue is that all precincts the data in are currently weighted equally, although the population of precincts is non-uniform. I predict that more urban, higher population precincts have comparatively more Democrats than smaller rural precincts, and hence turnout may be misrepresented in this analysis. Hence, a control for urbanization of the precinct may control for additional omitted variable bias not yet captured in the model.

This research demands further investigation before I can claim it disrupts the existing research. However, if these results are repeatable, it has contrary implications to the existing model of recall elections as a tool of accountability. In theory, this is a high urgency election for non-Democrats. I predicted that this urgency would increase young non-Democrat turnout due to the urgency outweighing the individual time costs of voting. However, contrary to my predictions, young Democratic voters had higher turnout than non-Democrats. This suggests that the recall election may not have been viewed as urgent by non-Democrats, who perceived that they would not win the recall effort. Instead of being a tool of accountability, the California recall election of 2021 may have been used as a tool of defamation against the Governor. If this is true, my results suggest raising the threshold required to initiate a recall, from 12% to a figure considerably higher. This election cost millions of dollars to hold, and hence only credible recalls should pass the minimum petition signature threshold. This recall could

set a precedent for future flimsy recall efforts, costing the state of California more time and money in the future.

Conclusion

My research question of interest is, "Did age have a unique effect upon voter turnout by political party in the 2021 California Special Election to recall Governor Gavin Newsom?" After running three models of multivariate ordinary least squares regressions, the interaction effect of age upon registered voter turnout rate caused by increasing proportions of registered voters of non-incumbent parties is statistically significant, suggesting that age did have a unique effect upon voter turnout by political party for the 2021 California Special Election. There are several proposed efforts to improve the internal validity of this study before its findings are generalizable, but the research demands a reconsideration of the low 12% threshold to initiate a recall election in California.

References

Alvarez, R. M., Melanie Goodrich, Hall, T. E., Kiewiet, D. R., & Sarah M. Sled. (2004). The Complexity of the California Recall Election. PS: Political Science and Politics, 37(1), 23–26. http://www.jstor.org/stable/4488756

California governor election results 2018: Live midterm map by county & analysis. POLITICO. (1970, November 7). Retrieved October 30, 2022, from https://www.politico.com/election-results/2018/california/governor/

Haman, M. (2021). Recall Elections: A Tool of Accountability? Evidence from Peru. Revista Desarrollo y Sociedad, (87), 73–111. https://doi.org/10.13043/dys.87.3

Harder, J., & Krosnick, J. A. (2008). Why Do People Vote? A Psychological Analysis of the Causes of Voter Turnout. Journal of Social Issues, 64(3), 525–549. https://doi.org/10.1111/j.1540-4560.2008.00576.x

Segura, G. M., & Fraga, L. R. (2008). Race and the Recall: Racial and Ethnic Polarization in the California Recall Election. American Journal of Political Science, 52(2), 421–435. http://www.jstor.org/stable/25193822

Williams, J. (2021, October 27). Results of Newsom recall certified; 12.8 million total votes cast. Ballotpedia News. Retrieved October 30, 2022, from https://news.ballotpedia.org/2021/10/27/results-of-newsom-recall-certified-12-8-million-total-votes-cast/

2021 Special Election Precinct Data. Statewide Database. (2021). Retrieved December

14, 2022, from https://statewidedatabase.org/d20/s21.html