Bella Davies and Devin Clark

Professor Shulamit

QST QM 222

February 21, 2023

Project 1 Report

In this project, we examine the “Bond and Tax Election Dataset,” uploaded to Kaggle by Ali Farahmandfar, sourced from the California State Treasurer. The dataset contains information about bond and tax elections in the state of California between 1986 and 2021. The information includes the agency county, the agency name, type of tax/debt, amount of bond/tax, purpose, percent of “yes” votes, percent of “no” votes, the result (pass/fail), election year, election type, and election date (Figure 1). There are 5,977 observations in the dataset, each representing a bond or tax election in California. To clean up the data, we dropped four observations. Three of these observations had zeros for the percent of “yes” and percent of “no” votes; one had n/a. After dropping these four observations, we were left with 5,973 rows of complete observations. The values of “Measure” were dropped since we are not using this information for our analysis. Additionally, "Election Date" and "Agency Name" were removed because we viewed them as duplicates of "Election Year" and "Agency County," making them unnecessary for our analysis. After cleaning the data, the descriptive statistics of the numerical variables, “% Yes,” “% No,” and “Election Year,” can be seen in Figures 2, 3, and 4, and the distributions of the categorical variables can be found in Figures 5-10.

When looking at the individual variables, we had a few interesting observations about some of the variables. In particular, when looking at the distribution of the “Agency County” values, the counties with the most tax elections were Los Angeles, Santa Clara, and San Diego

(Figure 5, 14). As these counties are highly populated urban areas, this is expected as they would likely have more taxes and tax elections than other less populated counties. Another interesting distribution of values was from the "Result" column. It is important to note that there were more passes than fails, with 65% of tax elections passing, around double the amount that failed (Figure 8). It is interesting to see the percentages on the distribution of results since we know that California is a state with many taxes. We also noticed an alternating pattern in the distribution of the "Election Year" values, with the number of referendums in even-numbered election years far outpacing the number in odd election year ballots (Figure 9). This is likely because voter turnout is always higher in even-year elections since they get more publicity. Another interesting variable was the election type; out of the 5,973 observations, 60% were on the ballot in a general election, 23% in a primary, 16% in a local election, and 1.54% were in a special election (Figure 10). This aligns with the sizes of these elections, as general elections have a higher turnout rate than primary, local, and special elections.

When examining the relationships between the variables, we chose to look at three relationships comparing different variables with the result of the elections, the first being the type of tax/debt and the result. Most observations were either PLF debts or general obligation bonds. These categories covered tax implementation and distributing money, respectively, and as expected, the latter has a narrower margin between pass and fail than the former (Figure 11). This is most likely because PLF debts were either raising or lowering taxes, leading to fail or pass votes, respectively, while for the bonds, handing out money is always popular for the jobs and benefits it generates regardless of where in the county it is going.

Another interesting relationship was between the purpose of the tax and the result. The purpose with the most passes was taxes for K-12 school facilities or college and university

facilities (Figure 12). It is nice to see that a large majority of taxes in California are going toward the school systems and education. It is also important to note the purposes that fail more often. These purposes included public safety, public buildings, recreation and sports facilities, street construction, and prisons/jails (Figure 12). The relationship between the purpose of the tax election and its result is important in identifying the priorities of California and where the state tends to direct money.

Finally, we examined the relationship between the result and election type. Most of these referendums were on general election ballots, which we expected since they are the highest-turnout elections. Though unexpectedly, for every election type, there were more passed referendums than failed ones (Figure 13). From this relationship, we can infer that this is because voters are more determined to turn out to vote if they strongly support or oppose the measure on the ballot. The shrinking turnouts mean this margin is narrowing due to voter awareness.

In conclusion, we have found several interesting observations and relationships in this dataset, which would have important implications for taxes in California. This dataset can be used to answer a variety of questions regarding California tax elections and their results. Voters may often wonder: Where is the money going? What is the money used for? How likely is a tax/bond ballot to pass? To learn where the money is going, one could look at the distribution of the values of Agency County (Figure 5, 14). To see what the money is being used for, we could look at the distribution of the purpose values and the relationship between purpose and result (Figure 7, 12). To see how likely it is that a referendum passes, one could look at the relationship between any of the variables and the result, or the distribution of results (Figure 8). Anyone from voters to strategists will find this dataset beneficial in answering these types of questions and providing clarity on how specific counties have historically voted in tax and bond referendums.

Appendix

I. Figure 1. Variable Definitions

Variable	Definition
Agency County	The county the referendum took place in.
Agency Name	The specific state level agency or organization that benefits from the change in tax or approval of a bond.
Type of Tax/Debt	The type of tax/debt being voted on
Amount of Bond/Tax	A description of what the measure actually does if it is implemented.
Purpose	Purpose of the tax/debt
% Yes	Percentage of yes votes in a particular election
% No	Percentage of no votes in a particular election
Result (Pass/Fail)	Result of the election, whether it passed or failed
Election Year	The year the election took place in.
Election Type	The type of election: general, primary, local, or special

II. Numeric Variables

Figure 2. "% Yes" Descriptive Statistics

% Yes	
Mean	62.1
Median	63.7
Standard Deviation	13
Minimum	5.8
Maximum	100
Missing Observations	0
Count	5973

Figure 3. "% No" Descriptive Statistics

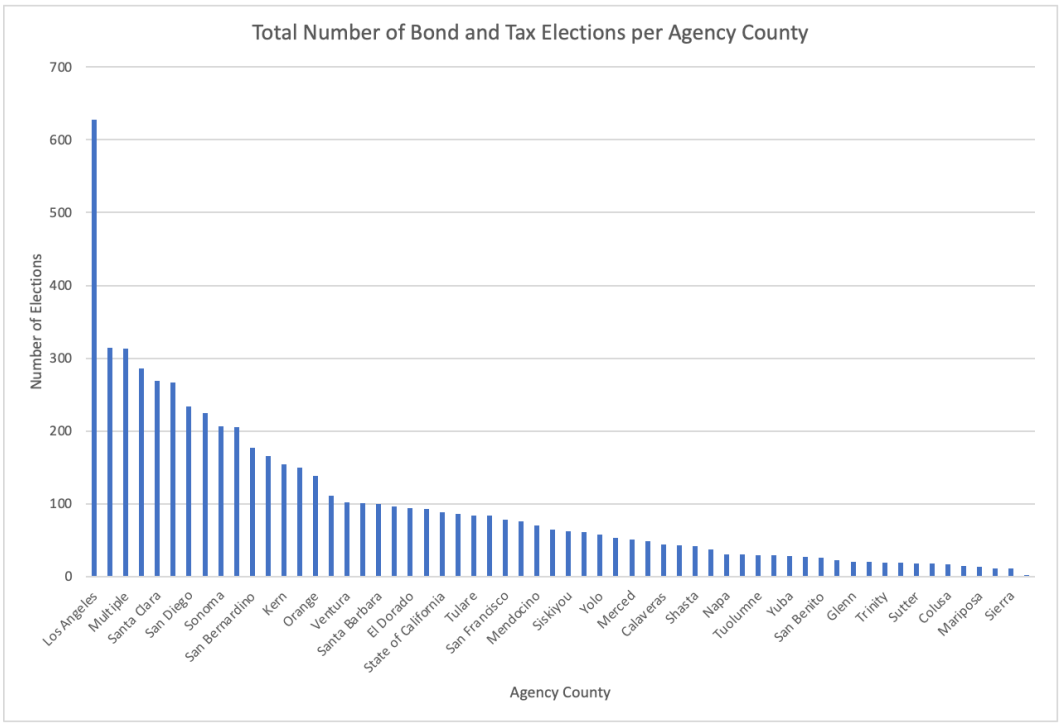
% No	
Mean	37.9
Median	36.3
Standard Deviation	13
Minimum	0
Maximum	94.2
Missing Observations	0
Count	5973

Figure 4. "Election Year" Descriptive Statistics

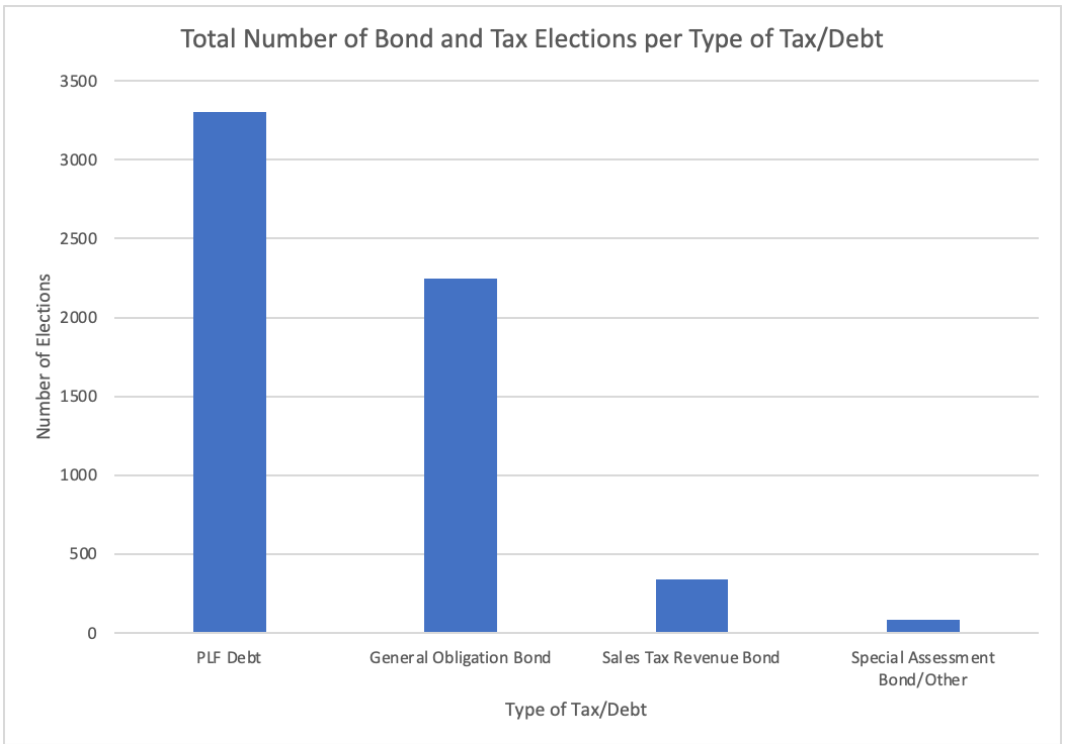
Election Year	
Mean	2008
Median	2009
Standard Deviation	8.5
Minimum	1986
Maximum	2021
Missing Observations	0
Count	5973

III. Categorical Variables

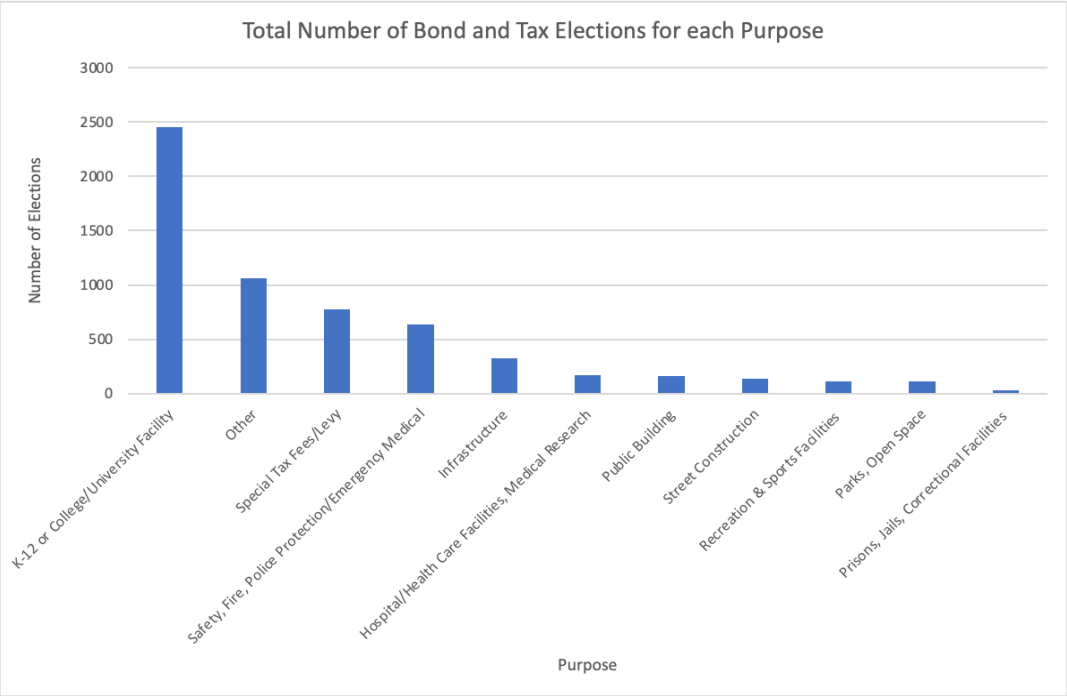
- Figure 5. Distribution of "Agency County"



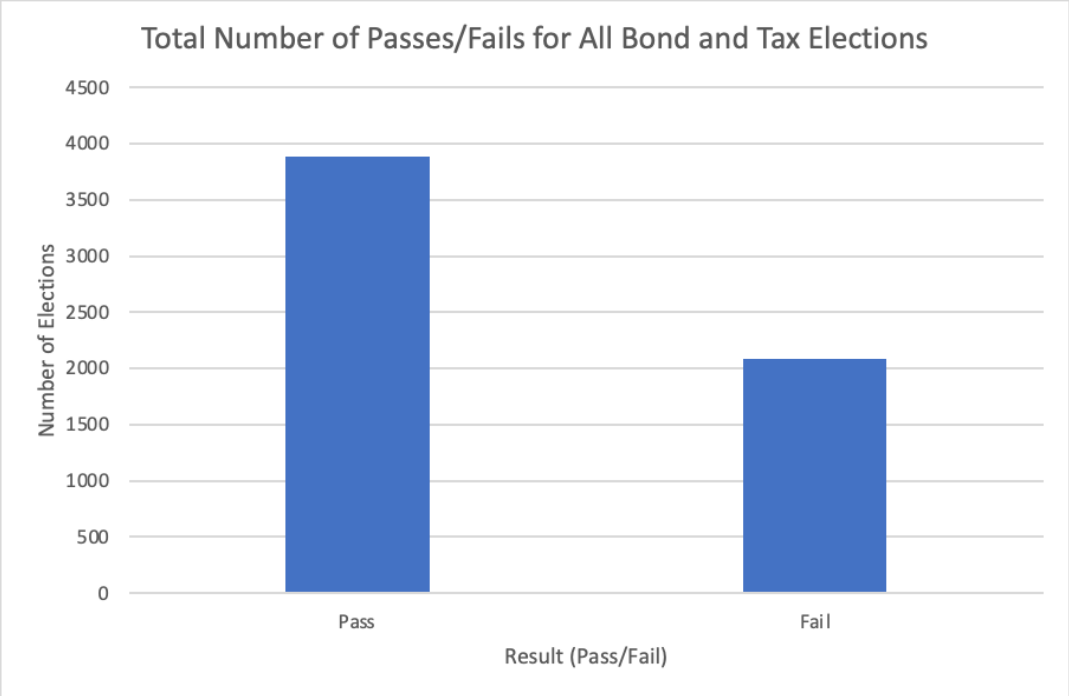
- Figure 6. Distribution of "Type of Tax/Debt"



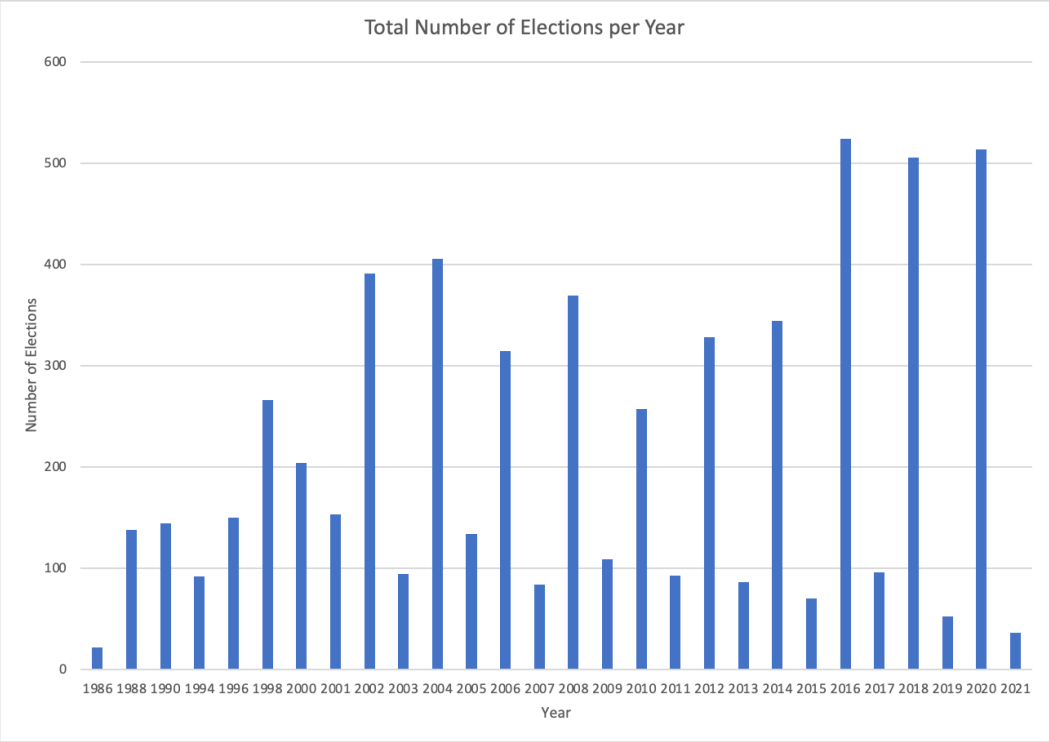
- Figure 7. Distribution of "Purpose"



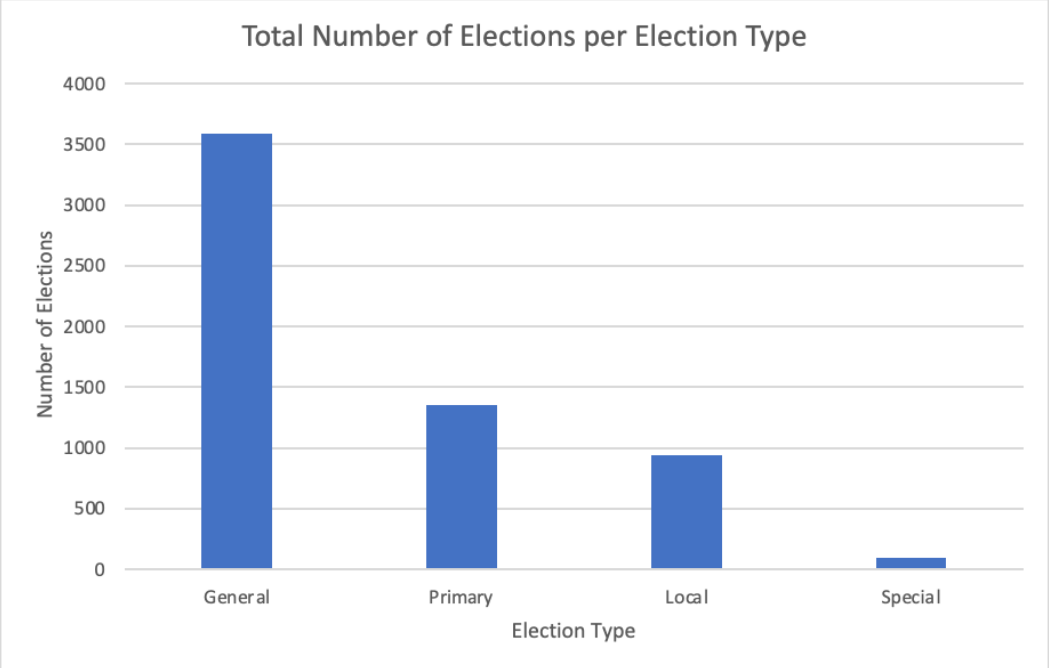
- Figure 8. Distribution of "Result"



- Figure 9. Distribution of "Election Year"

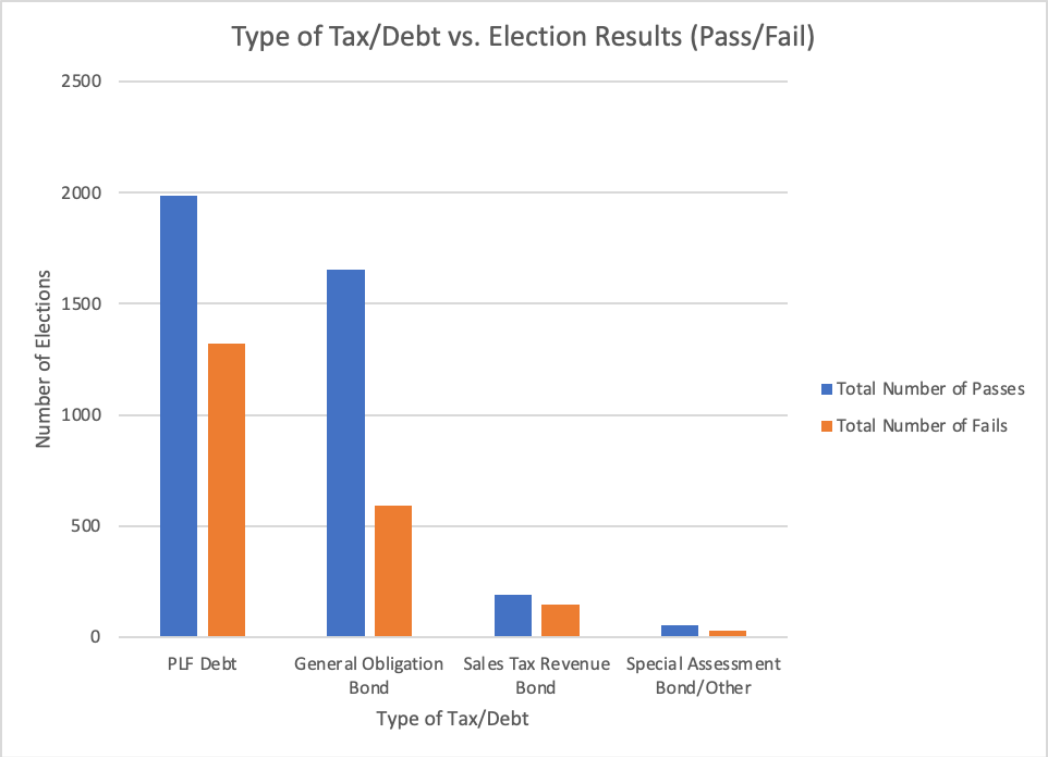


- Figure 10. Distribution of "Election Type"

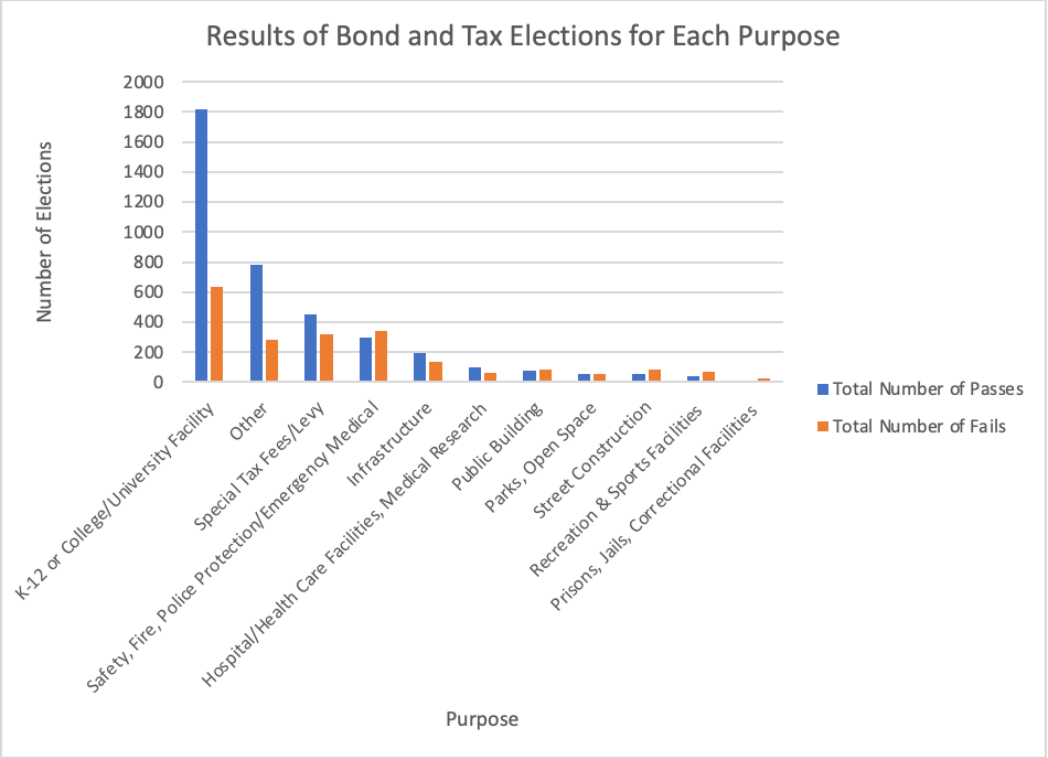


IV. Interesting relationships between variables

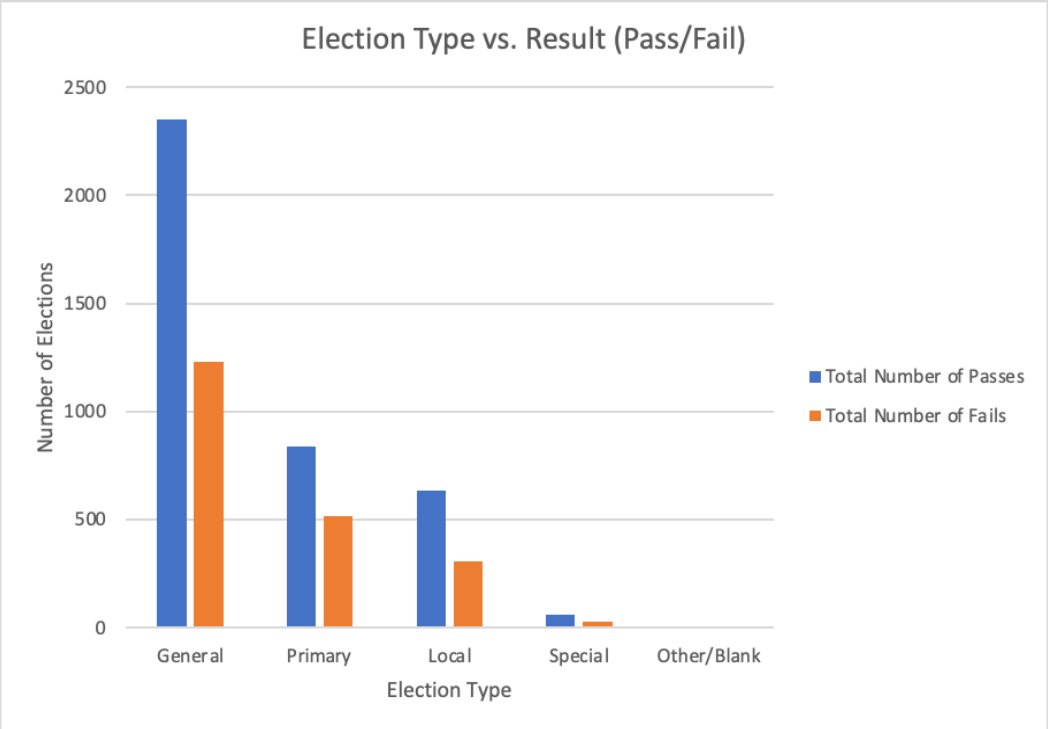
- Figure 11. Tax Type vs. Result



- Figure 12. Purpose vs. Result



- Figure 13. Election Type vs. Result



- Figure 14. Agency County vs. Number of Elections