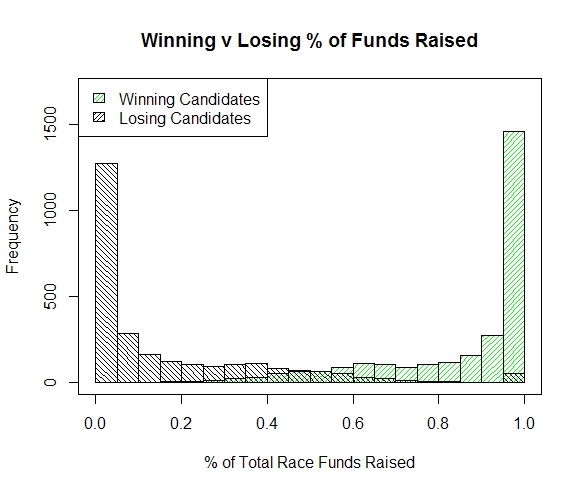
Question 1: Do campaign contributions predict electoral success?

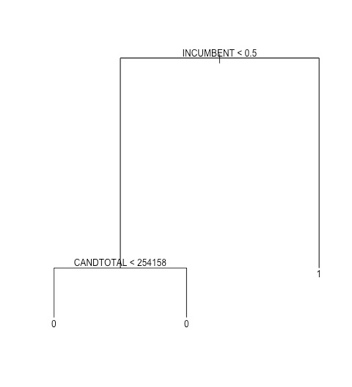
Our analysis found that campaign contributions are highly correlated with electoral success, both in terms of amount of funding and percentage of total funding in a given race received. The correlation coefficient between the percentage of funds received and the percentage of votes received is .87; for every additional percentage point of funds a candidate raised, he/she received an additional .87% of votes. In the graph of percentage of funds raised, it is clear that there are stark differences between the winning and losing candidates, with only a small bit of overlap.



This result was supported by the frequent itemset mining analysis. For this portion of the analysis, funds raised was binned into four categories (very low, low, mid-high, and high). Candidates who received a very low amount of funding tended to be election losers, while candidates who raised a high amount of funding tended to be election winners.

After concluding that campaign funding was highly correlated with electoral success, we ran some predictive analyses to determine how effective the campaign funding variable was in determining election winners. We set up a logistic regression model in which the winner variable was determined solely by the amount of money raised and the number of industries that contributed to a candidate, and found that these two variables alone gave us a classification rate of 81.8%, with most of the predictive power coming from the funds raised variable. We were a bit surprised by the high classification rate; although we came into the project with the idea that funding played a role in election outcomes, we were not expecting it’s impact to be so large.

We wanted to see if we could develop an even more accurate predictive model, so for the decision tree analysis we added the incumbency status variable. The prediction accuracy improved to about 90%, but, as the tree below shows, the incumbent variable dwarfed both other variables; once the model had a candidate’s incumbency status, nothing else was important in determining the outcome of the election. Basically, what is happening here is the incumbency variable subsumes the contribution amount; incumbents raise higher amounts of funding than non-incumbents, and since the incumbent variable is such a good predictor of winners on its own (over 90% of election winners in our dataset were incumbents), adding more info on top of the incumbency status doesn’t add any predictive power.



However, the fact that incumbency status is a better predictor of electoral success than amount of funding does not mean funding is meaningless. It is abundantly clear that the more money a candidate raises, the higher that candidate’s chances of being elected. An interesting extension of this analysis would be to closely examine outliers; candidates who raised a great deal of money yet still lost and candidates who raised very little money but were elected anyway, to see how they differ from their counterparts.