Final Report

Data Overview

The datasets that we used came from the Federal Election Commission and FiveThirtyEight, which were both provided in the project guidelines. For the causal inference analysis, we used the 2018 Primary Candidates Endorsements dataset from FiveThirtyEight. Since the data contained information of the whole population of candidates, we deemed this dataset to be a census, not a sample. For the Multiple Hypothesis Testing analysis, we used a total of six datasets, all retrieved from the FEC's 2022 Campaign Financing Data. Since the website contained a massive amount of bulk data on all of the campaigns, we deemed this to be a census as well.

Upon EDA, we saw that the dataset used for our causal inference analysis was missing the race of 20% of the candidates. We decided not to impute race (as this is incredibly difficult), so we dropped 20% of the dataset. However, we still decided to include race as a feature for the analysis, since we ultimately believe that race is an extremely important confounder for our analysis. This may have caused selection bias in the data. Furthermore, we did not include attributes encoding endorsements. This was because the majority of candidates were neither endorsed nor explicitly not endorsed by various groups. Thus, 80% of candidates were recorded as null values. Because a lack of opinion is not equivalent to active support or opposition, we did not include endorsements in our analysis. Extraneous columns were excluded either because they are not confounders or because they are redundant. As for granularity, each row in the dataset represents a primary candidate from the 2018 election, with information on their name, gender, race, party affiliation, state, primary votes percentage, STEM background, and more.

To conduct the multiple hypothesis analysis within a relevant scope of the Bipartisan Campaign Reform Act (BCRA), we extracted all "Individual Contributions" data from 1997 to 2006 and also included the data on all committees. Furthermore, since the current political climate is dominated by the Democratic and Republican parties, we filtered the data down to only these two parties. Therefore, we must preface that this may or may not have caused selection bias in the data. The reason why we included the committee's data was because it had information on each of the committees, whereas the individual contributions data only had information about which committee received it. We joined the two tables together and extracted columns we needed, including transaction amount, party affiliation, transaction dates, and committee names/ids. For data cleaning, we removed negative transaction amounts because we were unsure of what they represent (debts, loans, refunds, etc.). Additionally, we formatted transaction dates into the datetime datatype. Moreover, we removed tuples with no transaction date because date is crucial to our questions. We created a column categorizing each transaction as either 'Before' or 'After' the BCRA took effect. In total, we were left with around six million rows of data, post-processing. Each row represented a donation to a committee, with information on committee ids and names, transaction amount, party affiliation, transaction year, whether it was before or after the BCRA, and more.

Research Questions

Causal Inference Question

Does coming from a STEM background have a causal impact on the percentage of votes won by Democratic candidates?

Causal inference is a good fit for this question because it directly addresses the causal relationship between two variables, given that we can account for confounders. This question is asking for a causal relationship between primary votes and STEM background, which is perfect for causal inference. The only limitation is the inherent limitations of causal inference itself. It is difficult to account for confounding variables for a question like this, so the only plausible direction is to apply methods like IPW or matching to try and eliminate bias.

Multiple Hypothesis Questions

Q1. Did the BCRA have an effect on individual contributions? (Tests 1-3)
Q2. Is there a meaningful difference in when donations are made during election years versus non-election years? (Tests 4-7)

Multiple hypothesis testing is well-suited for these research questions because it enables the examination of multiple aspects of the data while controlling for false positives. For Q1, it can assess the BCRA's effects across different metrics of individual contributions. For Q2, it can identify differences in donation timing between election and non-election years. Its strengths include flexibility and robust error control, but its limitations include reduced power with conservative corrections, potential issues with correlated tests, and difficulty in isolating causality, particularly for the BCRA's effects. These challenges can affect the reliability of the results if not carefully addressed.

Prior Work

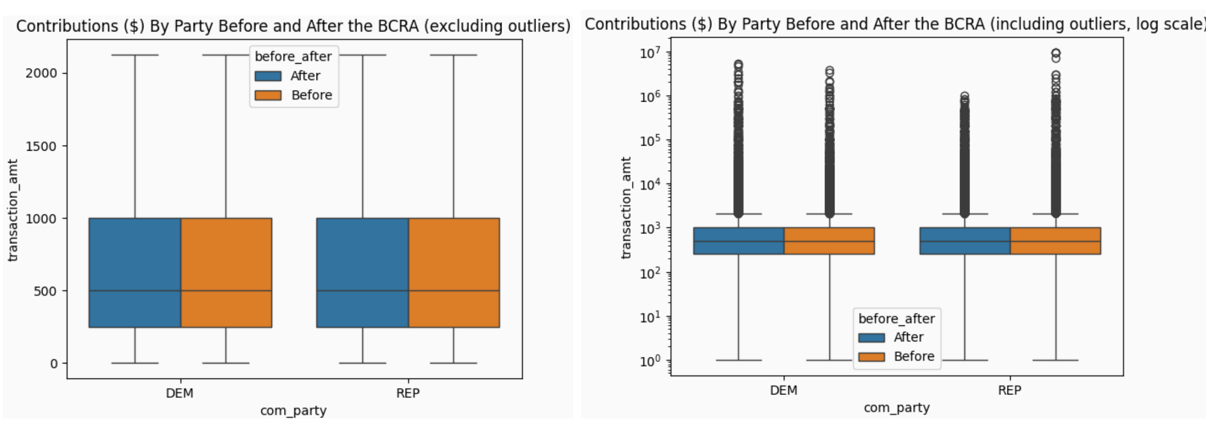
Krasno, Jonathan S, and Frank J Sorauf. "Evaluating the Bipartisan Campaign Reform Act." *Review of Law and Social Change. New York University*, vol. 28, no. 1, 1 Jan. 2003, pp. 121–181, www.researchgate.net/publication/338477546_Evaluating_the_Bipartisan_Campaign_Reform_Act.

Krasno and Sorauf explore the impacts of the BCRA in this sweeping and in-depth review of campaign finance. This article is a useful touchstone for our multiple hypothesis question, but provides little guidance on data methods or analysis as it is largely based on political science theory.

Murakami, Go. “Candidates’ Ethnic Backgrounds and Voter Choice in Elections.” *Library.ubc.ca*, 17 Dec. 2024, open.library.ubc.ca/soa/cIRcle/collections/ubctheses/24/items/1.0166940. Accessed 16 Dec. 2024.

Murakami conducts a parallel study to ours on the causal effect of ethnicity on voter choice in elections in Canada and Japan. Though useful as a similar theoretical framework for our causal inference question, Murakami created his own surveys to produce data, rather than working with observational data. As such the methods used cannot inform our observational data analysis.

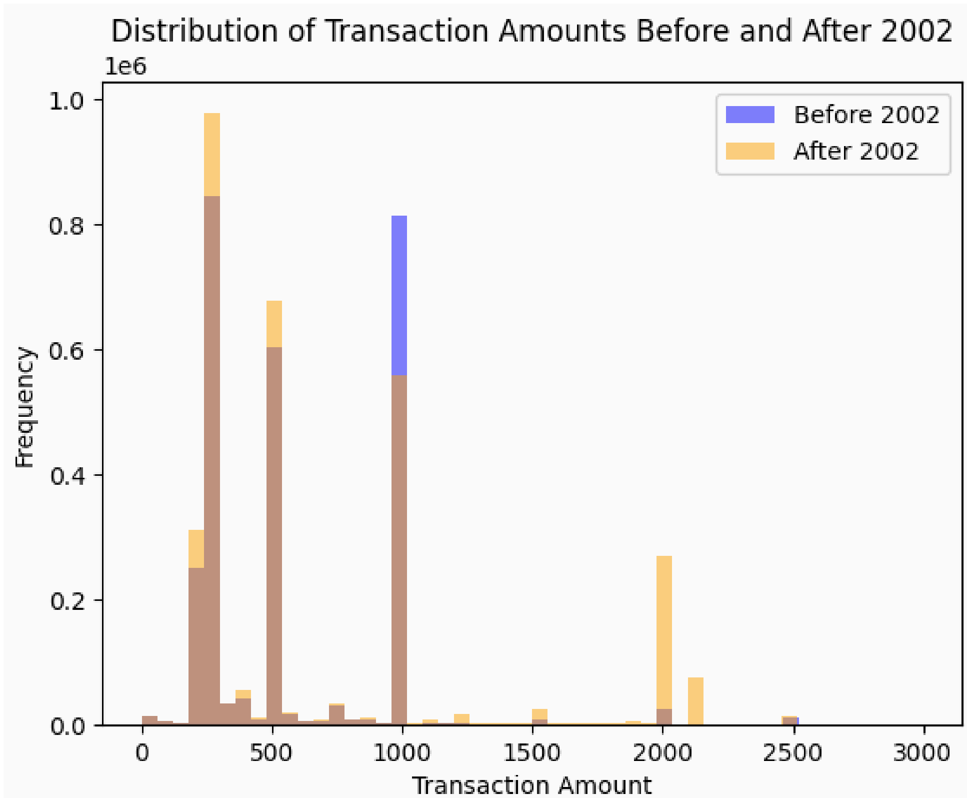
EDA




```
Average Democrat Contribution Before: 745.5963110857577 Average Democrat Contribution Before: 900.3784463737035
Average Republican Contribution Before: 711.2438791051231 Average Democrat Contribution Before: 910.8101448094283
```

Figure 1. Boxplots of transaction amounts and averages by party before and after the BRCA. The left diagram excludes outliers, while the right includes them.

These boxplots demonstrate that the median and quartiles are roughly distributed the same before and after passing the BRCA. However, upon checking the change in average transaction amount, we found that the average increases by over $150 for both Democrats and Republicans. This motivates some of our multiple hypothesis testing questions, which ask whether this difference in average transaction amount is significant or not.




```
length before trim: 6004416, length after trim: 5894896, percent data retained: 0.9817600912395144
```

Figure 2. Overlaid histograms of all transaction amounts less than $3000 before and after 2002.

This diagram shows that in general, the number of contributions decreases as the size of the contribution increases. It is expected that we see large peaks at round numbers, since people are more likely to donate a round number. It is also notable that for every visible peak besides $1000, the frequency of transaction amounts increased after 2002. This informs our question on whether individual transaction amounts changed after BRCA.

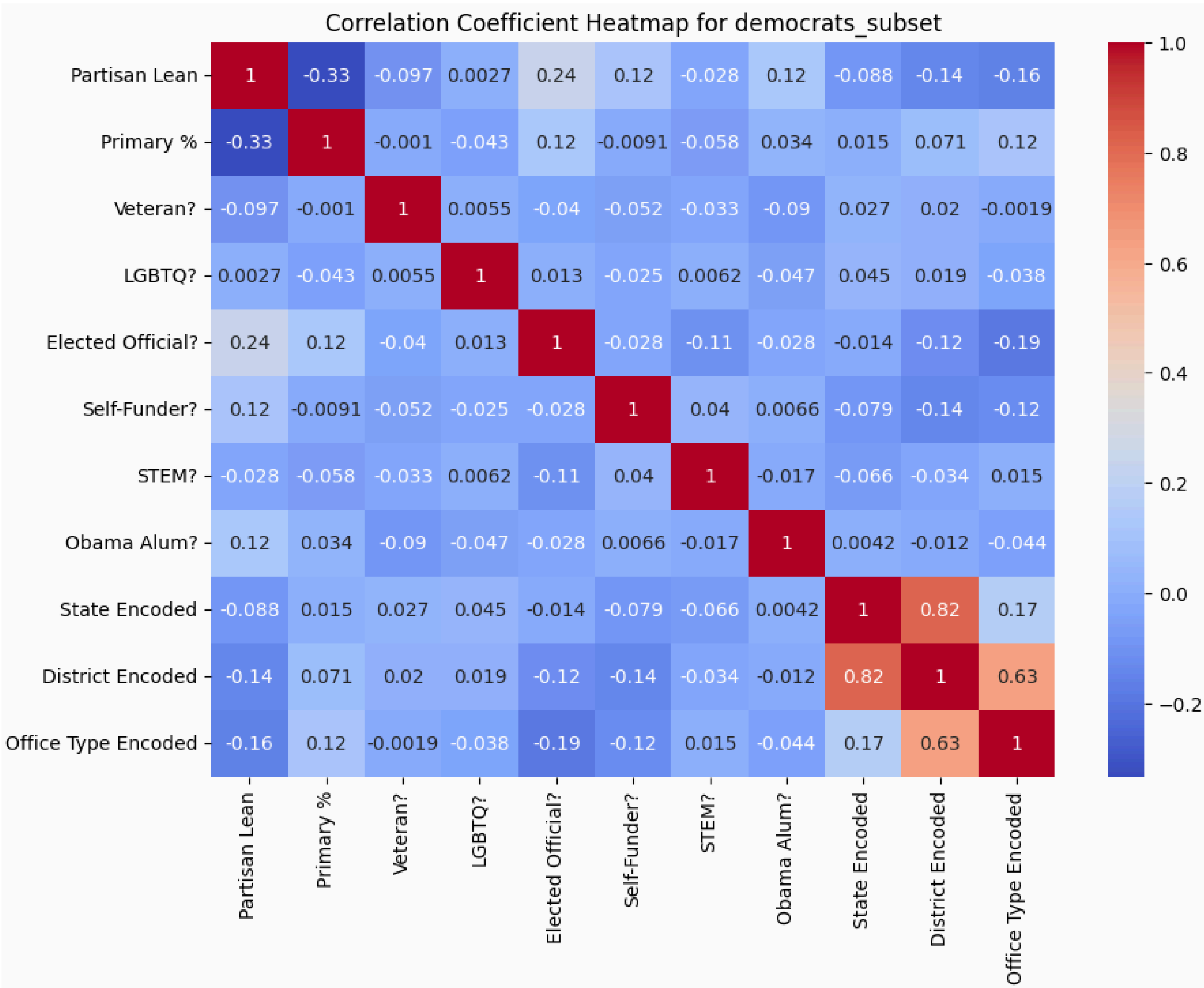


Figure 3. Heatmap of correlation coefficients for the Democrat candidate dataset.

This heatmap gives us an idea of what correlation having a STEM background has with other categorical variables, which informs our decisions for what columns to include in our logistic regression for IPW. This heatmap also shows a positive correlation between district and state, because what district a candidate is from encapsulates data about their state.

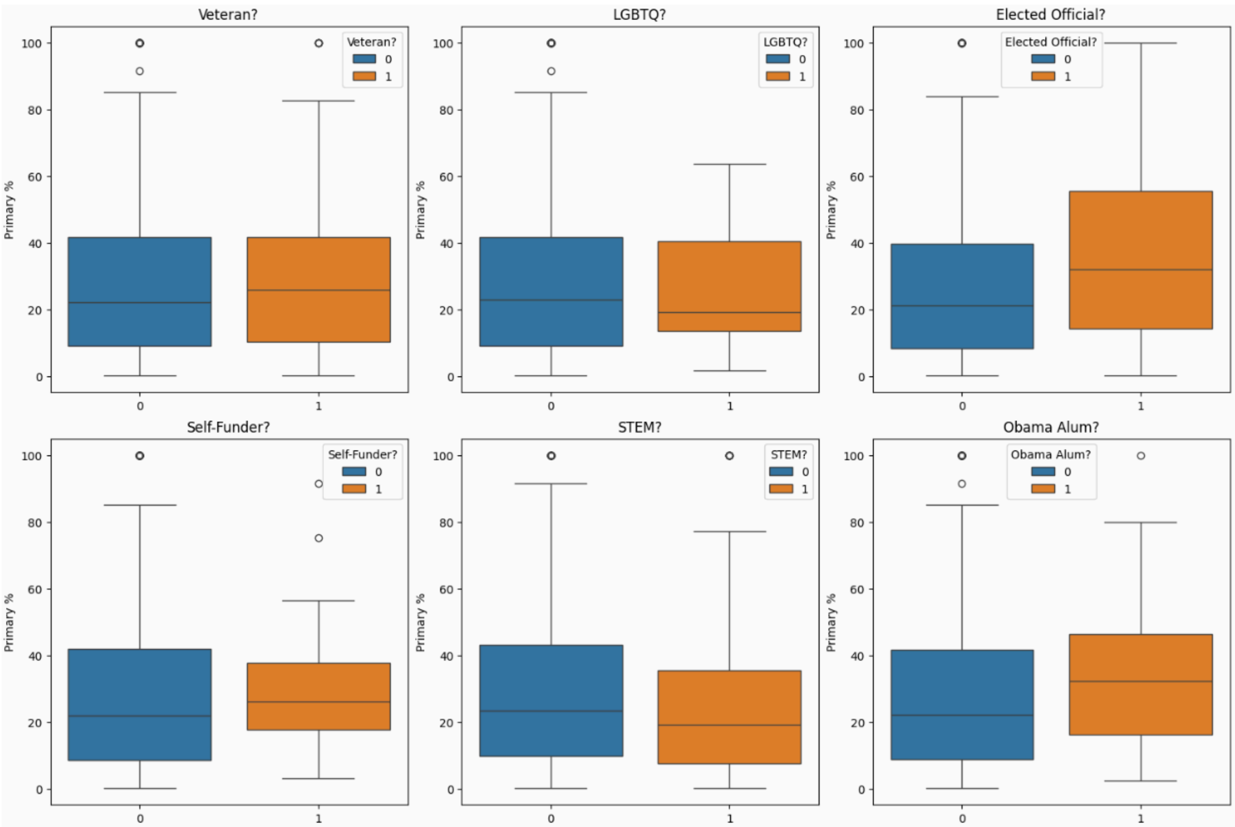


Figure 4. Boxplots of Primary % for each categorical variable.

These boxplots show shifts in primary % for categorical variables of interest, giving us an idea of whether a variable had a positive or negative effect on a candidate's votes. A relationship that is interesting to us is the seemingly negative effect a STEM background has on a candidate's popularity, since we would have predicted a positive effect.

Causal Inference - Research Question 1

Our causal inference question is: Does coming from a STEM background have a causal impact on the percentage of votes won by Democratic candidates? The 2018 Primary Candidates Endorsements dataset from FiveThirtyEight features data from the 2018 Primary Election for Democrats and Republicans, but only Democratic candidates have data on whether or not they come from a STEM background (captured by the 'STEM?' binary feature). As such, candidates with values of 'STEM?' equal to 1 represent our treatment group, and 0 our control group.

While many of these features were useful, we decided to omit several of them from our causal inference question on the bases described in Table 5 below.

'Candidate'	No	The candidate name is used as PID, but it is not relevant for our use of inverse propensity weighting
'State'	Yes	The state is related to both having a STEM background and your primary %. Some states have better public education and a stronger emphasis on STEM education. If you run in a really red state then the state impacts your primary %.
'District'	Yes	Same reasoning to the reasoning why the state impacts having a STEM background and primary %, but applied to the district you are running in.
'Office Type'	Yes	The demographics for each type of office are different. For instance, the Senate tends to be older, wealthier, and whiter. For that reason, this could be an informative feature for whether a candidate has a STEM background.
'Race Type',	No	This attribute indicates whether or not the race was a “regular” or “special” election. The special elections only account for 1.5% of data, so it is not assumed to confound significantly.
'Race Primary Election Date'	No	The election dates are assumed to not confound significantly.
'Primary Status'	No	We are eliminating this column because we are using Primary % as the outcome.
'Primary Runoff Status'	No	This is related directly to Primary Status, so we are not using this column.
'General Status'	No	This is related directly to Primary Status, so we are not using this column.
'Partisan Lean'	Yes	If you come from a historically Republican state, then there is a high correlation between red states and their funding and emphasis on public education. Thus, the partisan lean can definitely impact whether or not someone pursues a STEM field. This can also impact the primary % because if you come from a red state then it’s more likely they won’t vote for a democratic nominee, and vice versa.
'Primary %'	Yes	This is the outcome.

'Won Primary'	No	This is directly related to primary status, so we are not using this column.
'Race'	Yes	Upon EDA, we can see that the dataset is missing the race of 20% of the candidates. We decided not to impute race (as this is incredibly difficult), so we have decided to drop 20% of the dataset. We understand that this could potentially reduce statistical power and may lead to biased results if the missingness is not random. However, we ultimately believe that race is an extremely important confounder for having a STEM background and primary %, thus we have decided to include it.
'Veteran?'	Yes	There is likely a causal effect between veteran status and STEM background. For instance, those who served as young adults are less likely to have a higher education, or to have education provided by the military in a specific field.
'LGBTQ?	Yes	There is a relationship between being LGBTQ and having a STEM background. There could be many reasons but one particular one might be that many LGBTQ individuals migrate to urban areas for greater acceptance and community, where STEM education and career opportunities are often more abundant. Being LGBTQ can also impact how the primary % they receive.
'Elected Official?'	Yes	Since the majority of people in politics have backgrounds in the humanities, being an elected official could be informative for determining if someone has a STEM background.
'Self-Funder?'	Yes	Having a STEM background could impact how much an individual makes and thus affect an individual's ability to fund their own campaign.
'STEM?	Yes	This is the treatment variable.
'Obama Alum?	Yes	Being an Obama alum means that a candidate has a history of being involved in politics. Since the majority of people in politics have backgrounds in the humanities, being an Obama alum could be informative for determining if someone has a STEM background.

Table 5. Reasoning behind inclusion/exclusion of features from primary dataset.

As a general precaution, we decided not to include the attributes pertaining to whether or not a candidate had received an endorsement from politicians or issue organizations. This was because the majority of candidates were neither endorsed nor explicitly not endorsed by various groups. Thus, 80% of candidates were recorded as null values. Because a lack of opinion is not equivalent to active support or opposition, we did not include endorsements in our analysis.

Causal Inference Question — Methods

The causal relationship between STEM background and their primary vote percentage can be subject to a multitude of potential confounders, making it imprudent to assume 'unconfoundedness.' Moreover, we have excluded endorsements from our list of potential confounders, despite their possible significant impact on this causal inference question. Therefore, we cannot say that the unconfoundedness condition holds.

To adjust for confounders, we decided to use Inverse Propensity Weighting (IPW). Originally, we thought that "matching" would be a more optimal algorithm to deploy on this causal inference study. However, due to the lack of data, matching seemed undesirable, which led us to the conclusion of using IPW to control for confounders.

We assume that there are no colliders in the dataset. This is because our outcome variable, the percentage of votes a candidate received, is an endpoint of the election measured after all other features in the dataset have been determined. Therefore, because it is not possible for 'Primary %' to affect the other variables, there are no colliders.

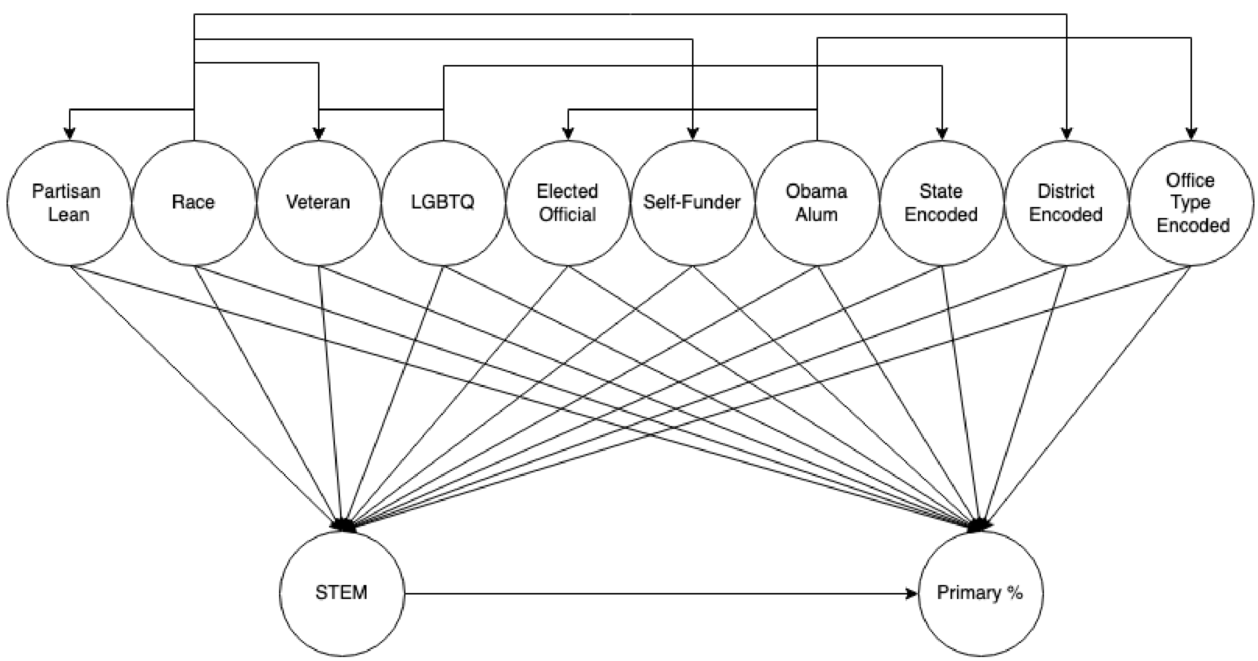


Figure 6. Causal directed acyclic graph of confounding relationships in our model.

Causal Inference Question — Results

After trimming propensity scores, our IPW model found that having a STEM background causes a 7.05 percentage point decrease for Democrats in Primary elections. This might be because people with STEM backgrounds are less likely to be career politicians and may have less political expertise as their peers. However, our model is not statistically significant enough to trust this result. In Figure 7, none of the features in the logistic regression model predicting STEM background are statistically significant, each containing 0 in their 95% confidence intervals.

```
Optimization terminated successfully.
         Current function value: 0.568809
         Iterations 28
```

Logit Regression Results

Dep. Variable:	STEM?	No. Observations:	54
Model:	Logit	Df Residuals:	46
Method:	MLE	Df Model:	7
Date:	Sun, 15 Dec 2024	Pseudo R-squ.:	0.1488
Time:	19:03:46	Log-Likelihood:	-30.716
converged:	True	LL-Null:	-36.085
Covariance Type:	nonrobust	LLR p-value:	0.1504

	coef	std err	z	P>|z|	[0.025	0.975]
Partisan Lean	-0.0774	0.095	-0.812	0.417	-0.264	0.109
State	0.1720	0.481	0.357	0.721	-0.771	1.115
District	-0.3243	0.218	-1.490	0.136	-0.751	0.102
Race	0.6802	0.663	1.026	0.305	-0.619	1.979
Veteran?	1.0218	0.952	1.074	0.283	-0.843	2.887
LGBTQ?	-293.8434	4.15e+63	-7.08e-62	1.000	-8.13e+63	8.13e+63
Elected Official?	0.5712	1.275	0.448	0.654	-1.927	3.070
Self-Funder?	1.1004	0.898	1.226	0.220	-0.659	2.860

Figure 7. Summary of logistic regression model used to calculate propensity scores.

Consulting distributions of the propensity scores produced for treatment and control groups illustrates the inadequacy of the data used in our model. If our IPW model was sound, we would expect control group propensity scores to cluster around 0, meaning the features in our data accurately predict STEM background. Instead, the distributions of propensity scores across our treatment and control groups do not look that different. Looking at these distributions, it is not surprising why the features are statistically insignificant.

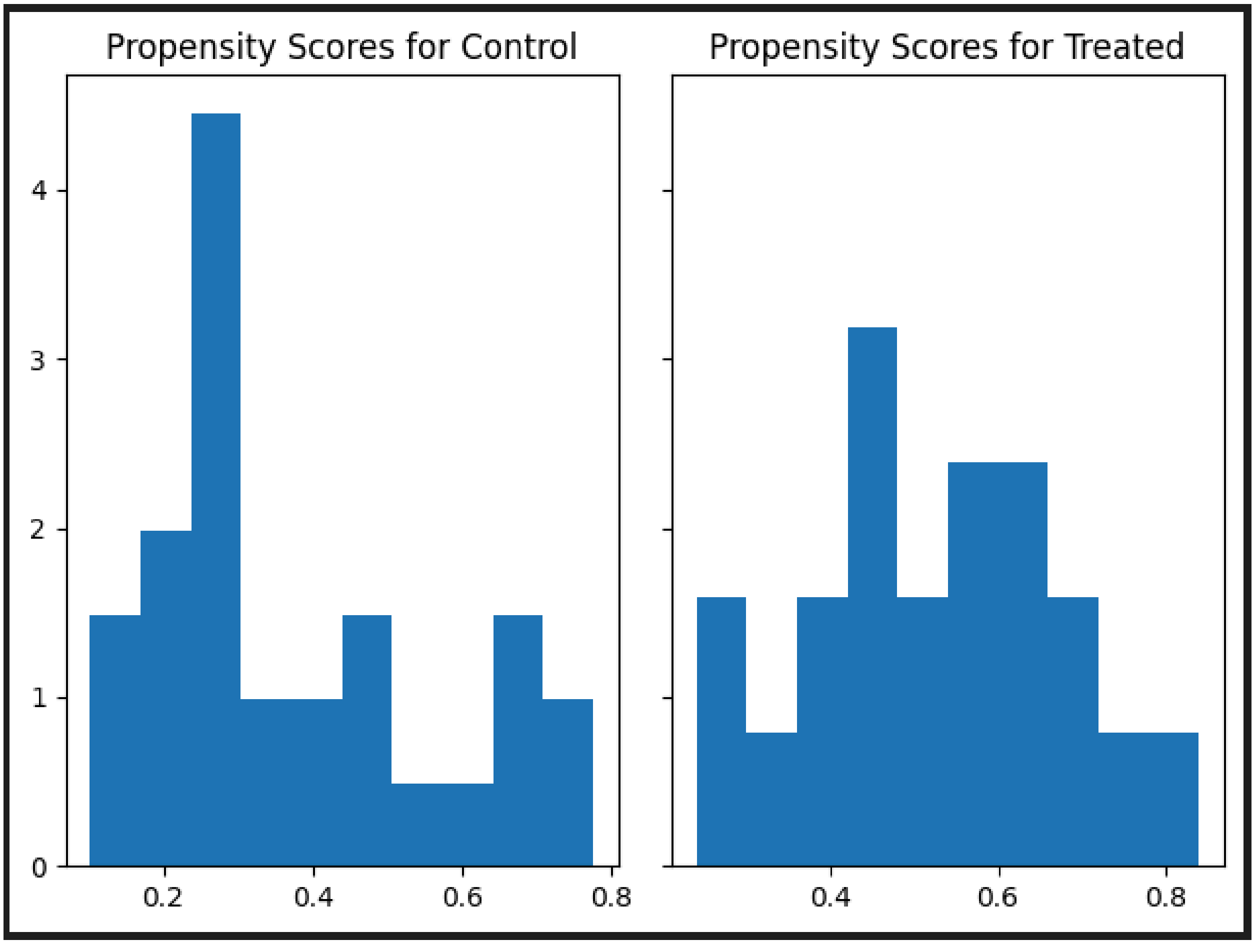


Figure 8. Distributions of propensity scores for treatment and control groups.

Causal Inference Question — Discussion

Our methods were limited by the data that we used. The features that we decided to include in our model did not provide enough variation and correlation to accurately predict if candidates had a STEM background, and as such we failed to produce a statistically significant model. Ideally, we would have had more features to choose from to better answer this question. For instance, data on a candidate's highest level of education, previous occupational history, or evidence of a bachelor’s or advanced degree in a STEM field could have been important indicators of STEM backgrounds. In general, greater demographic data on axes such as gender and age could have been usefully discriminating, unfortunately, due to the pervasive hegemony of STEM fields.

Our model is not statistically significant enough to declare that there is a causal relationship between STEM background and success in Democratic primaries. There may be an underlying causal relationship between STEM background and success with a Democratic electorate, as candidate backgrounds are scrutinized by voters, but we were unable to identify one with confidence. The effect observed in our data could be due to any number of over- or underestimations of the impact of STEM on the votes a candidate earns, including, but not limited to, confounders not captured by our data such as political history, personal history, and the country's state of affairs in 2018—a time undeniably marked by an increase in public discourse skeptical of science, truth, and political honesty.

Multiple Hypothesis Testing - Research Question 2

Multiple Hypothesis Testing — Methods

We aim to answer two overarching questions:

1. Did the BCRA have an effect on individual contributions? (Tests 1-3)
2. Is there a meaningful difference in when donations are made during election years versus non-election years? (Tests 4-7)

We have decided to explore these 7 hypothesis questions using our data:

1. Is there a significant difference in the donation amount ($) of individual contributions before and after the act was implemented?
2. Is there a significant difference in the average individual contribution donations spending to Democrats before and after the act was implemented?
3. Is there a significant difference in the average individual contribution donations spending to Republicans before and after the act was implemented?
4. Were individual distributions in 2002 uniformly distributed across the months of the year?
5. Were individual distributions in 2003 uniformly distributed across the months of the year?
6. Is the proportion of donations made in October (month 10) 2002 significantly greater than the quantity of donations made in October (month 10) 2003?
7. Is the proportion of donations made in October (month 10) 2004 significantly greater than the quantity of donations made in October (month 10) 2003?

It makes sense to use multiple hypotheses because understanding whether or not the BCRA had an impact on individual contributions is a broad question, one that could be evaluated over numerous metrics requiring specific test statistics. Multiple hypothesis testing allows us to identify more granular effects of sweeping policy initiatives on subsets of the population we are interested in.

We did not compute the power of an alternative hypothesis mainly because we did not have a predictive hypothesis question. We are mainly interested in the before and after of a treatment (the BCRA) on a specific group's behavior, and we are interested in the general trends of donations during specific years.

Hypothesis Test 1

Null Hypothesis: There is no significant difference between the mean donation amount of individual contributions before and after the act.

Alternative Hypothesis: There is a significant difference between the mean donation amount of individual contributions before and after the act.

Test Statistics: We used a t-test to compare the mean donation amount of the two groups. Because the t-test requires data to be normally distributed, we bootstrap sampled mean donation amounts from before and after the act to obtain a normal distribution.

Hypothesis Test 2

Null Hypothesis: There is no significant difference between the mean donation amount of individual contributions to Democrats before and after the act.

Alternative Hypothesis: There is a significant difference between the mean donation amount of individual contributions to Democrats before and after the act.

Test Statistics: We used a t-test to compare the mean donation amount of the two groups. Because the t-test requires data to be normally distributed, we bootstrap sampled mean donation amounts from before and after the act to obtain a normal distribution.

Hypothesis Test 3

Null Hypothesis: There is no significant difference between the mean donation amount of individual contributions to Republicans before and after the act.

Alternative Hypothesis: There is a significant difference between the mean donation amount of individual contributions to Republicans before and after the act.

Test Statistics: We used a t-test to compare the mean donation amount of the two groups. Because the t-test requires data to be normally distributed, we bootstrap sampled mean donation amounts from before and after the act to obtain a normal distribution.

Hypothesis Test 4

Null Hypothesis: The distribution of individual donations is uniformly distributed.

Alternative Hypothesis: The distribution of individual donations is not uniformly distributed.

Test Statistics: We used a chi-square test to determine if the distribution was uniform or not. The observed is the observed proportion of votes for each month, and the expected is 1/12 of the total amount of transactions in the entire year. We used a chi-square test to determine if the difference between observed and expected proportions is statistically significant.

Hypothesis Test 5

Null Hypothesis: The distribution of individual donations is uniformly distributed.

Alternative Hypothesis: The distribution of individual donations is not uniformly distributed.

Test Statistics: We used a chi-square test to determine if the distribution was uniform or not for the same reasoning as hypothesis test 4, as it is the same question for a different year.

Hypothesis Test 6

Null Hypothesis: The proportion of donations made in October of 2002 is not greater than the proportion of donations made in October 2003.

Alternative Hypothesis: The proportion of donations made in October of 2002 is greater than the proportion of donations made in October 2003.

Test Statistics: We used a one-sided z-test for proportions to determine if there were more donations in October 2002 than October 2003. This makes sense because we are testing whether the proportions of two independent populations are statistically significantly different in one direction. If it were multiple proportions instead of just two, then perhaps we would have used a chi-square test again.

Hypothesis Test 7

Null Hypothesis: The proportion of donations made in October of 2004 is not greater than the proportion of donations made in October 2004.

Alternative Hypothesis: The proportion of donations made in October of 2004 is greater than the proportion of donations made in October 2003.

Test Statistics: We used a one-sided z-test for proportions to determine if there were more donations in October 2004 than October 2003. The justification is the same as hypothesis 6 because it is the same question for different years.

Multiple Hypothesis Testing — Results

Hypothesis Test 1

T-test statistic: -77.67321370868123

p-value: 0.0

After performing the t-test, we got a chi-squared statistic of ~-77.67 with a p-value of 0. Thus, we reject the null hypothesis. This means that there is a significant difference between mean donation amount before and after the BCRA.

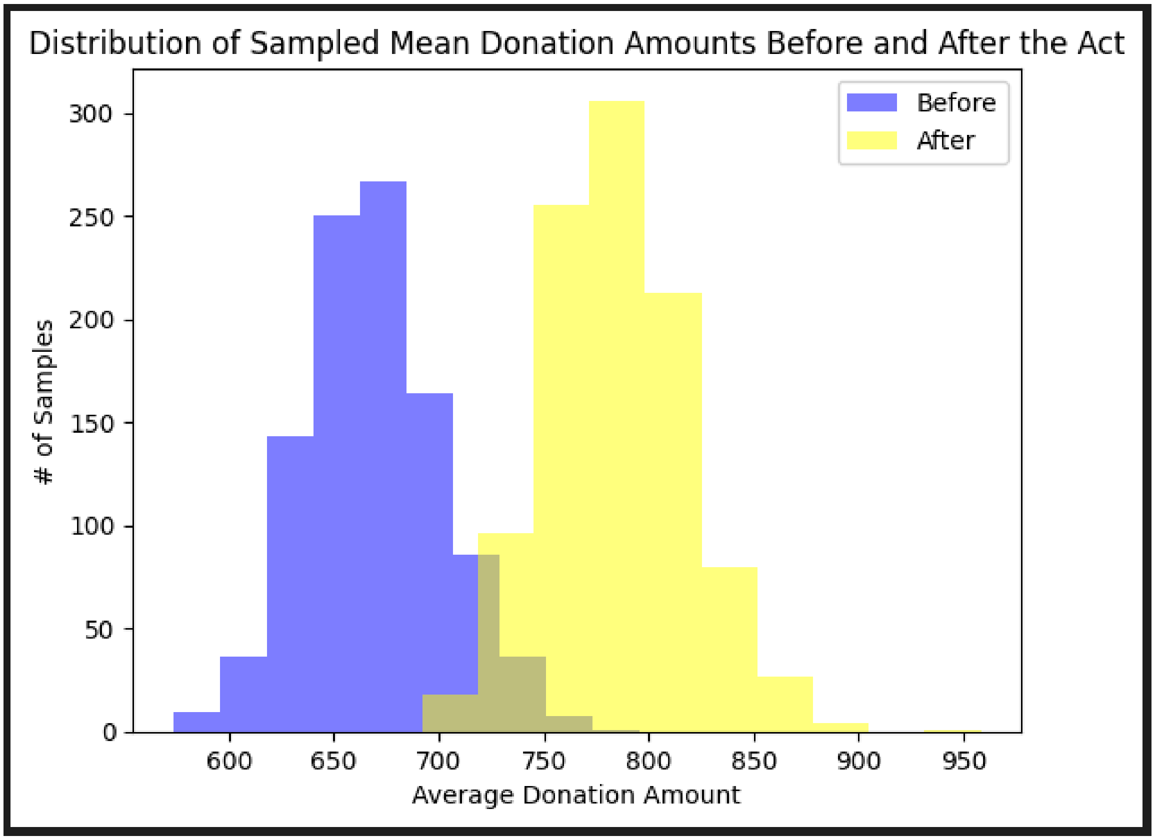


Figure 9. Distribution of Sampled Mean Donations Before and After BCRA

Hypothesis Test 2

T-test statistic: -73.62318723481036

p-value: 0.0

After performing the t-test, we got a chi-squared statistic of ~-73.62 with a p-value of 0. Thus, we reject the null hypothesis. This means that there is a significant difference between the mean donation amount to Democrats before and after the BCRA.

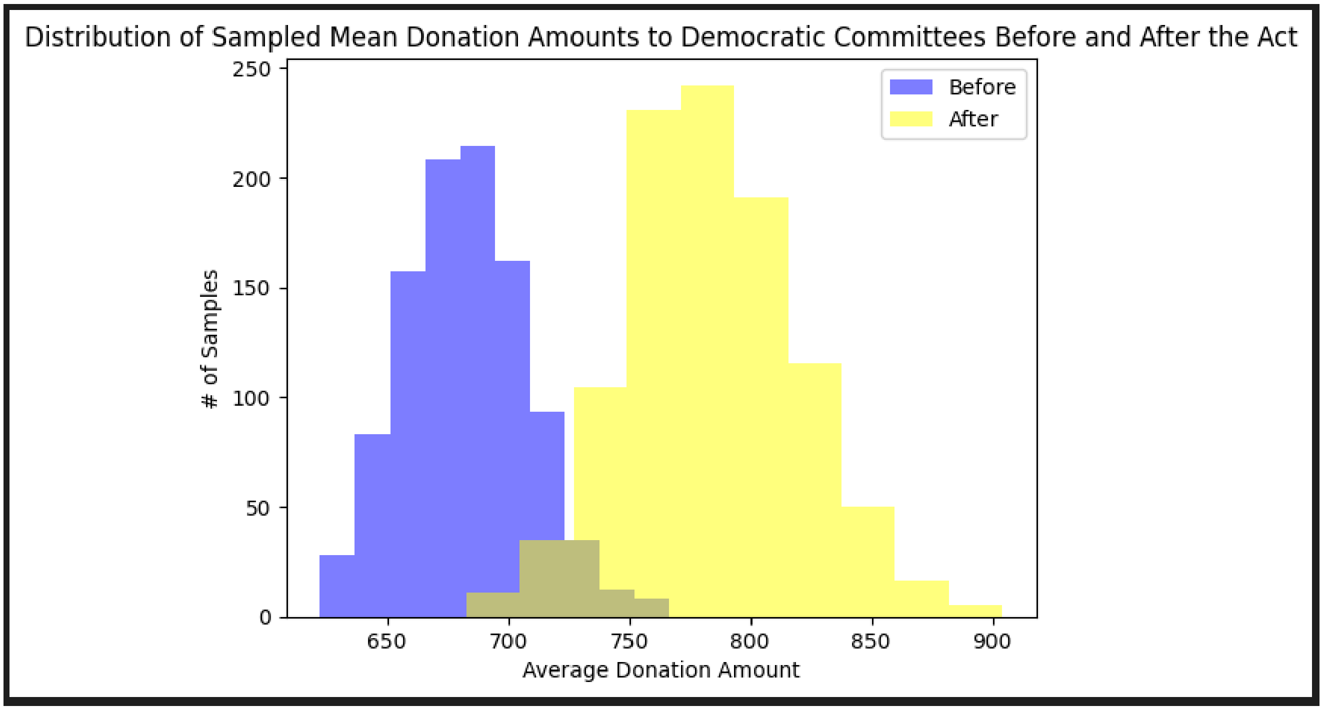


Figure 10. Distribution of Sampled Mean Contributions to Democratic Committees Before and After BCRA

Hypothesis Test 3

T-test statistic: -74.8249941229751

p-value: 0.0

After performing the t-test, we got a chi-squared statistic of ~-74.82 with a p-value of 0. Thus, we reject the null hypothesis. This means that there is a significant difference between the mean donation amount to Republicans before and after the BCRA.

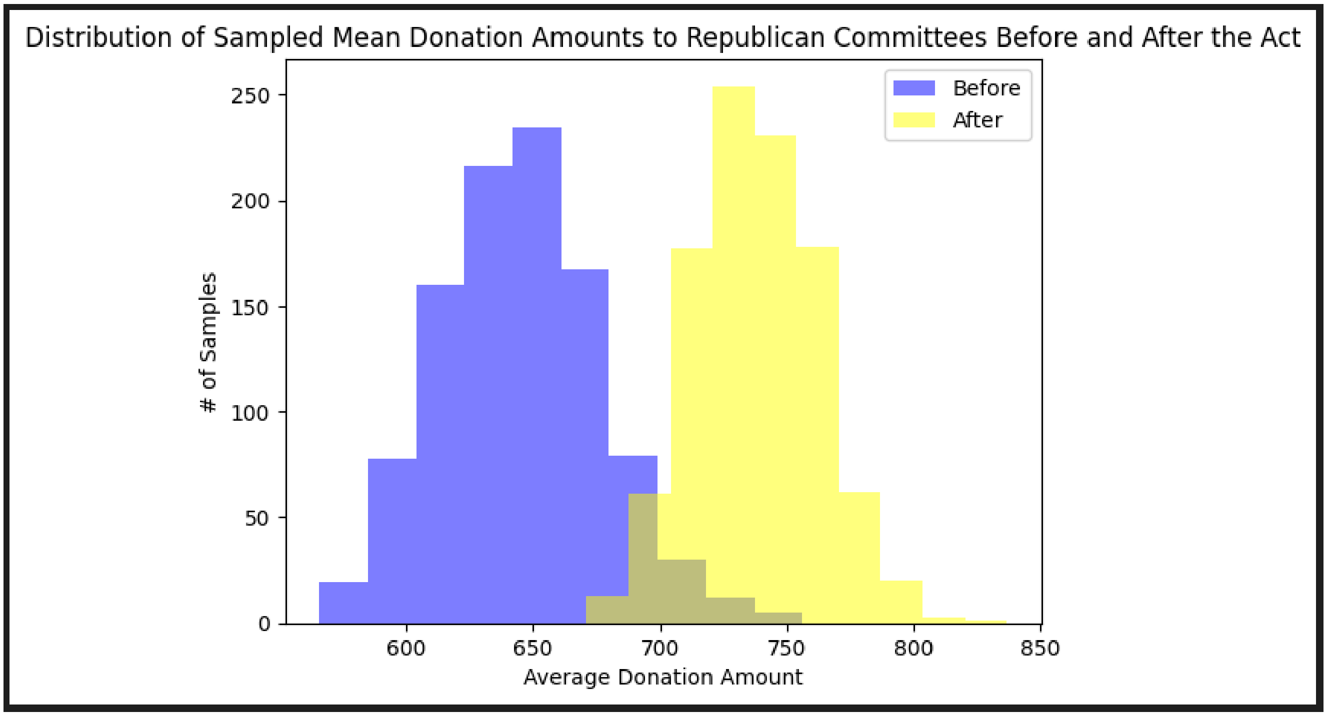


Figure 11. Distribution of Sampled Mean Contributions to Republican Committees Before and After BCRA

Hypothesis Test 4

Chi-square statistic: 106687.26977553572

p-value: 0.0

After performing the chi-squared test, we got a chi-squared statistic of ~106687.26 with a p-value of 0. Thus, we reject the null hypothesis. This means that one or more months in the year must have had more votes than the others.

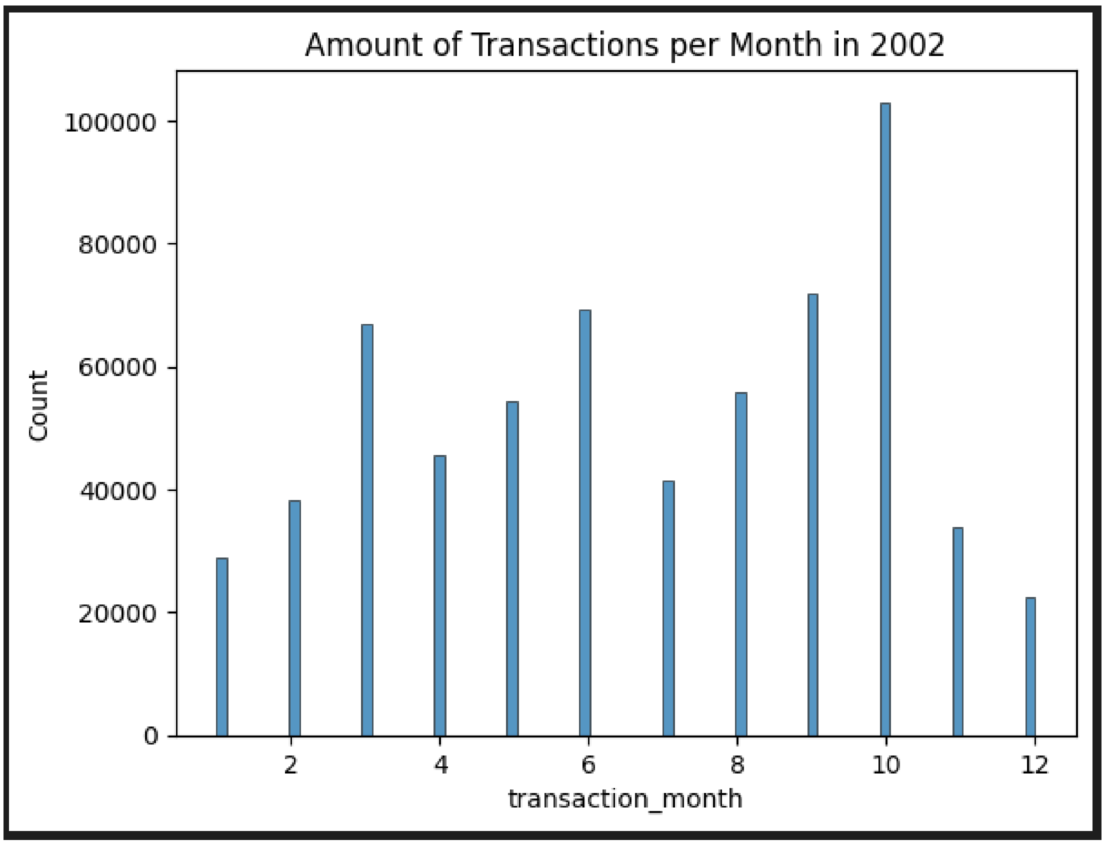


Figure 12. Distribution of Transactions Per Month in 2002

Hypothesis Test 5

Chi-square statistic: 102922.67018488757

p-value: 0.0

After performing the chi-squared test, we got a chi-squared statistic of ~102922.67 with a p-value of 0. Thus, we reject the null hypothesis. This means that one or more months in the year must have had more votes than the others.

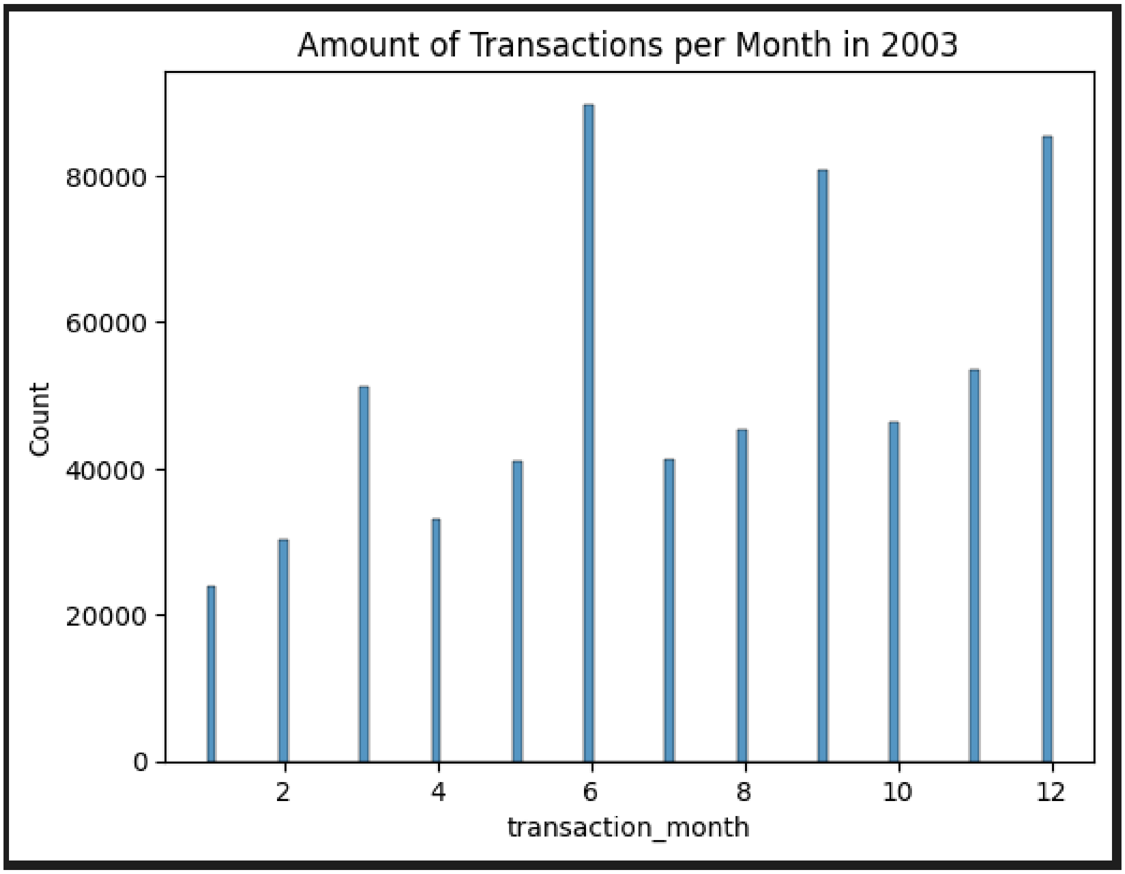


Figure 13. Distribution of Transactions Per Month in 2002

Hypothesis Test 6

Z-statistic: 152.89312765766016

p-value: 0.0

After performing the one-sided Z- test for proportions, we got a z-statistic of ~152.89 with a p-value of 0. Thus, we reject the null hypothesis. Since we used a one-sided test, we can say that the number of donations in October of 2002 is actually greater than the number of donations in October of 2003.

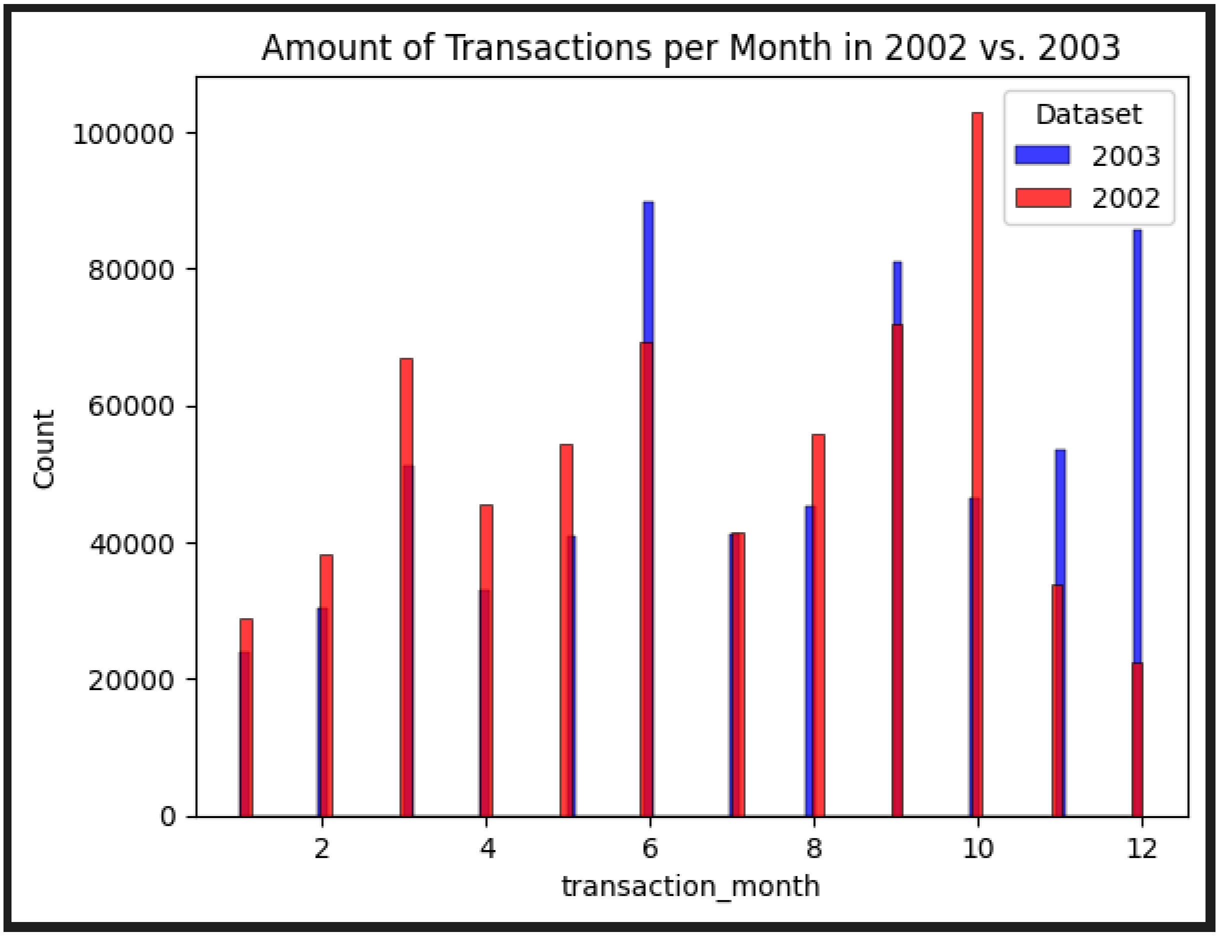


Figure 14. Comparison of Transactions Per Month in 2002 vs 2003

Hypothesis Test 7

Z-statistic: 158.39376512767294

p-value: 0.0

After performing the one-sided Z- test for proportions, we got a z-statistic of ~158.39 with a p-value of 0. Thus, we reject the null hypothesis. Since we used a one-sided test, we can say that the number of donations in October of 2004 is actually greater than the number of donations in October of 2003.

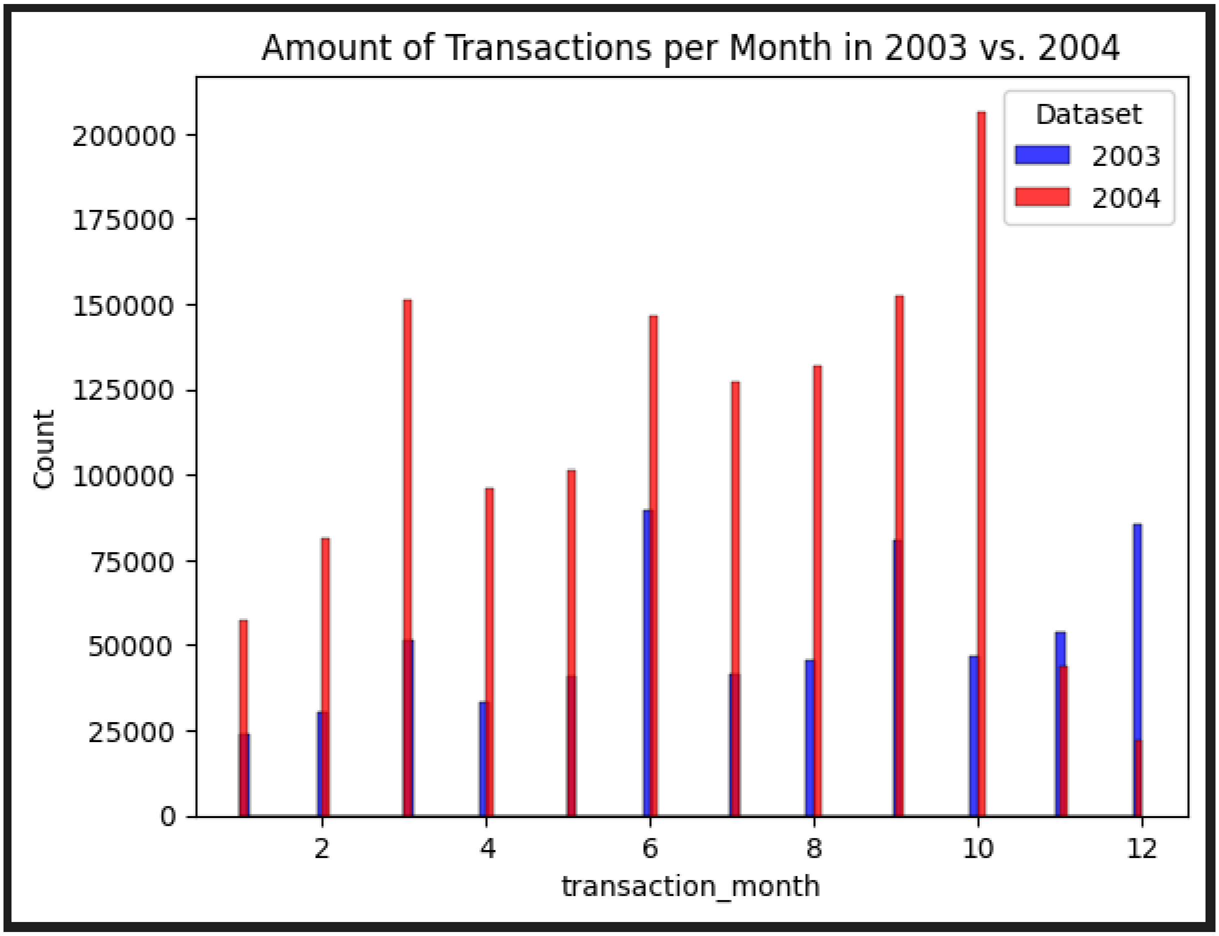


Figure 15. Comparison of Transactions Per Month in 2003 vs 2004

Tests 6 and 7 show us that donations in October of 2002 and 2004 were statistically significantly greater than the number of donations made in October of 2003. Contextually, 2002 and 2004 were election years, and October is the month before the primaries and presidential election for 2002 and 2003 respectively. These tests could tell us that during election years, people donate more money right before election results come out.

We conducted two sets of error corrections, one for each of the error evaluations for our multiple hypothesis tests. We split the tests into two sets because our tests were answering two different questions. The methods we used to perform this error correction were Bonferroni Correction and Benjamini-Hochberg Correction. Bonferroni Correction is a method that controls the FWER (family-wise error rate) to manipulate the probability of false discoveries among each test. It works by adjusting the significance threshold (α) to account for the number of tests being conducted. Specifically, the adjusted threshold is calculated as α/m, where α is the original significance level (e.g., 0.05) and m is the number of comparisons or tests. By using this more stringent criterion, the Bonferroni Correction ensures that the overall family-wise error rate (the probability of making one or more errors across all tests) remains at or below the desired significance level. Unlike the Bonferroni correction, which controls the probability of any false positive, the Benjamini-Hochberg Correction method focuses on limiting the proportion of false positives among all significant results. It ranks the p-values from smallest to largest and compares each p-value to a threshold $(k/m) \cdot \alpha$, where k is the rank, m is the total number of tests, and α is the desired FDR level. The largest p-value meeting the threshold and all smaller p-values are considered significant. This approach is less conservative than Bonferroni, offering greater statistical power while maintaining control over the expected proportion of false discoveries.

Discussion

Error Correction:

Question 1 — Bonferroni Corrections

Applying Bonferroni Correction on our first question required controlling for a Family Wise Error Rate of 0.05 / 5 = 0.01. Using this threshold on our five hypothesis tests, we made five discoveries.

Question 1 — Benjamini-Hochberg Corrections

Applying Benjamini-Hochberg Correction on our first question, we made five discoveries.

Question 2 — Bonferroni Corrections

Applying Bonferroni Correction on our second question required controlling for a Family Wise Error Rate of 0.05 / 2 = 0.025. Using this threshold on our two hypothesis tests, we made two discoveries.

Question 2 — Benjamini-Hochberg Corrections

Applying Benjamini-Hochberg Correction on our second question, we made two discoveries.

Unsurprisingly, these results make sense because all the p-values we got were 0, so for each of the two questions, both Bonferroni and Benjamini-Hochberg tests should include all of the p-values in that set.

The results of hypothesis questions 1, 2, and 3 inform us that there was a significant increase in the average transaction amount after 2002. Examining the results of our hypothesis testing informs us on the impact that the BCRA had on campaign financing and can push lawmakers to pursue future policy.

For hypothesis questions 6 and 7, we can see that in election years, there are a proportionally greater amount of donations that come in during October. A strategic decision could be that a candidate's campaign team could reserve funds and allocate additional resources for a final push in October, ensuring high visibility and turnout efforts in the critical pre-election period.

If we had more data, then we can perform tests to see whether the October spike varies across states, demographics, or donor categories. If we had more demographic data from who these donations came from, then they could further strategize and appeal to those donors during this time period. We could perform a chi-squared test for example on if geographical locations differ in this October spike.

If we had more data across multiple years, then we could also test this hypothesis in the same way across multiple election years vs. non election years to strengthen or reconsider our initial inference. We would again use a z-test for proportions.

Lastly, for a more complicated question, if we had more data about candidate favorability at the time of the donations, we could analyze whether increases in donations during October are correlated with shifts in candidate popularity. A correlation test or linear regression could help determine if rising poll numbers or media coverage leads to spikes in donations.

Conclusions

Our first research question was investigating whether coming from a STEM background has a causal impact on the percentage of votes won by Democratic candidates. We were unable to conclude if there is a causal relationship between coming from STEM and the percentage of votes won by Democratic candidates as our model is not statistically significant enough.

Our second research question was testing multiple hypotheses to gain insight on two questions:

1. Did the Bipartisan Campaign Reform Act of 2002 have an effect on individual contributions?
2. Is there a meaningful difference in when donations are made during election years versus non-election years?

Based on our data analysis, we can see that after BCRA was enacted, individual contributions increased across the board for Democrats and Republicans alike. We could see from our analysis that during election years, the proportion of donations made during October are higher than non-election years.

Causal Inference Question

Unfortunately, we had many limitations. As mentioned in our data overview, we had to exclude endorsements because null values are not equivalent to active support or opposition. We also had to drop 20% of the rows from our data that did not include self-reported race. This could have introduced significant bias, as self-reporting race is not random.

In terms of domain knowledge, we are not political experts. There are many factors that go into a candidate's success that we are not privy to. One key factor is the true involvement and impact of endorsements. One question we could ask a domain expert is "How should we interpret the absence of endorsement data? Do null values imply neutrality, an unknown position, or something else?" If they tell us that null values in endorsement data truly represent neutrality, then we could have included them as a separate feature rather than eliminating the attribute entirely.

Due to missing values, we cannot assume our model to be robust. Data on candidate race and the choice to not self-report race would likely yield interesting analyses. The excluded candidates might systematically differ in geography, ideology, or voter demographics. This leads to an incomplete understanding of the role race plays as a confounder. We could have treated this feature differently in our model by trying to impute race based on other candidate attributes. However, we fear that this would not fix our bias issues, as imputing race based on the other limited features would still introduce bias.

Our findings are not very generalizable. In terms of candidate context, we specifically focused on Democratic candidates who are running elections that do not include current incumbents with a STEM background. This may not apply to Republican or Third Party candidates. The time frame of our data was also only the 2018 primaries, so our conclusions may not extend to other election cycles.

A future study that could build upon our work is one that examines the relationship between a STEM background and election success across multiple political parties. If a study can find data that includes information from multiple elections and more attributes, then it can create a model that can confirm if having a STEM background influences primary results more generally.

One call to action could be making candidate information more transparent to the general public. This could include logging information that aligns with the features that were present in the dataset so that we can see the true relationships present. The Federal Election Commission or state-level election boards could implement this by requiring candidates to provide information, such as endorsements, financial backgrounds, and personal identifications. This would be helpful from a data science and social perspective because we could then see how factors like race, sexuality, and finances influence election results. This could provide data on how racism might still be a large factor in gaining power in the US or elucidate voting patterns in specific regions. However, it could be unethical to make candidates report these features. For the same reasons as to how these factors might impact voting, perhaps candidates choose not to report as they fear how they will be treated.

Multiple Hypothesis Testing

We encountered different limitations with our multiple hypothesis testing. This dataset was great to work with because it was the entire population of donations, but the attributes it contained were limited. The biggest limitation in the data we had was that a small proportion of the donation amounts were negative. There was no information on what that meant exactly, so we decided to drop those rows.

Domain knowledge that we are missing is the intricacies of the BCRA and understanding of how it was implemented. For example, the BCRA was highly polarized along partisan lines. Many Republicans opposed the BCRA, some even challenging this law in the Supreme Court case *McConnell v. Federal Election Commission*. One important question we can ask a domain expert would be “When laws like these are strongly contested, does that impact how fast the law is put into action?”. If the BCRA was hard to enforce, then perhaps the date cutoffs we used in our analysis should change. For instance, perhaps the effects of a law are put on pause when it is challenged in the Supreme Court. Thus, these details could change the cutoffs that we used.

Our conclusions are rather robust because our data is well populated, and this is the entire population not a sample. One modelling choice to change could be including the negative values as it could represent a bounced check of a previously counted donation. This could bias our results because this might not be what the values represent.

Our findings are not necessarily generalizable because we were analyzing the impact of one very specific law that was enacted but also contested and altered in the next coming years. It is likely that we cannot generalize those findings to other campaign reforms or other contexts. However, our questions about donations coming in during election years is likely very generalizable and broad for American politics.

One call to action is rooted in the need for additional research that explores the impact of the BCRA on more demographics such as age groups, geographic regions, and income levels. We could further confirm or refute our findings for these different groups of donors.

Another question that could further the inquiries spurred by our research is to gauge whether banning all non-regulated campaign financing changed the way that campaigns spent their funds. With all finance being regulated and visible in the federal data, campaigns might change the way that they spend the donated funds in politicized ways.

More importantly, a looming ethical concern is brought up with the ban of soft money that requires further inquiry. While eliminating unregulated campaign donations is a sensible task at face value, it also imposes significant visibility on political activity. Prior to the act, a citizen could make soft money contributions to support a political initiative and not have to worry about public scrutiny. Hard money contributions are federally regulated, and their records published publicly. The elimination of a method for relatively anonymous campaign donations in the form of soft money has an outsized impact on citizens who would like to participate in political activities anonymously. For instance, a donor might not want to expose their contributions to a party if they live in an area where politics are highly contentious, or if their employer would discriminate against them for having a certain political stance. Similarly, vulnerable populations hoping to support issue items on topics of sexuality, gender, race, and ideology are at disproportionate risk of real world harm should their contributions become public record, as is the case with hard money donations. This is an externality of the BCRA that we could not analyze with our data, but requires attention in future work and conversation about the impacts of the act.