

# Data 102 Final Project

Dora Du, Ricardo Reynoso, Alex Zheng, Hanson Wang

## Data Overview

We supplemented the **dem\_candidates**, **rep\_candidates**, and **lobbyist\_bundle** data with our own, in order to better understand the relationship between several metrics in our multiple hypothesis tests. Most of our supplemental data was government-sourced, eliminating potential biases such as selection bias and convenience sampling since government data typically includes comprehensive records of candidate information, populations, poverty rates, donations, etc. As a result, most of our data is census data rather than sampled data. Any data cleaning involved aligning the names of states to allow for easy merging (e.g. making a new column renaming states from their abbreviations to their full names). All supplemental data is explained below:

**dataset:** FY2023-STC-Category-Table-Transposed.csv

This dataset is also from the U.S. census bureau, and we downloaded this data straight from [data.census.gov](https://data.census.gov). We needed this dataset because we wanted information about the total taxes collected across the states to help answer one of the hypothesis tests in our first research question, which asked about the relationship between the tax rates and campaign finance.

**dataset:** dollars\_raised.csv

This dataset is from <https://www.followthemoney.org/>, which now belongs to a nonprofit organization called OpenSecrets. We manually inserted the data into the notebook because the original website did not have a download option. We needed this dataset because in addition to the dataset above that helped us answer one of the hypothesis tests, we also needed information on the changes in campaign finance after knowing how much taxes were collected among each state.

**dataset:** MortgageData.csv

This dataset is [from the Federal Housing Finance Agency's National Mortgage Database \(NMDB\)](#). There was no spreadsheet download option, so we had to manually go into the dashboard and copy the mortgage loan volume in every state. This was the only place we could find the total mortgage volume since all the other sources only provide the average mortgage amount.

**dataset:** Counties Per Capita Income.csv

This dataset was [sourced from Kaggle](#) and provides information about per capita income, household income, family income, and population across U.S. counties. We selected this dataset to analyze the relationship between poverty rates and campaign finance, as it offers a detailed breakdown of economic conditions at the county level.

**dataset:** Unemployment in America Per US State.csv

This dataset was [sourced from Kaggle](#) and includes monthly unemployment statistics for U.S. states, along with labor force and employment data. We used this dataset to investigate the impact of unemployment rates on campaign contributions and voter support, aligning with our goal of understanding how economic factors influence political behavior.

**dataset:** 2023 Small Business Profiles.csv

This dataset was manually entered and sourced [from the U.S. Small Business Administration](#), and includes statistics of U.S. businesses directly from the U.S. census bureau. We used this dataset to analyze the relationship between the density of local businesses in a state and the amount of dollars raised for campaigns in that state. This would allow us to determine if there is a relationship between small business density and campaign spending.

**dataset:** GDP\_Growth.csv

This dataset was taken [from the Bureau of Economic Analysis \(BEA\)](#) of the U.S. Department of Commerce, and contains percentages for each state measuring the rate of Gross Domestic Product growth within the first two quarters of 2024. We used this dataset to begin determining if economic growth rates of a state influence the number of small individual donations. Since our dataset did not contain the amount of small individual donations per state, we had to look further and concatenate our results with this dataset. Also, this data is percentage-based, which doesn't grow proportionally the way that the number of small donations would. (More on this in the next dataset.)

**dataset:** Totals by State, 2023-2024.csv

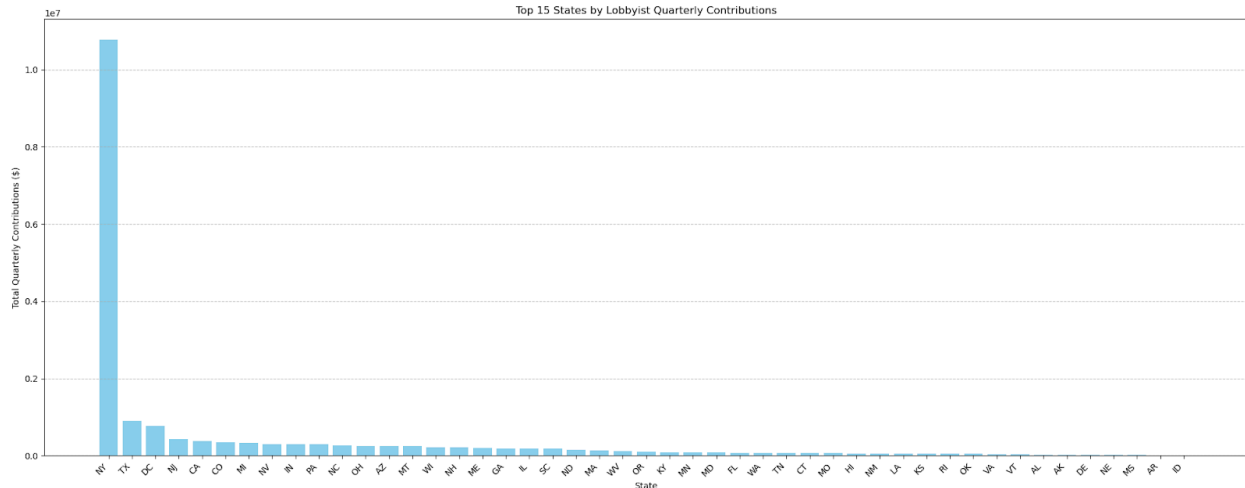
This dataset was taken [from the Federal Election Commission](#) and contains data of individual campaign contributions. It contains data for the donor's name, the state, the recipient, and the amount. This allows us to filter for small individual donations and answer our question about the relationship between economic growth rates and small individual donations. However, since this data is expected to grow proportionally by state population, it needed to be pre-processed to correlate more closely with GDP growth. We divided the small donation total by the state's population using the dataset below, then took the log.

**dataset:** Populations.csv

This dataset was taken [from the U.S. census bureau](#) and contains data of the total population of each state in 2020. The original data frame had the states as the indices so the data was transposed to better fit with the above datasets, which were concatenated with this one for analysis. This dataset allowed us to create a ratio between small individual donations in a state, and its population.

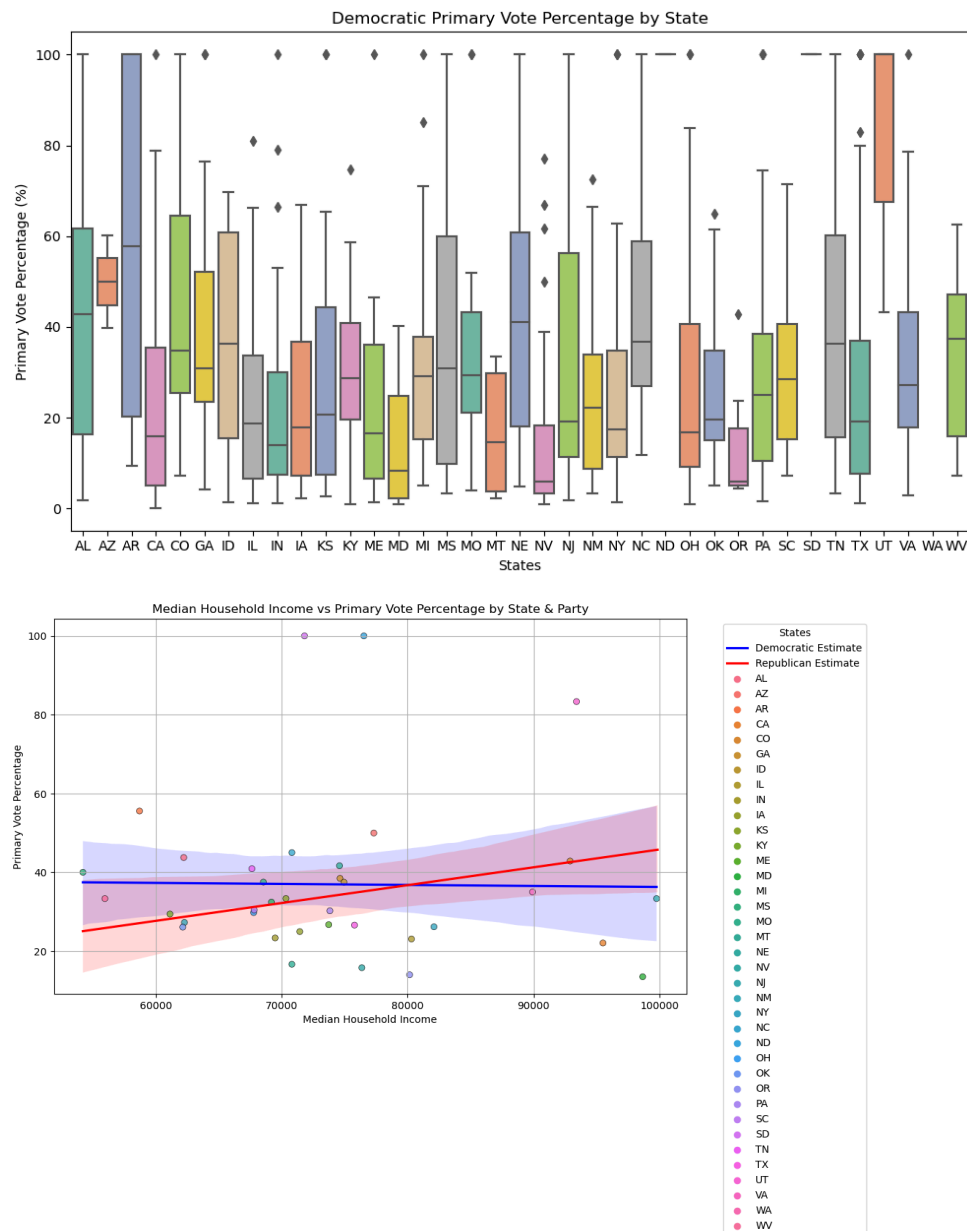
# EDA

**Quantitative Variable #1:** Quarterly Contributions (Lobbyist dataset): To investigate financial support levels.



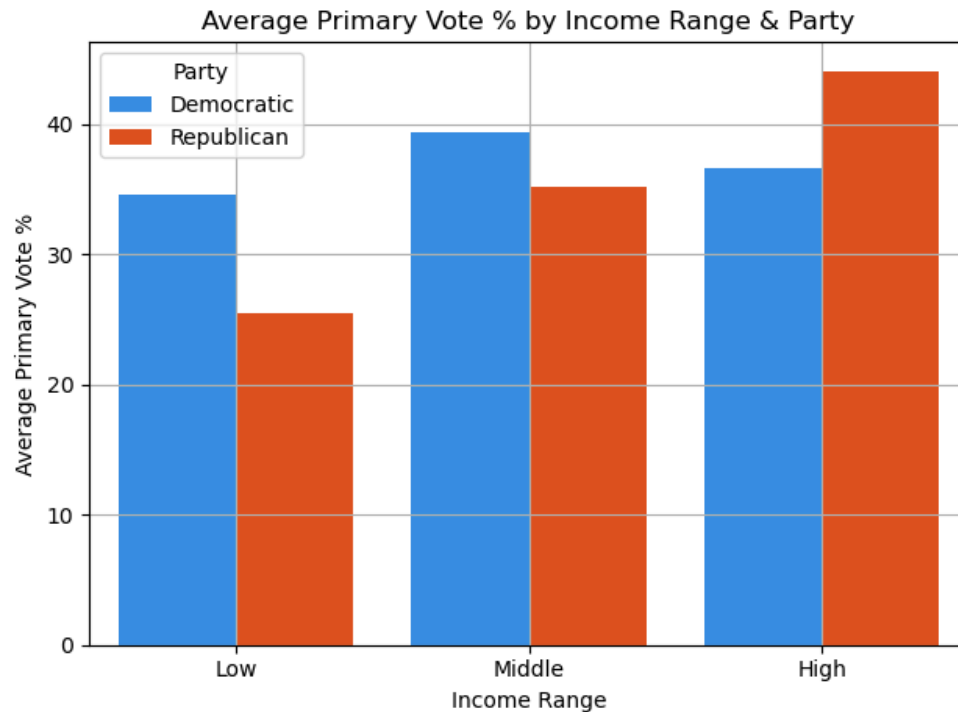
The bar chart shows the total quarterly lobbyist contributions for the top 15 states. New York dominates with significantly higher contributions compared to all other states, followed by Texas, Washington, D.C., and New Jersey, which have moderately high contributions. Beyond the top few states, contributions drop off rapidly. This suggests that there are geographic differences in funding, which may have a different impact on campaign success in each state. The graph highlights the geographic concentration of financial resources in a few key states, suggesting that candidates from these states may benefit from greater financial support. New York, in particular, stands out as a major financial hub, home to numerous corporations and advocacy groups who actively contribute to political campaigns. Moreover, New York's diverse and influential political landscape frequently attracts substantial funding from both local and national organizations aiming to influence policy decisions or support candidates who align with their interests.

**Quantitative Variable #2:** Median Household Income: To investigate how household income relates to voter turnout by state & party



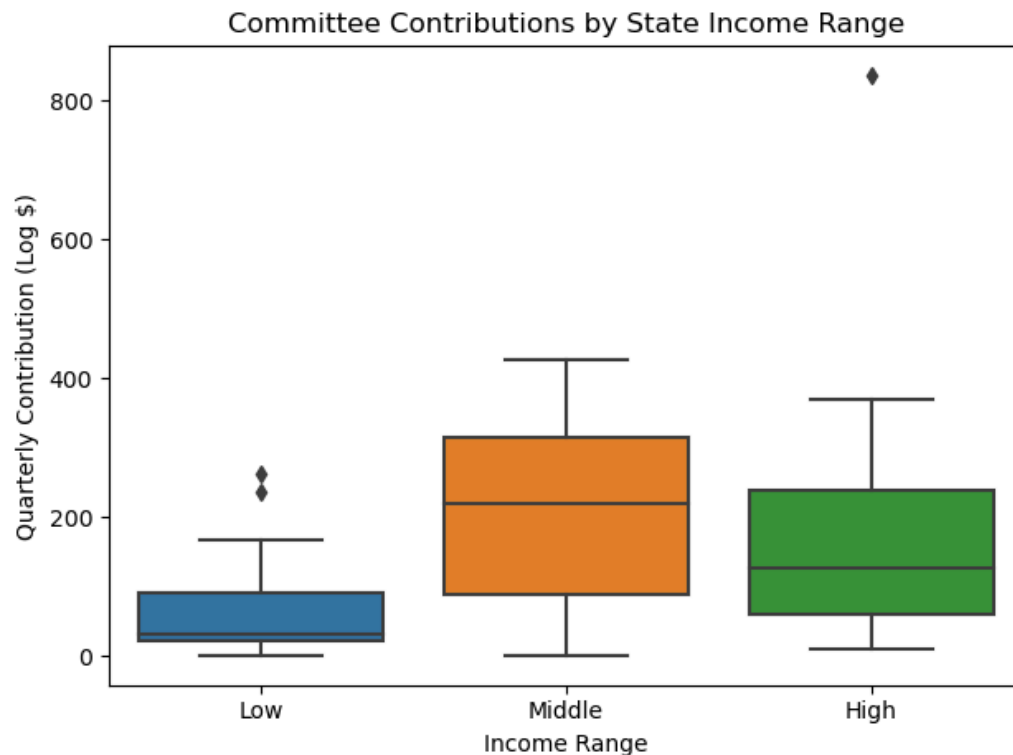
By combining the data from U.S. Census with the democratic candidates, Arkansas and Utah seem to have the highest primary vote percentage, Nevada and Oregon have the lowest primary vote percentage, where Utah has the highest median household income, and Arkansas has the lowest median household income. 1 interesting note is that California has a median household income higher than all these 4 states, but the primary vote percentage is pretty low, which can be shown by the second scatter plot above.

**Categorical Variable #1:** Income Brackets: We transformed the above into a categorical variable to see how income range correlates with primary voter turnout



We split each state equally into three income ranges, then found their average primary vote % by party. This bar plot suggests that states in the lower income range vote less overall, but we see something interesting in the middle and high ranges. For Republicans, the primary vote percentage rises the higher the state's income. However, for Democrats, the primary vote percentage actually decreases slightly when going from middle to high income states. This may be due to the way our income ranges were split, which might not be truly indicative of the socially accepted income threshold of each state. The interpretation is that higher-income Republicans are more politically active than higher-income Democrats, meaning that a state's economic conditions may not be the only influence in voter support, and political party should be considered a more significant factor. It is also worth noting that Republicans are more involved in PACs than Democrats, who lean more toward individual contributions.

**Categorical Variable #2:** Committee Contributions & State Income Range: To see if there is a relation between the income range of a state and the amount of committee contributions it delivers



This boxplot connects a state's income range to its committee contributions. We see the same trend we saw above with Democratic primary vote % by state income level. Contributions rise significantly from lower income range states to middle income range states, but then they fall for high income range states. Committees from middle income range states contribute the most financially compared to any other income range. This could be interpreted in a couple of ways. Grassroots support typically comes from middle-income voters who have enough finances to chip in to campaigns, which is typically the most popular form of financial support for a campaign regardless of party. Regardless of the medians, we see that high income states have extremely high outliers, and low income states also have relatively high outliers. This is likely due to contributions from billionaires who live in high-income states, and from wealthier individuals who live in low-income states and heavily overshadow the typically low donation.

## Research Questions

Our first research question is: “How do economic conditions within a state influence campaign spending and voter support for candidates?” By studying how various economic factors within a state (such as unemployment, income, and tax rates) affect campaign spending and voter support, political parties can optimize their funding and support based on the economic conditions of each region, ensuring that resources are allocated where they are most likely to have an impact. Multiple Hypothesis Testing is well-suited here because it allows researchers to systematically evaluate numerous economic indicators at once while minimizing false positives. A limitation of this method may be that multiple hypothesis testing reduces statistical power because it uses more conservative significance thresholds across multiple hypotheses, increasing the likelihood of missing true effects.

Our second research question is: “What types of endorsements (e.g. political organizations, public figures, social media, etc.) are most influential in determining a candidate’s success in primary elections?” By identifying which endorsements will actually move primary outcomes, campaigns can allocate resources more strategically to secure high-impact endorsements. Causal inference methods are well suited to this problem because they help to separate the effects of support from confounding and selection bias inherent in observational data. The limitations of what we might get are that they may rely on critical assumptions. Success depends on having valid instrumental variables that truly only affect support, and on capturing all major confounders in the propensity score model. If important variables are omitted, or if the instrumental variables directly affect the election outcome, the estimates may be biased.

## Prior Work

Akdim, K., & Belanche, D. (2022, November 26). Endorsement effectiveness of different social media influencers: The moderating effect of brand competence and warmth. *Journal of Business Research*. <https://www.sciencedirect.com/science/article/abs/pii/S0148296322009419>

This study specifically focused on social media influencers and celebrities. We wanted to analyze with a much broader scope to see how effective endorsements are by politicians as well as activist groups. We also wanted to analyze specifically with regard to the primaries. The primaries focused more on ideological purists.

Barber, M. (2017, April 1). Ideologically sophisticated donors: Which candidates do Indi. *American Journal of Political Science*.  
<https://ideas.repec.org/a/wly/amposc/v61y2017i2p271-288.html>

This study focuses on what type of demographics are most likely to donate to campaigns. It focuses only on individuals and not on PACs and organizations. We wanted to approach the situation from the other direction. Do economic demographics cause candidates to spend more on their campaigns? Only a small chunk of political donations come from individuals.

# Research Question #1: Multiple Hypothesis Testing

The question we seek to answer is: What economic conditions within a state influence campaign finance and voter support for candidates?

Testing multiple hypotheses is necessary because campaign finance is a complex phenomenon influenced by economic, demographic, and political factors. By testing multiple hypotheses and analyzing multiple variables like poverty, tax rates, economic growth, we can identify which factors play the most significant roles and how they interact with each other.

## 1st Hypothesis Test

- H0: state poverty rates do not affect campaign contributions.
- H1: higher state poverty rates reduce campaign finance.

The hypothesis explores the relationship between state unemployment rates and campaign spending. We aim to identify whether candidates in states with higher unemployment face financial disadvantages, which could impact their campaign effectiveness and overall competitiveness in elections. We aim to identify whether candidates in states with higher poverty are less likely to financially support political campaigns. We'll be using OLS regression analysis due to its simple interpretability and linearity since we're only quantifying for correlation based on the independent variable of poverty rate and dependent variable of quarterly contribution.

The regression shows a positive relationship between unemployment rates and campaign spending. For every 1% increase in unemployment rate, campaign spending increases by \$70,410. We reject the null hypothesis because the P-value is  $1.69e-05$ .

## 2nd Hypothesis Test

- H0: state income levels within a state have no effect on voter support for candidates.
- H1: higher state income levels can increase voter support for candidates.

The hypothesis examines whether voter support is influenced by the wealth of their region. Wealthier regions may have different levels of political engagement or priorities compared to less wealthy neighborhoods. This may affect how they vote and potential campaign strategies. We'll be using OLS again since we're only predicting based on two variables for a clear and readable correlation.

The regression shows per capita income has a significant positive relationship with voter support percentage (Primary %). For every additional dollar in per capita income, voter support increases by 0.0006 percentage points. We reject null hypothesis because P\_value is  $3.17e-148$ .

## 3rd Hypothesis Test

- H0: density of local businesses in a state has no effect on campaign finance.
- H1: density of local businesses in a state has a significant effect on campaign finance.



The hypothesis focuses mainly on the density of small businesses and how it influences campaign financing in a given state. By focusing on small businesses, we are able to determine if grassroots-level economic activity impacts campaign financing at large. Again, we'll be sticking to OLS due to only modeling on two variables and the interpretability of the correlation. Plus, the data is largely proportional so no manipulation is needed.

Our results indicate that the coefficient for Small Businesses is 140.1331, meaning that for every additional small business in a state, the total campaign spending increases by approximately \$140.13 on average. The p-value is  $2.31e-17$ , which is extremely small. With an alpha threshold of 0.05, we would be able to successfully reject the null hypothesis and are able to demonstrate a significant positive relationship between small businesses and campaign finance.

#### **4th Hypothesis Test**

- H0: economic growth rates in a state do not influence the number of small individual donations.
- H1: higher economic growth rates correlate with more small individual donations.

This hypothesis test seeks to question if there is a correlation between economic growth rates of a given state and the amount of small individual donations. Since GDP growth largely benefits every constituent, we believe it is worth focusing on smaller individual donations to get a sense of the general population's donating habits rather than PACs or committees'. Though this question requires more data manipulation since GDP growth is percentage-based and the margins are very low and we're modeling on total contributions (which isn't proportional), we'll still be sticking to OLS since we want to find a correlation between two variables.

Our results indicate that the coefficient for GDP Growth Percentage is -0.0660, meaning that for every additional percentage increase in GDP Growth of a state, the total log donation/population ratio of a state decreases by approximately 0.0660. Our graph supports this downwards trend. The p-value is 0.517, which unfortunately is too high for if we use an alpha threshold of 0.05. Thus we fail to reject the null hypothesis and so far cannot demonstrate a significant positive or negative relationship between economic growth rates and small individual donations.

#### **5th Hypothesis Test**

- H0: state tax rates (total taxes collected) have no effect on campaign finance.
- H1: state tax rates (total taxes collected) have effect on campaign finance.

Testing this hypothesis is to understand the significance of tax rates in influencing campaign finance, as a higher tax rate might lead to lower personal income, which can then influence the amount of money they can spend on the campaign. We use OLS to see how total taxes correlates with dollars raised, which should be clear, easy, and interpretable.

Based on the summary, the coefficient for total taxes is 3.0917, meaning that for every additional dollar in taxes, campaign spending increases by approximately \$3.09 on average. The p-value for total taxes is  $5.26e-15$ , which is highly significant. If we use a typical alpha 0.05, we can reject the null hypothesis and state that there is a significant positive relationship between total taxes collected and campaign spending, i.e. states with higher tax collections tend to have higher campaign spending.

### 6th Hypothesis Test

- H0: states with higher mortgages do not affect campaign finance.
- H1: states with higher mortgages can affect campaign finance.

Our goal with this test is to see how mortgages/loan volume affect campaign financing in a state, whether it be negatively or positively. This is important since home affordability can significantly influence the spending habits of homeowners to different degrees, like a "safety net." This could increase willingness to donate to campaigns with "leftover" money. We use OLS here since we are curious about the correlation between the two variables of loan volume and campaign finances, and we want a clear and interpretable coefficient.

The coefficient for total mortgage loan volume is 0.0004. That means for every additional dollar loaned, campaign raised increases by only 0.0004 dollars. For every 10K dollars loaned, the amount raised by a campaign increases by 4 dollars. The p-value is  $2.26e-21$  which is below the alpha value of 0.05. We reject the null hypothesis since the data is statistically significant. There is thus a significant relationship between mortgages and campaign finance.

### Compare and contrast FWER control and FDR control

The 2 methods we are using are Bonferroni and Benjamini-Hochberg. The Bonferroni correction controls the FWER, which is the probability of making at least one false positive among all the hypothesis tests. This method is very conservative, especially when the number of tests is large, and is designed to avoid any false positives at the cost of potentially missing true positives. The Benjamini-Hochberg correction controls the FDR, which is the proportion of false positives among all the rejected null hypotheses. It's less conservative than Bonferroni, allowing for more discoveries while controlling the rate of false positives among them.

The number of discoveries made under each correction method is 1. For this multiple hypothesis testing, both the Bonferroni correction and the Benjamini-Hochberg correction resulted in one discovery each, this indicates that both methods identified the same hypothesis as significant in this specific case, despite their different error control mechanisms.

The Bonferroni method controls the FWER, which is the probability of making at least one false positive across all hypothesis tests. This method is very conservative and minimizes the chance of false positives but may reduce statistical power.

The Benjamini-Hochberg method controls the FDR, which is the expected proportion of false positives among the rejected hypotheses. This method is less conservative than Bonferroni,

allowing for a higher number of discoveries while maintaining control over the proportion of false positives.

Our research question seeks to understand how various economic factors influence campaign finance and voter support across the states in the U.S., this means that identifying multiple significant factors is important, and since Benjamini-Hochberg method allows for discovering more potential relationships while maintaining a controlled proportion of false positives, it makes more sense to prefer Benjamini-Hochberg over Bonferroni in this case, even though they both returned the same number of discoveries for the hypothesis tests that we chose, as the Bonferroni correction may be too strict for this type of analysis, reducing the chances of detecting true relationships, especially if the evidence for significance isn't as strong as the cases we analyzed.

The results indicate that poverty rates, income levels, density of local businesses, tax rates, and mortgage loan volumes all have a significant effect on campaign spending, donations, or voter support. We also found out that GDP growth is not a significant factor in small individual donations. This is supported by our FDR (which controls for false positives) and FWER (which controls for at least 1 false positive).

## Discussion

After applying our correction procedures, we discovered that poverty rates, income levels, firm density, tax revenues, and mortgage markets remain significant.

From the individual tests, we gained several useful insights. We found a negative impact of poverty rates on campaign contributions. Candidates in less affluent districts may face challenges in securing sufficient funding and will need targeted fundraising to keep their campaigns on track. On the other hand, the positive relationship between tax revenues, mortgage loan volumes, and campaign spending shows how broader economic activity influences political financing. Overall, the state of the economy plays a key role in influencing the campaign.

In this analysis, we came up with a few key limitations. First, our dataset lacks sufficient information about certain variables, which may create blind spots in understanding specific relationships. Second, economic indicators such as income levels and taxes often overlap. This can lead to problems of multicollinearity, which masks the unique impact of each variable. Finally, assuming linear relationships may oversimplify the true interactions of these factors in the real world and reduce the precision of our results. We took a few steps to avoid p-hacking. We applied multiple-comparison corrections like Bonferroni and Benjamini-Hochberg to keep false positives in check. In addition, we disclosed all of our results(both significant and non-significant), so everything was on the table and transparent.

If more data were available, I would expand the multiple-hypothesis by adding additional economic and demographic variables, such as racial diversity and education attainment levels. We tested hypotheses for each factor separately to determine which factors most strongly predicted campaign contributions and voter support.

## Research Question #2: Causal Inference

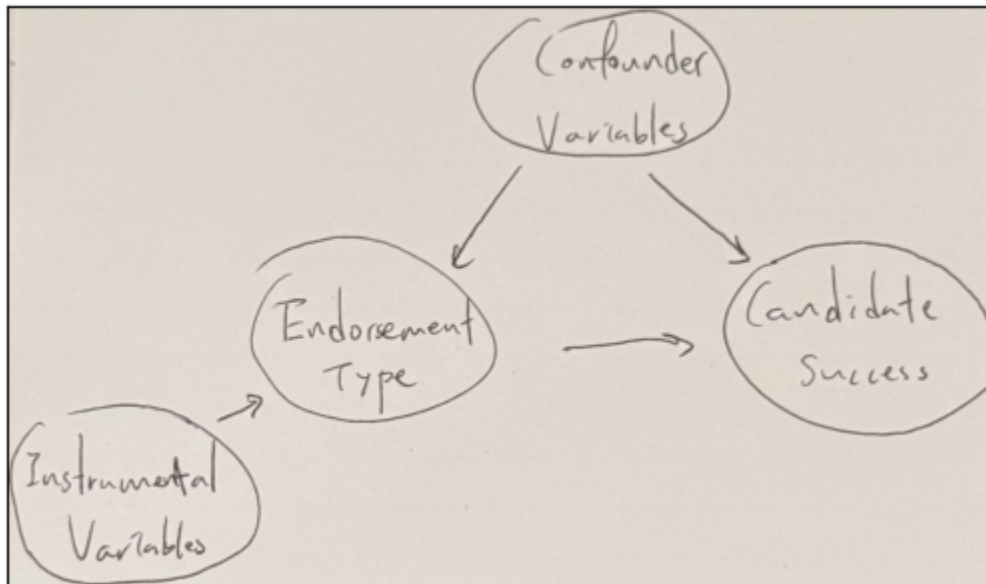
Our second research question is: What types of endorsements are most influential in determining a candidate's success in primary elections?

### Methods

For this part, we will use the data from FiveThirtyEight: 2018 Primary Candidate Endorsements. We will be using propensity score + matching to control for confounders and compare similar candidates with different endorsement types. Since the confounders influencing treatment assignment might not be fully captured, we have introduced some instrumental variables that affect endorsement types and don't directly affect candidate successes, but could affect the likelihood of receiving endorsements, like election race type (regular vs special election).

- outcome variable  $Y^{(i)}$  is the success of candidate  $i$  in the primary election.
- treatment variable  $Z^{(i)}$ : is the type of endorsements candidate  $i$  receives.
- instrumental variable  $W^{(i)}$  is the variable that influences endorsements but does not directly affect election success of candidate  $i$ .
- confounder  $X^{(i)}$  are candidate characteristics.

DAG:



### Results

(Please refer to the notebook for a clearer explanation)

Using lab 8 as a reference, we splitted our regression method into 2 parts. In the first stage, we predicted the treatment variables with the instrumental variables with logistic regression as all the features we are using are binary. Then, we regressed the candidate's success in the primary election onto the predicted treatment variables. Because there exist confounders, we addressed them using propensity scores and matching techniques.

We first estimated the causal effect of all types of endorsements in “dem\_candidates.csv” on the “Won Primary”.

A positive ATE indicates the treatment improves the outcome, while a negative ATE suggests it worsens the outcome. From the calculated table, we saw that democratic candidates with party support were 77% more likely to win the primary than similar candidates without party support, followed by endorsements from Indivisible and Elizabeth Warren. On the other hand, endorsements from Our Revolution and Justice Democrats were 42% less likely to win the primary than other democratic candidates.

Similarly, we went on to process “rep\_candidates.csv” on the “Won Primary”. However, because this dataset did not provide any other variables that could be identified as confounders or instrumental variables, we assumed unconfoundedness and solely estimated the treatment effect on the outcome without additional adjustments.

Based on the coefficients from the returned table, we saw that candidates endorsed by Donald Trump had a significant positive effect on the likelihood of winning the primary election, followed by endorsements from Koch brothers’ political network and National Republican Senatorial Committee, and it looked like endorsements from Susan B. Anthony List slightly reduced the likelihood of winning the primary election.

## Discussion

Throughout the process, we were faced with some limitations involving our propensity scoring, unconfoundedness assumption, and the generalizability of our results.

Our propensity scoring for the Democratic dataset depends on matching, which may prove difficult since our data contained several different types of endorsements. This large number of variables limits the pool of candidates with the exact same endorsements, which makes matching more difficult to achieve and thus our propensity scoring loses some reliability.

We also relied on our unconfoundedness assumption for the results, which could be very overzealous as there are several potential confounding factors like media exposure, net worth, and the social media sentiment of a candidate that our data – particularly the Republican dataset – did not account for. By not including these as confounders, our results may lose some degree of validity.

Also, since we are only working with data from the 2018 Primary Endorsements dataset, we are unable to make conclusions about election data from other years, and for other elections such as midterm and general elections from 2019 to the present. By generalizing our data, we may lose out on new data or nuances (such as political climate shifts) that were not present during the 2018 election season.

Some additional data that would be useful for answering this question of what endorsements are more useful would include: campaign contributions by type (e.g. PACs vs small individual donations), social media sentiment, net worth, amount of rallies attended. Contributions by type may affect endorsements and candidate success, since PACs have large influence over endorsement and funding is already a proven factor in success rates. Social media sentiment may motivate figures to endorse, and they also affect success. Net worth similarly indicates larger resources that may imply a larger support network for candidates, and rallies attended may cause more figures/groups to endorse while simultaneously increasing success. All this additional data can therefore be useful since they are all potential confounders, and by including them in our causal inference we can eliminate some false associations in our results.

We are moderately confident that there is a causal relationship between these endorsements and a candidate's success (at varying degrees) since the use of propensity scoring and instrument variables address confoundedness to some extent. For Democratic candidates, party support had the highest ATE, indicating a causal relationship. Another explanation for this is that the Democratic party as a whole, regardless of high-profile endorsements, may have been largely mobilized in 2018 due to discontent with the Trump administration, which also explains the massive voter turnout in the subsequent 2020 election. For Republican candidates, a Trump endorsement had the largest coefficient, and thus causal relationship to candidate success. Given that Trump was the sitting president in 2018, his endorsement likely held very significant value for the success of a candidate in his party.

## Conclusion

### Outcomes summary

The key finding of our research reveals that regional economic conditions, such as state income levels and GDP, play a significant role in shaping campaign financing and voter support. Candidates in economically prosperous regions often have better financial backing, which correlates with a higher likelihood of success in primary elections. Additionally, endorsements from influential figures and organizations were shown to have a strong impact on election outcomes, with high-profile endorsements providing a noticeable advantage. However, the findings also highlight a transparency gap in how campaign contributions are utilized and their effectiveness in swaying voter support. This underscores the need for clearer public disclosures about campaign spending and endorsement practices. Improving transparency would not only enhance donor trust but also empower voters to make more informed decisions, ultimately contributing to a fairer electoral process.

### Critical Evaluation

Our analysis found a positive correlation between tax revenues and mortgage loan volumes with campaign contributions. We also found a negative correlation between poverty rates and campaign contributions. This shows that economic factors play a key role in affecting political behavior. People with large loans and that pay high taxes seem to be more civically engaged

because any economic shift would affect them significantly. For example, an individual losing their job because of a recession could default on their mortgage. The main issue with our data is multicollinearity. Tax revenues, mortgage loan volumes, and poverty rates very often overlap. It is hard to isolate and determine how many of them influence campaign contributions. The domain experts we should consult are political scientists. We would ask them what is the base fixed cost to running a campaign. We feel like some of the lower populated states' campaign spending are skewed compared to medium and larger states. We would also ask them whether higher property taxes would mean less home ownership. This could mean many people don't have any mortgage loans since they are just renting. Would total housing costs be a better metric to analyze? Moreover, there is a positive relationship between endorsements and winning the primary. One issue with our data is that not every person or organization endorses everyone. For example, Main Street only endorsed certain candidates which could skew the efficacy of their endorsements. We need more background research if these groups have any methodology to determine their endorsements. The domain experts we should consult with are political scientists. We would ask them if the endorsed candidates values and activist groups (NRA, Right to Life, etc) values match. This way we could determine whether or not they are just endorsing safe seats. We could also see if their values are reflected in the electorate. For example, are candidates actually voting for the most pro-gun candidates endorsed by the NRA. We also need to adjust for endorsements for federal and state candidates. Voters select state candidates based on much more local issues compared to federal elections. States have a lot of power in America because of our decentralized nature. For example, 90% of public school funding comes from the state while 10% comes from federal. A voter that cares about public schools has different priorities when voting for a federal candidate. Candidates endorsed by federal politicians should also have a larger impact. For example, a Trump endorsement for a House candidate should have a larger impact than an endorsement for a gubernatorial candidate. The House rep will be directly responsible for enacting Trump's agenda while the governor runs his own state. Our results are generalizable because we take into account candidates and data from every state. We also use official government data when it comes to economic and campaign spending data.

## Recommendations

A potential follow-up study could examine what types of campaign spending influence the likelihood of winning an election. For example, it could analyze whether campaigns that allocate more funds to advertising or maintaining an active and trending online presence are more likely to gain voter support and secure a victory compared to those with lower expenditures in these areas.

Based on our findings, a key recommendation is to enhance the transparency of campaign financing and endorsements as we found that it was very difficult to find a clear dataset that describes this from a trustworthy source. This would allow individual donors to better understand how their contributions are being used, whether the funds are effectively aiding the campaign's success, and to what extent they align with the campaign's stated goals. While government bodies would oversee implementation of requirements for detailed public disclosures about the sources and uses of campaign contributions, some political groups might

raise concerns about privacy or the administrative challenges involved, and transparency measures might spark debate over donor privacy, the risk of revealing strategic details to opponents, and varying interpretations of what constitutes sufficient disclosure. Nevertheless, clearer information could be considered to benefit voters by enabling more informed decisions to provide fairness to the elections and help them better understand what they are paying for .