American Partisanship in Relation to National Attention and Citizen Engagement

Rida Anjum, Ron Barzilay, Anna-Maria Gueorguieva, Sifath Mannan

1. Introduction

Partisan polarization in the United States is widely believed to be consequential and growing, with Americans from the political elite to average voting citizens becoming more extreme in their political opinions. One question that arises from increasing polarization is whether national attention fixates on elections in areas with higher extremism or more moderation in political viewpoints. We will attempt to answer this question on two levels: elite and ground-level.

One way that political elites communicate their attention and care for an election is through endorsements. Some elites push forward entire political ideologies while others represent stances on single issues. We will use hypothesis testing to look at how their choice to make endorsements varies based on the partisan lean of a district.

On the ground level, average citizens express their care for an election by turning out to vote. There is no clear consensus on whether voters pay more attention to elections in more extreme or moderate districts. Extreme districts tend to have more politically passionate citizens, but voters in moderate districts tend to have much more sway in toss-up elections. We will use causal inference techniques to try to address this debate by answering whether partisan lean causes changes in voter turnout.

2. Data Overview

2.1 FiveThirtyEight primary-candidates-2018 Data

This dataset is a census of all 811 candidates who appeared in Democratic party primary elections and 774 candidates who appeared in Republican Party primary elections in 2018 - not counting races featuring a same-party incumbent. A valuable feature within this data set is the endorsements that candidates received from various groups and individuals. The dataset excludes races with an incumbent running due to the fact that there are many primaries where incumbents run unopposed or have no serious challengers. In short, this dataset is focused on "competitive" primaries; this may be beneficial to our research question as previous research has shown that partisan voting is often viewed as a competition or rivalry.[1]

Details on the elections are public data, and candidates are likely aware of the endorsements they received. Each row represents a candidate and their endorsements, though we will go on to group the data by Congressional districts. Each data point will represent a district, meaning that we can make conclusions about Congressional districts at large.

---

[1] Cooper, Chad, and Patrick Miller. "Partisan voters treat politics and elections like a competitive sports rivalry."

We have no concerns about selection bias, measurement error, etc. because this is a census of public data.

2.2 FiveThirtyEight 2018 partisan-lean Data

This dataset is a census of all 435 US Congressional districts' FiveThirtyEight partisan lean metric in 2018. Partisan lean is calculated as "the average difference between how a district votes and how the country votes overall, with 2016 presidential election results weighted 50 percent, 2012 presidential election results weighted 25 percent and results from elections for the state legislature weighted 25 percent." We chose to add this source because while the original FiveThirtyEight data included some of this information, it only had the partisan lean for the districts with Democratic primaries. We found and downloaded this data from the exact same GitHub repository as the FiveThirtyEight primary candidates dataset.

This dataset excludes no districts, and it stems from public data. Each row represents a Congressional district, allowing us to make conclusions on a district level. Because of the fact that this is a census, we have no concerns about selection bias, measurement error, or convenience sampling.

2.3 US Census Data

This dataset is the information collected on the US census grouped by district. It includes data on population, voting rates in the 2018 general election, age, sex, poverty, education, and race. It is generated from the 2010 US Census and 2018 American Community Survey. Both are generated using a combination of internet and mail questionnaires, with telephone and personal visit used as follow-up to nonresponse. The former (as indicated by its name) is a census while the latter includes estimations made from samples, though it has data on each Congressional district.

We chose to add this data to be able to account for confounders in our analysis, as age, race, poverty, etc. were quite obvious ones that we could account for on a district level. We downloaded the dataset from the official US Census website as several separate Excel tables.

With US Census data, there are always questions about systematic exclusion, with undocumented residents being the primary concern. Even in the aggregate there may be misleading data due to their exclusion. Census participants voluntarily answer questions so they are aware of the collection of their data and which parts are made public. Each row represents a Congressional district, so the interpretation is the same as for the other datasets. The Census also usually has concerns about selection bias and measurement error because the type of people that avoid answering the Census may have systematic characteristics and people may lie in their answers (e.g. lying about age, poverty).

3. Research Questions

Research Question #1: Is the partisan leaning of a district associated with receiving specific types of endorsements? (Hypothesis Testing)

This research question aims to understand the relationship between partisanship and national attention on elections, where we use the usage of endorsements as a proxy for national attention. Individuals and groups give endorsements to the most salient elections; our curiosity is whether the most salient elections are in districts with more or less partisan lean. There also may be variance in the perception of how salient an issue is from the perspective of different interest groups, so we will test the association of partisan lean and the probability of garnering an endorsement by multiple "endorsement groups."

One real-life application of this research question is in planning election campaigns. If a candidate runs in a district with a partisan lean level associated with receiving endorsements from certain types of organizations, they should account for the effects of extra media coverage on that issue in planning their advertising strategy.

Hypothesis testing is the best way to answer this question because we want to not only know whether there are associations between partisan lean and endorsement types, but also whether they are statistically significant, i.e. not spurious. We use multiple hypotheses given that there are many different categories of endorsements that can be created, and one simple hypothesis such as 'partisan lean is/is not associated with endorsements' is not able to capture the nuances of different types of endorsements.

We will group the endorsements in the FiveThirtyEight dataset both by districts and by types of endorsements. For example, if in California District 17 the NRA endorsed any candidate running in either the Republican or Democratic Primary, we would indicate this with a 1. We make no distinctions as to who the endorsements are for or whether they were made for a Democratic or Republican primary, and we assume that these distinctions are irrelevant for our research question. We are essentially asking: did any (specific type of) endorsers bother to make a statement about this district's primary election?

These are the eight different endorsement categories we created to explore partisanship and national attention:

1. Any organization.
2. Gun-rights related organizations. Examples: the NRA, Gun Sense.
3. Abortion related organizations. Examples: Susan B. Anthony, Right to Life, Emily's List.
4. Left-leaning political figures. Examples: Bernie Sanders, Joe Biden, Elizabeth Warren.
5. Right-leaning political figures. Examples: Donald Trump, Steve Bannon.
6. Any political figure.
7. Progressive organizations. Examples: Justice Dems, Our Revolution, PCCC, Indivisible, WFP.
8. Conservative organizations. Examples: Main Street, Koch Brothers, Tea Party.

We chose these groups because, besides being the most prominent groups in the FiveThirtyEight dataset, they represent two critical issue topics (abortion and guns) as well as political figures and organizations divided along party lines, which are some of the most relevant endorsers. We also combined some endorsement groups (e.g. left- and right- leaning political figures) to have a balance of granularity and generalization.

Research Question #2: Does greater partisan leaning cause increased voter turnout? (Causal Inference)

There are debates about whether partisan polarization makes voters more alienated from politics because people tend to be moderate in their political opinions, or makes voters more engaged in politics because they are passionate about their party affiliation. This research question will attempt to address this debate. Political parties could use the conclusions from this research question to determine which districts need the most help with voter turnout (i.e. higher vs. lower partisan-leaning districts). They can also choose either moderate or extreme candidates in order to garner more votes.

Causal inference is a good fit for this question because it will allow us to directly see the influence of partisan lean on voter turnout in the general election by district after accounting for confounding variables, so that we can see the effect of partisanship alone. Measures like average treatment effect will give us clear, interpretable metrics for the effect of partisan lean.

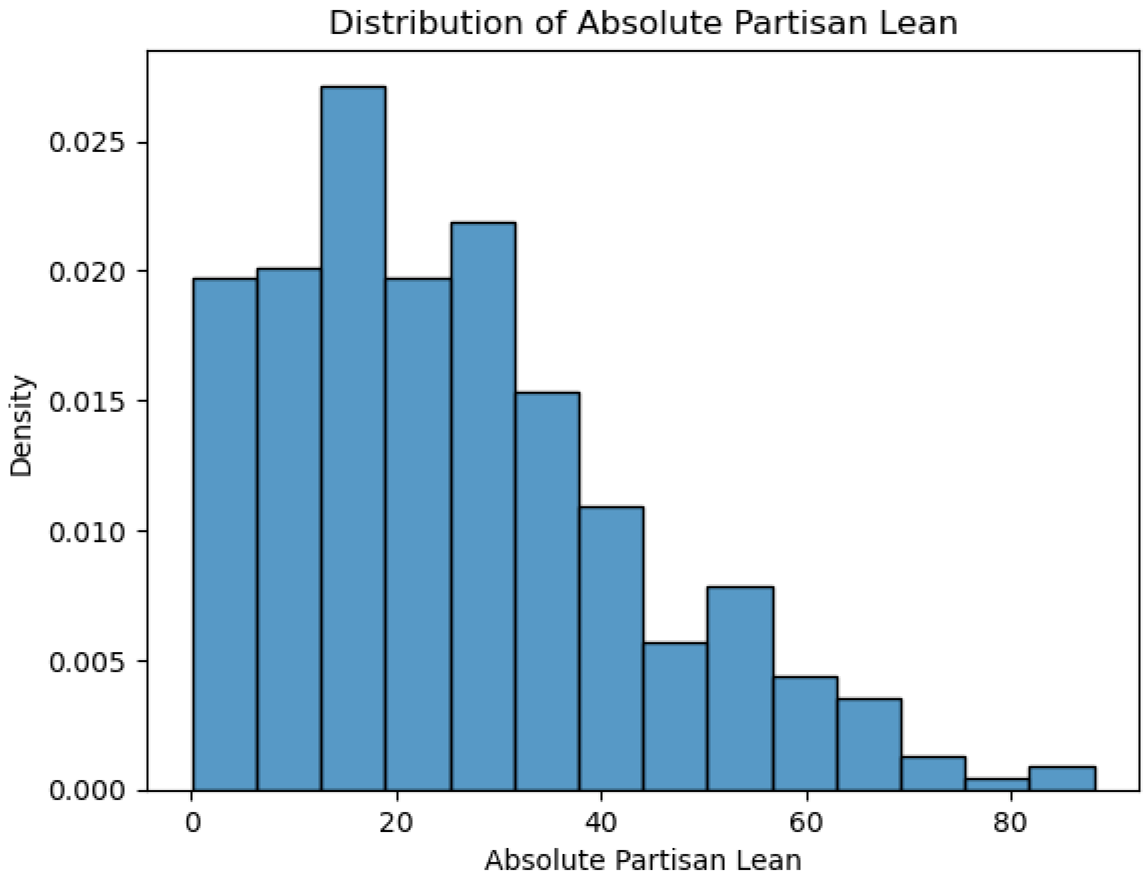
4. Exploratory Data Analysis

4.1 Data Cleaning

The bulk of the data cleaning consisted of converting the outside census data (which includes voting rates and distributions of race, age, education, etc. by district) into a usable csv format. Then, we merged the FiveThirtyEight endorsement data for Democratic and Republican primaries, followed by creating the endorsement types/groups that we wanted to study (e.g. abortion organizations, gun-related organizations) and creating binary columns to indicate whether any candidates in the district received any endorsement from such an organization. We also added the full partisan lean column missing from the original FiveThirtyEight. Then we joined the endorsement and census data, dropping rows that had no data for voter turnout (districts 10, 14, 21, and 24 in Florida and district 9 in North Carolina).

The key takeaway from data cleaning is that we have no data on districts that had no primary elections (Democratic or Republican) in 2018, about 70 districts. It is not immediately obvious what this should mean, but we should move forward understanding that we are not including districts that are uncompetitive for either party. Other than that, we have full partisan lean, endorsement, and demographic data for 363 districts, which should be a sufficient sample to make statistically significant conclusions.

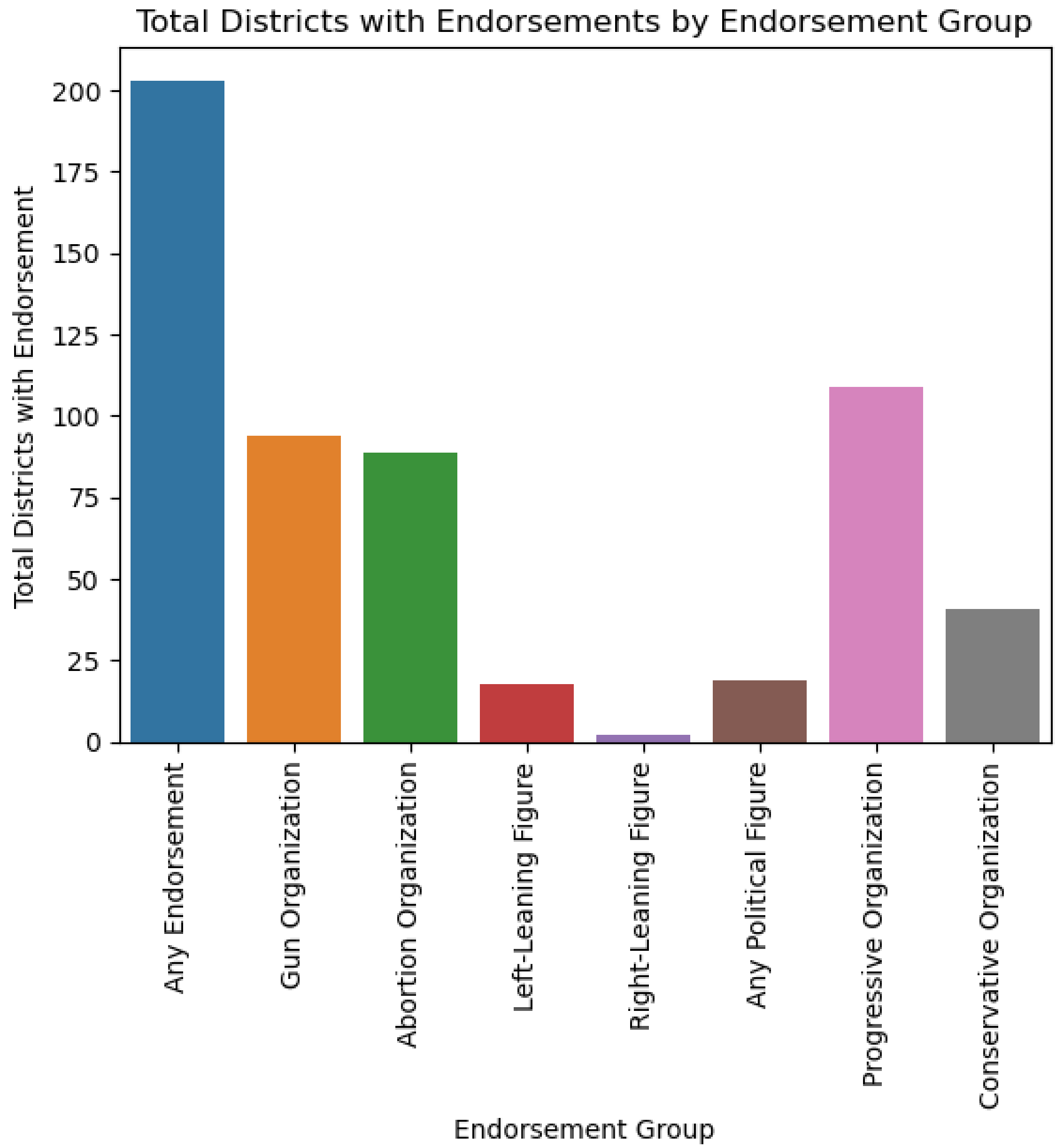
4.2 Partisan Lean



We begin by looking at partisan lean, which is essentially the explanatory variable in both of our research questions. The key metric is the absolute value of partisan lean; whether a district is left-leaning or right-leaning is not relevant for our research question; we are more interested in the idea of extreme versus moderate. The distribution of absolute partisan lean is right-skewed, with many values close to zero and around 20, and fewer stretching all the way to about 80. We will use partisan lean as a basis for our understanding of partisanship and extremism, i.e. the higher the absolute partisan lean of a district, the less moderate and more polarized a district is. As we can see, we have a sizable sample of low and high absolute partisan lean, which should allow us to answer our research questions.

One important decision we had to make later on was to decide a threshold for “normal” and “high” absolute partisan lean, so that we could make a binary variable for causal inference via inverse propensity weighting in research question #2. While we considered simply using the median, it is only around 22, which we did not think reflected the distribution of “normal” and “unusually high” partisan lean. We saw that any observation between 0 and 40 was quite typical, while a tail of outliers started to form around 40; as such, we decided to use 40 as a threshold: anything higher we labeled as a 1 for a treatment of “high partisan lean.”

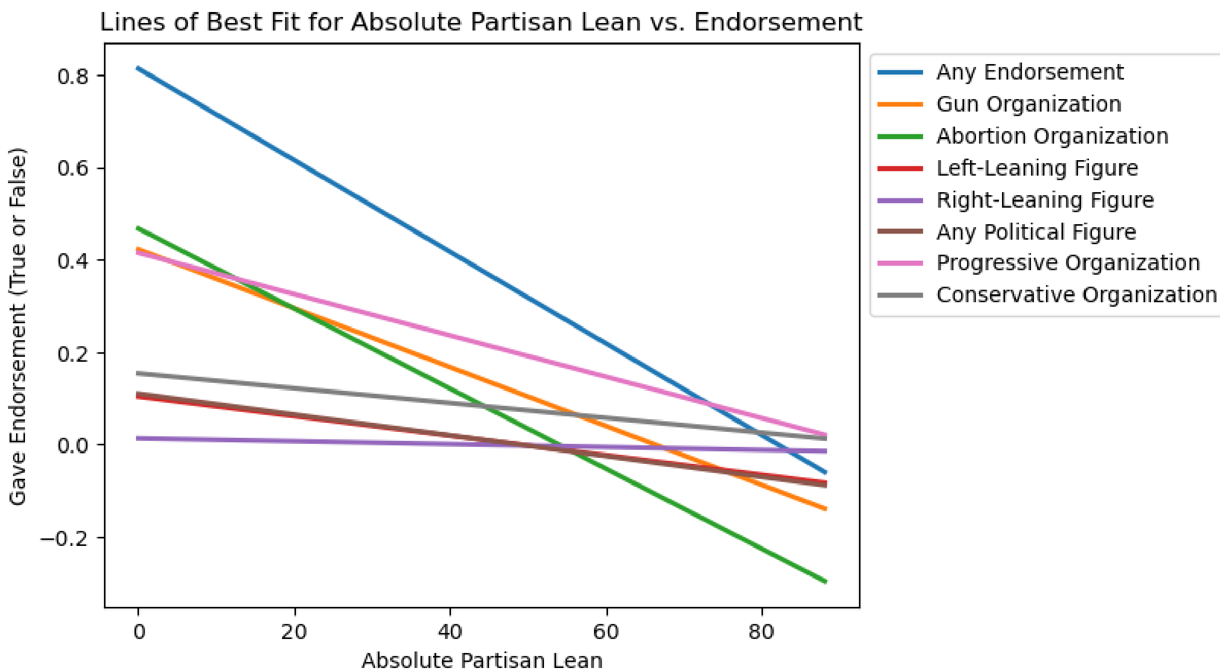
4.3 Endorsement Groups



This graph looks into the endorsement groups that we defined for the first research question and looks at the total quantity of each of these endorsement groups across all districts in the dataset. In our data cleaning we accounted for all endorsements given to candidates and grouped by district, so rather than showing a too granular view of per candidate endorsements, we can aggregate by district and effectively compare it to partisan lean later on.

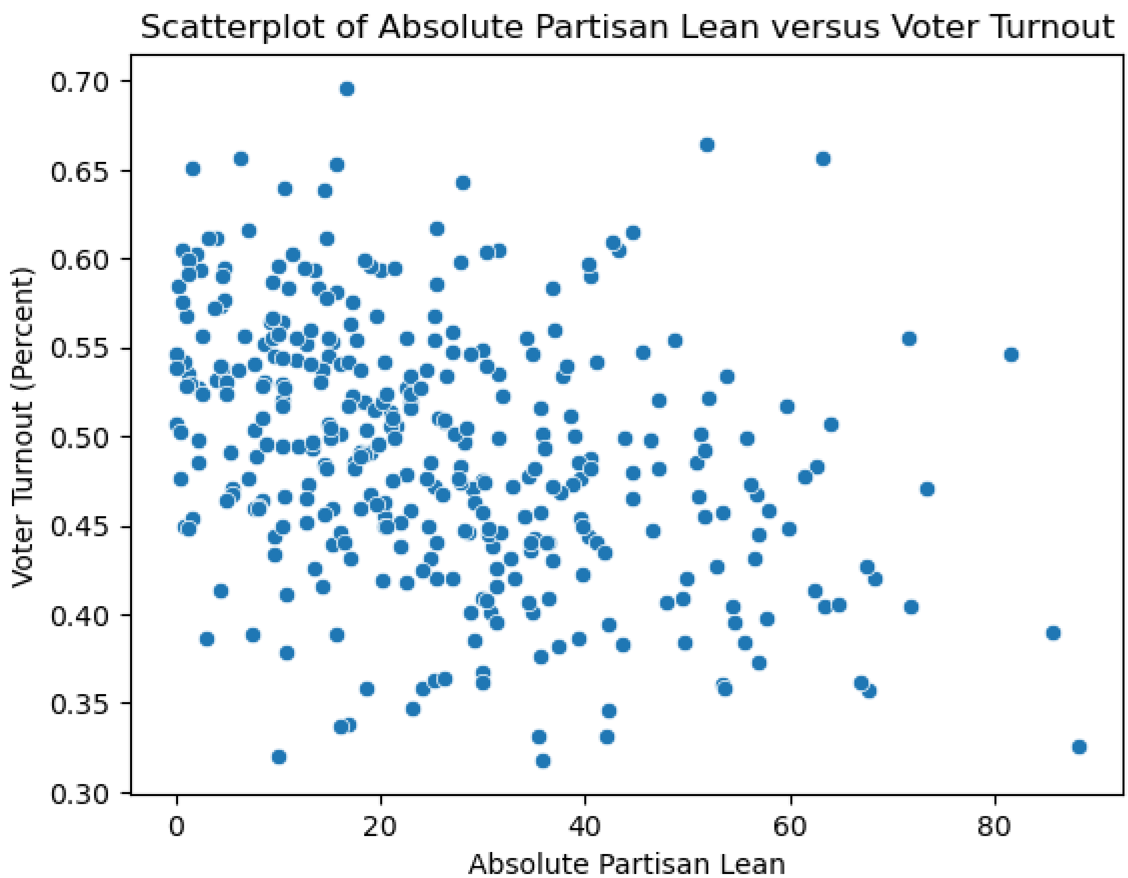
A noticeable observation in this graph is the much higher number of progressive endorsements given compared to conservative ones, and a similar trend is followed by left-leaning political figures giving more endorsements than conservative political figures. There are very few political figure endorsements in general. This will likely affect observations for the first research question; results for right-leaning figures and conservative organizations could be misleading since there are so few overall.

4.4 Partisan-Lean/Endorsement Relationship by Endorsement Group



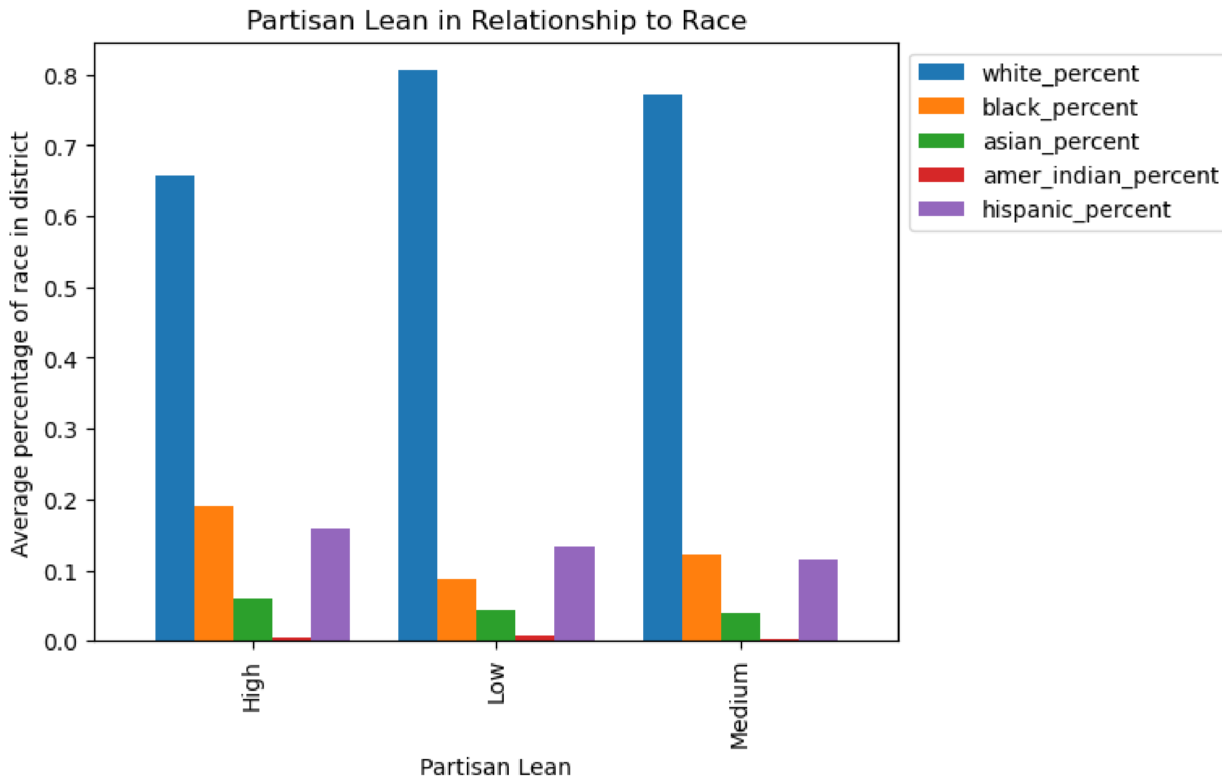
This plot shows the line of best fit where the x-axis represents partisan lean and the y-axis is a binary variable representing whether that district got an endorsement from one of the relevant groups. A slope farther from zero means a stronger association between partisan lean and endorsements. As predicted, the relationships for conservative organizations, right-leaning figures, and political figures in general are extremely minimal, likely because there is so little data overall. However, the associations between partisan lean and endorsements for gun organizations, abortion organizations, and progressive organizations appear to be negative. In this case, this means that less partisan lean is associated with the likelihood of an endorsement in that district. We will investigate whether any of these relationships are statistically significant in our first research question.

4.5 Association for Partisan Lean/Voter Turnout



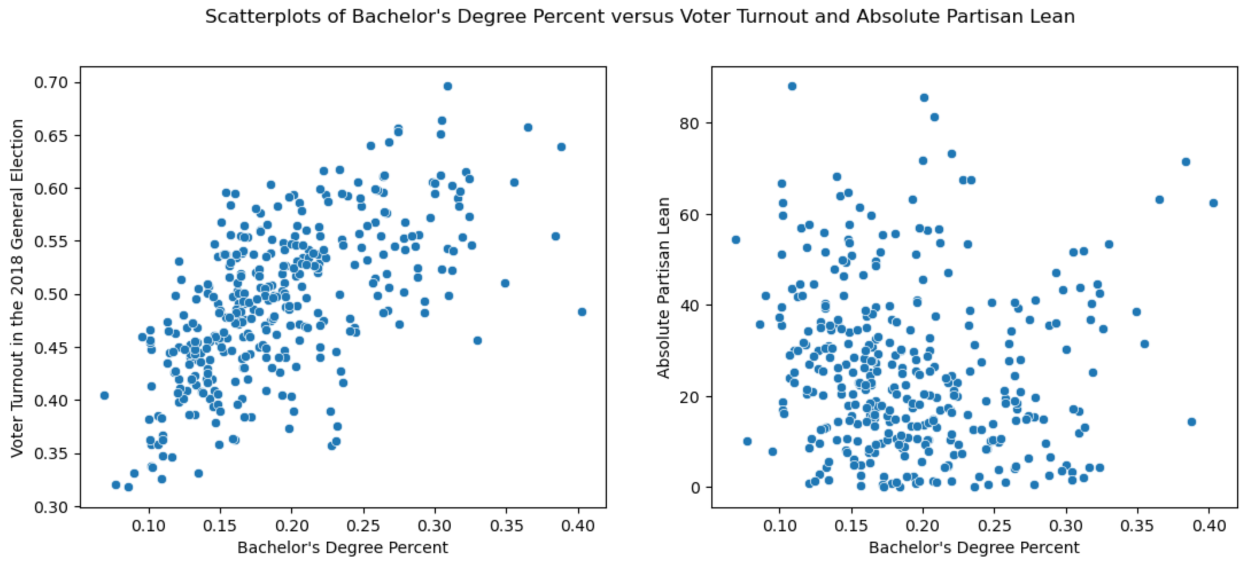
Starting to look at voter turnout which relates to our second research question, we can see a weak but noticeable negative association between partisan lean and voter turnout. This suggests a basis for our second research question, which is whether partisan lean causes changes in voter turnout. Initially, it appears as though less partisan lean is associated with higher voter turnout, but we must account for our confounders later to make any causal claims.

4.6 One Confounder: Race



Our identities are relevant to how we vote, thus race is a confounding factor to voter turnout. Studying racial demographics in relation to partisan lean may show interesting insight about how racial identities shape larger voting trends. The above graph separates the continuous variable of partisan lean into three categories: high, medium, and low. In general, there is no obvious difference in the distribution of races for different amounts of partisan lean, so perhaps race is not too strong of a confounder. This EDA most likely fails to capture the granularity of race and mixed-race nuances, however, it is a starting point to understand if race is an interesting feature that impacts partisan lean when taken across all districts.

4.7 Another Confounder: Education



In this graph, we see another clear confounder in the relationship between absolute partisan lean and voter turnout in the general election: education. It is plausible that the distribution of education levels in a Congressional district affects both the partisanship and voter turnout in that district given the importance of education in public engagement. Here, we see that there is a clear relationship between the proportion of citizens with a bachelor's degree in a district and the voter turnout, and a more subtle relationship between the same proportion and absolute partisan lean. We will account for this confounder in our second research question, along with many more.

Another important observation is that the effect of this confounder is not obviously linear. While it is tempting to take the results of OLS outcome regression at face value, we have to consider that the confounders have nonlinear relationships with the treatment and outcome variables. This is the impetus for why we will use an alternative method: inverse propensity weighting.

5. Research Question #1:

Is the partisan leaning of a district associated with receiving specific types of endorsements?

5.1 Methods

To determine whether partisan lean of a district is associated with different types of endorsements, we performed hypothesis testing on the Pearson correlation coefficient between the two variables. The hypotheses are as such:

H_0 (null): there is no association (association is zero) between partisan lean and the presence of a specific endorsement type (e.g. abortion-based organization, conservative group, left-leaning politician) and any apparent association is due to random chance.

H_1 (alternative): there is an association (positive or negative) between partisan lean and the presence of a specific endorsement type.

We use correlation as our measure because it is a relatively easy-to-interpret metric for analyzing the positive, negative, or nonexistent link between partisan lean and endorsement types. We will first use a standard p-value cutoff of $\alpha = 0.05$.

One known problem for our approach is that our p-values are not independent of each other. Clearly, the correlation of partisan lean and all political figures and partisan lean is dependent on the correlation of partisan lean and left-leaning figures. However, we do not see an issue with this since we are simply combining categories to create more broad ones, such that we will be able to see both granular and generalized views of similar relationships.

To account for any spurious statistically significant correlations of any of the endorsement groups and partisan lean, we used two correction methods: the Bonferroni correction and Benjamini-Hochberg Procedure. The Bonferroni correction controls for family wise error rate (FWER), or the probability of having at least one false positive. The Benjamini-Hochberg procedure controls for false discovery rate (FDR), which is the expected value of the false discovery proportion, taken with respect to the randomness of the eight decisions. For the Bonferroni correction, we will implement it such that FWER < 0.05. For the Benjamini-Hochberg procedure, we will implement it such that FDR < 0.05.

5.2 Results

Hypothesis Test Results
For α = 0.05

Endorsement Group	Pearson Correlation with Absolute Partisan Lean	p-value	Rejected at $p < 0.05$?	Rejected with Bonferroni correction?	Rejected with Benjamini-Hochberg Procedure
Any Endorsement	-0.353727	3.8699e-12	Yes	Yes	Yes
Gun-related Organization	-0.257660	6.4771e-07	Yes	Yes	Yes
Abortion-related Organization	-0.356720	2.4739e-12	Yes	Yes	Yes
Left-leaning figure	-0.172480	0.0009682	Yes	Yes	Yes
Right-leaning figure	-0.075207	0.1527	No	No	No
Any political figure	-0.179029	0.0006104	Yes	Yes	Yes
Progressive organization	-0.172731	0.0009515	Yes	Yes	Yes
Conservative Organization	-0.089606	0.08823	No	No	No

As seen in the above tables, absolute partisan lean is negatively correlated with the presence of an endorsement from all endorsement groups. In other words, an increase in partisan lean is associated with less likelihood to have an endorsement from a group. At the simple p-value threshold of $\alpha = 0.05$, we can say that these relationships are statistically significant for endorsements in general, gun-related organizations, abortion-related organizations, left-leaning figures, political figures in general, and progressive organizations. In other words, it is unlikely that these associations are actually zero and we only see negative correlations by random chance. The same is not true for right-leaning figures and conservative organizations. We can explain the minimal relationship for right-leaning figures by the lack of right-leaning endorsements there were overall, but it is interesting that there were still a sizable number of endorsements from conservative organizations, but they were not associated with partisan lean.

As previously mentioned, since we did hypothesis testing for multiple groups at once, we controlled for family-wise error rate and false discoveries using Bonferroni correction and the Benjamini-Hochberg procedure, respectively. For the Bonferroni correction, we used $\alpha = 0.05$ — i.e. ensuring that the probability of having at least one false positive (FWER) in our eight hypothesis tests is less than 0.05. For the Benjamini-Hochberg procedure, we used $\alpha = 0.05$ — i.e. ensuring that the expected false discovery proportion in our eight hypothesis tests is less than 0.05.

5.3 Discussion

After applying the correction methods, we report that the Bonferroni correction did not change any of the rejected hypotheses; the same associations between endorsement groups and partisan leans remained. Seeing as the Benjamini-Hochberg procedure has a less strict cutoff than Bonferroni, it also did not reverse any of the rejections. We can say with relative confidence that the associations we found are real and not due to spurious p-values found because of testing multiple hypotheses at once.

Since endorsements ostensibly come long after the partisan lean of a district has settled (such shifts occur over very long periods of time), these negative correlations could be interpreted to mean that higher partisanship leads to fewer endorsements in a district. So, endorsers focus on elections in “swing” or “moderate” districts as opposed to heavily polarized ones, at least for the statistically significant endorsement groups. As per the aggregate of the tests, candidates running in highly partisan elections should not anticipate organizations making endorsements. Per our individual tests, this result is especially true for abortion organizations and gun-related organizations, and somewhat true for left-leaning figures and progressive organizations. As for the lack of an association for conservative organizations, we can argue that conservative organizations make endorsements with less regard for the partisanship of a district than progressive ones.

The primary limitation of our analysis is the lack of data on endorsements by political figures. The FiveThirtyEight data was sparse overall, indicating that many districts do not get endorsements from large national organizations. It was especially sparse for endorsements from political figures; there were few figures and they made few endorsements around the country, so any conclusions about their associations with partisan lean should be taken with suspicion. The associations are also generally weak, so higher partisan lean should not be taken to guarantee fewer endorsements in a district. However, we are confident that p-hacking was not an issue in our analysis, considering that both of our correction procedures did not alter any results.

If we had more data, our first turn of focus would be on other types of endorsements, and the inclusion of more endorsers (especially political figures). It would also be valuable to perform the same tests including data on primary elections with incumbents (which FiveThirtyEight excluded). It would be interesting to see if these relationships hold on less competitive elections.

6. Research Question #2:

Does greater partisan leaning cause increased voter turnout?

6.1 Methods

In this analysis, we used the absolute partisan lean of a district as our treatment variable, and the proportion of voters who turned out in the 2018 US general election as our outcome variable. However, we used the treatment variable in two different ways for two different approaches. First, we used absolute partisan lean as a continuous variable for an OLS regression approach. However, we also wanted to adjust for confounders in a way that would recognize their nonlinearity; confounders like the percent of people in poverty in a Congressional district or the distribution of educational level (which we saw some of in section 4.7) do not necessarily have a linear relationship with partisan lean or voter turnout. Inverse propensity weighting (IPW) would allow us to adjust for confounding variables in a nonlinear manner. Since IPW as shown in Data 102 is incompatible with continuous treatments, we decided to also transform absolute partisan lean into a binary categorical treatment. As discussed in section 4.2, any district with an absolute partisan lean higher than 40 was labeled with a 1, and any district with less than that was labeled 0.

As for the confounders that we included, we thought that the most important ones would have to do with the identities of citizens in a district. We included the following confounders, based on distributions from the US census:

- Age distribution
- Sex distribution
- Percent in poverty
- Education distribution
- Racial distribution

We recognize the potential of more confounders. Wealth distribution in a district is one that comes to mind immediately (although we hope something like education is related enough). We hope that the district-level census data will cover some of the clearest confounders.

As previously mentioned, we will use two different methods to adjust for the confounders: outcome regression and IPW. In outcome regression, we will simply include the confounders as explanatory variables alongside absolute partisan lean in an OLS regression model to predict voter turnout. In the IPW method, we will give each district a propensity score (using the causalinference package in python, i.e. the probability of getting the treatment (a district with partisan lean > 40) given the confounders. Then, we will use those propensity scores in calculating the average treatment effect (ATE), i.e. the estimated difference in voter turnout for a district with high vs. normal partisan lean. As is standard, we will also remove any values with a propensity score below 0.1 and above 0.9 so as to reduce the variance of the estimate. We do not recognize any clear colliders in our dataset; we could not think of any ways that partisan lean or voter turnout could affect something like a district's age distribution.

6.2 Results

Here is the output from the OLS Regression:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          voting_percent   R-squared:                       0.778
Model:                             OLS   Adj. R-squared:                  0.764
Method:                  Least Squares   F-statistic:                     54.29
Date:                 Thu, 08 Dec 2022   Prob (F-statistic):           3.53e-97
Time:                         11:17:53   Log-Likelihood:                 613.92
No. Observations:                  363   AIC:                            -1182.
Df Residuals:                      340   BIC:                            -1092.
Df Model:                           22
Covariance Type:             nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.0037      0.021      0.175      0.861      -0.038       0.045
18-29_percent            -3.0354      4.496     -0.675      0.500     -11.879       5.808
30-44_percent            -2.8194      4.502     -0.626      0.532     -11.674       6.035
45-64_percent            -2.5736      4.500     -0.572      0.568     -11.426       6.278
65_plus_percent          -2.6400      4.491     -0.588      0.557     -11.473       6.193
men_percent               9.6803      5.790      1.672      0.095      -1.709      21.069
women_percent             9.2430      5.775      1.600      0.110      -2.117      20.603
poverty_percent           0.0792      0.124      0.640      0.523      -0.164       0.323
less_9th_percent         -6.6106      3.281     -2.015      0.045     -13.064      -0.157
9th-12th_percent         -6.1326      3.314     -1.851      0.065     -12.651       0.386
high_school_percent      -5.9069      3.281     -1.800      0.073     -12.360       0.546
some_college_percent     -5.5180      3.280     -1.682      0.093     -11.970       0.934
associates_percent       -5.2172      3.281     -1.590      0.113     -11.670       1.236
bachelors_percent        -5.4308      3.290     -1.650      0.100     -11.903       1.041
graduate_percent         -5.3350      3.275     -1.629      0.104     -11.777       1.107
white_percent            -0.5814      0.666     -0.873      0.383      -1.892       0.729
black_percent            -0.5775      0.663     -0.871      0.384      -1.881       0.726
asian_percent            -0.8792      0.673     -1.306      0.193      -2.204       0.445
amer_indian_percent      -1.0924      0.635     -1.720      0.086      -2.342       0.157
nat_hawaiian_percent     -1.3743      0.829     -1.658      0.098      -3.005       0.256
other_race_percent       -0.5940      0.675     -0.880      0.379      -1.921       0.733
two_more_race_percent    -0.3145      0.758     -0.415      0.679      -1.806       1.177
abs_partisan_lean        -0.0002      0.000     -0.876      0.382      -0.000       0.000
==============================================================================
Omnibus:                      308.239   Durbin-Watson:                   1.422
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18159.260
Skew:                          -3.062   Prob(JB):                         0.00
Kurtosis:                      37.104   Cond. No.                     1.58e+05
==============================================================================
```

As we can see, when accounting for our chosen confounders, the coefficient for absolute partisan lean is close to 0, has a confidence interval that includes zero, and is not statistically significant (p = 0.382). So, we cannot say that with any reasonable level of confidence that the absolute partisan lean of a district causes any change in voter turnout.

Here is the output from the IPW ATE estimate, once again provided by the causalinference package, which uses a doubly-robust version of the Horvitz-Thompson weighting estimator:

```
Treatment Effect Estimates: Weighting

                     Est.       S.e.          z       P>|z|      [95% Conf. int.]
--------------------------------------------------------------------------------
           ATE      0.006      0.008      0.685      0.493     -0.010      0.022
```

For the sake of thoroughness, we also used the standard inverse propensity weighting formula presented in Data 102 which garnered similar results: an estimated ATE of 0.0098 and a 95% confidence interval of {-0.055, 0.075}[2]. Once again, we cannot claim that partisan lean causes a change in voter turnout with confidence. The average treatment effect on voter turnout for high partisan vs. normal partisan lean is close to zero, has a confidence interval that includes zero, and has a high p-value ($p = 0.493$ in the Horvitz-Thompson weighting estimator).

6.3 Discussion

Our approach is limited in that it is only connected to the 2018 general election cycle. One plausible issue is that there are cross-district effects; for example, if we are in a generally high era of polarization, the partisan lean of other districts could still affect the voter turnout of a district. We are also limited by the meaningfulness of the partisan lean metric; it is an inherently imperfect formula that does not capture everything about polarization, moderation, extremism, etc. It is extremely difficult to capture such effects in a single number.

Data on polarization and voting rates across different elections (probably going as far back as the 1970s if we want to capture multiple eras of American politics) would help us address the issue of the study only covering 2018. A more robust partisan lean metric would help as well would improve on the simplicity of FiveThirtyEight's. Data on confounders not present in the census (like wealth in a district) would improve the model further.

In summary, we are not at all confident that there is a causal relationship between partisan lean and voter turnout; the data do not present a clear case for a district being more polarized causing more or less people to vote in the general election, after adjusting for many potential confounders. The coefficient in an OLS regression as well as the estimated ATE using IPW are both close to zero and include zero in their 95% confidence intervals.

---

[2] https://stats.stackexchange.com/questions/132509/confidence-interval-for-average-treatment-effect-from-propensity-score-weighting

7. Conclusion

7.1. Limitations

This work allows us to find insights on the 2018 election cycle. While the methods used could be applied to other election cycles, it may be difficult to create generalizable conclusions about voting patterns using just one election term. If attempting to draw conclusions from our data and use it in some form of real world application, it would probably require more robust data from many elections across a long time period before attempting to create generalizable voting pattern claims. Additionally, endorsements are one way of understanding national attention, however, including other proxies for attention (such as television interviews or mentions in newspapers) could provide more information. Another limitation is that there is no time-based analysis to see how partisanship changes over time; we may not see a relationship between partisan lean and voter turnout because the small sliver of time we were exploring may just be a high-partisan era with high polarization and cross-party resentment visible to the public, so people don't show up to vote in moderate districts. A final limitation is the difficulty in accounting for all confounding variables. Given that our data exists within a political and social process, there is arguably an infinite amount of confounding variables, many of which can't be quantified.

7.2. Future Work

Expanding this analysis to more election cycles would allow us to build a more robust understanding of partisanship patterns using time-based analysis. For example, if we had the same features but coming from elections ranging from 1970 to 2022, we would be able to make more generalizable claims about modern elections and partisanship. Future work could also include using social science methods to gain a better understanding of confounders. Interviews with politicians, political experts, and activists could provide insight into other important confounders that should be accounted for within our causal inference.

7.3. Results and Real World Application

We discovered valuable insight for political campaigns. Using multiple hypothesis testing on the relationship between partisanship and endorsements we saw that endorsers focus on elections in "swing" districts as opposed to heavily polarized ones. Additionally, we saw that candidates running in highly partisan elections should not anticipate organizations making endorsements. Using causal inference, we were not able to confidently find a causal relationship between partisan lean and voter turnout through either OLS and IPW methods. However, political scientists have theorized a potential link between these factors.[3] Potentially, integrating more social science methodology such as interviews and surveys and expanding our data set to include many more

[3] Fraga, B.L., Moskowitz, D.J. & Schneer, B. Partisan Alignment Increases Voter Turnout: Evidence from Redistricting.

election years could have provided better insight into the relationship between partisan lean and voter turnout.

Insights like the ones above are useful to political campaigns that must attempt to gain the most amount of votes and positive attention with the most efficient use of resources. Especially for campaigns at the federal level, it is important to understand general trends across America and within specific districts. Applying our analysis to real world campaigns could save resources and money by directing campaign organizers attention to where their efforts would be most effective. A social consideration here is the potential power that data science has to create a separation between individuals and their political representatives, turning individuals within districts into numbers rather than real people that are looking for real representation from their elected government officials. Another consideration is the limitations and reliance on data; it is important to question the data analysis and if it is able to give us truly robust information on voting. While it should be one tool in the political toolkit, it is valuable to maintain other social science methods that interact directly with the individuals that our political processes aim to help.

Based on both of our research questions, we believe that there is valuable political insights to be found through analysis of endorsements and voter turnout, however, these insights should be scaffolded with the understanding that the data aims to model an incredibly complex system where it is nearly impossible to identify and quantify all confounders. Accompanying our data analysis insights with other social science methods could contribute to a better understanding of political partisanship in order to shape our democratic institutions and political processes in a way that produces the most voter engagement and promotes an inclusive democracy.

Works Cited

"Confidence interval for average treatment effect from propensity score weighting?" *Cross Validated*, 7 January 2015, https://stats.stackexchange.com/questions/132509/confidence-interval-for-average-treatment-effect-from-propensity-score-weighting.

Cooper, Chad, and Patrick Miller. "Partisan voters treat politics and elections like a competitive sports rivalry. | USAPP." *LSE Blogs*, 24 April 2015, https://blogs.lse.ac.uk/usappblog/2015/04/24/partisan-voters-treat-politics-and-elections-like-a-competitive-sports-rivalry/.

Fraga, B.L., Moskowitz, D.J. & Schneer, B. Partisan Alignment Increases Voter Turnout: Evidence from Redistricting. *Polit Behav* 44, 1883–1910 (2022). https://doi.org/10.1007/s11109-021-09685-y

Appendix

Datasets used:

Citizen Voting-Age Population and Voting Rates for Congressional Districts: 2018
https://www.census.gov/data/tables/time-series/demo/voting-and-registration/congressional-voting-tables.html

FiveThirtyEight's Partisan Lean (2018)
https://github.com/fivethirtyeight/data/tree/master/partisan-lean/2018

Primary Candidates 2018
https://github.com/fivethirtyeight/data/tree/master/primary-candidates-2018