Candidate Quality and Demographic Trends in the 2022 United States House of Representatives elections

Kevin Ordet

December 16, 2022

Introduction

Background Information

In the wake of the 2020 presidential election, where President Joe Biden denied incumbent Donald Trump a second term, then-President Trump and a significant number of Republicans made false claims regarding the election's legitimacy. After Trump's defeat, a mob of his supporters attacked the U.S. capitol building to stop the counting of electoral votes. In the months that followed, this unprecedented attempt to subvert American democracy has continued to take form in state legislatures throughout the country, with eyes on overturning the 2024 election.

The 2022 midterms are the first major national election to follow these events. As FiveThirtyEight found, as many as six in ten Americans had an election denier on the ballot in 2022. While the conventional wisdom suggests that the president's party struggles in midterms, how would these developments, along with the Supreme Court's overturning of *Roe*, change the political calculus?

In an era where negative partisanship runs deep and split-ticket voting is at an all-time low, it may seem reasonable to conclude that candidate quality and other factors do not matter much. After all, the majority of voters will always check the box for whichever candidate has a "D" or "R" next to their name. However, in every election cycle, there are a handful of candidates who manage to defy the political gravity: a Democrat in rural Alaska, a Republican in the New York metro. This analysis aims to evaluate not only traditional metrics of candidate quality, such as incumbency and fundraising, but also the element of election denialism: how did these candidates perform at the ballot box?

This study focuses on the 2022 House of Representatives elections, which are contested in 435 single-member districts across the country. This choice was made to access the greatest amount of data, recognizing that statewide Senate and gubernatorial races each have distinct features. For purposes of this analysis, elections must only involve a single Democrat and single Republican; although, there may be third party candidates. While most states hold partisan primaries and first-past-the-post general elections, there are some exceptions. California and Alaska each utilize non-partisan blanket (or jungle) primaries, in which multiple candidates from the same party can advance to the general election. The majority of the California seats are contested by both parties (and thus were included in this analysis), while Alaska's at-large district (AK-01) had two Republicans on the general election ballot (and thus was excluded). Further, parties sometimes choose not to nominate candidates for races that are deemed uncompetitive; these districts are also omitted from the study.

Data Description and Key Variables

Dataset	Source	Description
cook_popular_vote_tracker	Cook Political Report	2022 House election results by district
urbanization-index-2022	FiveThirtyEight	District partisan lean and urbanization
fivethirtyeight_election_deniers	FiveThirtyEight	Republican stances on the 2020 election

Dataset	Source	Description
candidate_summary_2022	Federal Election Commission	Candidate fundraising
educ_by_district	POLITICO	Percent college educated population by district
state_abbr	self	State names and postal codes

This analysis uses data from four nonpartisan sources: the Cook Political Report, FiveThirtyEight, the Federal Election Commission (FEC), and Politico. Additionally, there is a table of state names and postal codes to aid in merging the data from these different sources. The relevant variables from each dataset are merged into a single dataframe. Of the 435 congressional districts in total, 399 have complete data and satisfy the condition of having a single Democrat running against a single Republican.

Name	Type	Description
district	Categorical	The five digit alphanumeric code for the congressional district (e.g. NC-04 is North Carolina's 4th district)
state	Categorical	The state's postal code (e.g. NC)
pvi_22	Numeric	FiveThirtyEight's partisan lean of the district
margin	Numeric	Democratic percentage point margin in 2022 House election
incumbent	Categorical	The incumbent candidate's party (`dem` or `gop`) or `open` to specify open seats
dem_fundraising_pct	Numeric	The Democratic percentage of total fundraising in the district
pct_college	Numeric	The percentage of the district's population with a college degree
urbanindex	Numeric	FiveThirtyEight's urbanization index
gop_stance	Categorical	`denied`, `accepted`, or `unclear` position on the legitimacy of the 2020 election

An detailed explanation of the above variables follows:

- FiveThirtyEight's partisan lean `pvi_22` is defined as "50 percent the state or district's lean relative to the nation in the most recent presidential election, 25 percent its relative lean in the second-most-recent presidential election and 25 percent a custom state-legislative lean." By convention, positive values indicate Democratic-leaning seats and negative values indicate Republican-leaning seats. (It is also common to express partisan leans such as "D+3" to indicate a `pvi_22` of 3, or "R+5" to indicate a partisan lean `pvi_22` of -5.) Seats that favor a party by 15 or more points are said to be "Safe Democratic/Republican," while seats that favor a party by 5 to 15 points are considered "Lean Democratic/Republican." Districts with a `pvi_22` of no more than 5 points in either direction are categorized as "Toss-ups". By this metric, the most competitive seat in the country is TX-15 (`pvi_22` = -0.08), a right-trending district stretching from the Rio Grande Valley to the outskirts of San Antonio.
- FiveThirtyEight's urbanization index `urbanindex` is defined as "the natural logarithm of the average number of people living within a five-mile radius of every census tract in a given district." Lower values indicate more rural districts, and higher values indicate more urban districts. The most rural district, with an `urbanindex` of 8.1, is MN-07 (eastern Minnesota). The most urban district, with an `urbanindex` of 14.9, is NY-12 (parts of Manhattan, Brooklyn, and Queens).
- The 2022 House vote `margin` is calculated by subtracting the Democrat's vote share from the Republican's vote share. If the Democrat received 48.0 percent of the vote and the Republican received 51.0 percent,

the `margin` is -3.0.

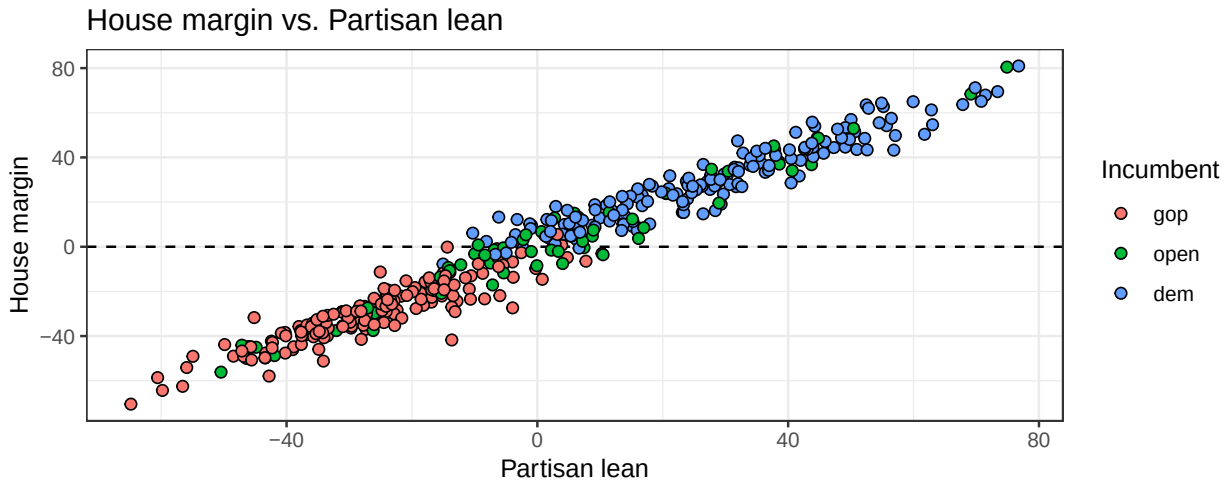
- The `fivethirtyeight_election_deniers.csv` dataset classifies Republican candidates into one of six stances on the 2020 election: fully denied, raised questions, no comment, avoided answering, accepted with reservations, or fully accepted. Candidates who "fully denied" the 2020 election "either clearly stated that the election was stolen from Trump or took legal action to overturn the results, such as voting not to certify election results or joining lawsuits that sought to overturn the election." Republicans who "fully accepted" the 2020 election agreed that President Biden won and have not raised concerns about the integrity of the election. Of the 399 districts in this analysis, 152 (38%) of the Republican candidates were "full deniers" and 47 (12%) were "full acceptors". The `gop_stance` variable was encoded as `denied` for "full deniers", `accepted` for "full acceptors", and `unclear` for all other stances.
- The Democratic fundraising share `dem_fundraising_pct` is the percentage of total individual contributions that went to the Democratic candidate. For example, if the Democrat raised $300,000 and the Republican raised $100,000, the `dem_fundraising_pct` is 75. This method is used to evaluate fundraising since FiveThirtyEight's forecasting models implement this calculation (with some adjustments), rather than the raw fundraising totals.
- The remaining variables are self-explanatory. The reader should note that "DEM" and "GOP" are commonly used abbreviations for Democrats and Republicans, respectively.

Research Aims

This study has three major goals:

- Evaluate how "district factors" such as partisanship and demographics impacted the results of the 2022 House elections. Specifically, this study observes the effects of population density and education, two measures which underpin polarization in modern American politics. In the last decade, high-educated suburbs have trended sharply towards Democrats, while low-educated rural areas have moved even further towards Republicans. While urban areas are largely Democratic strongholds, Republicans have poached some support from these areas in recent years. How would these trends hold up in 2022?
- Evaluate how "candidate factors" such as incumbency, fundraising, and election idealism mattered in the 2022 House elections. In our hyper-polarized era, the importance of incumbency may be declining, but can still make a difference on the margins – after all, this is where elections are won and lost. Incumbents tend to out-fundraise their opponents, which is often attributed to electoral success. While many forecasting models tend to avoid subjective criteria, such as a candidate's stance, this analysis also aims to evaluate the effects of Republican election denialism at the ballot box.
- Highlight the candidates and districts that managed to buck the political gravity. In doing so, we will gain a better understanding of the factors listed above and, hopefully, an appreciation for the complexity of electoral politics. While broader trends often dictate outcomes, the individual circumstances of each race still matter.

Exploratory Data Analysis



The plot above shows the relationship between FiveThirtyEight's partisan lean `pvi_22` and the actual House vote `margin`. The points are colored by the incumbent party `inc`: blue for Democratic incumbents, red for Republican incumbents, and green for open seats. Points above the dashed line represent districts won by Democrats, while points below it indicate districts won by Republicans.

Glaringly, the variation in partisan lean `pvi_22` predicts 96% of the variation in `margin`. Although partisanship dominates, there are several examples of candidates defying expectations across the map.

From here on, Democratic performance over expectation `poe` is defined as the difference between `margin` and `pvi_22`. Positive values of `poe` indicate Democratic outperformances, while negative values represent Republican ones. By comparing the actual margin to the partisanship of each district, we can evaluate how each party performed relative to expectations.

Table 3: Democratic outperformances

district	pvi_22	margin	poe
OH-09	-6.2	13.3	19.4
ME-02	-10.3	6.1	16.4
HI-01	31.9	47.4	15.5
KS-03	-3.2	12.2	15.4
CT-02	2.9	18.0	15.1

Table 4: Democratic outperformances

district	pvi_22	margin	poe
FL-26	-13.6	-41.7	-28.1
FL-28	-3.9	-27.4	-23.4
LA-05	-34.1	-51.2	-17.1
KY-06	-13.1	-29.1	-15.9
NY-02	-6.0	-21.9	-15.9

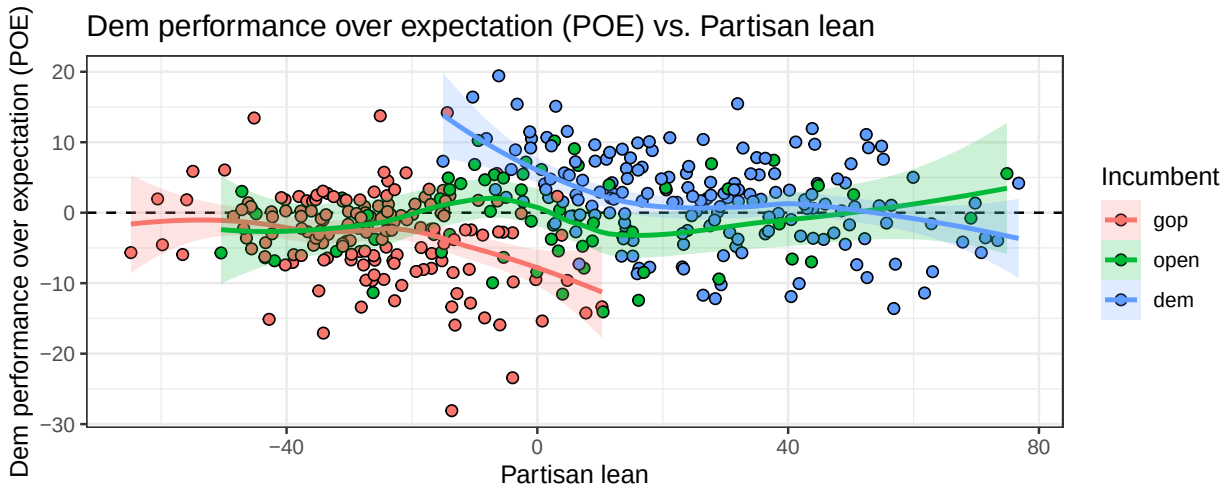
The tables above show the five largest outperformances on each side of the aisle.

Three notable Democratic victories occurred in Republican-leaning districts, as incumbents Marcy Kaptur (OH-09), Jared Golden (ME-02), and Sharice Davids (KS-03) each won re-election by at least 6 points.

Also worth mentioning is Mary Peltola's AK-01, which was not included in this analysis due to Alaska's non-partisan primary; the incumbent Democrat won Alaska's lone district by 10 points, even as Trump carried the state by double digits two years prior.

On the Republican side, Mario Diaz-Balart (FL-26) and Carlos Gimenez (FL-28) each trounced to re-election in formerly competitive south Florida seats, which continued to march rightward. Both districts backed Hillary Clinton just six years ago, and yet were won by Republicans by 42 and 27 points, respectively. In NY-02, moderate Republican Andrew Garbarino won a second term by 22 points, even as Biden nearly carried the seat in 2020.

The major theme, to be explored further, is that although partisanship reigns supreme, candidate quality and other factors still matter. The first of these factors to be analyzed is incumbency.



The plot above shows the Democratic performance over expectation`poe` of each district plotted against the partisan lean, with the points again colored by incumbent party `inc`. Democrats overperformed in districts plotted above the dashed line and underperformed in districts plotted below the dashed line.

The trendlines highlight the importance of incumbency in swing districts. In "crossover districts" that leaned towards the one party but were held by the other, incumbents tended to defy expectations by 5 to 10 points or more. This trend can be partially attributed to self-selection; only the largest outperformers are going to hold crossover districts in the first place.

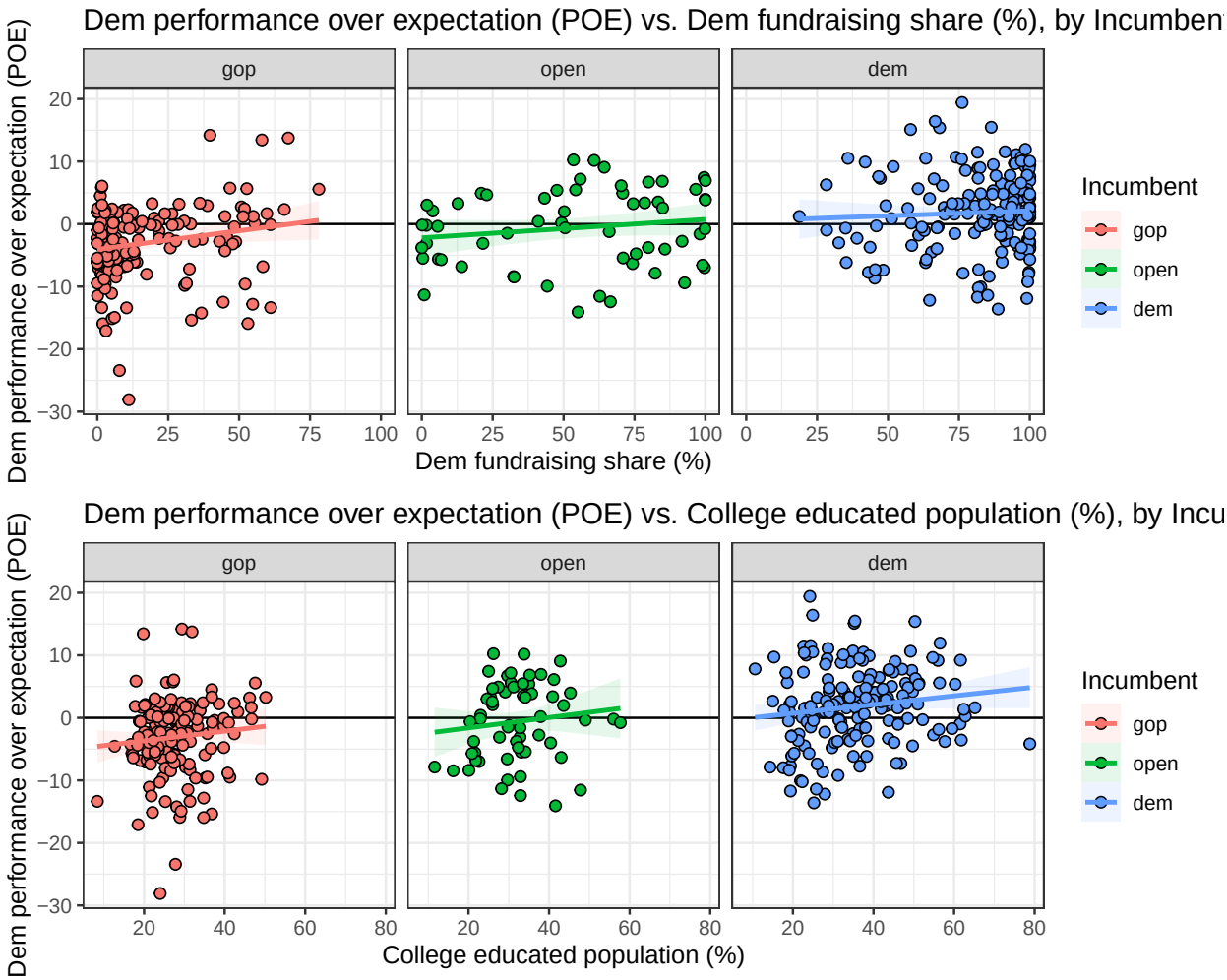
Another interesting point is the relative success of Democrats in Republican-tilting open seats. In winning the national popular vote, Republicans overperformed partisanship in the majority of open seats. However, Democrats actually tended to beat partisan lean in open seats with a lean between R+5 and R+15 (i.e. "Lean R" seats). Further analysis of the qualities of these districts and candidates will help explain this phenomenon.

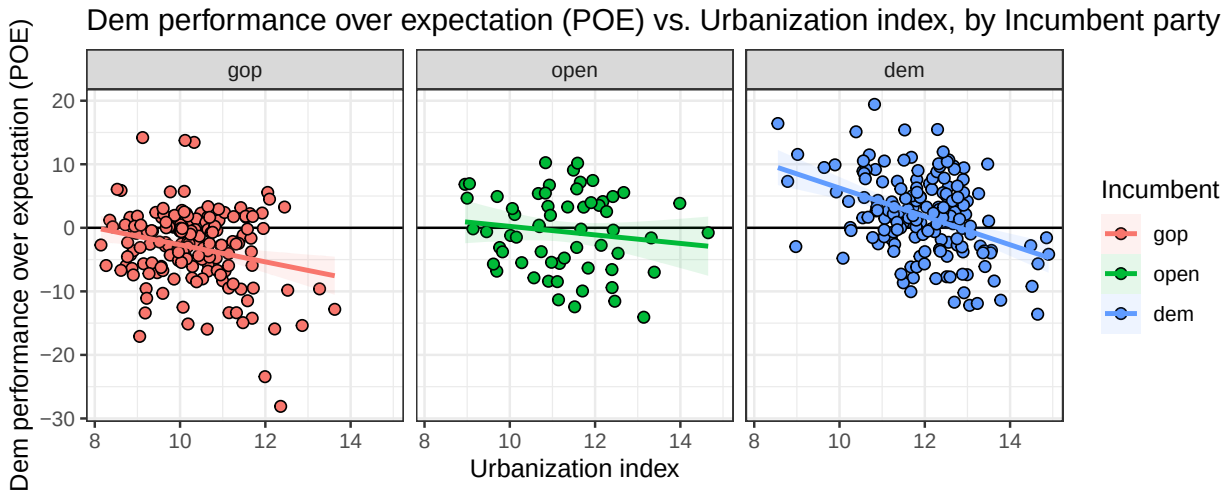
Table 5: Performance over Expectation (POE) in open Lean R seats (PVI: R+15 to R+5)

district	gop_stance	pvi_22	margin	poe
WA-03	denied	-9.4	0.8	10.2
MT-01	unclear	-10.0	-3.2	6.8
AZ-06	unclear	-6.9	-1.5	5.4
NC-11	unclear	-14.2	-9.3	4.9
MI-10	unclear	-5.4	-0.5	4.9
WI-03	denied	-8.4	-3.7	4.7
FL-13	denied	-12.2	-8.1	4.1
OH-07	unclear	-14.0	-10.7	3.3

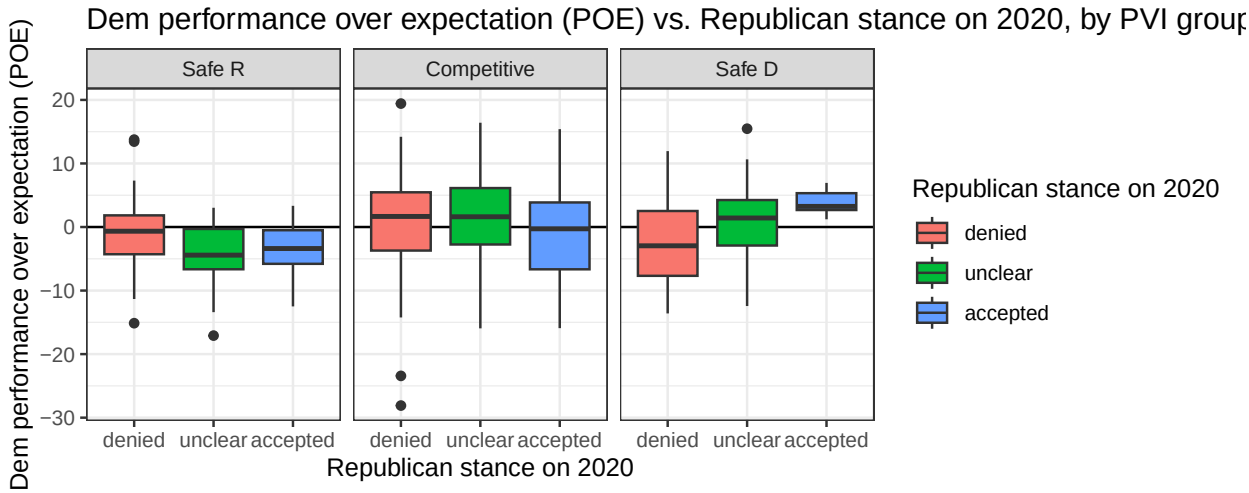
district	gop_stance	pvi_22	margin	poe
CA-03	unclear	-7.5	-7.3	0.2
FL-07	denied	-14.0	-17.1	-3.1
NY-01	unclear	-5.4	-11.7	-6.3
FL-15	accepted	-7.1	-17.1	-9.9

The table above shows the Republican-leaning (`pvi_22` between -15 and -5) open seats sorted by Democratic outperformance `poe`. Of note, Democrat Marie Gluesenkamp Perez defeated election-denying Republican Joe Kent in WA-03, which leans 9 points Republican, in what has been dubbed "the upset of the cycle." In total, Democrats outran expectations in 9 of the 12 districts and by an average of 2.1 points. It should be noted that all six Republican election deniers contesting these seats ran behind partisan lean; the only full acceptor ran almost 10 points ahead of partisanship.

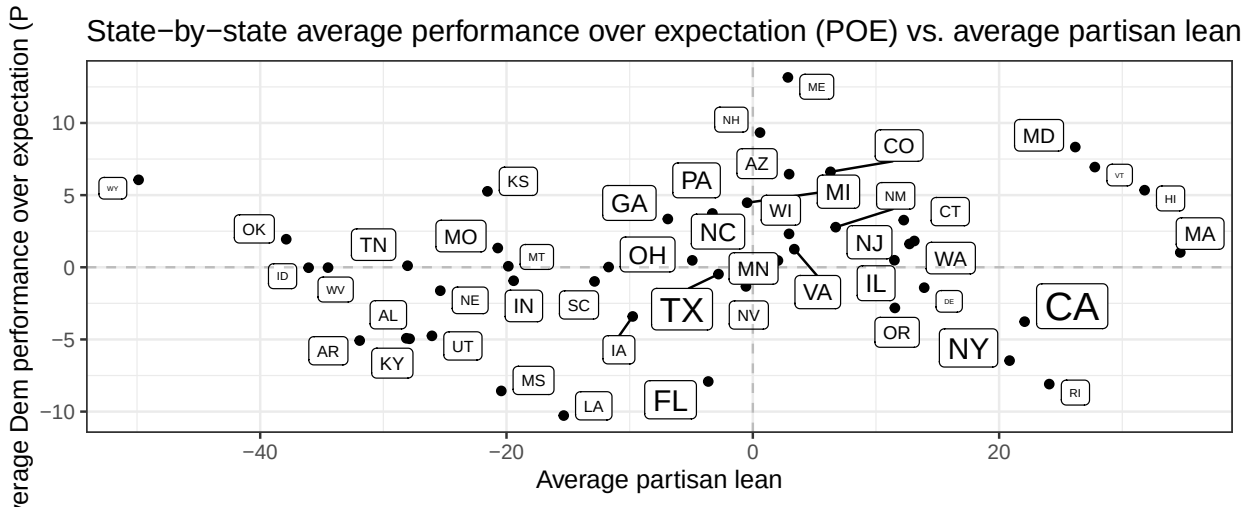




The plots above show the Democratic outperformance `poe` plotted against Democratic fundrasing share `dem_fundriasing_pct`, percent of population with a college degree `pct_college`, and FiveThirtyEight's urban index. The plots are also grouped by incumbent type `inc`. The first major observation is that Democrats tended to outperform partisan lean in rural and high college-educated districts, while Republicans did so in urban and low college-educated seats. Second, increased fundraising tended to result in overperformance of partisanship,and incumbents dominated fundraising.



The boxplot above shows the Democratic outperformance `poe` by the `gop_stance` on the 2020 election, grouped by the partisan lean of the district. "Competitive" districts include all districts with a partisan lean `pvi_22` less than 15 points in either direction (that is, "Lean R", "Lean D", and "Tossup" seats). Note that the bars are colored by `gop_stance` rather than `inc`. While election denying Republicans actually ran ahead of expectation in Safe Democratic seats, they underperformed acceptors in both Safe Republican an Competitive districts, and they underperformed candidates with an unclear position in Safe Republican races. There are several possible explanations for this occurrence; strong Republican voters may have been more likely to turn out for election deniers in uncompetitive races they otherwise would not have. Regardless, the critical takeaway is the relative success of election-accepting Republicans in competitive districts, which ultimately decide the outcome of the election.



The plot above shows the average Democratic overperformance `poe` by state plotted againt the state's average partisan lean `pvi_22`. The labels are sized by the number of districts in the state.

There is little correlation between partisan lean of the state and average Democratic outperformance ($r^2 = 0.04$). However, there were some extreme cases of defying expectations throughout a state:

- In Florida, Republicans outran partisan lean in 23 of 25 analyzed districts by an average of 7.9 points.
- In Colorado, Democrats outran partisan lean in all 8 districts by an average of 6.6 points.
- In New York, Republicans outran partisan lean in 23 of 24 analyzed districts by an average of 6.5 points.
- In Michigan, Democrats outran partisan lean in 12 of 13 districts by an average of 4.5 points.

With the exception of New York, these strong localized environments have largely been attributed to popular governors running for re-election (Ron DeSantis in FL, Jared Polis in CO, and Gretchen Whitmer in MI). In Michigan, an abortion rights referendum also has been credited with boosting Democratic enthusiasm. On the other hand, Democratic turnout was depressed in New York, where Democratic voters were lukewarm about scandal-ridden Governor Andrew Cuomo's appointed replacement, Kathy Hochul.

Methodology

The relevant variables, which are listed as follows, were merged into a single dataframe by `district`:

- The response variable `margin` was calculated by subtracting the `dem_pct` from the `gop_pct` in `cook_popular_vote_tracker.csv`. The dataset's `incumbent` variable is also taken, which uses a `1` for Democratic incumbents, `-1` for Republican incumbents, `0` for open seats. In the variable `inc`, these values are translated to `dem`, `gop`, and `open`, respectively.
- The FiveThirtyEight partisan lean `pvi_22` and urbanization index `urbanindex` were taken `urbanization-index-2022.csv`.
- The Republican candidate's stance on the 2022 election was taken from `fivethirtyeight_election_deniers.csv`. The `gop_stance` variable was encoded as `denied` for "full deniers", `accepted` for "full acceptors", and `unclear` for all other stances.
- The percent college educated population `pct_college` was taken from `educ.by.district.csv`.
- Using `candidate_summary_2022.csv`, the highest fundraising candidate for each party was determined for each district to obtain `raised_dem` and `raised_gop` (the premise being that the lower fundraising candidates were eliminated in the primary). From there, the `raised_dem` was divided by the sum of `raised_dem` and `raised_gop` and multiplied by 100. The result of this calculation is the `dem_fundraising_pct`

- The district's `state` was simply pulled from the first two digits of the alphanumeric code for each `district`.

Before proceeding, three of the variables were centered to allow for better interpretation of results:

- `centered_dem_fundraising_pct` is `dem_fundraising_pct` minus 50.
- `centered_urban_index` and `centered_college_pct` are the mean-centered `urbanindex` and `pct_college`, respectively.

The final dataset consists of 399 observations, representing the districts that satisfied the criteria for inclusion (contested seats with one Democrat and one Republican, having complete fundraising data). To predict the `margin`, a linear model was fit using the predictors `pvi_22`, `inc`, `centered_dem_fundraising_pct`, `centered_urban_index`, `centered_college_pct`, and `denier`. To account for dependence within states, a random intercept was included for `state`. The full equation for the model is as follows:

Results

Table 6: Model fixed effects

	Estimate	Standard error	t-value
Intercept	0.061	0.814	0.076
FiveThirtyEight Partisan Lean (PVI)	0.918	0.017	53.049
Democratic incumbent	1.999	0.736	2.716
Republican incumbent	-2.802	0.773	-3.623
Democratic percentage of fundraising (centered)	0.100	0.013	7.940
FiveThirtyEight Urbanization Index (centered)	-1.043	0.297	-3.509
College graduate percentage (centered)	0.109	0.024	4.534
Republican "fully denies" legitimacy of 2020 election	0.990	0.525	1.885
Republican "fully accepts" legitimacy of 2020 election	-0.770	0.715	-1.077

For an open seat with even partisanship, equal fundraising, the national average for population density and educational attainment, and a Republican candidate who neither fully denied nor accepted the legitimacy of the 2020 election, the expected outcome was a Democratic win by 0.06 percentage points.

Holding all other variables constant and including random intercepts by state:

- For every one percentage point increase in the partisan lean (i.e. more Democratic), a district was expected to vote 0.918 points more net Democratic.
- Districts with elected Republican incumbents were expected to vote 2.802 points more Republican than open seats. Similarly, districts with elected Democratic incumbents were expected to vote 1.999 points more net Democratic than open seats.
- For every one percentage point increase in the Democratic share of combined fundraising, a district was expected to vote 0.100 points more net Democratic.
- For every one point increase in the urbanization index, a district was expected to vote 1.043 points more net Republican. (For reference, the most urban districts have urbanization indices just under 15, while the most rural districts have values slightly above 8.)
- For every one percentage point increase in the proportion of population with college degrees, a district was expected to vote 0.109 points more net Democratic. (For reference, districts ranged from 8.4% to 78.8% college educated, and the middle 80% of districts fall between 20.3% and 46.9% college educated.)
- Districts with election-denying Republicans on the ballot were expected to vote 0.990 points more net Democratic than districts with Republicans who neither fully denied nor accepted the legitimacy of the 2020 election.

- Districts with election-accepting Republicans on the ballot were expected to vote 0.770 points more net Republican than districts with Republicans who neither fully denied nor accepted the legitimacy of the 2020 election.

Discussion

The model's outcome is consistent with the overall picture of the 2022 House elections. Excluding uncontested seats, the average swing from the 2020 presidential election was 5.09 points towards Republicans, implying a national environment that leans Republican by 0.64 points (since Democrats won the presidency by 4.45 points nationwide). This figure is similar to the near neutral implied by the model's intercept.

While the incumbency advantage is ever-declining, it still yields a significant advantage, even before accounting for fundraising. On average, Democratic incumbents garnered 82% of a district's total fundraising, while Republican incumbents hauled in 84% of the total. These figures would suggest that the average incumbent actually outperformed partisan lean by 5-6 points, rather than the 2-3 points assuming neutral fundraising.

Exit polls suggested weak Democratic turnout, but an unusually strong performance with independent voters in an in-party midterm. Per CNN's polling, registered Republicans had a 36% to 33% turnout advantage over registered Democrats, but independents actually broke *for Democrats* by a 2-point margin. The model's output confirms a collapse in Democratic base turnout in urban areas (especially in Florida and New York), but highlights their success in persuading rural, suburban, and college-educated voters.

Election denialism has negative electoral consequences, with the model suggesting that Republicans who "fully denied" the results of the 2020 election lost about a point on net margin. This may seem insignificant, but consider that Democrats only needed to win five more seats to retain House control and lost five seats by fewer than 0.7 percentage points. In the closest race of the cycle, Democrat Adam Frisch came within 600 votes (0.16%) of flipping the 14-point Republican leaning CO-03 held by election-denying Lauren Boebert.

Conversely, Republicans who rejected extremism found success on election day. In the only two Republican victories in districts leaning at least 10 points towards Democrats, both Republican candidates were "fully accepting" of the 2020 election. In arguably Republicans' best performance of the cycle, Anthony D'Esposito won his first term in NY-04, which had backed Biden by almost 15 points. In CA-22, incumbent David Valadao won re-election in a seat that voted for Biden by over 13 points. In doing so, Valadao was one of only two Republican Trump-impeachers to secure another term (the other being Dan Newhouse in WA-04, who ran 12.5 points ahead of base partisanship).

As a further point, the aforementioned model coefficients may understate the impact of extremism on the Republican Party as a whole. The historically weak performance by Republicans in an out-party midterm suggests that presence of extremist candidates thwarted even more moderate members of the party. It seems plausible to suggest that a non-election denying Republican Party may have yielded a more favorable national environment (which would only be reflected in the model's intercept).

Appendix

Table 7: Demographic and fundraising summary of PVI groups

pvi_group	districts	pct_college	urban_index	dem_fundraising_pct
Safe D	130	35.7	12.3	87.3
Lean D	45	35.7	11.6	71.1
Toss-up	40	33.8	11.2	61.3
Lean R	40	30.6	10.8	36.4
Safe R	144	27.1	10.1	12.2

References