# Data 102 Final Project: Primary Election Data

## Data Overview

We chose to use data of 2018 primary election candidates, which includes a dataframe for Democrats and one for Republicans. This data was created by fivethirtyeight using ballot data and doing general research about candidates to analyze and report on, which was done in articles such as "How's The Progressive Wing Doing In Democratic Primaries So Far?"[1]. The data only includes candidates from a certain timeframe for only a few kinds of elections and doesn't include candidates in races that include a democratic incumbent. As a result it is only a sample of our population.

Our democrats data is about 70% white (non-hispanic) and 30% nonwhite. The true US population is about 60% white (non-hispanic) and 40% nonwhite[2]. There is a pretty large difference here, although we are unsure whether it could be due to white people more commonly running for office. This outcome does imply some bias in our results, as we now know that the sample is likely not as representative of the overall population as we might have hoped.

Participants were not aware of their use in this dataset, however, because they are public figures running to represent their country, they are in the public eye and aware that people are recording information about them. The datasets are relatively fine because each row represents a different candidate running in the 2018 primary elections. This will impact the interpretation of our findings, and make it based on a candidate level. We can look at specific attributes on a candidate level and draw conclusions..

We are not overly concerned about selection bias as the dataset contains everyone in the group they were trying to capture (primary candidates in non-incumbent races). We are not overly concerned about measurement error as the numbers for elections are well recorded and regulated, as well as the data on endorsements. There are concerns of convenience sampling in regards to the collection of the race column. The fact that they limited their search to only 2 minutes means that they were doing this out of convenience and there may be concerns of it being easier to identify some groups than others online.

Our dataset was not modified for differential privacy. This was a public election, so all data is publicly available. We wish that we had demographic data (Race, LGBTQ, veteran) for the Republican dataset, as it would allow us to attempt to answer the same questions about these columns that we do with the democrat dataset and would allow for interesting comparisons between the two groups.

There are multiple columns with null values in our dataset. For the race column, values are missing if they were not able to identify their race within 2 minutes of searching online news reports. For the veteran, LGBTQ, and STEM columns, values are missing if the candidate did not have a website. For the endorsements, the fields are left blank if the group did not weigh in on that race. Because it is the main feature in our analysis, we chose to drop the rows where there was no value for 'won primary'. Only five rows were dropped. For veterans, LGBTQ, and race, we chose to impute values by choosing either yes or no randomly from a Bernoulli distribution where $p = proportion\ of\ yes/1\ in\ given\ column$ since it is functionally similar to imputing the mean. For endorsements, we fill missing values with no, since we can reasonably assume the candidate did not get endorsed.

We applied multiple preprocessing techniques to prepare our data for analysis. For the hypothesis tests we removed candidates that ran unopposed by removing rows where the race only had one candidate. We chose to do this because if a candidate is running unopposed, they will win no matter what so their attributes may be less representative of what makes a candidate more favorable. In addition, we replaced 'Yes' and 'No' as well as 'Nonwhite' and 'White' with 1 and 0 in order for everything to be in the same format and to allow us to analyze the data quantitatively. We also indicated whether or not a candidate was endorsed by taking the sum across endorsements and checking if they sum to at least 1. Adding this column allows us to analyze whether having any sort of endorsement is associated with primary election win rate.

# Research Questions

**Research Question 1**

When looking at the data, we immediately wondered about the different factors that influenced the outcome of the primary election. This led to our first research question of "Do various factors (veteran status, race, LGBTQ, elected official) have an association on primary election win rates for Democrats and Republicans individually?" We decided that Multiple Hypothesis testing (MHT) was the best way to answer this question, since we want to explore the relationship between several independent variables (LGBT, Race, Veteran?) and a dependent variable (Primary Won). Using MHT allows for the evaluation of these variables simultaneously. We can also control for FWER and FDR using Bonferroni correction and Benjamini-Hochberg. Our six Hypotheses are:

Hypothesis 1:

Alternative: Veteran status is associated with primary election win rates for Democrats

Null: Veteran status has no association with primary election win rates for Democrats

Hypothesis 2:

Alternative: Race has an association with election primary win rates for Democrats

Null: Race has no association with election primary win rates for Democrats

Hypothesis 3:

Alternative: Being LGBTQ is associated with primary election win rates for Democrats

Null: Being LGBTQ has no association with primary election win rates for Democrats

Hypothesis 4:

Alternative: Having any kind of endorsement is associated with primary election win rates for Democrats

Null: Having any kind of endorsement has no association with primary election win rates for Democrats

Hypothesis 5:

Alternative: Having any sort of endorsement is associated with primary election win rate for Republicans

Null: Having any sort of endorsement has no association with primary election win rate for Republicans

Hypothesis 6:

Alternative: Being a Democrat is associated with different win rates among all endorsed candidates

Null: Being a Democrat is not associated with different win rates among endorsed candidates

By answering this question, candidates can be better informed of what factors affect their chance of winning. For example, whether or not it is important to seek out an endorsement. It can also help candidates look at their opponents within their specific party and gauge their chances against them.

In our use of multiple hypothesis testing using permutation testing, there are some limitations. This test requires that the data points are independent of one another, but this may not be the case. If one candidate wins, then it means that the other candidates in the same race do not win, thus they are not dependent. This limits the effectiveness of hypothesis testing. It may not go well if the points are very dependent on each other, and this can make the results inaccurate.

**Research Question 2**

Our second research question posed looks specifically at endorsements: "How do different types of endorsements affect the likelihood of winning the primary election for Democrats?" To answer this question we decided Bayesian Hierarchical Modeling (BHM) was suitable since it allows for the incorporation of prior knowledge, even when there is limited prior information or uncertainty about the relationships being studied (i.e an endorsements affect on the likelihood of winning the primary election). In this case, we didn't want to assume strong prior beliefs or biases since there weren't many endorsed candidates in the dataset so this Bayesian method allows us to reflect that by using a uniform prior. Additionally, BHM provides a natural framework that handles uncertainty pretty well. Political elections are inherently uncertain, and Bayesian models can quantify and express uncertainty in the estimates.

Answering this question could have real-world implications for political candidates in terms of strategic campaign planning. Political candidates could use the research findings by securing and prioritizing endorsements that have been shown to be more influential in winning primary elections. This could involve dedicating time, effort, and allocating financial resources to encourage relationships with endorsers who are likely to sway votes.

In our use of BHM, there are some limitations. Including a sparse amount of endorsement data (i.e. only 5 democratic candidates were endorsed by Warren). To account for this, we couldn't establish strong prior beliefs in our BHM. As a result, the limited data for endorsed candidates may have resulted in less informative posterior distributions and less nuanced estimates in our model, making it challenging to draw robust conclusions. For instance, if a highly influential endorsement is not adequately captured in the prior, it may lead to less nuanced and imprecise estimates in the model.

# EDA

The key features of our dataset are Won Primary, Race, Veteran?, LGBTQ, and all the endorsements. We wanted to find out which endorsements helped candidates with the primary election, so our first step was to see how many candidates were endorsed by each individual endorser. We learned that it was uncommon to be endorsed by certain endorsers. For example, out of about 800 democratic candidates, only nine were endorsed by Bernie Sanders. Also, there were a lot of NaN values for endorsements and we replaced these with no endorsement.

We then looked at the candidates endorsed by multiple endorsers and whether or not that has an impact on the "Won Primary" outcome (Figure 1). We observed that with zero endorsements, the number of those who lost the primary is far lower than the number of those who won. When candidates receive one endorsement, this difference almost completely levels out. At three or more endorsements, the number of candidates who wins the primary is actually greater than the number of those who lost. Although the values at this point are pretty small, so we don't have a lot of data to work off of, the difference observed at these higher numbers of endorsements imply that a greater number of endorsements means a lot when it comes to a candidate's chances of winning their race. This has massive implications for our second research question, which is about whether or not different types of endorsements have an effect on the likelihood of winning the election. As aforementioned, from the data in figure one, we can observe that it is likely that an endorsement has a pretty large effect on chances of winning, so this plot does a lot to motivate our question about searching to see how specific endorsements affect the chances of winning.

Our EDA continued with looking into the distribution of partisan lean split on race, because we were curious whether or not being white or nonwhite would make a candidate more likely to win. This could suggest a possible alternative answer to what our hypothesis tests would suggest in that there may be a confounding factor if people would run when they are more likely to win. If being Non-White is more likely to make a candidate win in a certain partisan leaning state, we may have a confounder on our hands.

We observe from Figure 2 that there may be some discrepancies in the tails in that there is a spike in “nonwhite” in the larger partisan lean areas, implying that nonwhite people may be more likely to run in more democrat leaning areas. On the other hand, in the left tail, there appears to be a higher concentration of whites in the lower partisan leanings, implying that whites may be more likely to run in more conservative areas. This could also be due to the demographics of counties, possibly suggesting that there may be more whites in conservative leaning areas and nonwhites in democrat leaning areas. This visualization is relevant to our research question regarding if race has an effect on a democrat winning a primary in that it suggests a possible answer to the question. This graph suggests that there may be no effect of race in whether a democrat wins a primary **in general**, but that there may be differences based on the partisan lean of the area they run in.

In Figure 3, we decided it would be beneficial to look at the distribution of primary vote percentage (primary percentage), split on race. From the graph, we observe that the distribution of primary percentage for white democrats is slightly more skewed left than it is for non-white democrats. This suggests that white democrats may be more likely to receive a higher percentage of primary votes, which could also suggest that white democrats are more likely to win a primary. The nonwhite candidates have a higher peak density at a lower primary vote percentage, suggesting that nonwhite candidates receive a smaller percentage of votes on average.

In Figure 4, we can see that the proportion of white democratic candidates who won the election (.322) is slightly higher than the proportion of nonwhite candidates who won their election (.28). The difference, however, is small enough to consider negligible upon our first observation. As such, we can see that both the proportion of nonwhites and whites who win their primary election are about equivalent. This suggests that this is the outcome we will find from our hypothesis test checking whether race has an effect on a democrat's chances of winning.

We then were curious about democratic veteran status and whether or not the candidates won the primary, so we made another countplot (Figure 5). This bar chart shows the candidates based on their vet status. While there are less veteran candidates, some of them have had success in their primary elections.We see that in both vet and non-vet the proportions seem similar, this may suggest there is not a difference between being a vet and non-vet and winning the primary. The count of non-vet who won is substantial, but smaller compared to non-vets who lost. The ratios of the winers to total population of vet and non-vet are pretty much identical and it would be interesting to see if there is change with hypothesis testing.

## Multiple Hypothesis Testing

### Methods

The specifics of our six hypotheses are mentioned above in the ‘Research Questions’ section, however they all revolve around one central idea: Do various factors have an association on primary election win rates for democrats and republicans individually. It makes sense to use multiple hypotheses to answer our question because we want to observe the associations of multiple different facts of our dataset. We tested the power of our fifth hypothesis, whose alternative is that being endorsed is associated with primary election win rates for Republicans. The test showed a power of about 0.955, which indicates that it will pretty reliably reject the null hypothesis with the alternative being true.

We will test all of our hypotheses using permutation tests, with the test statistic being the difference of means between two groups (ex. vet vs non-vet). We did not base any of our hypotheses on a certain direction (positive or negative association) since we were more interested in seeing if there was any kind of deviation from the overall groups, so we will be using the absolute value of the differences when making comparisons.

The two ways for which we will be correcting for multiple hypothesis tests are Bonferroni correction, which controls family wise error rate, and the Benjamini-Hochberg method, which controls the false discovery rate. Both our FWER controlling method (Bonferroni) and our FDR controlling method (Benjamini-Hochberg) made one discovery. It is important to note that the BH method requires hypotheses to be independent. Because we are analyzing the effects of wildly different personal attributes, we made the assumption that our hypothesis tests are independent, but we have included more discussion on this limitation below. For our specific purposes in hypothesis testing, we would prefer to use Benjamini-Hochberg since it is less strict and has higher power in detecting true positives. At this stage, we are more interested in identifying potential patterns as opposed to making solid predictions about our candidates and their chances at winning.

**Results**

Our hypothesis tests resulted in the following p-values: 0.6831, 0.2846, 0.0548, 0.0, 0.0199, and 0.7557. A lower p-value indicates that there is some kind of association between the tested feature and primary election win rates. The smaller the value, the more we would expect there to be an association. For example, hypothesis one, the larger value indicates that there is little association between veteran status and primary election win rates for democrats. The small p-value for hypothesis 4 indicates that there is a great amount of association between having been endorsed and primary election win rates for democrats.

Bonferroni helps us control failure-wise error rate (FWER), by setting our threshold to be $\alpha/n$, with our desired FWER threshold being $\alpha$. FWER is the chance of a false positive. Benjamini-Hochberg helps us control the false discovery rate (FDR) in a slightly more intricate process, where $\alpha$ is our desired FDR threshold. FDR is essentially FP / (FP + TP). Bonferroni is generally more strict.

**Discussion**

After applying Bonferroni correction, we received the output $[False,\ False,\ False,\ True,\ False,\ False]$, indicating that only the p-value for the fourth test remained significant. After applying Benjamini-Hochberg we received the same output. If we look at each test by itself using naive thresholding we would reject the null for hypothesis 4 and 5, indicating that there is some kind of association between any kind of endorsement and primary election win rates for democrats, as well as some kind of association between endorsement and primary election win rate for republicans. From this result, we might decide that endorsements are good for all parties, with the p-values from their analyses being under the threshold for both Democrats and Republicans. When working with multiple p-values, however, it is important to use correction techniques. As aforementioned, when applying the correction techniques, only the p-value for hypothesis 4 remained significant. This might make us uneasy towards making a blanket statement about endorsements for both parties and only gives us increased insight about associations with win rate for Democrats.

One major limitation of our analysis is that, although we know our data points are not independent from each other when they should have been, there was no way to correct for this. It may have been possible with a larger dataset, but the dataset was too small for us to sample from and still get consistent results. Furthermore, although we assert independence between our hypothesis test for the purpose of running Benjamini-Hochberg, we cannot entirely guarantee that factors like being LGBTQ and veteran status have no interaction with each other.

Given more data, we'd like to conduct all the same kinds of tests we did for democrats on Republicans. Oddly enough, the Republican dataset does not contain as many columns that detail additional information about candidates, like race, whether or not they are LGBTQ, or whether or not they are a veteran. It would be interesting to see if the republicans differed from democrats in any of the tests involving this information. It would also be interesting to run tests involving things such as the candidate's highest level of education and more information about the people voting in each of the elections.

## Bayesian Hierarchical Modeling

### Methods

We decided to use a Bayesian Hierarchical model to infer the chances of winning the election given each endorsement. Figure 6 depicts the model we used. The alpha and beta are the parameters for our priors on endorsements, which were originally intended to be in a Beta but we decided against it, so these can now be interpreted as the mean and SD of a uniform. Our Thetas are the chances of winning as a democrat with a certain endorsement, N is the total number of people who ran for office with that endorsement, and X is the total number of democrats who won with that specific endorsement. Our specific mixture model is not being used to identify groups in the dataset, instead we are using it to infer the chances of winning the primary election given various endorsements.

When we started the process of choosing a prior for the distribution of primary election win rates, we wanted to use the most commonly given endorsements to find a specific beta distribution that aligned with our observed mean and variance values. Unfortunately, our mean and variance values led us to values for alpha and beta that were essentially zero, meaning our prior using this method would have been beta(0,0). We felt that this did not do a good job of encoding any of the information we wanted to and as a result, decided it was not a good option. This led us to choosing a uniform(0, 1) distribution instead for the prior of primary election win rates. The benefit of this is also that we would reasonably expect the most given endorsements to have different effects than the lesser given ones, meaning our initial idea for a prior was not very representative of the whole set from the beginning.

### Results

After building out our PyMC model, our results (figure 7) show that endorsements from Joe Biden and Elizabeth Warren were associated with a far larger increase in election win rates than the other kinds of endorsements observed. It is also interesting to note that the empirical mean of the Biden and Warren endorsements was 1, meaning everyone who was endorsed by them won their election. Our posterior mean was not at 1, which is good considering that we could not reasonably expect these endorsements to cause a win 100% of the time. Our Revolution and Justice Democrats endorsements were weakly associated with a higher win rate than the overall mean, while Emily's list, Sanders, PCCC, Indivisible, and WFP endorsements were all a fair bit better than the mean of the whole dataset. In general, it appears that all endorsements did better than the mean election win rate of the whole dataset, although some were very small differences.. The most surprising finding was how unexpectedly powerful Biden and Warren’s endorsements were at the time, given that the positions they held at that time were not much more prominent that one like Sanders’.

To make it easy to read and understand the uncertainty in our estimates, we decided to provide them in this manner:

- For candidates endorsed by Emily's List, the posterior mean is 0.715 with a standard deviation of 0.059. This implies that most values are between 0.656 and 0.774, but not outside of the bounds of (0,1)
- For candidates endorsed by Joe Biden, the posterior mean is 0.917 with a standard deviation of 0.077. This implies that most values are between 0.84 and 0.994, but not outside of the bounds of (0,1)
- For candidates endorsed by Elizabeth Warren, the posterior mean is 0.855 with a standard deviation of 0.121. This implies that most values are between 0.734 and 0.976, but not outside of the bounds of (0,1)
- For candidates endorsed by Bernie Sanders, the posterior mean is 0.545 with a standard deviation of 0.14. This implies that most values are between 0.405 and 0.685, but not outside of the bounds of (0,1)
- For candidates endorsed by Our Revolution, the posterior mean is 0.323 with a standard deviation of 0.05. This implies that most values are between 0.273 and 0.373, but not outside of the bounds of (0,1)
- For candidates endorsed by Justice Dems, the posterior mean is 0.325 with a standard deviation of 0.066. This implies that most values are between 0.259 and 0.391, but not outside of the bounds of (0,1)
- For candidates endorsed by PCCC, the posterior mean is 0.643 with a standard deviation of 0.11. This implies that most values are between 0.533 and 0.753, but not outside of the bounds of (0,1)
- For candidates endorsed by Indivisible, the posterior mean is 0.647 with a standard deviation of 0.067. This implies that most values are between 0.58 and 0.714, but not outside of the bounds of (0,1)
- For candidates endorsed by WFP, the posterior mean is 0.498 with a standard deviation of 0.086. This implies that most values are between 0.412 and 0.584, but not outside of the bounds of (0,1)
- For candidates endorsed by VoteVets, the posterior mean is 0.568 with a standard deviation of 0.085. This implies that most values are between 0.483 and 0.653, but not outside of the bounds of (0,1)

**Discussion**

The main limitation of our model is our weak uniform prior. We wanted to use a Beta Prior to represent the different endorsements starting distribution, however when calculating the parameters, they came out to almost zero, which is not okay for the Beta distribution. So we had to use an uninformative prior instead. If we had more data in general, we could calculate a more informative prior and have less uncertainty in general, so it would be nice to get a couple thousand more data points.

# Conclusion

Starting with our hypothesis tests, we found that when we use naive thresholding at p=0.05, Hypothesis 4 and 5 (Having any kind of endorsement is associated with primary election win rates for Democrats and Having any sort of endorsement is associated with primary election win rate for Republicans) are statistically significant. However, when we use Bonferroni correction to control for Family Wise Error Rate or Benjamini-Hochberg to control for False Discovery Rate, the only discovery that we make is for hypothesis 4, meaning that we reject the null hypothesis that having any kind of endorsement is not associated with election win rates for democrats.

Continuing with our BHM, Because we used a uniform(0,1), the posterior means were pushed slightly toward 50%. This is a very weak prior and did not have a strong effect, however we see that the bayesian modeling is effective for endorsements that are very strong and do not have that many data points. For example, empirically, when candidates receive a Biden or Warren endorsement, they win 100% of the time. However, we know that this does not 100% guarantee a win, so the posterior means are lower and show that they are still strong endorsements, but nothing is 100%. In general, we find that the posterior means are higher than the empirical means for all candidates, endorsed and not endorsed. This leads us to believe that candidates that have endorsements are more likely to win than those who don't. More specifically, Biden and Warren endorsements are the strongest, and Our Revolution and Justice Dems being the weakest.

While we believe that our methods are technically correct and that the results are believable, due to the limitations on the size of the dataset as well as the limitations in applying the models (dependance), we hesitate to say that the results are very generalizable and that they could be used for any meaningful purpose other than analysis. Our findings are broad, as we are looking at broad demographic statistics over primary elections in general, as opposed to looking at mch more specific things or a very specific kind of election in a specific state.

Our main finding in our analysis was the strength of endorsements in general and more specifically the strength of certain endorsements such as Biden and Warren. Based on these results we would suggest that primary candidates prioritize acquiring endorsements to maximize their chance of winning the election, and specifically that of well known politicians. At the time of this election, Biden and Warren were very influential and their endorsement meant a lot.

We did not merge different data sources and felt that we had sufficient information in the provided dataset.

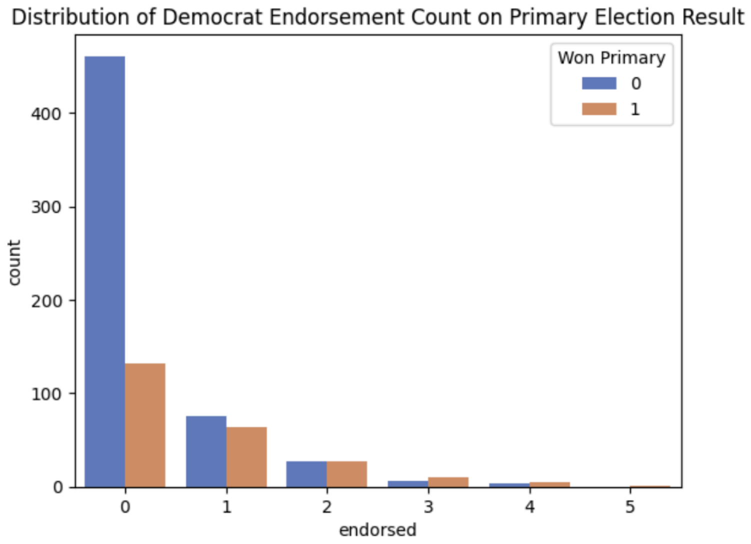
One major limitation of this dataset is its size. There are not that many elections so gathering large amounts of data on them can prove to be a difficult task. Having limited data meant that some columns such as some of the endorsements had very few data points. In addition, not having data on incumbent races limits our analysis. While incumbents often are more likely to win elections, this still could have provided more data and more interesting analysis.

In the future, we think that looking at how different demographics have different effects in different areas of the country would be interesting. For example, if being a veteran helps your chances more in one state than another. In addition, looking at if endorsements have different effects on different demographics would also be interesting to look into.

During this project we learned how difficult it is to meet the requirements to apply models in the real world. Requirements such as independence of data limit the effectiveness of our analysis when they are not met. In addition, throughout all of our time at Berkeley we have been provided datasets that are easy to work with, as well as we are given specific tasks and models to use. However, in the real world you are not just told exactly what to do or if you meet the requirements or limitations, and these prove to be very difficult to figure out.

## Figures

**Figure 1**



Bar Graph of the number candidates who either won or lost the primary vote split upon how many endorsements the candidates have

**Figure 2**

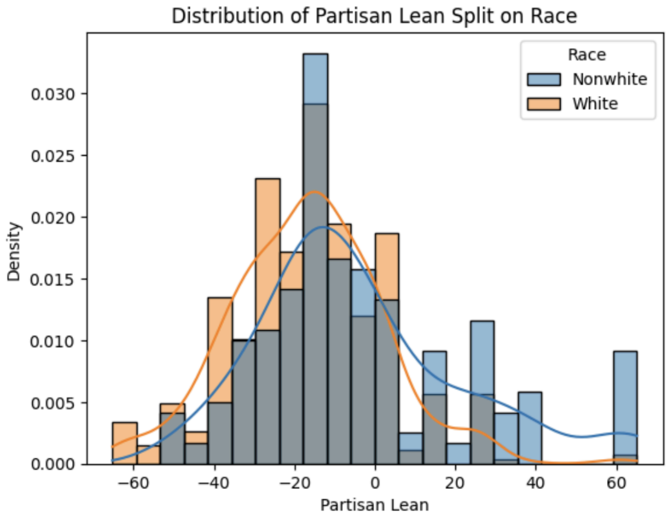


**Figure 3**

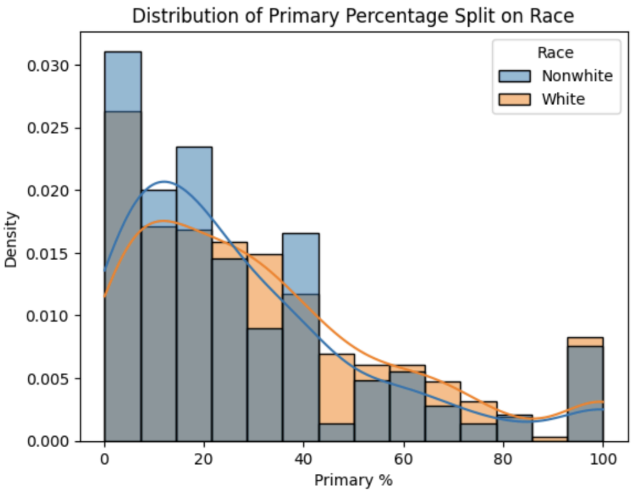


Figure 4

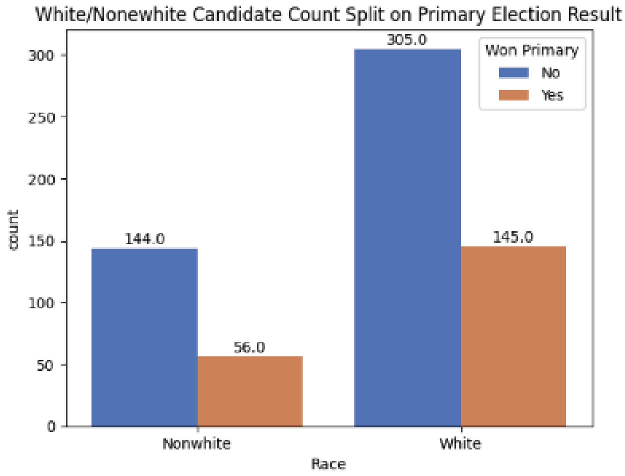


Figure 5

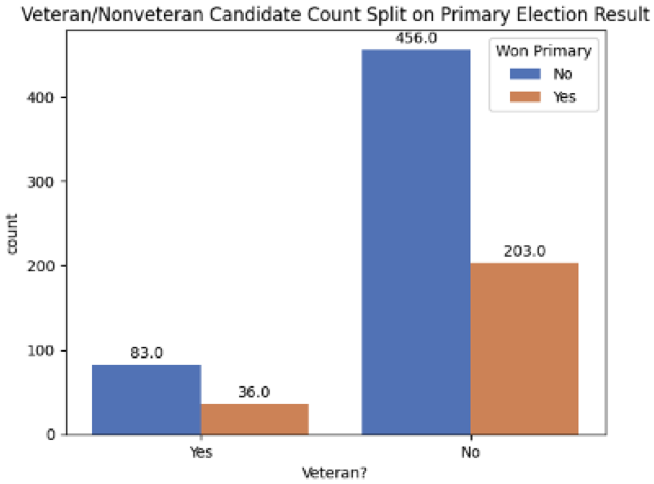
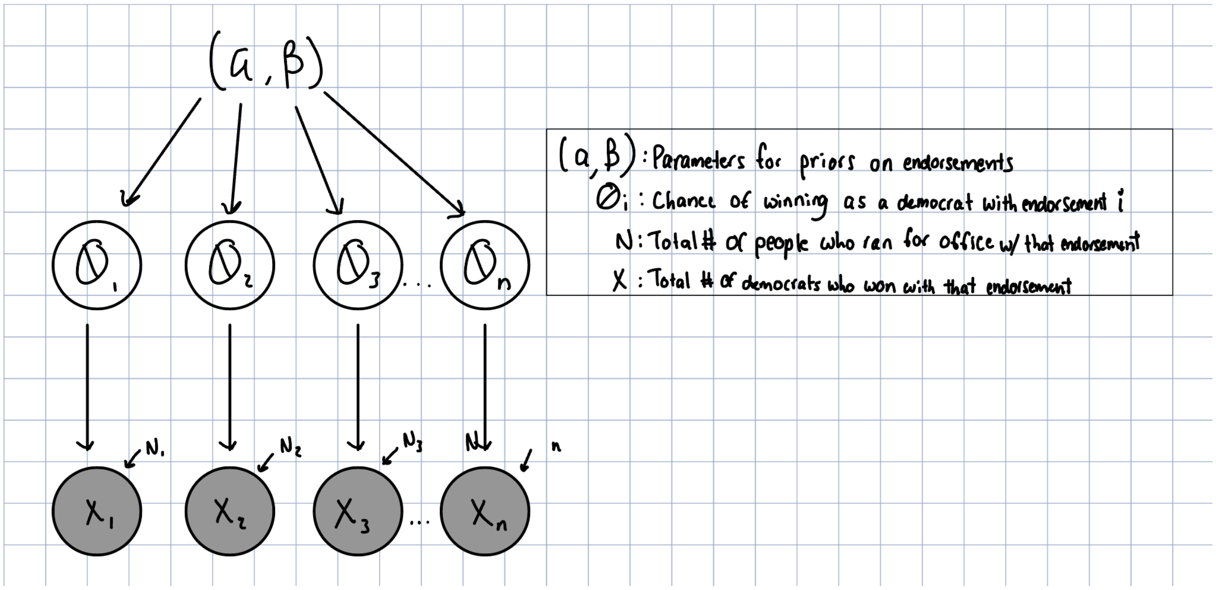
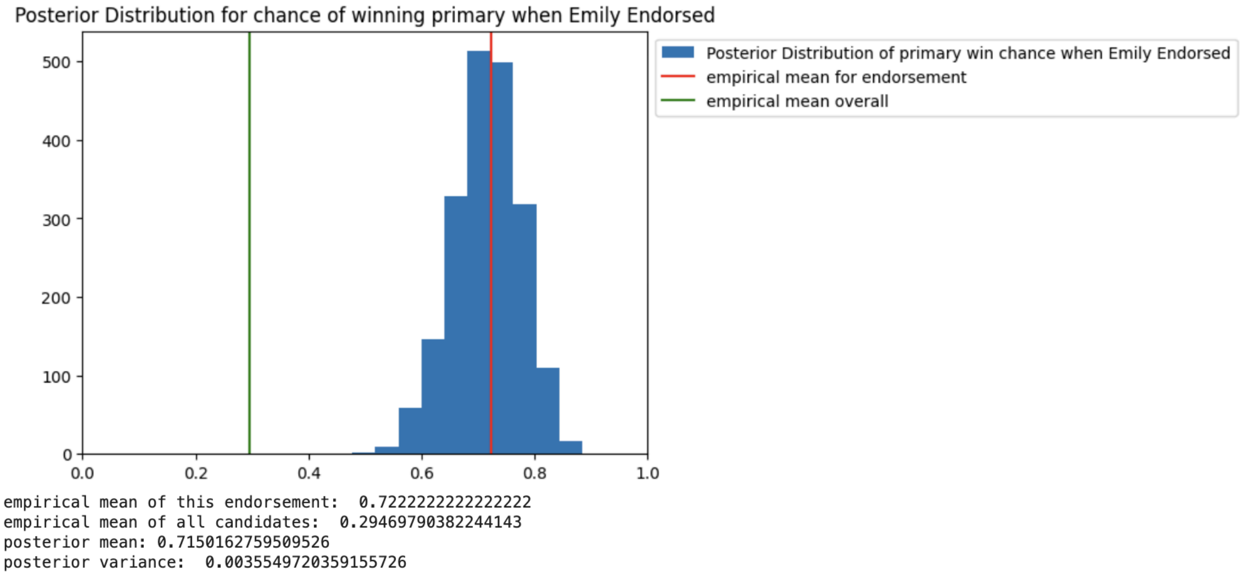


Figure 6


(α, β)
θ1
θ2
θ3
θn
N1
N2
N3
N
n
X1
X2
X3
Xn
(α, β): Parameters for priors on endorsements
θi : Chance of winning as a democrat with endorsement i
N: Total # of people who ran for office w/ that endorsement
X : Total # of democrats who won with that endorsement

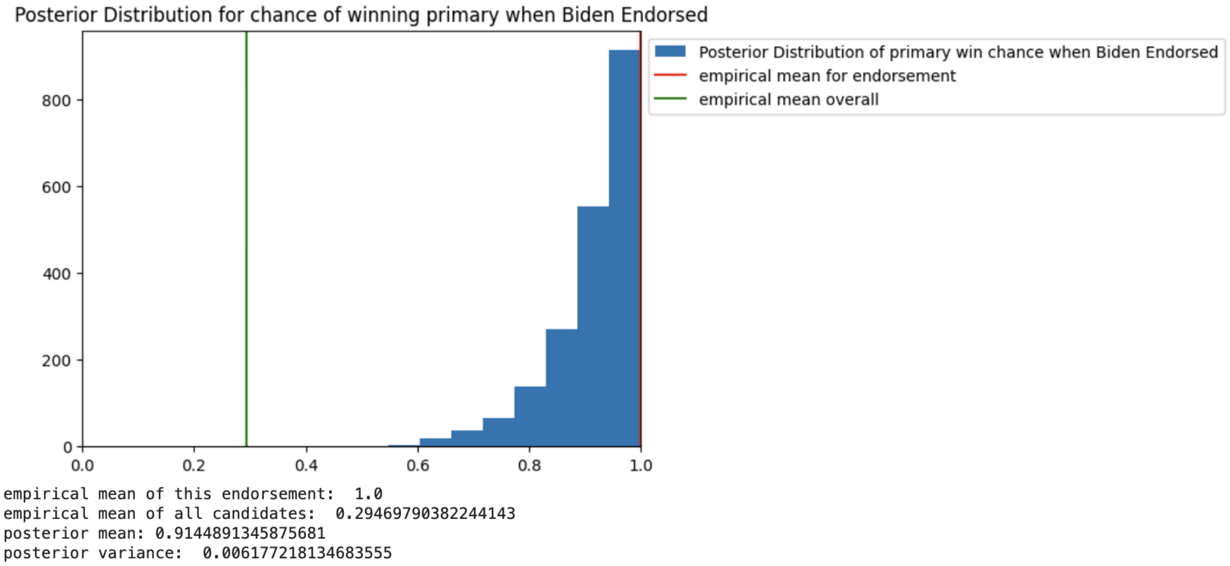
Our Bayesian Hierarchical model structure

**Figure 7**



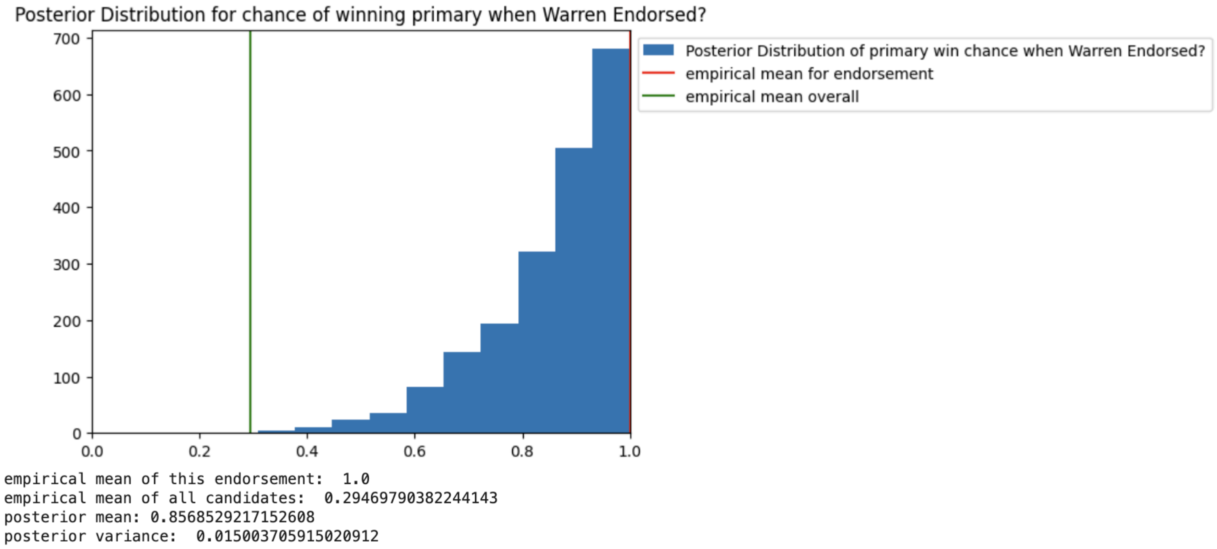
```
empirical mean of this endorsement:  0.7222222222222222
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.7150162759509526
posterior variance:  0.0035549720359155726
```

**Figure 8**



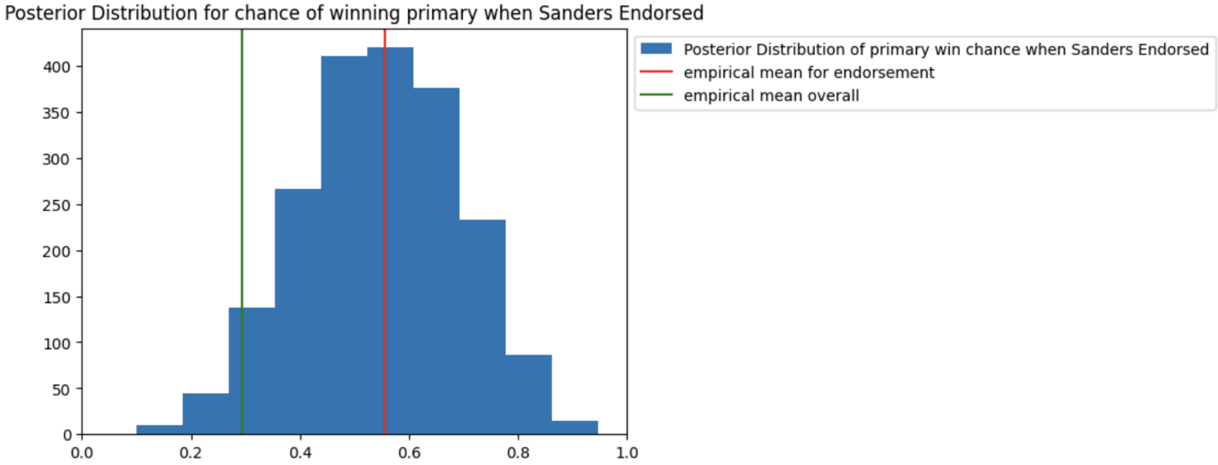
```
empirical mean of this endorsement:  1.0
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.9144891345875681
posterior variance:  0.006177218134683555
```

**Figure 9**



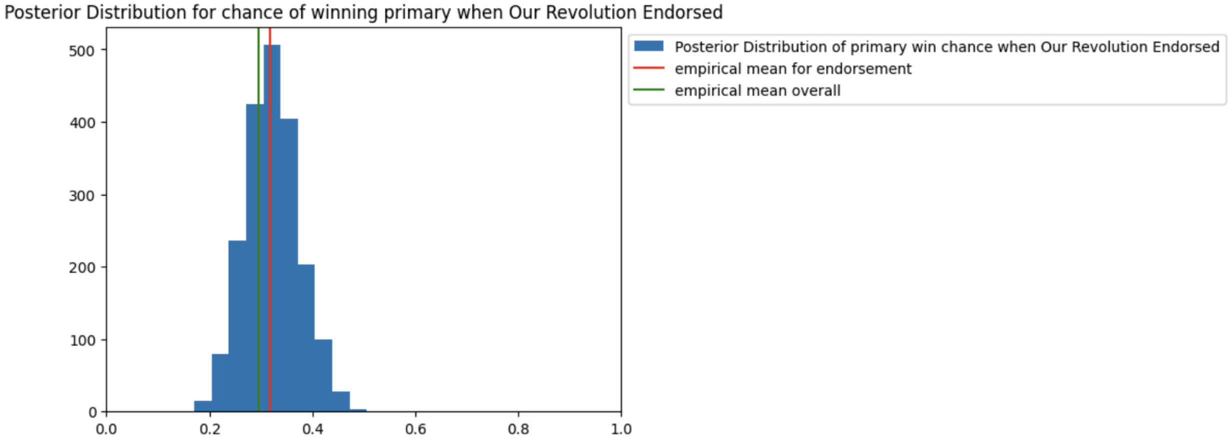
```
empirical mean of this endorsement:  1.0
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.8568529217152608
posterior variance:  0.015003705915020912
```

**Figure 10**



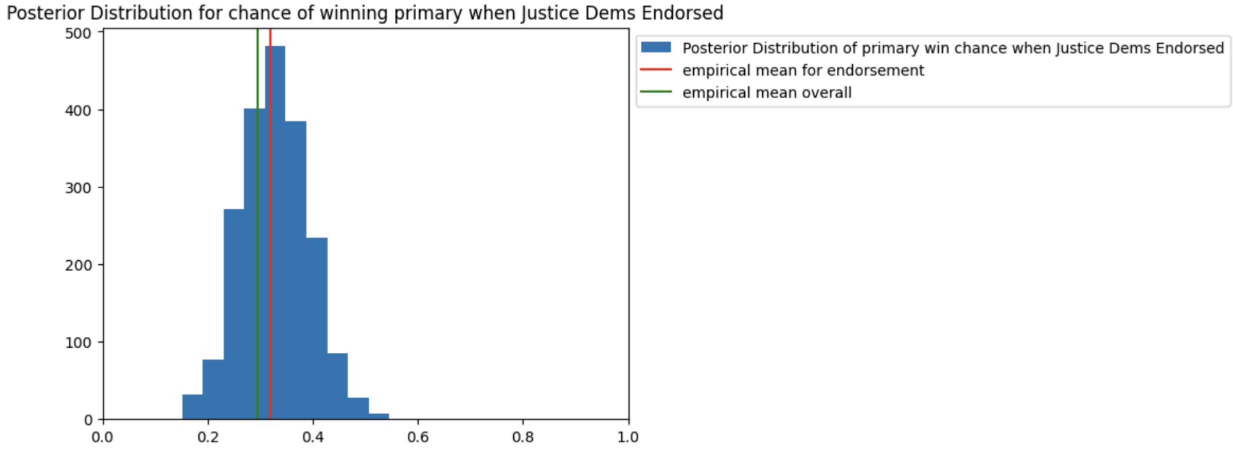
```
empirical mean of this endorsement:  0.5555555555555556
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.547853087113111
posterior variance:  0.020513589241731904
```

**Figure 11**



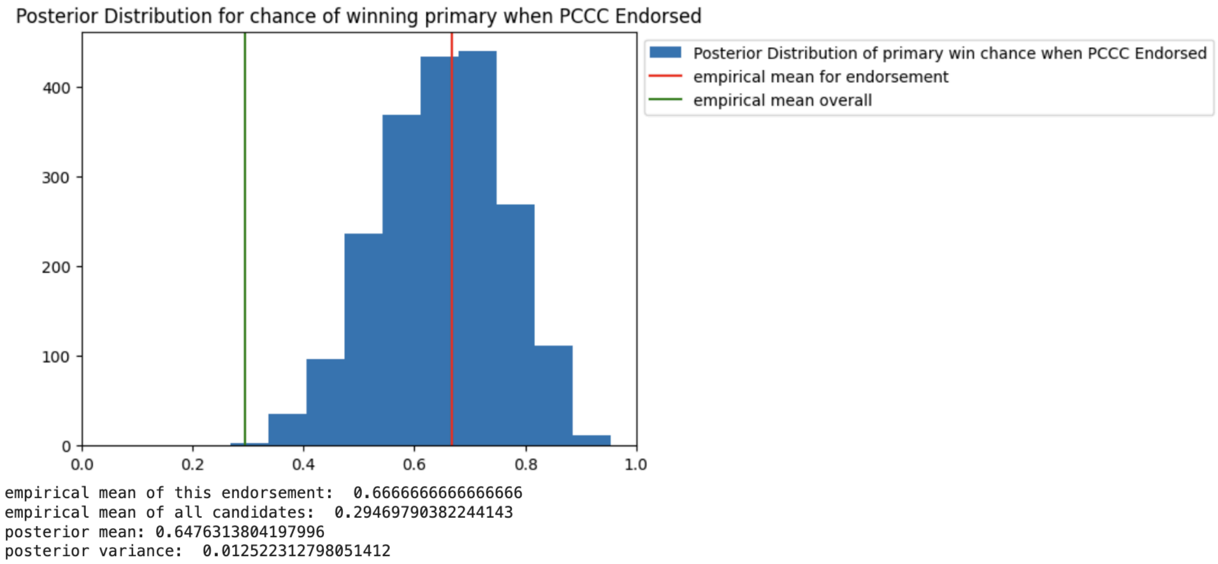
```
empirical mean of this endorsement:  0.3176470588235294
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.3222638274321191
posterior variance:  0.002667854709987896
```

**Figure 12**



```
empirical mean of this endorsement:  0.32
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.3275365436312887
posterior variance:  0.004052966790077538
```

**Figure 13**

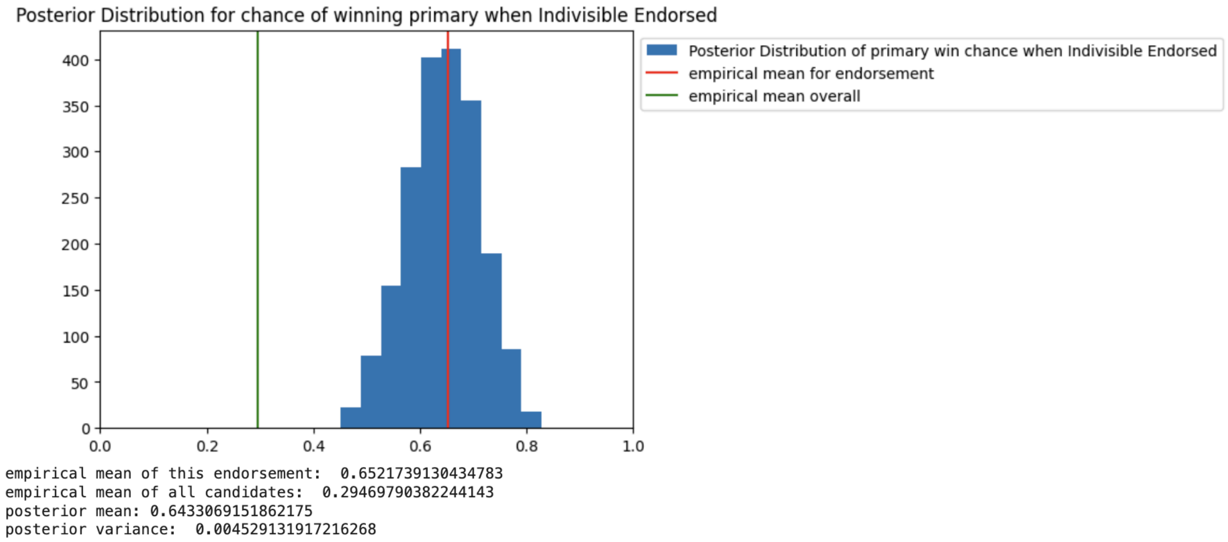


Posterior Distribution for chance of winning primary when PCCC Endorsed

Posterior Distribution of primary win chance when PCCC Endorsed
empirical mean for endorsement
empirical mean overall
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0

```
empirical mean of this endorsement:  0.6666666666666666
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.6476313804197996
posterior variance:  0.012522312798051412
```

**Figure 14**

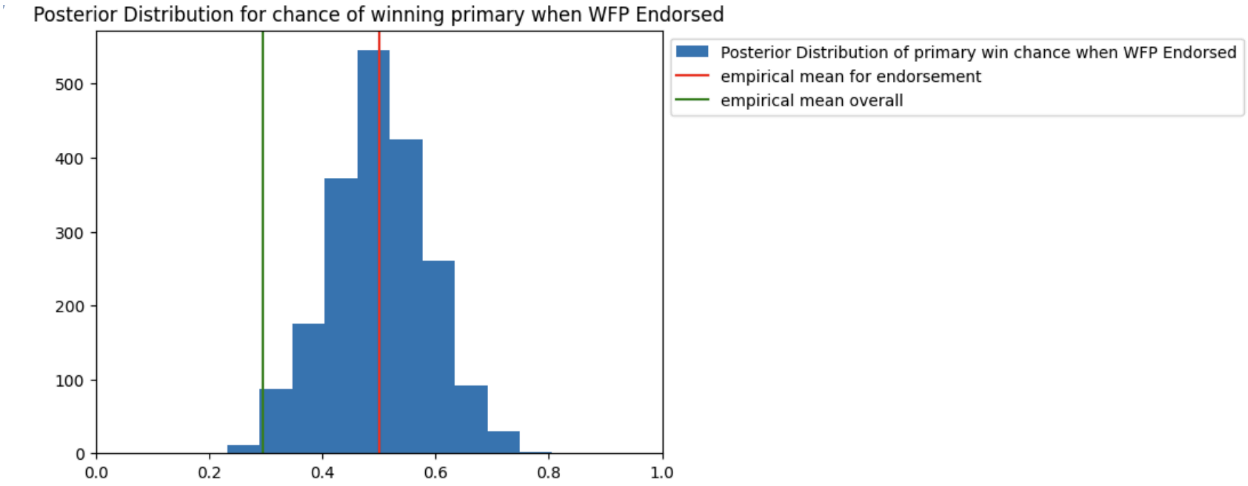


Posterior Distribution for chance of winning primary when Indivisible Endorsed

Posterior Distribution of primary win chance when Indivisible Endorsed
empirical mean for endorsement
empirical mean overall
0
50
100
150
200
250
300
350
400
0.0
0.2
0.4
0.6
0.8
1.0

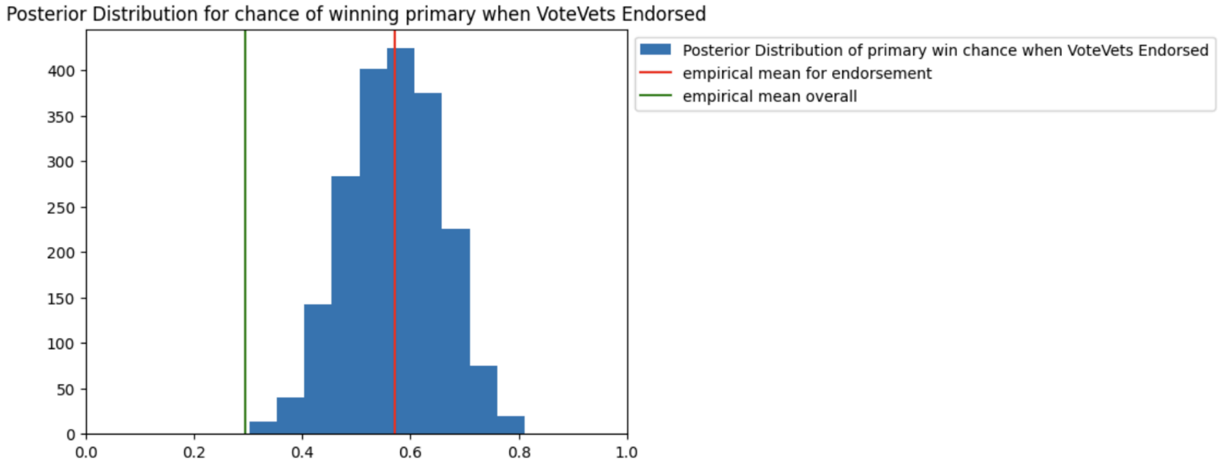
```
empirical mean of this endorsement:  0.6521739130434783
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.6433069151862175
posterior variance:  0.004529131917216268
```

**Figure 15**



```
empirical mean of this endorsement:  0.5
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.49958498156998693
posterior variance:  0.007695023798037745
```

**Figure 16**



```
empirical mean of this endorsement:  0.5714285714285714
empirical mean of all candidates:  0.29469790382244143
posterior mean: 0.5688360580419949
posterior variance:  0.007526045199406818
```

Works cited (all references were made on the first page):

[1] https://fivethirtyeight.com/features/the-establishment-is-beating-the-progressive-wing-in-democratic-primaries-so-far/

[2] https://www.census.gov/quickfacts/fact/table/US/PST045222