2020 U.S. Democratic Nomination: Analysis

Brad Reeves
CS Masters Degree Program
Central Washington University
Ellensburg, WA, USA
bradley.reeves@cwu.edu

Cheng Su
CS Masters Degree Program
Central Washington University
Ellensburg, WA, USA
suc@cwu.edu

Paul Collet
CS Masters Degree Program
Central Washington University
Ellensburg, WA, USA
paul.collet@cwu.edu

Tin Nguyen
CS Masters Degree Program
Central Washington University
Ellensburg, WA, USA
ntin@cwu.edu

Abstract—The effectiveness of opinion polls came under scrutiny after the U.S. presidential election in 2016. Most polls showed Hillary Clinton as winning, but Donald Trump won instead. This report takes real-time data to determine if opinion polls, fundraising data, and advertising expenses truly predict how people will vote.

Index Terms—Democratic Presidential Nomination, Polls, Funding, Advertising, Delegates

I. Introduction

Opinion polls are usually designed to represent the opinions of a population by conducting a series of questions and then extrapolating generalities in ratio or within confidence intervals [3]. Politicians and various political groups use such polls to determine what voters are willing to support. Campaigning for a cause, or person, in the U.S.A can be an expensive endeavor. In the last presidential race, in 2008, presidential candidates spent about $2.4 billion [1]. As of March 2020, over $1.5 million has been spent by the candidates [2]. The election is not until November 2020. With so much money involved, accurate polling could save candidates and interest groups several millions of dollars.

However, accurately polling a U.S.A presidential election cycle is extremely difficult. How candidates are selected in the country is more complex than most, if not all, other elective governments. Each state within the union has jurisdiction over how delegates are chosen to represent the state, as part of the Electoral College. Besides that lack of consistency, how each poll is administered depends on the organization conducting it. One of the biggest examples of polls not correctly reflecting the actual votes was during the 2008 presidential election. Almost every public poll had Hillary Clinton, the Democratic party nominee, winning the election when Donald Trump, the Republican nominee, actually won. When analysts studied why that was, they discovered that most of the polls wouldn't question the same people. Randomizing the samples for a poll is important to be unbiased, but so is tracking the change of opinions over time. It was the polls that tracked that change, that correctly predicted the election.

This report takes the polling data from Five-Thirty-Eight, a website that focuses on opinion poll analysis, politics, economics, and sports blogging [4]. They use GitHub, a common online storage hub, to store their data in a comma-separated values (CSV) format. The line graphs in this report use the following: https://projects.fivethirtyeight.com/polls-page/president_primary_polls.csv. The period covered spans from Jan 1, 2020, to when this report was written.

Funding data comes from the Federal Election Commission in CSV format. The time covers from 2019 Q1 to 2019 Q4. This project focus on the amount of funding for each committee from each state, and the type of funding that includes individuals, campaign committee, another committee, organization, political party committee, and Political action committee.

II. Hypothesis

- **Null Hypothesis**: The polling data reflects the presidential democratic primary outcomes

- **Alternative Hypothesis**: The polling data *does not* reflect the presidential democratic primary outcomes

III. Analysis

Figures 1 and 2 show the average candidate polling from February 1, 2020, to March 5, 2020. From the moderate rise and fall of Michael Bloomberg to Joe Biden's surprising March surge, many interesting events occurred in this time-frame.

Figure 1 shows polling averages in relation to voting days. As of March 5, five voting events took place: the Iowa Caucus, the New Hampshire Primary, the Nevada Caucus, the South Carolina Primary and Super Tuesday. Super Tuesday is a major event where voting occurs in 15 regions of the US.
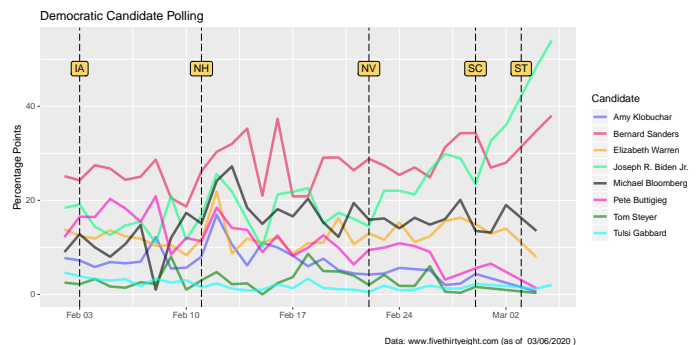

Fig. 1. Voting Events and Polls

Figure 2 shows polling averages in relation to democratic debates. During this period, the seventh, eight, ninth, and tenth debates happened. We can see that debates did have an influence on polling numbers.
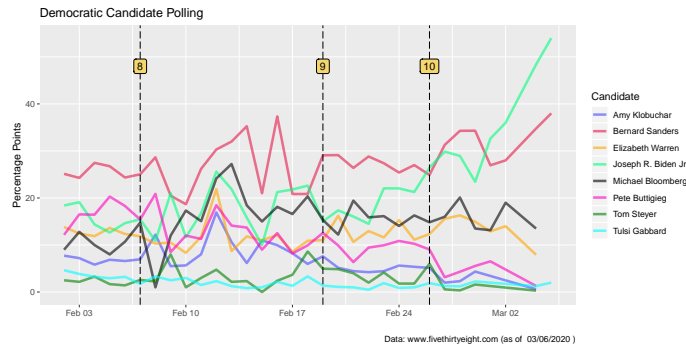

Fig. 2. Debates and Polls

These polls will be analyzed by event in more detail throughout the following sections.

A. Iowa Caucus

Iowa Caucus was the first national contest in the US democratic nomination cycle. With 41 delegates at stake, it provided candidates the opportunity to set themselves apart in a crowded field.
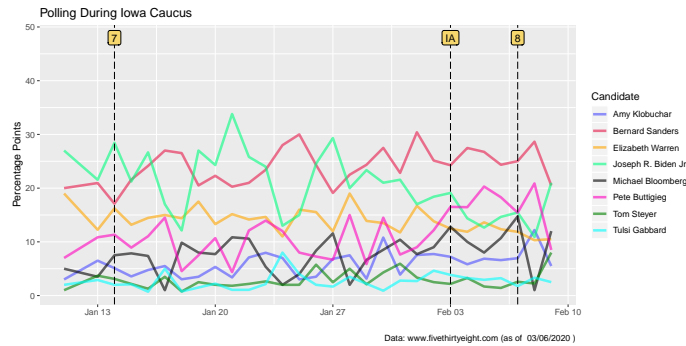
Polling

Fig. 3. Polling Near Iowa Caucus

Focusing on the events shown in Figure 3, the Iowa Caucus and the seventh and eighth Democratic Primary Debates, each event greatly influenced the public opinion of the candidates. The seventh debate negatively affected the public opinions for Joseph Biden, Elizabeth Warren, and Pete Buttigieg in relation to their competitors, while opinions of Bernard Sanders improved. This trend repeated itself in the Iowa Caucus.
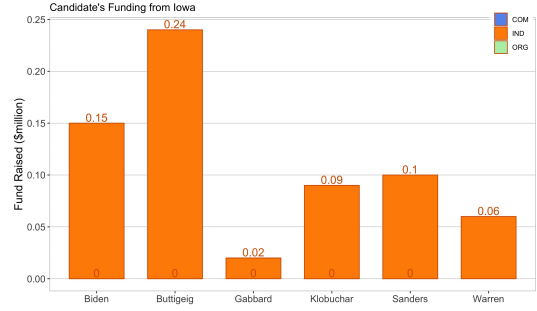
Fundraising

Fig. 4. Fundraising from Iowa Caucus

Fund raised from Iowa Caucus mostly comes from individuals.

Ads Spending

Ads spending data during Iowa Caucus from FiveThirtyEight [6] aren't time series data versus polling time series data which makes it very difficult to draw any correlation between money spent and candidate's pooling result. However, this data can be used to analyze if ad spending is a meaningful predictor of the outcome of this Caucus.

Let's first examine the ads spending data during Iowa Caucus of all primary candidates on the chart below:
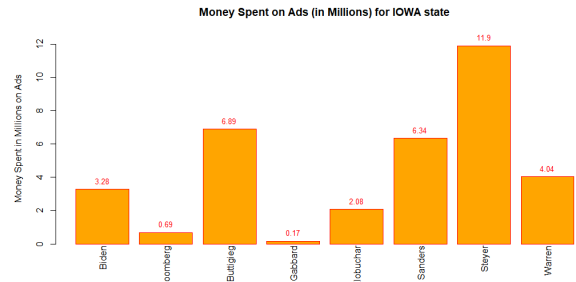

Fig. 5. Ads Spending During Iowa Caucus

Table 1: Top 5 candidates during Iowa Caucus

Candidates	Polling %
Sander	25%
Buttigieg	18%
Biden	15%
Warren	12%
Bloomberg	12%

Based on the two data which are presented above, if there is a positive correlation between money spent and polling result, Sander would be spending the most during this Caucus. However, the money Steyer spent on ads was almost double Sander or Buttigieg but he didn't even make it in the top five list. It is safe to say that there is no correlation between these data during this Iowa Caucus.

Outcome

Buttigieg ended up securing the most delegates from Iowa. He was followed by Sanders, Warren, then Biden. Leading up to the caucus, Sanders was polling the highest with Joe Biden closely following. We can see that the results did not reflect the polling averages.

B. New Hampshire Primary

The New Hampshire Primary is a relatively small contest with only 24 delegates on the table. Buttigieg won the most delegates in Iowa giving him momentum for this primary.
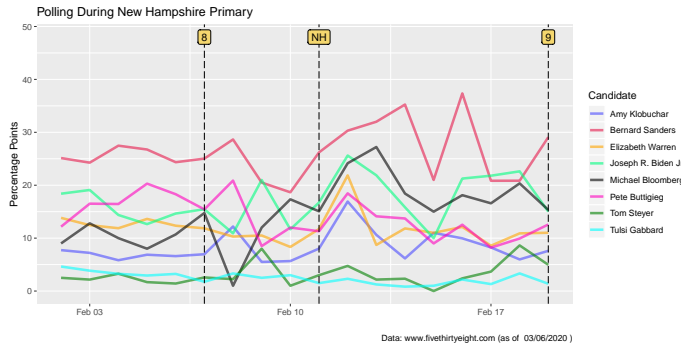
Polling

Fig. 6. Polling Near New Hampshire Primary

The eighth and ninth debates, as well as the New Hampshire primary, are shown in Figure 6. Opinions of Sanders, Buttigieg, and Klobuchar improved after the eighth debate, while opinions of Biden and Bloomberg decreased. Interestingly, the public opinions of all the candidates improved after the primary.
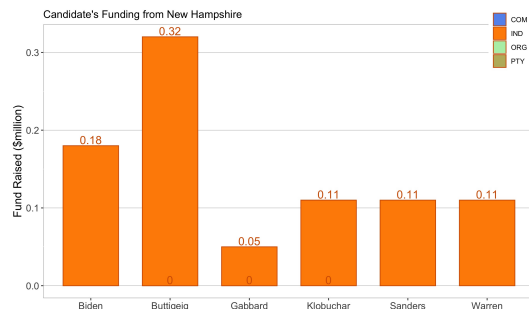
Fundraising

Fig. 7. Fundraising from New Hampshire

Fund raised from New Hampshire mostly comes from individuals.

Ads Spending

Similar to the ads spending data that was presented for the Iowa Caucus, Ads spending data during New Hampshire Primary which was obtained from FiveThirtyEight [6] also aren't

time series data. However, when analyzing this ad spending data versus the candidate polling, it provides some useful information to determine if ads spending is a meaningful predictor to determine the outcome of the hypothesis.

Let's first examine the ads spending data during New Hampshire Primary of all primary candidates on the chart below:
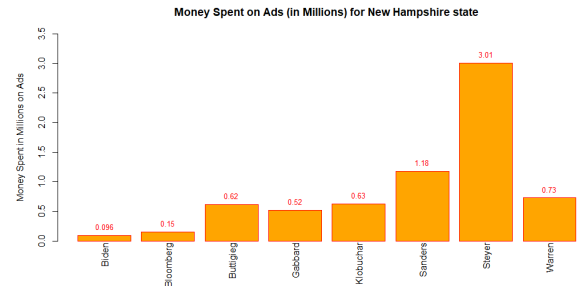

Fig. 8. Ads Spending During New Hampshire Primary

Table 2: Top 5 candidates during New Hampshire Primary

Candidates	Polling %
Sander	26%
Biden	19%
Bloomberg	18%
Warren	13%
Buttigieg	12%

Based on the two data which are presented above, if there is a positive correlation between money spent and polling results during New Hampshire Primary, Sander or Biden would be spending the most during this Caucus. However, the money Steyer spent on ads was almost triple Sander and so much more compare to Biden but Steyer wasn't even on top five lists during this Primary. It is safe to say that there is no correlation between these data during this Iowa Caucus.

Outcome

Buttigieg was able to carry momentum from Iowa into the New Hampshire contest where he and Sanders tied for the most delegates. Amy Klobuchar took the rest of the delegates up for grabs. Sanders was once again leading the polls so this was not surprising, however, Buttigieg and Klobuchar were both relatively low in polling.

C. Nevada Caucus

The Nevada Caucus is the first contest of the western US states. Being an early contest, it can have a lot of influence over who wins the party's nomination. Nevada had 36 total delegates at stake.
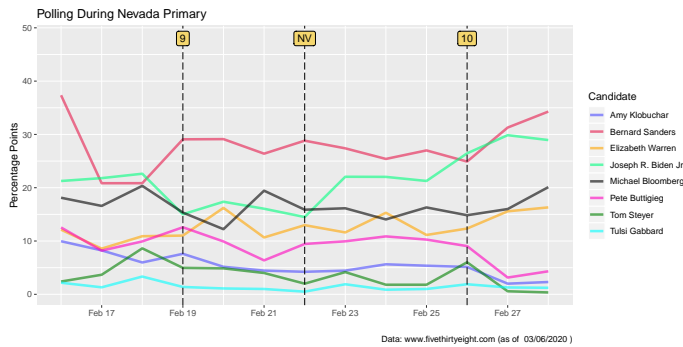
Polling

Fig. 9. Polling Near Nevada Caucus

The ninth and tenth debates, and the Nevada Primary, are shown in Figure 9. Opinions of Biden and Warren improved after the ninth debate, while opinions of Bloomberg, Buttigieg, and Klobuchar decreased. The public opinion for the other candidates seemed to be unaffected by the debate. After the Nevada Primary, however, opinions of Sanders, the front runner at the time, experienced a slight decrease, while opinions of Biden dramatically improved.
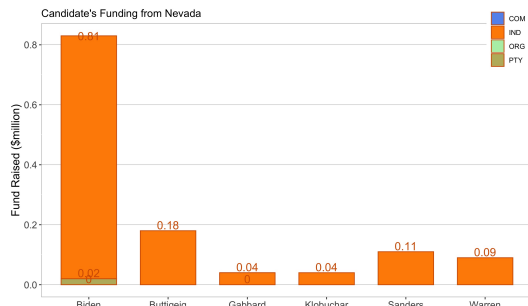
Fundraising

Fig. 10. Fundraising from Nevada

Fund raised from Nevada mostly comes from individuals.

Ads Spending

Ads spending data during Nevada Caucus from FiveThirtyEight [6] also aren't time series data. However, when analyzing this data versus the candidate polling, it is useful to observe if this ad spending data from the Nevada Caucus is a meaningful predictor to the outcome of the hypothesis.

Let's first examine the ads spending data during New Hampshire Primary of all primary candidates on the chart below:
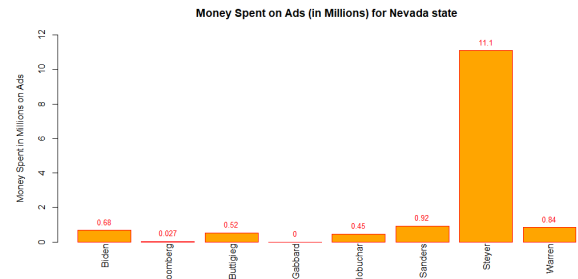

Fig. 11. Ads Spending During Nevada Caucus

Table 3: Top 5 candidates during Nevada Caucus

Candidates	Polling %
Sander	32%
Pete	25%
Biden	20%
Warren	15%
Bloomberg	13%

Based on the two data which are presented above, Sander is leading with 33%, Buttigieg is second with 25% but they didn't spend much money at all compare to Steyer on ads. Steyer was spending about 11 times more than the top 2 candidates but he didn't even make it to the top 5 list during the Nevada Caucus. It is safe to say that there is no correlation between these data during this Nevada Caucus and ads spending wasn't a meaningful predictor of the hypothesis.

Outcome

Sanders took the majority of delegates from Nevada. He was followed by Buttigieg and Biden. Sanders and Biden accurately reflected polling, but Buttigieg was still very low in the polls.

D. South Carolina Primary

The South Carolina Primary is the most important voting contest leading into Super Tuesday. With 54 delegates at stake, it can sway the momentum. Winning here gives candidates a sharp advantage moving forward.
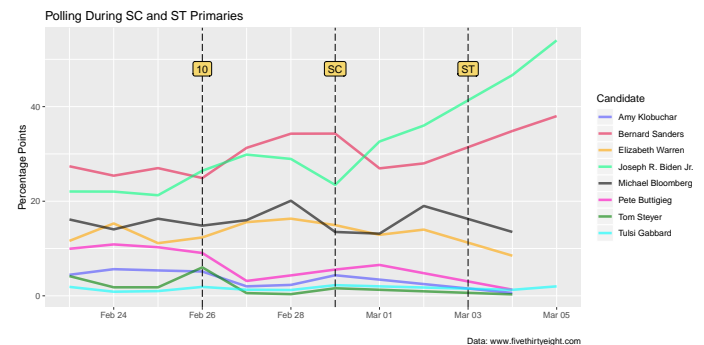
Polling

Fig. 12. Polling Near South Carolina Primary

The tenth, and final Democratic Primary Debate, as well as the South Carolina and Super Tuesday (ST) primaries, are shown in Figure 12. ST will be analyzed in the following section. Opinions of Sanders, Biden, Bloomberg, and Warren improved after the tenth debate, with opinions of Sanders seeing the highest rate of improvement. The opinions of the rest of the candidates decreased after the debate. That changed after the South Carolina Primary. Opinions of Biden improved dramatically after the SC primary, while opinions of Sanders, who kept the lead in the polls for much of 2020, decreased. The opinions of the other candidates were marginal by comparison.
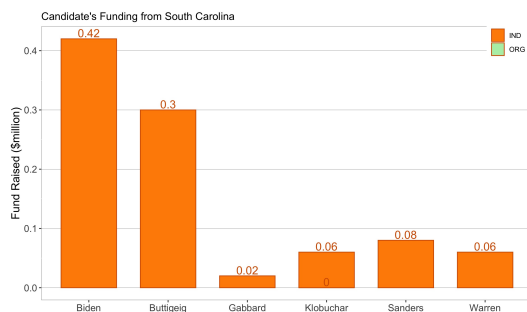
Fundraising

Fig. 13. Fundraising from South Carolina

Fund raised from South Carolina mostly comes from individuals.

Ads Spending

Ads spending data during South Carolina Primary from FiveThirtyEight [6] also aren't time series data. However, when analyzing this data versus the candidate polling, it is useful to observe if this ad spending data from South Carolina is a meaningful predictor of the outcome of the hypothesis.

Let's first examine the ads spending data during South Carolina Primary of all primary candidates on the chart below:
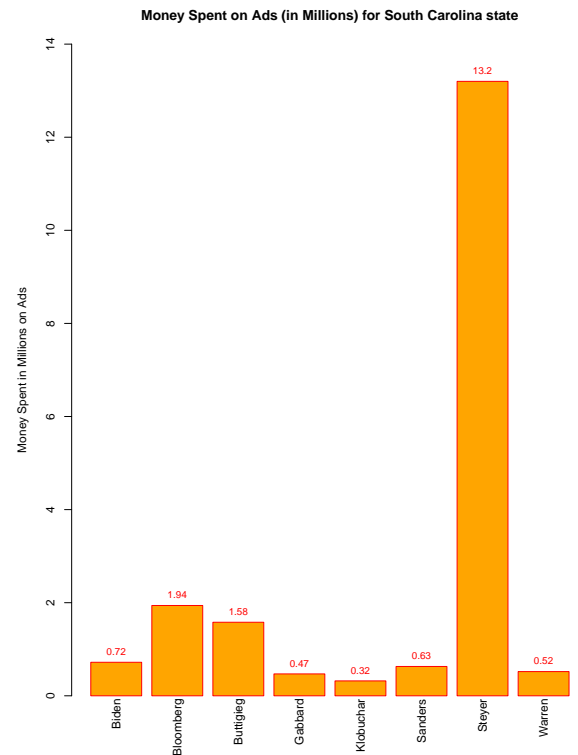

Fig. 14. Ads Spending During South Carolina Primary

Table 4: Top 5 candidates during South Carolina Primary

Candidates	Polling %
Sander	34%
Biden	24%
Warren	15%
Bloomberg	13%
Pete	7%

Based on the two data which are presented above, Sander is leading with 34%, Biden is second with 24% but they didn't spend much money at all compare to Steyer on ads. The top three candidates during South Carolina almost spent the least of money. Steyer was spending about 13 times more than the top 2 candidates but he didn't even make it to the top 5 list during this Nevada Caucus. It is safe to say that there is no correlation between these data during this South Carolina Primary and ads spending wasn't a meaningful predictor to the hypothesis.

Outcome

Biden took a majority of delegates in South Carolina. The only other candidate to win any delegates in this state was Bernie Sanders. At this point, they were both heavily favored in the polls so this was expected. This contest was interesting because it set the tone for Super Tuesday.

E. Super Tuesday

Super Tuesday (ST) refers to the first Tuesday in March in which the largest number of state primaries, for all political

parties, occur simultaneously. The following states have their primaries on ST in 2020: Alabama, Arkansas, California, Colorado, Democrats Abroad, Maine, Massachusetts, Minnesota, North Carolina, Oklahoma, Tennessee, Texas, Utah, Vermont, and Virginia.

ST is a "make it or break it" moment for most candidates. Up until this point, only 155 delegates were awarded in total. ST accounts for a whopping $1,357$ delegates. In 2020, ST narrowed the field to just a couple of candidates.
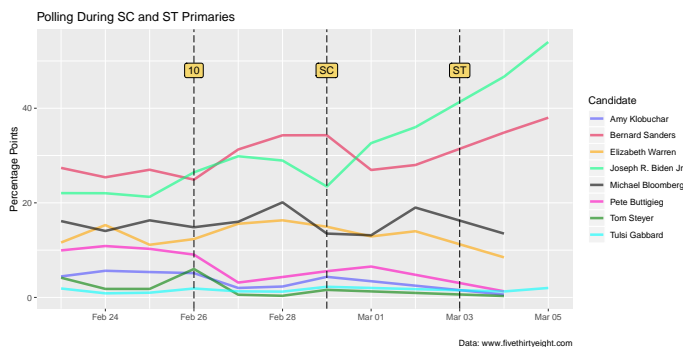
Polling

Fig. 15. Polling Near Super Tuesday

Leading into Super Tuesday, opinions of Sanders were already improving and that trend continued during ST. However, the opinions of Biden remained higher. The opinions of the other candidates either continued a decline or remain unaffected. It should be noted, however, that Steyer, Buttigieg, and Klobuchar, though still in the polls, had dropped from the race before ST.
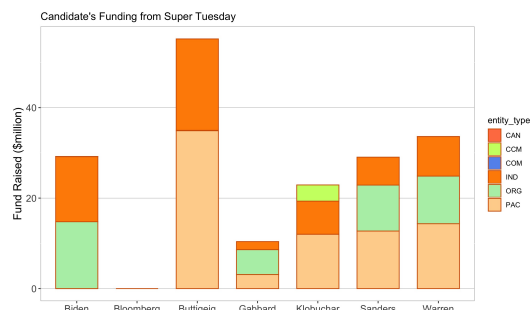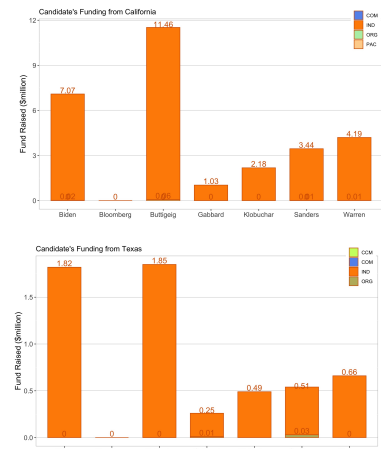
Fundraising

Fig. 16. Fundraising from Super Tuesday states

Fund raised from Super Tuesday states has more variety of fund sources. Fund-raise graph for California and Texas are selected as representation of important states of Super Tuesday.

Ads Spending

Ads spending data during Super Tuesday will only focus on two states which have the largest impact on the primary result which are California and Texas. The data will also be collected on FiveThirtyEight website FiveThirtyEight [6] and they are not series data. However, when analyzing this data versus the candidate polling, it is useful to observe if this ad spending data from Super Tuesday is a meaningful predictor of the outcome of the hypothesis.

Let's first examine the ads spending data from California of all primary candidates on the chart below:
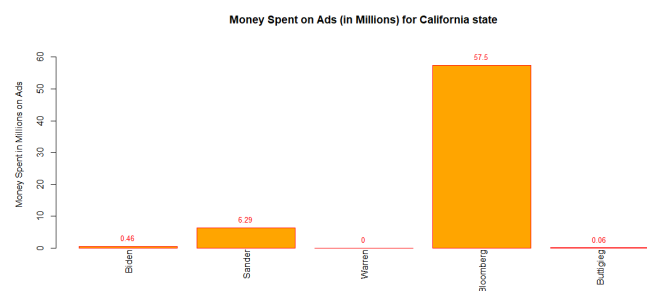

Fig. 17. Ads Spending from California

examine the ads spending data from Texas of all primary candidates on the chart below:
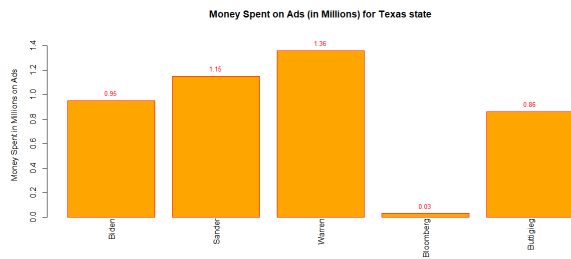
Fig. 18. Ads Spending from Texas

Table 5: Top 5 candidates during Super Tuesday

Candidates	Polling %
Biden	45%
Bernie	35%
Bloomberg	13%
Warren	8%
Pete	2%

Based on the results of the top two candidates after Super Tuesday, the data shows that there is a negative relationship between money spent on ads and the number of delegates each of the top two candidates received. Compare to Sander, Biden spent a lot less money on ads on California which was 0.5 million compares to 6.29 million Sander spending, and 0.95 million from Biden compare to 1.15 million from Sanders in Texas. And Biden received about 45% of total delegates after Super Tuesday compare to 35% of total delegates of Sander. Based on this data, it is safe to say that there is no correlation between these data during this South Carolina Primary and ads spending wasn't a meaningful predictor to the hypothesis because spending more money on ads should not hurt the results of the primary.

Outcome

Joe Biden had the strongest showing on Super Tuesday. He finished with 76 more delegates than Bernie Sanders. All other candidates received so few delegates that they dropped out of the race. Gabbard is still competing even though she has only secured 2 delegates. After South Carolina, Sanders dropped in the polls while Biden climbed. It was interesting to see the shift in public opinion during the days leading up to and after the South Carolina primary. Before this, it looked like Sanders would run away with the nomination. Now, it looks like Joe Biden has a better chance of winning.

F. National/Overall Summary

Polling

During a presidential campaign, opinion polls can show how different events in time affect people's perceptions of candidates. Up until the ninth debate, opinions of the candidates would fluctuate up to 20 percentage points within a week. After the ninth debate, there were fewer fluctuations and more discernible trends. After the tenth debate, only

data associated with Biden and Sanders showed much change.
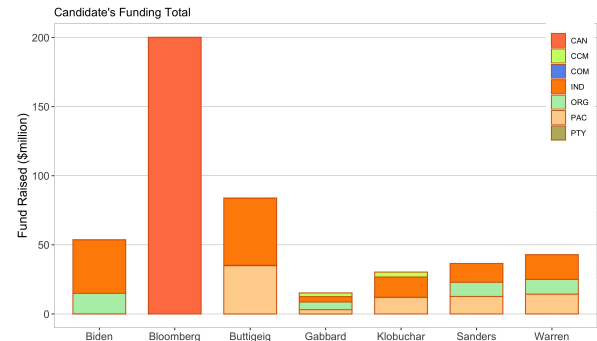
Fundraising

Fig. 19. Candidates fundraising for all states
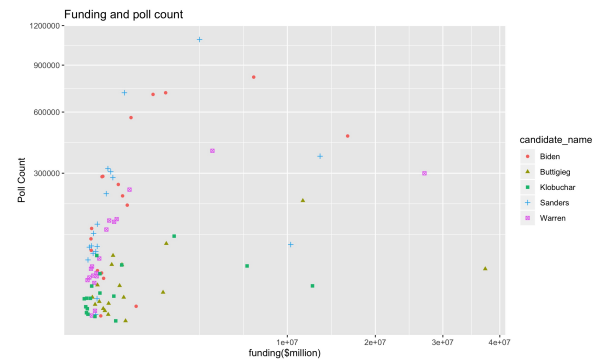
Fig. 20. Candidates fundraising for all poll states

The graph above shows the relationship between poll count and fundraising for each candidate for all the polled states. Candidates need to be mentioned is Bloomberg, who is self-funded and would be less meaningful to see the relationship between his fund and his poll.

Polynomial curve fitting and logarithmic curve fitting were tested on the data for each candidate covered above. Polynomial curve fitting has an overall smaller p-value compared to the logarithmic curve for all of the candidates, which shows a significance in the relationship, and adjusted R^2 in between 0.6 and 0.9 that more than half of the data could be explained by the fit.

Below is an example of fundraising for Biden:

More things could be tested in this experiment to help improving the models using percentage of polls and percentage of fund from each state instead to see the relationship between poll and fund-raise, and separating the types of fund source to see if there is a difference in affect of fund on delegates and total polls.

```
le(formula = Count ~ log(contribution_receipt_amount), data = biden_amount)

Residuals:
    Min      1Q  Median      3Q     Max
-360067 -135845    6577  124539  281980

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -1338760     371277  -3.606 0.002370 **
log(contribution_receipt_amount)  124832      28583   4.367 0.000479 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 188100 on 16 degrees of freedom
Multiple R-squared:  0.5438,   Adjusted R-squared:  0.5153
F-statistic: 19.07 on 1 and 16 DF,  p-value: 0.0004786
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-381795  -71544    2603   47933  233675

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                           271128      36278   7.474   3e-06 ***
poly(contribution_receipt_amount, 3)1 498248     153915   3.237 0.005963 **
poly(contribution_receipt_amount, 3)2 -702270    153915  -4.563 0.000443 ***
poly(contribution_receipt_amount, 3)3 409662     153915   2.662 0.018603 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 153900 on 14 degrees of freedom
Multiple R-squared:  0.7327,   Adjusted R-squared:  0.6755
F-statistic: 12.79 on 3 and 14 DF,  p-value: 0.0002679
```

Ads Spending

Surprisingly, the ads spending data that were collected during the democratic campaign 2020 from FiveThirtyEight [6] don't have any correlation with the primary results, so it is not a meaningful predictor to test the hypothesis.

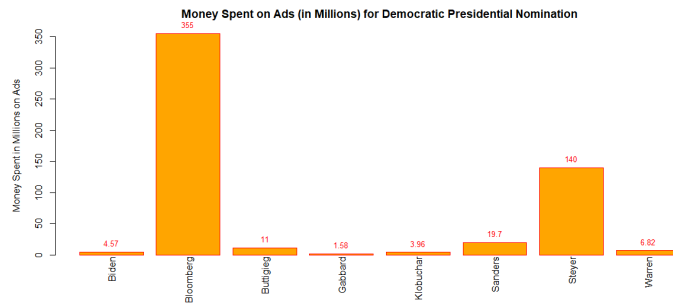Let's examine the total money spent on ads of all candidates:

Fig. 21. National Ads Spending - All Candidates

Table 6: Top 5 candidates in the Primary

Candidates	Polling %
Biden	45%
Bernie	35%
Bloomberg	13%
Warren	8%
Pete	2%

Bloomberg and Steyer were the big spenders in the democratic primary campaign (355 millions and 140 millions respectfully) versus all other candidates spent in average about 10 millions for their ads campaign, however money spent on ads doesn't reflect or have any contributions to the outcome of the primary results as discussed above.

Outcome

Surprisingly, Biden has accrued the most delegates so far with 627. Sanders is a close second with 551 delegates. For candidates to secure their nomination on the first ballot at the Democratic National Convention, they need 1991 delegates. If no candidate gets the necessary delegates needed, it could lead to contested or split convention. Additional rounds of voting will occur to select the democratic nominee. If Bernie Sanders has more delegates than Joe Biden at the convention, then we will likely see this. Sanders has publicly stated that if he is trailing in delegates at convention time, he will drop his candidacy. Figure 21 shows the delegate count just after Super Tuesday. All candidates have since dropped out except Joe Biden, Bernie Sanders, and Tulsi Gabbard.
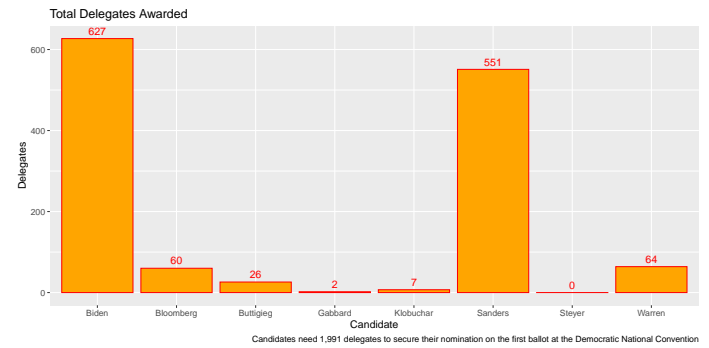

Fig. 22. Delegate Count as of March 5, 2020

IV. CONCLUSION

Because polling, spending, and funding data did not reflect outcomes of the primary, the null hypothesis is rejected and the alternative hypothesis is accepted.

REFERENCES

[1] Wikipedia Contributors (2019). Campaign finance in the United States. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Campaign_finance_in_the_United_States.
[2] OpenSecrets. (2016). 2020 Presidential Race. [online] Available at: http://www.opensecrets.org/2020-presidential-race.
[3] Wikipedia Contributors (2019). Opinion poll. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Opinion_poll.
[4] Wikipedia Contributors (2019). FiveThirtyEight. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/FiveThirtyEight [Accessed 11 Jun. 2019].
[5] How Many Delegates Do The 2020 Democratic Presidential Candidates Have? [online]. https://www.npr.org/2020/02/10/799979293/how-many-delegates-do-the-2020-presidential-democratic-candidates-have
[6] FiveThiryEight-Who's buying ads?. [online]. Available at: https://projects.fivethirtyeight.com/2020-campaign-ads/