2016 Clinton Presidential Campaign

Senior Meeting

horizontal line

4:30 PM / ROOM 436

# ATTENDEES

Ryne Schultz, Joe Schmo, Jane Doe, Wendy Writer, Ronny Reader, Abby Author

# AGENDA

## Last Meeting Follow-up

1. Review funding goals.
2. Updates since the last meeting.

## New Business

* Walkthrough of the new election model.
* Discuss the model’s findings:
  + Swing state categorization
  + Election predictions
  + Discussion of Grand Strategy

# RECOMMENDATIONS

Our model recommends the following “hit list” of swing states in which to campaign:

1. Florida (29)
2. Ohio (18)
3. North Carolina (15)
4. Arizona (11)
5. Iowa (6)
6. Nevada (6)

However, though our model predicts that Clinton will win, the model predicts that she will win by only 8 electoral votes (273 to 265), suggesting that her lead is tenuous. Therefore, we recommend that the Clinton campaign shore up support in those Democratic states that are the most “at risk” of defecting to Trump. These states include:

1. Pennsylvania (20)
2. Michigan (16)
3. Wisconsin (10)
4. Colorado (9)
5. New Hampshire (4)
6. Maine (4)

Our recommendation is that the Clinton campaign should dedicate most of its resources to these states - focusing primarily on Pennsylvania, Michigan, and Wisconsin, the largest “at risk” states - only dedicating resources to the “hit list” swing states if time and resources allow.

It is our estimation that by shoring up her base in these “at risk” states, Clinton has the greatest chances of becoming president.

# MODEL WALKTHROUGH

The data utilized in this model comes from FiveThirtyEight’s data page on [kaggle.com](https://www.kaggle.com/fivethirtyeight/2016-election-polls#presidential_polls.csv). The original file is a .csv file containing 27 fields. There were many superfluous columns in the dataset that we removed. The following is a list of those removed fields along with the reasoning behind each removal:

* **cycle** - all values were the same (i.e. set to "2016").
* **branch** - all values were the same (i.e. set to "Presidential").
* **type** - this determines which model FiveThirtyEight associates with each poll (i.e. "Now-Cast", "Polls-Only", or "Polls-Plus"), which is irrelevant for our model.
* **matchup** - all values were the same (i.e. set to "Clinton vs. Trump vs. Johnson")
* **forecastdate** - all values were the same (i.e. set to "11/1/2016")
* **rawpoll\_johnson** - our model only compares voting patterns between the two likeliest presidential candidates (i.e. Clinton and Trump) and therefore excludes data for both Johnson and McMullin.
* **rawpoll\_mcmullin** - our model only compares voting patterns between the two likeliest presidential candidates (i.e. Clinton and Trump) and therefore excludes data for both Johnson and McMullin.
* **adjpoll\_johnson** - our model only compares voting patterns between the two likeliest presidential candidates (i.e. Clinton and Trump) and therefore excludes data for both Johnson and McMullin.
* **adjpoll\_mcmullin** - our model only compares voting patterns between the two likeliest presidential candidates (i.e. Clinton and Trump) and therefore excludes data for both Johnson and McMullin.
* **multiversions** - this field denotes if there were multiple versions of the same poll. Our analysis will not make this distinction and will simply take each poll at face value.
* **url** - links to the original polling data location.
* **poll\_id** - unique identifier used by FiveThirtyEight to identify polls in other models.
* **question\_id** - unique identifier to determine the form of the poll question. We do not have access to the reference table for these questions so this will not be included in our dataset.
* **createddate** - the date the data was inputted to FiveThirtyEight's model, which is irrelevant for our data set.
* **timestamp** - date the model was rerun, incorporating all the data up to that point. The model was last updated November 1st, 2018 and is irrelevant to our data set.

The result was a new, clean dataset containing the following fields:

* **state** - the state from which the poll was taken (the value "U.S." signifies a national poll).
* **startdate** - the date on which the poll began.
* **enddate** - the date on which the poll ended.
* **pollster** - the organization that conducted the poll.
* **grade** - the accuracy rating of the polling organization. FiveThirtyEight determines this grade by analyzing a pollster's past predictions, comparing them to actual results. The more accurate the pollster the higher the grade.
* **samplesize** - the size of the sample for the poll (i.e. the "n" value).
* **population** - a unique regional identifier.
* **poll\_wt** - the weight assigned to each poll predicated on the polling organization's grade.
* **rawpoll\_clinton** - the raw percentage of respondents polled who plan on voting for Hillary Clinton.
* **rawpoll\_trump** - the raw percentage of respondents polled who plan on voting for Donald Trump.
* **adjpoll\_clinton** - the weight-adjusted percentage of respondents polled who plan on voting for Hillary Clinton.
* **adjpoll\_trump** - the weight-adjusted percentage of respondents polled who plan on voting for Donald Trump.

In order to get familiar with the election data, we decided to analyze national polling data patterns (i.e. where the state field is set to “U.S.”). These polls represent a subset of Americans from around the country, as opposed to a specific state.

To do so, we filtered the data even further to include only the following variables:

* **state**
* **enddate**
* **rawpoll\_clinton**
* **rawpoll\_trump**

A look at the summary of the data revealed some important, baseline statistics:

**state enddate rawpoll\_clinton rawpoll\_trump**

Length:3105 Min. :2015-11-16 Min. :31.00 Min. :28.80

Class :character 1st Qu.:2016-05-22 1st Qu.:42.00 1st Qu.:36.00

Mode :character Median :2016-08-09 Median :43.90 Median :38.50

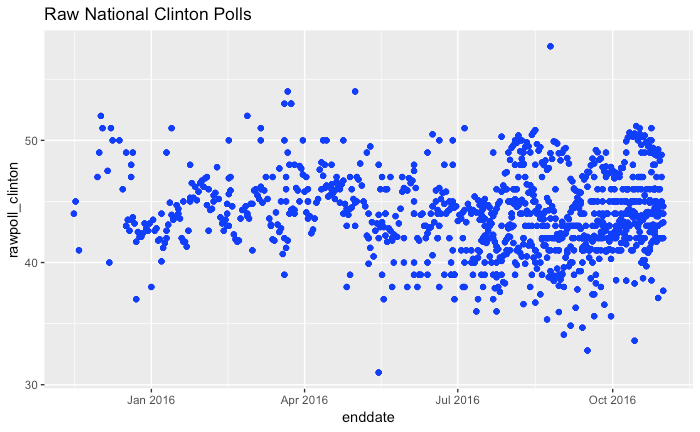
Mean :2016-07-16 Mean :43.91 Mean :39.43

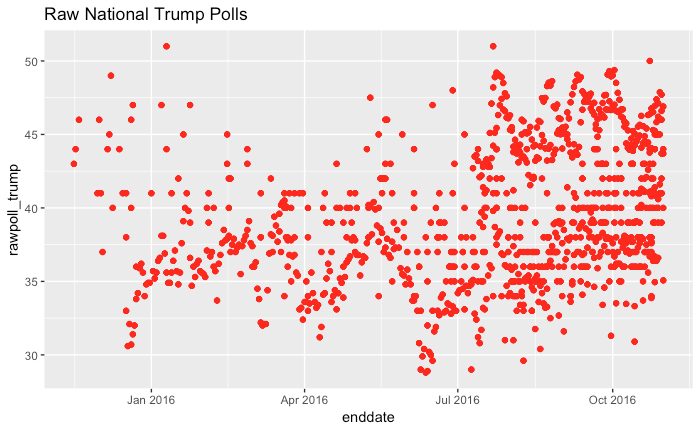
3rd Qu.:2016-09-25 3rd Qu.:46.00 3rd Qu.:43.00

Max. :2016-10-31 Max. :57.70 Max. :51.00

First, it provided us with the total number of observations - i.e. the number of national polls: 3,105. The data also provided us with the time frame over which these polls were conducted: November 11th, 2016 through October 31st, 2016 - about 1 years' worth of data. The data also showed that Clinton's raw mean polling data was higher than Trump's: 43.91% vs. 39.43%, respectively. Finally, the median raw polling data for both Clinton and Trump were not very far from their means - signifying a lack of extreme outliers.

Now that we understood the data a little bit better, we decided to create a simple scatter plot of the rawpoll\_clinton and rawpoll\_trump variables over time (i.e. by enddate) to see if there were any obvious patterns. Below is a plot of Clinton's and Trump’s raw polling data:





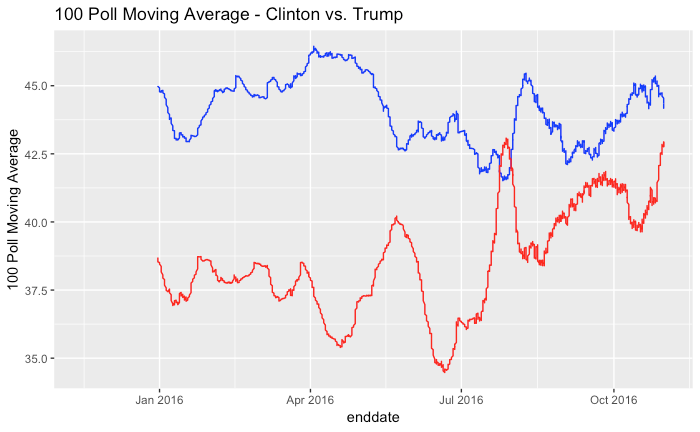
One obvious pattern that stands out is that the number of polls increases over time. This makes sense since as we get closer to election day, more and more polls are conducted in order to determine who the likely winner will be.

Another obvious pattern is that the variation in the polls also increases over time - like a funnel that widens over time.

Perhaps the most important point to make about the data is that it is quite "noisy" - even temporally close polls can drastically differ in predicted outcome. Smoothing out the data could be helpful in teasing out a more meaningful pattern.

One useful smoothing technique is a moving average - a rolling average of a predefined number of polls. A moving average can help us find the "signal" amongst the "noise".

Therefore, we applied a 100-poll moving average to Clinton's and Trump’s raw polling data and plotted them over time, providing us with the below visualization:



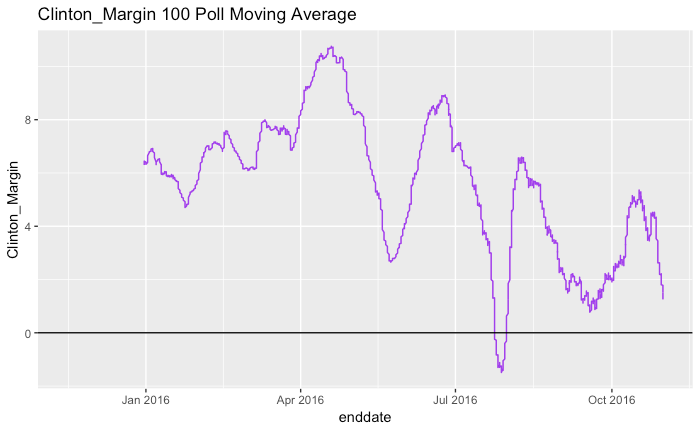
After smoothing out the data we can see some clear patterns. Early on in the cycle, it appears that Clinton had an early and substantial lead in the national polls (though this is tempered by the fact that there were fewer polls during this period).

However, this early and substantial lead appears to erode over the summer, improving in the fall, but never fully regaining those early highs. In contrast, Trump's polling data improved over time, starting around July, and spiking in the final weeks of the campaign.

**Margin of Victory**

Up to this point, the data had been separated into two columns, containing polling data for each candidate. Though this is useful, what really matters in elections is which candidate wins a higher percentage of the vote. In other words, what we want to know is the *margin* between Clinton and Trump.

Therefore, we created a new moving average variable by subtracting Trump's polling figures from Clinton's to arrive at a new **Clinton\_Margin** variable. A positive figure indicates that Clinton leads Trump in the national polls and a negative figure indicates that Clinton is trailing Trump in the national polls. Below is a visualization of this new margin variable:



Using this visualization of the margin we can see a clearer pattern of Clinton's chances of winning the presidency. We see an early rise in her chances as she widened her margin from January through May. However, her chances deteriorated throughout the Summer and Fall - reaching their lowest point in August (where it was actually net-negative for the first and only time during the campaign). Her margin then increased in mid-September through mid-October, only to precipitously fall back down in late October. If this late, falling trend were extended through election day (i.e. November 6th, 2016), it's entirely plausible that Clinton would have a net negative margin on election day.

**This is NOT a Popularity Contest**

In American presidential elections, the president is chosen *not* through the popular vote but through the Electoral College. In the Electoral College, each state receives a total number of electoral votes equal to its representation in Congress.

More specifically, each state's total number of electors is equal to the combined total of the state's membership in the Senate and House of Representatives; currently, there are 100 senators (2 for each state) and 435 representatives. Additionally, the Twenty-third Amendment of the U.S. Constitution dictates that the District of Columbia (which is *not* a state) is entitled to a number of electors no greater or less than that of the least populous state (i.e. 3 electoral votes), for a total of 538 electoral votes.

Therefore, to win, an American presidential candidate must win an absolute majority of 270 electoral votes to become president. Therefore, in order to create a model that accurately predicts who will become president, and *not* who will win the most votes, we need to predict the winner in each state and allocate each state's electoral votes to the predicted winner. The first candidate to 270 wins.

**Swing for the Fences**

Therefore, the most critical job of this presidential campaign is to determine *where* to campaign. Every campaign is constrained by its limited resources (volunteers, money, advertising, ect.), so we must make strategic decisions about *which* states to campaign in.

It does not make sense for the campaign to commit resources to a state that is highly unlikely to vote for Clinton, nor does it make sense for it to do the opposite - i.e. to commit resources to a state that is already highly likely to vote for Clinton. Therefore, the optimal use of our limited campaign resources is to campaign in "swing" states - states that could vote for *either* Clinton *or* Trump.

Toward this end, an unsupervised k-means clustering model would help us in determining which states are considered solidly Clinton, which states are considered solidly Trump, and - most importantly - which states are considered swing states.

K-means clustering attempts to divide data into *k* number of discrete groups and is highly effective at uncovering underlying data patterns.

The k-means clustering algorithm works by first splitting data into *k* number of clusters by initially randomly assigning a category or class to each data point.

The model then finds the *k* number of centroids - the points of central tendency (or mean value) for each category or class of data.

Next, it calculates each point’s Euclidean distance from each centroid, assigning points according to their *nearest* centroid.

Finally, it recalculates the centroids, repeating the whole process until the centroids no longer change and all the data points have been successfully classified.

To implement the k-means clustering model we first needed to create a State-specific data set which we called "State\_Polls\_Swing". This data set pulled in the same variables as the National polling data, only this time excluding national polls or district-specific polls[[1]](#footnote-1).

Next, we created a variable that enabled us to effectively categorize the states. Toward that end, we created another **Clinton\_Margin** variable - similar to our difference in moving averages margin - calculated by subtracting Trump's raw polling numbers from Clinton's.

Finally, we took care to only calculate our k-means clusters using *unique* polling data, so we modified our polling data to only include unique polls.

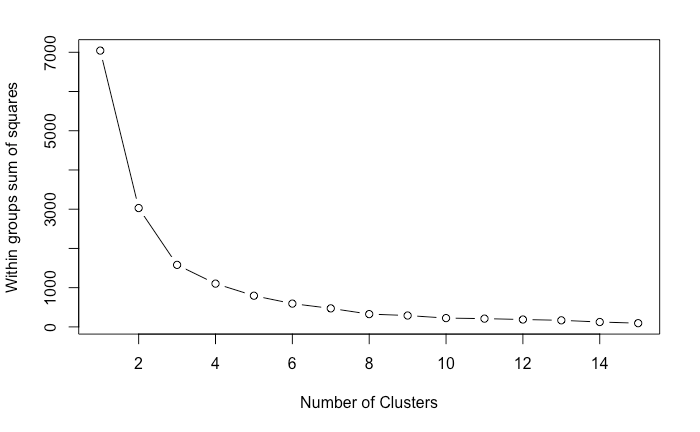
The end result was a data set containing 7,044 observations of five variables: **state**, **pollster**, **rawpoll\_clinton**, **rawpoll\_trump**, and our new **Clinton\_Margin** variable.

When it comes to selecting a *k* value for our k-means cluster model, we understood through pre-existing domain knowledge that states often fall into one of three categories: deep red or Republican states, deep blue or Democratic states, and swing state - states that will often swing back and forth between the two parties depending on the candidates. Therefore, our initial inclination is to set *k* equal to 3.

We can confirm or deny this inclination using a scree plot. A scree plot charts the degree of scattering or variance inside a cluster as the total number of clusters increases. A scree plot compares the Sum of Squared Errors (SSE) - measured as the sum of the squared distances between the centroids and the other neighbors inside the cluster. SSE tends to drop as more clusters are created.

In general, we should opt for a number of clusters where the SSE significantly decreases, but *before* it reaches a point of little to no change as the number of clusters increases (i.e. as we move right along the graph). If this dramatic drop (often referred to as an "elbow") occurs when the number of clusters is roughly equal to 3, we can be confident that our selection of *k* = 3 is legitimate.

**Scree Plot**



Sure enough, the elbow of our scree plot appears to reach its apex at 3 clusters. Therefore, the scree plot confirms our initial inclination that k should be set to 3. This will ensure that these three cluster solutions are distinct and will have a significant impact on our classification of the states.

Now that we have a *k* value and the new Clinton\_Margin variable we can calculate the Euclidean distance amongst Clinton's polling margins, set *k* equal to 3, and run the model.

# RESULTS

When we run the model we arrive at a cluster assignment for each each state’s polling margin. If we then take the mean of the cluster assignments and round the mean to the nearest whole number, a state can then be categorized as either a 1 (solidly Democratic), 2 (Swing), or 3 (solidly Republican).

Below is a map assigning each state to its model classification, with the color blue signifying solidly Democratic states, the color grey signifying Swing states, and the color red signifying solidly Republican states.

The model indicates that though Clinton holds what seems like a commanding lead (273 to 81), the substantial list of swing states indicates that this lead is tenuous at best. It is highly unlikely that every single Clinton leaning state will break for Clinton and all it takes is for one small Clinton leaning state to defect and for the swing states to break for Trump for Clinton to lose.

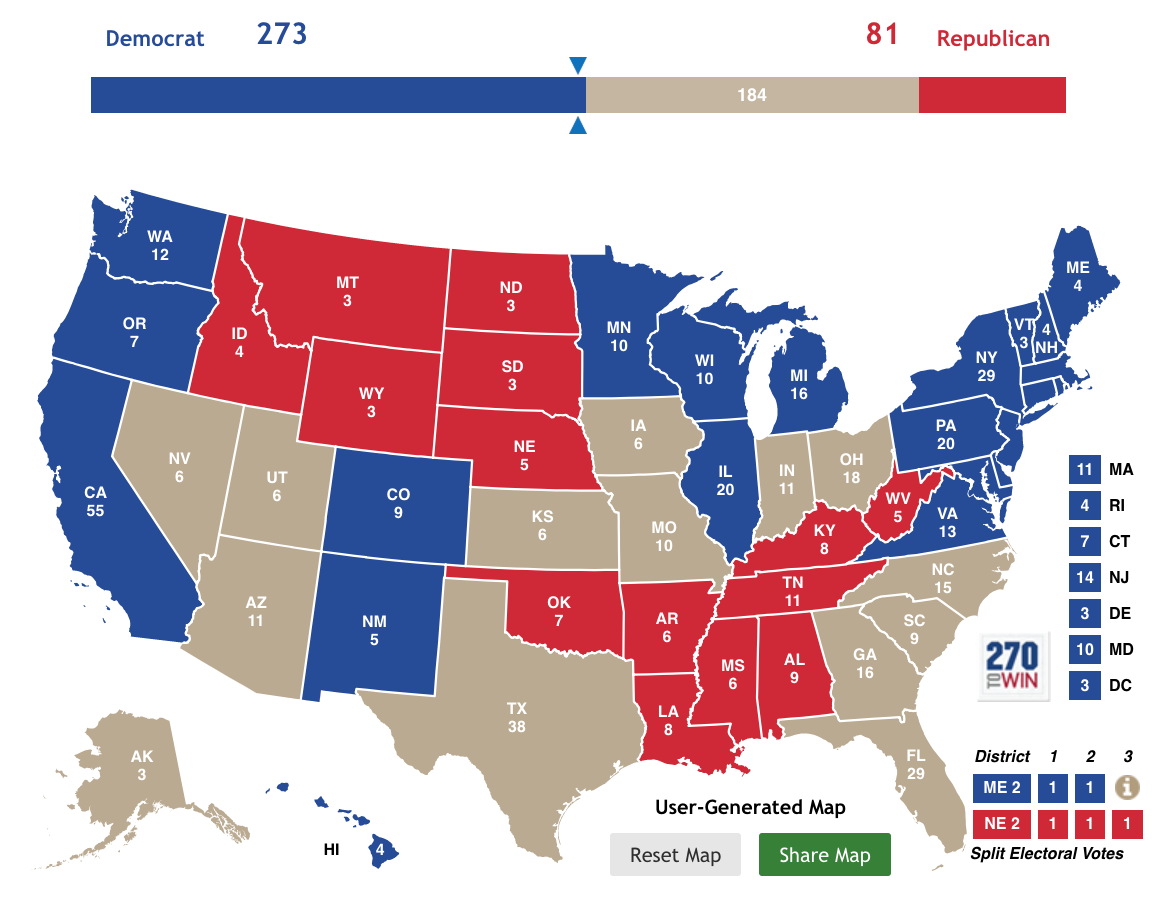
To better understand our swing states let's looks at their *unrounded* mean cluster group values to determine whether they are predominantly Clinton leaning or Trump leaning.

Let's look at these swing states in order of most Clinton leaning to least Clinton leaning (along with each state's electoral votes)[[2]](#footnote-2):

1. Florida (29)
2. Iowa (6)
3. Arizona (11)
4. Ohio (18)
5. North Carolina (15)
6. Nevada (6)
7. Georgia (16)\*
8. Kansas (6)\*
9. Alaska (3)\*
10. Missouri (10)\*
11. Utah (6)\*
12. South Carolina (9)\*
13. Indiana (11)\*
14. Texas (38)\*

\* Indicates that the state is Republican-leaning (i.e. has an unrounded mean cluster value greater than 2).

**Swing State Predictions**



If we eliminate the Republican-leaning states we are left with the following list of Democratic-leaning swing states - ordered by electoral votes:

1. **Florida (29)**
2. **Ohio (18)**
3. **North Carolina (15)**
4. **Arizona (11)**
5. Iowa (6)
6. Nevada (6)

This "hit list", represents the swing states that, according to the model, Clinton should campaign in order to improve her chances of reaching 270 electoral votes.

Thus, we have reached our first recommendation for the Clinton campaign: based on the polling data and our k-means clustering model, the Clinton campaign should focus on the swing states of Florida and Ohio, along with North Carolina and Arizona if time and resources permit.

It makes little sense for the Clinton campaign to waste time and resources on the relatively small number of electoral votes represented by the states of Iowa or Nevada, nor should the Clinton campaign commit resources to the Trump leaning swing states of Georgia, Kansas, Alaska, Missouri, Utah, South Carolina, Indiana, or Texas.

**But Who Will Win?**

What is elegantly simple about our k-means clustering model is that we can dial up or down the k-value depending on the question we want to answer. We have already answered the swing state question by choosing a k-value of 3, but if we want to answer our final question - i.e. who will win the election - we can modify our model by setting k equal to 2.

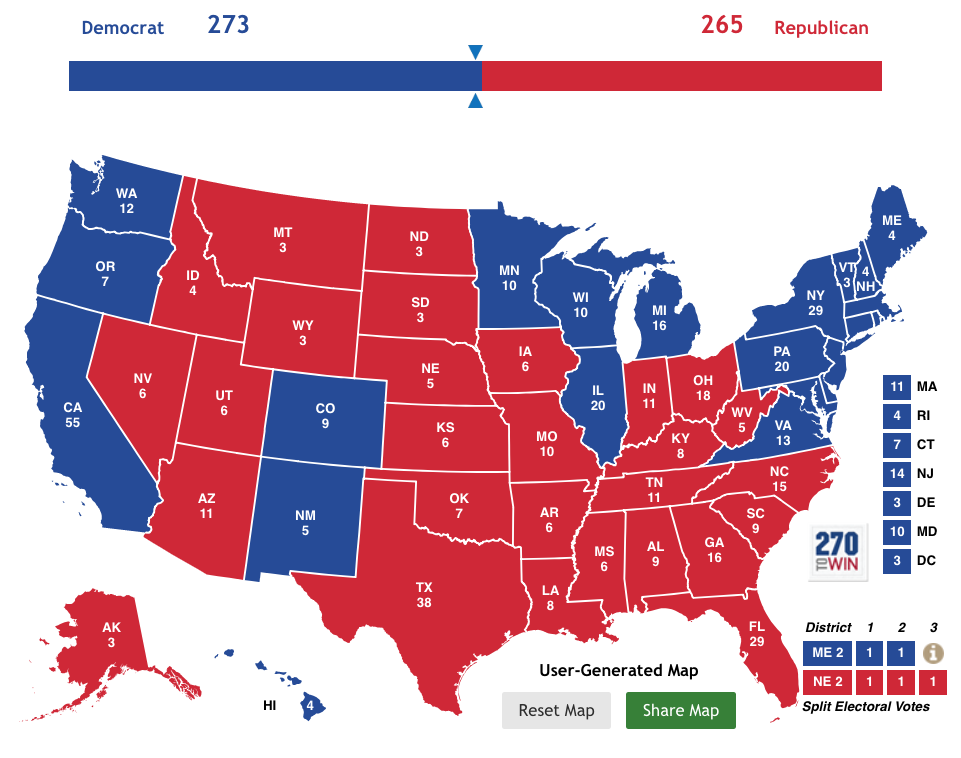
By choosing a k-value of 2, we split the decision into a binary choice: either Clinton wins a given state or she loses. By making this very simple modification we can rerun our model to determine a winner in each state and make a final prediction about who will win the election.

Though our scree plot indicated that the data is most accurately divided into 3 categories, the plot also shows that a k value of 2 is perhaps the second-best division of the data, so we can still be confident in our results.

Below is the map that was drawn by our modified model (again, the color blue represents Democratic states and the color red represents Republican states).

What is fascinating about the final results is that the list of solidly Clinton supporting states did *not* change. Though Clinton's coalition of states is stable, the fact that Trump ran the board on the swing states should be a cause for concern.

**Final Prediction**



# GRAND STRATEGY

The results of our model suggest to possible Grand Strategies for the campaign.

**Option 1: Defensive/Conservative Strategy**

Clinton should shore up her base in those states most likely to vote for her. This may not lead to an electoral blow-out, but it would be the strategy most likely to lead to victory.

If the campaign were to adopt this strategy it would be important to know which Democratic-leaning states are most likely to defect to Trump. The campaign should then devote resources to these “at risk” states in order to shore up Clinton's likelihood of reaching 270 electoral votes.

Let's take a look at these states’ unrounded mean cluster group scores in order of most at risk to least at risk:

1. Pennsylvania (20)\*
2. Michigan (16)\*
3. Colorado (9)\*
4. New Hampshire (4)\*
5. Maine (4)\*
6. Wisconsin (10)\*
7. New Mexico (5)
8. Virginia (13)
9. Delaware (3)
10. Oregon (7)
11. New Jersey (14)
12. Rhode Island (4)
13. Minnesota (10)
14. Washington (12)
15. Illinois (20)
16. California (55)
17. Connecticut (7)
18. District of Columbia (3)
19. Hawaii (4)
20. Maryland (10)
21. Massachusetts (11)
22. New York (29)
23. Vermont (3)

If we look only at those states with an unrounded mean cluster group score of 1.3 or higher (indicated by an \*), and order the list from greatest electoral votes to least, we get the following "at risk" list of states:

1. **Pennsylvania (20)**
2. **Michigan (16)**
3. **Wisconsin (10)**
4. Colorado (9)
5. New Hampshire (4)
6. Maine (4)

Total Electoral Votes: 63

If the campaign focuses on these states - particularly Pennsylvania, Michigan, and Wisconsin - Clinton will most likely have an electoral wall strong enough to defend against any encroachment by the Trump campaign. Though Clinton would win by a narrow electoral vote margin (only 8 electoral votes) her chances of winning would be more assured.

**Option 2: Offensive/Risky Strategy**

Under this alternative strategy, Clinton should make a play for our previously determined “hit list” of swing states, dedicating resources to those states with the greatest number of electoral votes. Though Clinton may lose some blue states the result may be a larger overall electoral vote total.

This strategy seeks to maximize electoral votes at greater electoral risk. Making a push into large swing states could result in a larger electoral vote margin, and may therefore give Clinton greater political capital once in office. In other words, the goal is not just to win but to win *big*.

If the senior campaign team chooses this strategy it makes the most sense to campaign in those swing states with the largest number of electoral votes. Once again, these states include:

1. Florida (29)
2. Ohio (18)
3. North Carolina (15)
4. Arizona (11)
5. Iowa (6)
6. Nevada (6)

Total Electoral Votes: 85

If Clinton were to lose all of her "at risk" states but gained all of these swing states, the result would be a net gain of 22 electoral votes, bringing her total up to 295. Perhaps this would change the perceptions of Congress, and provide her with additional political capital once in office.

However, if our goal is to attain the requisite number of electoral votes with the highest probability of success, then the only strategy that does both is the *conservative* strategy.

Therefore, it is the Data Analytics team’s recommendation that the campaign devotes resources to shoring up Clinton’s "at risk" states and to adopt a more defensive strategy.

# RESULTS & ACCURACY

Let's examine the results of our swing state analysis. Let us define a swing state as a state that had a +/-5% Clinton margin. Below is a list of these states, as well as Clinton's margin in each state:

1. Michigan (-0.22%)
2. New Hampshire (+0.37%)
3. Pennsylvania (-0.72%)
4. Wisconsin (-0.76%)
5. Florida (-1.20%)\*
6. Minnesota (+1.52%)
7. Nevada (+2.42%)\*
8. Maine (+2.96%)
9. Arizona (-3.55%)\*
10. North Carolina (-3.66%)\*
11. Colorado (+4.91%)

\*Indicates that the state was on the original list of swing states.

Next, let's examine the accuracy of our model when it came to categorizing solidly Clinton states. Below is a list of states that ended up being solid Clinton wins, which we will define as a Clinton margin greater than +5%:

1. District of Columbia (+86.78%)\*
2. Hawaii (+32.18%)\*
3. California (+30.11%)\*
4. Massachusetts (+27.20%)\*
5. Maryland (+26.42%)\*
6. Vermont (+26.41%)\*
7. New York (+22.49%)\*
8. Illinois (+17.06%)\*
9. Washington (+15.71%)\*
10. Rhode Island (+15.51%)\*
11. New Jersey (+14.10%)\*
12. Connecticut (+13.64%)\*
13. Delaware (+11.37%)\*
14. Oregon (+10.98%)\*
15. New Mexico (+8.21%)\*
16. Virginia (+5.32%)\*

\*Indicates that the state was on the original list of solidly Clinton states.

Finally, let's examine the accuracy of our model when it comes to categorizing solidly Trump states. Below is a list of states that ended up being solid Trump wins, which we will define as a Clinton margin less than -5%:

1. Wyoming (-46.30%)\*
2. West Virginia (-42.07%)\*
3. Oklahoma (-37.08%)\*
4. North Dakota (-35.73%)\*
5. Idaho (-31.77%)\*
6. Kentucky (-29.84%)\*
7. South Dakota (-29.79%)\*
8. Alabama (-27.73%)\*
9. Arkansas (-26.92%)\*
10. Tennessee (-26.01%)\*
11. Nebraska (-25.05%)\*
12. Kansas (-20.60%)
13. Montana (-20.42%)\*
14. Louisiana (-19.64%)\*
15. Indiana (-19.17%)
16. Missouri (-18.64%)
17. Utah (-18.08%)
18. Mississippi (-17.83%)\*
19. Alaska (-14.73%)
20. South Carolina (-14.27%)
21. Iowa (-9.41%)
22. Texas (-8.99%)
23. Ohio (-8.13%)
24. Georgia (-5.13%)

\*Indicates that the state was on the original list of solidly Trump states.

In terms of swing states, we called 4 states out of 11 (a ~36% success rate); in terms of solidly Clinton states, we called 17 out of 17 (a 100% success rate); and in terms of solidly Trump states, we called 14 out of 24 (a ~58% success rate). Overall, our model accurately categorized ~67% of the states.

If we had randomly assigned a 1/3 chance of assigning a state either as solidly Clinton, solidly Trump, or swing, then we would have expected a success rate of 1/3 or ~33%. Clearly, our model was more accurate than this baseline.

Next, let's turn to our final prediction. In the end, we accurately predicted the winner of every state except 4 - Michigan, Nevada, Pennsylvania, and Wisconsin - a ~92% accuracy rate. If we compare this a random coin toss, we handily beat this baseline assumption of 50%.

Additionally, if we compare the success rate of this model versus FiveThirtyEight’s “polls-only” forecast, our model wins again: FiveThirtyEight’s forecast correctly called every state except 5.

But note: of those four states we got wrong, three were on our "at risk" list of states: Pennsylvania, Michigan, and Wisconsin. As part of our conservative strategy recommendation, we urged the campaign to dedicate resources to these three states in order to shore up support and to prevent defection - especially given Clinton’s very narrow and extremely tenuous lead in the electoral college.

# AREAS FOR IMPROVEMENT/FURTHER EXPLORATION

One major limitation to the data is that the data is exclusive to 2016, meaning that we did not have previous election year data to work with. This limitation became abundantly clear early on as we attempted to create a logistic regression formula to predict the winner of each state based on Clinton and Trump’s raw polling data.

Not having previous election year data made it impossible to split the data into the requisite “training” set and “testing” set. It would have been helpful to use 2012’s election data, for example, to train the logistic regression on and then to apply it to the test set of 2016 data.

Another limitation of the data was its format. Though the data is tidy - in the sense that each variable is represented by a column in the data set, with each row being an observation - the data would not have allowed us to, for instance, test pollsters to determine which are the best at determining the winner. Again, this would have required previous election year data, but even if we *had* acquired 2012 data it would have been quite difficult and tedious to modify the format of the data to allow us to test the “pollster” variable - or many of the other variables for that matter.

Some areas for further exploration include the following:

* Testing FiveThirtyEight’s pollster grading system by incorporating the **adjpoll\_clinton** and **adjpoll\_trump** fields into a modified version of the model. If the model’s accuracy improves relative to the original model then the adjustments introduced by FiveThirtyEight may prove to be useful.
* It seems clear that as we approach the actual election date, the accuracy of polls must improve as undecided voters begin making final decisions about who to vote for. Perhaps weighting more recent polls greater than older polls would improve the accuracy of the model.
* Polls with a greater number of respondents are likely to be more accurate than polls with a relatively small number of respondents. Perhaps weighting polling data by respondent size (along with recency) could also improve the accuracy of the model.

1. Some states allocate electoral votes by congressional district instead of in a "winner-take-all" fashion. Currently, only two states do so - Maine and Nebraska - but for the sake of simplicity, we will assume that as the state goes, so go the congressional districts. [↑](#footnote-ref-1)
2. It is important to note that when we say “Clinton Leaning” we do *not* mean that, if the election were held today, Clinton would be the most likely winner of that state. The unrounded mean cluster values simply provide a *relative* measure rather than an *absolute* prediction. For example, though Florida is the most Clinton leaning of the swing states, this does *not* mean that Clinton is the most predicted winner of Florida. The model is simply showing that *of the swing states*, Florida is the most Clinton leaning. [↑](#footnote-ref-2)