

# 2016 Presidential Elections Based on Party Platforms

Tidewater Big Data Enthusiasts

Chuck Cartledge

Developer

April 23, 2020 at 12:36 Noon

## Contents

<b>List of Figures</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Analysis</b>	<b>1</b>
2.1 Assumptions . . . . .	1
2.2 Analysis of party platforms . . . . .	1
2.3 Analysis of electoral college . . . . .	3
2.4 Contractions . . . . .	3
<b>3 Conclusion</b>	<b>10</b>
<b>4 References</b>	<b>11</b>
<b>A Misc. files</b>	<b>12</b>

## List of Figures

1	Relative sentiments of party platforms. . . . .	2
2	Democratic platform word cloud. . . . .	4
3	Libertarian platform word cloud. . . . .	5
4	Republican platform word cloud. . . . .	6
5	Platform vocabulary by grade. . . . .	7
6	Predicted electoral college based on different voter participation percentages.	7

# 1 Introduction

The 2016 Presidential election cycle has been long and acrimonious. Some have characterized them as being the most divisive in recent memory. Innumerable different predictions have been made about the outcome, by many different “authoritative” sources. This report is another prediction, not by an authoritative source, but by a person interested in all aspects of Big Data, looking at different sources of data and how the pieces can fit together.

We will bring together the platforms for the three major parties: Democratic, Libertarian, and Republican<sup>1</sup>, execute grade level and sentiment analysis on the platform, and based on those analyses, distribute undecided voters.

## 2 Analysis

### 2.1 Assumptions

We will make a simple prediction based on the following ideas:

1. Each party has a “core” constituency that Will vote the “party” ticket. We have assumed that this is 65% of the registered party members.
2. Each party’s platform contains a collection of positive, negative, and neutral sentiments. These sentiments will sway undecided voters.
3. For those states where party registration is not readily available, an average based on the reporting states will be used.
4. All 50 states and the District of Columbia will be considered for the electoral college.
5. The percentage of the US population over 18 that will vote is unknown, so various percentages will be considered.

Source files used in this analysis are listed in the appendix (see Section A on page 12).

### 2.2 Analysis of party platforms

The text from each party’s platform was analyzed in two different ways and presented in three different manners. First each platform was analyzed for sentiment by computing the percentage of positive and negative words in the platform. These percentages were then used as coordinates and plotted on an X-Y plot (see Figure 1 on the next page). Each platform had approximately the same percentage of negative words, but the positive percentages were different. These positive sentiments were used to decide how undecided voters would vote.

---

<sup>1</sup>The Green and Constitution parties were considered, no included because their apparent representation was too low to be considered significant.

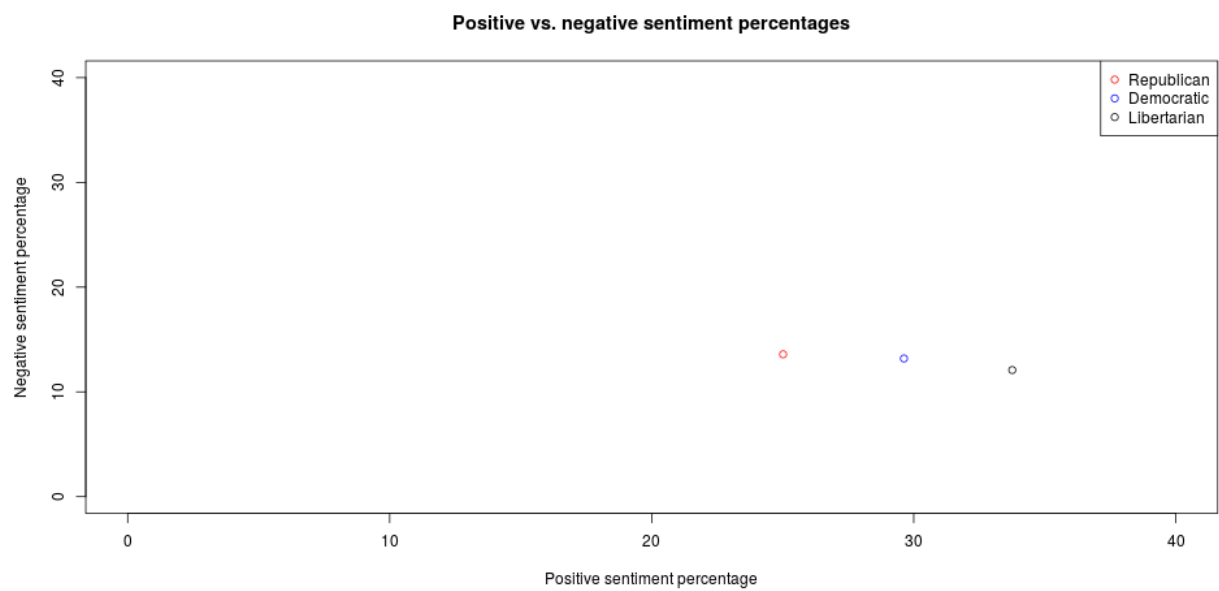


Figure 1: Relative sentiments of party platforms. Each important word in platform was compared to lists of positive and negative words. The number of positive words compared to the total number of platform words was used to compute the positive percentage. Similar processing was used for negative words.

Plotting sentiment analysis is only one way of analyzing the platforms. Another way is using color coded word clouds. A word cloud is a visual representation of the words in a corpus where the size of the text reflects how often the word is used. A word cloud was created for each of the platforms, where the color of the word reflects whether the word is positive (in green), negative (in red), or neutral (in black) (see Figures 2 on the following page through 4 on page 6).

The vocabulary used in each platform is something that is talked about in the media from time to time. Each platform's vocabulary was measured against a set of words that should be understood by a student completing that grade. The grades went from Kindergarten, to post graduate college (see Figure 5 on page 7). Contrary to what different pundits have said about how the opposing party is "writing down" to their constituents, the plots for all platforms are very similar.

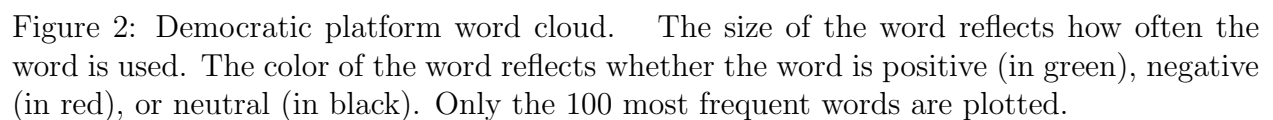
## 2.3 Analysis of electoral college

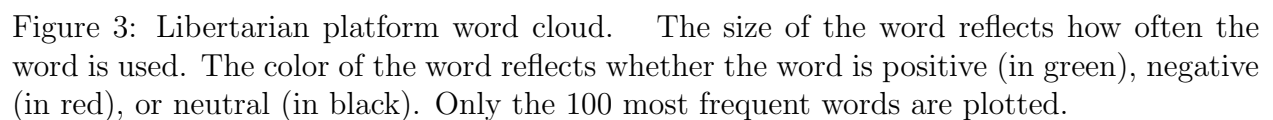
The purpose of analyzing the various party platforms was to predict (using simple assumptions), which party would win the Presidential Election. We made a number of simplifying assumptions (see Section 2.1 on page 1) in order to be able to make a prediction based on readily available data. Along with the things that are fixed, is the realization that the percentage of participating voters is unknown. Therefore we will vary the participating percentage and see how what the range of outcomes are (see Figure 6 on page 7). There are a number of interesting things happening in the plot:

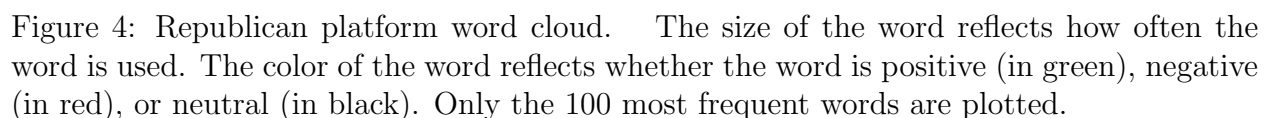
1. If the participation by eligible voters remains low, then the Democrats will accumulate the number of required electoral votes.
2. If the participation by eligible voters is extremely high, then third parties (like the Libertarians) will have a significant voice in the outcome.

## 2.4 Contractions

One of the things that "popped" out of the exploration is how words were changed based on removing punctuation marks. The analysis program uses the R `tm` library function `removePunctuation()` to remove punctuation marks. This function does exactly what its name says in that all punctuation marks are removed from the character vector. This can result in changing contracted words entirely (see Table 1 on page 8)[2].







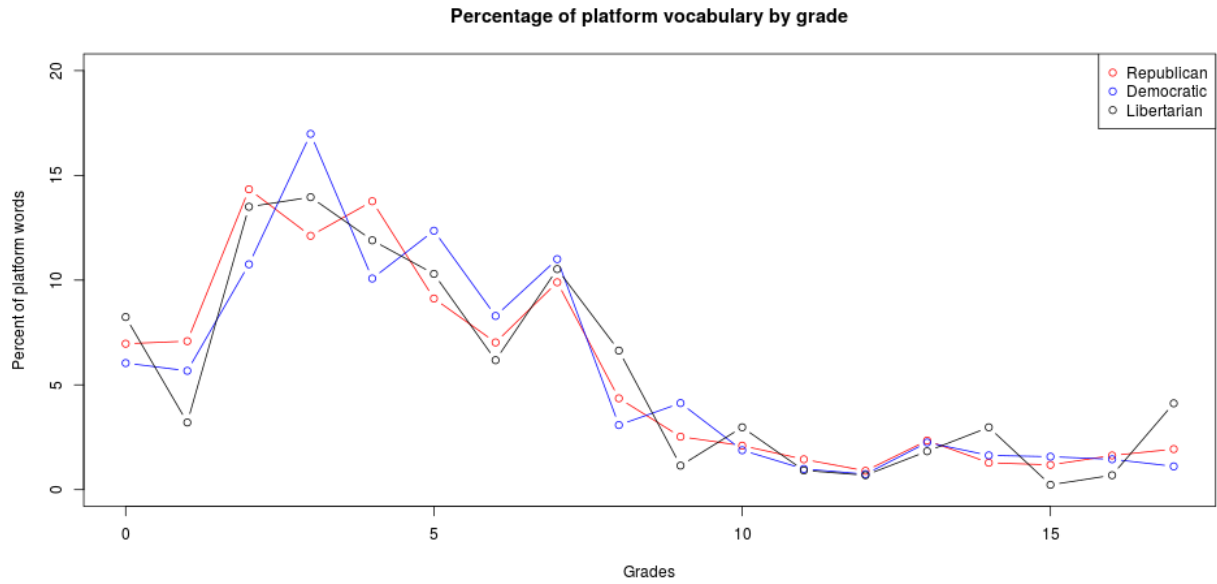


Figure 5: Platform vocabulary by grade.

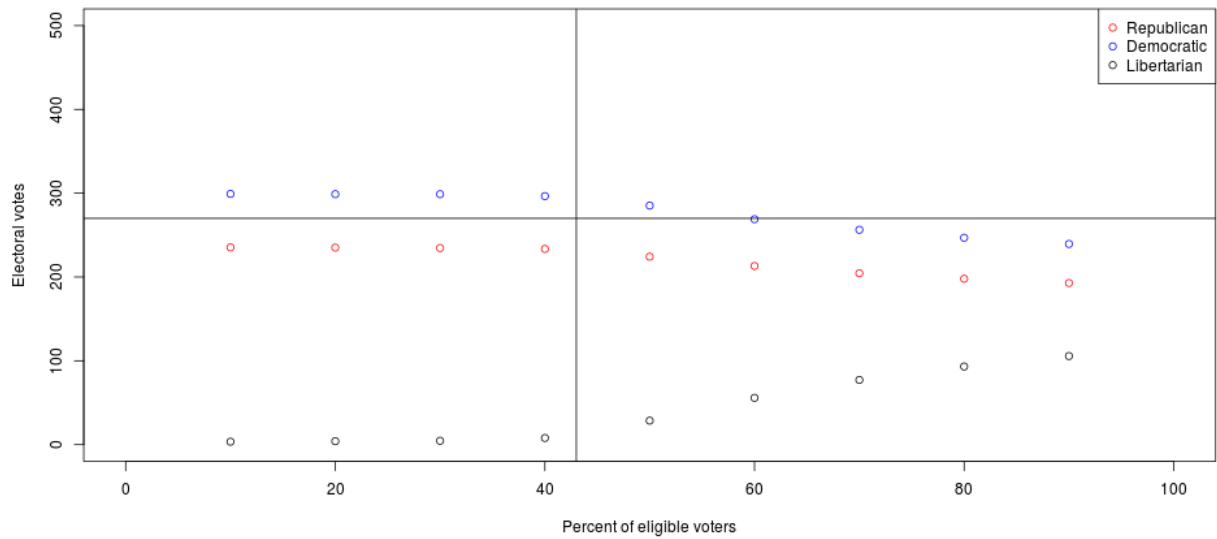


Figure 6: Predicted electoral college based on different voter participation percentages. The horizontal line marks the 270 electoral college votes needed to win the presidential election. The vertical line at 43% represents the historical percentage of voters from past elections.



Table 1: Common English language contractions. Various common contractions are given, along with their expanded meaning. Where the contraction with the apostrophe removed results in a properly spelled word, it is indicated.

Contraction	Meaning	Result	
aren't	are not	arent	
can't	cannot	cant	<b>Different word</b>
couldn't	could not	couldnt	
didn't	did not	didnt	
doesn't	does not	doesnt	
don't	do not	dont	
hadn't	had not	hadnt	
hasn't	has not	hasnt	
haven't	have not	havent	
he'd	he had; he would	hed	
he'll	he will; he shall	hell	<b>Different word</b>
he's	he is; he has	hes	<b>Different word</b>
I'd	I had; I would	Id	<b>Different word</b>
I'll	I will; I shall	Ill	<b>Different word</b>
I'm	I am	Im	
I've	I have	Ive	
isn't	is not	isnt	
it's	it is; it has	its	<b>Different word</b>
let's	let us	lets	<b>Different word</b>
mightn't	might not	mightnt	
mustn't	must not	mustnt	
shan't	shall not	shant	
she'd	she had; she would	shed	<b>Different word</b>
she'll	she will; she shall	shell	<b>Different word</b>
she's	she is; she has	shes	<b>Different word</b>
shouldn't	should not	shouldnt	
that's	that is; that has	thats	<b>Different word</b>
there's	there is; there has	theres	

(Continued on the next page.)

Table 1. (Continued from the previous page.)

<b>Contraction</b>	<b>Meaning</b>	<b>Result</b>	
they'd	they had; they would	theyd	
they'll	they will; they shall	theyll	
they're	they are	theyre	
they've	they have	theyve	
we'd	we had; we would	wed	<b>Different word</b>
we're	we are	were	<b>Different word</b>
we've	we have	weve	
weren't	were not	werent	
what'll	what will; what shall	whatll	
what're	what are	whatre	
what's	what is; what has; what does	whats	<b>Different word</b>
what've	what have	whatve	
where's	where is; where has	wheres	<b>Different word</b>
who'd	who had; who would	whod	
who'll	who will; who shall	wholl	
who's	who is; who has	whos	<b>Different word</b>
who've	who have	whove	
won't	will not	wont	<b>Different word</b>
wouldn't	would not	wouldnt	
you'd	you had; you would	youd	
you'll	you will; you shall	youll	
you're	you are	youre	
you've	you have	youve	

(Last page.)

### 3 Conclusion












We made a simple model to predict the outcome of 2016 Presidential Election using sentiment analysis on the major the platforms published by the Democratic, Libertarian, and Republican parties, and assumptions on how “non-core” party members would vote. Based on the model, and assuming various percentages for eligible voter turn out, the model says that the Democratic party will likely win the election. If voter turn out is higher than expected, then the Libertarian party may have major influence in the final outcome. The model is extremely simple, and does not include real-time events (tweets, Congressional investigations, changes in party strategies, etc.). The model shows what can be done using Big Data techniques on freely available data.

## 4 References

- [1] Mark Blumenthal and Ariel Edwards-Levy, *Huffpollster: A state-by-state guide to party registration*, [http://www.huffingtonpost.com/2014/05/27/state-party-registration\\_n\\_5399977.html](http://www.huffingtonpost.com/2014/05/27/state-party-registration_n_5399977.html), 2014.
- [2] Richard Nordquist, *A list of contractions commonly used in informal english*, <http://grammar.about.com/od/words/a/EnglishContractions.htm>, 2016.
- [3] Democratic Party, *2016 democratic party platform*, <https://www.demconvention.com/wp-content/uploads/2016/07/Democratic-Party-Platform-7.21.16-no-lines.pdf>, 2016.
- [4] Libertarian Party, *Libertarian party platform*, <http://www.lp.org/files/2016LPPlatform.pdf>, 2016.
- [5] Republican Party, *Republican platform 2016*, [https://prod-staticngop-pbl.s3.amazonaws.com/media/documents/DRAFT\\_12\\_FINAL\[1\]-ben\\_1468872234.pdf](https://prod-staticngop-pbl.s3.amazonaws.com/media/documents/DRAFT_12_FINAL[1]-ben_1468872234.pdf), 2016.
- [6] BUSD Staff, *Busd grade level academic vocabulary*, [http://www.berkeleyschools.net/wp-content/uploads/2013/05/BUSD\\_Academic\\_Vocabulary.pdf](http://www.berkeleyschools.net/wp-content/uploads/2013/05/BUSD_Academic_Vocabulary.pdf), 2016.
- [7] Census Staff, *Annual estimates of the resident population: April 1, 2010 to july 1, 2015 more information 2015 population estimates*, [http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP\\_2015\\_PEPANNRES&src=pt](http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2015_PEPANNRES&src=pt), 2016.
- [8] ———, *State totals: Vintage 2015*, <https://www.census.gov/popest/data/state/totals/2015/index.html>, 2016.
- [9] MajorTests Staff, *Sat word lists*, <http://www.majortests.com/sat/wordlist.php>, 2016.

## A Misc. files

The files used to create all these figures are attached to this report. They are:

1. analyzePlatforms.R - the R program used to analyze the platforms and create the figures.
2. BUSD\_Academic\_Vocabulary.pdf - a collection of grade specific words[6].
3. ElectoralPopArea.xls - an Excel spreadsheet with electoral votes and population.
4. lexicon.csv - a file with a collection of words categorized as positive or negative.
5. library.R - a support library for the analyzePlatforms.R program.
6. NST-EST2015-01.xlsx - an Excel spreadsheet with estimated population[8].
7. PEP\_2015\_PEPAGESEX.xls - an Excel spreadsheet with estimated population by age[7].
8. platform-Democratic.pdf - the Democratic party platform[3].
9. platform-Libertarian.pdf - the Libertarian party platform[4].
10. platform-Republican.pdf - the Republican party platform[5].
11. satWords.txt - a collection of SAT words[9].
12. State Total Party Registration Data.xlsx - an Excel spreadsheet with party registration by state[1].
13. stopwords.txt - a collection of US English stop words.