

In this report, we show the presidential election data from Moodwire showed that Donald Trump will be the next president even months ago.

Let's first download the summary data using Moodwire's Summary API. We will use the UUIDs for both **Donald Trump** (5516321c9b724739f3649835) and **Hillary Clinton** (55166a299b724739f364f02c). We are using *rjson* library to download and parse data from Moodwire API.

Next, we convert the results into a R data frame. Let's look at the head and tail of the data frame.

ID	Name	type	Buzz	Mood	Positives	Negatives	Neutrals	D
5516321c9b724739f3649835	Donald Trump	person	213788	-18457	23154	34485	156149	2008
55166a299b724739f364f02c	Hillary Clinton	person	71981	-4868	9355	12789	49837	2008
5516321c9b724739f3649835	Donald Trump	person	169995	-8237	21347	28015	120633	2008
55166a299b724739f364f02c	Hillary Clinton	person	79232	-4752	9264	12936	57032	2008
5516321c9b724739f3649835	Donald Trump	person	180916	-14780	17399	28096	135421	2008
55166a299b724739f364f02c	Hillary Clinton	person	86084	-8937	8167	14735	63182	2008

	ID	Name	type	Buzz	Mood	Positives	Negatives	Neutra
95	5516321c9b724739f3649835	Donald Trump	person	156475	-6223	16441	21454	118580
96	55166a299b724739f364f02c	Hillary Clinton	person	106515	-7867	10215	16137	80163
97	5516321c9b724739f3649835	Donald Trump	person	176369	-10283	16684	24462	135220
98	55166a299b724739f364f02c	Hillary Clinton	person	101077	-6470	10091	15304	75682
99	5516321c9b724739f3649835	Donald Trump	person	141282	-6212	14439	19290	107550
		Hillary						

100	55166a299b724739f364f02c	Clinton	person	78818	-7917	7467	13588	57763
-----	--------------------------	---------	--------	-------	-------	------	-------	-------

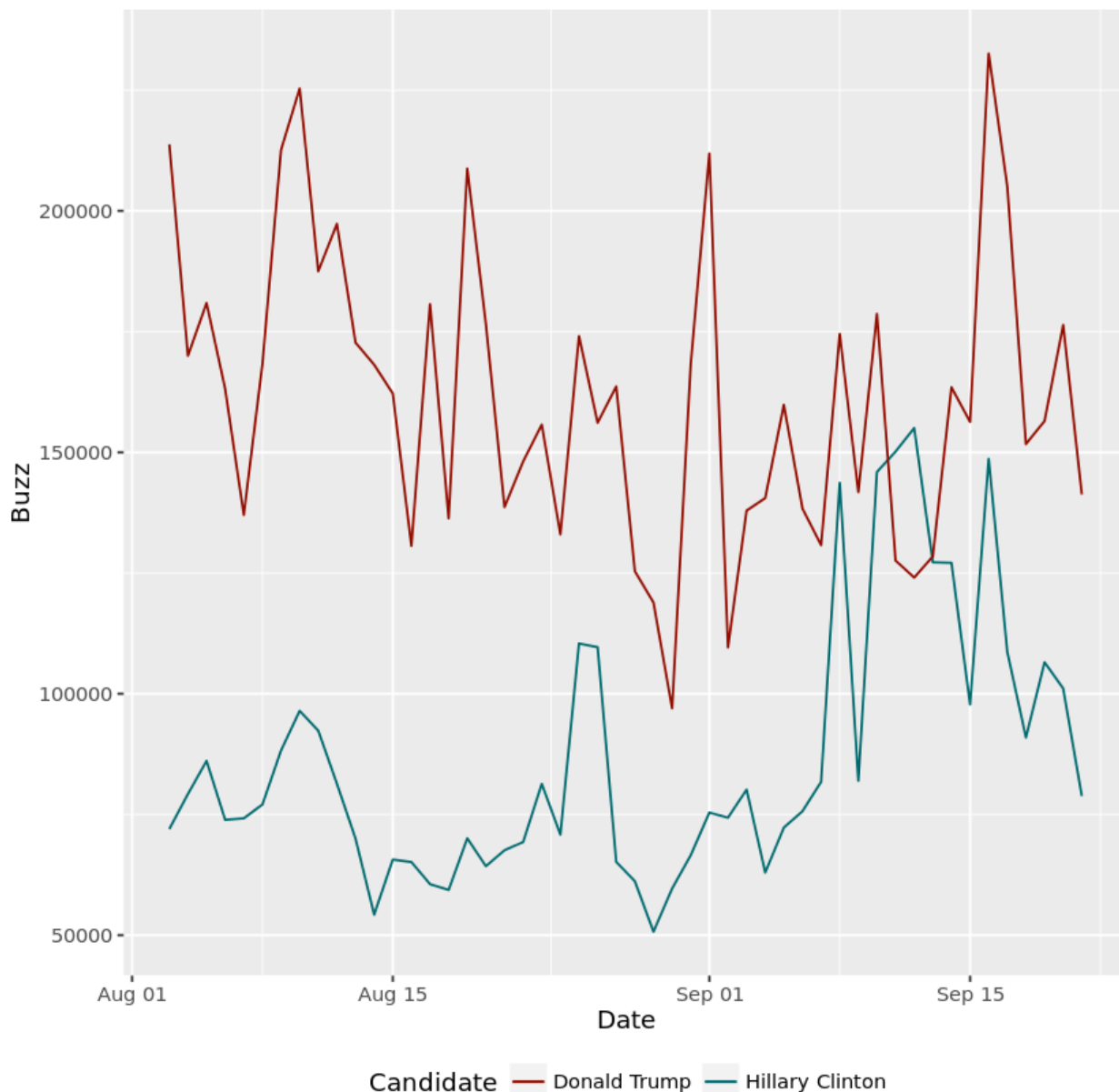
Next, we split this dataframe into two different dataframes, one for Hillary Clinton and one for Donald Trump. Let's look at if the data is split as expected. We will look at the head of the dataframe for Donald Trump.

	ID	Name	type	Buzz	Mood	Positives	Negatives	Neutrals
1	5516321c9b724739f3649835	Donald Trump	person	213788	-18457	23154	34485	156149
3	5516321c9b724739f3649835	Donald Trump	person	169995	-8237	21347	28015	120633
5	5516321c9b724739f3649835	Donald Trump	person	180916	-14780	17399	28096	135421
7	5516321c9b724739f3649835	Donald Trump	person	163111	-9611	19558	25820	117733
9	5516321c9b724739f3649835	Donald Trump	person	137028	-7691	16295	21841	98892
11	5516321c9b724739f3649835	Donald Trump	person	168309	5818	26326	22734	119249

Data looks good. Now we can start our analysis. We will look at the daily buzz and sentiment distribution for both candidates and compare. We are going to use ggplot library for our charts.

Buzz Comparison

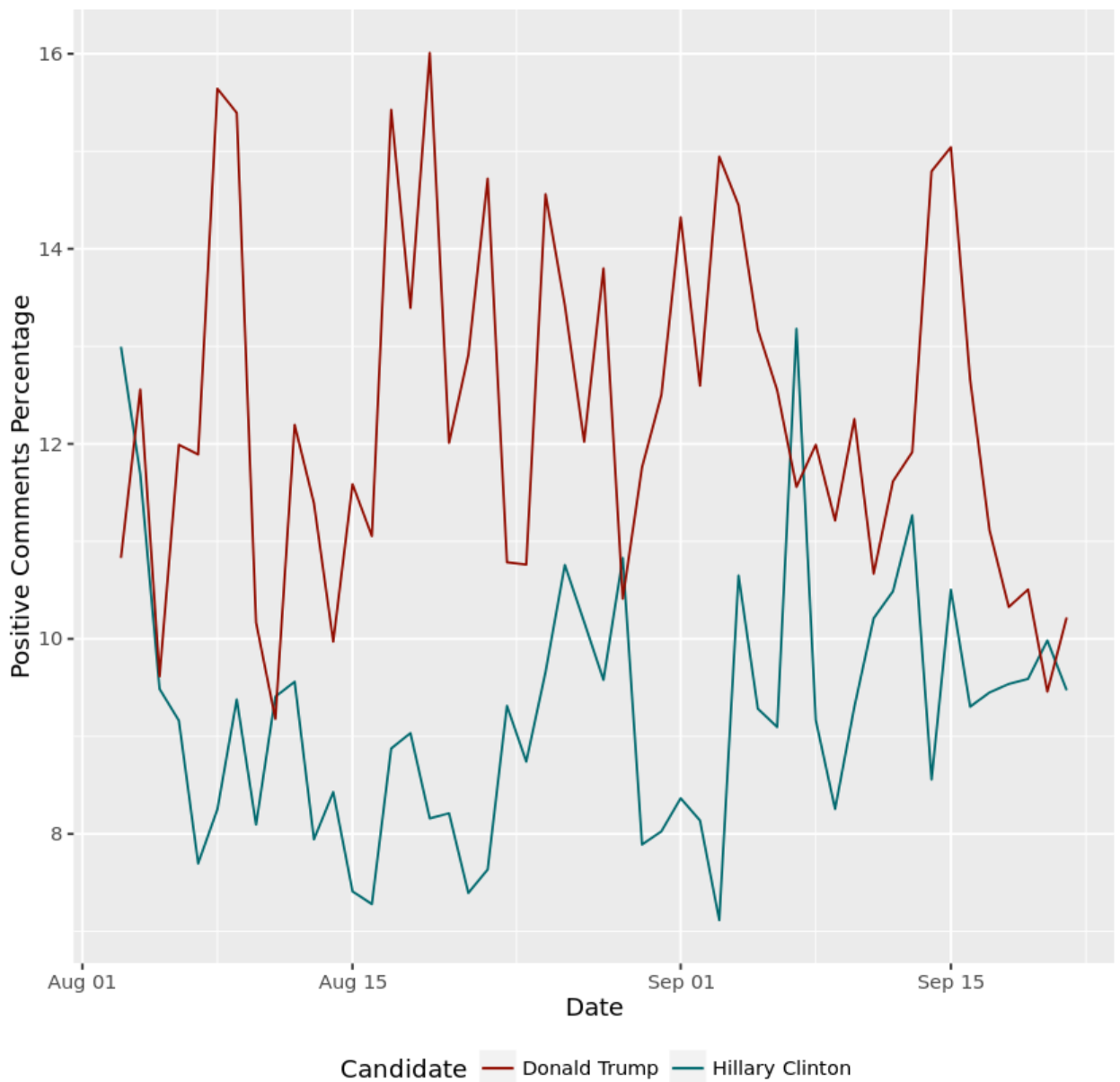
Throughout the campaign Donald Trump was the more talked about candidate. If his daily buzz goes below a certain threshold; he would say something controversial and his buzz would jump again.



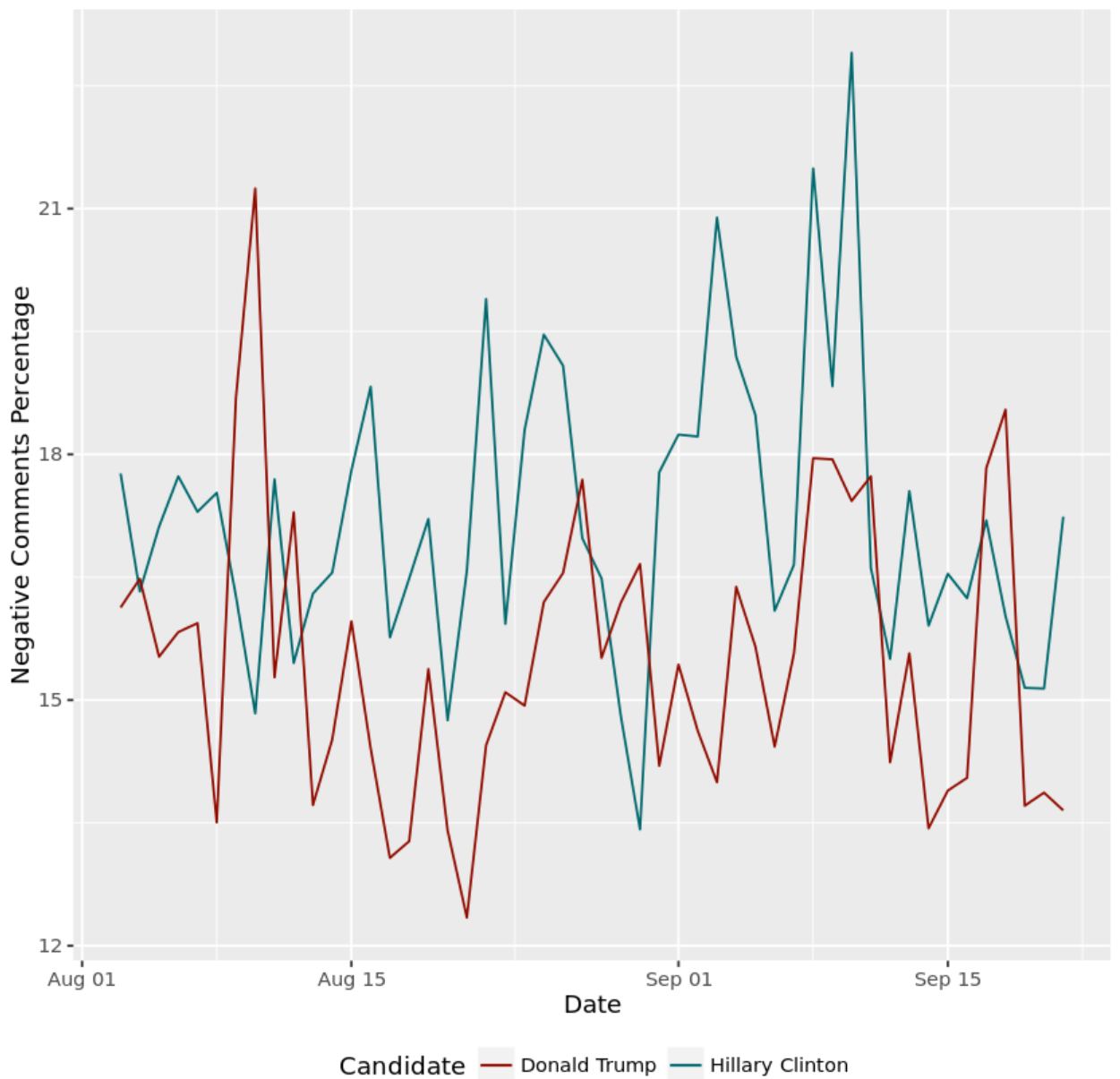
Sentiment Comparison

As the above graph shows Donald Trump was more popular than Hillary Clinton. Let's see if this popularity also meant he was liked more than Hillary Clinton. In order to do that, we will compare the sentiment of candidates. Let's first look at the positive sentiment percentage distribution. Positive sentiment percentage is computed as the ratio of the number of positive posts to the number of total posts.

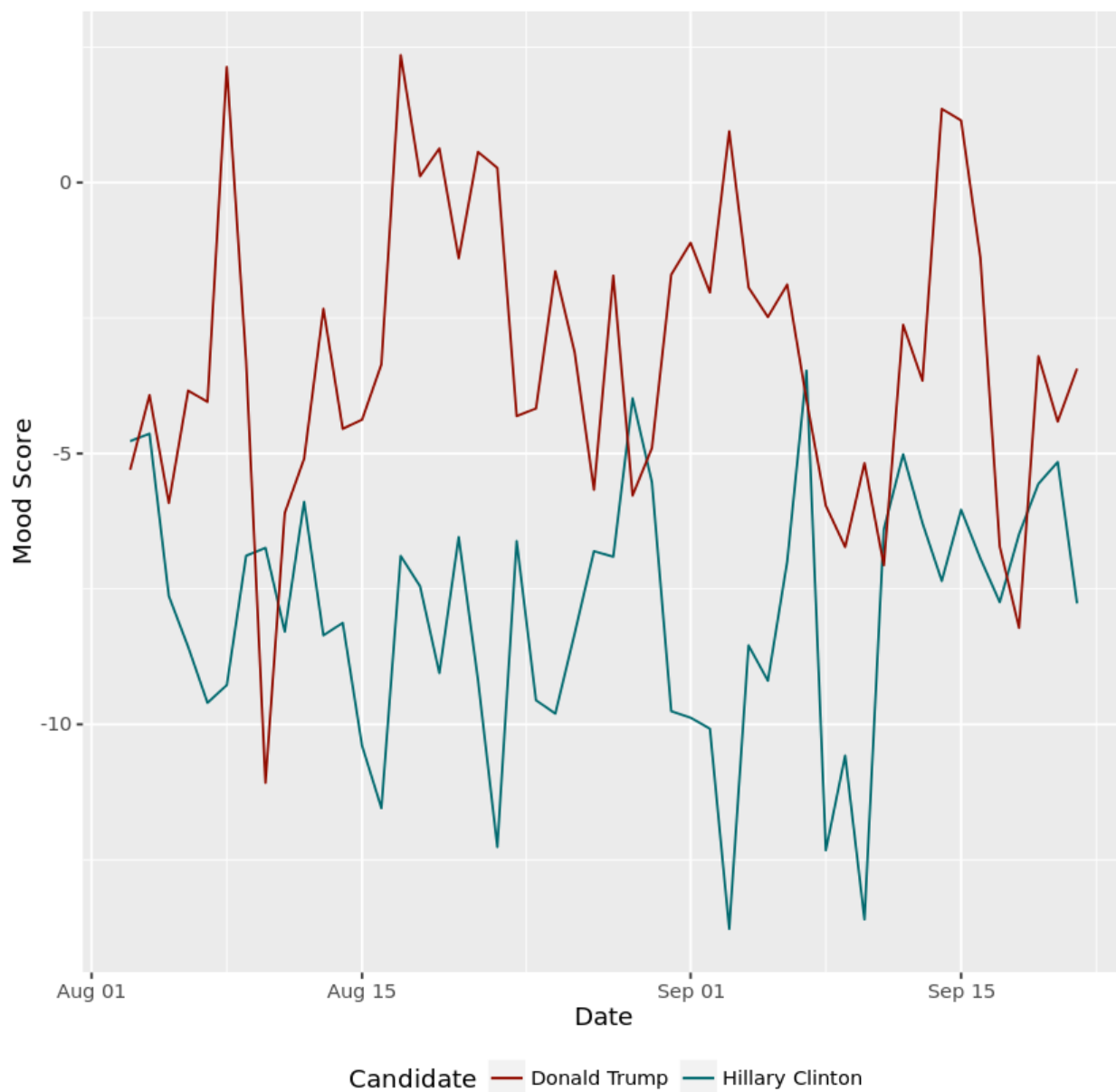
Positive Sentiment Percentage comparison, Donald Trump sentiment on social media is more positive than the Hillary Clinton Sentiment. However, notice that the positive sentiment for both candidates is very low. Maximum positive sentiment percentage was 16% of all the posts



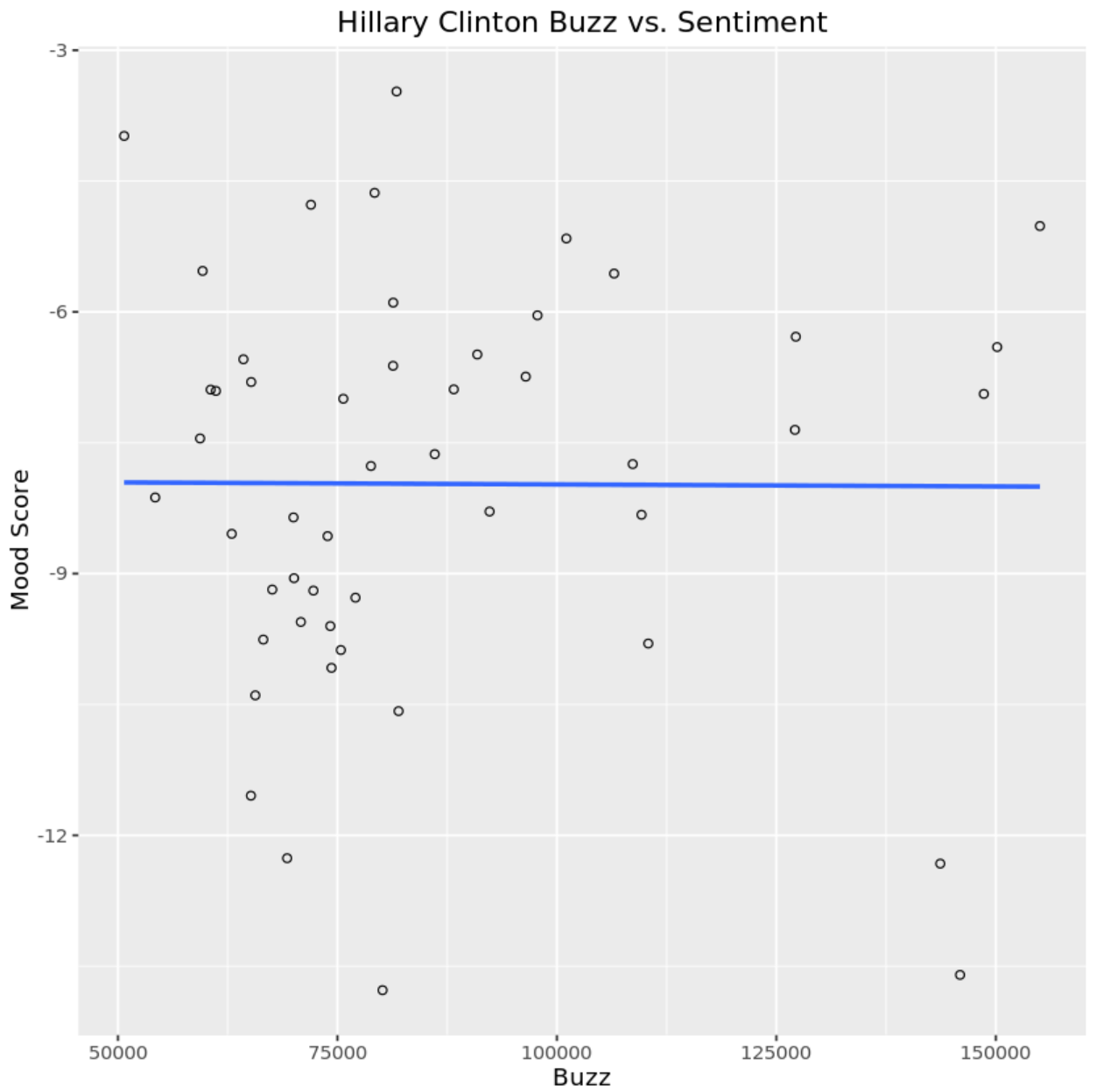
Another way to look at data, is to compare the candidates' negative sentiment. As the below graph shows Sentiment on Hillary Clinton was more negative than Donald Trump. Negative sentiment comments were more than positive sentiment comments, and Hillary Clinton comments were more negative.



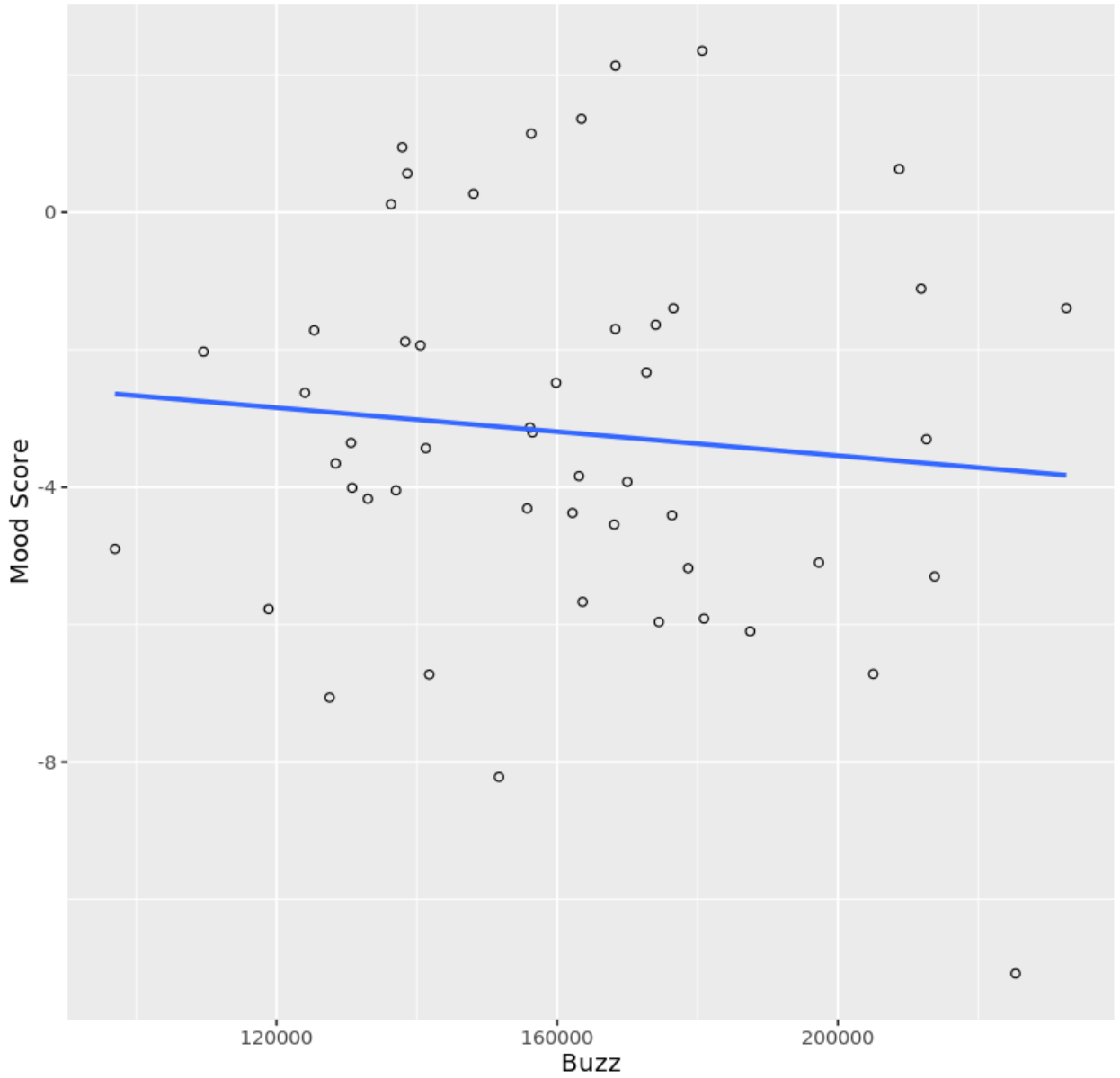
Let's compute the sentiment score of candidates using the $(\text{Positives} - \text{Negatives}) * 100 / \text{Buzz}$ formula. As the graph below shows sentiment on both candidates is mostly negative and overall Hillary Clinton is more negative. Voters talked more about **Donald Trump** and when they talked about him it was more positive than the time they talked about **Hillary Clinton**.



Another question we asked looking at this data, when the buzz on a candidate increases, does the sentiment also increase? In order to see that we can look at the scatter plot of sentiment vs. buzz. Below we see that the sentiment and buzz does not correlate for Hillary Clinton; however, for Donald Trump, sentiment and buzz has a negative correlation.



Donald Trump Buzz vs. Sentiment



Let's see how the United States map looked like before elections.

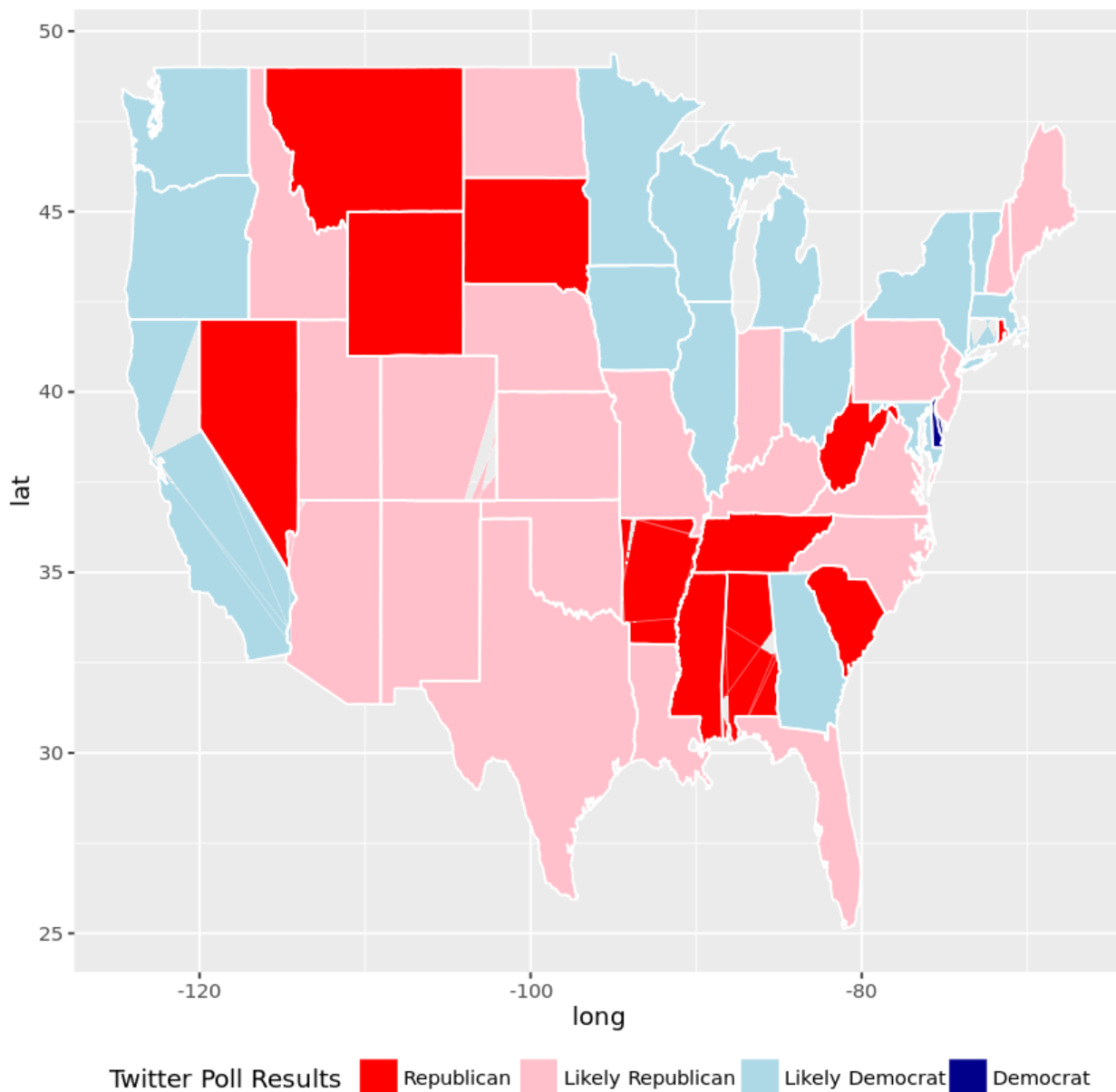
ID	Name	type	buzz	mood_sum	positives	negatives	neutral:
5516321c9b724739f3649835	Donald Trump	person	40319	-1731	4487	5810	30022
5516321c9b724739f3649835	Donald Trump	person	8823	-380	987	1264	6572
5516321c9b724739f3649835	Donald Trump	person	75070	-3663	8299	11159	55612
5516321c9b724739f3649835	Donald Trump	person	19544	-464	2244	2657	14643

5516321c9b724739f3649835	Donald Trump	person	363755	-26054	38230	57080	268445
5516321c9b724739f3649835	Donald Trump	person	53255	-3423	5680	8152	39423

	state_name	positive_precentage
51	alabama	-0.08861451
52	alaska	-0.07658820
53	arizona	-0.08679737
54	arkansas	-0.08077781
55	california	-0.05069828
56	colorado	-0.07150584

region	hill_mood	donald_mood	electoral_vote	mood
alabama	-0.08861451	-0.03281331	9	-0.05580120
alaska	-0.07658820	-0.03139522	3	-0.04519298
arizona	-0.08679737	-0.03809778	11	-0.04869960
arkansas	-0.08077781	-0.02113181	6	-0.05964600
california	-0.05069828	-0.05182059	55	0.00112231
colorado	-0.07150584	-0.04641818	9	-0.02508766

Many of the states where Donald Trump won was already decided, however, most undecided states were closer to Republican candidate than Hillary Clinton.



If we just assume the states, where the sentiment score for Hillary Clinton is higher than the sentiment score for Donald Trump, then Hillary Clinton will win those states. We will get the following states as the states Democratic candidate will won:

```
[1] " california , delaware , illinois , maryland , massachusetts , washington , "
```

and the states that will be won by Donald Trump are:

```
[1] " alabama , alaska , arizona , arkansas , colorado , connecticut , florida , georgia , hawaii , idaho , indiana , iowa , kansas , kentucky , louisiana , maine , michigan , minnesota , mississippi , missouri , montana , nebraska , nevada , new hampshire , new jersey , new mexico , new york , north carolina , north dakota , ohio , oklahoma , oregon , pennsylvania , rhode island , south carolina , south dakota , tennessee , texas , utah , "
```

```
vermont , virginia , west virginia , wisconsin , wyoming , "
```

These are the states where Donald Trump was talked more positively but won by Hillary Clinton.

```
'colorado' 'connecticut' 'district of columbia' 'hawaii' 'maine' 'michigan'  
'minnesota' 'nevada' 'new hampshire' 'new jersey' 'new mexico' 'new york'  
'oregon' 'rhode island' 'vermont' 'virgina'
```

Accuracy of our prediction using this assumption is show below.

```
[1] "Recall: 0.272727272727273 "  
[1] "Precision: 1"
```

```
[1] "Recall: 0.931034482758621 "  
[1] "Precision: 0.613636363636364"
```

0.66

Next, we can assume that all the light blue and blue states voted for Clinton and pink and red states voted for Trump. Then the precision and recall for Hillary Clinton predictions is as follows:

```
[1] "Recall: 0.545454545454545 "  
[1] "Precision: 0.75"
```

and the precision and recall for Donald Trump predictions is as follows:

```
[1] "Recall: 0.758620689655172 "  
[1] "Precision: 0.6875"
```

0.68