

# **TWITTER SENTIMENT ANALYSIS: HAS SENTIMENT TOWARDS TRUMP SHIFTED AFTER ELECTION?**

Bloomiya Kurian

Foundations of Data Science Workshop

May 2017

# THE PROBLEM

- Soon after Trump became President his Policies and Views changed the tone of national conversation
- Executive Orders such as immigration ban, travel ban, Muslim ban etc. sparked debate on both sides of the aisle
- Identify if the supporters who voted for Trump also supported his policies or has the sentiment shifted after election



# THE APPROACH

- Collect tweets containing hashtags\* and search terms ('immigrationban', 'travelban', 'muslimban', 'BanIslam', 'resist', 'DeleteUber', 'IResist', 'TheResistance') for each state in US
- Clean-up tweet texts and calculate a polarity level (-1 to +1) sentiment score for each text
- Aggregate (mean) sentiment score on State level

\*These search terms and hashtags were selected based on the major actions (executive orders on travel ban, immigration etc.) President Trump implemented immediately after becoming the President. Some articles such as [this one](#) listed the most common hashtags people used to tweet their opinions about President Trump.

## THE APPROACH (2)

- Visualize sentiment polarity on US map and compare state wise support and resistance to Trump
- Statistically prove if there is a relationship between voting trend and sentiment after election across the US states



# DATA EXTRACTION

- Used Twitter Rest API
- 'searchTwitter' R function
- Search query with search terms, geocodes and radius was executed for main cities of all US states except Idaho
- Results of all the queries were consolidated, deduped and saved to a .csv file

# DATA SET

G2																			NA	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R		
	Text	Favorite	Favorit	ReplyToS	Created	Truncate	ReplyToSI	ID	ReplyToU	StatusSou	ScreenNam	Retwee	IsRetwe	Retweete	Longitu	Latitud	City	State		
1	Some play chess, others play 5	FALSE	0	NA	3/2/2017 6:06	FALSE	NA	8.37E+17	NA	<a href="h	Deadlyalpaca	0	FALSE	FALSE	NA	NA	Birmingham	AL		
3	@wef Bc the less educated the	FALSE	0	wef	3/2/2017 6:05	FALSE	8.37E+17	8.37E+17	5120691	<a href="h	SwannDives	0	FALSE	FALSE	NA	NA	Birmingham	AL		
4	#TheResistance My #Dog Baile	FALSE	1	NA	3/2/2017 6:05	TRUE	NA	8.37E+17	NA	<a href="h	heatherhan	0	FALSE	FALSE	NA	NA	Birmingham	AL		
5	How many	FALSE	2	NA	3/2/2017 6:01	FALSE	NA	8.37E+17	NA	<a href="h	QueenBeez	0	FALSE	FALSE	NA	NA	Birmingham	AL		
6	RT @MeeestarCoke: @ALT_DC	FALSE	0	NA	3/2/2017 5:52	FALSE	NA	8.37E+17	NA	<a href="h	starrhaus	1	TRUE	FALSE	NA	NA	Birmingham	AL		
7	@ALT_DOJ @bluedutchess726	FALSE	0	ALT_DOJ	3/2/2017 5:49	TRUE	8.36E+17	8.37E+17	8.24E+17	<a href="h	MeeestarCo	1	FALSE	FALSE	NA	NA	Birmingham	AL		
8	Every surrogate follower voter	FALSE	0	NA	3/2/2017 5:47	TRUE	NA	8.37E+17	NA	<a href="h	kat1063	0	FALSE	FALSE	NA	NA	Birmingham	AL		
9	@HillaryClinton Please tell us	FALSE	0	HillaryClin	3/2/2017 5:38	FALSE	NA	8.37E+17	1.34E+09	<a href="h	QueenBeez	0	FALSE	FALSE	NA	NA	Birmingham	AL		
10	.@realDonaldTrump @POTUS	FALSE	0	NA	3/2/2017 5:35	TRUE	NA	8.37E+17	NA	<a href="h	Raynman67	0	FALSE	FALSE	NA	NA	Birmingham	AL		
11	.@SenRonJohnson Step down	FALSE	0	NA	3/2/2017 5:34	TRUE	NA	8.37E+17	NA	<a href="h	janicenall	0	FALSE	FALSE	NA	NA	Birmingham	AL		
12	.@RonJohnsonWI WTF?? Resig	FALSE	0	NA	3/2/2017 5:33	TRUE	NA	8.37E+17	NA	<a href="h	janicenall	0	FALSE	FALSE	NA	NA	Birmingham	AL		
13	Is this "the failing New York	FALSE	1	NA	3/2/2017 5:31	FALSE	NA	8.37E+17	NA	<a href="h	shelbybb	0	FALSE	FALSE	NA	NA	Birmingham	AL		
14	RT @MeeestarCoke:	FALSE	0	NA	3/2/2017 5:27	FALSE	NA	8.37E+17	NA	<a href="h	Amplitude3	2	TRUE	FALSE	NA	NA	Birmingham	AL		
15	RT @Deadlyalpaca: @NancyPe	FALSE	0	NA	3/2/2017 5:17	FALSE	NA	8.37E+17	NA	<a href="h	Drewtejano	2	TRUE	FALSE	NA	NA	Birmingham	AL		
16	RT @Deadlyalpaca: @NancyPe	FALSE	0	NA	3/2/2017 5:16	FALSE	NA	8.37E+17	NA	<a href="h	gjos24	2	TRUE	FALSE	NA	NA	Birmingham	AL		
17	Homegrown anti-Semitism. #Si	FALSE	0	NA	3/2/2017 5:15	FALSE	NA	8.37E+17	NA	<a href="h	prochoiceAl	0	FALSE	FALSE	NA	NA	Birmingham	AL		
18	<ed><U+00A0><U+00BD><ed>	FALSE	2	NA	3/2/2017 5:15	TRUE	NA	8.37E+17	NA	<a href="h	MeeestarCo	2	FALSE	FALSE	NA	NA	Birmingham	AL		
19	I was there that morning &amp	FALSE	0	NA	3/2/2017 5:12	TRUE	NA	8.37E+17	NA	<a href="h	NotJesseBro	0	FALSE	FALSE	NA	NA	Birmingham	AL		
20	RT @artsyboy1985: Why is it t	FALSE	0	NA	3/2/2017 5:07	FALSE	NA	8.37E+17	NA	<a href="h	bobaloo24u	10	TRUE	FALSE	NA	NA	Birmingham	AL		
21	RT @artsyboy1985: Why is it t	FALSE	0	NA	3/2/2017 5:05	FALSE	NA	8.37E+17	NA	<a href="h	feminist_w	10	TRUE	FALSE	NA	NA	Birmingham	AL		



# INITIAL DATA EXPLORATION

- Total Data Count : **65,7511**
- State with minimum number of tweets: **Delaware**
- State with maximum number of tweets: **New York**
- Missing State: **Idaho**
- Varying count of tweets per State -> **Uneven population**

State	tweet_count
AK	2115
AL	5389
AR	1158
AZ	3993
CA	45198
CO	4686
CT	1820
DC	69355
DE	160
FL	10312
GA	11968
IA	1242
IL	15416
IN	4502
KS	1678
KY	1288
LA	2735
MA	16064
MD	73861
ME	1899
MI	1710
MN	3081
MO	24626
MS	749
MT	484
NC	10878
ND	3142
NE	1192
NH	1090
NJ	23276
NM	8537
NV	5499
NY	122622
OH	23140
OK	1838
OR	4342
PA	38733
RI	540
SC	3968
SD	1092
TN	1507
TX	29015
UT	1313
VA	53578
VT	310
WA	9002
WI	6124
WV	948
WY	336

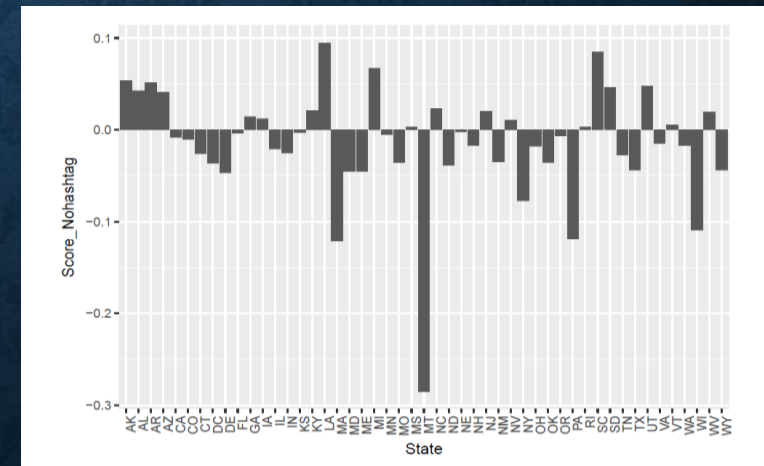
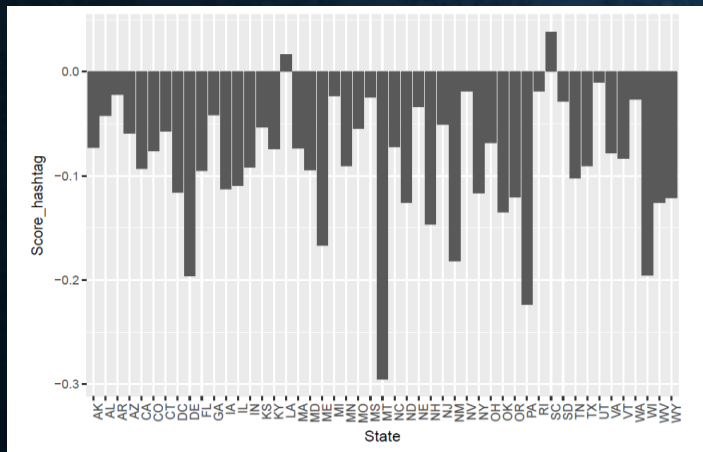
# SENTIMENT SCORE

- Sentiment Classification Package: '**sentimentr**'
- Calculated sentiment on sentence level rather than counting number of negative and positive words in the sentence
- Score was generated for texts '**with**' and '**without**' hashtags
- Correlation Analysis between the two scores showed a strong positive association with a coefficient of **0.8766415**



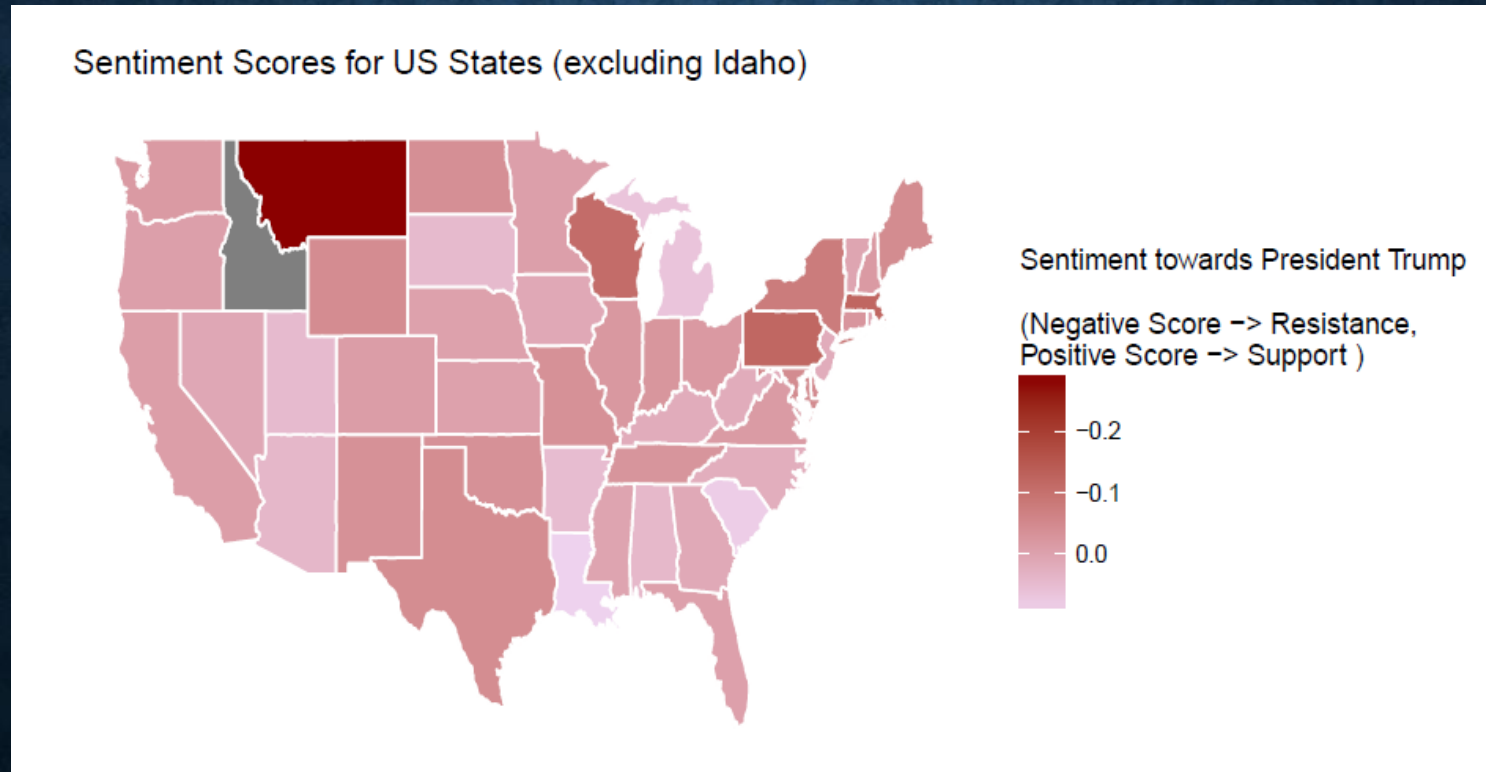
# SENTIMENT SCORE DISTRIBUTION

- Texts without hashtag gave similar number of positive (supporting) and negative (resisting) scores
- Texts with hashtag gave negative scores for all states except two
- **Hence scores generated by texts without hashtag were selected to perform further analysis**
- Distribution **with** hashtag
- Distribution **without** hashtag



# US MAP WITH SENTIMENT SCORE

Mean score for all States except Idaho





# STATISTICAL ANALYSIS

- Sentiment score and popular vote election results for each state were compared to identify if there is a relation between a state's people's voting trend and their sentiment towards the president's actions post election
- Trump's Victory Margin for each state was used as election results
- **Victory Margin:** Difference between total number of votes received by a candidate and total number of votes received by the winning person (or if the candidate is the winner for the state, the difference between his votes and the second place candidate's votes)

# ELECTION RESULTS DATA SET

State	Total '16				VictoryMargin
	Clinton (D)	Trump (R)	Others	Votes	
Alabama	729,547	1,318,255	75,570	2,123,372	27.70%
Arizona	1,161,167.00	1,252,401.00	159,597	2,573,165	3.50%
Arkansas	380,494.00	684,872.00	65,269	1,130,635	26.90%
California	8,753,788	4,483,810	943,997	14,181,595	-30.10%
Colorado	1,338,870	1,202,484	238,866	2,780,220	-4.90%
Connecticut	897,572	673,215	74,133	1,644,920	-13.60%
Delaware	235,603	185,127	20,860	441,590	-11.40%
District of Columbia	282,830	12,723	15,715	311,268	-86.80%
Florida	4,504,975.00	4,617,886.00	297,178	9,420,039	1.20%
Georgia	1,877,963	2,089,104	125,306	4,092,373	5.20%
Idaho	189,765	409,055	91,435	690,255	31.80%
Illinois	3,090,729	2,146,015	299,680	5,536,424	-17.10%
Indiana	1,033,126.00	1,557,286.00	144,546	2,734,958	19.20%
Iowa	653,669	800,983	111,379	1,566,031	9.40%
Kansas	427,005	671,018	86,379	1,184,402	20.60%
Kentucky	628,854	1,202,971	92,324	1,924,149	29.80%
Louisiana	780,154	1,178,638	70,240	2,029,032	19.60%
Maine	357,735.00	335,593.00	54,599	747,927	-3.00%
Maryland	1,677,928.00	943,169	160,349	2,781,446	-26.40%
Massachusetts	1,995,196	1,090,893	238,957	3,325,046	-27.20%
Michigan	2,268,839	2,279,543	250,902	4,799,284	0.20%
Minnesota	1,367,716	1,322,951	254,146	2,944,813	-1.50%
Mississippi	485,131	700,714	23,512	1,209,357	17.80%
Missouri	1,071,068	1,594,511	143,026	2,808,605	18.60%
Montana	177,709	279,240	40,198	497,147	20.40%
Nebraska	284,494	495,961	63,772	844,227	25.00%

State	Total '16				VictoryMargin
	Clinton (D)	Trump (R)	Others	Votes	
Nevada	539,260.00	512,058.00	74,067	1,125,385	-2.40%
New Hampshire	348,526	345,790	49,842	744,158	-0.40%
New Jersey	2,148,278	1,601,933	123,835	3,874,046	-14.10%
New Mexico	385,234.00	319,666.00	93,418	798,318	-8.20%
New York	4,556,124	2,819,534	345,795	7,721,453	-22.50%
North Carolina	2,189,316.00	2,362,631.00	189,617	4,741,564	3.70%
North Dakota	93,758	216,794	33,808	344,360	35.70%
Ohio	2,394,164	2,841,005	261,318	5,496,487	8.10%
Oklahoma	420,375	949,136	83,481	1,452,992	36.40%
Oregon	1,002,106.00	782,403.00	216,827	2,001,336	-11.00%
Pennsylvania	2,926,441	2,970,733	218,228	6,115,402	0.70%
Rhode Island	252,525	180,543	31,076	464,144	-15.50%
South Carolina	855,373	1,155,389	92,265	2,103,027	14.30%
South Dakota	117,458	227,721	24,914	370,093	29.80%
Tennessee	870,695	1,522,925	114,407	2,508,027	26.00%
Texas	3,877,868	4,685,047	406,311	8,969,226	9.00%
Utah	310,676	515,231	305,523	1,131,430	18.10%
Vermont	178,573	95,369	41,125	315,067	-26.40%
Virginia	1,981,473	1,769,443	231,836	3,982,752	-5.30%
Washington	1,742,718	1,221,747	401,179	3,365,644	-15.50%
West Virginia	188,794	489,371	34,886	713,051	42.20%
Wisconsin	1,382,536	1,405,284	188,330	2,976,150	0.80%
Wyoming	55,973	174,419	25,457	255,849	46.30%



# REGRESSION ANALYSIS

- Dependent Variable: **Sentiment Score**
- Independent Variable: **Victory Margin**
- Two outliers (removed):

*Montana (Sentiment Score) = -0.285*

*DC (Victory Margin) = -0.868*

# REGRESSION RESULTS

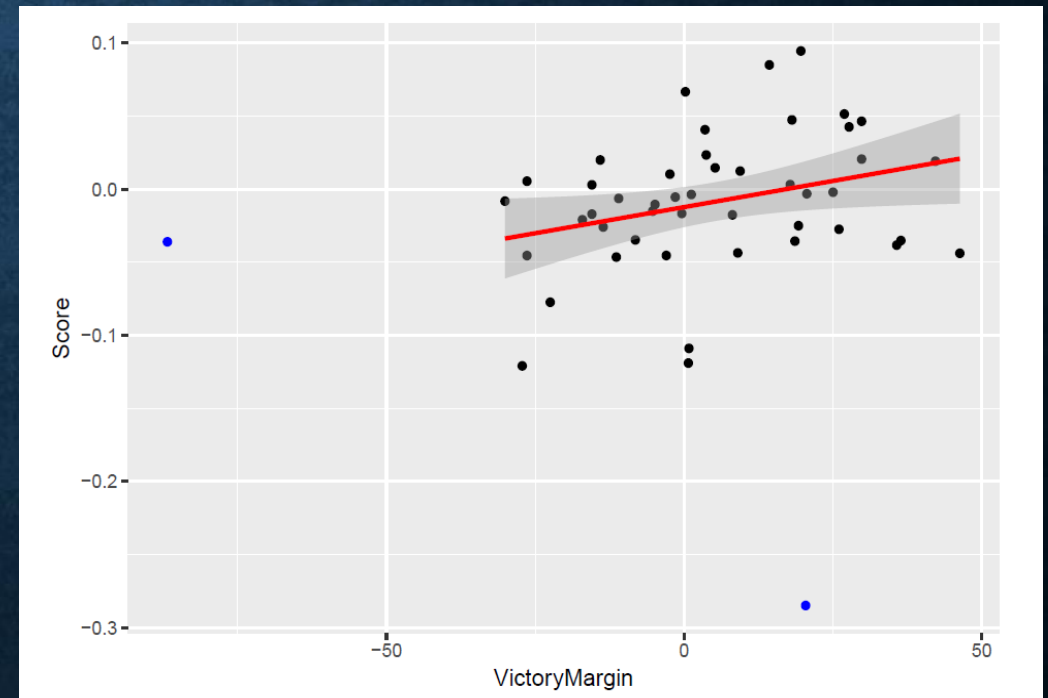
- Roughly 9.4% of the variance found in the Sentiment Score can be explained by Victory Margin
- P-value less than the alpha value .05 indicates that the relationship between Victory Margin and Sentiment Score is significant

```
Call:
lm(formula = Score ~ VictoryMargin, data = Final_Data)

Residuals:
    Min       1Q   Median       3Q      Max
-0.107239 -0.026371  0.003121  0.026138  0.092801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0122861  0.0067768  -1.813   0.0767 .
VictoryMargin  0.0007156  0.0003348   2.137   0.0382 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04442 on 44 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.09406,    Adjusted R-squared:  0.07347
F-statistic: 4.568 on 1 and 44 DF,  p-value: 0.03817
```





# CONCLUSION

- State with highest resistance: Montana and State with highest support: Louisiana
- Based on regression analysis, only 9.4% of the variation in Sentiment Score can be explained a linear relationship with the Victory Margin
- A p-value less than .05 indicates that the null hypothesis can be rejected and can conclude that there is a significant linear relationship between Victory Margin and Sentiment Score
- Tweets collected show that the sentiment towards President Trump has not changed much after the election

**THANK YOU TO:**

**CHRIS BISHOP**

**SPRINGBOARD TEAM**



# REFERENCES

- <https://extranewsfeed.com/what-does-twitter-have-to-say-about-trumps-executive-order-on-immigration-602f45eeef32>