

Interpretation for clusters of Twitter accounts (Z-score word cloud)

READ CAREFULLY: The U.S Presidential Elections in 2016 were accompanied by many controversies. One of them is that the Russian government allegedly tried to influence result of the elections by spreading fake news, propaganda and misinformation. The U.S. Congress released a list of 2,848 Twitter accounts suspected of being tied to Russian government. We call those accounts "Troll accounts".

The aim of our study is to determine how to cluster* those Troll accounts according to the similarity of tweets that were posted by them. We created a clustering model that clusters Troll accounts into six clusters. Due to amount of data, we had to use computational Natural Language Processing** algorithms to perform the clustering. As a consequence, the result of the clustering is not directly understandable to humans.

To facilitate communication of the results, we propose several ways of visualization of the clusters. In this study, we would like to evaluate these alternative visualization for their interpretability and comprehensibility.

* Refer to https://en.wikipedia.org/wiki/Cluster_analysis (https://en.wikipedia.org/wiki/Cluster_analysis) for more information about clustering.

** Refer to https://en.wikipedia.org/wiki/Natural_language_processing (https://en.wikipedia.org/wiki/Natural_language_processing) for more information about NLP.

There are 29 questions in this survey.

Prolific ID

Please, enter your Prolific ID: *

Please write your answer here:

Prior knowledge

In this section, we will ask you about your prior knowledge of topics related to this survey.

On scale of **1** to **10** how would you rate your knowledge about the following topics?

(**1** - no knowledge at all; **10** - highly knowledgeable about this topic) *

Please choose the appropriate response for each item:

	1	2	3	4	5	6	7	8	9	10
Clustering / Segmentation (unsupervised data mining algorithms)	<input type="radio"/>									
NLP (natural language processing)	<input type="radio"/>									
Concept of Z-scores (Standardization technique)	<input type="radio"/>									
Russian interference on U.S. Presidential Election in 2016	<input type="radio"/>									

With which U.S. political party is **Hilary Clinton** affiliated? *

Please choose **only one** of the following:

- Democrats
- Republicans
- Other

With which U.S. political party is **Donald Trump** affiliated? *

Please choose **only one** of the following:

- Democrats
- Republicans
- Other

Which candidate won the U.S. Presidential Election in 2016?

*

Please choose **only one** of the following:

- Hillary Clinton
- Ted Cruz
- Donald Trump
- Bernie Sanders

Classify Troll accounts

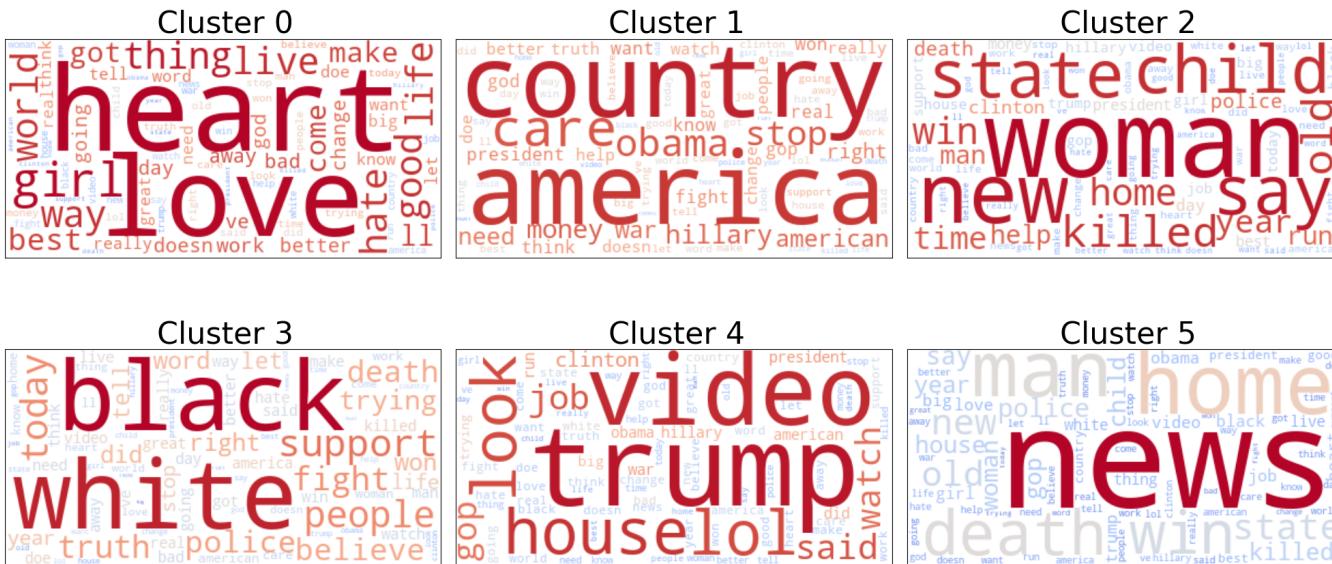
Z-score word cloud visualization of the clustering model is shown below. Word clouds are often used to outline the most significant keywords in some domain. This word cloud is using z-score normalization* of words' weights for calculation of font size and font color. Therefore, words that are huge and **red** occur in a cluster more frequently than in the whole data set. Similarly, words that are small and **blue** occur in a cluster less frequently than in the whole data set.

In the picture below, you can see a Z-score word cloud visualization for each cluster of Troll accounts. We ask you to assign each of the six Troll accounts presented below into one of the clusters. There are six clusters** to choose from. Each cluster is presented by one of the word clouds. None of the six Troll accounts presented below belongs to the same cluster.

Troll accounts are represented by an example consisting of five tweets posted by them.

**In statistics, the z-score is the signed fractional number of standard deviations by which the value of an observation or data point is above the mean value of what is being observed or measured. Observed values above the mean have positive z-scores, while values below the mean have negative z-scores [Wikipedia].*

***Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics [Wikipedia].*



To which cluster would you assign a Troll account that posted the following tweets?

- #BlackHistoryMonth #BlackEveryDay #BlackIsBeautiful #AmericanHistoryIsBlack
- I woke up in a cold sweat. Can't go back to sleep. My People, always on my mind. #BlackLivesMatter
#BlackPowerBaby <https://t.co/OxEsCEzzip> (<https://t.co/OxEsCEzzip>)
- We need justice. Police are not who can help with it. #policebrutality #WearHoodieForTrayvon
<https://t.co/RNCJoa9tlz> (<https://t.co/RNCJoa9tlz>)
- Any black journalist worth a damn would've researched race of the #OscarsSoWhite supporters, managers
- Get \$0.05 for every download. \$5.00 for every biz. <https://t.co/jtsItkOFRk> (<https://t.co/jtsItkOFRk>)
#blackowned #BuyBlack #blacktwitter <https://t.co/FTpIQbHH8a> (<https://t.co/FTpIQbHH8a>)

*

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

To which cluster would you assign a Troll account that posted the following tweets?

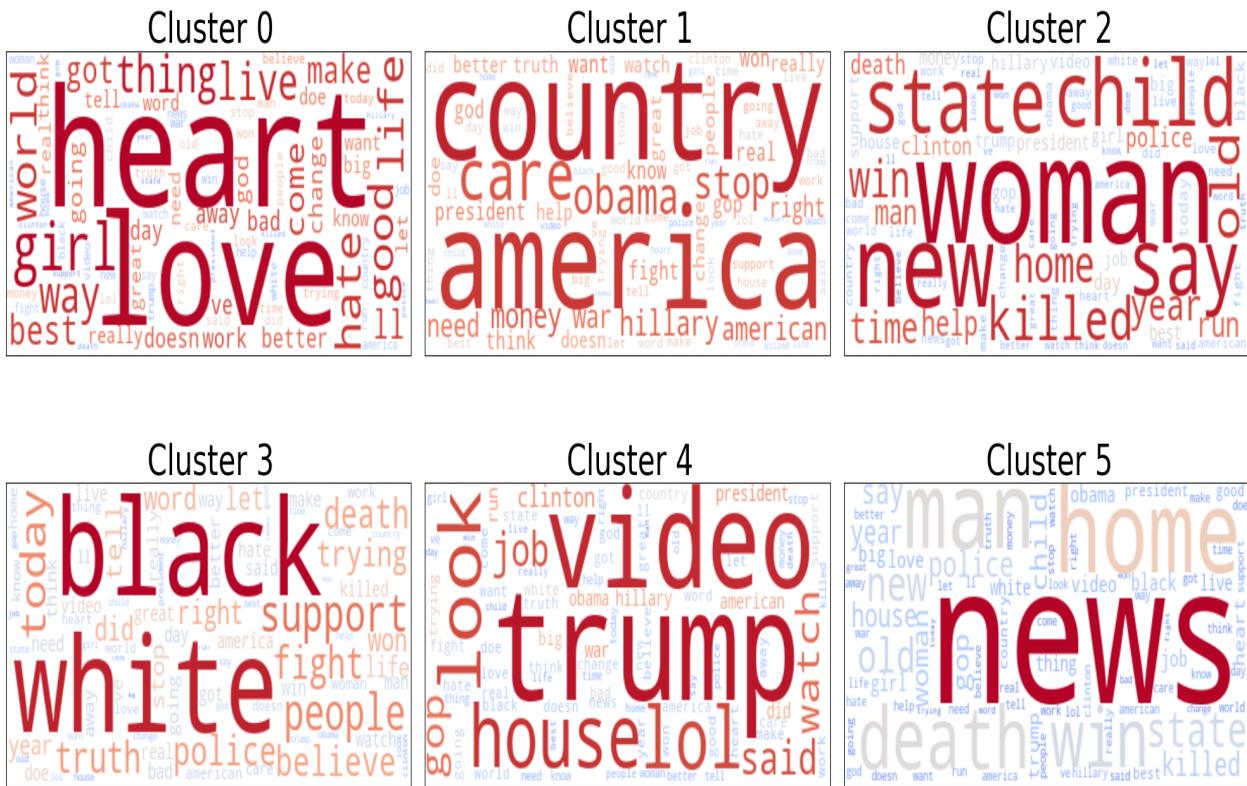
- #Politics #News U.S. Senate passes bill renewing federal terrorism insurance program: WASHINGTON (Reuters) - The U.S. Senate on Thurs...
- #Politics #News Paris attack complicates Republican plans on U.S. security funding: WASHINGTON (Reuters) - The deadly attack in Paris...
- #News #Crime Police seize 2,700 marijuana plants at grow houses in San Jose, San Leandro and Fairfield (San Jo... <http://t.co/lfliGnccl8> (<http://t.co/lfliGnccl8>)
- #News #US U.S. military Twitter account hacked (CBS News) <http://t.co/lbEbt7Taq6> (<http://t.co/lbEbt7Taq6>)
- #TopNews Short courses offer hope to U.S. education companies <http://t.co/GQOhi63qbH> (<http://t.co/GQOhi63qbH>)

*

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

Same visualization repeated for easier reference:



To which cluster would you assign a Troll account that posted the following tweets?

- Of course, You can make America great again #VegasGOPDebate
- BREAKING: #HILLARY AUDITIONING TO BECOME 3RD MEMBER OF CIRCUS ACT. #MakeAmericaGreatAgain
<https://t.co/4rrRSnLebD> (<https://t.co/4rrRSnLebD>)
- Of course, Obama has shown us how the dems are going to rule our country, and I don't like it
#VegasGOPDebate
- Donald Trump routinely lies, and the media lets him. Why? via @FortuneMagazine
<https://t.co/VfDHW6cgdr> (<https://t.co/VfDHW6cgdr>) #CCOT #TeaParty #Nhprimary
- #demndebate What will you do with #Obamacare? #DemDebate

*

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4

Cluster 5

To which cluster would you assign a Troll account that posted the following tweets?

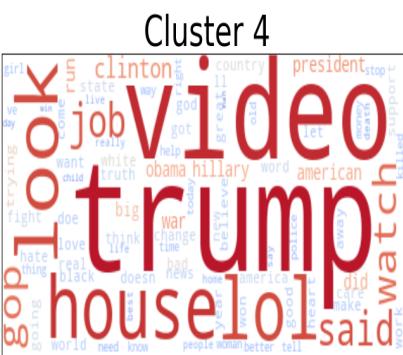
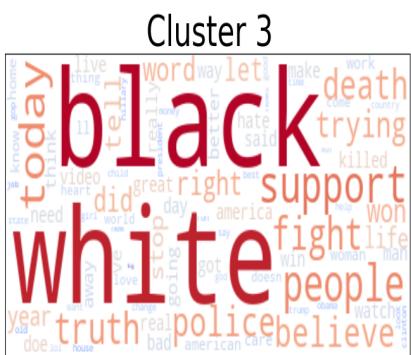
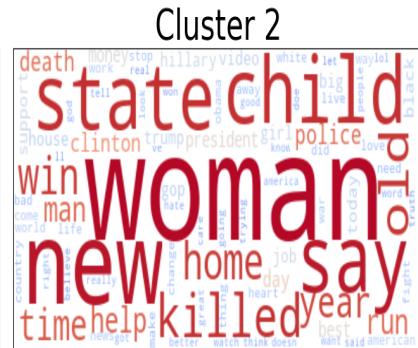
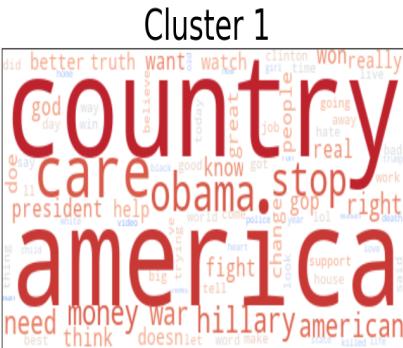
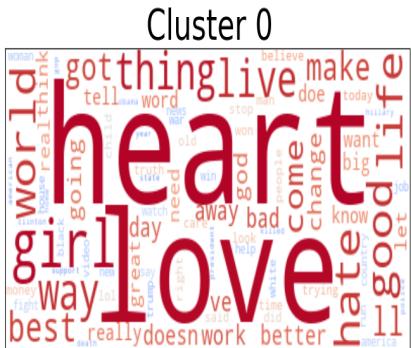
- don't fall in love. fall off a bridge, it hurts less.
- I'm not gonna change for anyone, I don't care what people think, because I am me, and proud of it.
- If you truly love someone, then they never leave your heart, only your side. #iHQ
- "I like the way I feel when he looks at me. Like I wanna believe in myself." -Serena van der Woodsen
#IAMONFIRE
- "Shut the fuck up and have the shot." -#GirlCode #lamonfire

*

Please choose **only one** of the following:

 Cluster 0
 Cluster 1
 Cluster 2
 Cluster 3
 Cluster 4
 Cluster 5

Same visualization repeated for easier reference:



To which cluster would you assign a Troll account that posted the following tweets?

- 3 Dead, 4 Wounded in Mass Shooting on Rochester's Southwest Side
- Captured New York inmate pleads not guilty to escape charges
- New York Prison Escapee David Sweat Pleads Not Guilty to Escape an
- St. Louis Protests Sparked by 18-Year-Old Killed by Police
- Growing Washington wildfires are an unprecedented cataclysm, governor says

*

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

To which cluster would you assign a Troll account that posted the following tweets?

- #ameliss Update : Google Fires Employee for Criticizing Forced Diversity [\(https://t.co/EAYtQHy6kh\) #amelin \[\\(https://t.co/xps1PXad6M\\)\]\(https://t.co/xps1PXad6M\)](https://t.co/EAYtQHy6kh)
- #ameliss VIDEO : Corey SLAMS GOP Traitors Not on Board with Trump Agenda [\(https://t.co/MVL7rFdi4Z\) #amelin \[\\(https://t.co/HuM2Ys8e9Y\\)\]\(https://t.co/HuM2Ys8e9Y\)](https://t.co/MVL7rFdi4Z)
- #ameliss Trump TRIGGERS Climate Change Freaks With THIS Policy Change [\(https://t.co/TT3LJ1tP6c\) #amelin \[\\(https://t.co/j8BjZgtsKP\\)\]\(https://t.co/j8BjZgtsKP\)](https://t.co/TT3LJ1tP6c)
- #ameliss Anti-Trump Nevada Senator Dean Heller Just Got VERY BAD NEWS [\(https://t.co/ippFCnaac\) #amelin \[\\(https://t.co/rmHIPPVOhq\\)\]\(https://t.co/rmHIPPVOhq\)](https://t.co/ippFCnaac)
- #ameliss BREAKING : Michelle Obama's Lunch Program Connected to Indicted Criminal [\(https://t.co/tYznOdgwYS\) #amelin \[\\(https://t.co/faVpOVTVTL\\)\]\(https://t.co/faVpOVTVTL\)](https://t.co/tYznOdgwYS)

*

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

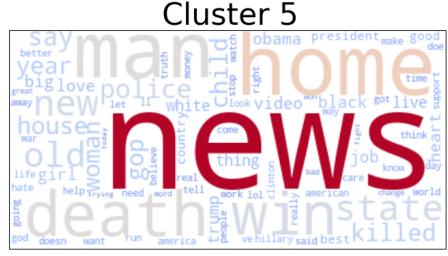
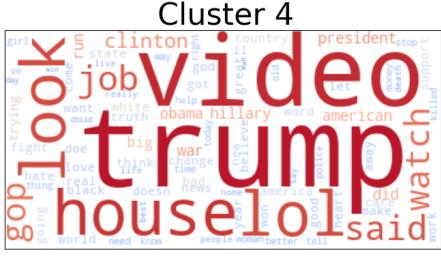
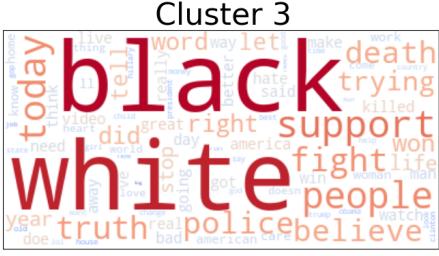
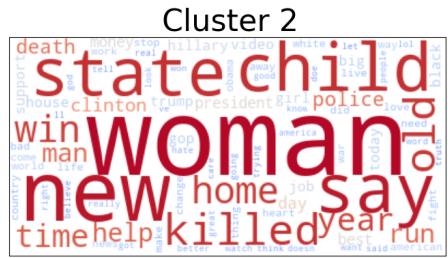
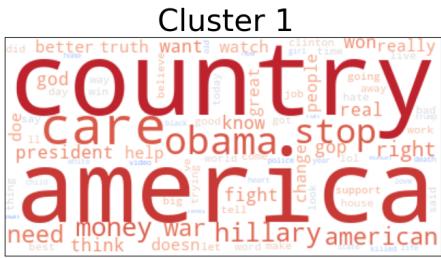
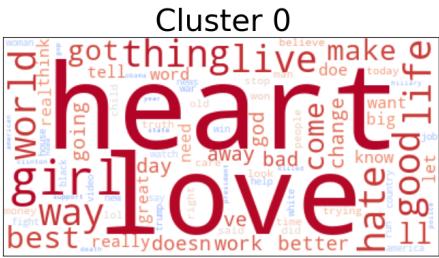
Explain clusters

Z-score word cloud visualization of the clustering model is shown below. Word clouds are often used to outline the most significant keywords in some domain. This word cloud is using z-score normalization* of words' weights for calculation of font size and font color. Therefore, words that are huge and **red** occur in a cluster more frequently than in the whole data set. Similarly, words that are small and **blue** occur in a cluster less frequently than in the whole data set.

In the picture below, you can see a Z-score word cloud visualization for each cluster of Troll accounts. We ask you to answer eight general questions about visualized clusters*.

* In statistics, the z-score is the signed fractional number of standard deviations by which the value of an observation or data point is above the mean value of what is being observed or measured. Observed values above the mean have positive z-scores, while values below the mean have negative z-scores [Wikipedia].

** Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics [Wikipedia].



Which three words are the most characteristic for the **cluster 0**? *

Please choose **only one** of the following:

- police, Trump, Obama
- love, heart, thing
- black, police, white
- video, look, Obama

Which three words are the most characteristic for the **cluster 3**? *

Please choose **only one** of the following:

- police, Trump, Obama
- love, heart, thing
- black, police, white
- video, look, Obama

Which word would have to be added to the following tweet to most increase the probability of it becoming a **cluster 5 member?**

- 3 Dead, 4 Wounded in Mass Shooting on Rochesters Southwest Side

*

Please choose **only one** of the following:

- police
- black
- news
- death

Which word would have to be added to the following tweet to most increase the probability of it becoming a **cluster 4 member?**

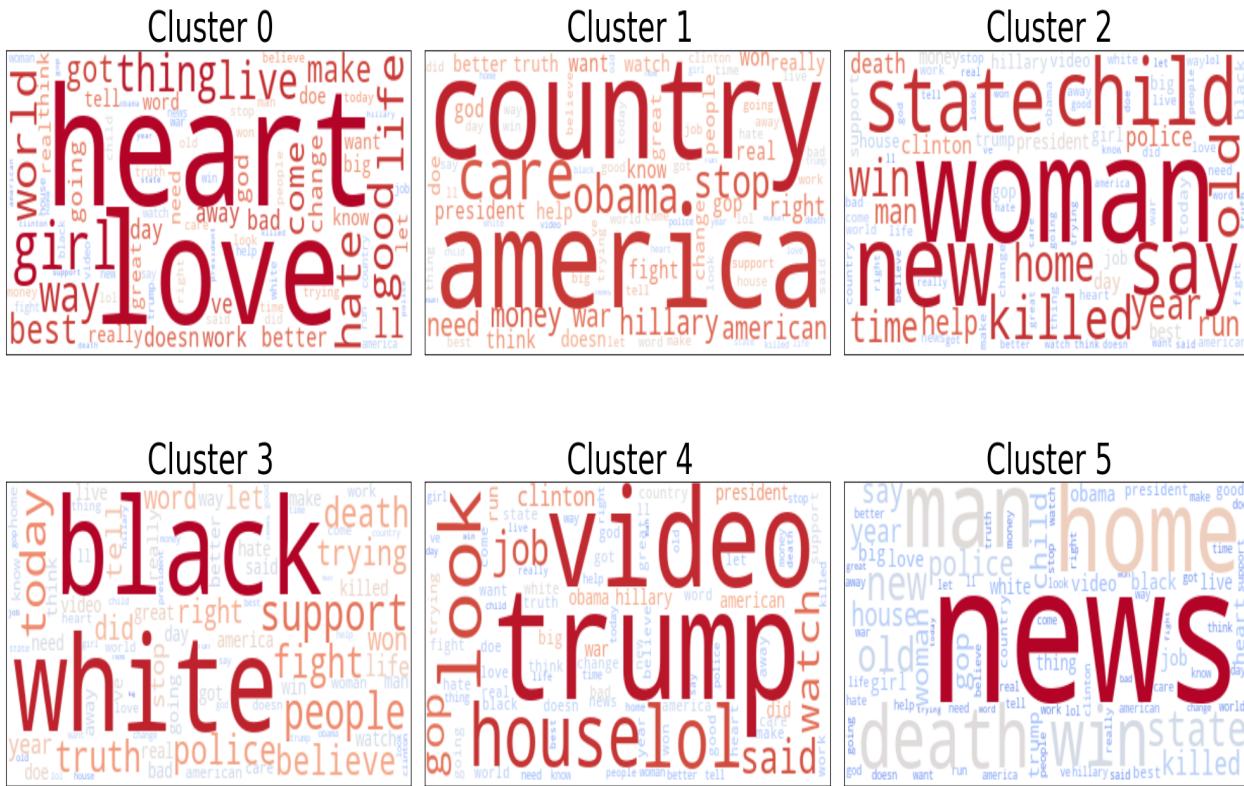
- We need justice. Police are not who can help with it. #policebrutality #WearHoodieForTrayvon
<https://t.co/RNCJoa9tlz> (<https://t.co/RNCJoa9tlz>)

*

Please choose **only one** of the following:

- new
- trump
- look
- obama

Same visualization repeated for easier reference:



Which cluster could be described as accounts that tweet about racial issues?

*

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

Which cluster could be described as accounts that tweet about news stories? *

Please choose **only one** of the following:

- Cluster 0
- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Cluster 5

What is the best suitable description for the **cluster 2?** *

Please choose **only one** of the following:

- accounts that tweet about crime
- accounts that tweet about news stories
- accounts that tweet about Donald Trump
- accounts that tweet American presidents

What is the best suitable description for the **cluster 1?** *

Please choose **only one** of the following:

- accounts that tweet about Donald Trump
- accounts that tweet about American presidents
- accounts that tweet about crime
- accounts that tweet about news stories

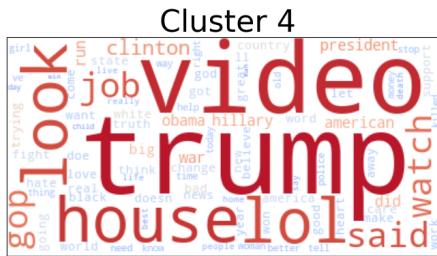
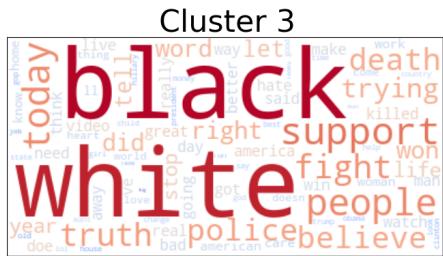
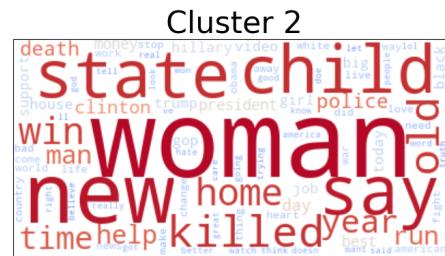
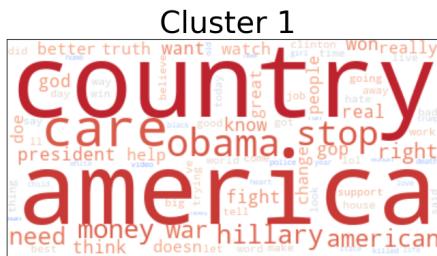
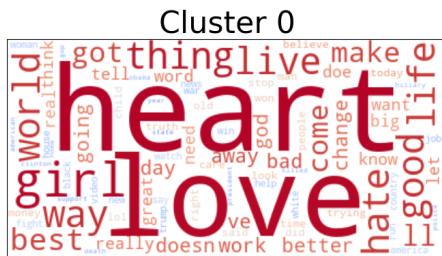
True/false questions

Z-score word cloud visualization of the clustering model is shown below. Word clouds are often used to outline the most significant keywords in some domain. This word cloud is using z-score normalization* of words' weights for calculation of font size and font color. Therefore, words that are huge and **red** occur in a cluster more frequently than in the whole data set. Similarly, words that are small and **blue** occur in a cluster less frequently than in the whole data set.

In the picture below, you can see a Z-score word cloud visualization for each cluster of Troll accounts. We would like to ask you eight true/false questions about the visualizations.

* In statistics, the z-score is the signed fractional number of standard deviations by which the value of an observation or data point is above the mean value of what is being observed or measured. Observed values above the mean have positive z-scores, while values below the mean have negative z-scores [Wikipedia].

** Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, and computer graphics [Wikipedia].



Could **cluster 4** be described as "accounts that often tweet about Donald Trump"? *

Please choose **only one** of the following:

- True
- False

Could **cluster 0** be described as "accounts that often tweet about American presidents"? *

Please choose **only one** of the following:

- True
- False

Is it true that accounts in **clusters 1 and 5** often talk about America and presidents? *

Please choose **only one** of the following:

- True
- False

Is it true that accounts in **clusters 0 and 2** often talk about crime and violence? *

Please choose **only one** of the following:

- True
- False

Is it true that tweets in **clusters 3 and 5** rarely mention love? *

Please choose **only one** of the following:

- True
- False

Is it true that tweets in **clusters 2 and 5** rarely mention crime and violence? *

Please choose **only one** of the following:

- True
- False

Comprehensibility

In this section, we ask you questions about comprehensibility of word cloud visualization for representing results of our clustering model.

What is your opinion on the comprehensibility of the Z-score word cloud visualization for communicating the results of tweet clustering?

*

Please choose **only one** of the following:

- very easy to comprehend
- easy to comprehend
- comprehensible
- difficult to comprehend
- very difficult to comprehend

Make a comment on word cloud comprehensibility here:

Please write your answer here:

Submit your survey.

Thank you for completing this survey.