

Project 1: Social Media Analytic

By Group 17

Abishek Ganesh

Mahmood M Naumani

Sri Kavya Ravella

Contents

Problem Statement.....	3
Data collection	3
Geographic variations	3
Word cloud.....	5
Word Cloud generated using data collected from the 5 regions	5
Word Cloud for Washington and Oregon	6
Word cloud for Texas	6
Word Cloud for Georgia and Florida.....	7
Word cloud for California.....	7
Word cloud for Colorado	8
Topic modeling.....	8
NMF.....	8
Topic Modelling for All 5 States	8
Topic Modelling for California.....	9
Topic Modelling for Washington and Oregon.....	9
Topic Modelling for Texas	9
Topic Modelling for Georgia and Florida	9
Topic Modelling for Colarado	10
Gensim	10
For all 5 Regions.....	10
Florida and Georgia.....	11
Colorado.....	12
California.....	14
Washington and Oregon.....	15
Texas	16
Insights and Conclusion	17
Considerations	18

Problem Statement

Donald J Trump is one of the most controversial presidents of United States, after being elected as the President he has signed many executive orders and presidential memorandums which has brought him lot of criticism from people within and outside United States.

Our project focuses on using text mining approaches to conduct social media analytics to understand people's sentiment about the new administration within the different states within United States .We use twitter as our source of information. We perform sentimental analysis using this information, which helps us measure how positive /negative the tweets are. This will give us an overall idea of people's opinion on President Trump.

Data collection

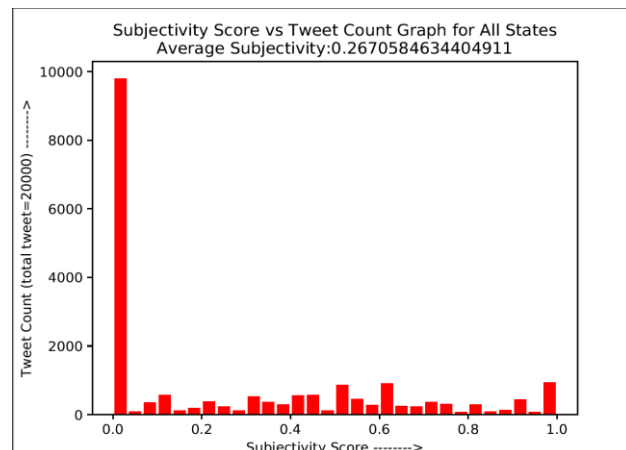
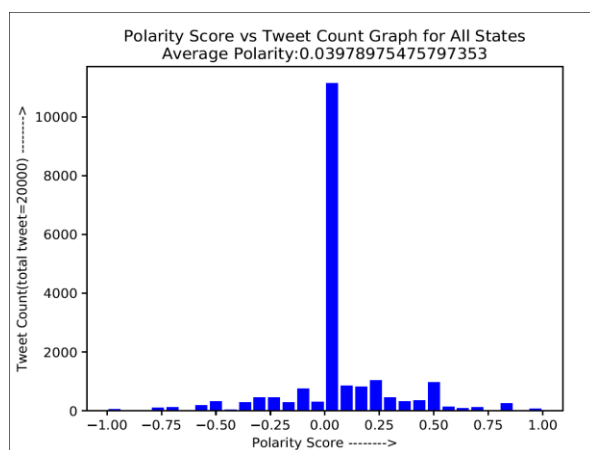
Using twitters steaming Api -Twython, we collected 10k tweets for the key word Trump from five regions which are Texas, California, Colorado and for the fourth region we combined Oregon & Washington and for the fifth region we combined Georgia & Florida. We have collected tweets as a Json and stored them as pickle file after cleaning and removing punctuations, digits and Unicode characters.

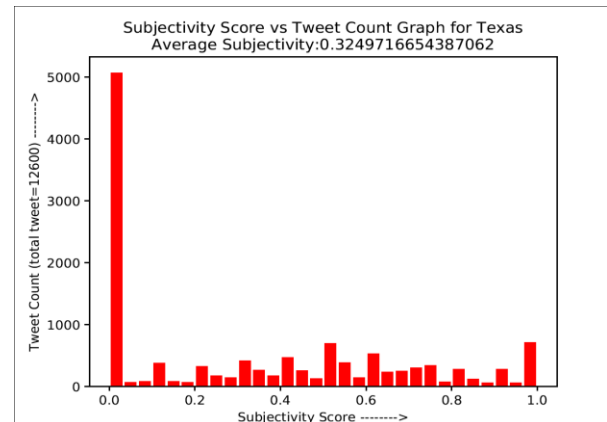
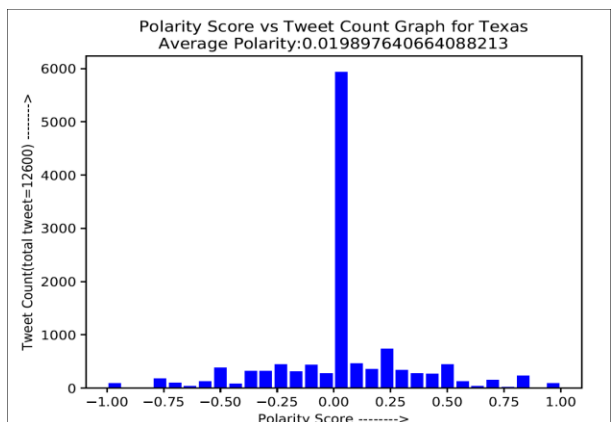
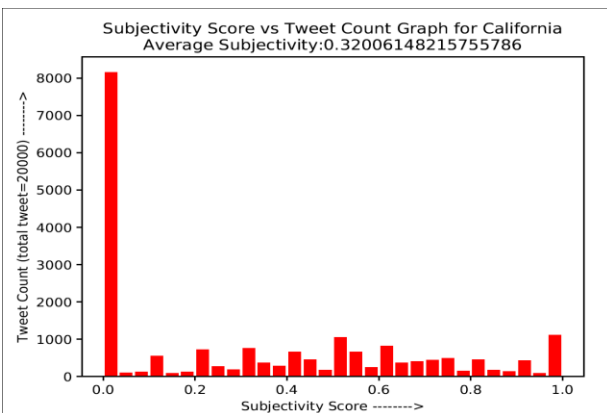
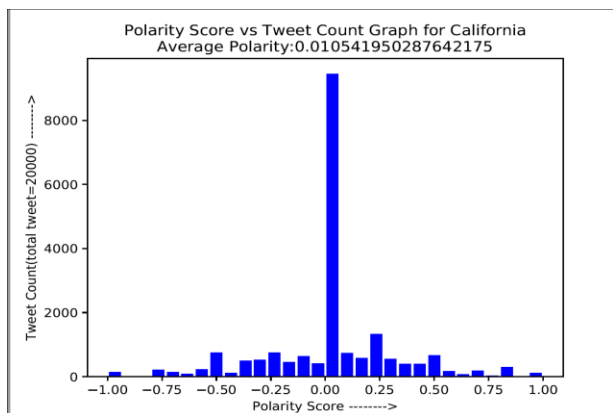
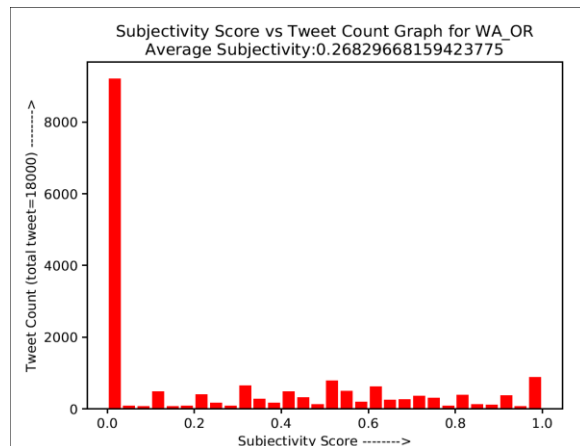
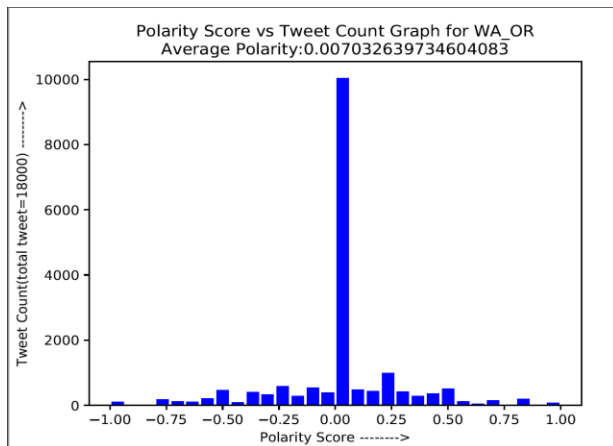
Geographic variations

To understand the variations depending on their geographic locations, we collect tweets from specific states using latitude and longitude and then filter out trump specific tweets with the help of 'if' statements.

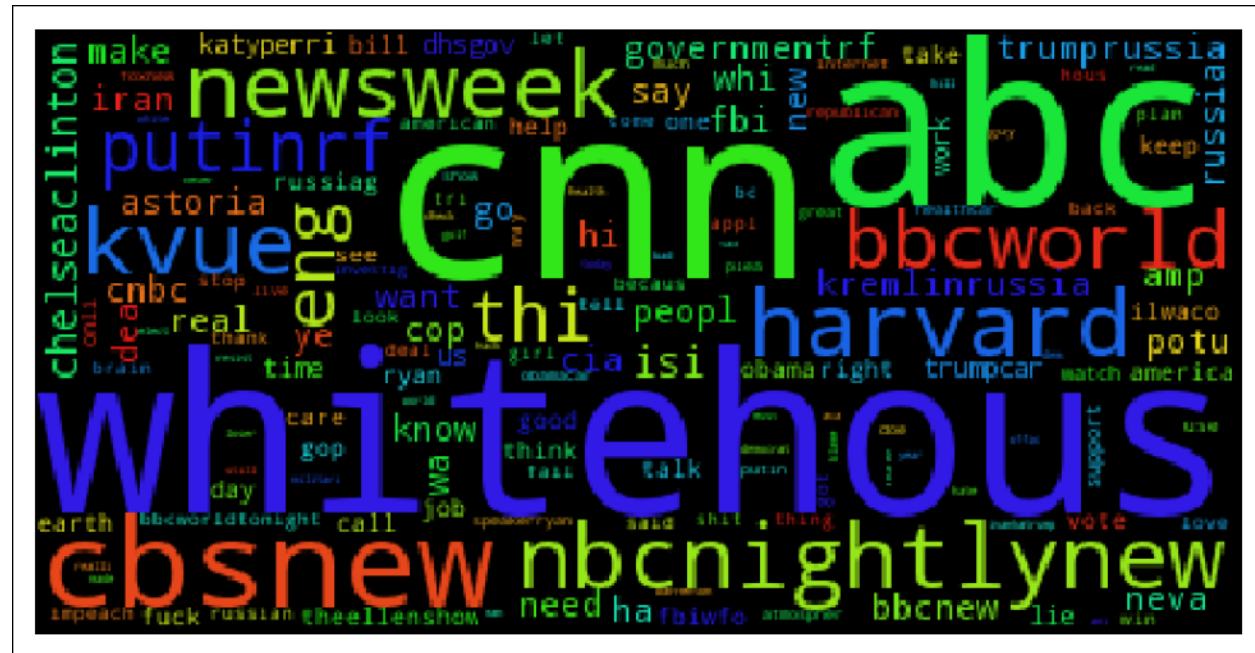
Sentiment Analysis: We used TextBlob on the collected data to generate polarity and subjectivity scores for each tweet. We have used these scores to plot histograms for a better understanding.

The Polarity scores are measured in a scale of -1 to + 1, where -1 implies highly negative opinions, and subjectivity is measured in scale of 0 to 1, measures how many emotional words are used in a document

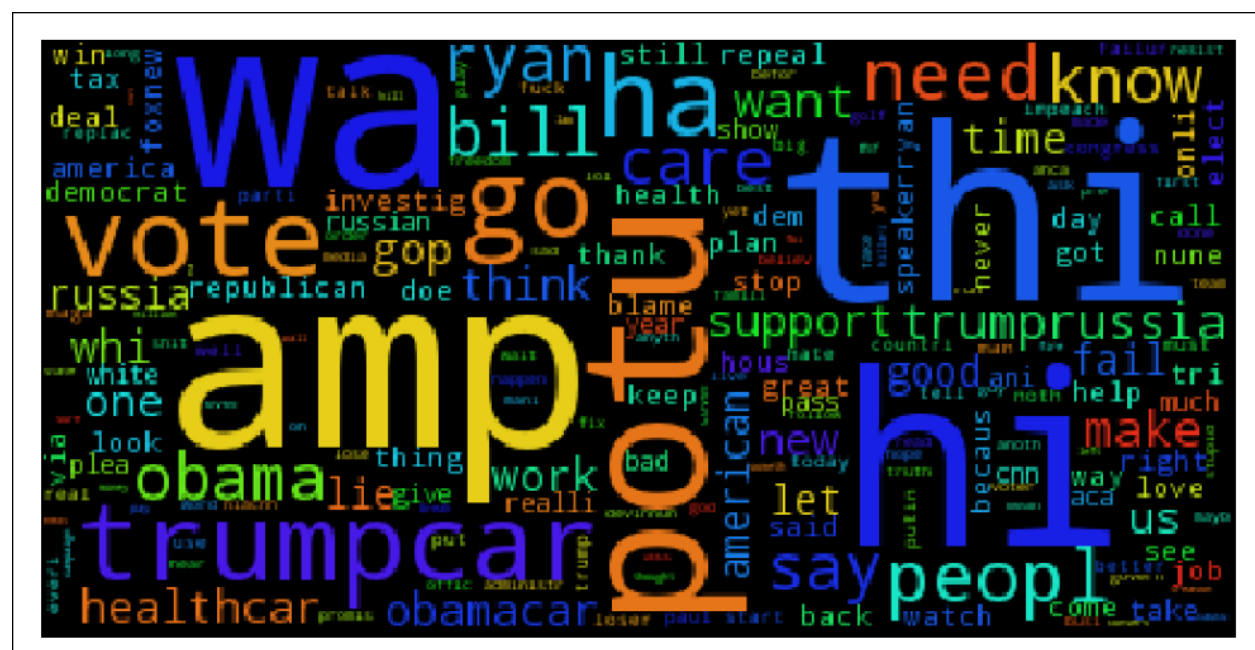




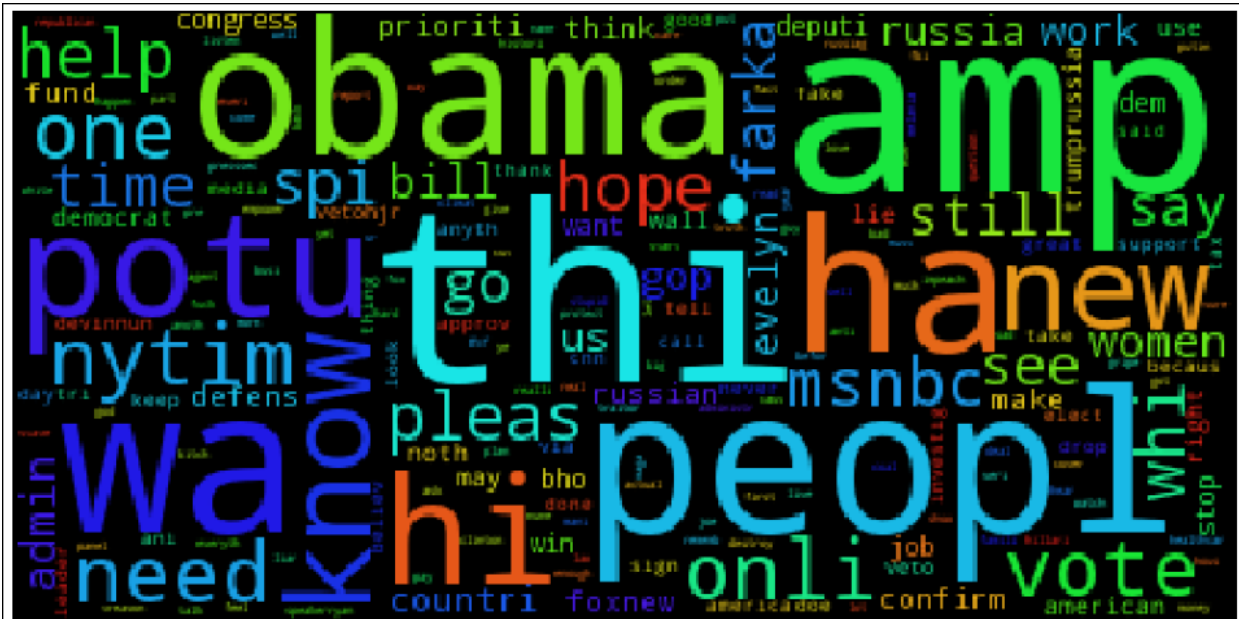
Word Cloud for Washington and Oregon



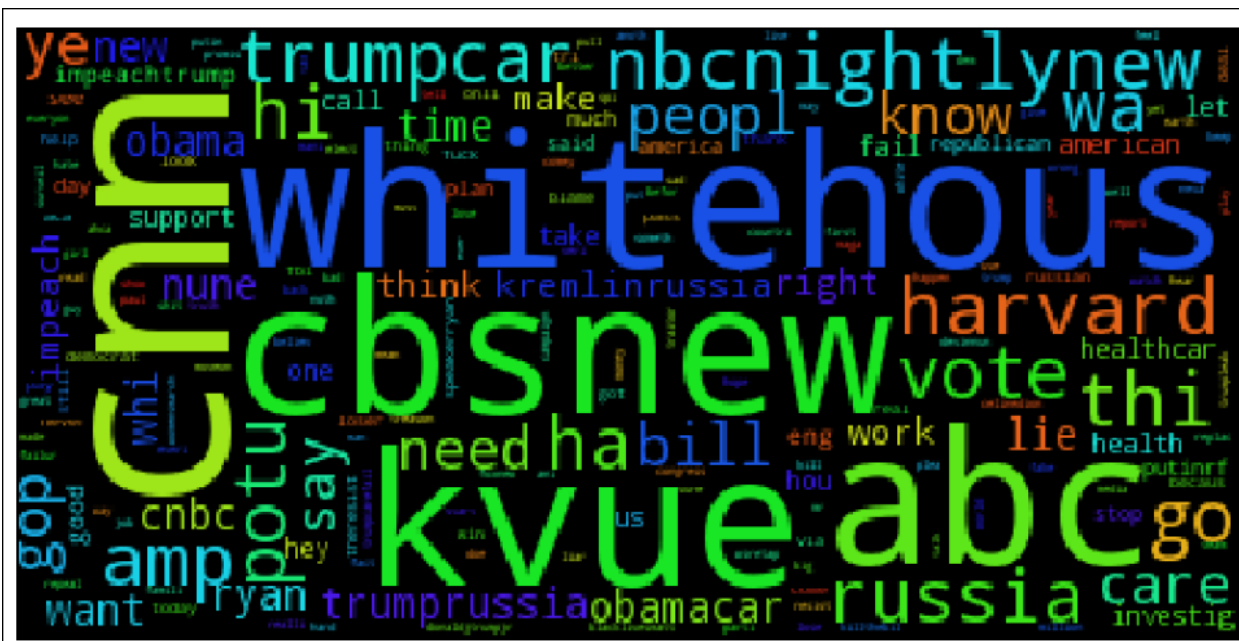
Word cloud for Texas



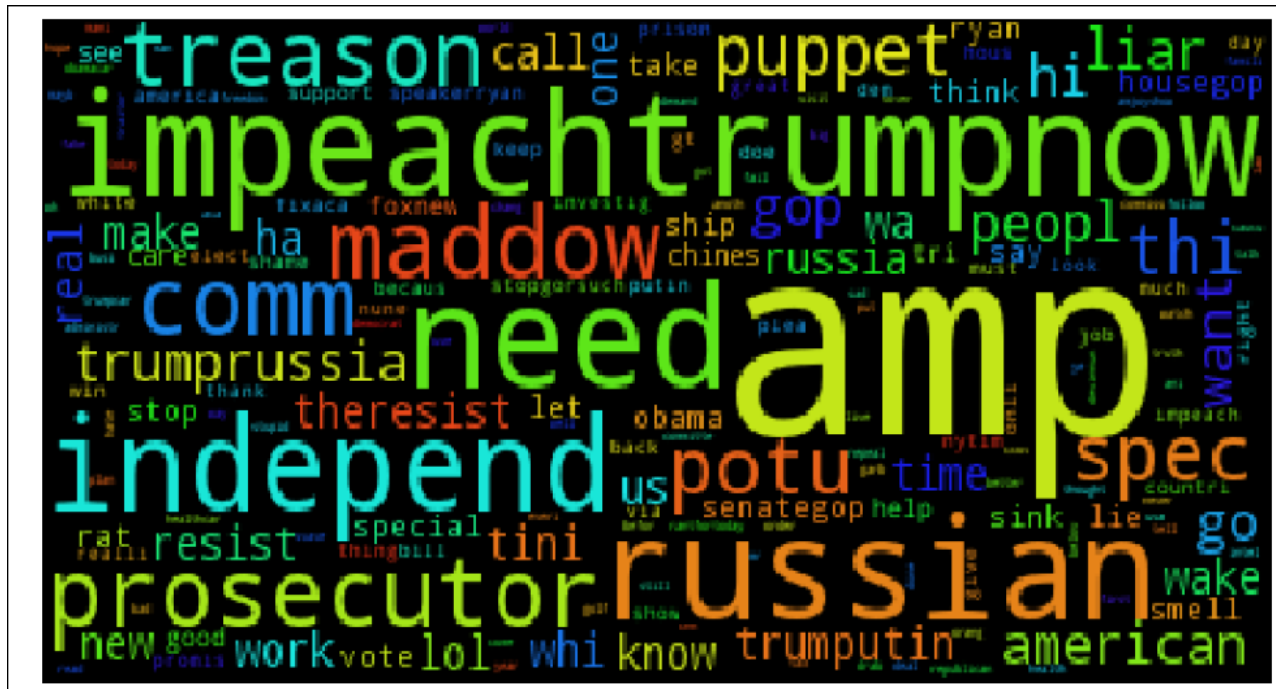
Word Cloud for Georgia and Florida



Word cloud for California



Word cloud for Colorado



Topic modeling

We perform topic analysis with the help of NMF (non negative matrix factorization) to find out the popular topics. NMF is a group of algorithms in multivariate analysis and linear algebra where a matrix is factorized into 2 matrices, with a property where all three have no negative elements. We have loaded the inaugural text corpus from NLTK, we have vectorized the text after removing the stop words and generated the term matrix and create a list of unique words which will help us get the topic.

NMF

Topic Modelling for All 5 States

Topic 1: lawrence agree resistance merriannedean damning surely

Topic 2: tomi lahren suspended job fbi investigated

Topic 3: inappropriate gone photo truly viral love

Topic 4: approved torrainewalker december funds requested congress

Topic 5: departments shame bidding local won aclu

Topic 6: kaepernick wheels meals colin million starving

Topic 7: people wiretapped whoa turns arrested tower

Topic 8: vote legitimate clouds suspicion gray finds

Topic 9: memes point love gabriel trbq confucius

Topic 10: money laundering russian linked lent scam

Topic Modelling for California

- Topic 1: cnn abc harvard kvue nbcnightlynews cbsnews
- Topic 2: trumprussia theresistance trumpleaks blacklivesmatter moveon womensmarch
- Topic 3: hey pay impeachtrump russia ties campaign
- Topic 4: trumpcare vote dead ahca obamacare ryancare
- Topic 5: amp potus people like know good
- Topic 6: yes neva cnn real abc cbs
- Topic 7: president right mr thank love time
- Topic 8: cnbc kremlinrussia cnn kvue abc nbcnightlynews
- Topic 9: nunes investigation fbi cooperating surveillance camp
- Topic 10: care health gop ryan vote plan

Topic Modelling for Washington and Oregon

- Topic 1: chelseaclinton bbcworld newsweek bbcnews kvue nbcnightlynews
- Topic 2: harvard kvue cnn abc nbcnightlynews cbsnews
- Topic 3: cnbc kremlinrussia kvue harvard cnn abc
- Topic 4: katyperry dhsgov theellenshow cia fbiwfo zeldawilliams
- Topic 5: isis claim ilwaco dea astoria claims
- Topic 6: iran bbcworld newsweek bbcworldtonight harvard cbsnews
- Topic 7: trumprussia russiagate trending keeping bc thread
- Topic 8: yes real neva cbs cnn nbc
- Topic 9: governmentrf eng putinrf fbi newsweek abc
- Topic 10: amp like president potus people trumpcare

Topic Modelling for Texas

- Topic 1: amp repeal obamacare replace said dems
- Topic 2: president mr thank time obama god
- Topic 3: sees online surge african pride store
- Topic 4: trumprussia right fakehealthcare devinnunes theresistance russiagate
- Topic 5: trumpcare vote gop speakerryan dead healthcare
- Topic 6: like looks look people mean going
- Topic 7: potus niacnn foxnews gloriaborger ac speakerryan
- Topic 8: care health people obama vote american
- Topic 9: ryan paul house obamacare speaker blame
- Topic 10: america great make good deal healthcare

Topic Modelling for Georgia and Florida

- Topic 1: deputy spying evelyn farkas defense bho
- Topic 2: hope veto vetohjres npr kathymaddox justmeelc
- Topic 3: priorities gop voted drop wall democratic
- Topic 4: people help country know congress gripe
- Topic 5: president mr god america work bless

Topic 6: amp russia tell dems news time
 Topic 7: nytimes remember thanks wednesdaywisdom times time
 Topic 8: potus like media stop work foxnews
 Topic 9: women empowerment panel womenshistorymonth melania courage
 Topic 10: trumprussia russiagate house russia nunes investigation

Topic Modelling for Colorado

Topic 1: amp spec puppet treasonous comm russian
 Topic 2: sinking smell chinese rat ship lol
 Topic 3: housegop senategop wake russians theresistance gop
 Topic 4: commission amjoyshow special americans maddow independent
 Topic 5: potus prison garb drab beloved visits
 Topic 6: trumprussia resist stopgorsuch fixaca real russiangop
 Topic 7: people like know amp work care
 Topic 8: demand special intel committee comm prosecutor
 Topic 9: read nytimes sadd constitution amd try
 Topic 10: president obama mr credit veto sjresolution

Gensim

We use LDA with genism, LDA is a generative model where each topic is a distribution over words and each word is drawn from one of those topics. Gensim is used to compute the distribution condition of documents based on probabilistic approach to get the most repetitive topics.

NO of Topics/ Model Perplexity	All States	California	Colorado	Washtington Orogan	Georgia Florida	Texas
3	- 8.040584413	- 8.378932865	- 8.032686904	- 7.259173422	- 8.461435135	- 8.752490902
5	- 7.999287605	- 8.288832854	- 8.119599067	- 7.422619313	- 8.706308714	- 8.938879828
7	-7.93840008	- 7.173832849	- 8.199627429	- 7.449814532	-8.8721239	- 9.093444249

For all 5 Regions

500 Passes and 3 Topics

Model Perplexity: -8.04058441292
 (0, u'0.007*"fbi" + 0.006*"u2026" + 0.006*"got" + 0.005*"russian" +
 0.005*"investigated" + 0.004*"money" + 0.004*"gorsuch" + 0.004*"job" +
 0.004*"trumprussia" + 0.004*"vote"')(1, u'0.004*"people" +
 0.003*"white" + 0.003*"would" + 0.003*"u2026" + 0.003*"https" +
 0.003*"tower" + 0.003*"russians" + 0.003*"get" + 0.003*"news" +
 0.003*"know"')(2, u'0.014*"agree" + 0.014*"lawrence" + 0.008*"love" +
 0.006*"obama" + 0.006*"point" + 0.005*"president" + 0.005*"memes" +
 0.005*"truly" + 0.005*"viral" + 0.005*"photo"')

500 Passes and 5 Topics

Model Perplexity: -7.99928760498

(0, u'0.007*"president" + 0.005*"vote" + 0.005*"million" + 0.005*"nasa" + 0.005*"remember" + 0.005*"kaepernick" + 0.005*"year" + 0.005*"wants" + 0.004*"colin" + 0.004*"last")(1, u'0.005*"russians" + 0.005*"people" + 0.005*"tower" + 0.004*"turns" + 0.004*"wiretapped" + 0.004*"guess" + 0.004*"arrested" + 0.004*"whoa" + 0.004*"shaunking" + 0.004*"c\u2026")(2, u'0.011*"love" + 0.009*"point" + 0.008*"memes" + 0.008*"truly" + 0.008*"viral" + 0.008*"photo" + 0.008*"gone" + 0.008*"inappropriate" + 0.007*"russia" + 0.007*"c\u2026")(3, u'0.021*"agree" + 0.020*"lawrence" + 0.009*"tomi" + 0.009*"lahren" + 0.008*"suspended" + 0.008*"fbi" + 0.008*"job" + 0.007*"got" + 0.007*"investigated" + 0.006*"find")(4, u'0.008*"obama" + 0.006*"approved" + 0.006*"december" + 0.006*"lie" + 0.006*"may" + 0.006*"requested" + 0.006*"funds" + 0.006*"torrainewalker" + 0.006*"congress" + 0.006*"president")

500 passes and 7 Topics

Model Perplexity: -7.93840007988

(0, u'0.041*"lawrence" + 0.040*"agree" + 0.010*"kaepernick" + 0.009*"colin" + 0.008*"last" + 0.008*"wheels" + 0.008*"meals" + 0.008*"cut" + 0.007*"wants" + 0.007*"mikel")(1, u'0.006*"hope" + 0.005*"lied" + 0.005*"one" + 0.005*"oh" + 0.004*"really" + 0.004*"kylegriffin" + 0.004*"took" + 0.004*"day" + 0.003*"first" + 0.003*"cuts")(2, u'0.013*"got" + 0.012*"investigated" + 0.012*"tomi" + 0.011*"lahren" + 0.011*"job" + 0.011*"find" + 0.011*"media" + 0.011*"suspended" + 0.010*"police" + 0.010*"tybroke sign")(3, u'0.014*"photo" + 0.007*"white" + 0.007*"shit" + 0.007*"office" + 0.007*"would" + 0.006*"thought" + 0.006*"day" + 0.006*"case" + 0.006*"ivanka" + 0.006*"take")(4, u'0.008*"vote" + 0.007*"remember" + 0.007*"president" + 0.007*"show" + 0.007*"presidency" + 0.006*"on\u2026" + 0.006*"care" + 0.006*"impeachment" + 0.006*"finds" + 0.006*"support")(5, u'0.013*"gone" + 0.010*"russia" + 0.007*"probe" + 0.006*"c\u2026" + 0.006*"election" + 0.006*"news" + 0.006*"kremlin" + 0.005*"gorsuch" + 0.005*"defend" + 0.005*"ties")(6, u'0.020*"love" + 0.015*"point" + 0.014*"memes" + 0.014*"truly" + 0.014*"inappropriate" + 0.013*"viral" + 0.009*"people" + 0.008*"c\u2026" + 0.008*"fact" + 0.008*"turns")(7, u'0.011*"president" + 0.010*"approved" + 0.010*"funds" + 0.010*"december" + 0.010*"torrainewalker" + 0.010*"may" + 0.010*"lie" + 0.010*"russian" + 0.010*"requested" + 0.010*"congress")(8, u'0.007*"campaign" + 0.007*"ukraine" + 0.006*"fbi" + 0.006*"russia" + 0.005*"new" + 0.005*"investigation" + 0.005*"attack" + 0.005*"paul" + 0.005*"investigating" + 0.005*"manafort")

Florida and Georgia

500 Passes and 3 Topics

Model Perplexity: -8.46143513536

(0, u'0.004*"please" + 0.004*"still" + 0.004*"hope" + 0.004*"priorities" + 0.004*"see" + 0.004*"bill" + 0.004*"president" + 0.003*"vetohjres" + 0.003*"veto" + 0.003*"voted")(1, u'0.004*"people" + 0.004*"know" + 0.003*"anything" + 0.003*"done" + 0.003*"country" +

0.003*"one" + 0.003*"help" + 0.003*"amp" + 0.003*"nothing" +
0.002*"congress"')(2, u'0.004*"obama" + 0.004*"farkas" +
0.004*"evelyn" + 0.004*"admin" + 0.004*"msnbc" + 0.004*"spying" +
0.004*"defense" + 0.004*"deputy" + 0.004*"bho" + 0.003*"confirms"')

500 Passes and 5 Topics

Model Perplexity: -8.70630871426

(0, u'0.004*"president" + 0.003*"get" + 0.002*"fuck" + 0.002*"nytimes"
+ 0.002*"mr" + 0.001*"nunes" + 0.001*"hard" + 0.001*"trumprussia" +
0.001*"u" + 0.001*"say"')(1, u'0.003*"see" + 0.003*"media" +
0.003*"like" + 0.002*"amp" + 0.002*"nytimes" + 0.002*"president" +
0.002*"us" + 0.002*"climate" + 0.002*"think" + 0.002*"potus"')(2,
u'0.008*"people" + 0.006*"priorities" + 0.004*"nothing" +
0.004*"country" + 0.004*"may" + 0.004*"one" + 0.004*"anything" +
0.004*"gop" + 0.004*"know" + 0.003*"voted"')(3, u'0.003*"president" +
0.002*"amp" + 0.002*"bless" + 0.002*"get" + 0.002*"trumprussia" +
0.002*"needs" + 0.002*"bad" + 0.002*"would" + 0.002*"god" +
0.002*"russian"')(4, u'0.006*"farkas" + 0.006*"evelyn" + 0.006*"obama"
+ 0.006*"admin" + 0.006*"hope" + 0.006*"spying" + 0.006*"defense" +
0.005*"please" + 0.005*"msnbc" + 0.005*"deputy"')

500 Passes and 7 Topics

Model Perplexity: -8.87212390033

(0, u'0.003*"like" + 0.002*"get" + 0.002*"via" + 0.002*"president" +
0.002*"amp" + 0.002*"things" + 0.002*"enough" + 0.002*"see" +
0.002*"need" + 0.002*"u"')(1, u'0.008*"people" + 0.006*"know" +
0.006*"help" + 0.005*"anything" + 0.005*"congress" + 0.005*"done" +
0.005*"one" + 0.005*"nothing" + 0.005*"country" + 0.004*"amp"')(2,
u'0.003*"president" + 0.002*"russia" + 0.002*"media" + 0.002*"us" +
0.002*"amp" + 0.002*"another" + 0.002*"maga" + 0.002*"said" +
0.002*"working" + 0.002*"like"')(3, u'0.009*"priorities" +
0.008*"please" + 0.008*"veto" + 0.008*"vetohjres" + 0.007*"hope" +
0.007*"bill" + 0.007*"still" + 0.005*"funding" + 0.005*"voted" +
0.005*"gop"')(4, u'0.003*"potus" + 0.002*"would" + 0.002*"believe" +
0.002*"cheif" + 0.002*"bad" + 0.002*"news" + 0.002*"amp" +
0.002*"remember" + 0.002*"mr" + 0.002*"real"')(5, u'0.004*"president"
+ 0.002*"amp" + 0.002*"bless" + 0.002*"like" + 0.002*"god" +
0.002*"nytimes" + 0.002*"tell" + 0.002*"melania" + 0.002*"never" +
0.002*"free"')(6, u'0.009*"obama" + 0.008*"spying" + 0.008*"admin" +
0.008*"defense" + 0.008*"farkas" + 0.008*"deputy" + 0.008*"evelyn" +
0.008*"bho" + 0.007*"confirms" + 0.007*"msnbc"')

Colorado

500 Passes and 3 Topics

Model Perplexity: -8.03268690435

(0, u'0.007*"gop" + 0.007*"wake" + 0.007*"russians" +
0.007*"senategop" + 0.007*"housegop" + 0.007*"americans" + 0.007*"u" +
0.007*"r" + 0.006*"theresistance" + 0.004*"impeachtrumpon"')(1,

u'0.008*"potus" + 0.008*"trumputin" + 0.008*"liar" +
0.008*"\U0001f602\U0001f602" + 0.008*"want" + 0.008*"lol" +
0.008*"tiny" + 0.007*"\u2026" + 0.007*"go" + 0.007*"ship"')(2,
u'0.005*"trumprussia" + 0.003*"resist" + 0.003*"president" +
0.003*"us" + 0.003*"nytimes" + 0.003*"read" + 0.003*"like" +
0.003*"house" + 0.003*"much" + 0.003*"white"')

500 Passes and 5 Topics

Model Perplexity: -8.11959906725

(0, u'0.042*"amp" + 0.033*"impeachtrumpnow" + 0.033*"independent" +
0.033*"prosecutor" + 0.030*"comm" + 0.029*"treasonous" +
0.029*"maddow\u2026" + 0.029*"puppet" + 0.029*"spec" +
0.028*"russian"')(1, u'0.006*"nytimes" + 0.004*"read" +
0.003*"trumprussia" + 0.003*"us" + 0.003*"rantfortoday" + 0.002*"ccot" +
0.002*"potus" + 0.002*"people" + 0.002*"like" + 0.002*"nunes"')(2,
u'0.010*"commission" + 0.008*"special" + 0.003*"work" + 0.002*"might" +
0.002*"russia" + 0.002*"trumprussia" + 0.002*"people" +
0.002*"trumpet" + 0.002*"would" + 0.002*"obamacare"')(3,
u'0.003*"stop" + 0.002*"get" + 0.002*"respect" + 0.002*"please" +
0.002*"president" + 0.002*"win" + 0.002*"ivanka" + 0.002*"fuck" +
0.002*"job" + 0.002*"people"')(4, u'0.003*"like" + 0.003*"never" +
0.003*"promises" + 0.003*"president" + 0.003*"trumprussia" +
0.003*"get" + 0.003*"foxnews" + 0.003*"know" + 0.003*"elected" +
0.002*"trumptastic"')

500 Passes and 7 Topics

Model Perplexity: -8.19962742858

(0, u'0.014*"special" + 0.010*"amjoyshow\u2026" + 0.010*"maddow" +
0.008*"americans" + 0.008*"r" + 0.008*"intel" + 0.007*"call" +
0.007*"committee" + 0.007*"demand" + 0.004*"\u2026"')(1, u'0.012*"u" +
0.012*"russians" + 0.012*"senategop" + 0.012*"wake" + 0.011*"gop" +
0.011*"housegop" + 0.011*"r" + 0.011*"americans" +
0.010*"theresistance" + 0.007*"impeachtrumpnow"')(2,
u'0.009*"trumprussia" + 0.009*"real" + 0.008*"potus" + 0.008*"resist" +
0.006*"stopgorsuch" + 0.006*"nytimes" + 0.006*"fixaca" +
0.006*"liar" + 0.006*"trumputin" + 0.005*"tiny"')(3, u'0.003*"think" +
0.003*"news" + 0.003*"evil" + 0.003*"approval" + 0.003*"got" +
0.003*"president" + 0.003*"devinnunes" + 0.003*"like" + 0.002*"time" +
0.002*"rino"')(4, u'0.052*"amp" + 0.040*"prosecutor" +
0.040*"independent" + 0.037*"comm" + 0.036*"impeachtrumpnow" +
0.036*"treasonous" + 0.036*"maddow\u2026" + 0.036*"puppet" +
0.036*"spec" + 0.033*"russian"')(5, u'0.004*"jobs" + 0.003*"back" +
0.003*"failing" + 0.003*"like" + 0.003*"god" + 0.002*"would" +
0.002*"turn" + 0.002*"give" + 0.002*"foxandfriends" +
0.002*"people"')(6, u'0.003*"impeach" + 0.003*"get" + 0.003*"wh" +
0.003*"women" + 0.002*"really" + 0.002*"going" + 0.002*"future" +
0.002*"gt" + 0.002*"fam" + 0.002*"nunes"')

California

500 Passes and 3 Topics

Model Perplexity: -8.37893286543

```
(0, u' 0.005*"nunes" + 0.003*"amp" + 0.003*"russia" +  
0.003*"investigation" + 0.002*"devinnunes" + 0.002*"obama" +  
0.002*"fbi" + 0.002*"potus" + 0.002*"presi  
dent" + 0.002*"russian"')  
(1, u'0.006*"trumpcare" + 0.004*"u" + 0.003*"president" +  
0.003*"bill" + 0.003*"care" + 0.003*"trumprussia" +  
0.003*"obamacare" + 0.003*"amp" + 0.003*"get"  
+ 0.003*"vote"')  
(2, u'0.020*"cnn" + 0.018*"abc" + 0.018*"whitehouse" +  
0.018*"kvue" + 0.018*"cbsnews" + 0.018*"nbcnightlynews" +  
0.018*"harvard" + 0.009*"yes" + 0.008*"cnbc  
" + 0.008*"e"')
```

500 Passes and 5 Topics

Model Perplexity: -8.288832853873

```
(0, u'0.007*"foxnews" + 0.006*"russia" + 0.006*"niacnn" +  
0.004*"ac" + 0.004*"adamschiffca" + 0.004*"keep" +  
0.004*"gloriaborger" + 0.004*"must" + 0.004*"yes" +  
0.004*"trumpru  
ssia"') (1, u'0.008*"president" + 0.008*"potus" + 0.006*"good" +  
0.005*"cnn" + 0.005*"please" + 0.005*"like" + 0.005*"news" +  
0.005*"mr" + 0.005*"never" + 0.005*"obama"') (2, u'  
0.007*"u" + 0.006*"amp" + 0.006*"need" + 0.006*"speakerryan" +  
0.006*"let" + 0.006*"gop" + 0.005*"get" + 0.005*"bill" +  
0.005*"want" + 0.005*"trumpcare"')  
(3, u'0.007*"campaign" + 0.007*"ukraine" + 0.006*"fbi" +  
0.006*"russia" + 0.005*"new" + 0.005*"investigation" +  
0.005*"attack" + 0.005*"paul" + 0.005*"investigating" +  
0.005*" Manafort"') (4, '0.02*"amp" + 0.03*"independent" +  
0.03*"prosecutor" + 0.022*"impeachtrumpnow" + 0.021*"comm" +  
0.01*"need" + 0.01*"russian" + 0.02*"treasonous" +  
0.02*"maddow..." + 0.02*"spec"')
```

500 Passes and 7 Topics

Model Perplexity: -7.17383284923

```
(0, u'0.004*"loser" + 0.004*"trumpcarefail" + 0.003*"yeah" +  
0.003*"air" + 0.003*"ivanka" + 0.003*"like" + 0.002*"flynn" +  
0.002*"bring" + 0.002*"keeps" + 0.002*"prison"') (1,  
u'0.004*"gloriaborger" + 0.004*"potus" + 0.004*"caucus" +  
0.004*"niacnn" + 0.003*"care" + 0.003*"u26a1\u201c" +  
0.003*"dumb" + 0.003*"ac" + 0.003*"joke" + 0.003*"obama"')
```

```
(2, u'0.004*"nunes" + 0.003*"amp" + 0.003*"russia" +
0.003*"investigation" + 0.002*"president" + 0.002*"fbi" +
0.002*"obama" + 0.002*"potus" + 0.002*"russian" +
0.002*"devinnunes"')(3, u'0.007*"trumpcare" + 0.004*"u" +
0.004*"trumprussia" + 0.004*"bill" + 0.003*"care" + 0.003*"vote"
+ 0.003*"president" + 0.003*"amp" + 0.003*"theresistance" +
0.003*"impeach"')(4, u'0.020*"cnn" + 0.018*"whitehouse" +
0.018*"abc" + 0.017*"kvue" + 0.017*"cbsnews" +
0.017*"nbcnightlynews" + 0.017*"harvard" + 0.011*"cnbc" +
0.010*"e" + 0.010*"kremlinrussia"')(5, u'0.004*"nunes" +
0.003*"amp" + 0.003*"russia" + 0.003*"investigation" +
0.002*"president" + 0.002*"fbi" + 0.002*"obama" + 0.002*"potus"
+ 0.002*"russian" + 0.002*"devinnunes"')(6, u'0.009*"trending" +
0.004*"dead" + 0.004*"potus" + 0.004*"easy" + 0.003*"loser" +
0.003*"trumpgenius" + 0.003*"neron" + 0.003*"cali" +
0.003*"alverezamanda" + 0.002*"navy"')(7, u'0.007*"ivankatrump"
+ 0.007*"donaldjtrumpjr" + 0.006*"erictrump" + 0.005*"chris" +
0.003*"theresistance" + 0.003*"idiot" + 0.003*"media" +
0.003*"ties" + 0.003*"art" + 0.003*"course"')
```

Washington and Oregon

500 Passes and 3 Topics

Model Perplexity: -7.25917342169

```
(0, u'0.006*"trumprussia" + 0.004*"russiagate" + 0.004*"amp" +
0.002*"u" + 0.002*"keeping" + 0.002*"people" + 0.002*"U0001f447" +
0.002*"potus" + 0.002*"trending" + 0.002*"like"')(1, u'0.003*"like" +
0.003*"president" + 0.003*"potus" + 0.002*"amp" + 0.002*"news" +
0.002*"get" + 0.002*"ivankatrump" + 0.002*"trumpcare" + 0.002*"us" +
0.002*"donaldjtrumpjr"')(2, u'0.023*"cnn" + 0.023*"abc" +
0.022*"nbcnightlynews" + 0.022*"cbsnews" + 0.022*"whitehouse" +
0.022*"kvue" + 0.021*"newsweek" + 0.021*"harvard" + 0.016*"bbcworld" +
0.015*"putinrf"')
```

500 Passes and 5 Topics

Model Perplexity: -7.4226193126

```
(0, u'0.006*"bc" + 0.004*"job" + 0.003*"find" + 0.003*"light" + 0.003*"winning" +
0.003*"camp" + 0.003*"gonna" + 0.003*"days" + 0.002*"tired" + 0.002*"rilea"')(1,
u'0.003*"gov" + 0.003*"iphone" + 0.003*"un" + 0.003*"jobs" + 0.003*"potus" + 0.003*"amp" +
0.002*"power" + 0.002*"u" + 0.002*"dead" + 0.002*"work"')(2, u'0.006*"apple" +
0.004*"amp" + 0.004*"president" + 0.004*"people" + 0.004*"like" + 0.004*"u" + 0.003*"ryan"
+ 0.003*"america" + 0.003*"google" + 0.003*"would"')(3, u'0.006*"dev" + 0.005*"guard" +
0.005*"coast" + 0.004*"radio" + 0.003*"magatrump" + 0.003*"science" + 0.003*"watching" +
0.003*"wow" + 0.003*"pink" + 0.003*"realdonaldtrump\u2026"')(4, u'0.014*"katyperry" +
0.013*"dhsgov" + 0.013*"theellenshow" + 0.013*"fbiwfo" + 0.013*"cia" +
```


0.005*"zeldawilliams" + 0.005*"donaldjtrumpjr" + 0.005*"ivankatrump" + 0.004*"erictrump" + 0.004*"eugenepolice")

500 Passes and 7 Topics

Model Perplexity: -7.44981453177

(0, u'0.015*"russiagate" + 0.015*"trumprussia" + 0.010*"bc" + 0.009*"keeping" + 0.009*"U0001f447" + 0.005*"job" + 0.004*"renealane" + 0.004*"jobs" + 0.004*"james" + 0.003*"hiring"')(1, u'0.007*"china" + 0.006*"control" + 0.005*"please" + 0.005*"mind" + 0.005*"sattilite" + 0.005*"love" + 0.005*"war" + 0.004*"realdonaldtrump\u2026" + 0.004*"stop" + 0.004*"usa"')(2, u'0.004*"amp" + 0.003*"pink" + 0.003*"trying" + 0.003*"us" + 0.003*"like" + 0.003*"russia" + 0.003*"climate" + 0.003*"devinnunes" + 0.003*"year" + 0.003*"freedom"')(3, u'0.031*"cnn" + 0.030*"abc" + 0.030*"newsweek" + 0.030*"nbcnightlynews" + 0.030*"cbsnews" + 0.030*"whitehouse" + 0.029*"kvue" + 0.027*"harvard" + 0.023*"bbcworld" + 0.022*"putinrf"')(4, u'0.004*"fucking" + 0.003*"friends" + 0.003*"happen" + 0.003*"wtf" + 0.003*"sure" + 0.003*"plane" + 0.003*"processor" + 0.003*"jail" + 0.003*"entire" + 0.003*"elections"')(5, u'0.015*"fbiwfo" + 0.008*"zeldawilliams" + 0.005*"hillaryclinton\u2026" + 0.005*"day" + 0.004*"god" + 0.004*"president" + 0.003*"nike" + 0.003*"whoopigoldberg" + 0.003*"said" + 0.003*"wait"')(6, u'0.009*"trending" + 0.004*"dead" + 0.004*"potus" + 0.004*"easy" + 0.003*"loser" + 0.003*"trumpgenius" + 0.003*"neron" + 0.003*"cali" + 0.003*"alverezamanda" + 0.002*"navy"')

Texas

500 Passes and 3 Topics

Model Perplexity: -8.75249090173

(0, u'0.004*"amp" + 0.004*"president" + 0.004*"get" + 0.003*"trumpcare" + 0.003*"people" + 0.003*"potus" + 0.003*"ryan" + 0.003*"like" + 0.003*"bill" + 0.003*"u"')(1, u'0.003*"read" + 0.003*"lol" + 0.002*"trumprussia" + 0.002*"order" + 0.002*"coal" + 0.002*"U0001f610" + 0.002*"via" + 0.002*"swamp" + 0.002*"supporter" + 0.002*"war"')(2, u'0.006*"sees" + 0.006*"age" + 0.006*"african" + 0.006*"store" + 0.006*"surge" + 0.006*"online" + 0.006*"pride" + 0.003*"fake" + 0.003*"god" + 0.003*"niacnn")

500 Passes and 5 Topics

Model Perplexity: -8.93887982783

(0, u'0.003*"trumprussia" + 0.003*"get" + 0.003*"read" + 0.003*"amp" + 0.002*"potus" + 0.002*"nytimes" + 0.002*"trumpcare" + 0.002*"blame" + 0.002*"supporter" + 0.002*"house"')(1, u'0.007*"sees" + 0.007*"age" + 0.007*"african" + 0.007*"store" + 0.007*"surge" + 0.007*"online" + 0.007*"pride" + 0.003*"people" + 0.002*"potus" + 0.002*"gloriaborger"')(2, u'0.005*"president" + 0.005*"amp" + 0.004*"bill" + 0.004*"trumpcare" + 0.004*"care" + 0.004*"obamacare" + 0.004*"healthcare" + 0.003*"u" + 0.003*"like" + 0.003*"repeal"')(3, u'0.004*"russia" + 0.003*"news" + 0.003*"ties" +

0.003*"via" + 0.002*"investigation" + 0.002*"u26a1\u002f" + 0.002*"youtube" +
0.002*"devinnunes" + 0.002*"thank" + 0.002*"repspeier")(4, u'0.003*"stupid" +
0.002*"obama" + 0.002*"says" + 0.002*"trumprussia" + 0.002*"via" + 0.002*"idiot" +
0.002*"trumpcarefail" + 0.002*"russians" + 0.002*"bitch" + 0.002*"world")

500 Passes and 7 Topics

Model Perplexity: -9.09344424943

(0, u'0.014*"african" + 0.014*"surge" + 0.014*"online" + 0.014*"pride" + 0.003*"executive" +
0.003*"clean" + 0.003*"cannot" + 0.003*"crazy" + 0.002*"pro" + 0.002*"repadamschiff")(1,
u'0.003*"corrupt" + 0.003*"impeach" + 0.003*"worry" + 0.003*"lago" + 0.003*"mar" +
0.003*"approval" + 0.003*"seriously" + 0.003*"days" + 0.002*"pray" + 0.002*"use")(2,
u'0.004*"nytimes" + 0.003*"us" + 0.003*"judgejeanine" + 0.003*"amp" + 0.003*"failing" +
0.003*"climate" + 0.003*"leader" + 0.003*"face" + 0.003*"war" + 0.003*"gop")(3,
u'0.003*"ha" + 0.003*"small" + 0.003*"finally" + 0.003*"asking" + 0.002*"follow" +
0.002*"hands" + 0.002*"son" + 0.002*"demand" + 0.002*"deserve" + 0.002*"answer")(4,
u'0.007*"president" + 0.006*"amp" + 0.006*"sees" + 0.006*"store" + 0.005*"potus" +
0.005*"ryan" + 0.005*"get" + 0.005*"people" + 0.004*"u" + 0.004*"bill")(5,
u'0.003*"maxinewaters" + 0.003*"world" + 0.003*"jobs" + 0.003*"defeat" + 0.003*"mexico" +
0.003*"russian" + 0.003*"ties" + 0.002*"fuck" + 0.002*"constitution" + 0.002*"paying")(6,
u'0.011*"age" + 0.006*"trumprussia" + 0.004*"trumpafamousquote" + 0.004*"via" +
0.003*"president" + 0.003*"U0001f610" + 0.003*"wall" + 0.003*"winning" + 0.003*"russia" +
0.003*"like")

Insights and Conclusion

Colorado has the most positive sentiments towards President Trump where as other states i.e California, Texas, Washington, Oregon, Florida and Georgia show an almost equal number of positive and negative sentiments towards President Trump's agenda.

President Trumps Russian allegation and Obamas Wire Tapping incident were the most tweeted topics. The revoking of Obama care was a popular topic in Texas as seen from word cloud and topic modelling. With regards to the hashtags being used GOP (Grand Old Party, referring to Republican Party) was under discussion.

Despite the geographic distance between the states. California, Texas and Colorado had very similar topics regarding Russian hacking allegations and Trump Care.

After completing the analysis we find that there is a balance in the polarity of opinions and though this is not an in-depth analysis but it gives us an overview of the sentiments of people using twitter.

Considerations

- The data being collected seems to be too unstructured as the hashtags contain a lot of information but we have not considered them separately i.e for example: #impeachtrump would have to be split into “impeach trump” to make it close to actual English statement. The user names being used is also being populated to the most frequent words, removing these is sometimes better but other times we need to keep them example: @realDonaldTrump is the user handle for President Trump, this needs to be kept as it refers to Trump, @poba is a twitter user and this user handle is of no use to us.
- The emoji used in twitter contain a lot of emotional information but we have not considered them for the same, a possible solution for this is to assign a polarity values for each emoji or replace them with a suitable word that expresses a similar emotion.
- The number of tweets that are received have the same content i.e retweets, these are not considered for evaluation. These should be considered for polarity as the retweets also express the people’s sentiment.
- LDA algorithm is used using Gensim , here to vary the number of topic and the passes improves the accuracy of the model. For our model we only have chosen 3 values for the number of topics which are 3, 5, 7 and the runs were fixed at 500. This is a low estimation for model prediction. The ideal scenario would need more than 10,000 runs for a particular number of topics and the number of topics must also be varied to find the threshold topics.
- For Gensim in the after the runs we have chosen the model with minimum perplexity