#VASen Analysis

Georges Colbert

07/15/2018

# Period of July 08 to July 14

Our goal is to cover the conversations being had about the candidates across multiple platforms, interpret it and visualize the information in a way that can help drive decision making. The Graphs shown and metrics calculated are best estimation of what can be most useful to the candidate. Through a process of constant feedback between us and the campaign, the most relevant information will be identified and facilitate the creation of a personalized dashboard that the campaign can access 24/7. For now, multiple visualizations of the same information could be provided as we try to identify the best way for the campaign to retain the information and identify the KPIs amongst the myriad of metrics available.

In the case of Twitter, it is really fortunate that twitter provides an api on which developers can request an acquire a substantial amount of tweets from the previous ten days before the API call.

For more information on the twitter API the link is:

<https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html>

To facilitate the process of Data mining the tweets, here’s a list of packages used:

* [TwitteR Package](https://cran.r-project.org/web/packages/twitteR/twitteR.pdf)
* [TM package](https://cran.r-project.org/web/packages/tm/index.html)
* [ggplot2 package](https://cran.r-project.org/web/packages/ggplot2/index.html)
* [wordcloud package](https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf)
* [SnowballC package](https://cran.r-project.org/web/packages/SnowballC/SnowballC.pdf)
* [SentimentR package](https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf)

For this analysis, a request was made to the API for up to 30,000 tweets that used the hashtag #VASEN covering the period of July 08-14. All the tweets returned by the API are in the zip folder.

# OVERVIEW

# Total Number of tweets

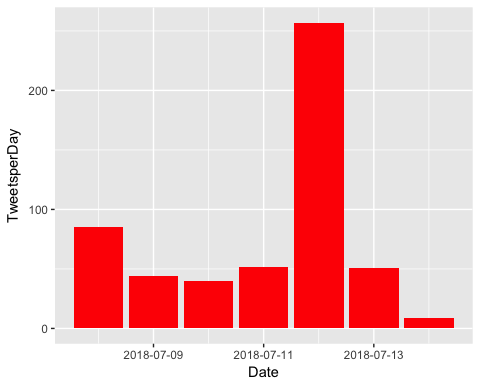
nrow(VA.tweetdf)

## [1] 538

# Number of Tweets per Day

##   
## Attaching package: 'scales'

## The following object is masked from 'package:readr':  
##   
## col\_factor

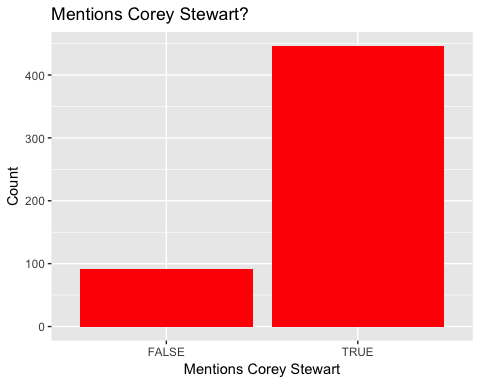


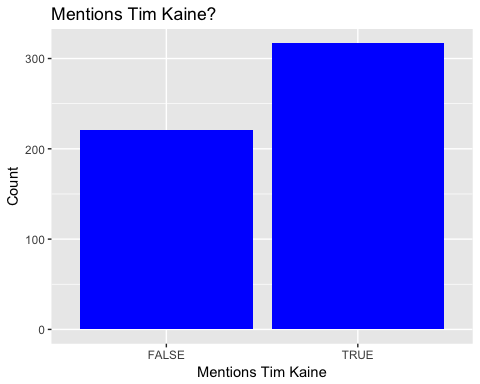
# How many are retweets?

##### How many are retweets?  
   
 table(VA.tweetdf$isRetweet)

##   
## FALSE TRUE   
## 70 468

#VASen is viewed as a neutral hashtag that supporters of either candidate can use to tweet about the race. Assessing which candidate’s name occurs more often amongst the tweets can be used as a proxy for which side is more activated.





**ENGAGEMENT**

Engagement is a measurement of how people on the platform are interacting with the candidate’s profile. Tracking engagement allows the campaign to assess how intensively are potential voters discussing or following the candidate, and identify which topics/tweets are “riling the base” by comparing popular tweets to ones that are less so.

As a Measure of direct engagement, here is a time series graph created with the ggplot2 package that shows, per day, how often people tweeted direct replies to one of the Candidates tweets.

### list of top 6 favorite tweets of the period

## text  
## 1 Check out our new ad: Kaine attended an "Abolish ICE" rally on June 30 and has done nothing in D.C. but raise taxes… https://t.co/lrN6xz7oWM  
## 2 .@realDonaldTrump has brought America's economy back. \n\n&gt;Econ Growth is at 4%. \n&gt;Unemployment is at its lowest in 1… https://t.co/k5R0FN1yyp  
## 3 NEW: @timkaine raised $2.2M in Q2, bringing total raised this cycle to $18.2M. We've also made an early $5M investm… https://t.co/cvaH51ukyp  
## 4 My first map! 2014 #VASEN -- Huge thanks to @cinyc9 @JMilesColeman @SenhorRaposa @liberaljoy for helping out. https://t.co/Rk4uRAFMnQ  
## 5 .@TimKaine blocking out TV advertising time in #VaSen elex, booking $5M for the fall -- and it's not even August. D… https://t.co/trxxG8NyMW  
## 6 Glad to see Sen. @timkaine focusing on rural VA. #VASen https://t.co/Bnl8JDcYWw  
## favoriteCount created screenName  
## 1 467 2018-07-12 02:52:09 CoreyStewartVA  
## 2 241 2018-07-08 16:37:45 CoreyStewartVA  
## 3 55 2018-07-11 18:49:22 IanSams  
## 4 33 2018-07-10 07:14:18 ChazNuttycombe  
## 5 24 2018-07-11 18:25:35 RTDSchapiro  
## 6 20 2018-07-10 02:47:34 politico\_chris

# top 6 retweeted tweets of the period

## text  
## 1 Check out our new ad: Kaine attended an "Abolish ICE" rally on June 30 and has done nothing in D.C. but raise taxes… https://t.co/lrN6xz7oWM  
## 2 .@realDonaldTrump has brought America's economy back. \n\n&gt;Econ Growth is at 4%. \n&gt;Unemployment is at its lowest in 1… https://t.co/k5R0FN1yyp  
## 3 NEW: @timkaine raised $2.2M in Q2, bringing total raised this cycle to $18.2M. We've also made an early $5M investm… https://t.co/cvaH51ukyp  
## 4 .@TimKaine blocking out TV advertising time in #VaSen elex, booking $5M for the fall -- and it's not even August. D… https://t.co/trxxG8NyMW  
## 5 Today @jfradioshow asked @DaveBratVA7th if he would campaign with @CoreyStewartVA or not.\n\nBrat spent \*\*\*4 minutes\*… https://t.co/uQ0cAR0Emp  
## 6 My first map! 2014 #VASEN -- Huge thanks to @cinyc9 @JMilesColeman @SenhorRaposa @liberaljoy for helping out. https://t.co/Rk4uRAFMnQ  
## retweetCount created screenName  
## 1 298 2018-07-12 02:52:09 CoreyStewartVA  
## 2 108 2018-07-08 16:37:45 CoreyStewartVA  
## 3 12 2018-07-11 18:49:22 IanSams  
## 4 8 2018-07-11 18:25:35 RTDSchapiro  
## 5 8 2018-07-09 18:01:57 AndrewBatesNC  
## 6 6 2018-07-10 07:14:18 ChazNuttycombe

### Top ten retweets overall

In this case, there were tweets that were creeated before the week that were still being shared on Twitter and accumulating retweets.

## text  
## 1 RT @CoreyStewartVA: Check out our new ad: Kaine attended an "Abolish ICE" rally on June 30 and has done nothing in D.C. but raise taxes, su…  
## 2 Check out our new ad: Kaine attended an "Abolish ICE" rally on June 30 and has done nothing in D.C. but raise taxes… https://t.co/lrN6xz7oWM  
## 3 RT @CoreyStewartVA: There are more jobs than people out of work, something the American economy has never experienced before!\n\nThank you, @…  
## 4 RT @CoreyStewartVA: .@realDonaldTrump has brought America's economy back. \n\n&gt;Econ Growth is at 4%. \n&gt;Unemployment is at its lowest in 18 ye…  
## 5 .@realDonaldTrump has brought America's economy back. \n\n&gt;Econ Growth is at 4%. \n&gt;Unemployment is at its lowest in 1… https://t.co/k5R0FN1yyp  
## 6 RT @SiriusXMPatriot: Fresh off his victory in the Republican #VASen Primary .@CoreyStewartVA says the #Democrats and @timkaine have betraye…  
## 7 RT @CoreyStewartVA: Yesterday, we rallied outside of the Red Hen with @realDonaldTrump supporters and @trump\_bikers in Lexington!\n\nWatch ou…  
## 8 RT @CoreyStewartVA: At the Dale City Parade celebrating Independence Day\n\n #VaSen #MAGA #Virginia #July4th https://t.co/pxnJVQ1XlG  
## 9 RT @CoreyStewartVA: .@realDonaldTrump &amp; House unveiled the largest tax cut in history, bringing jobs back. Will Senate support? @TimKaine w…  
## 10 RT @CoreyStewartVA: Huge support at the Purcellville parade!\n\n#VASen #MAGA https://t.co/qCMzi4wSd0  
## retweetCount created screenName isRetweet  
## 1 298 2018-07-14 12:34:35 CarrierTo TRUE  
## 2 298 2018-07-12 02:52:09 CoreyStewartVA FALSE  
## 3 167 2018-07-12 02:41:34 slewfan TRUE  
## 4 108 2018-07-13 13:42:58 HDowning113 TRUE  
## 5 108 2018-07-08 16:37:45 CoreyStewartVA FALSE  
## 6 101 2018-07-13 19:31:52 RealLHunterJ TRUE  
## 7 78 2018-07-12 17:31:30 ggmsmolly TRUE  
## 8 65 2018-07-13 13:43:13 HDowning113 TRUE  
## 9 62 2018-07-10 00:44:55 R98250729 TRUE  
## 10 46 2018-07-12 03:07:22 mindset2growth TRUE

## accounts that have the most total Retweets

## screenName x  
## 8 CoreyStewartVA 406  
## 5 ChazNuttycombe 14  
## 16 IanSams 14  
## 2 AndrewBatesNC 8  
## 9 Croatansound 8  
## 26 RTDSchapiro 8

## Most Active Accounts (including Retweets)

## Var1 Freq  
## 73 Croatansound 15  
## 58 ChinWusu 9  
## 351 SackComedian 6  
## 50 CarrierTo 5  
## 56 ChazNuttycombe 5  
## 417 winegirl73 5

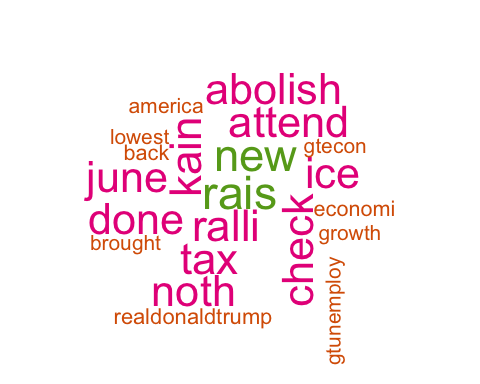
## Most Active Accounts

## Var1 Freq  
## 6 ChinWusu 8  
## 9 Croatansound 8  
## 28 SackComedian 5  
## 5 ChazNuttycombe 4  
## 16 IanSams 4  
## 31 TweetrHandle 4

# TEXT MINING

# What are people saying? The goal is to capture what words are being used to describe the candidate. We expect this section to be the one of the most interesting, as the team’s NLP capabilities get better we expect this section to expand.

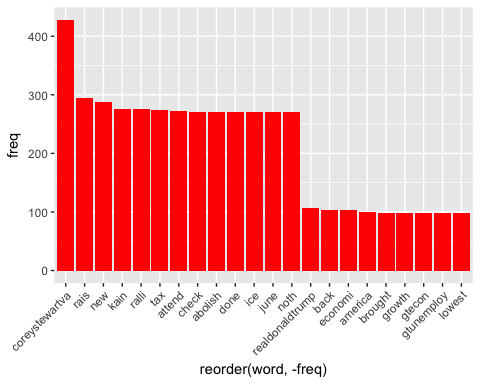
# WORDCLOUD of most common used words



# List of The Top 15 Frequently used Words

## word freq  
## coreystewartva coreystewartva 428  
## rais rais 294  
## new new 288  
## kain kain 276  
## ralli ralli 275  
## tax tax 274  
## attend attend 272  
## check check 271  
## abolish abolish 270  
## done done 270  
## ice ice 270  
## june june 270  
## noth noth 270  
## realdonaldtrump realdonaldtrump 107  
## back back 104

# Plot of Words that Appear at least 75 times



## Words Associated with

To find out what words are associated with the candidates, I use a function called FindAssocs() on the dataset. For any given word, findAssocs() calculates its correlation with every other word in the dataset. Scores range from 0 to 1. A score of 1 means that two words always appear together, while a score of 0 means that they never appear together.

# which words are associated with "Corey Stewart"  
   
 findAssocs(tdm,"coreystewartva", 0.1)

## $coreystewartva  
## ralli abolish attend done   
## 0.51 0.50 0.50 0.50   
## ice june noth check   
## 0.50 0.50 0.50 0.49   
## tax kain new rais   
## 0.49 0.47 0.37 0.30   
## back economi realdonaldtrump brought   
## 0.24 0.24 0.23 0.23   
## growth gtecon gtunemploy lowest   
## 0.23 0.23 0.23 0.23   
## america   
## 0.22

findAssocs(tdm,"timkain", 0.1)

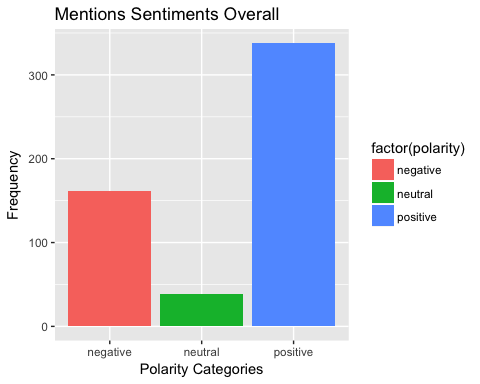
## $timkain  
## also bring cycl earli   
## 0.56 0.56 0.56 0.56   
## made total weve invest   
## 0.56 0.56 0.56 0.53   
## iansam democrat advertis august   
## 0.48 0.45 0.45 0.45   
## block book elex fall   
## 0.45 0.45 0.45 0.42   
## rtdschapiro even time jenniferwexton   
## 0.42 0.40 0.40 0.32   
## abolishic answer lesliecockburn spanbergerva   
## 0.28 0.28 0.28 0.28   
## still elaineluriava hawkdiesel vabeachdem   
## 0.28 0.28 0.28 0.28   
## vademocrat gerryconnolli hillaryclinton want   
## 0.28 0.22 0.22 0.22   
## sackcomedian virginianpilot garrenshipley accus   
## 0.22 0.20 0.18 0.16   
## karl marx betray fresh   
## 0.16 0.16 0.16 0.16   
## siriusxmpatriot victori add qanon   
## 0.16 0.16 0.16 0.16   
## around came frankoanderson region   
## 0.16 0.16 0.16 0.16   
## basicallynhpolit hey laurahbry yeah   
## 0.16 0.16 0.16 0.16   
## coreystewart elect firekain takebackva   
## 0.16 0.16 0.16 0.16   
## investm bostoninform massachusett focus   
## 0.16 0.16 0.16 0.16   
## rural sen come isnt   
## 0.16 0.16 0.16 0.16   
## may pass shock cut   
## 0.16 0.16 0.16 0.16   
## histori largest unveil amer   
## 0.16 0.16 0.16 0.16   
## comey haitifir hear peterstrzok   
## 0.16 0.16 0.16 0.16   
## thing scotttaylorva like primari   
## 0.16 0.13 0.13 0.10   
## other that thedemocrat north   
## 0.10 0.10 0.10 0.10   
## hous scotus violentleft   
## 0.10 0.10 0.10

findAssocs(tdm,"trump",0.1 )

## $trump  
## conspir hack judg report   
## 0.58 0.58 0.58 0.58   
## russian suit toss vase   
## 0.58 0.58 0.58 0.58   
## cadinet harass lib mood   
## 0.58 0.58 0.58 0.58   
## reg repmaxinewat anoth cuckservativ   
## 0.58 0.58 0.58 0.58   
## deafen kavanaugh wonder silenc   
## 0.58 0.58 0.58 0.41   
## alleg potus pick dem   
## 0.41 0.41 0.33 0.28   
## therussiahoax this just call   
## 0.28 0.28 0.25 0.25   
## muellerwitchhunt think even like   
## 0.23 0.23 0.17 0.17   
## croatansound campaign amp   
## 0.16 0.14 0.11

**Sentiment Analysis**

Using the (SentimentR) package, every tweet is scored to assess whether they are negative, positive or neutral.

 # **WorldCloud of words used in top 50 most negative and top 50 most positive tweets**

