#VASen Analysis

Georges Colbert

07/10/2018

# Period of June 30 to July 07

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 June 30-July 07. All the tweets returned by the API are in the zip folder.

# OVERVIEW

# Total Number of tweets

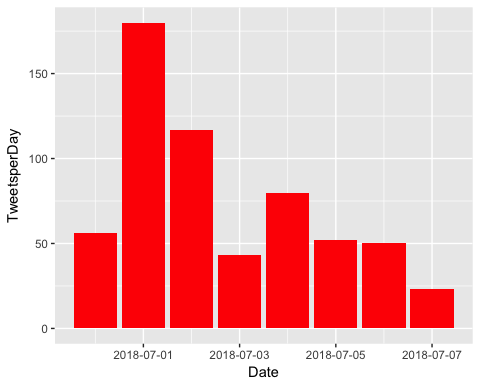
nrow(VA.tweetdf)

## [1] 601

# Number of Tweets per Day

##   
## Attaching package: 'scales'

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

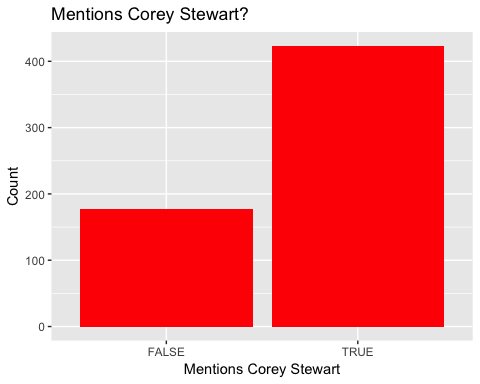


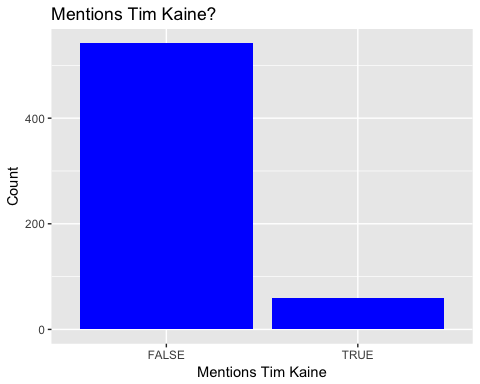
# How many are retweets?

##### How many are retweets?

##   
## FALSE TRUE   
## 127 474

#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 There are more jobs than people out of work, something the American economy has never experienced before!\n\nThank yo… https://t.co/g7TxIjSoTu  
## 2 At the Dale City Parade celebrating Independence Day\n\n #VaSen #MAGA #Virginia #July4th https://t.co/pxnJVQ1XlG  
## 3 Yesterday, we rallied outside of the Red Hen with @realDonaldTrump supporters and @trump\_bikers in Lexington!\n\nWatc… https://t.co/Ev76XZxhT6  
## 4 When selecting the new @VA\_GOP Chairman, party leadership should choose a strong leader who will fight for Presiden… https://t.co/7mLsr81TG2  
## 5 Huge support at the Purcellville parade!\n\n#VASen #MAGA https://t.co/qCMzi4wSd0  
## 6 Virginians everywhere need to know that @timkaine will never: give you up, let you down, run around, desert you, ma… https://t.co/yssL8WJvym  
## favoriteCount created screenName  
## 1 297 2018-07-01 18:04:44 CoreyStewartVA  
## 2 199 2018-07-04 15:10:54 CoreyStewartVA  
## 3 196 2018-07-01 14:34:32 CoreyStewartVA  
## 4 161 2018-06-30 18:04:57 CoreyStewartVA  
## 5 156 2018-07-04 22:16:01 CoreyStewartVA  
## 6 67 2018-07-01 21:00:37 dcaflight757

# top 6 retweeted tweets of the period

## text  
## 1 There are more jobs than people out of work, something the American economy has never experienced before!\n\nThank yo… https://t.co/g7TxIjSoTu  
## 2 Yesterday, we rallied outside of the Red Hen with @realDonaldTrump supporters and @trump\_bikers in Lexington!\n\nWatc… https://t.co/Ev76XZxhT6  
## 3 At the Dale City Parade celebrating Independence Day\n\n #VaSen #MAGA #Virginia #July4th https://t.co/pxnJVQ1XlG  
## 4 When selecting the new @VA\_GOP Chairman, party leadership should choose a strong leader who will fight for Presiden… https://t.co/7mLsr81TG2  
## 5 Huge support at the Purcellville parade!\n\n#VASen #MAGA https://t.co/qCMzi4wSd0  
## 6 Virginians everywhere need to know that @timkaine will never: give you up, let you down, run around, desert you, ma… https://t.co/yssL8WJvym  
## retweetCount created screenName  
## 1 167 2018-07-01 18:04:44 CoreyStewartVA  
## 2 76 2018-07-01 14:34:32 CoreyStewartVA  
## 3 63 2018-07-04 15:10:54 CoreyStewartVA  
## 4 62 2018-06-30 18:04:57 CoreyStewartVA  
## 5 45 2018-07-04 22:16:01 CoreyStewartVA  
## 6 19 2018-07-01 21:00:37 dcaflight757

### Top ten retweets overall

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

## text  
## 1 RT @JTHVerhovek: The NRSC is declining to comment on Corey Stewart's victory in the #VASen GOP primary\n\nEarlier tonight his general electio…  
## 2 RT @CoreyStewartVA: There are more jobs than people out of work, something the American economy has never experienced before!\n\nThank you, @…  
## 3 There are more jobs than people out of work, something the American economy has never experienced before!\n\nThank yo… https://t.co/g7TxIjSoTu  
## 4 RT @CoreyStewartVA: .@realDonaldTrump continues to find solutions while Democrats like @timkaine want lawlessnes and open borders. \n\n#VASen…  
## 5 RT @SuMoh7: RT! #Election2018 #Virginia #WinBlue\n\nSupport Sen. TIM KAINE #VAsen,&amp; @DemsWork4USA\n\xed\xa0\xbd\xed\xb1\x89https://t.co/f49Oa2551o\n\n#YesWeCan #KeepIt…  
## 6 RT @CoreyStewartVA: Yesterday, we rallied outside of the Red Hen with @realDonaldTrump supporters and @trump\_bikers in Lexington!\n\nWatch ou…  
## 7 Yesterday, we rallied outside of the Red Hen with @realDonaldTrump supporters and @trump\_bikers in Lexington!\n\nWatc… https://t.co/Ev76XZxhT6  
## 8 RT @CoreyStewartVA: At the Dale City Parade celebrating Independence Day\n\n #VaSen #MAGA #Virginia #July4th https://t.co/pxnJVQ1XlG  
## 9 At the Dale City Parade celebrating Independence Day\n\n #VaSen #MAGA #Virginia #July4th https://t.co/pxnJVQ1XlG  
## 10 RT @CoreyStewartVA: When selecting the new @VA\_GOP Chairman, party leadership should choose a strong leader who will fight for President @r…  
## retweetCount created screenName isRetweet  
## 1 238 2018-07-06 15:08:03 kstreet111 TRUE  
## 2 167 2018-07-05 00:24:27 atxflyboy TRUE  
## 3 167 2018-07-01 18:04:44 CoreyStewartVA FALSE  
## 4 149 2018-07-05 00:27:12 atxflyboy TRUE  
## 5 124 2018-07-06 23:58:11 harmonyis1 TRUE  
## 6 76 2018-07-05 00:25:02 atxflyboy TRUE  
## 7 76 2018-07-01 14:34:32 CoreyStewartVA FALSE  
## 8 63 2018-07-06 00:27:51 WilliamDaleDan1 TRUE  
## 9 63 2018-07-04 15:10:54 CoreyStewartVA FALSE  
## 10 62 2018-07-03 18:00:29 HDowning113 TRUE

## accounts that have the most total Retweets

## screenName x  
## 6 CoreyStewartVA 413  
## 21 IanSams 27  
## 9 dcaflight757 23  
## 7 Croatansound 14  
## 12 frankoanderson 8  
## 47 vabeachdems 8

## Most Active Accounts (including Retweets)

## Var1 Freq  
## 63 Croatansound 22  
## 300 SackComedian 13  
## 113 GarrenShipley 9  
## 46 ChazNuttycombe 7  
## 38 CarlosV64910815 6  
## 72 dcaflight757 6

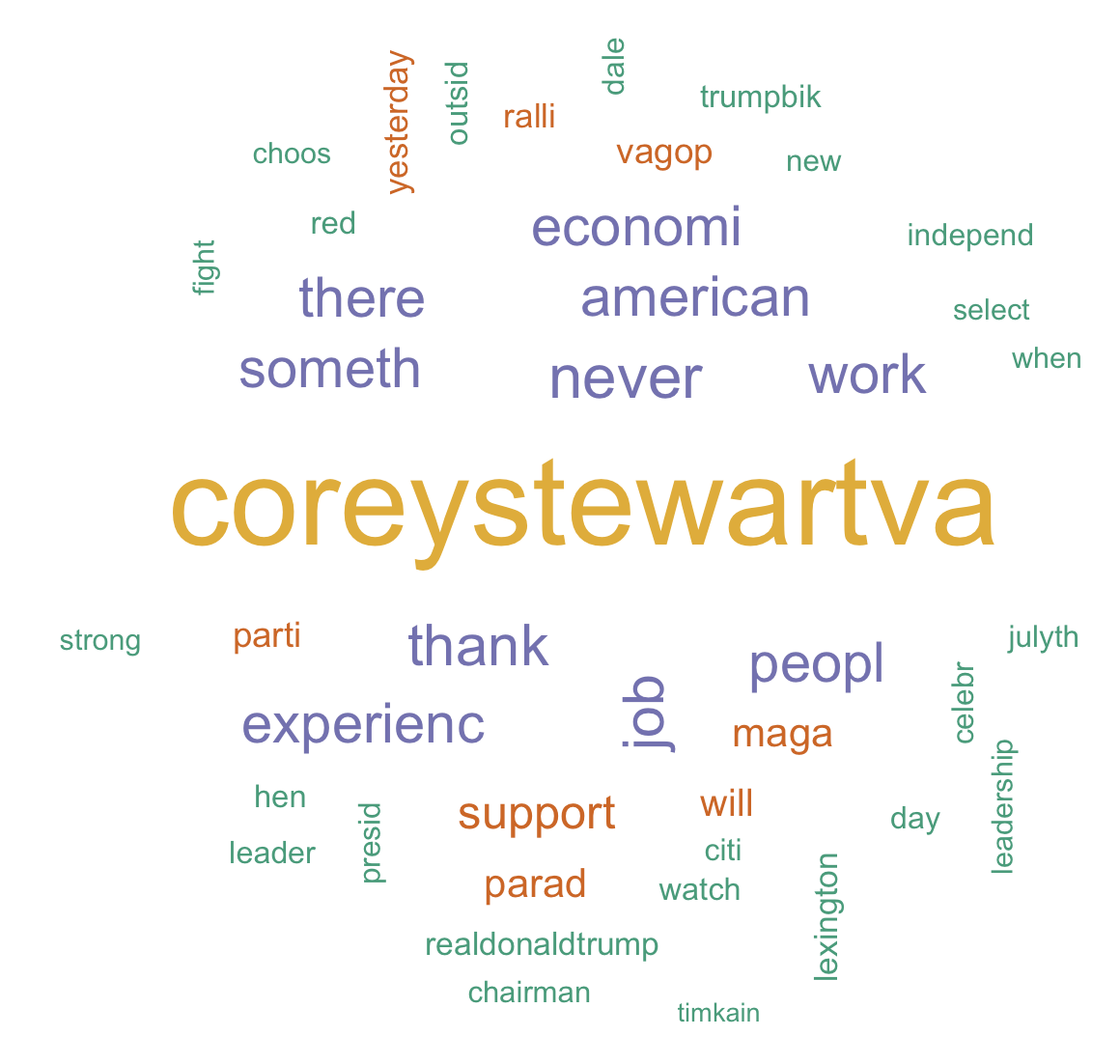
## Most Active Accounts

## Var1 Freq  
## 7 Croatansound 17  
## 40 SackComedian 12  
## 17 GarrenShipley 9  
## 4 ChazNuttycombe 6  
## 9 dcaflight757 6  
## 21 IanSams 6

# 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.

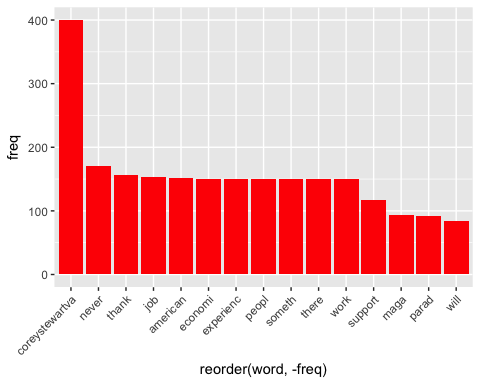
# WORDCLOUD of most common used words



# List of The Top 15 Frequently used Words

## word freq  
## coreystewartva coreystewartva 400  
## never never 171  
## thank thank 157  
## job job 154  
## american american 151  
## economi economi 150  
## experienc experienc 150  
## peopl peopl 150  
## someth someth 150  
## there there 150  
## work work 150  
## support support 117  
## maga maga 93  
## parad parad 92  
## will will 84

# 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  
## job american economi experienc   
## 0.39 0.39 0.39 0.39   
## peopl someth there work   
## 0.39 0.39 0.39 0.39   
## thank parad never maga   
## 0.36 0.28 0.27 0.25   
## watch leader realdonaldtrump hen   
## 0.23 0.22 0.22 0.22   
## outsid trumpbik celebr citi   
## 0.22 0.22 0.21 0.21   
## dale julyth red new   
## 0.21 0.21 0.21 0.21   
## independ choos fight leadership   
## 0.20 0.20 0.20 0.20   
## select strong when day   
## 0.20 0.20 0.20 0.19   
## lexington presid chairman purcellvill   
## 0.19 0.18 0.18 0.16   
## yesterday ralli pictur vagop   
## 0.16 0.15 0.15 0.13   
## huge support barbaracomstock anyway   
## 0.13 0.12 0.12 0.12   
## dont head strang onelonelybigot   
## 0.12 0.12 0.12 0.11   
## wompwomp bigot bless heart   
## 0.11 0.11 0.11 0.11   
## shaftan racist screenshot   
## 0.11 0.10 0.10

findAssocs(tdm,"timkain", 0.1)

## $timkain  
## around desert everywher give need   
## 0.80 0.68 0.68 0.68 0.68   
## let know virginian dcaflight make   
## 0.66 0.64 0.62 0.60 0.59   
## run jenniferwexton came gerryconnolli region   
## 0.55 0.46 0.41 0.41 0.41   
## local frankoanderson democrat lesliecockburn will   
## 0.35 0.34 0.27 0.27 0.24   
## what novemb team chesterfield field   
## 0.24 0.22 0.22 0.22 0.22   
## got jam pack offic spanbergerva   
## 0.22 0.22 0.22 0.17 0.17   
## fish irrevelv mock pond tbforlif   
## 0.15 0.15 0.15 0.15 0.15   
## tweet write written year child   
## 0.15 0.15 0.15 0.15 0.15   
## petul senat trade war tha   
## 0.15 0.15 0.15 0.15 0.15   
## find lawlessn solut abort badcathol   
## 0.15 0.15 0.15 0.15 0.15   
## demand karawavi parenthood sameol wavynew   
## 0.15 0.15 0.15 0.15 0.15   
## cimartinez congratul secur hasnt jerrywkilgor   
## 0.15 0.15 0.15 0.15 0.15   
## racistnazi help repres worst violentleft   
## 0.15 0.15 0.15 0.15 0.14   
## juli say garrenshipley not yet   
## 0.14 0.14 0.14 0.14 0.12   
## hey even volunt never small   
## 0.12 0.11 0.11 0.11 0.10   
## clear old victori servic join   
## 0.10 0.10 0.10 0.10 0.10

findAssocs(tdm,"trump",0.1 )

## $trump  
## alleg conspir hack judg   
## 0.50 0.50 0.50 0.50   
## muellerwitchhunt russian suit therussiahoax   
## 0.50 0.50 0.50 0.50   
## toss bikersfortrump demonstr potus   
## 0.50 0.50 0.50 0.50   
## redhen croatansound controversi want   
## 0.50 0.49 0.43 0.40   
## buildthewal iran more much   
## 0.40 0.40 0.40 0.40   
## paid securethebord than wall   
## 0.40 0.40 0.40 0.40   
## fall great massiv montana   
## 0.40 0.40 0.40 0.40   
## speech teapart fish irrevelv   
## 0.40 0.40 0.29 0.29   
## mock pond mts secu   
## 0.29 0.29 0.29 0.29   
## chariti dinner forthofjuli militari   
## 0.29 0.29 0.29 0.29   
## salut westvirginia wvsen wake   
## 0.29 0.29 0.29 0.28   
## report obama small openbord   
## 0.28 0.25 0.21 0.21   
## campaign independenceday servic ralli   
## 0.20 0.20 0.20 0.14   
## violentleft speak the vapol   
## 0.13 0.13 0.12 0.12   
## vase democrat   
## 0.11 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

