

Exploratory Analysis with Data Mined from Twitter

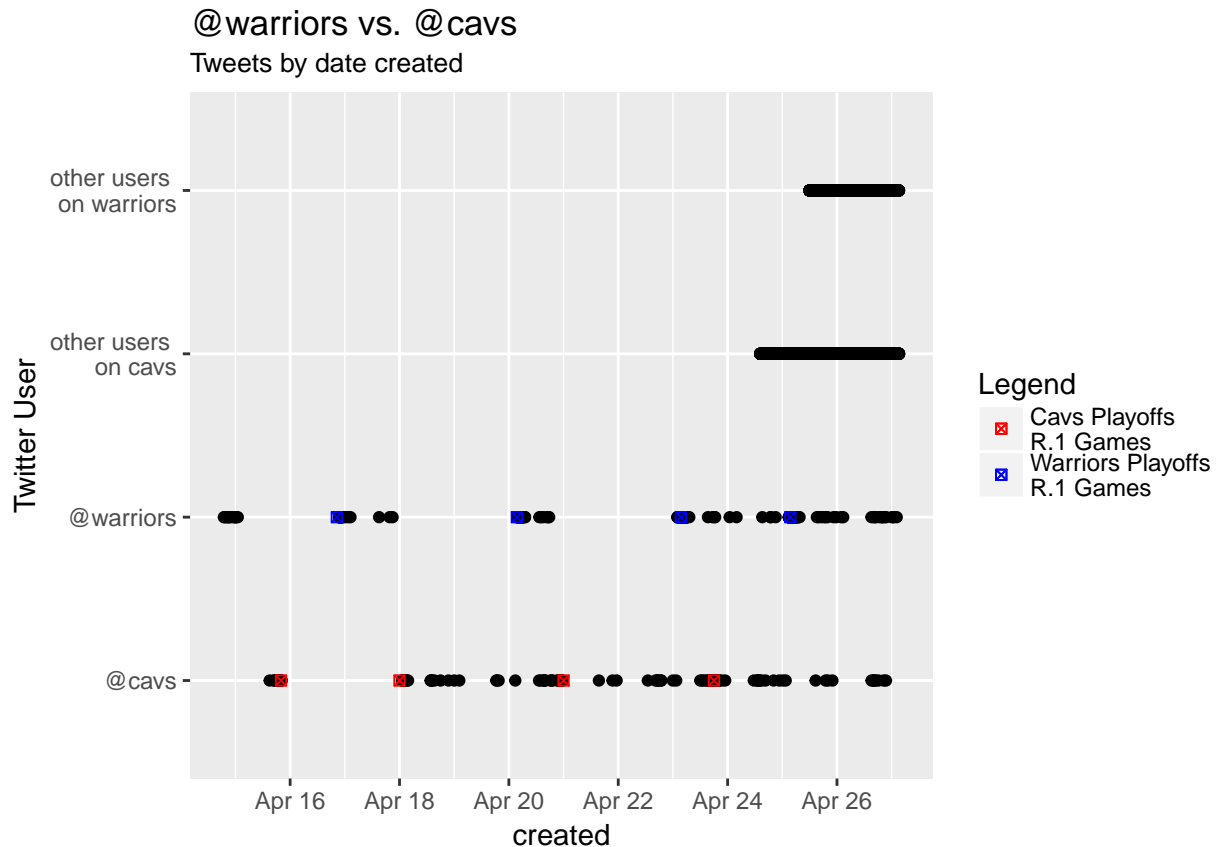
This R notebook is to showcase examples of analyses that can be done based on Twitter-data and Twitter-meta-data.

Let's focus on a more relevant topic, say the currently ongoing NBA Playoffs, since Twitter only makes recent tweets available. Additionally, not all teams are equal, in the sense that more popular teams will have more relevant Twitter activity and less popular teams will have less. So, let's compare the Twitter activity of the two teams in last year's NBA Playoff Finals - Cleveland Cavaliers & Golden State Warriors.

Note: While Twitter allows a maximum of 3600 tweets to be grabbed at once, this is a slow process. Thus, we choose to grab $n=3000$ tweets whenever `searchTwitter()` from the `twitterR` package is called. Further, Twitter has restricted access to its API, meaning too many requests to access Twitter's API in a short period of time leads to an error. There is no way to get around that, so we must be careful how many calls to access Twitter's API in a certain period of time we make. Otherwise, we have to wait, before trying again.

Charting When Tweets Were Created:

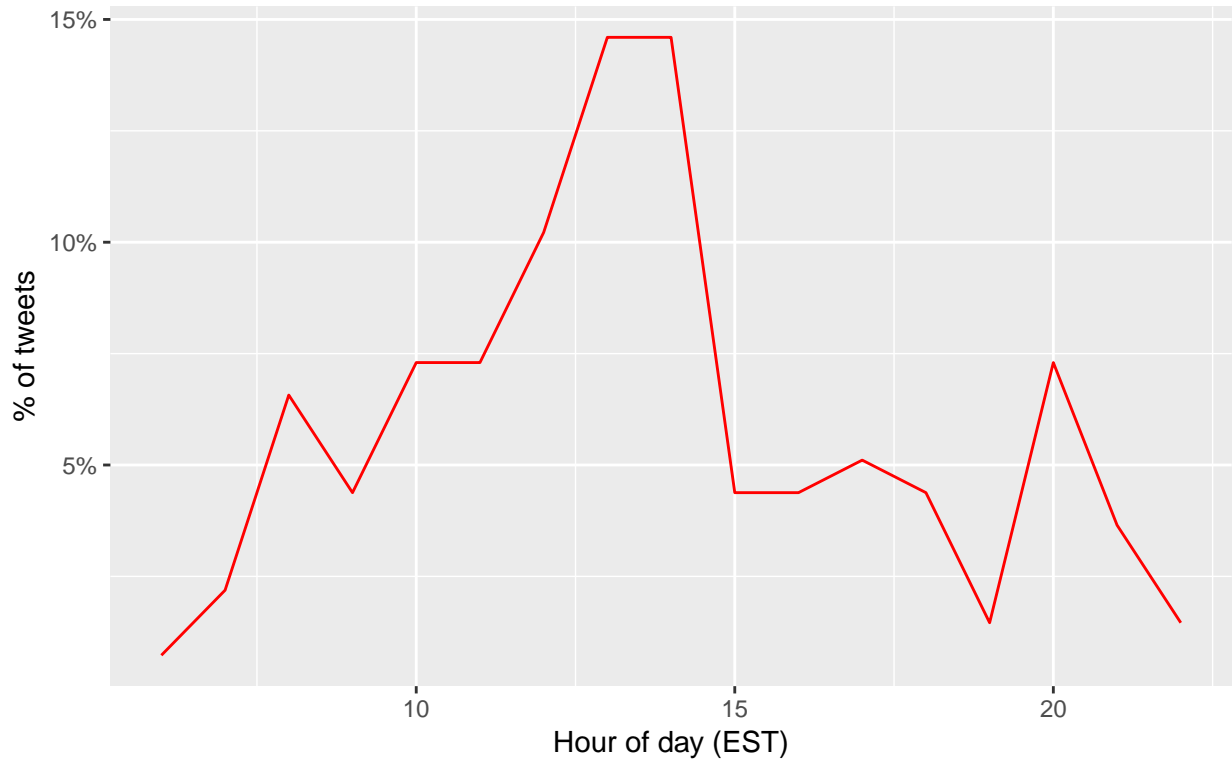
We may want to chronicle tweets from a given user's Timeline. Below, we chronicle the recent 3000 tweets on the official Cleveland Cavaliers Twitter and the recent 3000 tweets on the official Golden State Warriors Twitter. Also, we marked the dates corresponding to each team's playoff games. For the Warriors, as far as the last 4 games go, Twitter activity by official @warriors seems to be concentrated in the time immediately following a game. For the Cavs, as far as the last 4 games go, Twitter activity by official @cavs seems to more averagely spread across the time before, during, and after a game. Those tweets in between games, for both official team accounts, can probably be accounted for by promotional tweets, e.g. tweets promoting the sale of tickets. Seeing as how this is only the beginning of the Playoffs, i.e., Round 1, of course there is a high volume of tweets in the last few days reflecting the momentum of social media activity as hype for the successive games rises. As for the spread of tweets referencing either team by non-team-official accounts, it is expected that they would concentrate in the last few days, given the restriction of the number of tweets we're able to access. So, there is no significant conclusion that can be drawn from chronicling all Twitter user tweets about the Golden State Warriors and Cleveland Cavaliers.



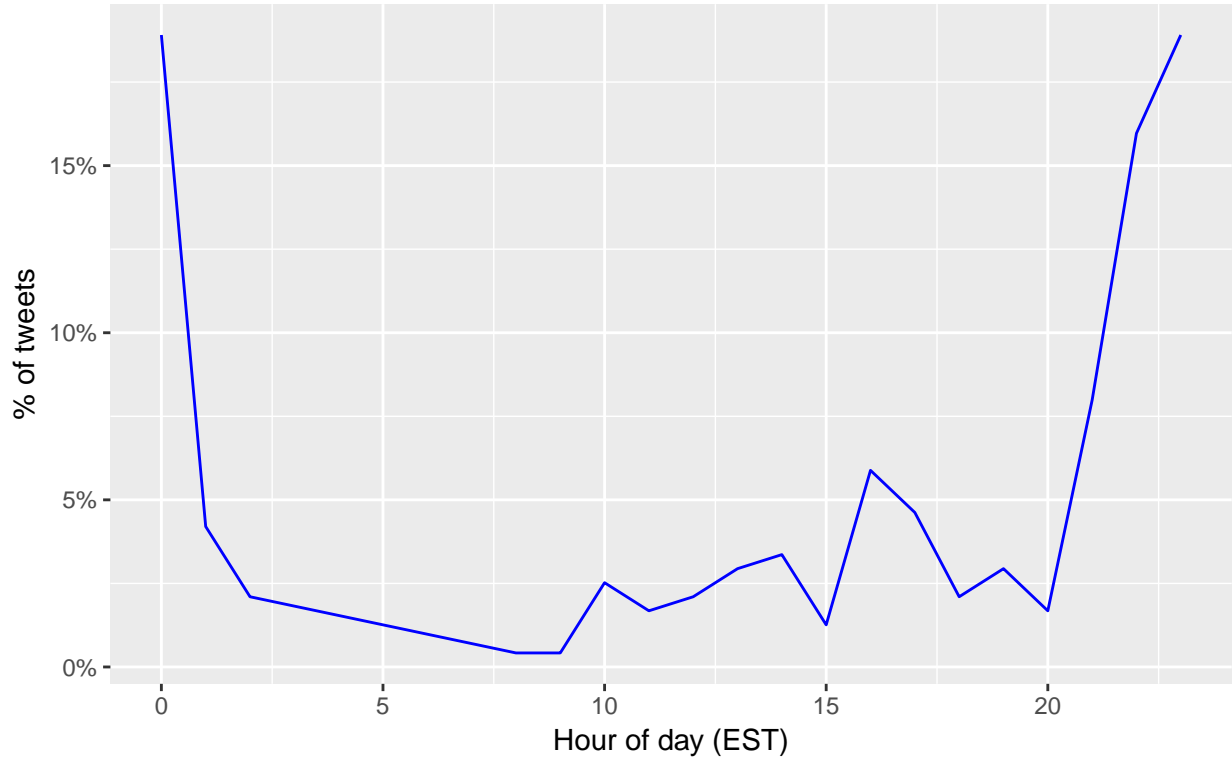
Breakdown of Tweets by the Hour:

We can look more in detail at tweets from the official @warriors and @cavs accounts based on hour of day. Again, from the 3000 tweets grabbed from Twitter, we plot the percentage of tweets as a function of the hour of the day (in Eastern Standard Time). Notice that the official @cavs has its highest percentage of tweets posted between 1pm and 2pm. This seems to be related to the fact that 2 of the 4 Cavs games have been at 1pm and 3pm. Additionally, the second peak about 8pm is related to the other 2 games at 7pm. The higher peak between 1pm and 2pm can be attributed to the fact that those 2 early afternoon games were on the weekend. Thus, it makes sense that more people are free to watch weekend, rather than weekday, games. So, official at @cavs may tweet more on the weekends to reach a larger audience. The same logic applies to the trend we see in the graph of percentage of tweets by hour for the Warriors. Since 3 of the 4 Warriors games were at 10:30pm, the greatest peak around midnight makes sense. The second peak is attributed to the 4th game that went on at 3pm. Officials for both the Cavs and Warriors are probably tweeting game highlights and recaps.

Official Tweets @cavs
Breakdown by Hour



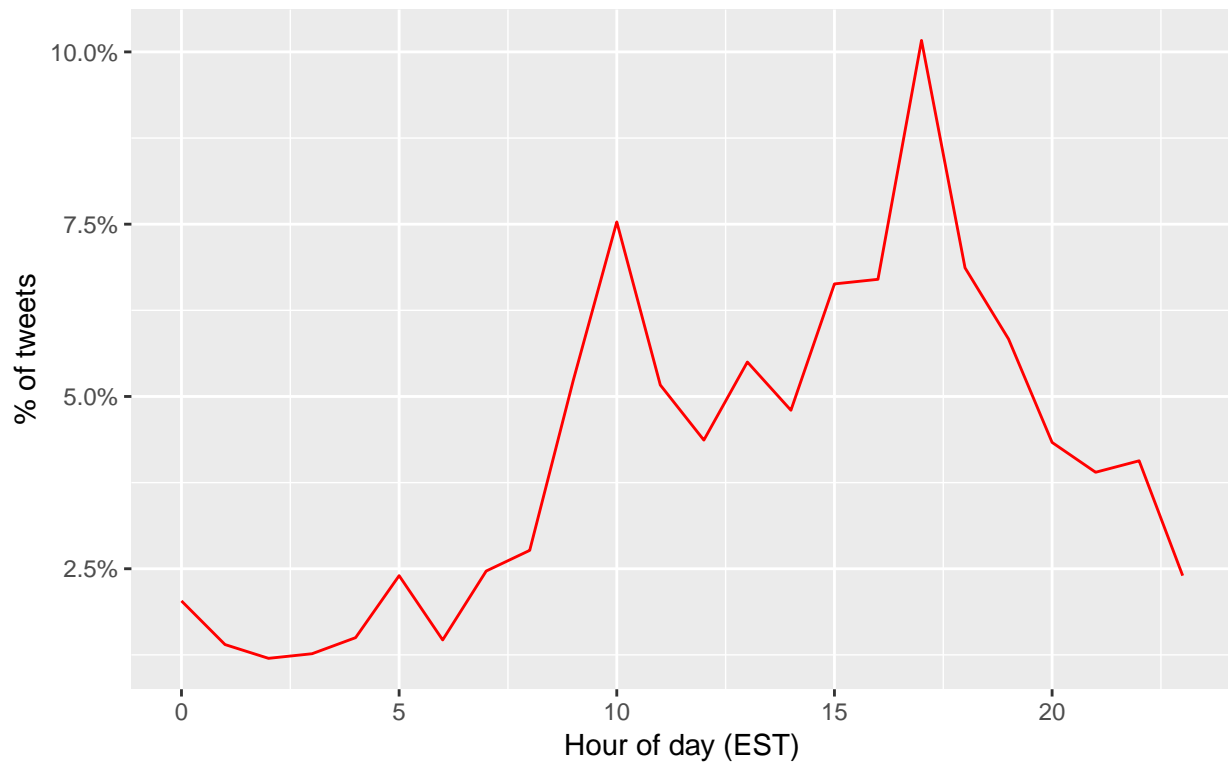
Official Tweets @warriors
Breakdown by Hour



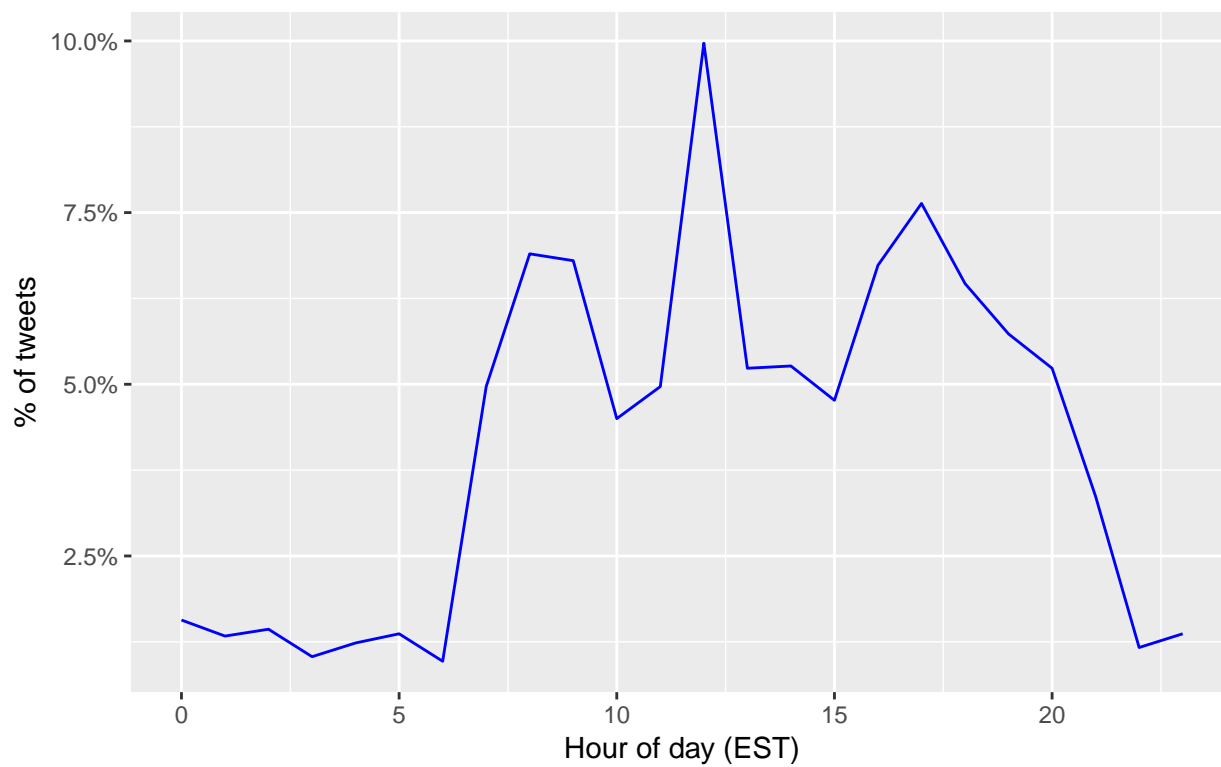
Now here is the breakdown of tweets from non-team-official Twitter users. The Cavs won all four of their Round 1 games against the Pacers. But their first game on April 15 at 3pm was the closest, as Cavs beat Pacers by 1 point (109-108). So the hype of the first game and the close score would explain the highest peak of Twitter activity, except that tweets were from April 24th-26th. Without further information about Cavs fans (or at least those who tweet about the Cavs), we can't conclude anything about twitter users tweeting about the Cavs. Further, for those tweeting about the Warriors, we may speculate that the time difference between East Coast and West Coast contribute to the highest peak at noon EST. Since most fans of the Warriors are probably on Pacific Standard Time, for the context of when they're tweeting in the day, we should consider tweets in PST. So then, at around 9am PST, tweets about the Warriors are at the highest level. Perhaps, many fans don't watch the games in real-time, and watch recaps the next morning. That would certainly account for the morning tweets about the Warriors. Again, we have no way of knowing for certain without additional information about Twitter users' tweeting behavior.

All Tweets about Cavaliers

Breakdown by Hour

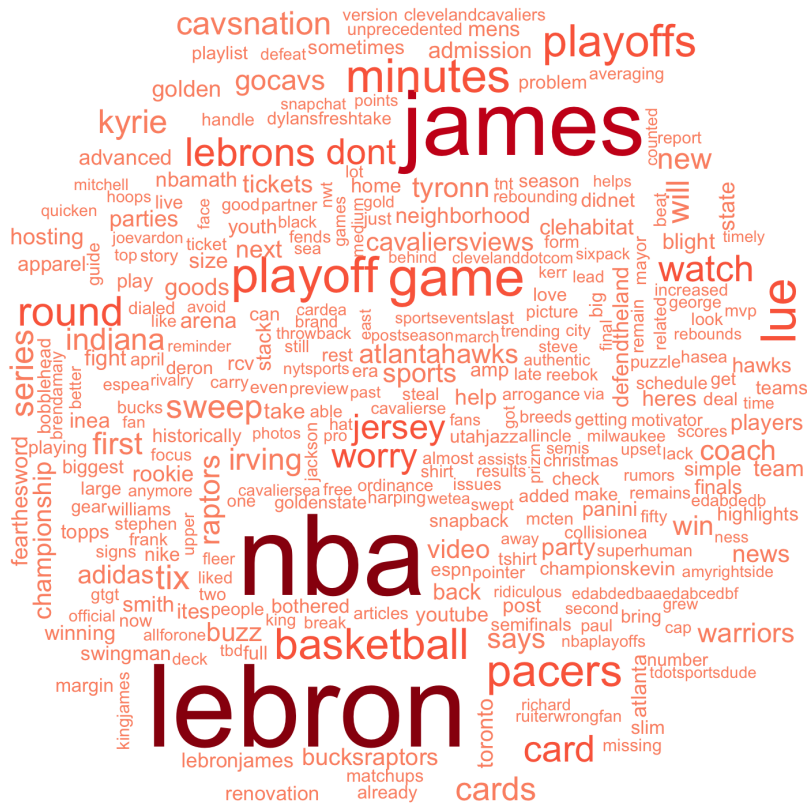


All Tweets about Warriors
Breakdown by Hour



Wordcloud:

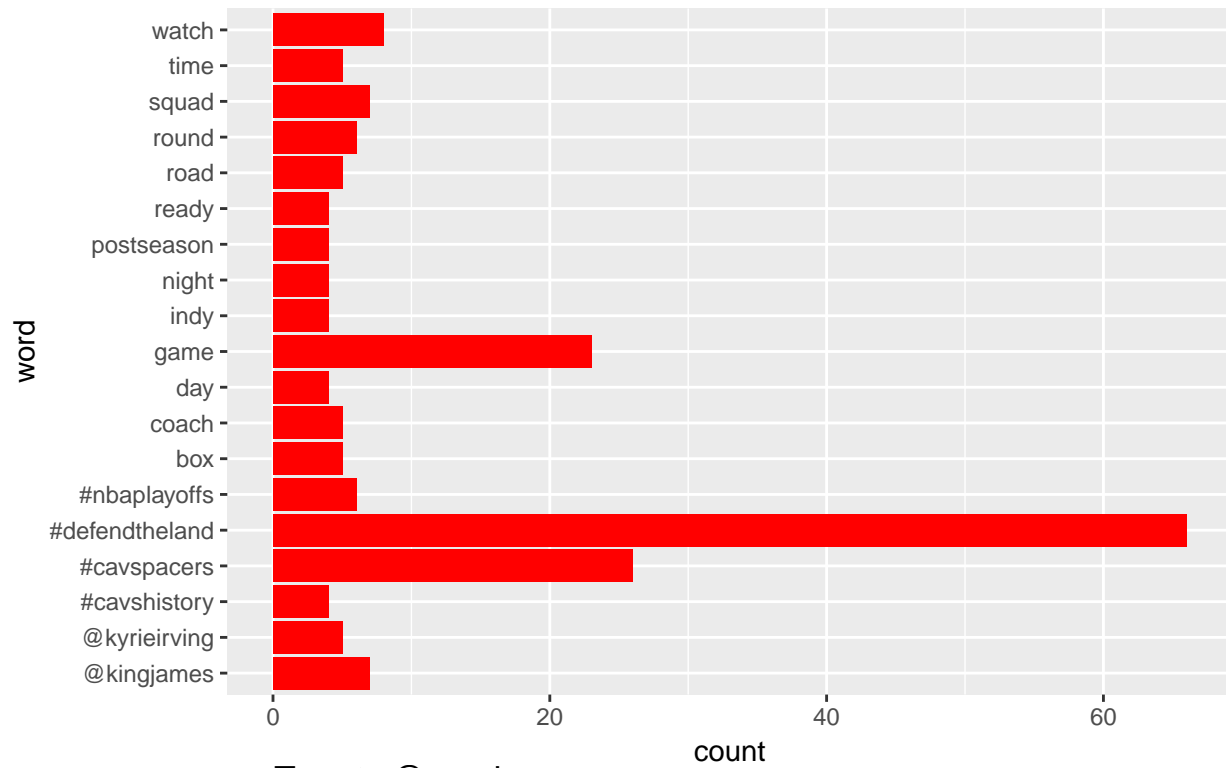
The following is a word cloud based on tweets about the Cavaliers:



The following is a word cloud based on tweets about the Warriors:

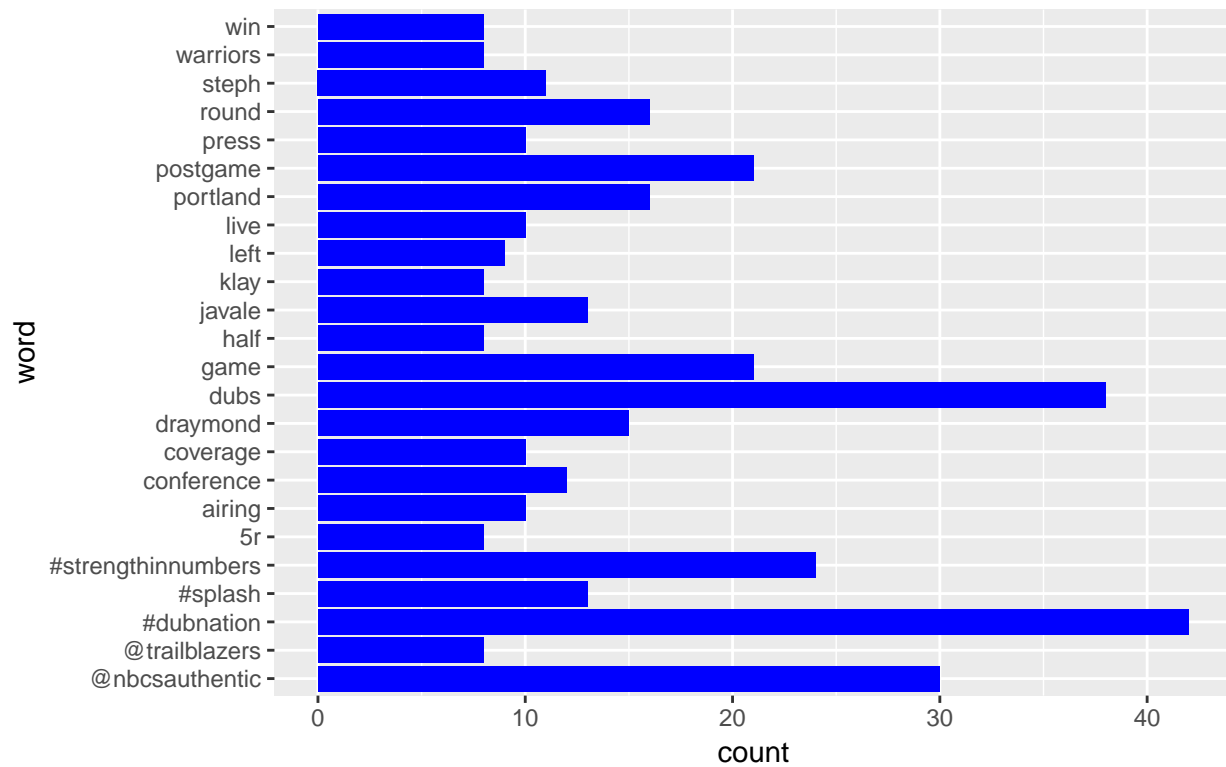
Tweets @cavs

breakdown by word count



Tweets @warriors

breakdown by word count



Sentiment Analysis: (NRC Lexicon; 10 sentiments)

First, let's look at the **sentiment analysis of tweets about the Cleveland Cavaliers**. Users tweeting about the Cavs use first and foremost positive words though negative words are a close third. Clearly, we can deduce that positive sentiments are from Cavs fans. However, negative sentiment may come from fans of rival teams (perhaps specifically fans of currently opposing team, the Pacers) and disappointed fans. The second most common sentiment in tweets about the Cavs is anticipation, which is not surprising as playoff games bring about a build-up of energy and emotions. The two weakest sentiments are anger and surprise which makes sense, since the Cavs won all 4 of their Round 1 games against the Pacers. Fans don't really have anything to be angry about and fans ought not be surprised since they by virtue of being fans expect their team to win.

##	sentiment	num_words
## 1	anger	369
## 2	anticipation	1101
## 3	disgust	207
## 4	fear	968
## 5	joy	579
## 6	negative	1032
## 7	positive	1333
## 8	sadness	544
## 9	surprise	343
## 10	trust	954

Here are the ten most common positive words in the 3000 tweets about the Cavaliers:

##	word	freq
## 16	basketball	126
## 104	love	65
## 166	top	48
## 184	youth	44
## 142	renovation	37
## 96	king	30
## 115	partner	29
## 71	gear	28
## 81	guide	26
## 111	mayor	25

Here are the ten most common negative words in the 3000 tweets about the Cavaliers:

##	word	freq
## 79	raptors	220
## 115	worry	161
## 12	collision	27
## 66	missing	27
## 1	avoid	24
## 21	defeat	24
## 99	steal	24
## 108	upset	24
## 4	black	21
## 55	late	21

Now, let's look at the **sentiment analysis of tweets about the Golden State Warriors**. By an overwhelming amount, the positive sentiment is strongest across all the tweets about the Warriors (even stronger than positive sentiment among tweets about the Cavs, at least in terms of number of positive words). Though unlike with the tweets about the Cavs that had the sentiment negative as the third strongest sentiment, tweets about the Warriors have the sentiment joy as the third strongest sentiment. This further

affirms that the dialogue surrounding the Warriors is generally more positive and happy than that regarding the Cavs. Similar to tweets about the Cavs, tweets about the Warriors have anticipation as the second strongest sentiment. Further notable is that trust is tied for the third strongest sentiment (along with joy) here. So, it seems those who tweet about the Warriors, presumably fans, are more confident and pleased, than those who tweet about the Cavs.

```
##      sentiment num_words
## 1      anger      364
## 2 anticipation  1151
## 3      disgust   106
## 4       fear    581
## 5       joy    1025
## 6    negative   764
## 7    positive  1836
## 8     sadness   311
## 9     surprise   516
## 10     trust   1021
```

Here are the ten most common positive words in the 3000 tweets about the Warriors:

```
##      word freq
## 36  confident 254
## 16  basketball 122
## 126 popularity 97
## 48   defense  63
## 88   hope    50
## 103  lead    49
## 186  wise    43
## 156  share   42
## 86   hilarious 41
## 81   green    39
```

Here are the ten most common negative words in the 3000 tweets about the Warriors:

```
##      word freq
## 136  wait  119
## 22  breakdown 86
## 114  shooting 50
## 83   loss  45
## 119  split  30
## 59   fool  27
## 95  outburst 27
## 35  defeat  16
## 72  injured  16
## 11  beating 12
```

Extension:

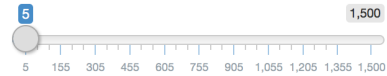
Analyzing Twitter data and meta-data allows us some insight into what people talk about and how they talk about them. Although we focused on the behavior and emotions of Cavaliers fans and Warriors fan, similar analyses can extend to other segments of Twitter users. We created a short interactive html notebook ([Twitter Mining in R.Rmd](#)) that takes a user-supplied term and searches Twitter's API for the user-specified number of tweets mentioning the term. The output is a data table of the latest 5 relevant tweets and a wordcloud based on the tweets. Additionally, the user can download a .csv file with data from `status` objects returned by `searchTwitter()`. Below are screenshots of a few examples:

1) Term = “NBA”; nTweets = 5

Twitter Mining in R

Enter a term

Select a number of tweets



Fetch Tweets

Last 5 tweets

text

RT @NBA: The @spurs (4-2) advance to the Western Conference Semifinals! #NBAPlayoffs <https://t.co/kyCh45i7no>

RT @NBA: SA closes on a 22-8 run to take the series 4-2! Tony Parker steps up with 27 points. GAME 6 FINAL - @spurs: 103 @memgrizz: 96 [htt...](https://t.co/kyCh45i7no)

RT @NBA: The @spurs (4-2) advance to the Western Conference Semifinals! #NBAPlayoffs <https://t.co/kyCh45i7no>

RT @NBA: SA closes on a 22-8 run to take the series 4-2! Tony Parker steps up with 27 points. GAME 6 FINAL - @spurs: 103 @memgrizz: 96 [htt...](https://t.co/kyCh45i7no)

great series hands down <https://t.co/5DlooJlphD>



Click 'Download' to download a .csv file with data on tweets requested

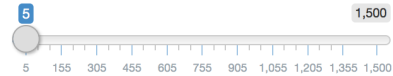
 Download

2) Term = “life”; nTweets = 5

Twitter Mining in R

Enter a term

Select a number of tweets



Fetch Tweets

Last 5 tweets

text

RT @OhioStateFB: "Coming from Ohio State, they prepare you mentally and physically." @shonrp2 talks that Buckeye life in his first intervi...

RT @extensively: IF YOU HAVE SOMEONE IN YOUR LIFE WHO MAKES YOUR SMILE BIGGER AND YOUR LAUGH A LITTLE LOUDER THEN YOU'RE LUCKY AF & HOLD ON...


RT @HugotDre: Life is too short to waste one single second with someone who doesn't appreciate and value you.

RT @i_desi_boy: Life is not a music player to listen your favourite songs.. It's a radio You must adjust yourself to every frequency & enjo...

@ColBannister @FoxNews @LouDobbs Trump thought being president would be easier than his old life. He's a buffoon &...
<https://t.co/vS64gvzkiF>

someone
&

Click 'Download' to download a .csv file with data on tweets requested

 Download

3) Term = "life"; nTweets = 123

Twitter Mining in R

Enter a term

life

Select a number of tweets

5

123

1,500

Fetch Tweets

Last 5 tweets

text

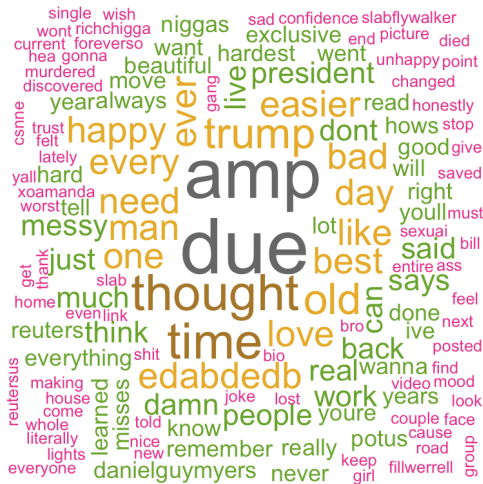
RT @deleteanxiety: Breathe. It's just a bad day, not a bad life.

So, he murdered a 19 year old girl, went to prison on a life sentence, escaped & murdered a 57 year old man? Well,...
<https://t.co/xpTn9hYxvO>

RT @kookieae: Everyone wants to be the sun that lights up ur life.But I'd rather be ur moon,so I can shine on you during ur darkest hour #J...

RT @_irenecl4: Biggest life problem <https://t.co/qAYgjesESq>

RT @marlboros: Do u ever look back at a really good point in ur life and wish you'd cherished it more



Click 'Download' to download a .csv file with data on tweets requested

Download