## 615 final

Kailun Huang 12/11/2017

For code description, Please see my Rmd File. This is just the final report

set up twitteR

## [1] "Using direct authentication"

First of all, we want to collect data from R and search players information on tweeter. Then used 'gsub' function to clean the data. Finally save the data frame to a csv file.

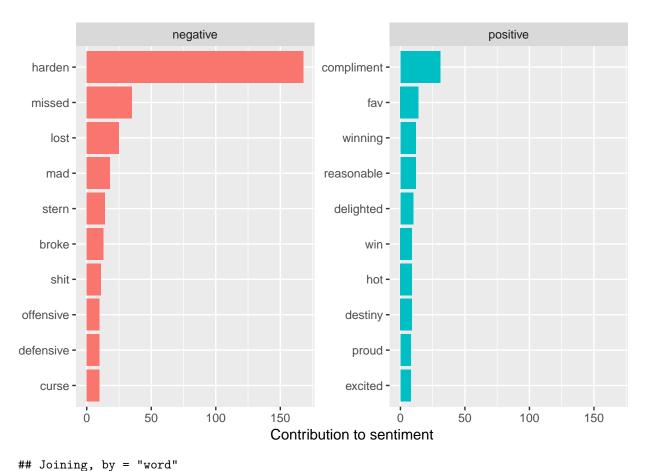
```
sure watch talk demand gop & contenders nightstime team withand status got lebron taxbeat harden know taxbeat harden know to company to company to company taxbeat harden know taxbeat
```

### Then we need to read the csv file and make word cloud. The more a specific word appears, the bigger and bolder the word will appear in the visualization. The most frequent words with Chris Paul is rockets, james and harden. It is normal because he plays in rockets and his team mates is James Harden. The key words relate to James Harden are "lebron" and "mvp". These words mean people compliment Harden performance.

### Sentiment Analysis

## Joining, by = "word"

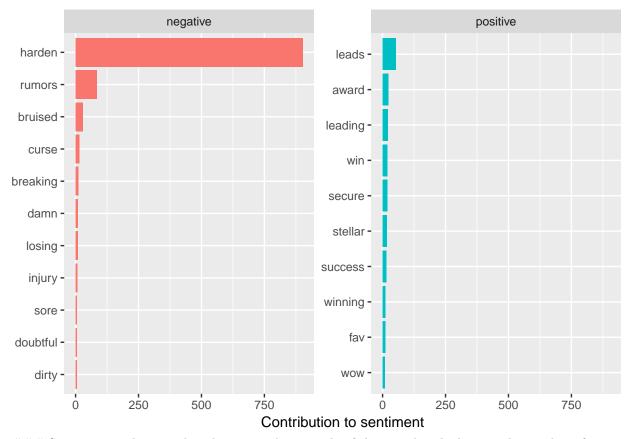
```
## # A tibble: 1,632 x 2
##
          word
##
          <chr> <int>
## 1
           paul
                  655
## 2
          chris
## 3
       rockets
                 299
## 4
          james
## 5
         harden
                  168
## 6
         lebron
                  93
## 7
          beat
                  91
## 8
           team
                  91
## 9
        chances
                   80
## 10 contenders
                   80
## # ... with 1,622 more rows
## Joining, by = "word"
## Joining, by = "word"
## # A tibble: 149 x 3
##
          word sentiment
##
          <chr> <chr> <int>
##
  1
         harden negative
                           168
         missed negative
##
   3 compliment positive
                            31
## 4
        lost negative
                            25
## 5
                            18
          mad negative
## 6
           fav positive
                            14
## 7
          stern negative
                            14
          broke negative
                            13
## 9 reasonable positive
                            12
        winning positive
                            12
## # ... with 139 more rows
## Selecting by n
```



```
## # A tibble: 998 x 2
##
               word
##
              <chr> <int>
##
   1
              james 1041
    2
##
             harden
                      902
   3
             lebron
                      266
##
##
   4
                mvp
                      162
##
   5
            rockets
                      155
##
    6
            giannis
                      130
   7 antetokounmpo
                      127
##
##
   8
              kyrie
                      125
   9
             irving
                      122
##
## 10
              total
## # ... with 988 more rows
## Joining, by = "word"
## Joining, by = "word"
## # A tibble: 95 x 3
##
         word sentiment
##
        <chr>
                  <chr> <int>
                          902
##
    1 harden negative
                           85
##
    2 rumors negative
##
        leads positive
                           52
##
   4 bruised negative
                           28
##
        award positive
                           23
```

```
##
    6 leading positive
                            21
##
    7
          win
               positive
                            19
##
       secure
               positive
                            18
               positive
                            16
##
    9 stellar
        curse
               negative
                            15
   # ... with 85 more rows
```

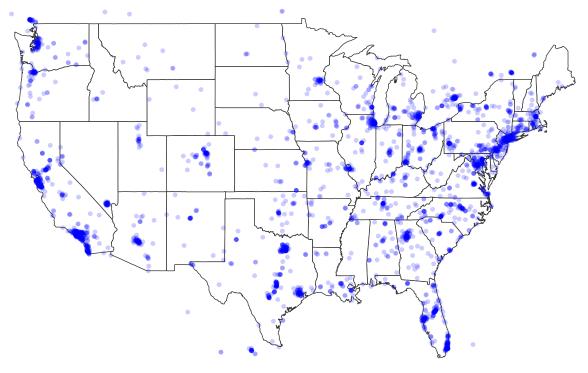
## ## Selecting by n



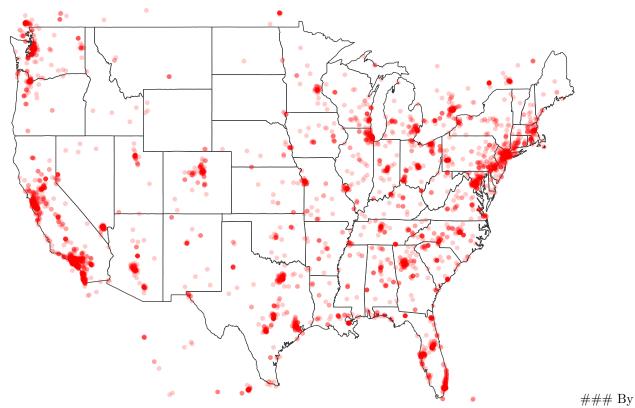
### Sentiment analysis used to determine the attitude of the text data looking at the number of positive and negative words. The most frequent negative word toward Chris Paul is harden. The most frequent positive word toward Chris Paul is compliment. The most frequent negative word toward Harden is harden. The most frequent positive word toward Harden is lead.

## ggmap for two players

- ## Capturing tweets...
- $\hbox{\tt\#\# Connection to Twitter stream was closed after 20 seconds with up to 656 tweets downloaded.}$
- ## 21104 tweets have been parsed.



- ## Capturing tweets...
- ## Connection to Twitter stream was closed after 20 seconds with up to 874 tweets downloaded.
- ## 44297 tweets have been parsed.



making ggplots of the United States, we can know where users from the states comment on these two players. Thus, we know which part in the country are interested in these two players.