

# NBA Match twitter 分析

- 資料: NBA Finals 2018 Warriors vs Cavaliers tweets
  - 發文者地區分佈
  - 發文內容分析
- 
- 載入需求函式庫

```
# Load require library
library(tidyverse)
library(leaflet)
library(ggmap)
library(igraph)
library(tm)
library(wordcloud)
```

- 繪製發文者地區分佈圖

```
# Load the location data
locations <- read.csv('data/locations.csv')

# Create Interactive map
leaflet(data=locations) %>%
  addTiles() %>%
  addCircles (lat=locations$lat, lng=locations$lon)
```



- 產生發文內容文字雲

```
# Load tweets data
tweets.df <- read.csv('data/TweetsNBA.csv')
tweets.df <- tweets.df %>%
  mutate_at(vars(text), as.character) %>%
  mutate_at(vars(lang), factor) %>%
  mutate(lang=recode(lang, en="English", es="Spanish"))
```

```

# Most frequent word's wordcloud
# English Tweets
en_tweets <- tweets.df %>%
  filter(lang=="English")

# Function to clean the corpus
removeURL <- function(x) gsub("http[[:alnum:]]*", "", x)
cleanCorpus <- function(corpus){

  corpus.tmp <- tm_map(corpus, removePunctuation)
  corpus.tmp <- tm_map(corpus.tmp, stripWhitespace)
  corpus.tmp <- tm_map(corpus.tmp, content_transformer(tolower))
  corpus.tmp <- tm_map(corpus.tmp, content_transformer(removeURL))
  v_stopwords <- c(stopwords("english"),
    "thats", "weve", "hes", "theres", "ive",
    "im", "will", "can", "cant",
    "dont", "youve", "us", "youre", "youll", "theyre", "whats", "didnt")
  corpus.tmp <- tm_map(corpus.tmp, removeWords, v_stopwords)
  corpus.tmp <- tm_map(corpus.tmp, removeNumbers)
  return(corpus.tmp)

}

# Function to find the frequency of each word
frequentTerms <- function(text){

  s.cor <- Corpus(VectorSource(text))
  s.cor.cl <- cleanCorpus(s.cor)
  s.tdm <- TermDocumentMatrix(s.cor.cl)
  s.tdm <- removeSparseTerms(s.tdm, 0.999)
  m <- as.matrix(s.tdm)
  word_freqs <- sort(rowSums(m), decreasing=TRUE)
  dm <- data.frame(word=names(word_freqs), freq=word_freqs)
  return(dm)

}

# Generate wordcloud
dm <- frequentTerms(tweets.df$description)
dm2 <- frequentTerms(en_tweets$text)
wordcloud(dm2$word, dm2$freq, min.freq=30, colors=brewer.pal(8,"Dark2"),
max.words=200)

```

