Twitter Analysis of AlphaGo

Liuqian Zhang

Dec 4th, 2017

1. Project Description

This is a project analyzing twitter data containing the keyword 'AlphaGo' from 2017-11-28 to 2017-12-04. In order to explore the brand perception of AlphaGo on Twitter, word frequency analysis, sentiment anlysis, day part analysis have been conducted.

2. Data Importing

2.1 Connect to twitter api

```
api_key <-  "Y96v0OOcyQfNb6Xtdq5OMKh65"
api_secret <- "QMamXhqcqyfGMDMtHhZVECAdvEvDJ0t2Tq406382qaUVA7YWgl"
access_token <- "927637446975770624-0HMkuqqvY4uONQH0Hk61UGjDvjSoyRV"
access_token_secret <- "VdyedkmyhuCLV1aOM9FvxRBT4EO1bAcb8BRYBA7tJdElA"

twitteR::setup_twitter_oauth(api_key,
                     api_secret,
                     access_token,
                     access_token_secret)
```

```
## [1] "Using direct authentication"
```

2.2 read data

```
data <- searchTwitter('alphago',since='2017-01-01',lang = "en", n=1000) %>% twListToDF()
head(data)
```

```
##
## 1 RT @carlcarrie: About Alpha Go Zero \n\nPrimers\n¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª¡ª\n\nPresenta

## 2                                                                The Cyberlaw Podcast: Did AlphaGo Lau
## 3            RT @DeepMindAI: Explore how #AlphaGo rates thousands of the most popular opening moves i

## 4            In contrast to #AlphaGo, the shocking #AI news here isn't the ease which with which #Alp
https://t.co/PqN9mIXcUI
## 5
## 6            RT @tegmark: In contrast to #AlphaGo, the shocking #AI news here isn't the ease which wi

##   favorited favoriteCount replyToSN             created truncated
## 1     FALSE             0      <NA> 2017-12-19 04:06:53     FALSE
## 2     FALSE             0      <NA> 2017-12-19 03:45:00     FALSE
## 3     FALSE             0      <NA> 2017-12-19 03:43:49     FALSE
## 4     FALSE             0      <NA> 2017-12-19 03:22:09      TRUE
```

```
## 5     FALSE             0   PragTob 2017-12-19 03:16:56     FALSE
## 6     FALSE             0      <NA> 2017-12-19 03:16:49     FALSE
##   replyToSID                 id replyToUID
## 1       <NA> 942969461732024320       <NA>
## 2       <NA> 942963951960297472       <NA>
## 3       <NA> 942963655234203649       <NA>
## 4       <NA> 942958204518129665       <NA>
## 5       <NA> 942956888353681409  334768270
## 6       <NA> 942956861505937408       <NA>
##                                                                          statusSource
## 1  <a href="https://about.twitter.com/products/tweetdeck" rel="nofollow">TweetDeck</a>
## 2                            <a href="http://bufferapp.com" rel="nofollow">Buffer</a>
## 3 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
## 4          <a href="http://convey.pro" rel="nofollow">Convey: Make it post for you</a>
## 5                   <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
## 6                   <a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>
##       screenName retweetCount isRetweet retweeted longitude latitude
## 1    seldomawake            2      TRUE     FALSE        NA       NA
## 2 PatrickCMiller            0     FALSE     FALSE        NA       NA
## 3        fbiosje          366      TRUE     FALSE        NA       NA
## 4 Aaron_J_Hardin            0     FALSE     FALSE        NA       NA
## 5 TheNewTreasury            0     FALSE     FALSE        NA       NA
## 6   janemccourt1          295      TRUE     FALSE        NA       NA
```

```
# write.csv(data, "AlphaGo.csv")

# read the previously saved twitter data in order to keep the analysis consistant
data <- read.csv("AlphaGo.csv", stringsAsFactors = FALSE)
```

3. Analysis

3.1 Word Frequency

```
## tokenize

tidy_data <- data %>% unnest_tokens(word,text)

## remove stop words

data(stop_words)

tidy_data <-  tidy_data %>%
  anti_join(stop_words)
```

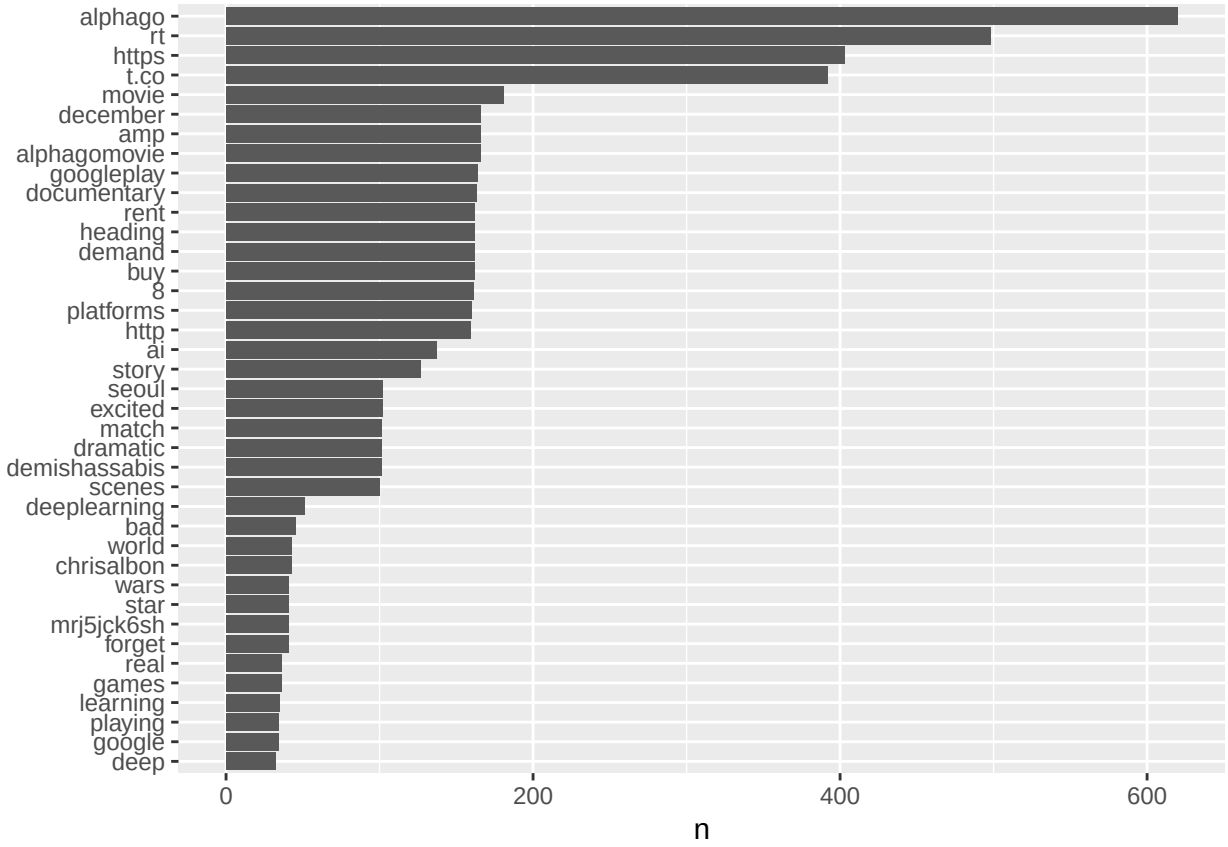
```
## Joining, by = "word"
```

```
tidy_data %>%  count(word, sort=TRUE)
```

```
## # A tibble: 1,114 x 2
##            word     n
##           <chr> <int>
##  1      alphago   620
##  2           rt   498
```

```
##  3        https   403
##  4         t.co   392
##  5        movie   181
##  6 alphagomovie   166
##  7          amp   166
##  8     december   166
##  9   googleplay   164
## 10  documentary   163
## # ... with 1,104 more rows
```

```
# word frequency list
tidy_data %>%
  count(word, sort = TRUE) %>%
  filter(n > 30) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip()
```

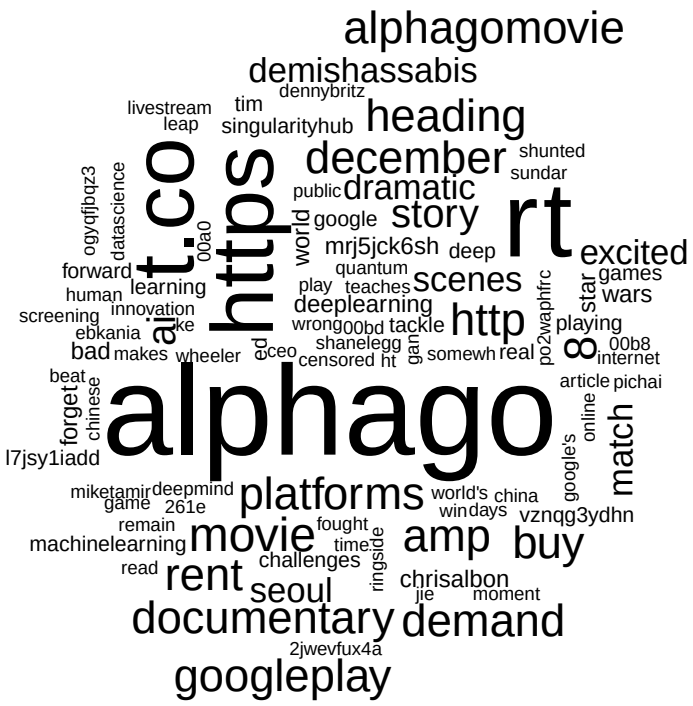


From the frequency list, we can see besides the word "alphago", the website address "https://t.co/mRJ5JcK6Sh" appeared around 41 times in this week. It is a link leads to the alphago movie official site.

The words "movie", "alphagomovie", "documentary", "december","googleplay", "buy", "demand", "heading", "platforms", and "rent" each showed up around 160 times. This pattern looks like a tweet being retweeted many times, or seem to be a marketing campaign. This guess will be validated in section 3.3.

```
library(wordcloud)
tidy_data %>%
  anti_join(stop_words) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100))
```

```
## Joining, by = "word"
```



##3.2 Sentiment Analysis

```
# get_sentiments("afinn")
#
# get_sentiments("bing")
#
# get_sentiments("nrc")
```

3.2.1 Fear Words and Joy Words Frequencies

```
nrcfear <- get_sentiments("nrc") %>%
  filter(sentiment == "fear")

tidy_data %>%
  inner_join(nrcfear) %>%
  count(word, sort = TRUE)
```

```
## Joining, by = "word"
```

```
## # A tibble: 27 x 2
##            word     n
##           <chr> <int>
##  1          bad    45
##  2 intelligence     8
##  3     powerful     7
##  4    challenge     5
##  5        watch     3
##  6          foe     2
##  7      advance     1
##  8        alien     1
##  9         bang     1
## 10      beating     1
## # ... with 17 more rows
```

```
nrcjoy <- get_sentiments("nrc") %>%
  filter(sentiment == "joy")

tidy_data %>%
  inner_join(nrcjoy) %>%
  count(word, sort = TRUE)
```

```
## Joining, by = "word"
```

```
## # A tibble: 33 x 2
##            word     n
##           <chr> <int>
##  1      excited   102
##  2         star    41
##  3 intelligence     8
##  4     powerful     7
##  5     champion     3
##  6     festival     3
##  7         love     3
##  8        teach     3
##  9         deal     2
## 10     exciting     2
## # ... with 23 more rows
```

The word "mother" appear in both lists, so we need to fix that by adding it to stopword list. The word "learning" appear in positive words, but in this context, it mostly related to machine learning, which is a neutural word. So we add it to positive stop words. On the other hand, the word "player" appears in negative lists. I don't consider this as negative word in this case, because it can be used to describe players of the Go game. Need to add it to the negative stop words list.

```
## adjust by removing custom stop words
custom_stop_words <- bind_rows(data_frame(word = c("mother", "player","learning","intelligence"),
                                          lexicon = c("custom")),
                               stop_words)

(fearlist <- tidy_data %>%
  anti_join(custom_stop_words) %>%
  inner_join(nrcfear) %>%
  count(word, sort = TRUE))
```

```
## Joining, by = "word"
## Joining, by = "word"
```

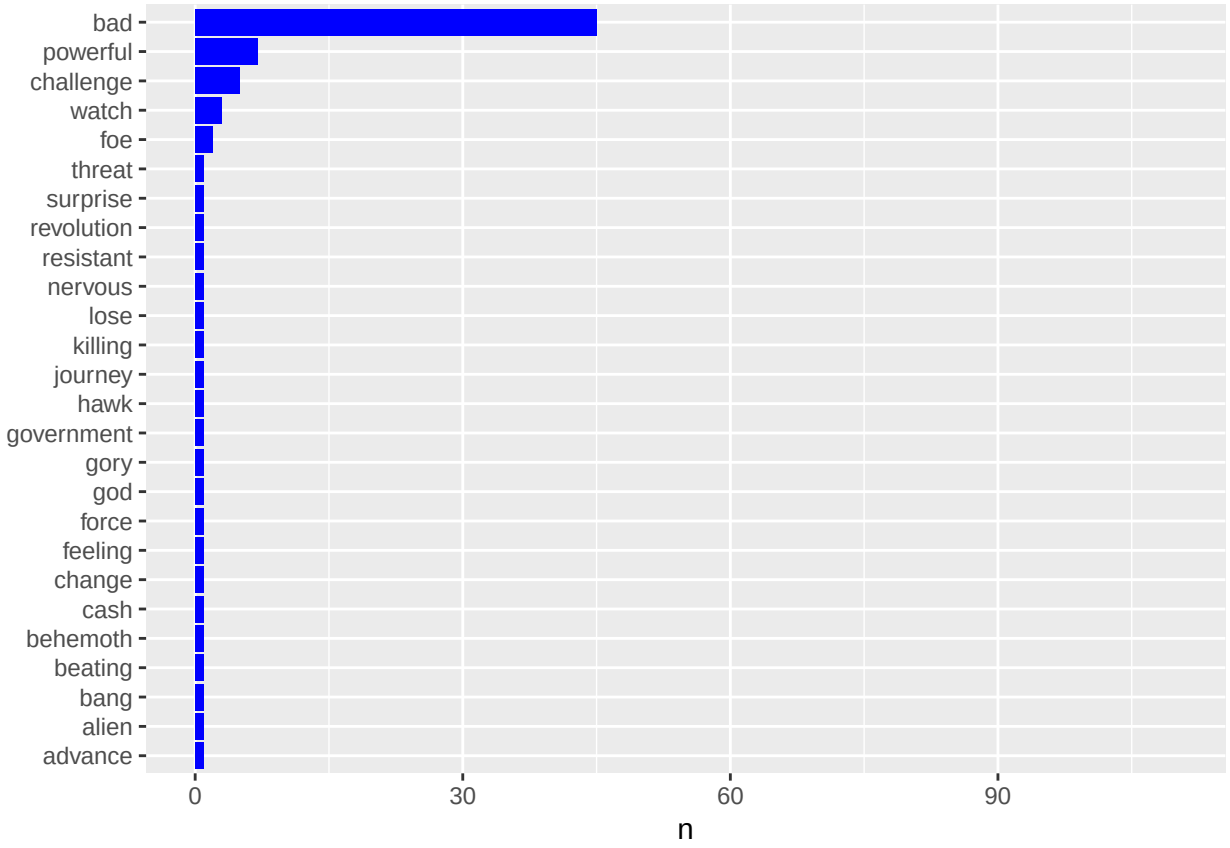
```
## # A tibble: 26 x 2
##         word     n
##        <chr> <int>
##  1       bad    45
##  2  powerful     7
##  3 challenge     5
##  4     watch     3
##  5       foe     2
##  6   advance     1
##  7     alien     1
##  8      bang     1
##  9   beating     1
## 10  behemoth     1
## # ... with 16 more rows
```

```
(joylist <- tidy_data %>%
  anti_join(custom_stop_words) %>%
  inner_join(nrcjoy) %>%
  count(word, sort = TRUE))
```

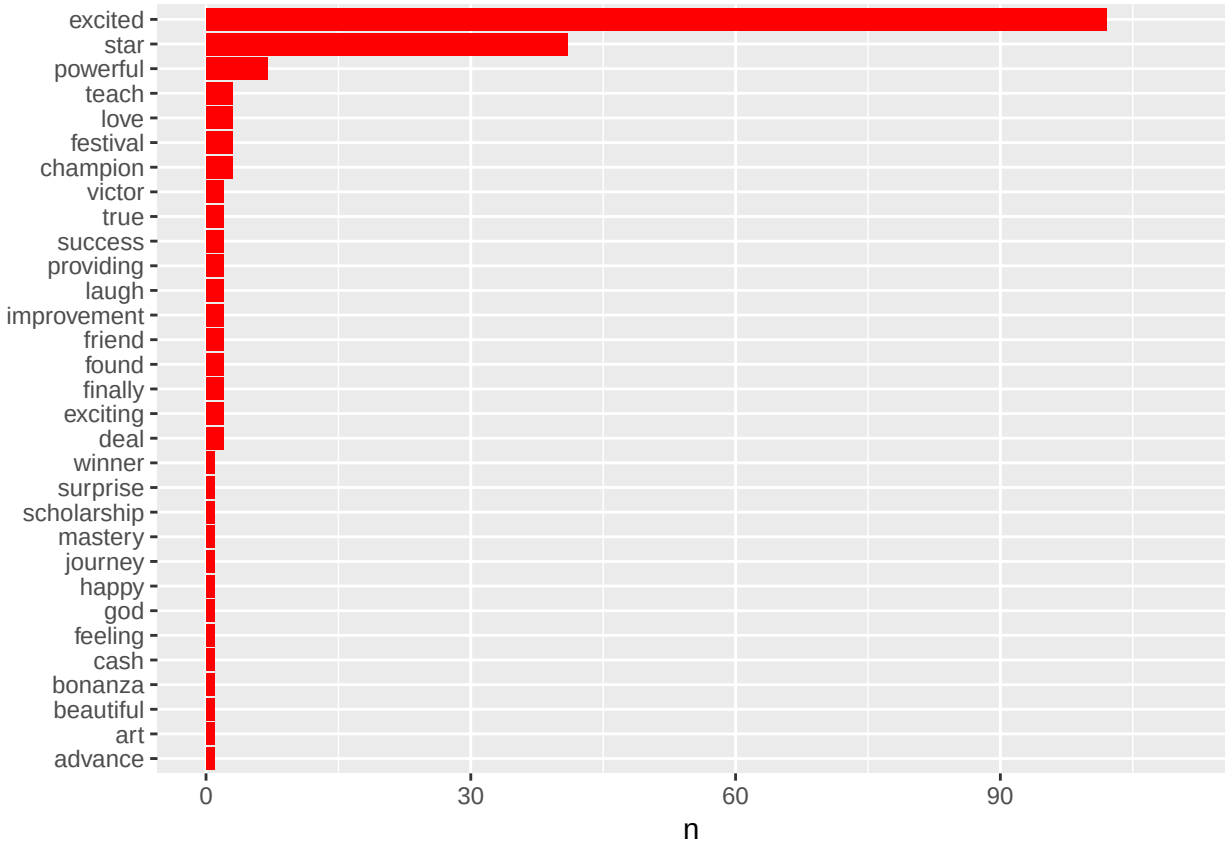
```
## Joining, by = "word"
## Joining, by = "word"
```

```
## # A tibble: 31 x 2
##        word     n
##       <chr> <int>
##  1  excited   102
##  2     star    41
##  3 powerful     7
##  4 champion     3
##  5 festival     3
##  6     love     3
##  7    teach     3
##  8     deal     2
##  9 exciting     2
## 10  finally     2
## # ... with 21 more rows
```

```
fearlist %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col(fill="blue") +
  xlab(NULL) +
  coord_flip() +
  scale_y_continuous(limits = c(0,110))
```




```
joylist %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col(fill="red") +
  xlab(NULL) +
  coord_flip() +
  scale_y_continuous(limits = c(0,110))
```



From the tweets, the word "excited" are mentioned over 100 time, indicating it is an important sentiment of people towards AlphaGo.

The word "bad" appeared over 40 times. This is also an important sentiment here. Some words like "threat", and "nervous" may indicate something about people's worry. But because the sample size of the twitters is not large enough, these words only appear once. We cannot say whether people feel fear about AlphaGo.

3.2.2 Positive and Negative Words Frequencies

```
bingpos <- get_sentiments("bing") %>%
  filter(sentiment == "positive")

(poslist <- tidy_data %>%
  anti_join(custom_stop_words) %>%
  inner_join(bingpos) %>%
  count(word, sort = TRUE))
```

```
## Joining, by = "word"
## Joining, by = "word"
```

```
## # A tibble: 50 x 2
##            word     n
##           <chr> <int>
##  1      excited   102
##  2   innovation    14
##  3          win    10
##  4       master     7
```

```
##  5     powerful     7
##  6      amazing     5
##  7          top     4
##  8 breakthrough     3
##  9     champion     3
## 10         cool     3
## # ... with 40 more rows
```

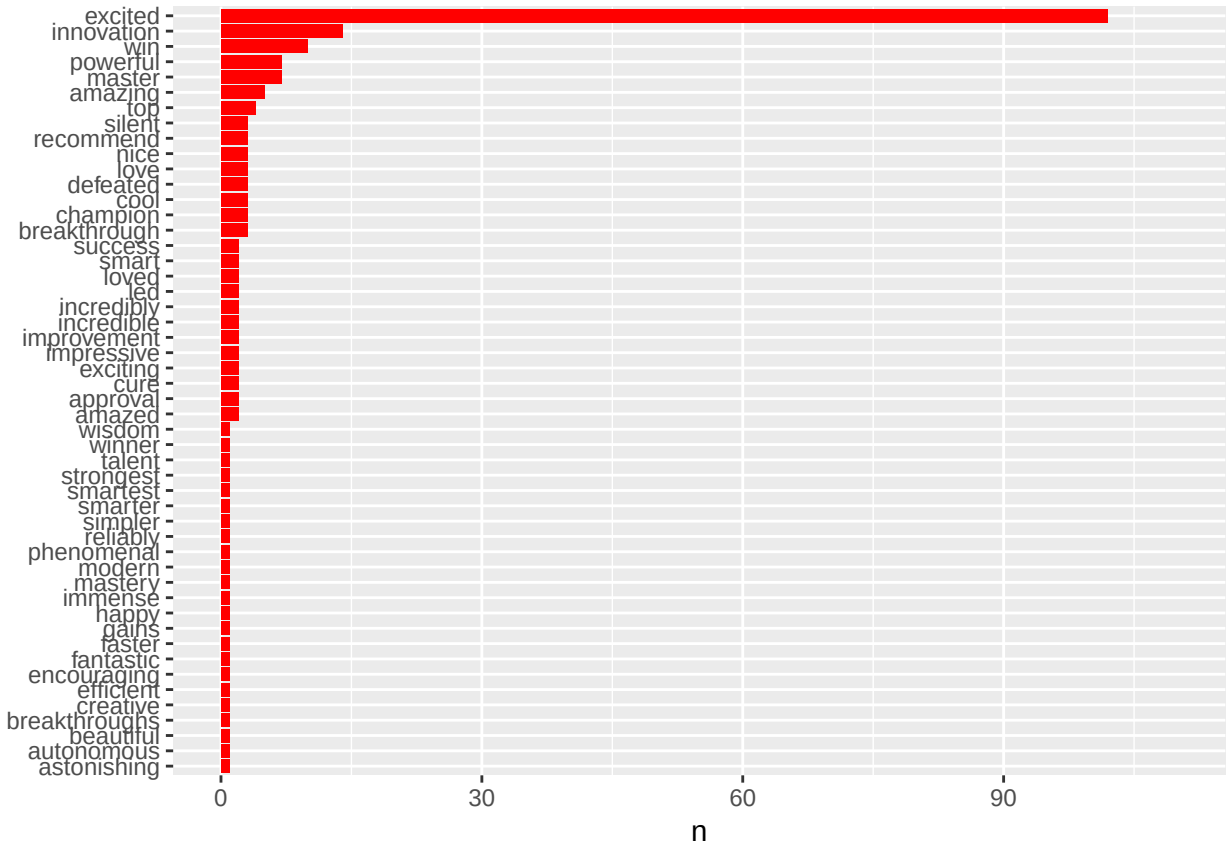
```
bingneg <- get_sentiments("bing") %>%
  filter(sentiment == "negative")

(neglist <- (tidy_data %>%
  inner_join(bingneg) %>%
  anti_join(custom_stop_words) %>%
  count(word, sort = TRUE)))
```

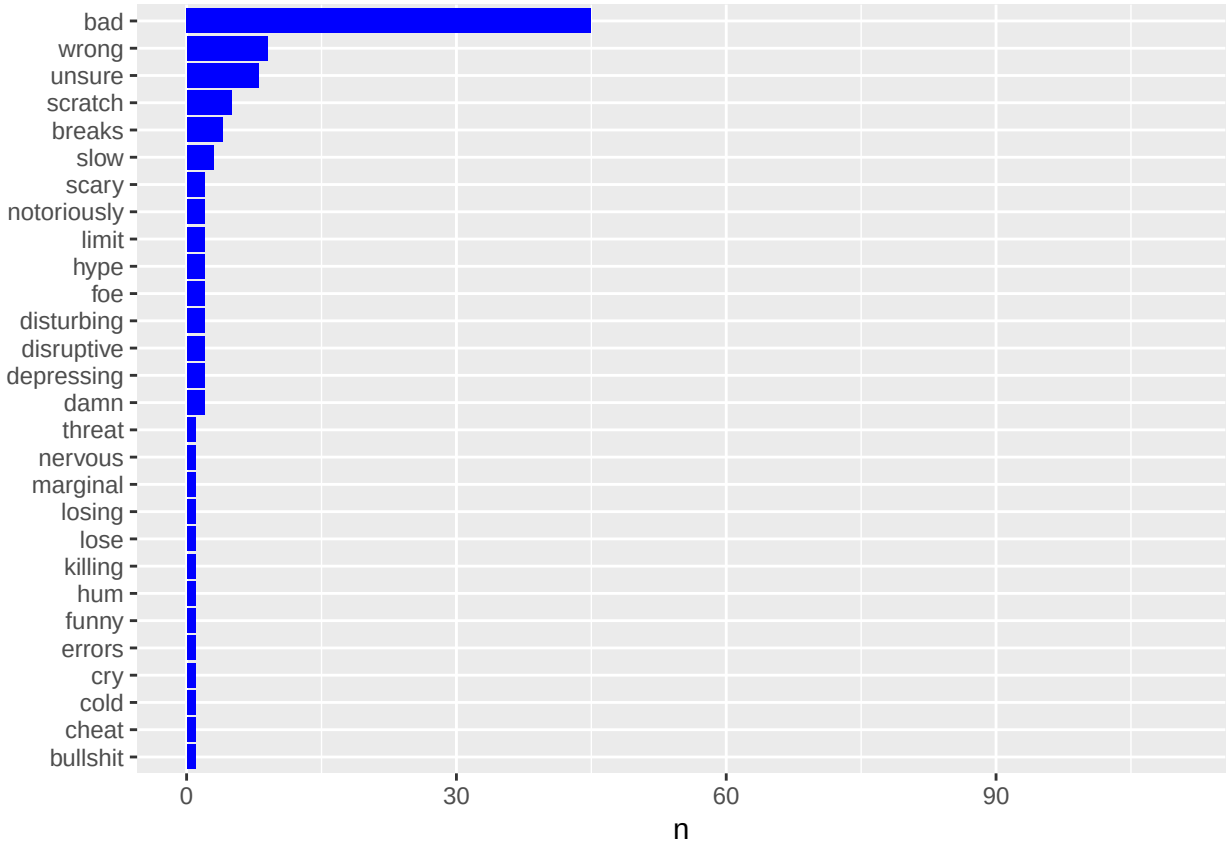
```
## Joining, by = "word"
## Joining, by = "word"
```

```
## # A tibble: 28 x 2
##          word     n
##         <chr> <int>
##  1        bad    45
##  2      wrong     9
##  3     unsure     8
##  4    scratch     5
##  5     breaks     4
##  6       slow     3
##  7       damn     2
##  8 depressing     2
##  9 disruptive     2
## 10 disturbing     2
## # ... with 18 more rows
```

```
poslist %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col(fill="red") +
  xlab(NULL) +
  coord_flip() +
  scale_y_continuous(limits = c(0,110))
```




```
neglist %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col(fill="blue") +
  xlab(NULL) +
  coord_flip() +
  scale_y_continuous(limits = c(0,110))
```



3.2.3 Word Clouds of Positive and Negative Words

```
tidy_data %>%
anti_join(custom_stop_words) %>%
inner_join(get_sentiments("bing")) %>%
count(word, sentiment, sort = TRUE) %>%
acast(word ~ sentiment, value.var = "n", fill = 0) %>%
comparison.cloud(colors = c("#F8766D", "#00BFC4"),
                 max.words = 100)
```

```
## Joining, by = "word"
## Joining, by = "word"
```

From the first word cloud, we can see during this period, people talk a lot about the alphago documentary. And the url of the movie appears many times.

3.3 Topic Hotness by Date

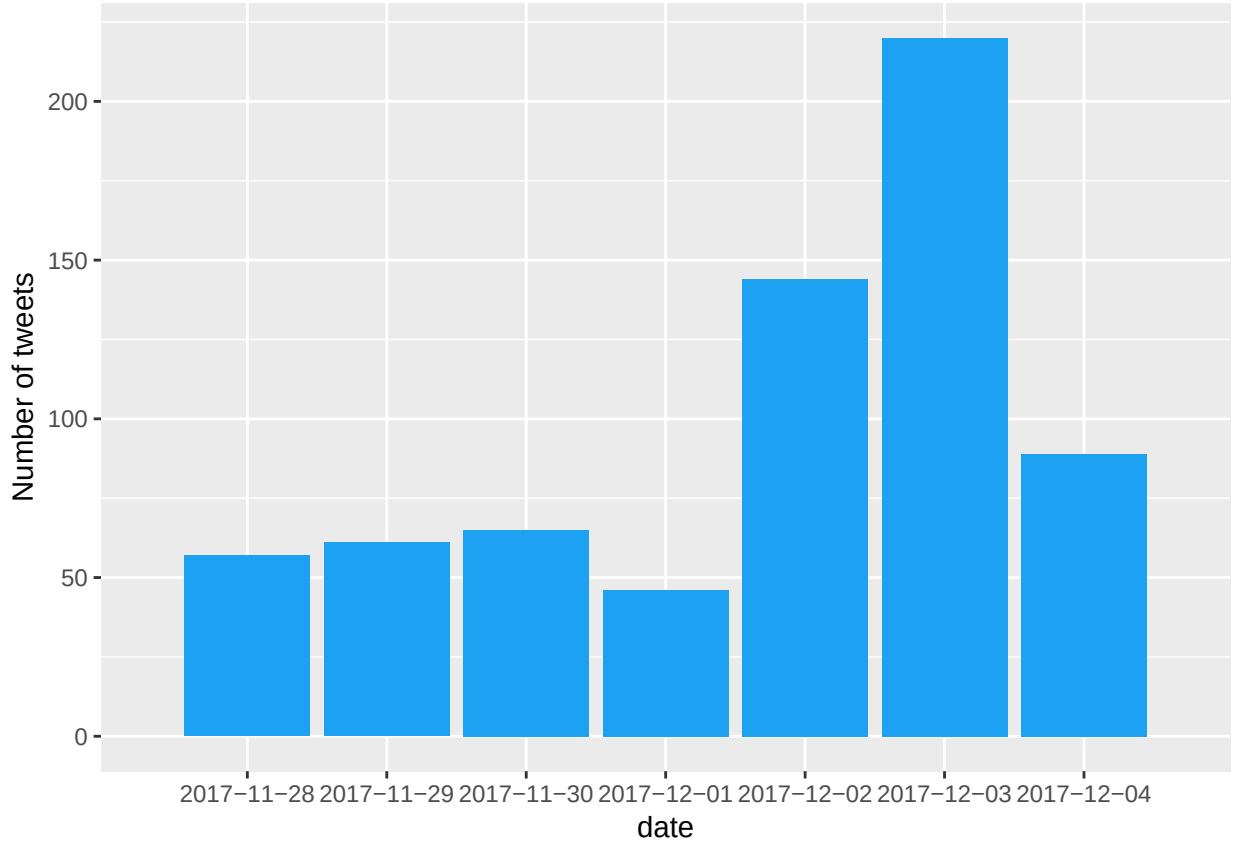
3.3.1 EDA

```
# by date
(dates <- data %>%
  mutate(date=date(created)) %>%
  group_by(date) %>%
  summarise(date_n = n()))
```

```
## # A tibble: 7 x 2
##         date date_n
##       <date>  <int>
## 1 2017-11-28     57
## 2 2017-11-29     61
## 3 2017-11-30     65
## 4 2017-12-01     46
## 5 2017-12-02    144
## 6 2017-12-03    220
## 7 2017-12-04     89
```

```
ggplot() +
  geom_bar(data=dates,
```

```
          aes(date,date_n),
          stat = "identity",
          fill="#1DA1F2") +
  ylab("Number of tweets") +
  scale_x_discrete(limits=lubridate::date(dates$date),labels=lubridate::date(dates$date))
```



The number of tweets about AlphaGo in December 2nd and 3rd are higher than other dates. ###3.3.2 Analysis of Percentages of Popular Tweets The following is anlyzing which tweets or topics are pushing the volumns high in December 2nd and 3rd. The number of tweets in dec 2nd (showing first 20 rows)

```
# The number of tweets in dec 2nd (showing first 20 rows)

data %>%
  filter(date(created)==c("2017-12-02")) %>%
  .$text %>%
  head(20)
```

```
##  [1] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
##  [2] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
##  [3] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
##  [4] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
##  [5] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
##  [6] "RT @ShaneLegg: The story behind AlphaGo is now available online :-) https://t.co/OgYQfjbqZ3"
##  [7] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
##  [8] "Got around to watching the AlphaGo doco ¨C loved it! Esp seeing how nervous @DeepMindAI were b
##  [9] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
```

```
## [10] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
## [11] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
## [12] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
## [13] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
## [14] "DeepMind creates an algorithm with AlphaGo! #AI #ReinforcementLearning #ML #Analytics #AlphaGo
## [15] "AlphaGo Zero shows how business is losing the innovation game https://t.co/pWXC7ewoiX via @ins
## [16] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
## [17] "AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see this so bad."
## [18] "\"[AlphaGo] was very impressive. [...] I think it did require a lot. It wouldn't surprise me i
## [19] "AlphaGo Zero vs. Master with Michael Redmond 9p: Game 2 https://t.co/tjfzwQ3vLX"
## [20] "Google's AlphaGo Zero taught itself to become the greatest Go player in history https://t.co/k
```

```
data %>%
  filter(date(created)==c("2017-12-02")) %>%
  nrow()
```

```
## [1] 144
```

Number of tweets with content "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh/n/nForget Star Wars, I want to see this so bad."

```
# number of tweets with content "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget

data %>%
  filter(date(created)==c("2017-12-02")) %>%
  filter(grepl("AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see this so
  nrow
```

```
## [1] 11
```

```
paste(round(11/144*100,2), "% of the tweets in Dec 2nd are retweeting the one from @chrisalbon")
```

```
## [1] "7.64 % of the tweets in Dec 2nd are retweeting the one from @chrisalbon"
```

Number of tweets containing GooglePlay in dec 2nd

```
# number of tweets containing GooglePlay in dec 2nd
data %>%
  filter(date(created)==c("2017-12-02")) %>%
  filter(grepl("#AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading to other o
  nrow
```

```
## [1] 85
```

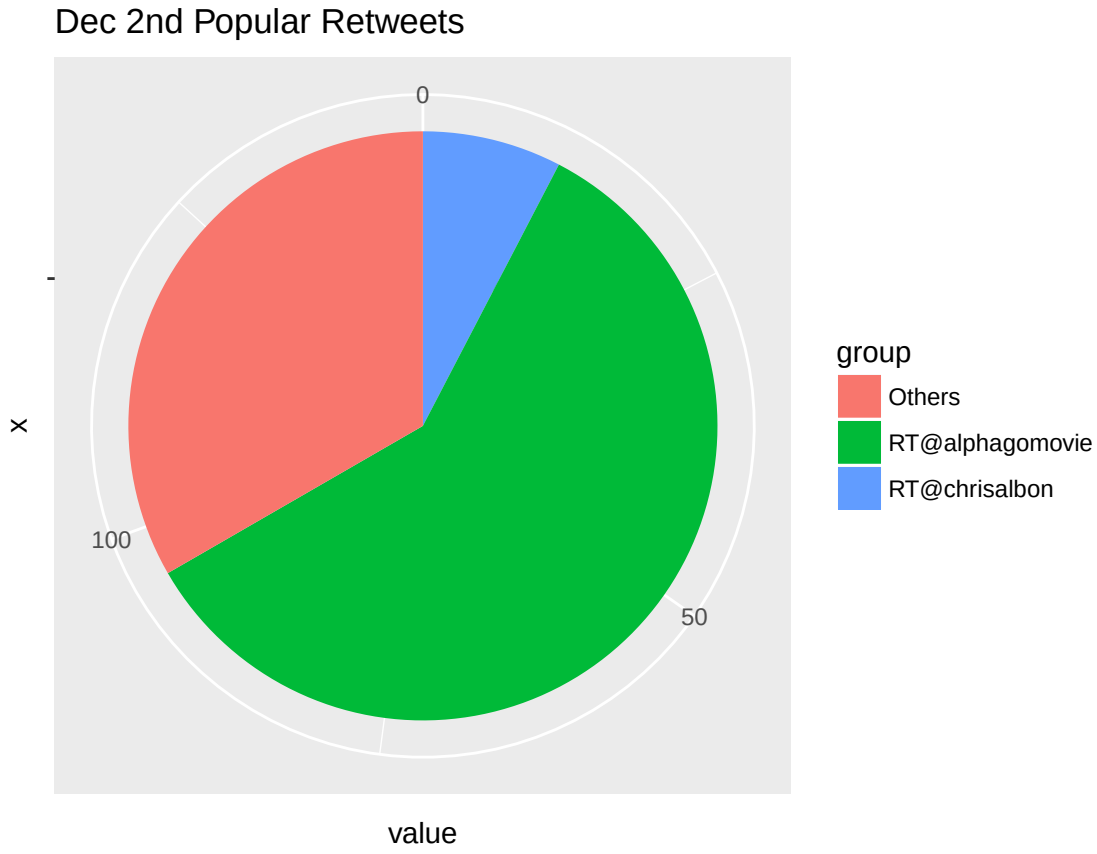
```
paste(round(85/144*100,2), "% of the tweets in Dec 2nd are retweeting the one created by @alphagomovie
```

```
## [1] "59.03 % of the tweets in Dec 2nd are retweeting the one created by @alphagomovie advertising it
```

pie chart of popular retweets

```
df1 <- data.frame(
  group = c("RT@alphagomovie", "RT@chrisalbon", "Others"),
  value = c(85, 11, 144-11-85)
  )

ggplot(df1, aes(x="", y=value, fill=group))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start=0) +
  ggtitle("Dec 2nd Popular Retweets")
```



Tweets on Dec 3rd.

```
# The number of tweets in dec 3rd (showing first 20 rows)

data %>%
  filter(date(created)==c("2017-12-03")) %>%
  .$text %>%
  head(20)
```

```
##  [1] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
##  [2] "@alphagomovie @DeepMindAI @GooglePlay A beautiful documentary: both Lee Sedol and the AlphaGo
https://t.co/x6FaD8nbHI"
##  [3] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
##  [4] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
##  [5] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
##  [6] "Google DeepMind knows when to hold ¡®em, when to fold ¡®em https://t.co/0sppzystJt"
##  [7] "RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes story of
"
##  [8] "RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes story of
"
##  [9] "RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes story of
"
## [10] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
```

```
"
## [11] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
## [12] "RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes story of
"
## [13] "RT @chrisalbon: AlphaGo documentary https://t.co/mRJ5JcK6Sh\n\nForget Star Wars, I want to see
## [14] "RT @mrkgrnao: Bad idea: I would love to see a Vim-using AI do to the editor what AlphaGo did t
"
## [15] "How big a deal is #Google's latest AlphaGo breakthrough? https://t.co/xCKQv7w6i5 https://t.co/
## [16] "RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes story of
"
## [17] "RT @TimandraHarknes: So AI can play games? Call me when AlphaGo Zero or IBM Watson even begins
"
## [18] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
## [19] "RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes story of
"
## [20] "RT @alphagomovie: #AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading
"
```

```
data %>%
  filter(date(created)==c("2017-12-03")) %>%
  nrow()
```

```
## [1] 220
```

Number of tweets about alphago movie on GooglePlay on dec 3rd

```
# number of tweets containing GooglePlay on dec 3rd
data %>%
  filter(date(created)==c("2017-12-03")) %>%
  filter(grepl("#AlphaGo Movie is now available to rent and buy on @GooglePlay &amp; heading to other o
  nrow
```

```
## [1] 65
```

```
paste(round(65/220*100,2), "% of the tweets on Dec 3rd are retweeting the one created by @alphagomovie a
```

```
## [1] "29.55 % of the tweets on Dec 3rd are retweeting the one created by @alphagomovie advertising it
```

Number of tweets retweeting the one from demishassabis

```
data %>%
  filter(date(created)==c("2017-12-03")) %>%
  filter(grepl("RT @demishassabis: So excited that the #AlphaGo documentary and the behind the scenes st
  nrow
```

```
## [1] 85
```

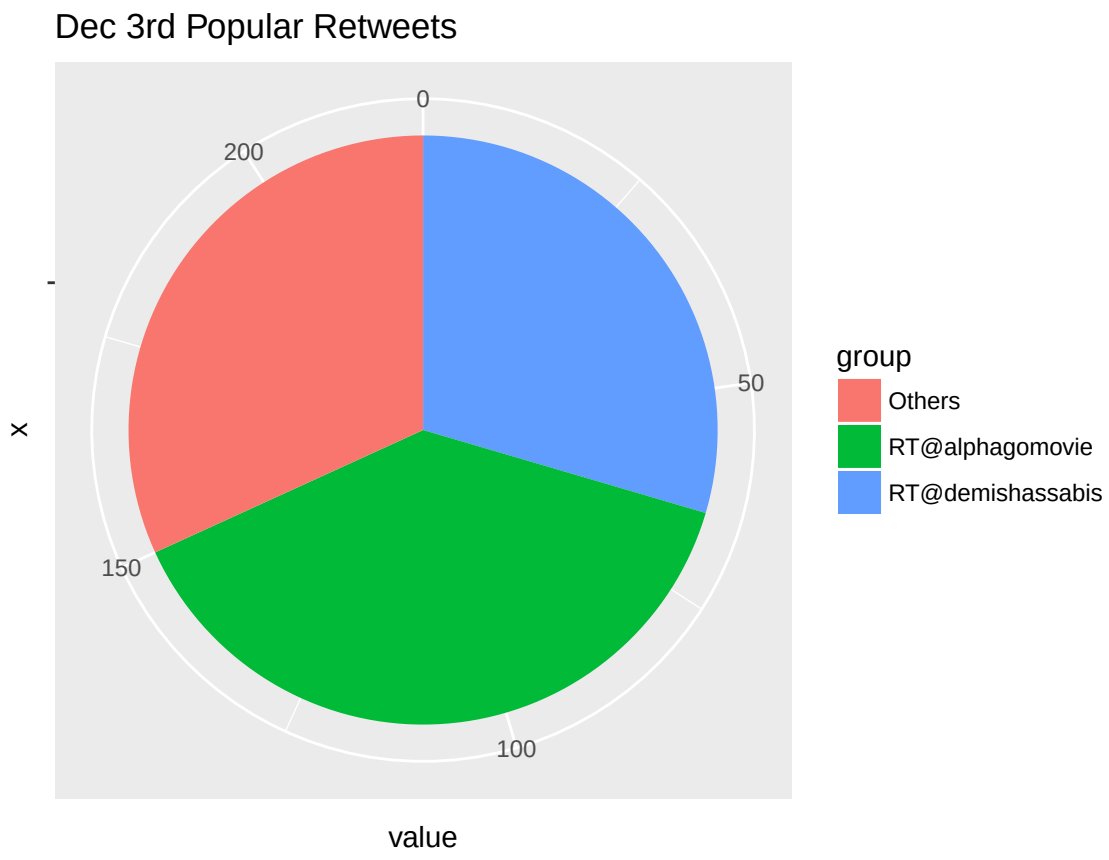
```
paste(round(85/220*100,2), "% of the tweets on Dec 3rd are retweeting the one created by @demishassabis
```

```
## [1] "38.64 % of the tweets on Dec 3rd are retweeting the one created by @demishassabis."
```

pie chart of popular retweets

```
df2 <- data.frame(
  group = c("RT@alphagomovie", "RT@demishassabis", "Others"),
  value = c(85, 65, 220-65-85)
  )
```

```
ggplot(df2, aes(x="", y=value, fill=group))+
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start=0) +
  ggtitle("Dec 3rd Popular Retweets")
```



3.4 Topic Hotness By Hour

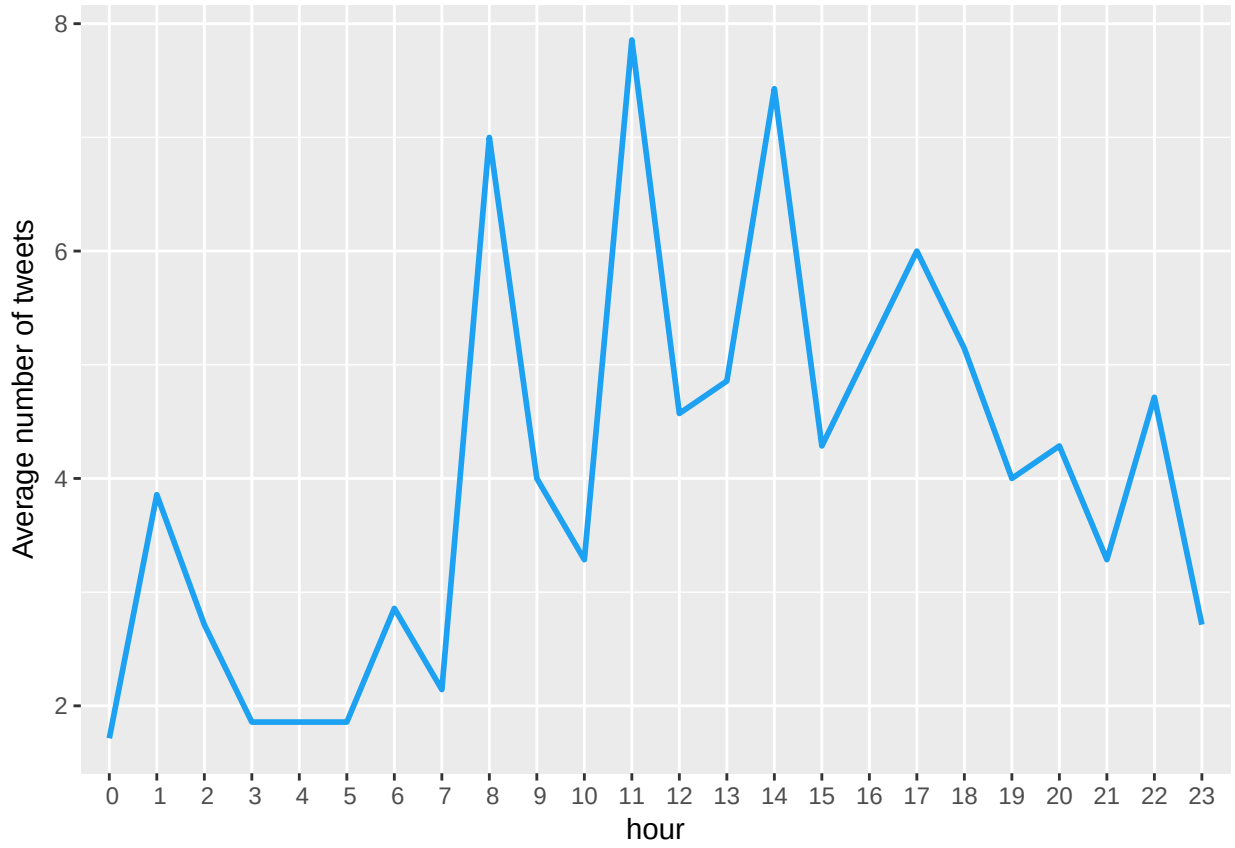
```
# by hour
(hourly <- data %>%
  mutate(hour=hour(created)) %>%
  group_by(hour) %>%
  summarise(hour_n = n()))
```

```
## # A tibble: 24 x 2
##     hour hour_n
##    <int>  <int>
##  1     0     12
##  2     1     27
##  3     2     19
##  4     3     13
##  5     4     13
##  6     5     13
##  7     6     20
##  8     7     15
```

```
##  9     8     49
## 10     9     28
## # ... with 14 more rows
```

```
label1 <-seq(0,23,1)

ggplot() +
  geom_line(data=hourly,
            aes(hour,hour_n/7),
            stat = "identity",
            color="#1DA1F2",
            size=1) +
  ylab("Average number of tweets") +
  scale_x_discrete(limits=label1)
```



The average number of tweets about AlphaGo is peaked at 8am, 11am, and 2pm.

4. Conclusion

From the word frequency analysis, I found out that the movie and the documentary are being talked about very frequently abour alphago in this week on Twitter. From the sentiment analysis, both joy and fear can be seen. From the topic hotness by date, we can see that on Dec 2nd and 3rd, the fequency of alphago tweets are about 3 times higher than other days in the week. Among them, the retweets of certain contents about the alphago movie took up more than half of the total amounts.

From this project, I think analysis of twitter data can be applied to monitoring online marketing campaigns,

because it is useful in reflecting the sentiments, word frequencies and number of retweets.