Blockchain on Twitter: A Sword of Damocles?

Shan Shan

12/18/2017

Blockchain is the world's leading software platform for digital assets. Google Trends shows that Blockchain has a boost in the year 2017. What is Twitterers attitudes toward blockchain? My project analyzes Twitter data with the keyword blockchain from 2017 01 01 to 2017 12 18. To explore the brand perception of blockchain on Twitter, I have conducted word frequency analysis, sentiment analysis, and mapping analysis.

```
#Set up Twitter

library(reshape)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following objects are masked from 'package:reshape':
##
##     colsplit, melt, recast
```

```
library(devtools)
library(twitteR)
library(tm)
```

```
## Warning: package 'tm' was built under R version 3.4.3
```

```
## Loading required package: NLP
```

```
library(stringr)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(tidytext)
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## annotate(): ggplot2, NLP
## expand():   tidyr, reshape
## filter():   dplyr, stats
## id():       dplyr, twitteR
## lag():      dplyr, stats
## location(): dplyr, twitteR
## rename():   dplyr, reshape
```

```
library(streamR)
```

```
## Loading required package: RCurl
```

```
## Loading required package: bitops
```

```
##
## Attaching package: 'RCurl'
```

```
## The following object is masked from 'package:tidyr':
##
##     complete
```

```
## Loading required package: rjson
```

```
library(ROAuth)
library(reshape)
library(dplyr)
library(ggplot2)
library(plotly)
```

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:reshape':
##
##     rename
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(grid)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:reshape':
##
##     stamp
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(graphTweets)
```

```
## help('graphTweets') for examples
```

```
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:lubridate':
##
##     %--%, union
```

```
## The following object is masked from 'package:plotly':
##
##     groups
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:purrr':
##
##     compose, simplify
```

```
## The following object is masked from 'package:tidyr':
##
##     crossing
```

```
## The following object is masked from 'package:tibble':
##
##     as_data_frame
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
library(streamR)
```

I. Read Twitter Data

- 1.1 Connect to Twitter API

```
api_key <- "Y9sY5tXFzxNGvUJYd1Um0Z1Ux"
api_secret <- "HHnbEWbyLLuSqg3tvjPhYfl5HWQnzh2Jxxqo1ZbmVv8FgPrJWW"
access_token <- "817556180717412352-nPKehdlSsJ0pdRzFI25uhzowIBMKqR7"
access_token_secret <- "bWyPN9Xsi3kSmhlKlk0ls9cDosDcleYMpSSjkH7Q7WnHF"

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

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

```
## Warning in strptime(x, fmt, tz = "GMT"): unknown timezone 'zone/tz/2017c.
## 1.0/zoneinfo/America/New_York'
```

- 1.2 Read data

```
blch <- searchTwitter('blockchain',since='2017-01-01', until='2017-12-18',lang = "en"
, n=1000) %>% twListToDF()
head(blch)
```

```
##
text
## 1                                                    RT @WePowerN: #Startupbootc
amp #Accelerator (SBC) chose us as one of 10 #energy companies for a prestigious #Aus
tralian energy track progra…
## 2 RT @AMLT_Token: Watch the video about the @Coinfirm_io AML/CTF Platform &amp;amp; @A
```

```
MLT_Token\n\xed\xa0\xbd\xed\xb3\xbahttps://t.co/t63X9pRInt\xed\xa0\xbd\xed\xb3\xba\n\
nAMLT by Coinfirm #TokenSal…
## 3                                                    RT @miniapps_pro: https://t
.co/8VBzempeAL founders Vitaly Gumirov and Dmitry Khan attended the biggest blockchai
n conference in Moscow http…
## 4                                                  RT @Azedolf1: IOTA on the Ethe
reum blockchain is here! \n\nAirdrop for only the first 5000 participants in the link
below: \nhttps://t.co/3M50…
## 5                                                                        I adde
d a video to a @YouTube playlist https://t.co/AyQbsIljy9 How to Mine 0.1 BTC in 40 mi
nutes - Blockchain Miner Pro
## 6                                        RT @WePowerN: \u26a1 #WePower Entering A
ustralia with the support of @sbcEnergyAus &amp;amp; local #energy companies!\u26a1 \nWe
are ready for great achieveme…
##   favorited favoriteCount replyToSN             created truncated
## 1     FALSE             0      <NA> 2017-12-17 23:59:57     FALSE
## 2     FALSE             0      <NA> 2017-12-17 23:59:55     FALSE
## 3     FALSE             0      <NA> 2017-12-17 23:59:54     FALSE
## 4     FALSE             0      <NA> 2017-12-17 23:59:54     FALSE
## 5     FALSE             0      <NA> 2017-12-17 23:59:54     FALSE
## 6     FALSE             0      <NA> 2017-12-17 23:59:53     FALSE
##   replyToSID                 id replyToUID
## 1       <NA> 942544931096350722       <NA>
## 2       <NA> 942544922044796928       <NA>
## 3       <NA> 942544918270144512       <NA>
## 4       <NA> 942544916953092096       <NA>
## 5       <NA> 942544914956607489       <NA>
## 6       <NA> 942544914482638848       <NA>
##                                                                        statusSo
urce
## 1 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android
</a>
## 2 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android
</a>
## 3                  <a href="http://twitter.com" rel="nofollow">Twitter Web Client
</a>
## 4   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone
</a>
## 5                          <a href="http://www.google.com/" rel="nofollow">Google
</a>
## 6 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android
</a>
##      screenName retweetCount isRetweet retweeted longitude latitude
## 1  Oleh78166119          407      TRUE     FALSE        NA       NA
## 2 Firmanassidiq          152      TRUE     FALSE        NA       NA
## 3       m1ntol6          228      TRUE     FALSE        NA       NA
## 4   gregore2112            2      TRUE     FALSE        NA       NA
```

```
## 5 rollotheclown             0     FALSE     FALSE        NA        NA
## 6  Oleh78166119           458      TRUE     FALSE        NA        NA
```

```
write.csv(blch, "blockchain.csv")
```

II.Word Frequency Blockchain – Sword of Damocles?

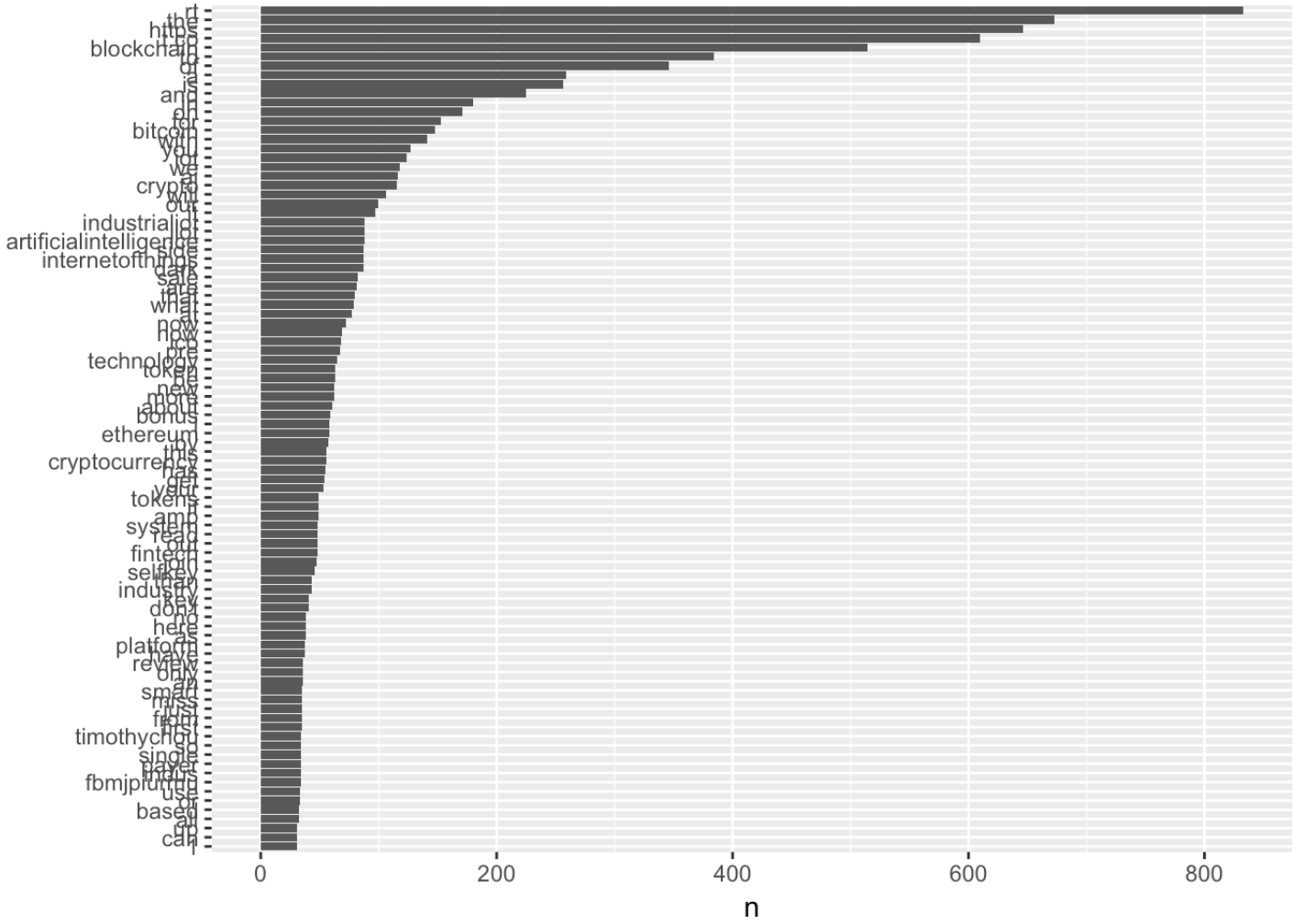
- 2.1 Text Analysis of Top Words (tokenize)

```
tidy_blch<- blch%>% unnest_tokens(word,text)

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

```
## # A tibble: 3,162 x 2
##          word     n
##         <chr> <int>
##  1         rt   833
##  2        the   673
##  3      https   646
##  4       t.co   610
##  5 blockchain   514
##  6         to   384
##  7         of   346
##  8          a   259
##  9         is   256
## 10        and   225
## # ... with 3,152 more rows
```

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



- 2.2 Sentiment Analysis sentiment analysis of positive and negtive words please refer to NRC Emotion Lexicon about the classification

```
get_sentiments("afinn")
```

```
## # A tibble: 2,476 x 2
##          word score
##         <chr> <int>
##  1    abandon    -2
##  2  abandoned    -2
##  3   abandons    -2
##  4   abducted    -2
##  5  abduction    -2
##  6 abductions    -2
##  7      abhor    -3
##  8   abhorred    -3
##  9  abhorrent    -3
## 10     abhors    -3
## # ... with 2,466 more rows
```

```
get_sentiments("bing")
```

```
## # A tibble: 6,788 x 2
##           word sentiment
##          <chr>     <chr>
##  1     2-faced  negative
##  2     2-faces  negative
##  3          a+  positive
##  4    abnormal  negative
##  5     abolish  negative
##  6  abominable  negative
##  7  abominably  negative
##  8   abominate  negative
##  9 abomination  negative
## 10       abort  negative
## # ... with 6,778 more rows
```

```
get_sentiments("nrc")
```

```
## # A tibble: 13,901 x 2
##           word sentiment
##          <chr>     <chr>
##  1      abacus     trust
##  2     abandon      fear
##  3     abandon  negative
##  4     abandon   sadness
##  5   abandoned     anger
##  6   abandoned      fear
##  7   abandoned  negative
##  8   abandoned   sadness
##  9 abandonment     anger
## 10 abandonment      fear
## # ... with 13,891 more rows
```

```
nrcpos <- get_sentiments("nrc") %>%
  filter(sentiment == "positive")

tidy_blch%>%
  inner_join(nrcpos) %>%
  count(word, sort = TRUE)
```

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

```
## # A tibble: 203 x 2
##           word     n
##          <chr> <int>
##  1  technology    65
##  2       bonus    59
##  3        join    47
##  4 improvement    28
##  5       worth    25
##  6   community    24
##  7      create    23
##  8       learn    23
##  9    solution    16
## 10        real    15
## # ... with 193 more rows
```

```
nrcneg <- get_sentiments("nrc") %>%
  filter(sentiment == "negative")

tidy_blch%>%
  inner_join(nrcneg) %>%
  count(word, sort = TRUE)
```

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

```
## # A tibble: 55 x 2
##          word     n
##         <chr> <int>
##  1 revolution    13
##  2      crypt    11
##  3      quote    10
##  4    payment     9
##  5      verge     6
##  6      fraud     5
##  7    problem     5
##  8 regulatory     5
##  9  anonymous     3
## 10     crisis     3
## # ... with 45 more rows
```

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

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

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

```
## # A tibble: 33 x 2
##          word     n
##         <chr> <int>
##  1      watch    19
##  2 revolution    13
##  3   powerful    12
##  4      crypt    11
##  5     change     6
##  6      verge     6
##  7    auditor     5
##  8    problem     5
##  9 regulatory     5
## 10      giant     4
## # ... with 23 more rows
```

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

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

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

```
## # A tibble: 65 x 2
##           word     n
##          <chr> <int>
##  1       bonus    59
##  2 improvement    28
##  3      create    23
##  4       white    14
##  5       enjoy    12
##  6        good    12
##  7    powerful    12
##  8     success    11
##  9        glad     9
## 10       music     9
## # ... with 55 more rows
```

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

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

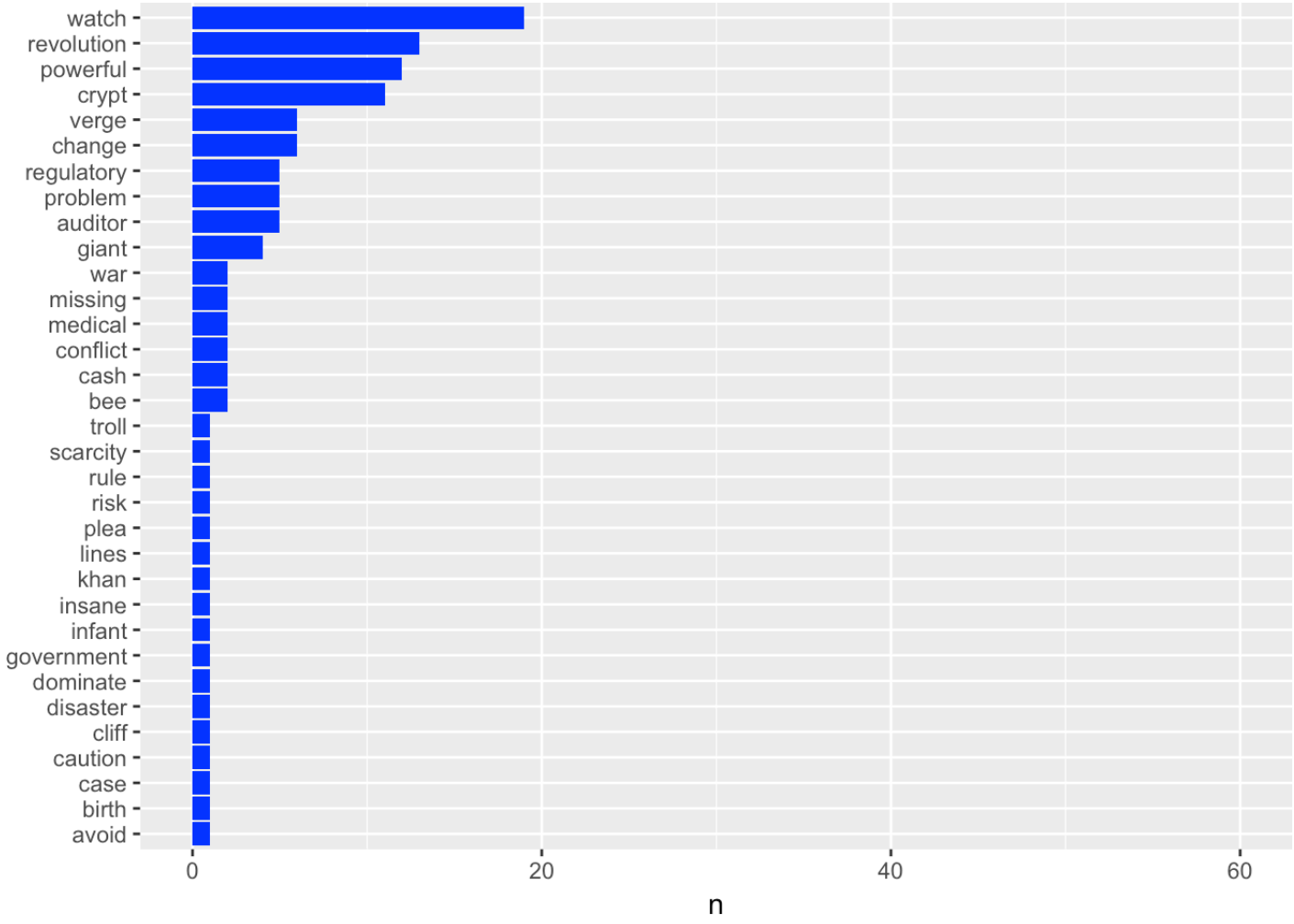
```
## # A tibble: 33 x 2
##          word     n
##         <chr> <int>
##  1      watch    19
##  2 revolution    13
##  3   powerful    12
##  4      crypt    11
##  5     change     6
##  6      verge     6
##  7    auditor     5
##  8    problem     5
##  9 regulatory     5
## 10      giant     4
## # ... with 23 more rows
```

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

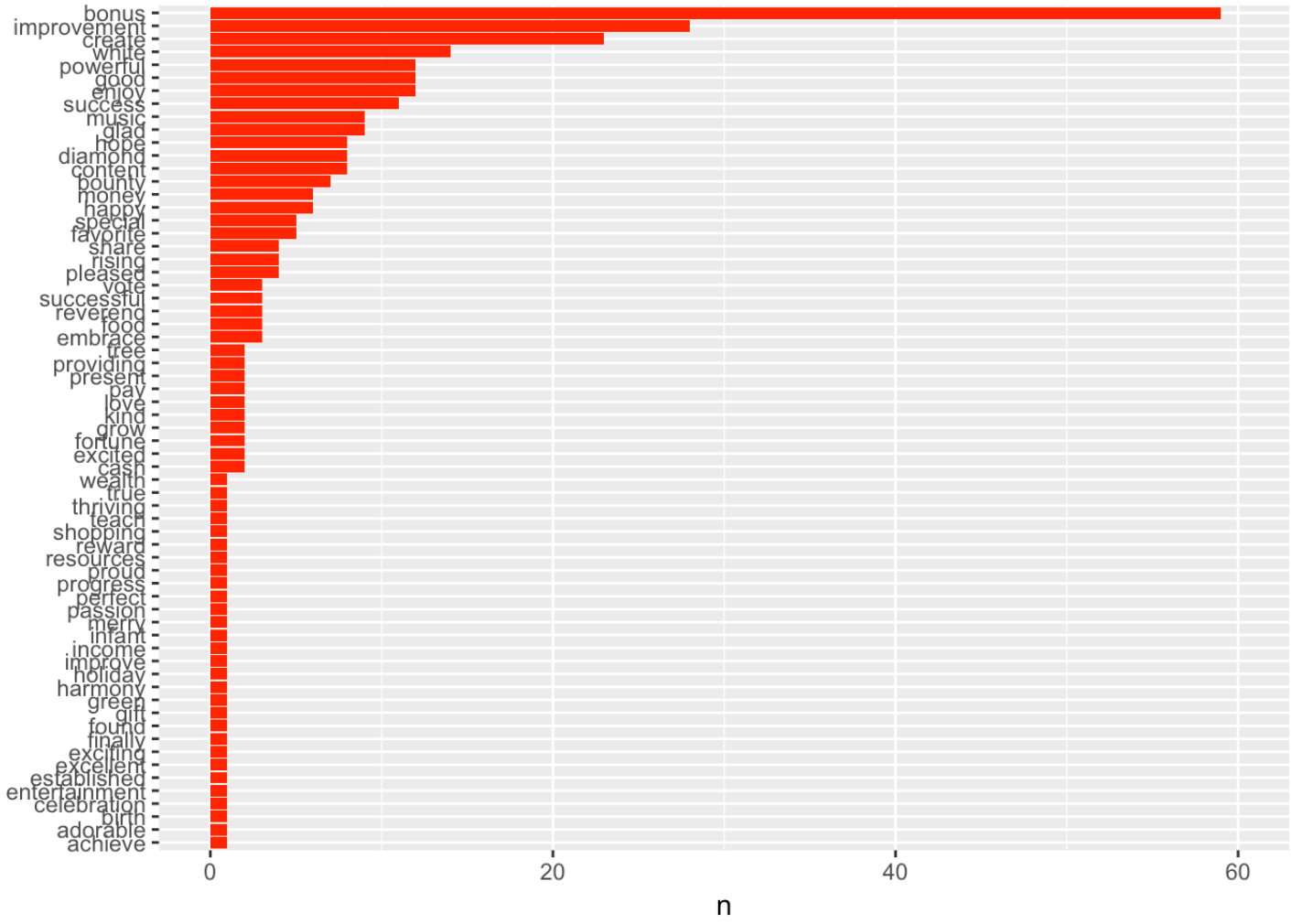
```
## Joining, by = "word"
```

```
## # A tibble: 65 x 2
##           word     n
##          <chr> <int>
##  1       bonus    59
##  2 improvement    28
##  3      create    23
##  4       white    14
##  5       enjoy    12
##  6        good    12
##  7    powerful    12
##  8     success    11
##  9        glad     9
## 10       music     9
## # ... with 55 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(0,60))
```




```
joylist %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col(fill="red") +
  xlab(NULL) +
  coord_flip() +
  scale_y_continuous(limits = c(0,60))
```



- 2.3 Summary

- From the Twitters, the word bonus are mentioned over 60 time, indicating it is an important sentiment of people towards blockchain.

- The word improvement appeared over 30 times. This is also an important sentiment here. Some words like disaster, and caution may indicate something about people's worry.

- 2.4 Sentiment analysis with bing

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

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

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

```
## # A tibble: 135 x 2
##           word     n
##          <chr> <int>
##  1       bonus    59
##  2       smart    35
##  3 improvement    28
##  4       great    25
##  5       worth    25
##  6       thank    16
##  7 interesting    13
##  8        like    13
##  9       enjoy    12
## 10        good    12
## # ... with 125 more rows
```

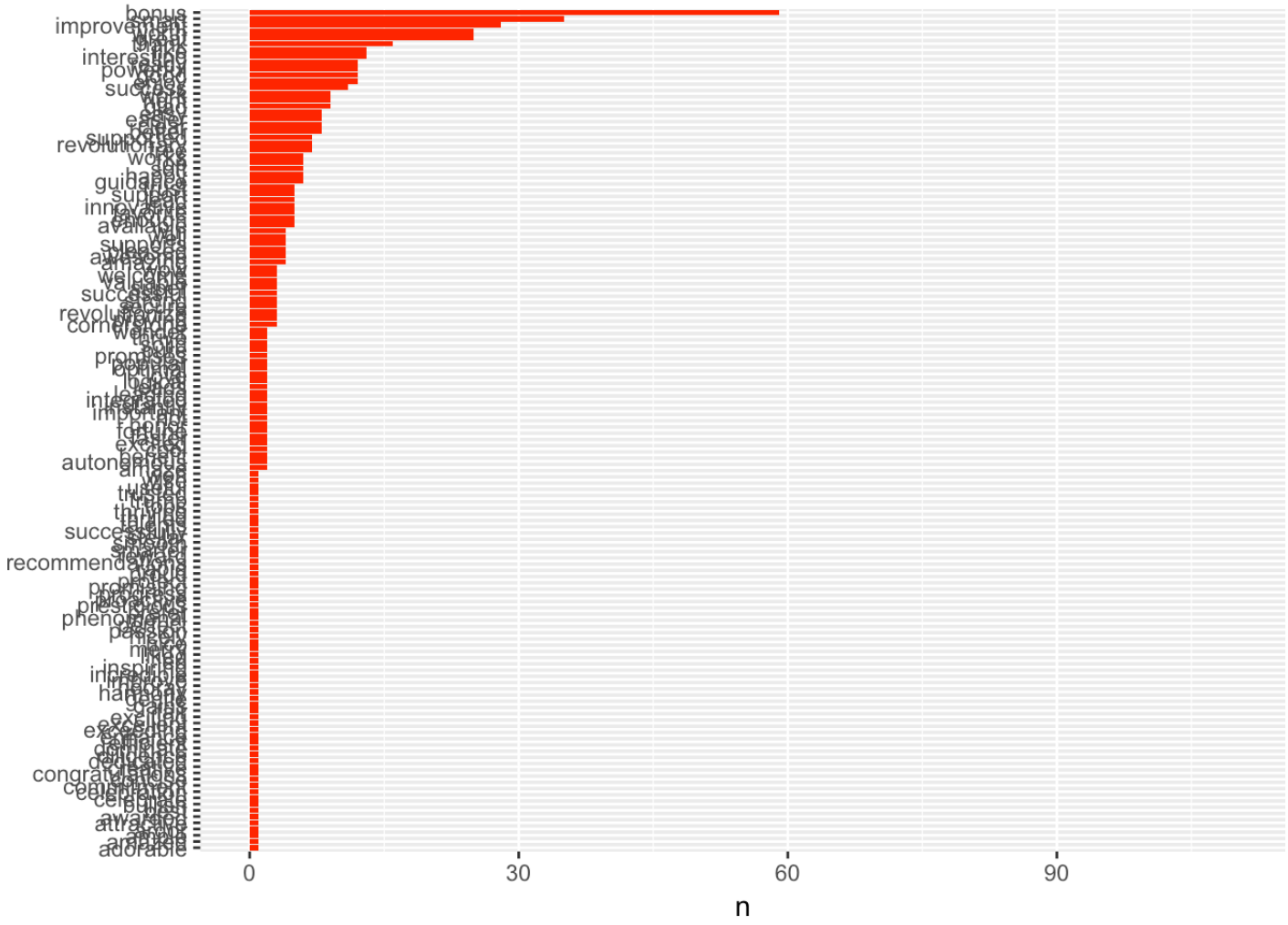
```
bingneg <- get_sentiments("bing") %>%
  filter(sentiment == "negative")

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

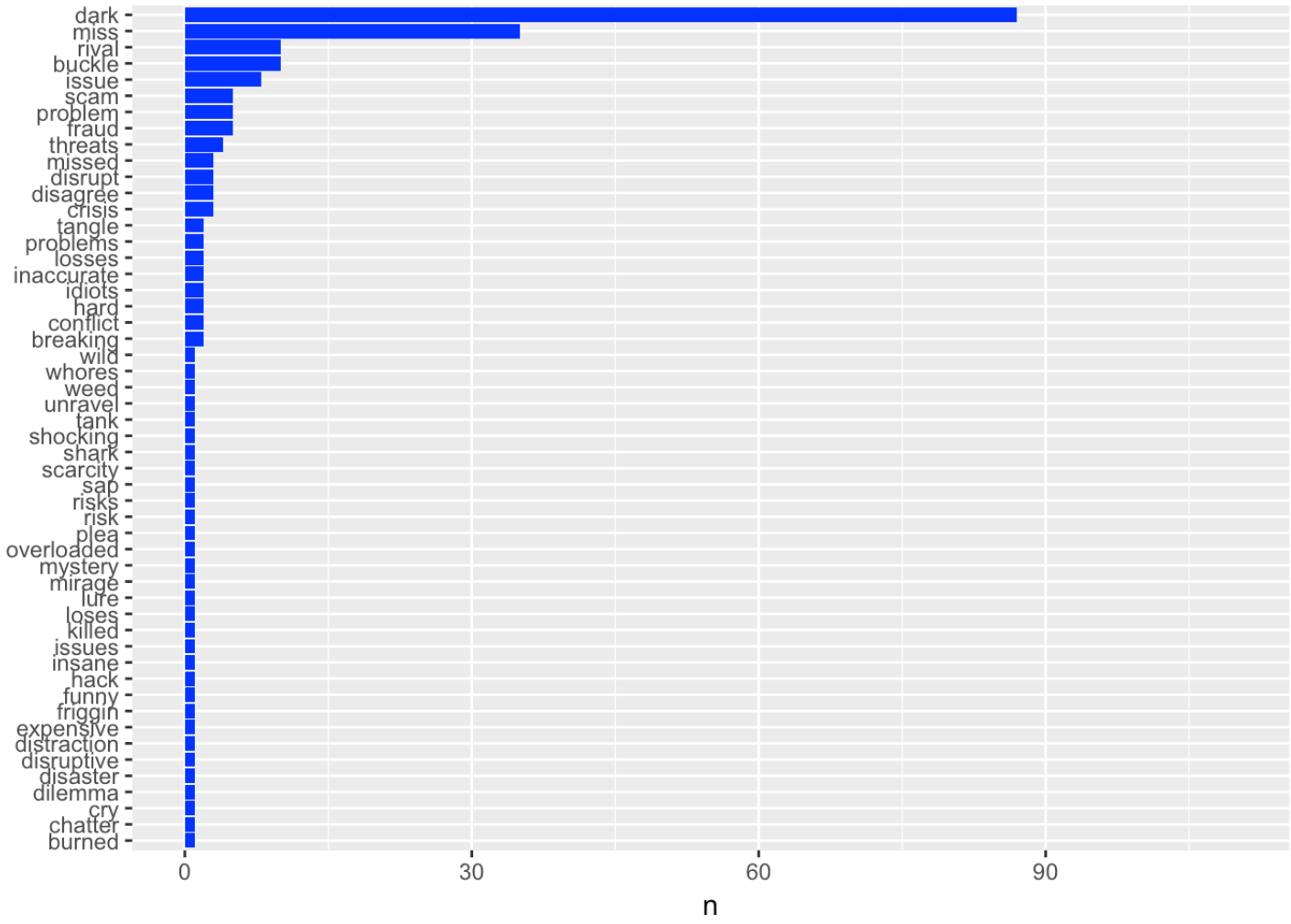
```
## Joining, by = "word"
```

```
## # A tibble: 52 x 2
##       word     n
##      <chr> <int>
##  1    dark    87
##  2    miss    35
##  3  buckle    10
##  4   rival    10
##  5   issue     8
##  6   fraud     5
##  7 problem     5
##  8    scam     5
##  9 threats     4
## 10  crisis     3
## # ... with 42 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))
```



- 2.5. Summary

- From the Twitters, the word dark are mentioned is 61 times, indicating it is an impressive negative sentiment of people towards blockchain.

- The word bonus appeared over 60 times. This is also an important sentiment here.

- Words like glad, and easier" may indicate something about people’s expectaiton of blockchain.

- Therefore, it seems that people holds a half-and-half attitudes towards blockchain.

- 2.6 Wordcloud

```
library(wordcloud)


tidy_blch%>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100))
```

```
tidy_blch %>%
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"
```

negative

positive

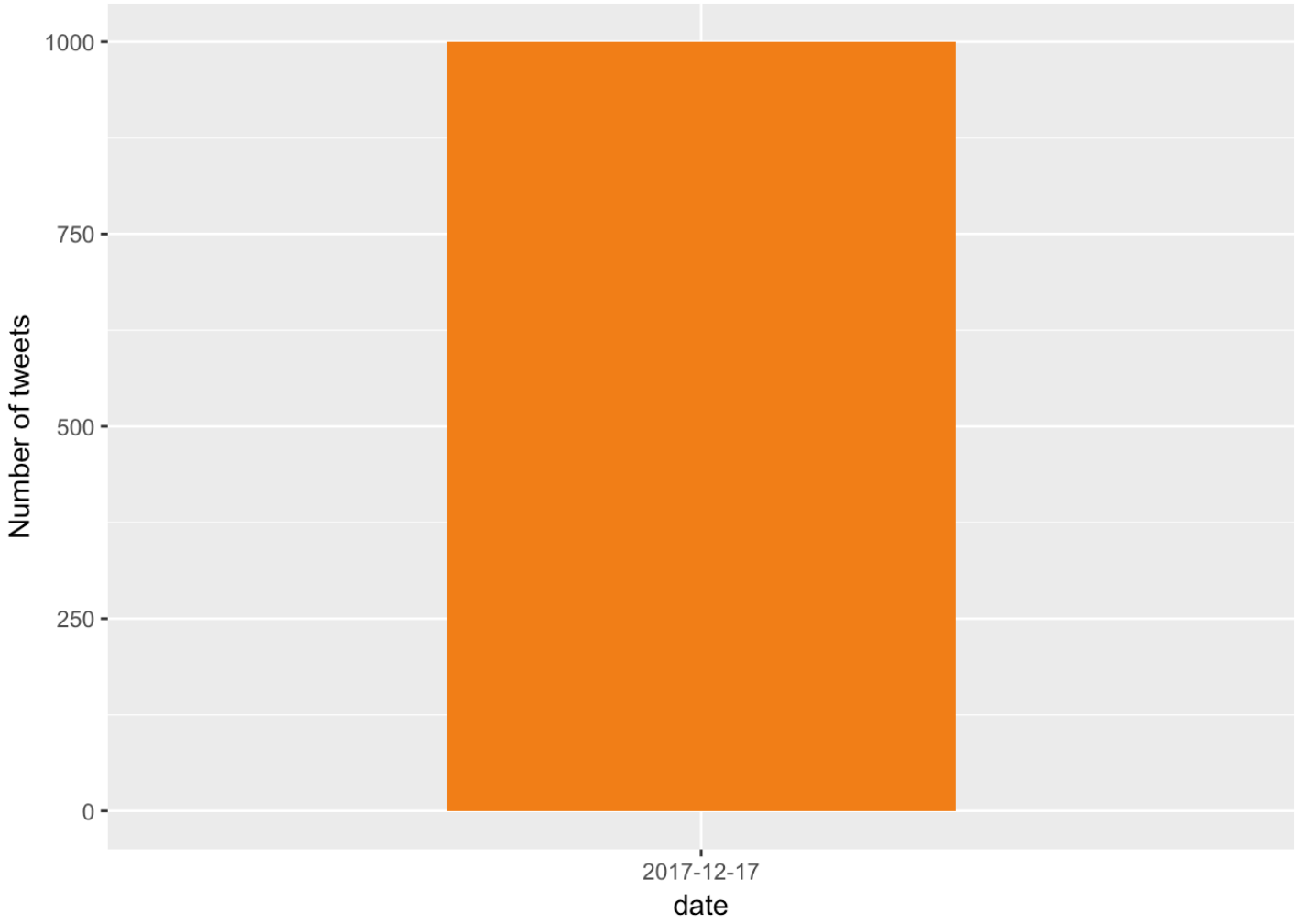
- The cloud figure supports my former analysis that people holds a half-and-half attitudes towards blockchain.

IV. A “summit” day– Dec 17 2017

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

```
## # A tibble: 1 x 2
##         date date_n
##       <date>  <int>
## 1 2017-12-17   1000
```

```
ggplot() +
  geom_bar(data=dates,
           aes(date,date_n),
           stat = "identity",
           fill="#f1801d") +
  ylab("Number of tweets") +
  scale_x_discrete(limits=lubridate::date(dates$date),labels=lubridate::date(dates$da
te))
```



- The number of tweets about blockchain in December 17 are higher than other dates.

```
# First 10 rows The tweets on dec 17

blch%>%
  filter(date(created)==c("2017-12-17")) %>%
  .$text %>%
  head(10)
```

```
##  [1] "RT @WePowerN: #Startupbootcamp #Accelerator (SBC) chose us as one of 10 #ene
rgy companies for a prestigious #Australian energy track progra…"
##  [2] "RT @AMLT_Token: Watch the video about the @Coinfirm_io AML/CTF Platform &amp
; @AMLT_Token\n\xed\xa0\xbd\xed\xb3\xbahttps://t.co/t63X9pRInt\xed\xa0\xbd\xed\xb3\xb
a\n\nAMLT by Coinfirm #TokenSal…"
##  [3] "RT @miniapps_pro: https://t.co/8VBzempeAL founders Vitaly Gumirov and Dmitry
Khan attended the biggest blockchain conference in Moscow http…"
##  [4] "RT @Azedolf1: IOTA on the Ethereum blockchain is here! \n\nAirdrop for only
the first 5000 participants in the link below: \nhttps://t.co/3M50…"
##  [5] "I added a video to a @YouTube playlist https://t.co/AyQbsIljy9 How to Mine 0
.1 BTC in 40 minutes - Blockchain Miner Pro"
##  [6] "RT @WePowerN: \u26a1 #WePower Entering Australia with the support of @sbcEne
rgyAus &amp;amp; local #energy companies!\u26a1 \nWe are ready for great achieveme…"
##  [7] "RT @CashaaLtd: In the second session with @kgauravitc at @BlockchainConio #I
ndia 2017 tomorrow Monday, learn all you need to know about #bl…"
##  [8] "RT @miniapps_pro: You can still purchase #MAT tokens to exchange them for a
#KRK coin with 50% #bonus.\nWe are developing Next Generation #B…"
##  [9] "RT @codemojoio: Any brand or publisher related configuration will be stored
as Smart Contracts in #blockchain https://t.co/bQItpWH9Vu #ALLO…"
## [10] "RT @JacBurns_Comext: Bitcoin-Based Ethereum Smart Contract &amp;amp; Sidechain R
ival RSK https://t.co/82Qj9kgmlF\n\n#Blockchain #Fintech #IoT #BigDat…"
```

```
blch%>%
  filter(date(created)==c("2017-12-17")) %>%
  nrow()
```

```
## [1] 1000
```

```
# number of tweets containing blockchain on Dec 17
blch%>%
  filter(date(created)==c("2017-12-17")) %>%
  filter(grepl("blockchain",text)) %>%
  nrow
```

```
## [1] 313
```

```
# number of tweets containing develop on Dec 17
blch%>%
  filter(date(created)==c("2017-12-17")) %>%
  filter(grepl("improve",text)) %>%
  nrow
```

```
## [1] 29
```

V. Where are the blockchain Twitterers?

```
load("my_oauth.Rdata")

filterStream("blockchainmap.json",
             track=c("blockchain"),
             locations = c(-125, 25, -66,50),
             timeout=200, oauth=my_oauth)
```

```
## Capturing tweets...
```

```
## Connection to Twitter stream was closed after 200 seconds with up to 9000 tweets d
ownloaded.
```

```
blockchainmap<-parseTweets("blockchainmap.json", verbose = TRUE)
```

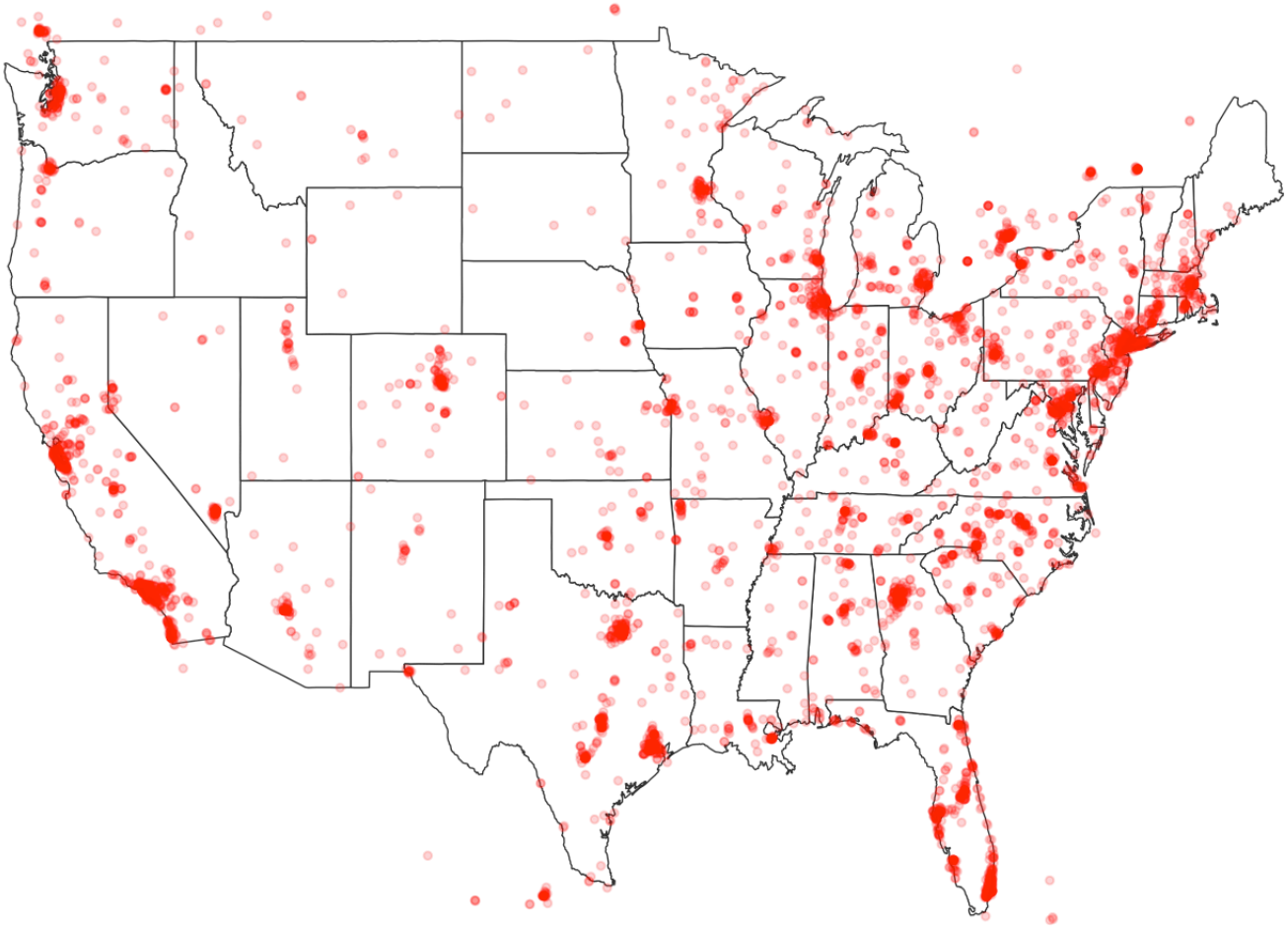
```
## 38672 tweets have been parsed.
```

```
ck1 <- sum(blockchainmap$lat>0, na.rm = TRUE)
ck2 <- sum(blockchainmap$place_lat>0, na.rm = TRUE)
ck3 <- sum(!is.na(blockchainmap$location))
map.data <- map_data("state")
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
netpoints <- data.frame(x = as.numeric(blockchainmap$lon),
                        y = as.numeric(blockchainmap$lat))
netpoints <- netpoints[netpoints$y > 25, ]
netpoints<-filter(netpoints,y>19&y<65,x>(-161.7)&x<(-68.01))
ggplot(map.data) +
  geom_map(aes(map_id = region),
           map = map.data,
           fill = "white",
           color = "grey20", size = 0.25) +
  expand_limits(x = map.data$long, y = map.data$lat) +
  theme(axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        axis.title = element_blank(),
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        plot.background = element_blank(),
        plot.margin = unit(0 * c( -1.5, -1.5, -1.5, -1.5), "lines")) +
        geom_point(data = netpoints,
        aes(x = x, y = y), size = 1,
        alpha = 1/5, color = "red")
```



- The map displays that east part of U.S., such as New York; and the west part like California, have the most people twittering the blockchain topic

General Findings

The sentiment analyses presents that some people regards blockchain as a improvement, while others regard it risky. In the negative chart, the word dark is mentioned is 35 times In the positive chart, the positive word bonus is also mentioned around 35 times.
The cloud text visualized this half-half attitude balance as well Therefore, we could assume that people holds a half-and-half attitudes towards blockchain
The number of tweets about blockchain on December 17 reached the highest amount than other dates. The map displays that east part of U.S., such as New York; and the west part like California, have the most people twittering the blockchain topic

Links

SHINY APP https://shshan.shinyapps.io/blockchain_on_twitter/ (https://shshan.shinyapps.io/blockchain_on_twitter/)

GIT HUB https://github.com/shan-bu-2017/Twitterers-and-Blockchain.git (https://github.com/shan-bu-2017/Twitterers-and-Blockchain.git)