Twitter zzh-gj-jz association

2022-05-25

In this analysis I will first extract tweets mentioning "#zhangzhehan", "#gongjun" or "#junzhe" and then check the overlap between accounts tweeting about these topics. After playing around a bit I decided to use only hashtags instead of full names to focus on dedicated tweets instead of casual mentions.

First let's get the tweets via Twitter API. Please note that there is a rate limit of up to 18000 tweets per day. Nonetheless, I never manage to get the maximum of 6000 tweets per query for some reason. Tweet search is also limited to the last 4 days.

I will then plot the accounts with the most tweets about these topics.

```
###extract tweets

tweets_extract_plot<-function(words,title){
  tweets<-search_tweets2(words,n = 6000,lang = "en",include_rts = FALSE)
  n_tweets<-nrow(tweets)
  stats<-paste(title,collapse=" ")
  stats<-paste(stats,", n=",n_tweets,"tweets")

  plot<-table(tweets$screen_name)%>%
    as.data.frame() %>%
    arrange(desc(Freq)) %>%
    top_n(15) %>%
    mutate(Var1 = reorder(Var1, Freq)) %>%
    ggplot(aes(x = Var1, y = Freq)) +
    geom_col() +
    xlab(NULL) +
    coord_flip() +
    ggtitle(paste("top accounts tweeting about",stats))
  print(plot)
  return(tweets)
}
zzh_tweets<-tweets_extract_plot(c("#zhangzhehan"),"#zhangzhehan")
```
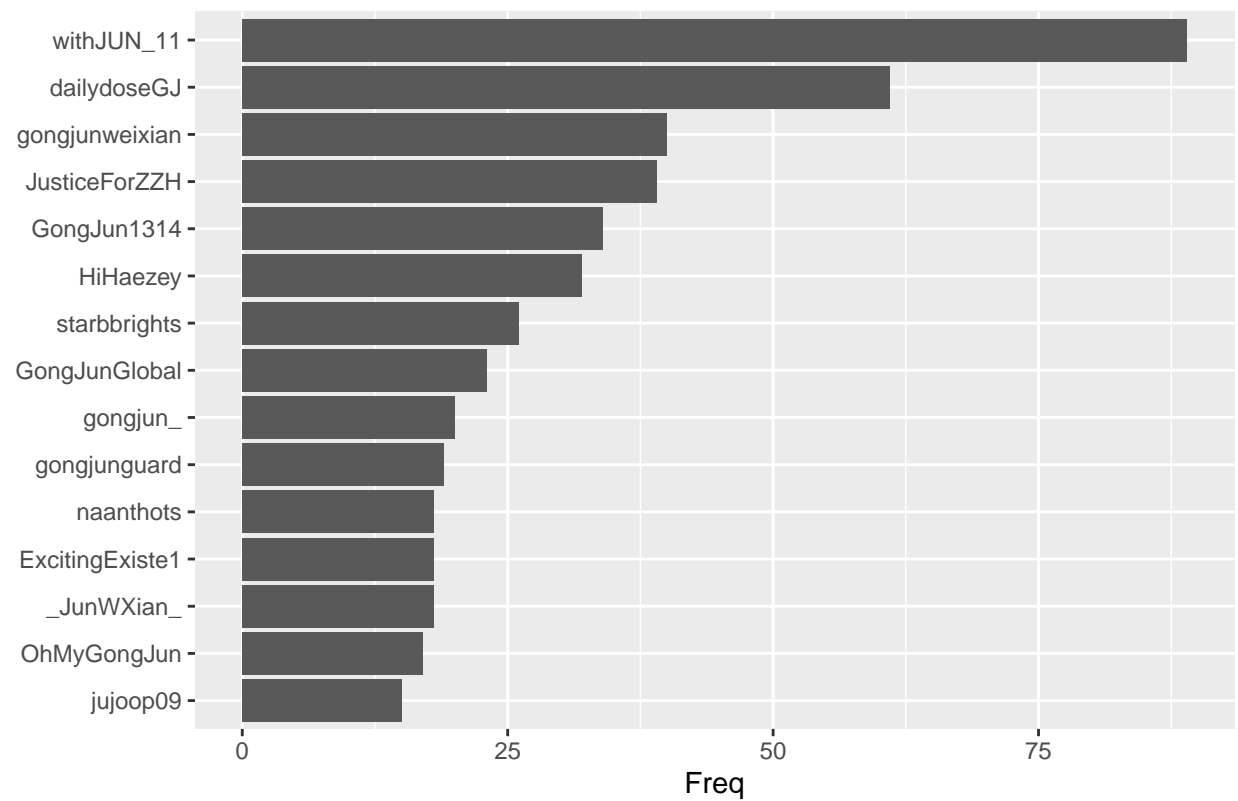
```
## Selecting by Freq
```

top accounts tweeting about #zhangzhehan , n= 1321 tweets
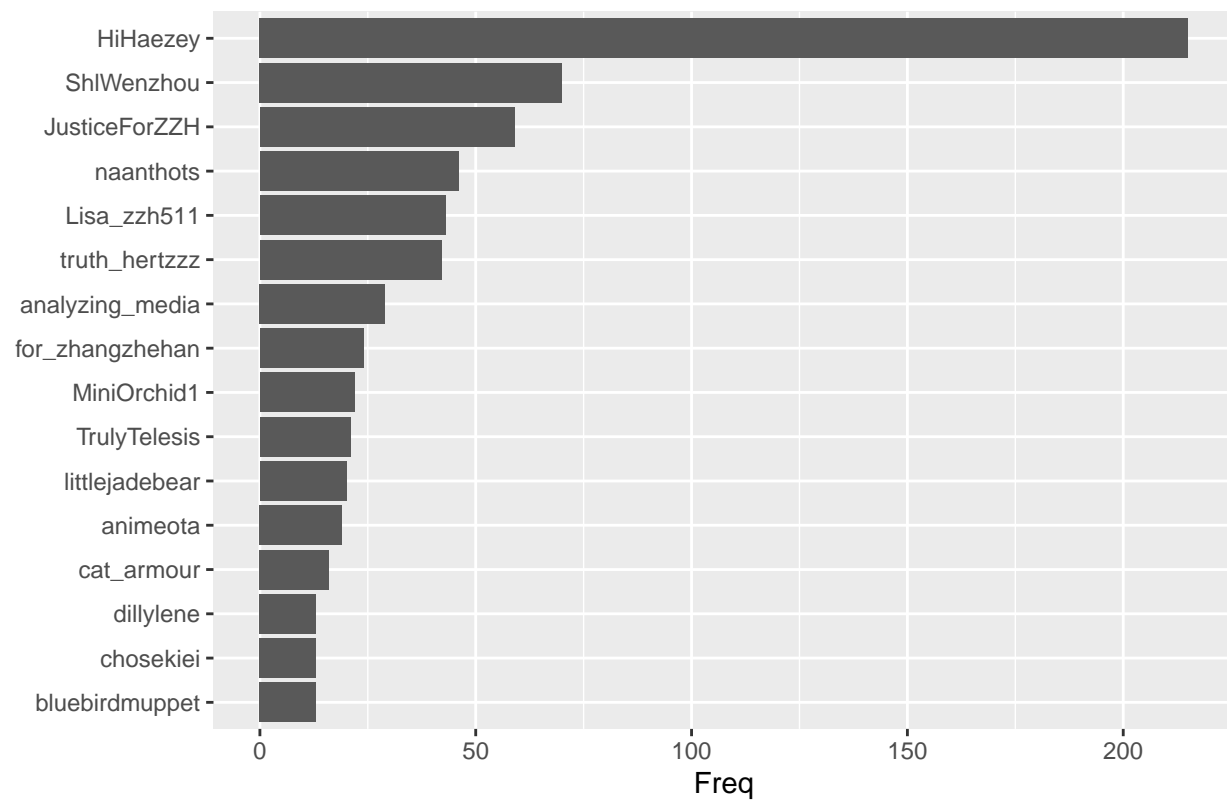

```
gj_tweets<-tweets_extract_plot(c("#gongjun"),"#gongjun")
```
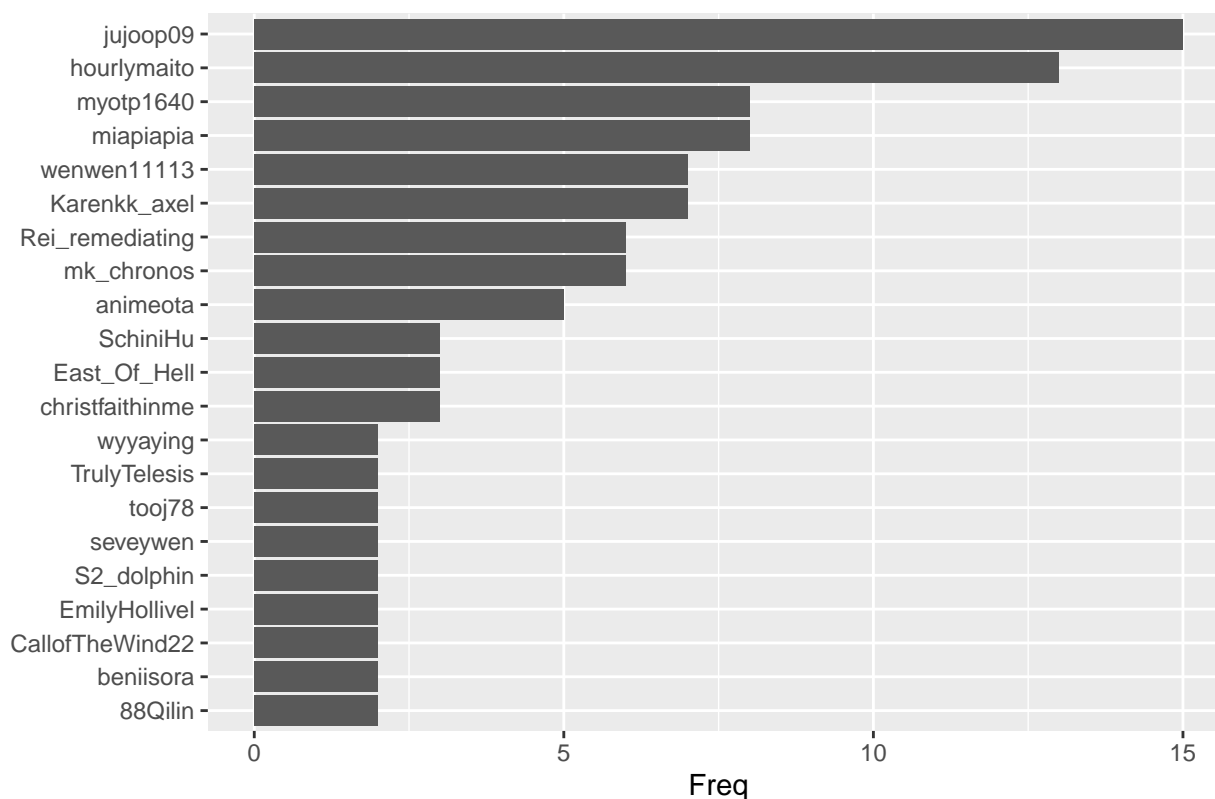
```
## Selecting by Freq
```

top accounts tweeting about #gongjun , n= 1014 tweets


```
jz_tweets<-tweets_extract_plot(c("#junzhe"),"#junzhe")
```

```
## Selecting by Freq
```

top accounts tweeting about #junzhe , n= 162 tweets

Next let's make a Venn diagram to see the overlap between accounts tweeting about "Junzhe" "Zhang Zhehan" or "Gong Jun". I have to use 2 Venn diagram packages because gplots gives me the numbers+intersections but only venneuler lets me make a beautiful diagram with proportional circle areas.

Note: This plot is not based on tweets containing the respective topics in combination. Instead it plots, e.g. how often accounts mentioning "ZZH" have also mentioned "Gong Jun", this can be in the same or different tweets.

```
###Venn diagram
library(gplots)
```

```
##
## Attaching package: 'gplots'
```

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

```
x <- list("ZZH" = zzh_tweets$screen_name,
          "JunZhe" = jz_tweets$screen_name,
          "GJ" = gj_tweets$screen_name)
v.table <- venn(x)
```
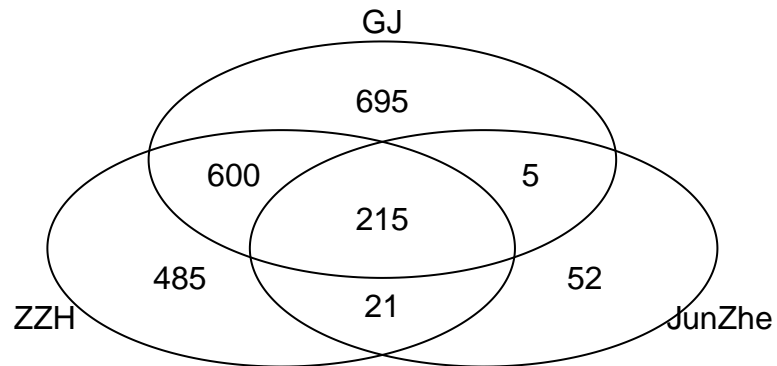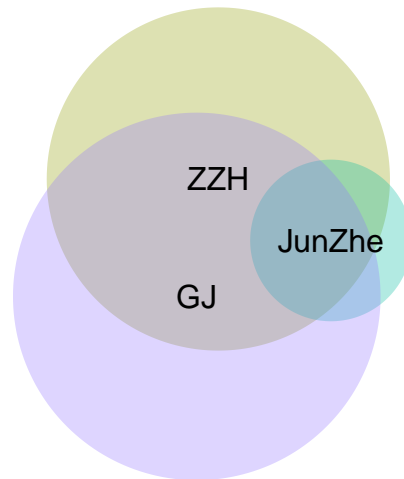
```
x <- attr(v.table,"intersections")
names_venn<-sapply(seq_along(x), function(i) paste(names(x)[[i]]))
numbers_venn<-lapply(seq_along(x), function(i) length(x[[i]]))
names_venn<-str_replace(names_venn, ":", "&")
names_venn<-str_replace(names_venn, ":", "&")
numbers_venn<-as.numeric(numbers_venn)
names(numbers_venn)<-names_venn


library(venneuler)
```

```
## Loading required package: rJava
```

```
v <- venneuler(numbers_venn)
plot(v)
```

Finally let's plot the top accounts tweeting about the respective topics+combinations. In these plots, for example the "Zhang Zhehan" accounts are the "pure" accounts not mentioning the other two topics.

```r
###top accounts n venn diagram slices
topgraph<-function(index,data){
  n_tweets<-length(data)
  table(data)%>%
    as.data.frame() %>%
    arrange(desc(Freq)) %>%
    top_n(15) %>%
    mutate(data = reorder(data, Freq)) %>%
    ggplot(aes(x = data, y = Freq)) +
    geom_col() +
    xlab(NULL) +
    coord_flip() +
    ggtitle(paste("top accounts tweeting about",index,", n=",n_tweets,"tweets"))
}


x <- attr(v.table,"intersections")
lapply(seq_along(x), function(i) topgraph(names(x)[[i]], x[[i]]))
```
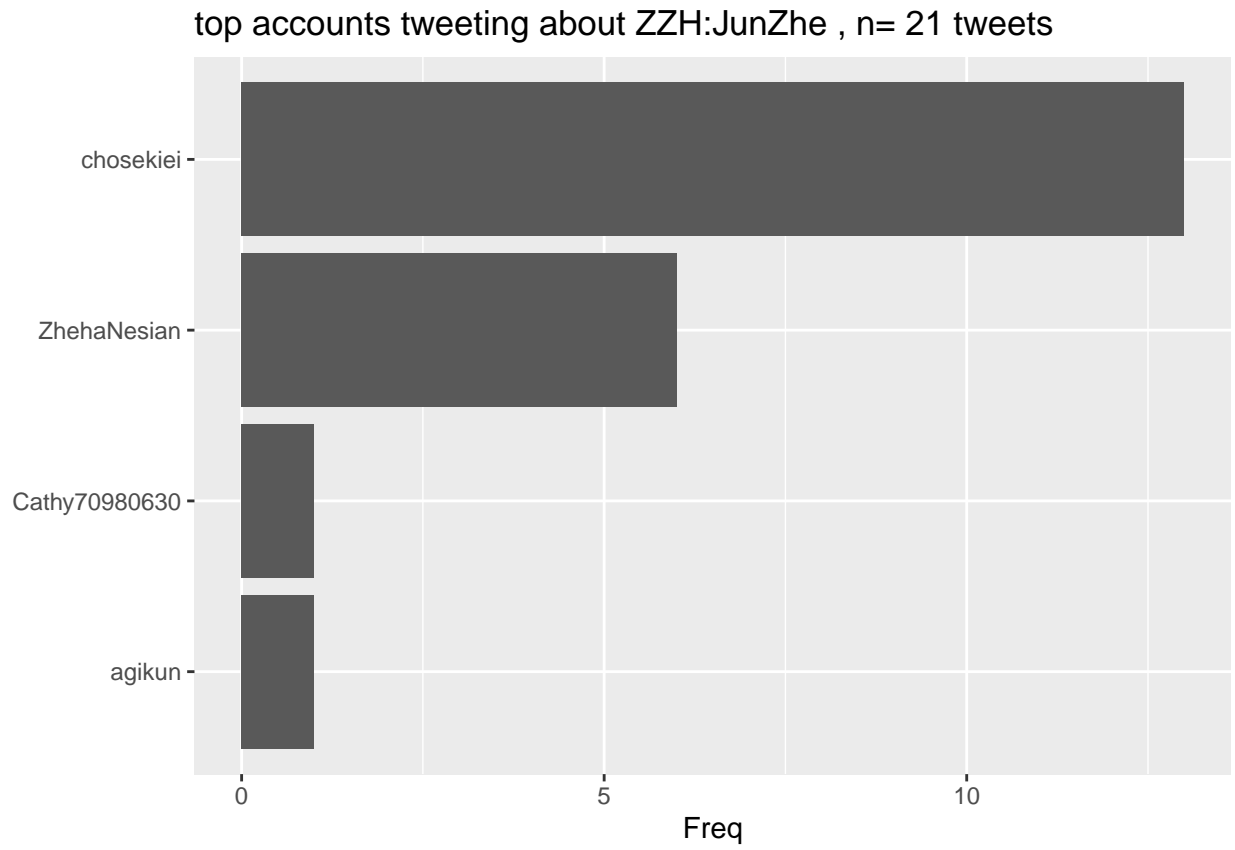
```
## Selecting by Freq
## Selecting by Freq
## Selecting by Freq
## Selecting by Freq
## Selecting by Freq
```
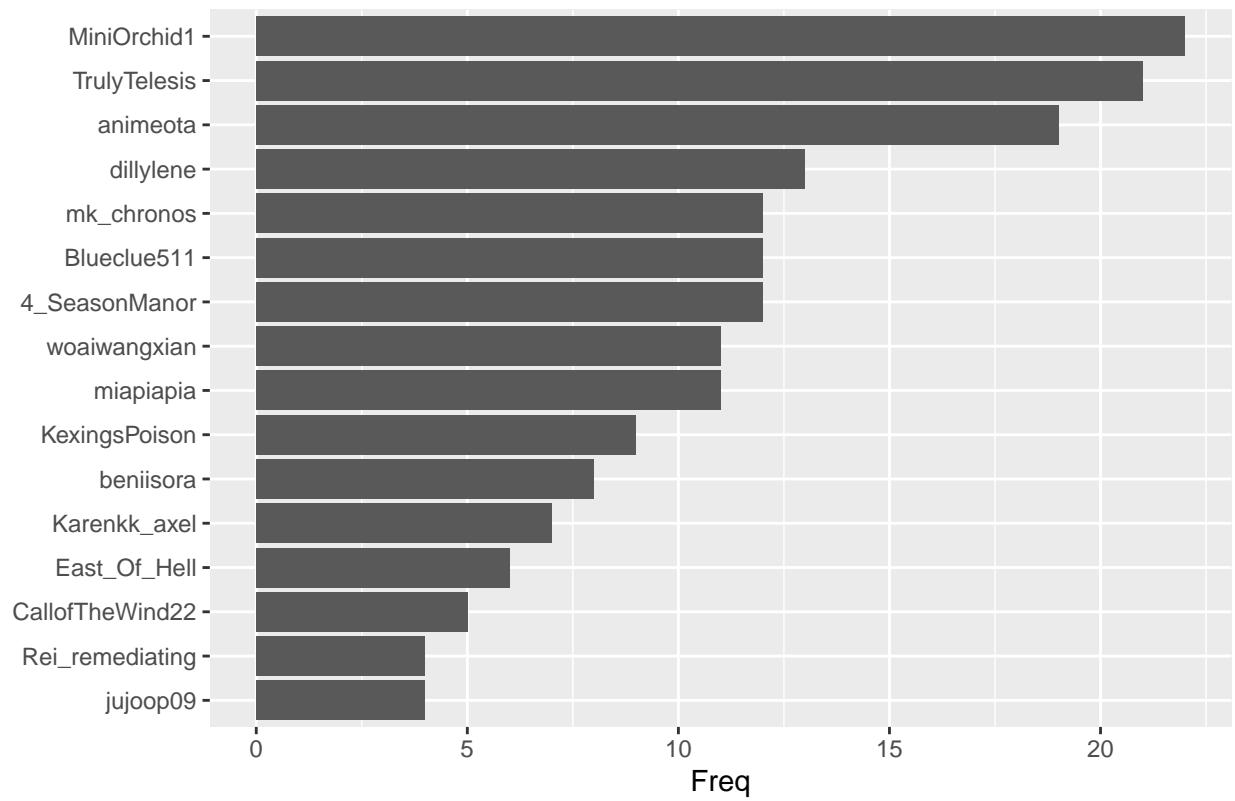
```
## Selecting by Freq
## Selecting by Freq
```
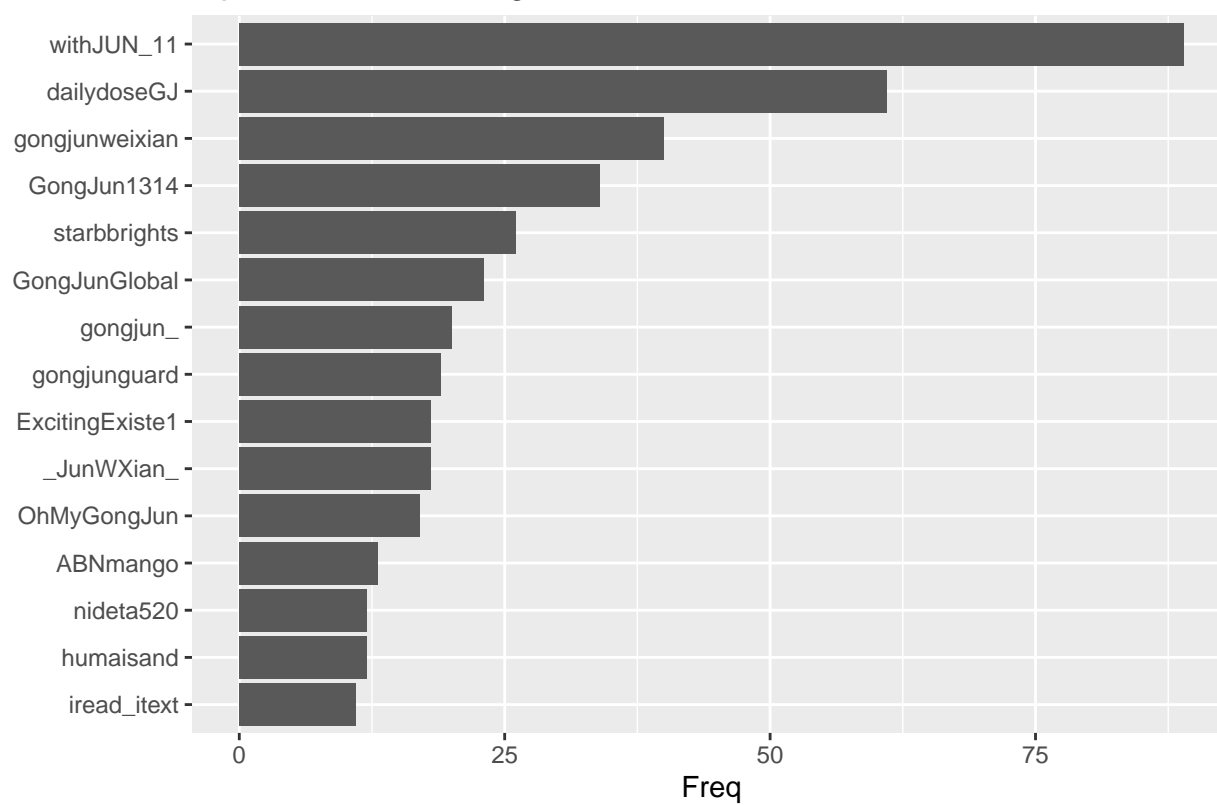
```
## [[1]]
```

top accounts tweeting about ZZH:JunZhe , n= 21 tweets


```
##
## [[2]]
```

top accounts tweeting about ZZH:JunZhe:GJ , n= 215 tweets
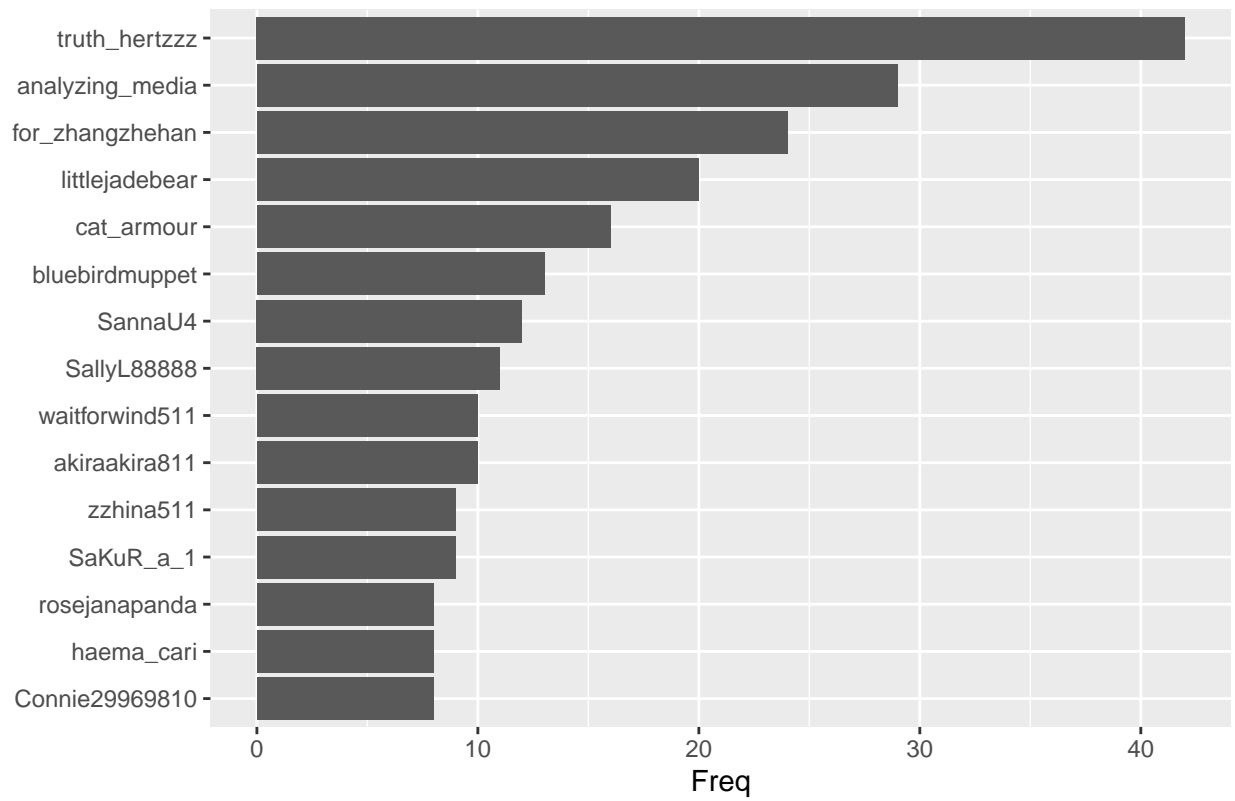
```
## 
## [[3]]
```
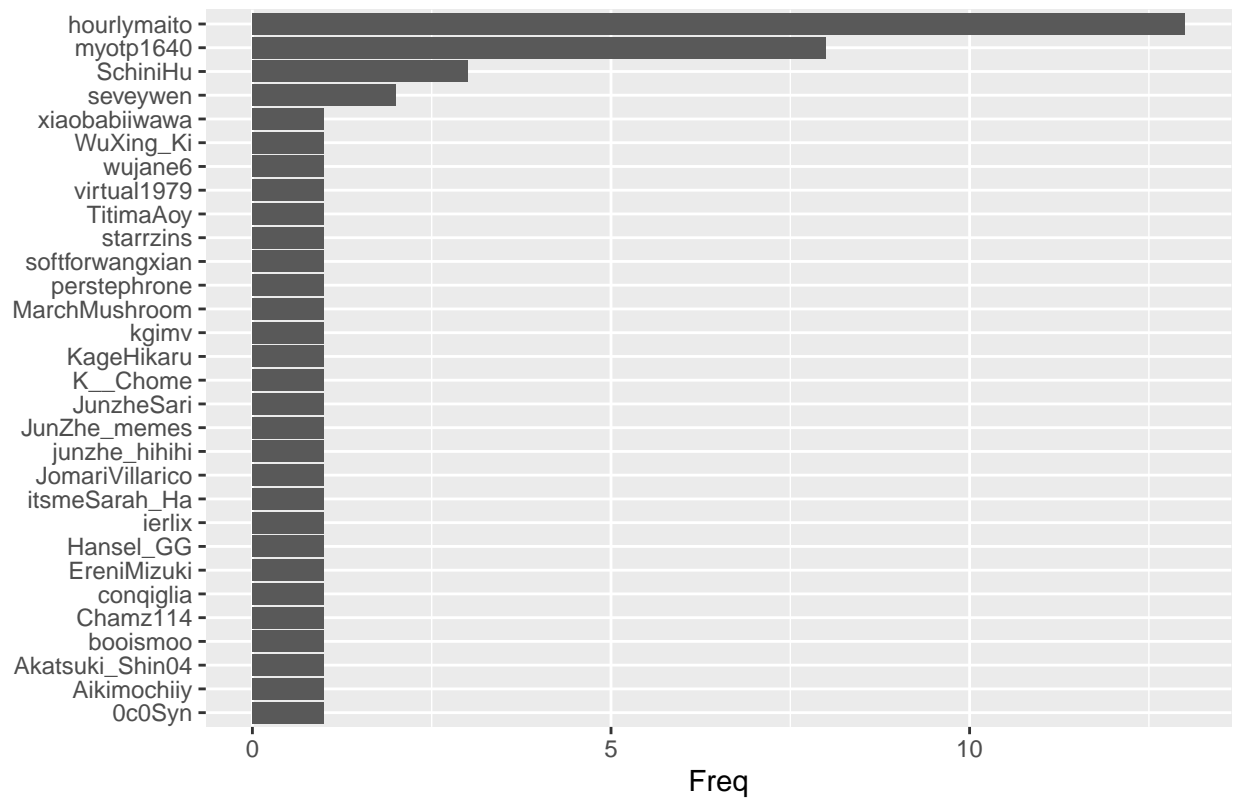
top accounts tweeting about GJ , n= 695 tweets


```
##
## [[4]]
```

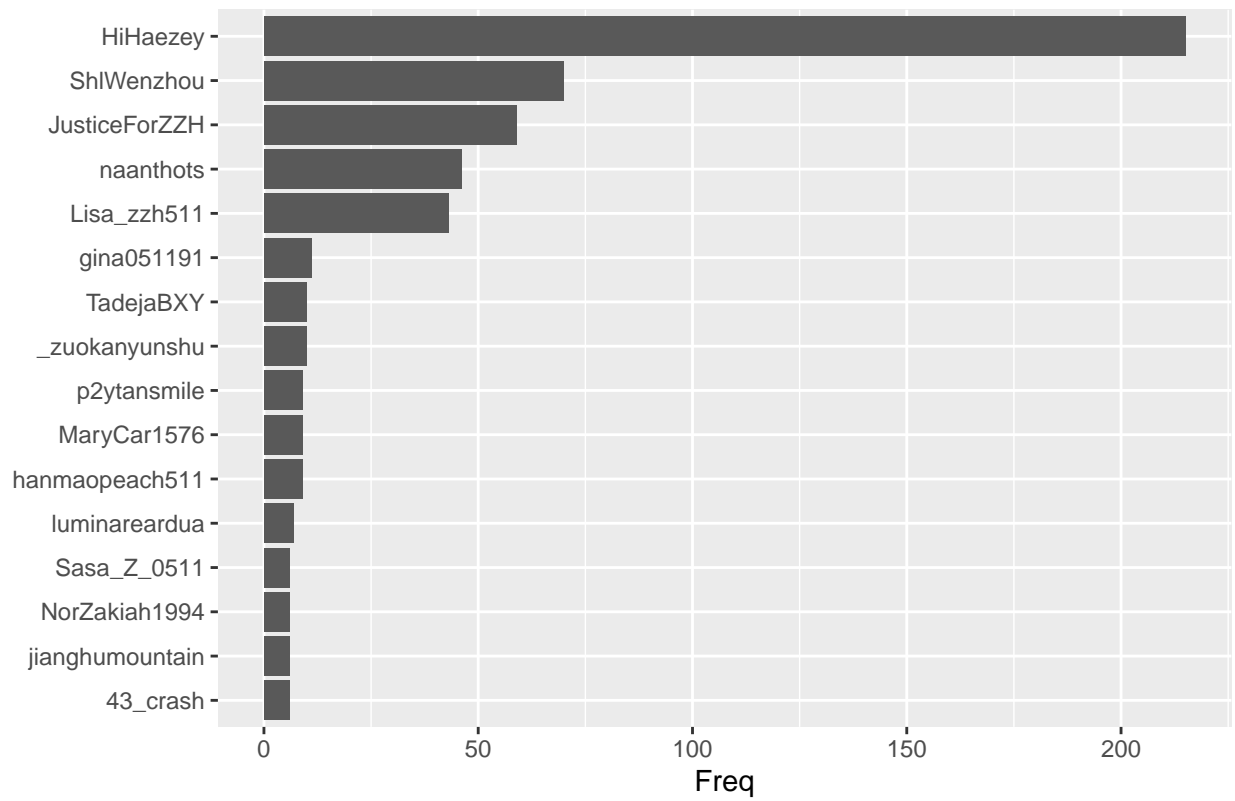## top accounts tweeting about ZZH , n= 485 tweets


```
##
## [[5]]
```

top accounts tweeting about JunZhe , n= 52 tweets

```
## 
## [[6]]
```

top accounts tweeting about ZZH:GJ , n= 600 tweets

```
##
## [[7]]
```

top accounts tweeting about JunZhe:GJ , n= 5 tweets