

Twitter zzh-gj-jz association

04 June, 2022

In this analysis I will first extract tweets mentioning “#zhangzhehan”, “#gongjun” or “#junzhe” and then check the overlap between accounts tweeting about these topics.

Edit 30/05/22: Added “zhang zhehan” and “gong jun” back to search keywords.

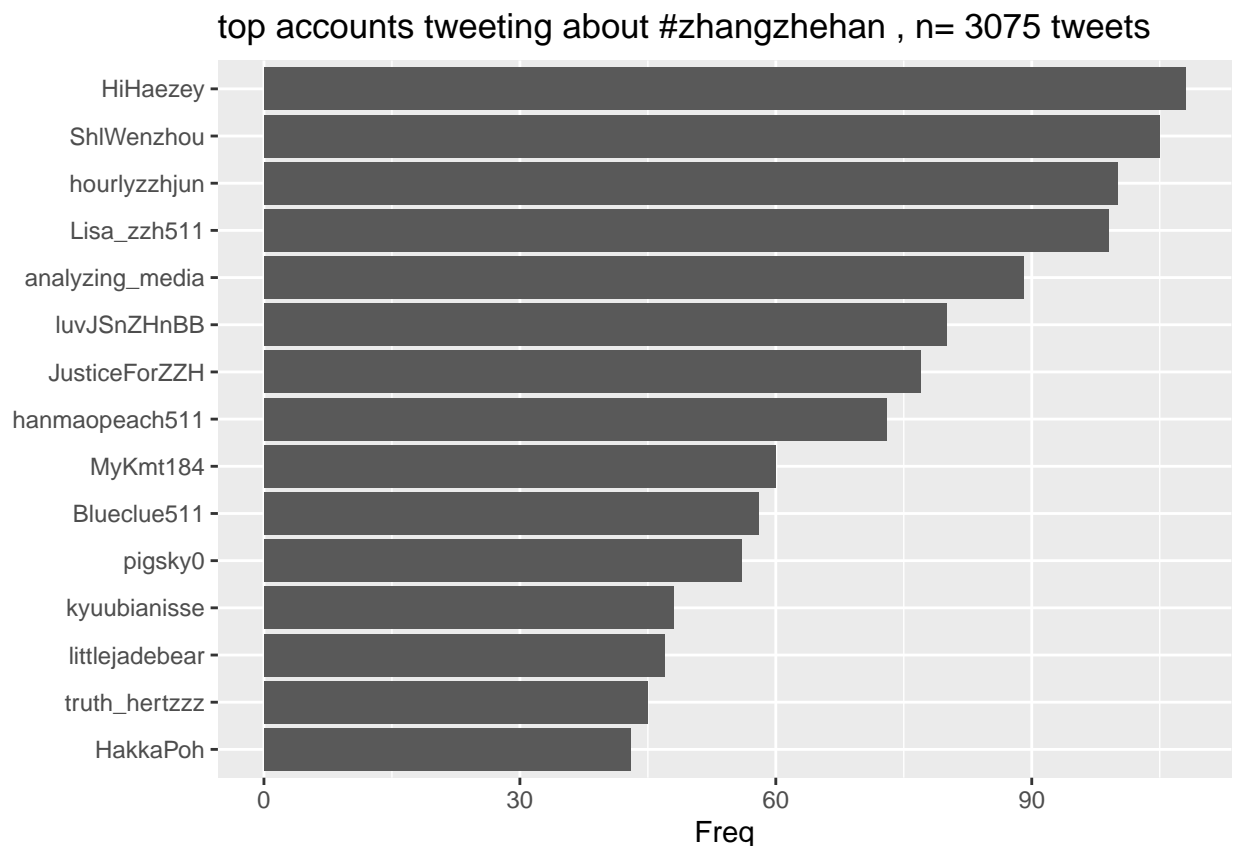
After playing around a bit I decided to use only hashtags instead of full names to focus on dedicated tweets instead of casual mentions.

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. Therefore this search is limited to the last couple of days.

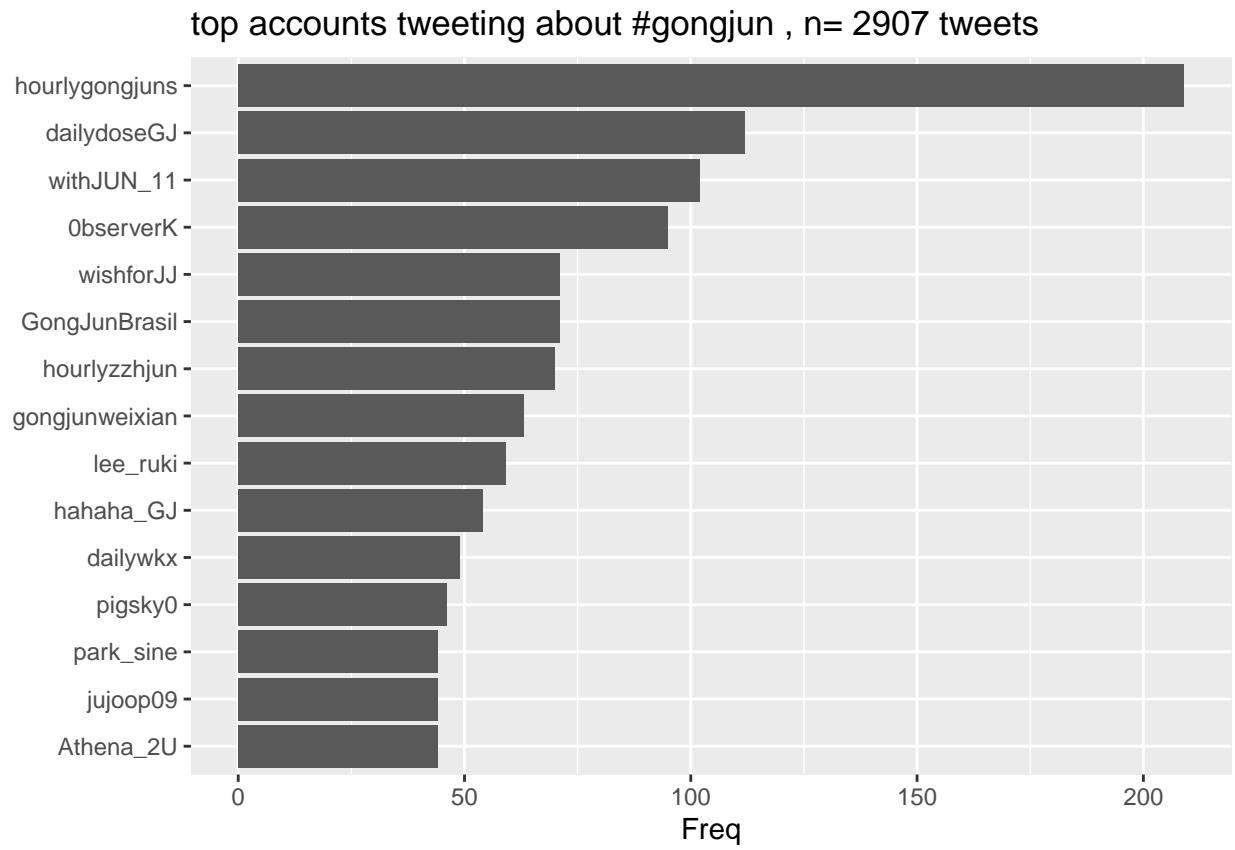
I will then plot the accounts with the most tweets about these topics.

Note: I previously filtered for tweets in English language, now no more language filter applied.

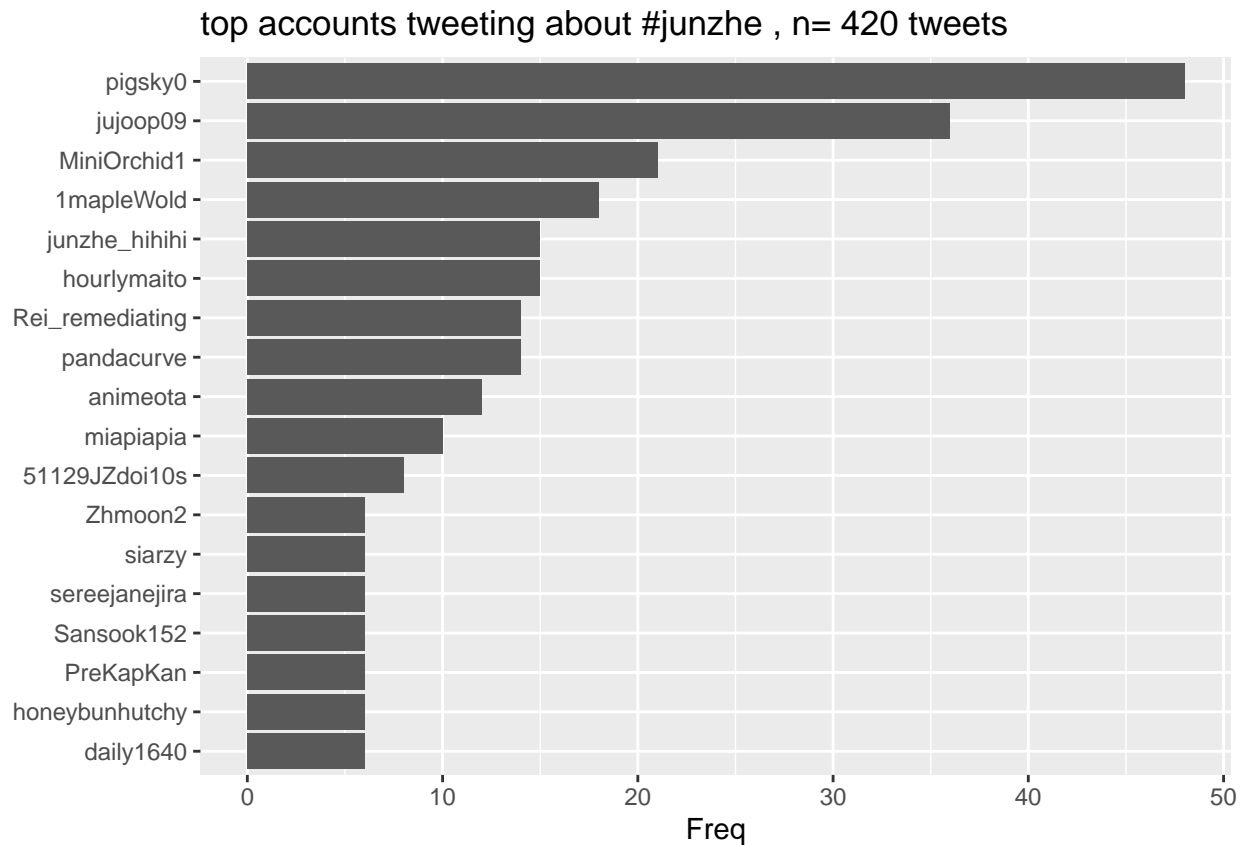
Selecting by Freq



Selecting by Freq



Selecting by Freq

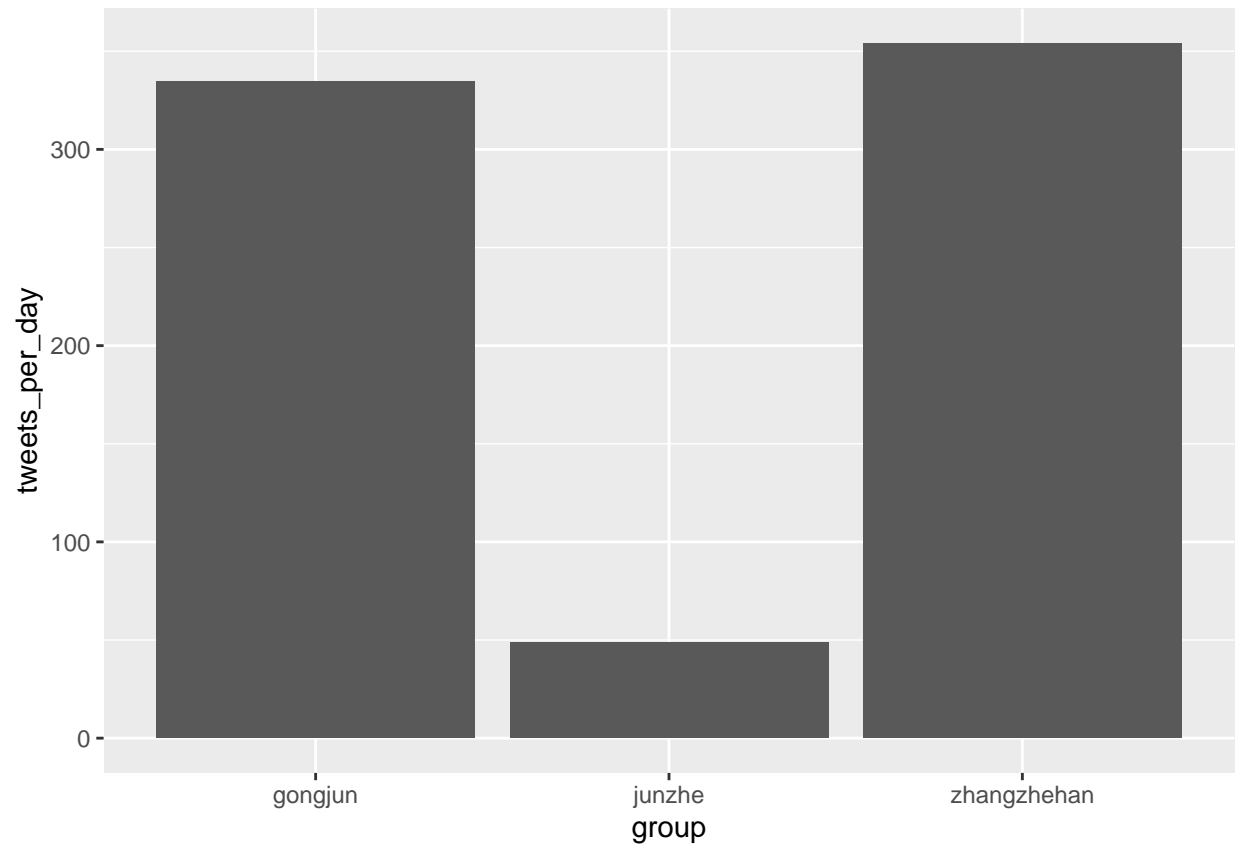


Quick info on which timespan the datasets cover and how many tweets per day

```
## [1] "dataset: #zhangzhehan"
## [1] "number of tweets: 3075"
## [1] "start: 2022-05-26 13:42:10"
## [1] "end: 2022-06-04 06:09:32"
## Time difference of 8.685671 days
## [1] "tweets per day: 354.03"
```

```
## [1] "dataset: #gongjun"
## [1] "number of tweets: 2907"
## [1] "start: 2022-05-26 13:42:09"
## [1] "end: 2022-06-04 06:08:42"
## Time difference of 8.685104 days
## [1] "tweets per day: 334.71"
```

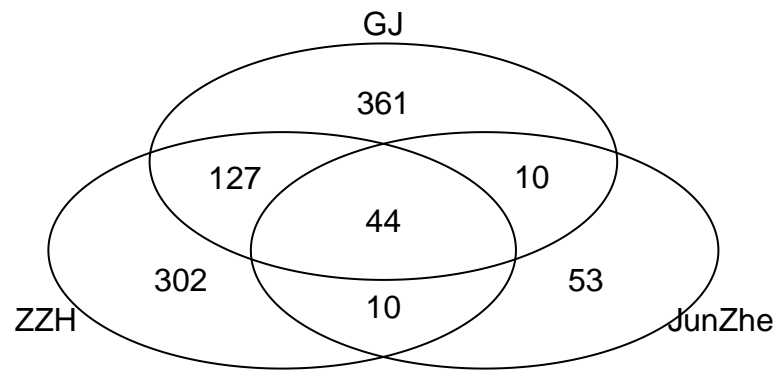
```
## [1] "dataset: #junzhe"
## [1] "number of tweets: 420"
## [1] "start: 2022-05-26 14:43:27"
## [1] "end: 2022-06-04 06:02:53"
## Time difference of 8.638495 days
## [1] "tweets per day: 48.62"
```



Next let's make a Venn diagram to see the overlap between accounts tweeting about #junzhe, #zhangzhehan or #gongjun. 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.

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

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

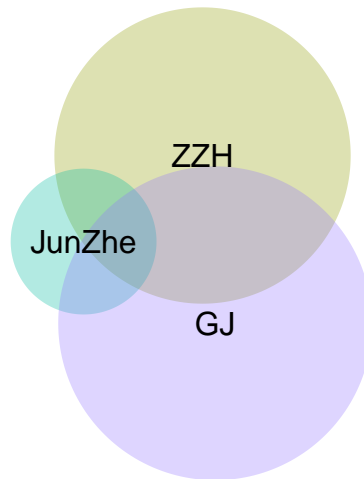


```
## [1] 483
```

```
## [1] 117
```

```
## [1] 542
```

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

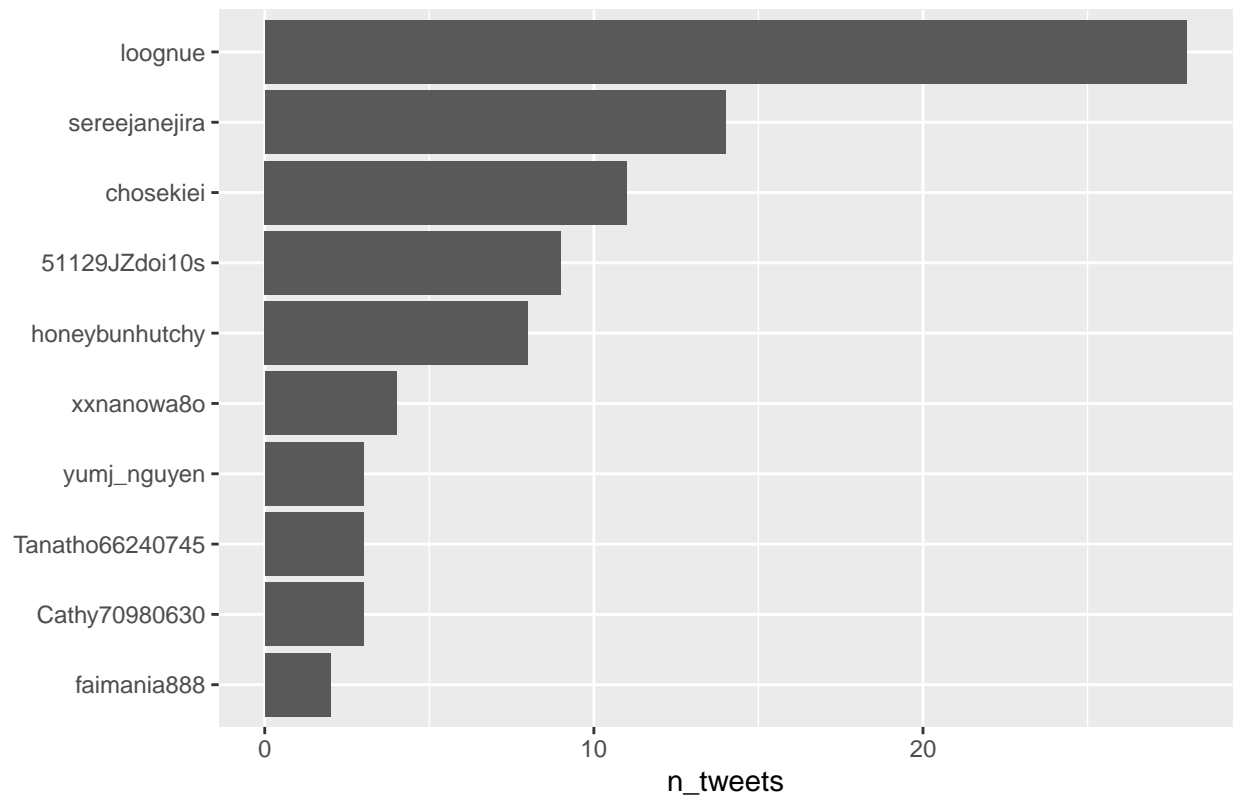


This Venn diagram only shows numbers of accounts tweeting about the three hashtags However I also want to see how many tweets are coming from each section of the Venn diagram.

```
## Selecting by n_tweets  
## Selecting by n_tweets  
## Selecting by n_tweets  
## Selecting by n_tweets  
## Selecting by n_tweets  
## Selecting by n_tweets  
## Selecting by n_tweets
```

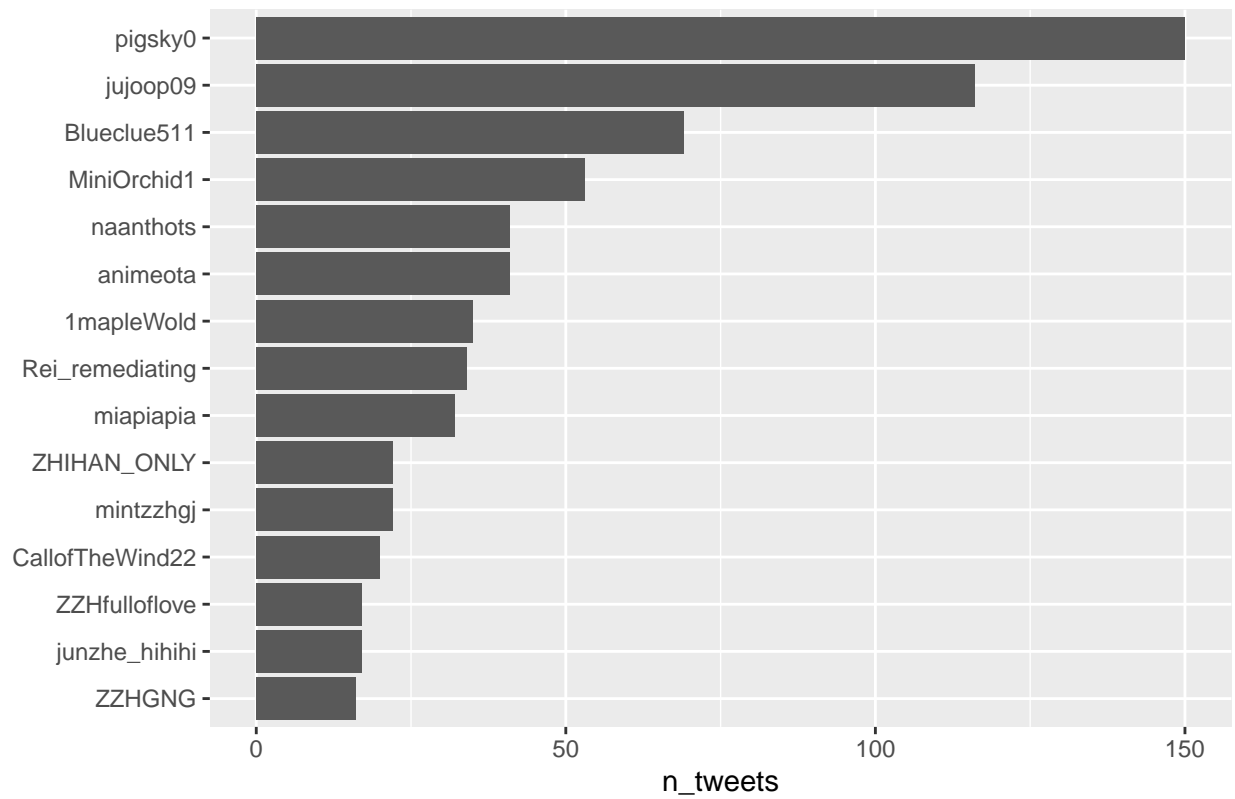
```
## [[1]]
```

top accounts tweeting about ZZH:JunZhe , n= 85 tweets



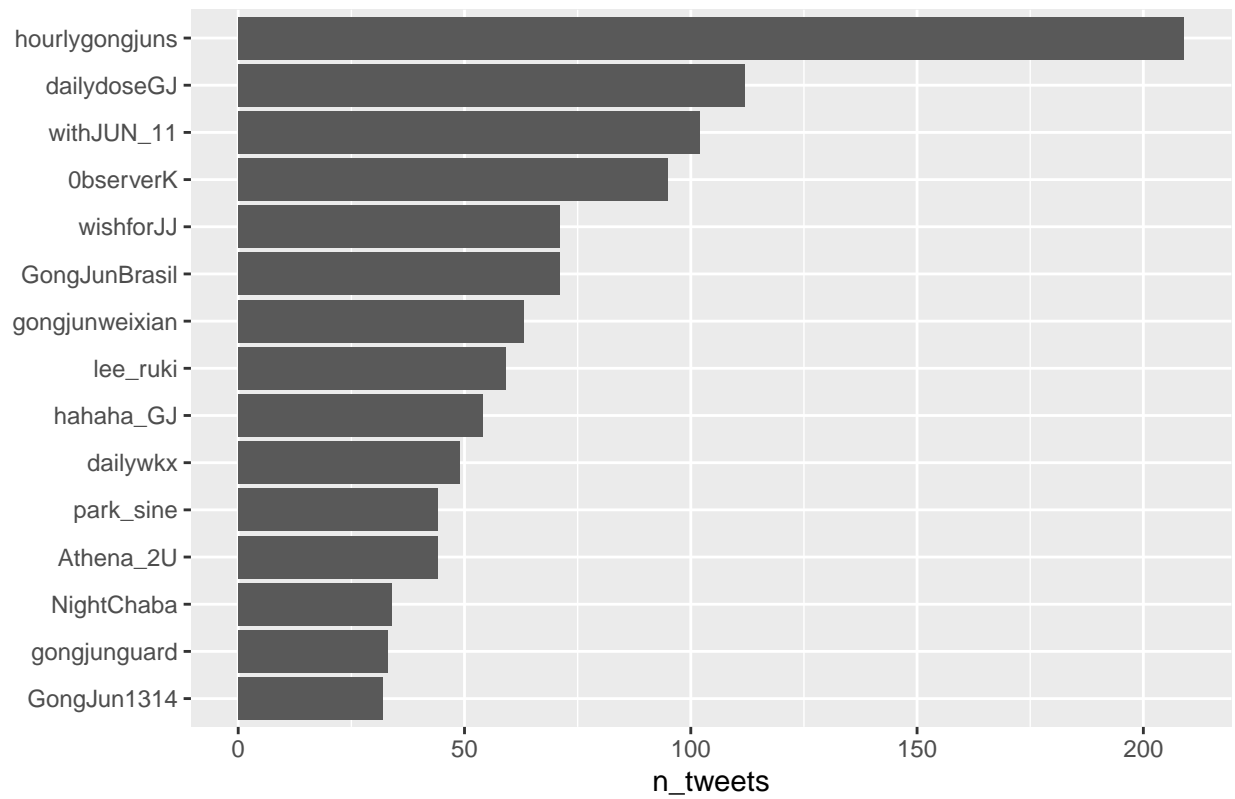
```
##  
## [[2]]
```

top accounts tweeting about ZZH:JunZhe:GJ , n= 884 tweets



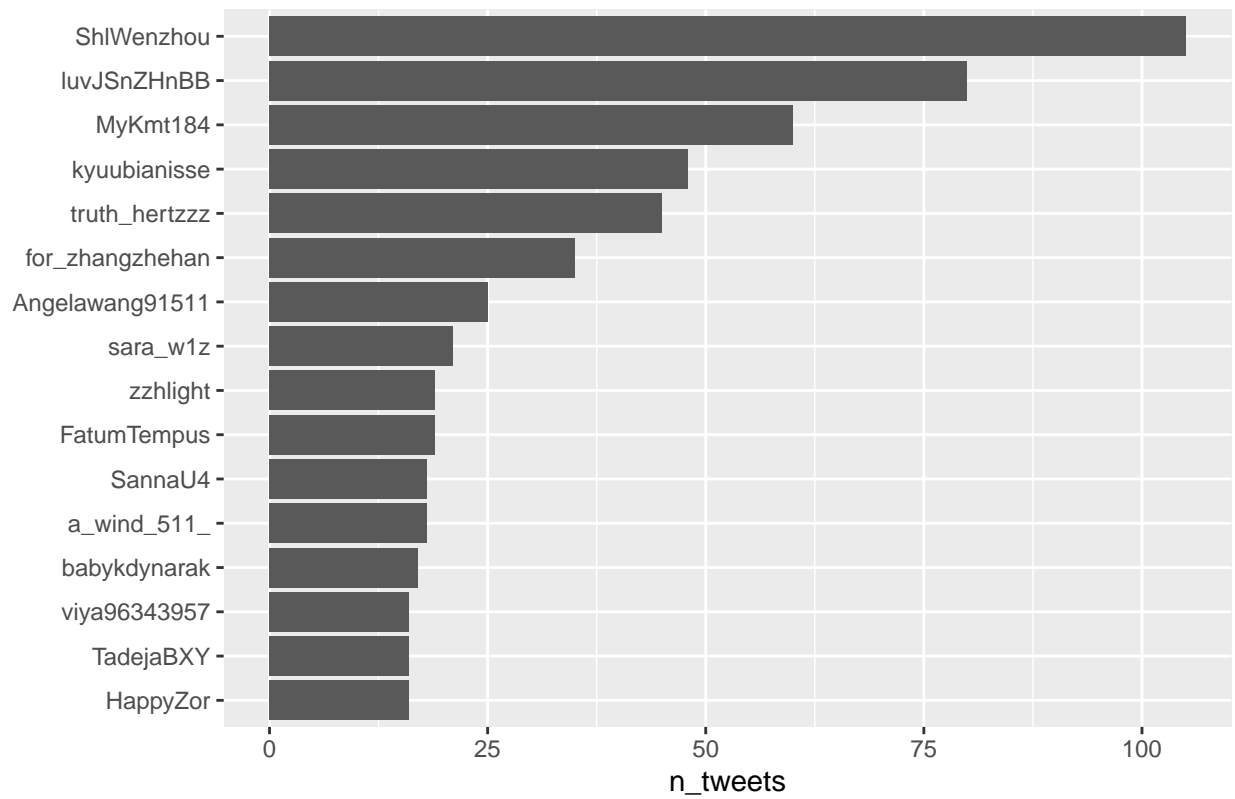
```
##  
## [[3]]
```


top accounts tweeting about GJ , n= 2067 tweets



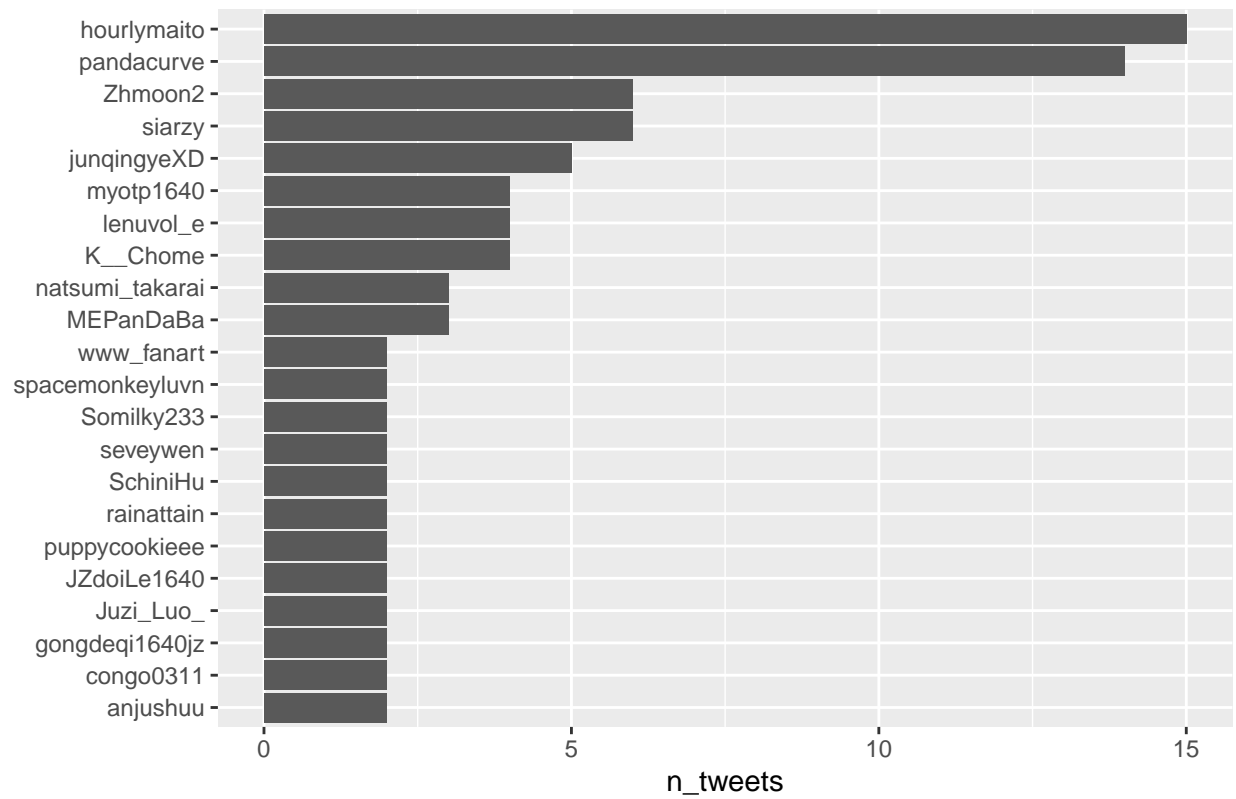
```
##  
## [[4]]
```

top accounts tweeting about ZZH , n= 1367 tweets



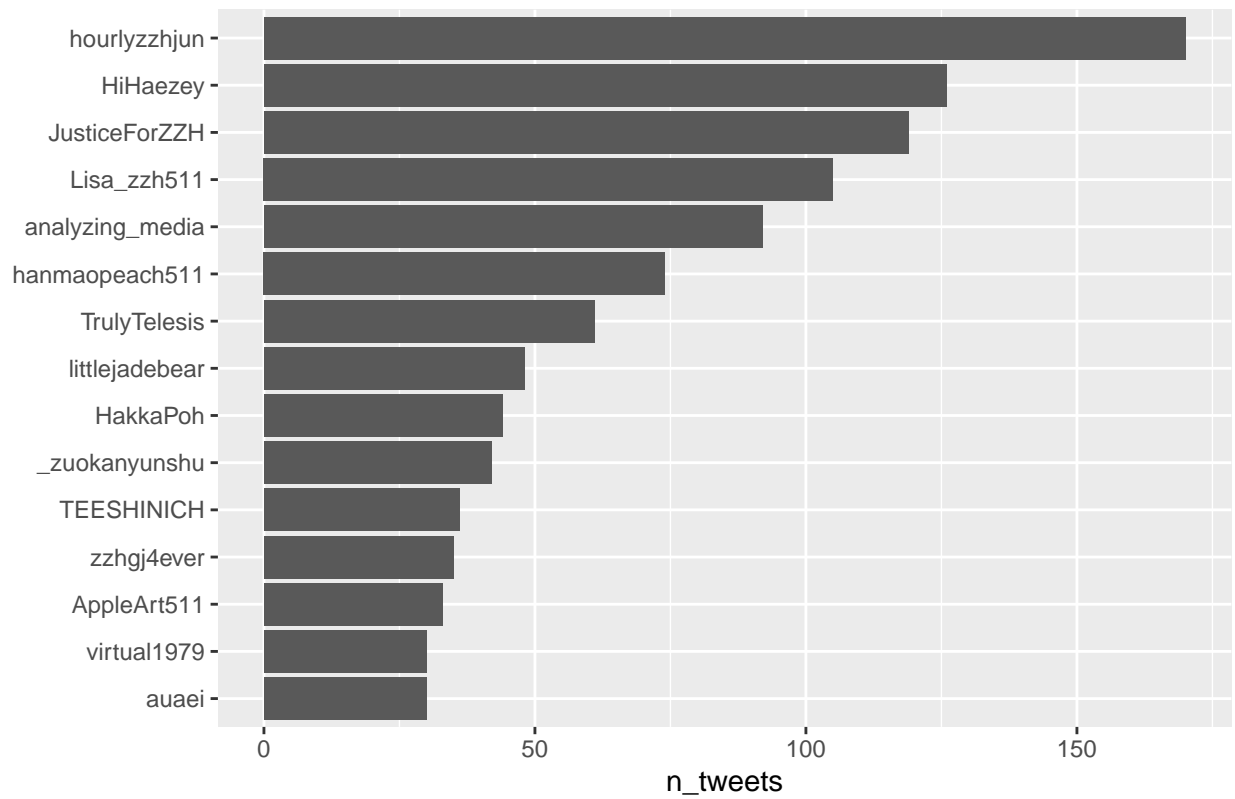
```
##  
## [[5]]
```

top accounts tweeting about JunZhe , n= 119 tweets

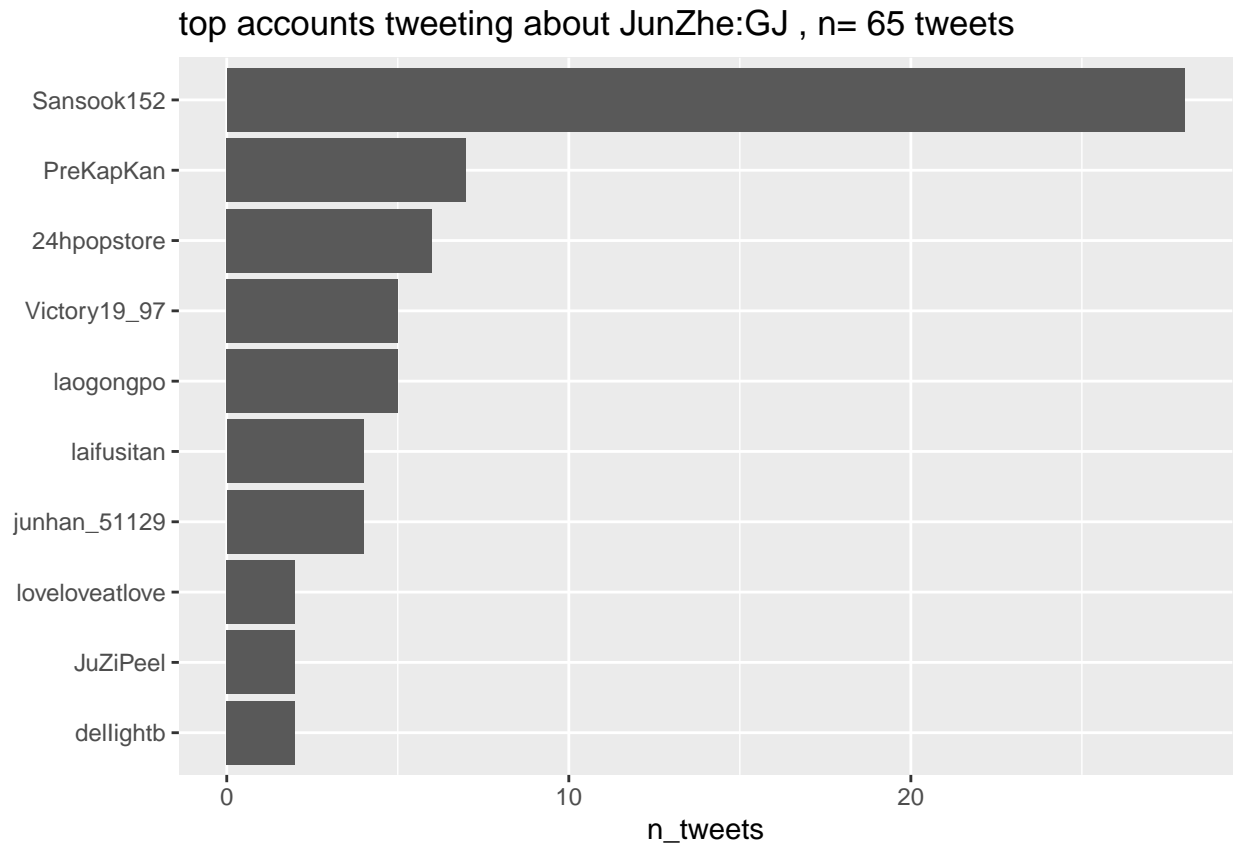


```
##  
## [[6]]
```

top accounts tweeting about ZZH:GJ , n= 1815 tweets



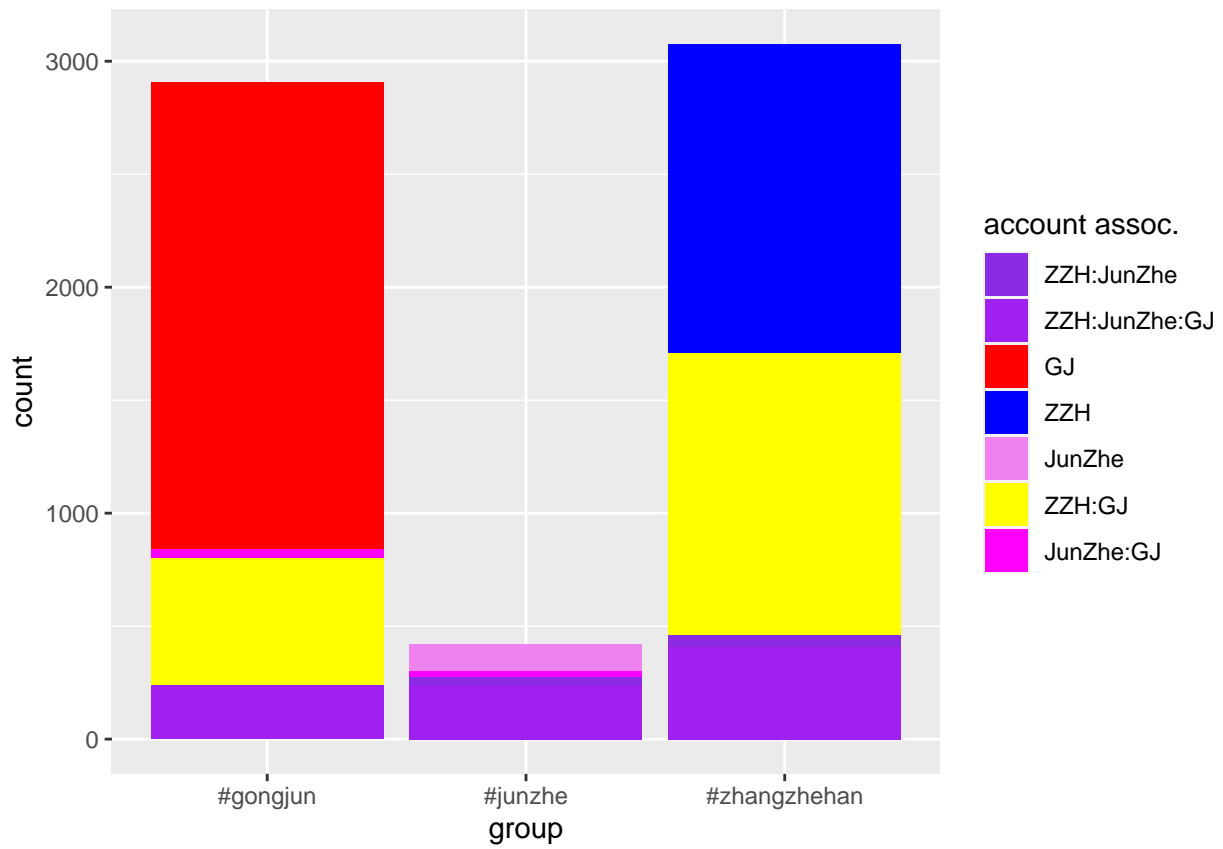
```
##  
## [[7]]
```



Now I have information on which accounts are also tweeting about the other two hashtags I looked at. Will now add this information to bar plot showing tweet counts per hashtag

```
## [1] "percent of ZZH tweets from 'pure' ZZH accounts"
## [1] 0.4445528
## [1] "percent of ZZH tweets from GJ-mentioning accounts"
## [1] 0.4061789
## [1] "percent of ZZH tweets from CP-related accounts"
## [1] 0.1492683
## [1] "percent of GJ tweets from 'pure' GJ accounts"
## [1] 0.7110423
## [1] "percent of GJ tweets from ZZH-mentioning accounts"
## [1] 0.1947024
## [1] "percent of GJ tweets from CP-related accounts"
```

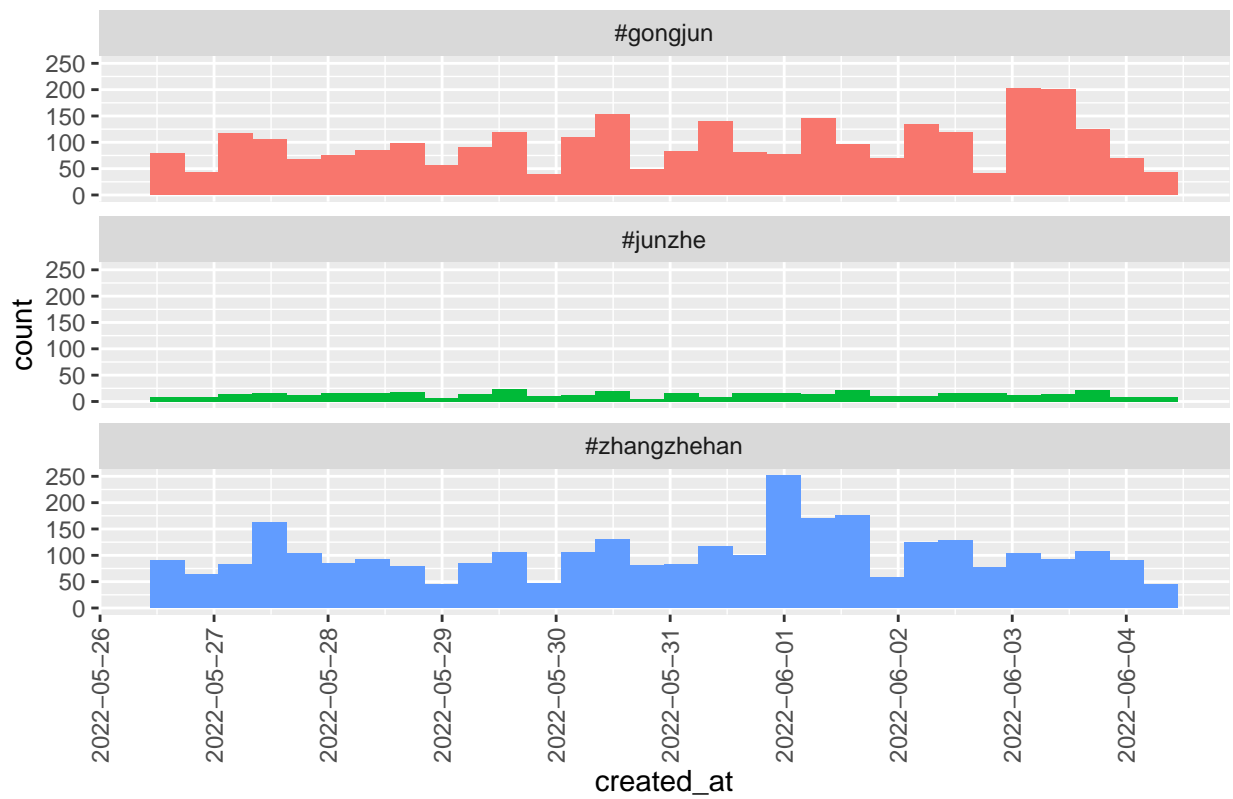
```
## [1] 0.09425525
```



Let's also look at tweets over time for all three hashtags

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

tweets over time



Same but this time fill color by venn slice

'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.

