Trendak_task

Salma Amr

Setup of twitteR package and authentication and changed the encodings:

```
#install.packages("twitteR")
#install.packages("dplyr")
library('dplyr')
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library('twitteR')
```

```
##
## Attaching package: 'twitteR'
```

```
## The following objects are masked from 'package:dplyr':
##
##     id, location
```

```
#install.packages("RCurl")
library('RCurl')
```

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

```
library('ggplot2')

Sys.setlocale("LC_ALL", "en_US.UTF-8")
```

```
## [1] "en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/C"
```

```
consumer_key <- 'igJt5cqmbSNfjnHXu0AUwvhe9'
consumer_secret <- 'XrvaWby4ZtebWk8cufosUic89lBJXhP1Pogn3Ll2LvaGxlnrY2'
access_token <- '554664043-7SSTDcrpejkM3IQNrtOdZKKAJnNhrkJg1QWURfdl'
access_secret <- 'hAxKsPvB9B919FDignD9WtipCDGPNOGADddTDgQ3Fw3YZ'
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

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

Search for @VodafoneEgypt tweets:

```
vodafone <- userTimeline(user = "VodafoneEgypt", n = 20000)
```

```
## Warning in statusBase(cmd, params, n, 3200, ...): statuses/user_timeline
## has a cap of 3200 statuses, clipping
```

```
vodafone %>% class()
```

```
## [1] "list"
```

```
vodafone_df <-  twListToDF(vodafone)

etisalat <- userTimeline(user = "EtisalatMisr", n = 20000)
```

```
## Warning in statusBase(cmd, params, n, 3200, ...): statuses/user_timeline
## has a cap of 3200 statuses, clipping
```

```
etisalat_df <-  twListToDF(etisalat)

orange <- userTimeline(user = "Orange_Egypt", n = 20000)
```

```
## Warning in statusBase(cmd, params, n, 3200, ...): statuses/user_timeline
## has a cap of 3200 statuses, clipping
```

```
## Warning in twInterfaceObj$doAPICall(cmd, params, method, ...): Rate limit
## encountered & retry limit reached - returning partial results
```

```
orange_df <-  twListToDF(orange)

orange_df = orange_df[-c(2)]
orange_df = orange_df[-c(5:10)]
orange_df = orange_df[-c(6:9)]

etisalat_df = etisalat_df[-c(2)]
etisalat_df = etisalat_df[-c(5:10)]
etisalat_df = etisalat_df[-c(6:9)]

vodafone_df = vodafone_df[-c(2)]
vodafone_df = vodafone_df[-c(5:10)]
vodafone_df = orange_df[-c(6:9)]
```

Save datasets into .CSV formats:

```
write.csv(etisalat_df, file = "etisalat.csv")
write.csv(vodafone_df, file = "vodafone.csv")
write.csv(orange_df, file = "orange.csv")
```

Do some processing on the tweets of each account:

```
vodafone_df$text <- lapply(vodafone_df$text, function(x) {
x = gsub('http\\S+\\s*', '', x) ## Remove URLs
x = gsub('#\\S+', '', x) ## Remove Hashtags
x = gsub('@\\S+', '', x) ## Remove Mentions
x = gsub("^[[:space:]]*","",x) ## Remove leading whitespaces
x = gsub("[[:space:]]*$","",x) ## Remove trailing whitespaces
x = gsub(' +',' ',x) ## Remove extra whitespaces
x = gsub('\\b+RT', '', x) ## Remove RT
x = gsub('[[:cntrl:]]', '', x) ## Remove Controls and special characters
x = gsub("\\d", '', x) ## Remove Controls and special characters
x = gsub('[[:punct:]]', '', x) ## Remove Punctuations
})


orange_df$text <- lapply(orange_df$text, function(x) {
x = gsub('http\\S+\\s*', '', x) ## Remove URLs
x = gsub('#\\S+', '', x) ## Remove Hashtags
x = gsub('@\\S+', '', x) ## Remove Mentions
x = gsub("^[[:space:]]*","",x) ## Remove leading whitespaces
x = gsub("[[:space:]]*$","",x) ## Remove trailing whitespaces
x = gsub(' +',' ',x) ## Remove extra whitespaces
x = gsub('\\b+RT', '', x) ## Remove RT
x = gsub('[[:cntrl:]]', '', x) ## Remove Controls and special characters
x = gsub("\\d", '', x) ## Remove Controls and special characters
x = gsub('[[:punct:]]', '', x) ## Remove Punctuations
})
```

```
etisalat_df$text <- lapply(etisalat_df$text, function(x) {
x = gsub('http\\S+\\s*', '', x) ## Remove URLs
x = gsub('#\\S+', '', x) ## Remove Hashtags
x = gsub('@\\S+', '', x) ## Remove Mentions
x = gsub("^[[:space:]]*","",x) ## Remove leading whitespaces
x = gsub("[[:space:]]*$","",x) ## Remove trailing whitespaces
x = gsub(' +',' ',x) ## Remove extra whitespaces
x = gsub('\\b+RT', '', x) ## Remove RT
x = gsub('[[:cntrl:]]', '', x) ## Remove Controls and special characters
x = gsub("\\d", '', x) ## Remove Controls and special characters
x = gsub('[[:punct:]]', '', x) ## Remove Punctuations
})
```

For announcements the “replyToSN = Null” for Orange and Etisalat user accounts only:

```
vodafone_ads <- vodafone_df %>% filter(is.na(replyToSN))
orange_ads <- orange_df %>% filter(is.na(replyToSN))
etisalat_ads <- etisalat_df %>% filter(is.na(replyToSN))
```

Trim data to have same time intervals:

Compare the total tweet counts through these time intervals and plot them:

```
vodafone_trim <- vodafone_df %>% filter(created > "2016-09-01")
etisalat_trim <- etisalat_df %>% filter(created > "2016-09-01")
orange_trim <- orange_df %>% filter(created > "2016-09-01")
```

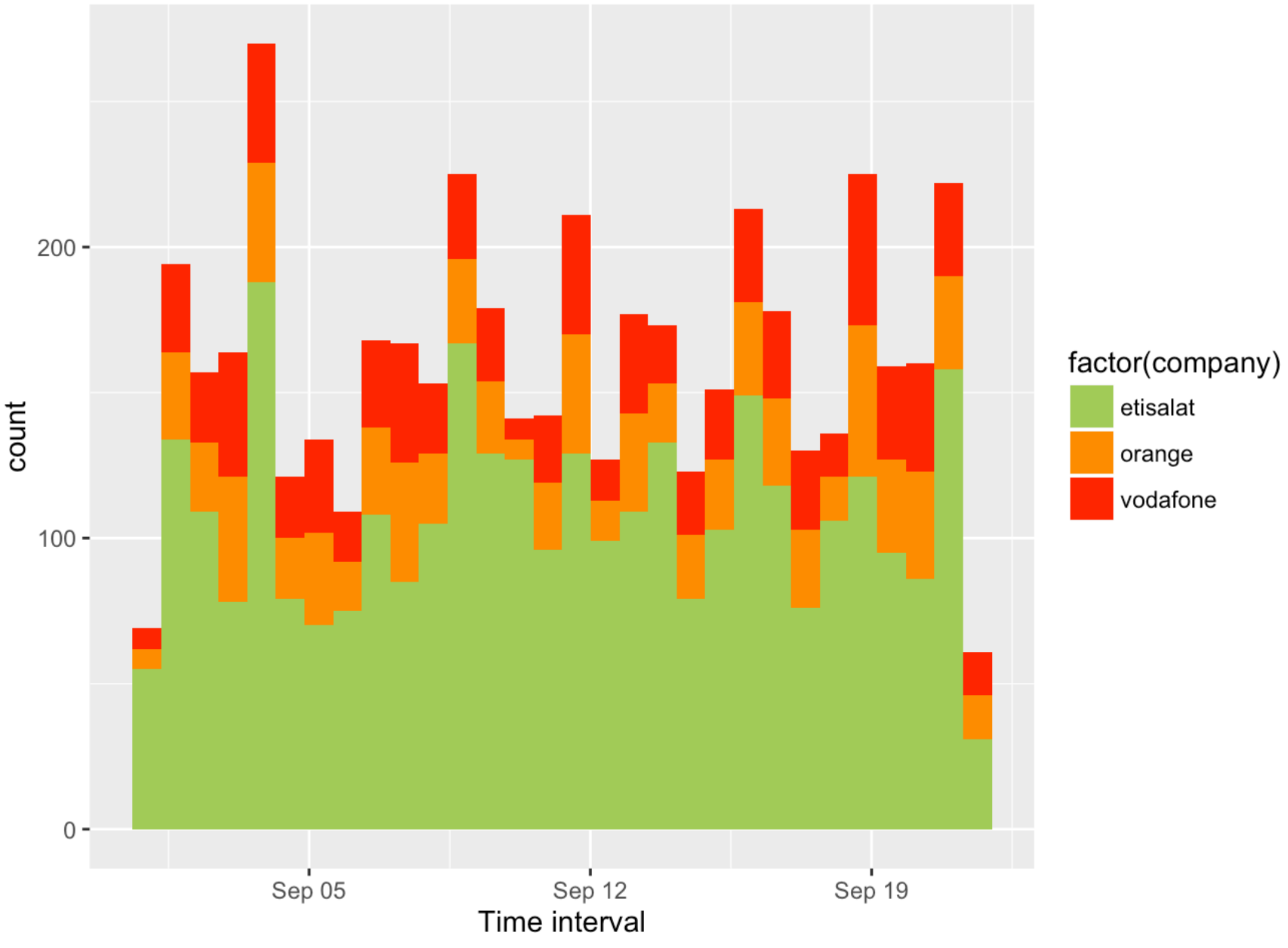
Get the account info of each user:

```
vodafone_user <- getUser('VodafoneEgypt')
etisalat_user <- getUser('EtisalatMisr')
orange_user <- getUser('Orange_Egypt')
```

Analysis on interval starting 1-9-2016 till 21-9-2016:

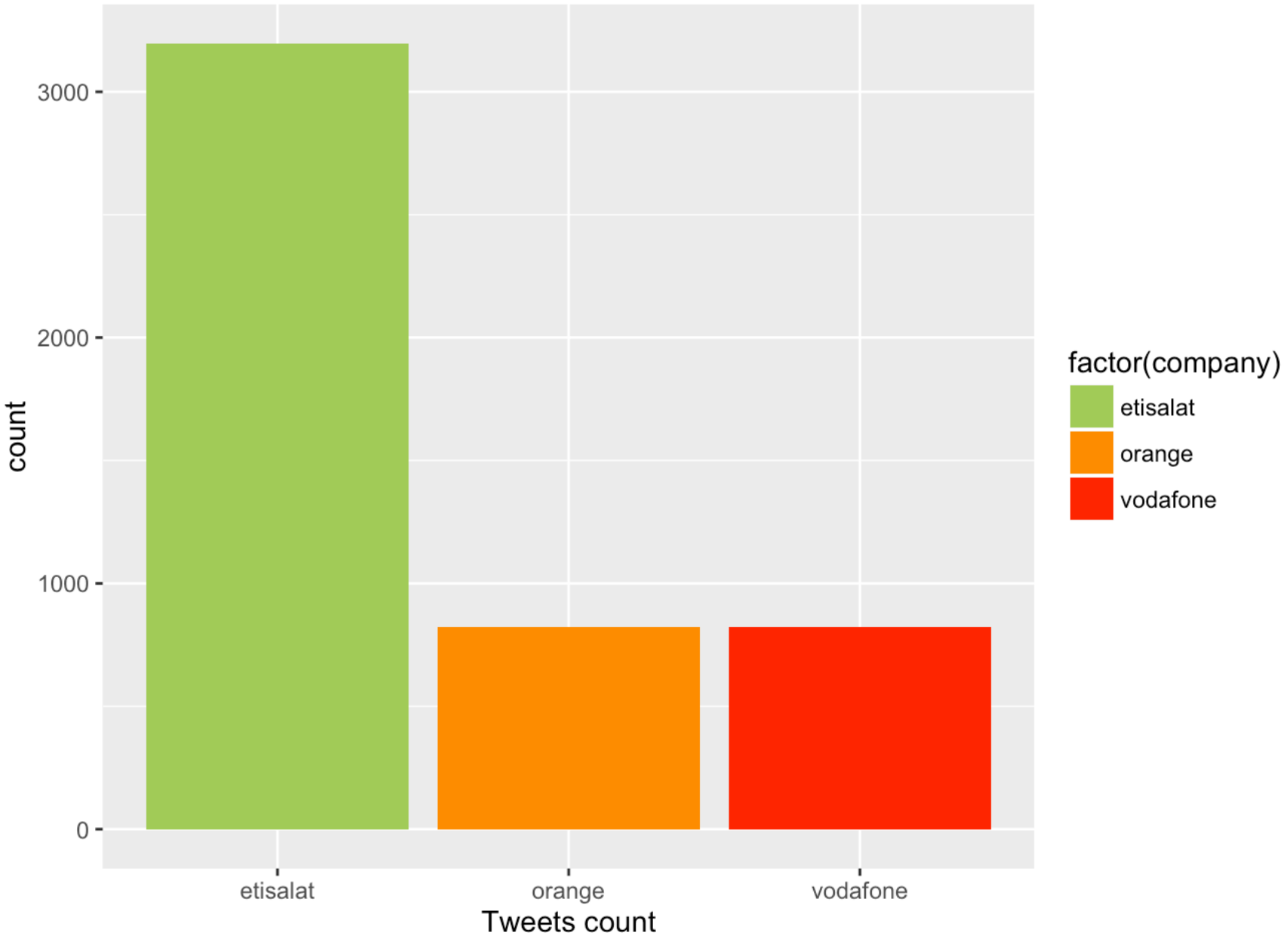
```
vodafone_trim <- vodafone_trim %>% mutate(company = "vodafone")
etisalat_trim <- etisalat_trim %>% mutate(company = "etisalat")
orange_trim <- orange_trim %>% mutate(company = "orange")

vodafone_trim$text <- as.character(vodafone_trim$text)
etisalat_trim$text <- as.character(etisalat_trim$text)
orange_trim$text <- as.character(orange_trim$text)

all_trim <- union(vodafone_trim, etisalat_trim) %>% union(orange_trim)

qplot(created, data=all_trim, fill=factor(company))  + scale_fill_manual(values=c(
"darkolivegreen3", "darkorange", "red")) + xlab("Time interval")
```

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



plot total number of tweets through the previously defined interval:

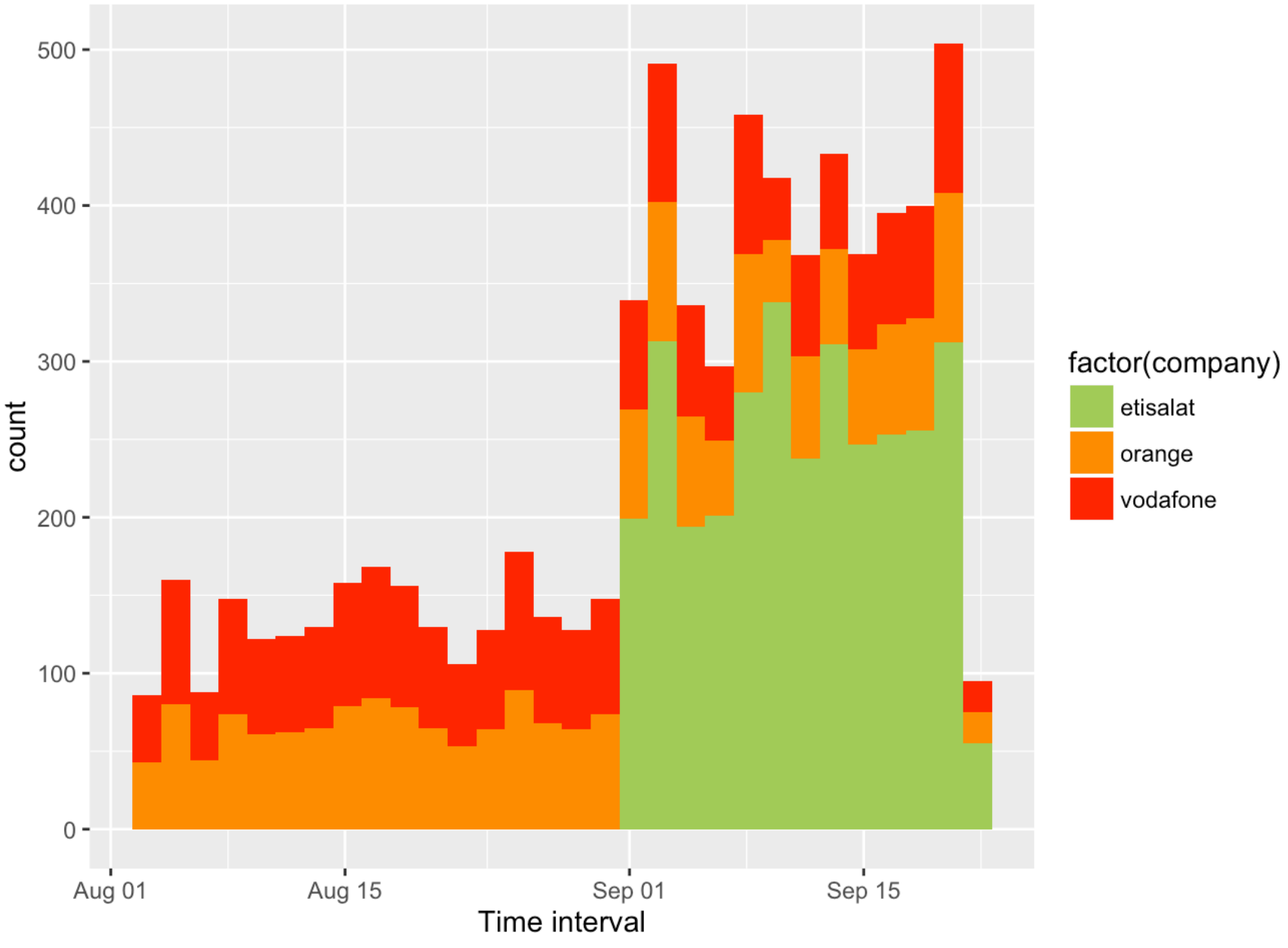
```
qplot(factor(company), data=all_trim, geom="bar", fill=factor(company)) + scale_fi
ll_manual(values=c("darkolivegreen3", "darkorange", "red")) + xlab("Tweets count")
```



Analysis on the most recent 3200 tweets:

```
vodafone_df <- vodafone_df %>% mutate(company = "vodafone")
etisalat_df <- etisalat_df %>% mutate(company = "etisalat")
orange_df <- orange_df %>% mutate(company = "orange")

vodafone_df$text <- as.character(vodafone_df$text)
etisalat_df$text <- as.character(etisalat_df$text)
orange_df$text <- as.character(orange_df$text)

all_df <- union(vodafone_df, etisalat_df) %>% union(orange_df)

qplot(created, data=all_df, fill=factor(company))  + scale_fill_manual(values=c("d
arkolivegreen3", "darkorange", "red")) + xlab("Time interval")
```

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



Extract Some info about three accounts:

Features	Vodafone	Mobinil	Etisalat
Followers count	1578352	974403	1270712
Tweets count	447649	134338	249237
Created Since	2009-07-13	2010-03-30	2010-04-26

Comments and Evaluation:

1. From the previous analysis, we can observe that Vodafone online twitter account is considered to be the most active account among the whole three accounts.
2. Considering the number of followers Vodafone is the highest, next comes Etisalat and Orange comes at the end.
3. Vodafone account is considered to be the oldest created one when considering the creation date, the comes Etisalat, least comes Orange.
4. The total number of tweets since the account was created is also the same and the previous rank, same when considering among the scpecified time interval from first of September till the fetch date.

5. Sentimate analysis could be furtherly done on the tweets forming a word cloud and thus we would be able to get more insights about the tweets contents and whether they fo reflect the satistaction of the customers or not.