The Chicken Group

Nitesh

07/10/2019

Load Libraries

library(rwhatsapp)  
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(ggplot2)  
library(lubridate)

##   
## Attaching package: 'lubridate'

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

library(tidytext)  
library(tidyr)  
library(stopwords)

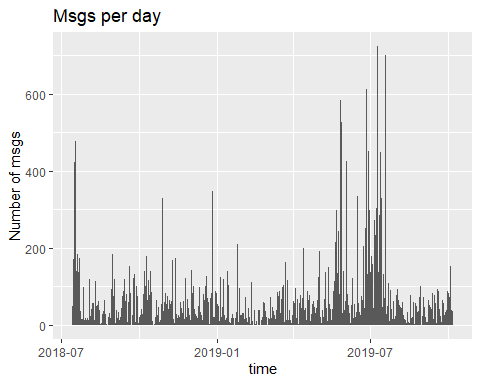
Importing files

chat <- rwa\_read("Chicken\_data.txt")  
chat <- na.omit(chat)  
levels(chat$author) <- c("Nitesh","Dusty","Kshitij","Kshitij","Nitesh","Paresh","Rajat","Tanwir")  
head(chat)

## # A tibble: 6 x 6  
## time author text source emoji emoji\_name  
## <dttm> <fct> <chr> <chr> <lis> <list>   
## 1 2018-07-14 20:07:32 Dusty Ye.. Argentina se ~ Chicken~ <chr~ <chr [0]>   
## 2 2018-07-14 20:17:32 Kshitij "\U0001f642" Chicken~ <chr~ <chr [1]>   
## 3 2018-07-14 20:18:32 Dusty "\U0001f923\U0001f~ Chicken~ <chr~ <chr [2]>   
## 4 2018-07-14 20:18:32 Dusty This emoji Chicken~ <chr~ <chr [0]>   
## 5 2018-07-14 20:18:32 Dusty <U+2665><U+2665><U+2665> Chicken~ <chr~ <chr [3]>   
## 6 2018-07-14 20:19:32 Kshitij Madharchod emoji h~ Chicken~ <chr~ <chr [0]>

**Lets first check the number of chit-chats in a day**

chat %>%  
 mutate(day=date(time)) %>%  
 count(day) %>%  
 ggplot(aes(x=day,y=n))+geom\_bar(stat="identity")+  
 ylab("Number of msgs")+xlab("time")+  
 ggtitle("Msgs per day")

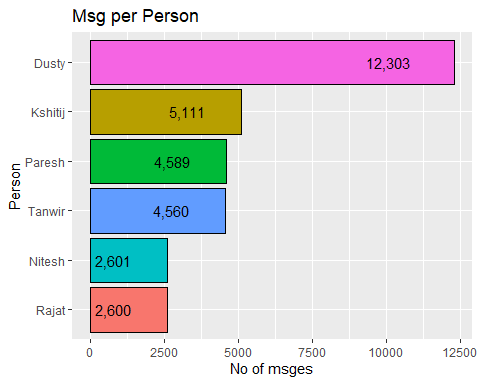


**umm chats have reached till 600 lets first see who is responsible of these talks ..also after 07/2018 we talked less until the placement season 07/2019 XD.**

**If you can find more information you are welcome**

Lets see total texts by person

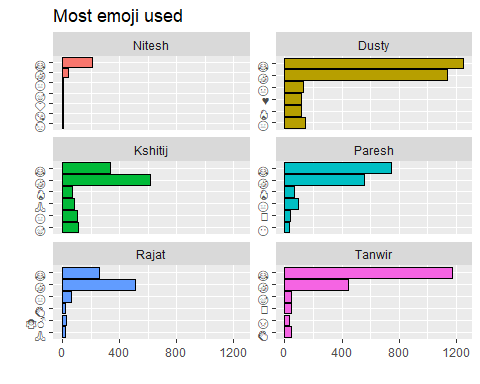
chat %>%  
 mutate(day=date(time)) %>%  
 count(author) %>%  
 ggplot(aes(x=reorder(author,n),y=n,fill=rainbow(6)))+  
 geom\_bar(stat="identity",col="black")+  
 ylab("No of msges")+xlab("Person")+  
 coord\_flip() + ggtitle("Msg per Person") + theme(legend.position = "none")+  
 geom\_text(aes(label = scales::comma(n)),hjust=2)



**Now we know who talks more** .

**Now its time to watch poeple fav emoji**

chat %>%  
 unnest(emoji) %>%  
 count(author,emoji,sort=T) %>%  
 group\_by(author)%>%  
 top\_n(n=6,n)%>%  
 ggplot(aes(x=reorder(emoji,n),y=n,fill=author))+  
 geom\_col(show.legend = F,col="black")+  
 ylab("")+xlab("")+coord\_flip()+  
 facet\_wrap(~author,ncol=2,scales="free\_y")+  
 ggtitle("Most emoji used")

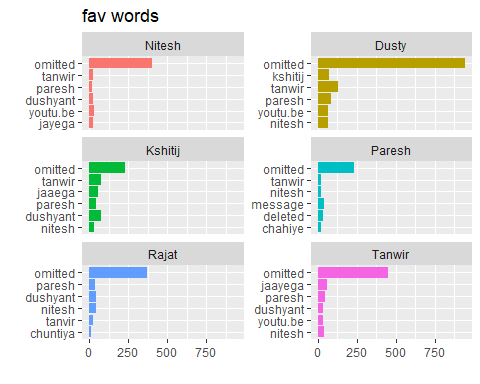


**You can judge personalities here …..**

**Only i kiss here looks like I M GAY**

**Okay lets look at favorate words to determine**

chat %>%  
 unnest\_tokens(input=text,output=word) %>%  
 count(author,word,sort=T)%>%  
 filter(nchar(word)>5) %>%  
 group\_by(author) %>%  
 top\_n(n=6,n) %>%  
 ggplot(aes(x=reorder(word,n),y=n,fill=author))+  
 geom\_col(show.legend = F)+  
 ylab("")+xlab("")+ggtitle("fav words")+  
 coord\_flip()+facet\_wrap(~author,ncol=2,scales="free\_y")

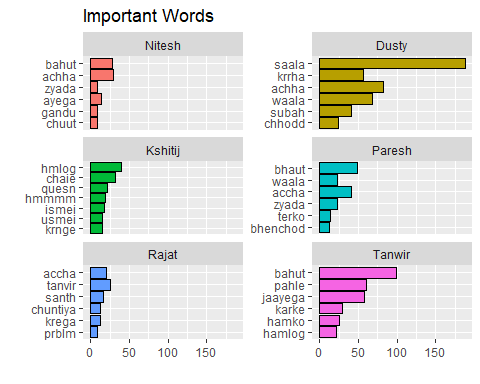


**Now i have a list of poeple who call each other .Also we share a lot of confidential documents** Me dushyant and tanwir are good with youtube <3 .Rajat has “chutiya” in his vocab ,god bless

Dushyant Likes tanwir;it has high frequency

**Getting words that are diffrent for each other will be of more value**

to\_remove <- c(stopwords("english"),"omitted","message","delete","media","nahi")  
chat %>%  
 unnest\_tokens(input=text,output=word)%>%  
 select(word,author) %>%  
 filter(nchar(word)>4) %>%  
 filter(!word %in% to\_remove) %>%  
 mutate(word=gsub(".com","",word))%>%  
 count(author,word,sort=T) %>%  
 bind\_tf\_idf(term=word,document=author,n=n)%>%  
 filter(n>7)%>%  
 group\_by(author)%>%  
 top\_n(n=6,tf\_idf)%>%  
 ggplot(aes(x=reorder(word,n),y=n,fill=author))+  
 geom\_col(show.legend = F,col="black")+xlab("")+ylab("")+  
 coord\_flip()+ggtitle("Important Words")+  
 facet\_wrap(~author,ncol=2,scales="free\_y")



**Tanwir and kshitj use less slangs**

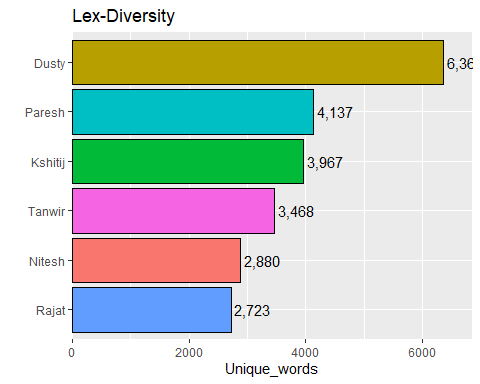
**Paresh is from delhi**

**dushant is frequent with saala**

**Nitesh is using ayega ya nahi ayega**

**Now lets check who uses more unique words**

chat %>%  
 unnest\_tokens(input=text,output=word)%>%  
 filter(!word %in% to\_remove) %>%  
 group\_by(author)%>%  
 summarise(lex\_diversity=n\_distinct(word))%>%  
 arrange(desc(lex\_diversity))%>%  
 ggplot(aes(x=reorder(author,lex\_diversity),y=lex\_diversity,fill=author)) + geom\_col(show.legend = F,col="black")+  
 xlab("")+ylab("Unique\_words")+  
 scale\_y\_continuous(expand=(multi = c(0,0,0,500)))+  
 coord\_flip()+ggtitle("Lex-Diversity")+  
 geom\_text(aes(label = scales::comma(lex\_diversity)), hjust = -0.1)



**Paresh was low on total word than kshitij,but he has vocab** **Also my total count and vocab was good ,respect me bitches**

**Any more suggestion of insights are welcome if you can find them** **I dont work for free**