WhatsApp Chat Analysis

18F-0421

12/20/2021

## R Markdown

library(stringr)  
library(ggplot2)  
library("rwhatsapp")

## Warning: package 'rwhatsapp' was built under R version 4.1.2

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("lubridate")

##   
## Attaching package: 'lubridate'

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

library("tidyr")  
library("tidytext")

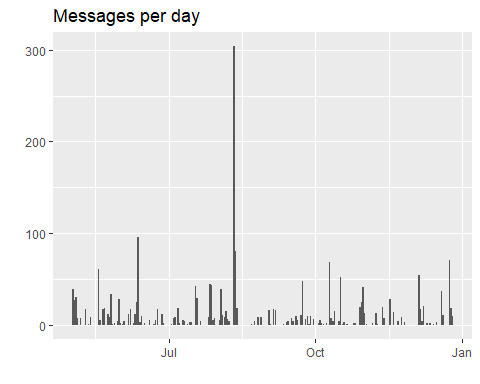
## Warning: package 'tidytext' was built under R version 4.1.2

library("stopwords")

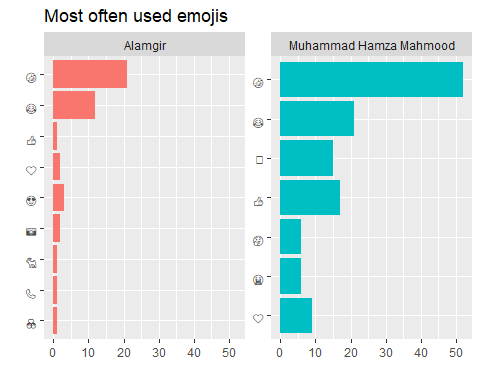
## Warning: package 'stopwords' was built under R version 4.1.2

## Including Plots

chat <- rwa\_read("WhatsAppChat.txt") %>%   
 filter(!is.na(author))   
  
chat %>%  
 mutate(day = date(time)) %>%  
 count(day) %>%  
 ggplot(aes(x = day, y = n)) +  
 geom\_bar(stat = "identity") +  
 ylab("") + xlab("") +  
 ggtitle("Messages per day")



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



chat %>%  
 unnest\_tokens(input = text,  
 output = word) %>%  
 count(author, word, sort = TRUE) %>%  
 group\_by(author) %>%  
 top\_n(n = 4, n) %>%  
 ggplot(aes(x = reorder\_within(word, n, author), y = n, fill = author)) +  
 geom\_col(show.legend = FALSE) +  
 ylab("") +  
 xlab("") +  
 coord\_flip() +  
 facet\_wrap(~author, ncol = 2, scales = "free\_y") +  
 scale\_x\_reordered() +  
 ggtitle("Most often used words")

