WhatsApp Chat Analysis in R

18F-0178\_M Mustafa Ali

12/7/2021

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

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

##   
## Attaching package: 'dplyr'

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

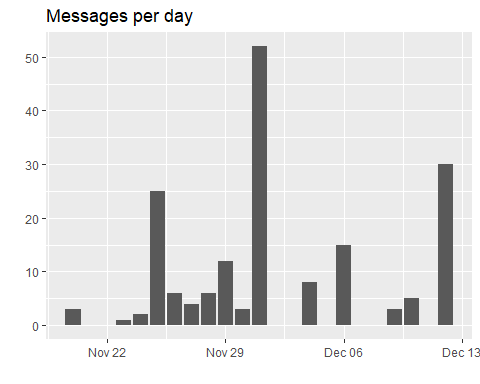
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library("lubridate")

##   
## Attaching package: 'lubridate'

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

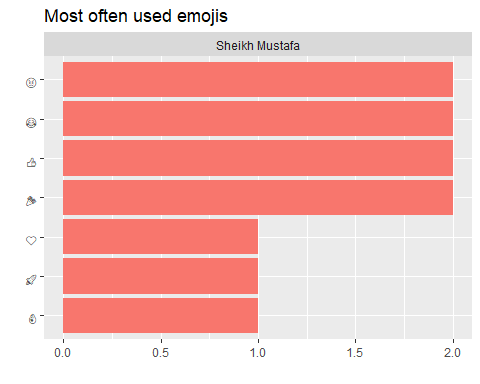
library("tidyr")  
library("tidytext")  
library("stopwords")  
  
chat <- rwa\_read("chat.txt")  
chat <- rwa\_read("chat.txt") %>%   
 filter(!is.na(author))   
  
chat %>%  
 mutate(day = date(time)) %>%  
 count(day) %>%  
 ggplot(aes(x = day, y = n)) +  
 geom\_bar(stat = "identity") +  
 ylab("")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



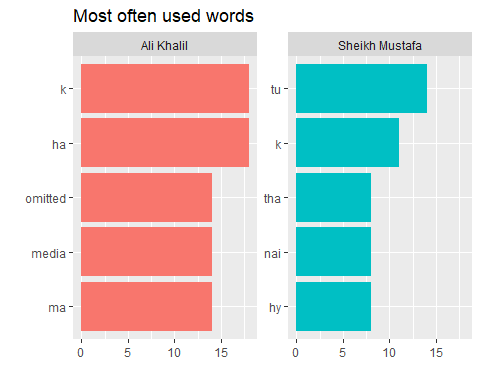
## Including Plots

You can also embed plots, for example:

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")



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.