DS Assignment   
WhatsApp  
  
library(rwhatsapp)

library(dplyr)

library("tidyr")

chat <- rwa\_read(file.choose())

View(chat)

library("ggplot2"); theme\_set(theme\_minimal())

library("lubridate")

chat %>%

+ mutate(day = date(time)) %>%

+ count(day) %>%

+ ggplot(aes(x = day, y = n)) +

+ geom\_bar(stat = "identity") +

+ ylab("") + xlab("") +

+ ggtitle("Messages per day")

chat %>%

mutate(day = date(time)) %>%

count(author) %>%

ggplot(aes(x = reorder(author, n), y = n)) +

geom\_bar(stat = "identity") +

ylab("") + xlab("") +

coord\_flip() +

ggtitle("Number of messages")

#to show emojis

library("ggimage")

emoji\_data <- rwhatsapp::emojis %>%

+ mutate(hex\_runes1 = gsub("\\s[[:alnum:]]+", "", hex\_runes)) %>%

+ mutate(emoji\_url = paste0("https://abs.twimg.com/emoji/v2/72x72/",

+ tolower(hex\_runes1), ".png"))

chat %>%

+ unnest(emoji) %>%

+ count(author, emoji, sort = TRUE) %>%

+ group\_by(author) %>%

+ top\_n(n = 6, n) %>%

+ left\_join(emoji\_data, by = "emoji") %>%

+ ggplot(aes(x = reorder(emoji, n), y = n, fill = author)) +

+ geom\_col(show.legend = FALSE) +

+ ylab("") +

+ xlab("") +

+ coord\_flip() +

+ geom\_image(aes(y = n + 20, image = emoji\_url)) +

+ facet\_wrap(~author, ncol = 2, scales = "free\_y") +

+ ggtitle("Most often used emojis") +

+ theme(axis.text.y = element\_blank(),

+ axis.ticks.y = element\_blank())

>

library("tidytext")

chat %>%

+ unnest\_tokens(input = text,

+ output = word) %>%

+ count(author, word, sort = TRUE) %>%

+ group\_by(author) %>%

+ top\_n(n = 6, 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")

to\_remove <- c(stopwords(language = "de"),

+ "media",

+ "omitted",

+ "i",

+ "you",

+ "it",

+ "the",

+ "and",

+ "to",

+ "a",

+ "i'm", "that", "but", "do", "me")

>

> chat %>%

+ unnest\_tokens(input = text,

+ output = word) %>%

+ filter(!word %in% to\_remove) %>%

+ count(author, word, sort = TRUE) %>%

+ group\_by(author) %>%

+ top\_n(n = 6, 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")

>

