WhatsAppChatAnalysisInR

18F-0114_Muhammad Zain Ul Abidin

12/27/2021

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")
```

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

```
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## 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")
```

```
## Warning: package 'lubridate' was built under R version 4.1.2
```

```
##
## Attaching package: 'lubridate'
```

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

```
library("tidyr")
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
library("tidytext")
```

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

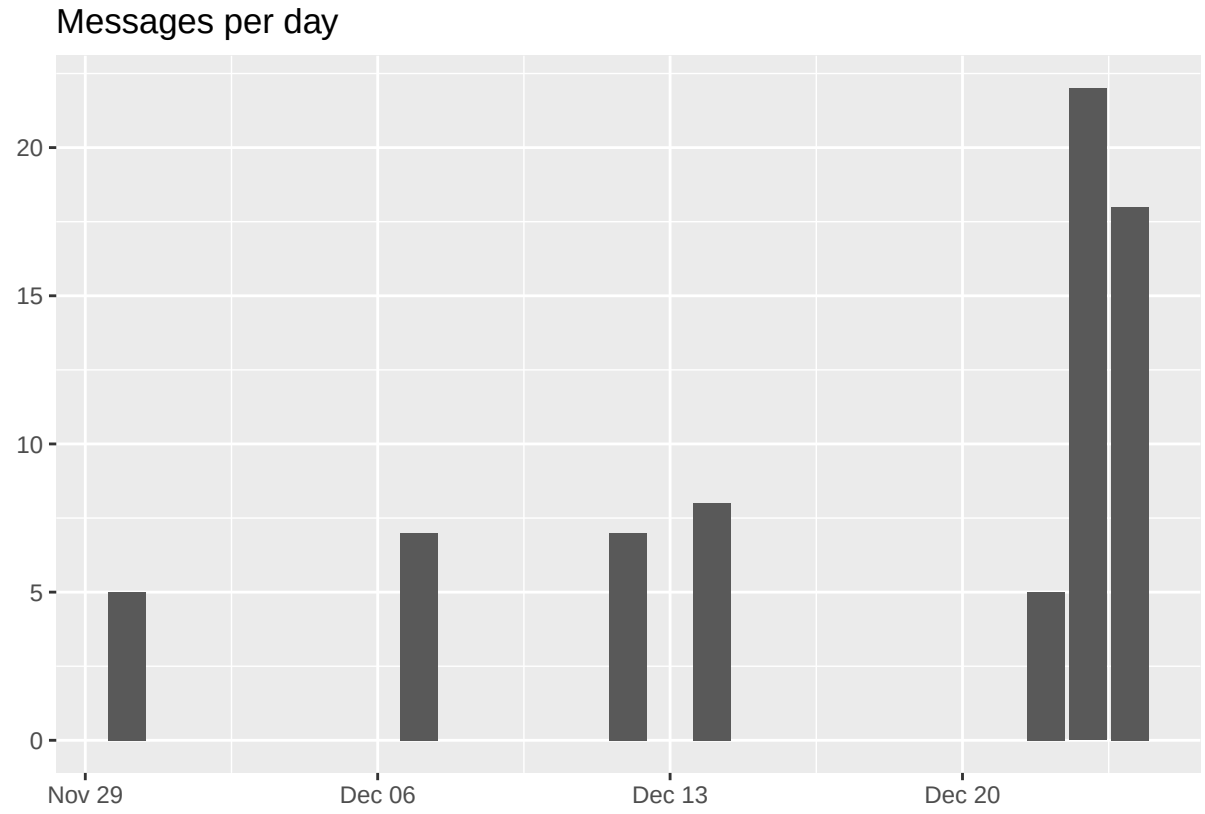
```
library("stopwords")
```

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

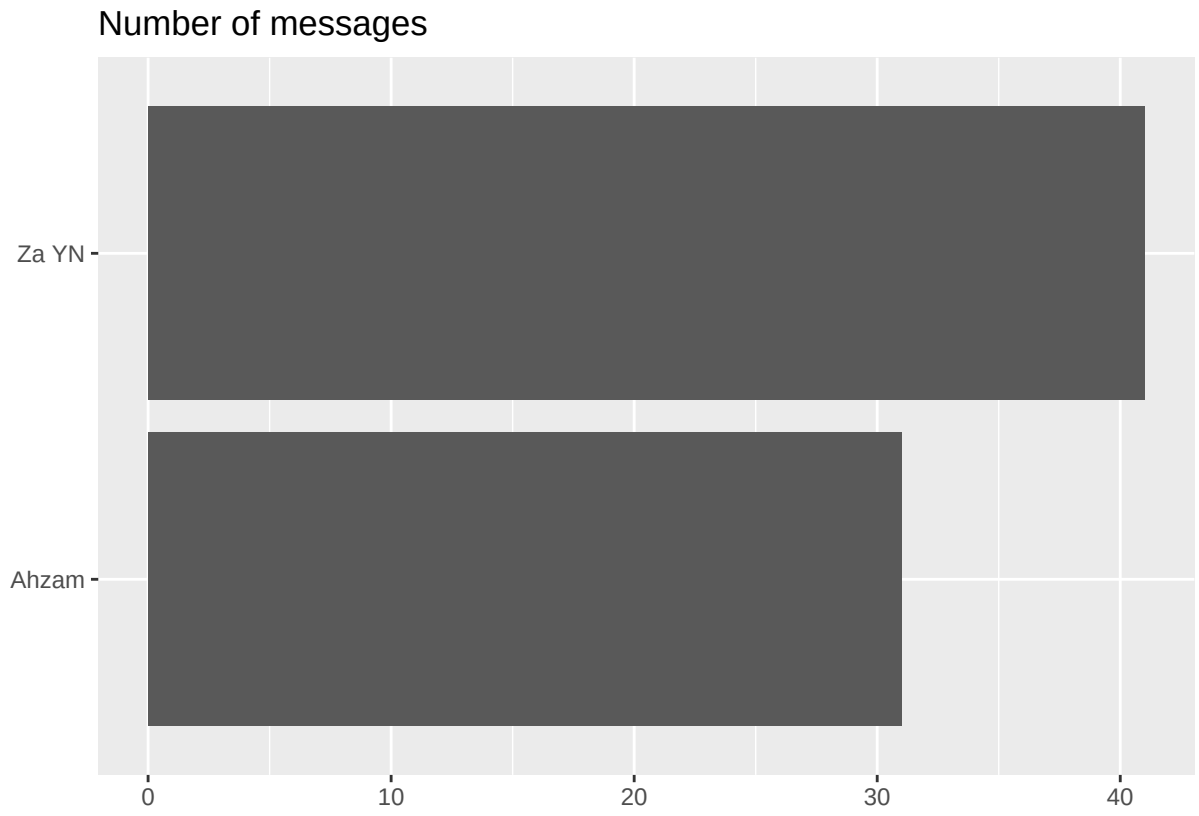
Including Plots

You can also embed plots, for example:

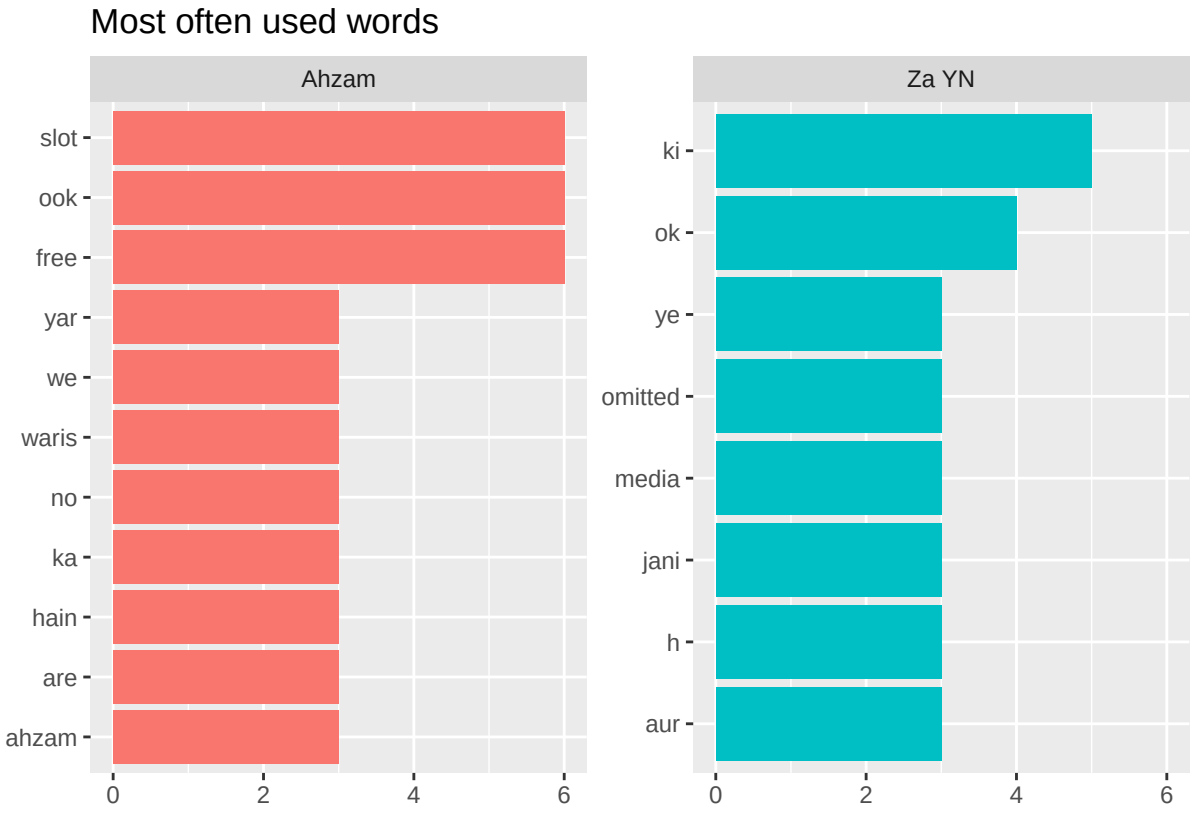
```
chat <- rwa_read("MyChat.txt") %>%
  filter(!is.na(author))

chat %>%
  mutate(day = date(time)) %>%
  count(day) %>%
  ggplot(aes(x = day, y = n)) +
  geom_bar(stat = "identity") +
  ylab("") + xlab("") +
  ggtitle("Messages per day")
```




```
chat %>%
  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")
```




```
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.