Word Frequency

Logan Calder

12/12/2017

```
library(lubridate)
```

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

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

```
library(tidytext)
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## as.difftime(): lubridate, base
## date():        lubridate, base
## filter():      dplyr, stats
## intersect():   lubridate, base
## lag():         dplyr, stats
## setdiff():     lubridate, base
## union():       lubridate, base
```

```
load("data_processed/strike_reports.RData")
```

```
dim(strike_reports)
```

```
## [1] 1026    2
```

```
n_reports = 1062
top_n_words = 15
```

```
words_ranked <- (strike_reports[1:n_reports,] %>%
    unnest_tokens(output = "word", input = "report_text") %>%
    anti_join(stop_words) %>%
    group_by(
        word
```

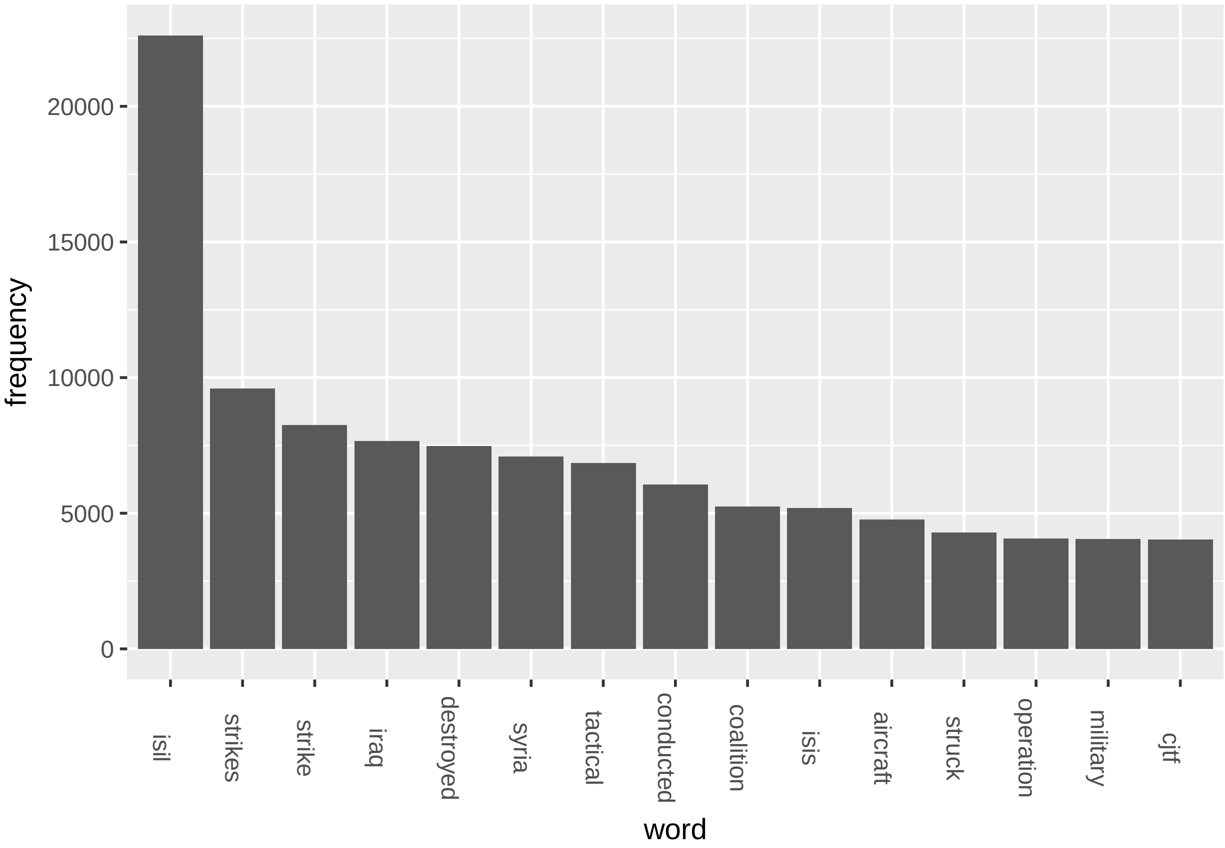
```r
    ) %>%
    summarize(
        frequency = length(word)
    ) %>%
    arrange(desc(frequency)) %>%
    mutate(
        word = factor(word, levels = word[1:top_n_words])
    ))[1:top_n_words,]
```

```
## Joining, by = "word"
```

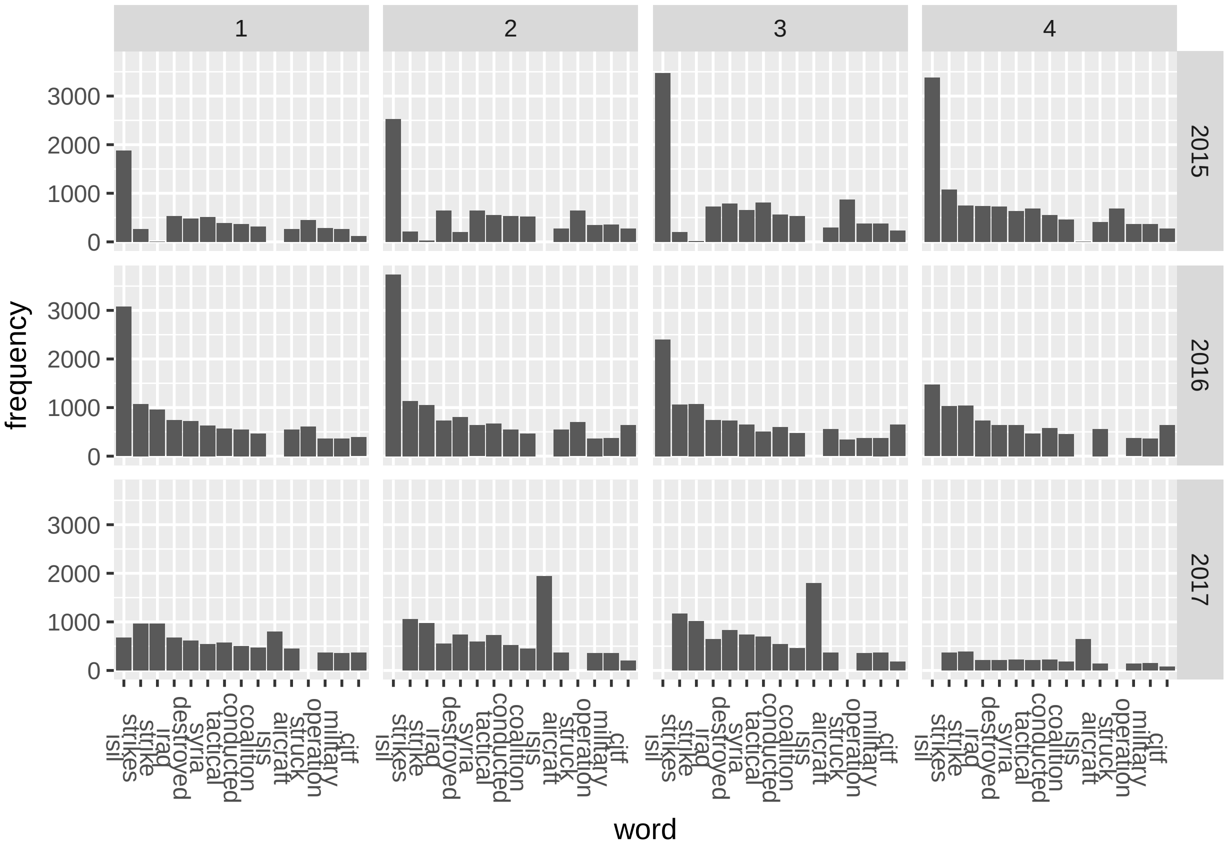
```r
words_ <-
    strike_reports[1:n_reports,] %>%
    unnest_tokens(output = "word", input = "report_text") %>%
    anti_join(stop_words) %>%
    transmute(
        year = year(report_created_date),
        quarter = quarter(report_created_date),
        word = factor(word, words_ranked$word)
    ) %>%
    filter(!is.na(word)) %>%
    group_by(
        year, quarter, word
    ) %>%
    summarize(
        frequency = length(word)
    )
```

```
## Joining, by = "word"
```

```r
ggplot(words_ranked, mapping = aes(x = word, y = frequency)) + geom_col() +
    theme(axis.text.x = element_text(angle = -90))
```



```
ggplot(words_, mapping = aes(x = word, y = frequency)) + geom_col() + facet_grid(year~quarter) +
    theme(axis.text.x = element_text(angle = -90))
```


1
2
3
4
2015
2016
2017
frequency
3000
2000
1000
0
isil
strikes
strike
iraq
destroyed
syria
tactical
conducted
coalition
isis
aircraft
struck
operation
military
cjtf
word