M4A_Project

Ross Pingatore

10/30/2020

```r
library(readxl)
medicare_for_all_1 <- read_excel("Data/medicare_for_all_1.xlsx")
single_payer_system_1 <- read_excel("Data/single_payer_system_1.xlsx")
universal_health_care_1 <- read_excel("Data/universal_health_care_1.xlsx")
medicare_for_all_2 <- read_excel("Data/medicare_for_all_2.xlsx")
single_payer_system_2 <- read_excel("Data/single_payer_system_2.xlsx")
universal_health_care_2 <- read_excel("Data/universal_health_care_2.xlsx")

data_names <- c(medicare_for_all_1,single_payer_system_1,universal_health_care_1,medicare_for_all_2,sing
```

```r
suppressPackageStartupMessages(library(tidyverse))
merged_data <- bind_rows(medicare_for_all_1,single_payer_system_1,universal_health_care_1,medicare_for_a

write_excel_csv(merged_data, 'Data/merged_data.xlsx')
view(merged_data)
```

```r
unique(merged_data$Language)
```

```
## [1] "en"  "und" "ja"  "tl"  "es"  "ro"  "zh"  "sv"  "nl"  "pt"
```

```r
library(tidytext)
library(tm)
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```r
library(dplyr)
library(NLP)
#https://stackoverflow.com/questions/36824296/r-remove-specific-word-in-a-txte-like-the-this

corpus <- Corpus(VectorSource(merged_data$Text))


corpus <- tm_map(corpus, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removePunctuation): transformation drops
## documents
```

```r
corpus <- tm_map(corpus, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(corpus, removeNumbers): transformation drops
## documents
corpus <- tm_map(corpus, tolower)

## Warning in tm_map.SimpleCorpus(corpus, tolower): transformation drops documents
corpus <- tm_map(corpus, removeWords, stopwords("english"))

## Warning in tm_map.SimpleCorpus(corpus, removeWords, stopwords("english")):
## transformation drops documents
corpus <- tm_map(corpus, stripWhitespace)

## Warning in tm_map.SimpleCorpus(corpus, stripWhitespace): transformation drops
## documents
corpus <- tm_map(corpus, stemDocument)

## Warning in tm_map.SimpleCorpus(corpus, stemDocument): transformation drops
## documents
doc_matrix <- TermDocumentMatrix(corpus)
doc_matrix_m <- as.matrix(doc_matrix)
doc_matrix_val <- sort(rowSums(doc_matrix_m), decreasing = T)
doc_matrix_df <- data.frame(word = names(doc_matrix_val), freq = doc_matrix_val)

top_20 <- head(doc_matrix_df, 20)

top_20%>%
  ggplot(aes(reorder(word,freq), freq)) + geom_bar(stat = "identity") + xlab("Term") + ylab("Frequency")
```


Most Frequent Terms Within Tweets Relating to Medicare For All