EDS 231: Assignment 3

Scout Leonard

04/20/2022

Load Libraries

```
packages=c("quanteda.sentiment",
           "quanteda.textstats",
           "tidyverse",
           "tidytext",
           "lubridate",
           "here",
           "wordcloud", #visualization of common words in the data set
           "reshape2",
           "quanteda") #devtools::install_github("quanteda/quanteda.sentiment") #not available currentl

for (i in packages) {
  if (require(i,character.only=TRUE)==FALSE) {
    install.packages(i,repos='http://cran.us.r-project.org')
  }
  else {
    require(i,character.only=TRUE)
  }
}
```

Import Data

```
raw_tweets <- read.csv("https://raw.githubusercontent.com/MaRo406/EDS_231-text-sentiment/main/dat/IPCC_

dat<- raw_tweets[,c(5,7)] # Extract Date and Title fields

tweets <- tibble(text = dat$Title,
                 id = seq(1:length(dat$Title)),
                 date = as.Date(dat$Date,'%m/%d/%y'))


#head(tweets$text, n = 10)
```

You will use the tweet data from class today for each part of the following assignment.

Part 1

Think about how to further clean a twitter data set. Let's assume that the mentions of twitter accounts is not useful to us. Remove them from the text field of the tweets tibble.

Part 2

Compare the ten most common terms in the tweets per day. Do you notice anything interesting?

Part 3

Adjust the wordcloud in the “wordcloud” chunk by coloring the positive and negative words so they are identifiable.

Part 4

Let's say we are interested in the most prominent entities in the Twitter discussion. Which are the top 10 most tagged accounts in the data set. Hint: the "explore_hashtags" chunk is a good starting point.

Part 5

The Twitter data download comes with a variable called "Sentiment" that must be calculated by Brandwatch. Use your own method to assign each tweet a polarity score (Positive, Negative, Neutral) and compare your classification to Brandwatch's (hint: you'll need to revisit the "raw_tweets" data frame).