Monica Vashu Kherajani
Person #50290424
mkheraja@buffalo.edu

Lab 1 Part 3

Step 1: Collection of tweets

The following describes the file names for particular keyword collection of tweets after cleaning. The rest of the files in the folder show some intermediate uncleaned tweets data.

1. Using 'flu' as keyword – fluKeywordCleanData2.csv (3418 Tweets) or fluKeywordCleanData.csv (495 Tweets)
2. Using 'influenza' as keyword – influenzaKeywordCleanData.csv (759 Tweets)
3. Using 'flu OR influenza' as a query argument - fluORInfluenzaKeywordCleanData.csv (3412 Tweets)
4. Using 'flu vaccine' as keyword – MyDataforfluvaccineCleaned.csv (484 Tweets)
5. Using 'influenza virus' as keyword - MyDataforinfluenzavirusCleaned.csv (62 Tweets)

Step 2: Cleaning Tweets/ data

1. To collect the data in a csv file, the data frame obtained from search_tweets of rTweet API was transformed. Some columns were in list format – unlisted them using paste and then stored the entire data frame in csv format.
2. The 'is_retweet' is checked for FALSE value to get only tweets and filter any retweets in data.
3. The list of US states obtained from: http://www.whypad.com/posts/excel-spreadsheet-of-us-states/583/ is used to get the correct locations of Users from location field.
4. The mapped states have three columns – state in uppercase, state, and the state abbreviations.
5. The location column in the obtained data frame are now in abbreviation of states format.
6. The states are mapped to the location field in the dataframe using sqldf and by joining using the common column with state abbreviations.
7. Finally, the cleaned data is written onto a csv file for plotting the heatmap.

The code for Step 1 and Step 2 are given in ***lab1p3.r.***
The code was repeated for various keywords, however, as a sample only two keywords code is shown. The keys for authentication have been removed for privacy concerns.

Step 3: Heat maps for comparison

1. Various heat maps of the above obtained 5 different cleaned keywords data are plotted.

Step 4: Shiny app

1. The comparison of various tweets from different keywords are plotting using a shiny app format of UI and server. The link is: https://monicakherajani.shinyapps.io/appsample/

This code is in ***app*** folder.

2. Using images obtained from the jupyter notebook are hosted using a Navigation bar. The link is: https://vaidehidhr.shinyapps.io/R_app/

This code is in ***app1*** folder.