Biafra Disinformation Project

The group we are interested in is Biafra separtists. Biafra existed as an secessionist state from May 1967 to January 1970 during the Nigerian Civil War. There was mass starvation in Biafra due to Nigerian blockade which killed almost 2 million Biafran civilans, mostly small children.

Many of those most active in the Biafran independence movement of today, like the leader of the Indigenous People of Biafra (IPOB), are too young to remember the war and do not currently live in Nigeria. In 2017, Nigeria designated IPOB as a terrorist organization.

Biafra separtists are known to heavily use social media to spread their message, and a significant amount of the message appears to be **disinformation**.

This project uses the Armed Conflict Database available in GECIQ along with Twitter data scrapped with Python's *Twint* to investigate the connection between sometimes dishonest social media posts and violence in Nigeria.

R Code part 1

In [ ]:
```
library(readxl)
library(data.table)
library(sqldf)
```

In [ ]:
```
GEC_Data <- read_excel("Nigeria Armed Conflicts.xlsx")
```

Look at the primary actors in the armed conflicts

In [ ]:
```
# Main actor fields

actor1 <- as.data.frame(table(GEC_Data$actor1))
actor1 <- actor1[order(actor1$Freq, decreasing = TRUE),]
head(actor1, 20)

actor2 <- as.data.frame(table(GEC_Data$actor2))
actor2 <- actor2[order(actor2$Freq, decreasing = TRUE),]
head(actor2, 20)

# More detailed actors

assocActor1 <- as.data.frame(table(GEC_Data$assocActor1))
assocActor1 <- assocActor1[order(assocActor1$Freq, decreasing = TRUE),]
head(assocActor1, 20)
```

```
assocActor2 <- as.data.frame(table(GEC_Data$assocActor2))
assocActor2 <- assocActor2[order(assocActor2$Freq, decreasing = TRUE),]
head(assocActor2, 20)
```

These sorted dataframes reveal Biafra conflicts aren't the most prolific among Nigeria's armed conflicts from 2000-2021. The only appearance of Biafra in the top 20 is MASSOB (Movement for the Actualization of a Sovereign State of Biafra) coming in at number 16 in the assocActor1 list. Although the Biafra separtists are more of a fringe group, their large online presence still makes them interesting to look into.

In [ ]:
```
# Subset to focus on group of interest

interest_group <- GEC_Data[(GEC_Data$actor1 %like% "Biafra" |
                              GEC_Data$actor2 %like% "Biafra" |
                              GEC_Data$assocActor1 %like% "Biafra" |
                              GEC_Data$assocActor2 %like% "Biafra"),]
```

In [ ]:
```
# Separate out the interest group data from the whole data set

non_interest <- sqldf("SELECT * FROM GEC_Data EXCEPT SELECT * FROM interest_group")
```

In [ ]:
```
# Compare non interest group data to interest group

# Fatalities
print("All Fatalities:")
summary(non_interest$fatalities)
print("Interest Group Fatalities:")
summary(interest_group$fatalities)

# Year
print("All Year:")
summary(non_interest$year)
print("Interest Group Year:")
summary(interest_group$year)

# Type of incident
type_all <- as.data.frame(table(non_interest$eventType))
type_all <- type_all[order(type_all$Freq, decreasing = TRUE),]
print("Type of incidents all:")
type_all

type_interest <- as.data.frame(table(interest_group$eventType))
type_interest <- type_interest[order(type_interest$Freq, decreasing = TRUE),]
print("Type of incidents interest group:")
type_interest

# News source
source_all <- as.data.frame(table(non_interest$source))
source_all <- source_all[order(source_all$Freq, decreasing = TRUE),]
print("Top 20 news sources all:")
head(source_all, 20)

source_interest <- as.data.frame(table(interest_group$source))
source_interest <- source_interest[order(source_interest$Freq, decreasing = TRUE),]
print("Top 20 news sources interest group:")
head(source_interest, 20)

# source scale (regional, national, international, etc.)
global_all <- as.data.frame(table(non_interest$sourceScale))
global_all <- global_all[order(global_all$Freq, decreasing = TRUE),]
print("News coverage scale all:")
global_all
```

```
global_interest <- as.data.frame(table(interest_group$sourceScale))
global_interest <- global_interest[order(global_interest$Freq, decreasing = TRUE),]
print("News coverage scale interest group:")
global_interest

# Sum fatalities
print("Sum of all fatalities")
sum(non_interest$fatalities)
print("Sum of interest group fatalities")
sum(interest_group$fatalities)
```

The comparisons above reveal that Biafran conflicts tend to have lower fatalities than Nigerian conflicts at large, and receive a lot less international attention.

Aggregate data and write .csv files that can later be ingested into Tableau

In [ ]:
```
# Write .csv of just interest data
write.csv(interest_group, "Biafra data.csv")

# write.csv of non-interest group
write.csv(non_interest, "non Biafra data.csv")

# Aggregate the Biafra data by date

interest_group$records <- 1

by_date <- aggregate(list(interest_group$records, interest_group$fatalities), by = list
                     sum)


# renames for clarity
names(by_date)[1] <- "Armed Conflict Data Event Date"
names(by_date)[2] <- "Armed Conflict Data Records"
names(by_date)[3] <- "Armed Conflict Data Fatalities"

by_date$`Armed Conflict Data Event Date` <- as.character(by_date$`Armed Conflict Data E
```

Python Twitter Scrapping Code

At this point, we switch from R to Python to scrap twitter for tweets that meet our search criteria

For this search, we want English tweets that contain the term "Biafra", were tweeted between 2000-2020 (the end date did not work for some reason, and it returned tweets through 2021), and were retweeted at least 100 times. This should ensure the tweet had some traction, and power to influence attendance at the conflict events identified in the GEC-IQ data.

In [ ]:
```
# Scrape Twitter Biafra

!pip3 install --user --upgrade git+https://github.com/twintproject/twint.git@origin/mas
!pip install nest_asyncio

# Avoid runtime errors
import nest_asyncio
nest_asyncio.apply()

import sys
# if pip3 install above gives warning that the twint.exe file is not on the path, must
# uncomment and update filepath below with where it was saved
#sys.path.append('C:/Users/d-jasmine.boatner/AppData/Roaming/Python/Python38/Scripts')
```

```
import twint
import pandas
import re
import os

# Configure
c = twint.Config()
c.Lang = "en"
c.Hide_output = True
c.Search = ['Biafra']
c.Min_retweets = 100
c.Since = '2000-01-01'
c.until= '2020-12-31'
c.Pandas = True

# Run
twint.run.Search(c)

Tweets_df = twint.storage.panda.Tweets_df

# Print to .csv
Tweets_df.to_csv("biafra tweets.csv")
```

About 8000 tweets were returned, and at this point we go back to R to manipulate this data and merge with previous dataset.

R code Part 2

In [ ]:

```
# Aggregate Tweets

library(readxl)
interest_tweets <- read_excel("biafra tweets.xlsx")

# Make a new column that is sum of retweets, comments, and likes on a tweet

interest_tweets$twitter_engagement <- interest_tweets$nretweets + interest_tweets$nlike

# Drop time stamps from dates, and make format match armed conflict format
interest_tweets$date <- format(as.POSIXct(interest_tweets$date, format = '%m/%d/%Y %H:%

# Count for number of records
interest_tweets$records <- 1

# Aggregate by date

by_date_twitter <- aggregate(list(interest_tweets$records, interest_tweets$twitter_enga
                             by = list(interest_tweets$date),
                         sum)

# renames for clarity
names(by_date_twitter)[1] <- "Twitter Date"
names(by_date_twitter)[2] <- "Twitter Records"
names(by_date_twitter)[3] <- "Twitter Engagement"
```

To avoid having to blend in Tableau, which has given me some issues in the past, I merged the GEC-IQ data and the Twitter data in R

In [ ]:

```
# Merge with armed conflict data by date
```

```
by_date_merge <- merge(by_date_twitter, by_date, by.x = "Twitter Date", by.y = "Armed C
                       all.x = TRUE, all.y = TRUE)

# write.csv
write.csv(by_date_merge, "biafra twitter state data.csv", na= "")
```

In addition to loking at the twitter data by date, look at the data by username. There are a few usernames who dominate the social media landscape and are the ones spreading a lot of the disinformation.

In [ ]:

```
# Aggregate by username
by_username <- aggregate(list(interest_tweets$records, interest_tweets$twitter_engageme
                         by = list(interest_tweets$username, interest_tweets$name),
                         sum)

# renames for clarity
names(by_username)[1] <- "Twitter Username"
names(by_username)[2] <- "Name"
names(by_username)[3] <- "Twitter Records"
names(by_username)[4] <- "Twitter Engagement"

# write.csv
write.csv(by_username, "biafra by username.csv", na="")

write.csv(interest_tweets, "biafra tweets tableau.csv", na="")
```

End Code

The ouput from the scripts above was ingested into Tableau to create the final products of the project, interactive Tableau dashboards.