



Predicting Terror Attacks in 2020 & Beyond

Sarah Qin, Valeriya Antonova

June 2019

Problem Statement

- “Terrorism has become a festering wound. It is an enemy of humanity.” - Atal Bihari Vajpayee
- “What separates us from the animals, what separates us from the chaos, is our ability to mourn people we’ve never met.”-David Levithan, Love is Higher Law
- The U.S. Code of Federal Regulations defines terrorism as ‘the unlawful use of force and violence against persons or property to intimidate or coerce a government, the civilian population, or any segment thereof, in furtherance of political or social objectives” (28 C.F.R Section 0.85).
- Since October 2018, some terroristic acts which occurred in the world:
 - 3 synagogue shootings in the US
 - 2 mosque shootings in New Zealand
 - Bombing in Sri Lanka on Easter

Data

The Global Terrorism Database (GTD) is an open-source database including information on terrorist events. It is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism.

Time Frame: 1969 through 2017 (missing 1993)

Scope: International

Number of Cases: > 180,000

Has the following features:

Year, Month, Day, City, Country, Region, Latitude, Longitude, Attack Type, Killed, Wounded, Target, Summary, Target Type, Weapon Type, Motive, etc.

“Information in the GTD is drawn entirely from publicly available, open-source materials. These include electronic news archives, existing data sets, secondary source materials such as books and journals, and legal documents.”

“GTD does include failed attacks, it does not include foiled or failed plots, the distinction being that the attack must actually be attempted to qualify for inclusion in the database. Likewise, the GTD does include attacks in which violence is threatened as a means of coercion, but does not include threats to attack where no action is taken.”

Assumptions

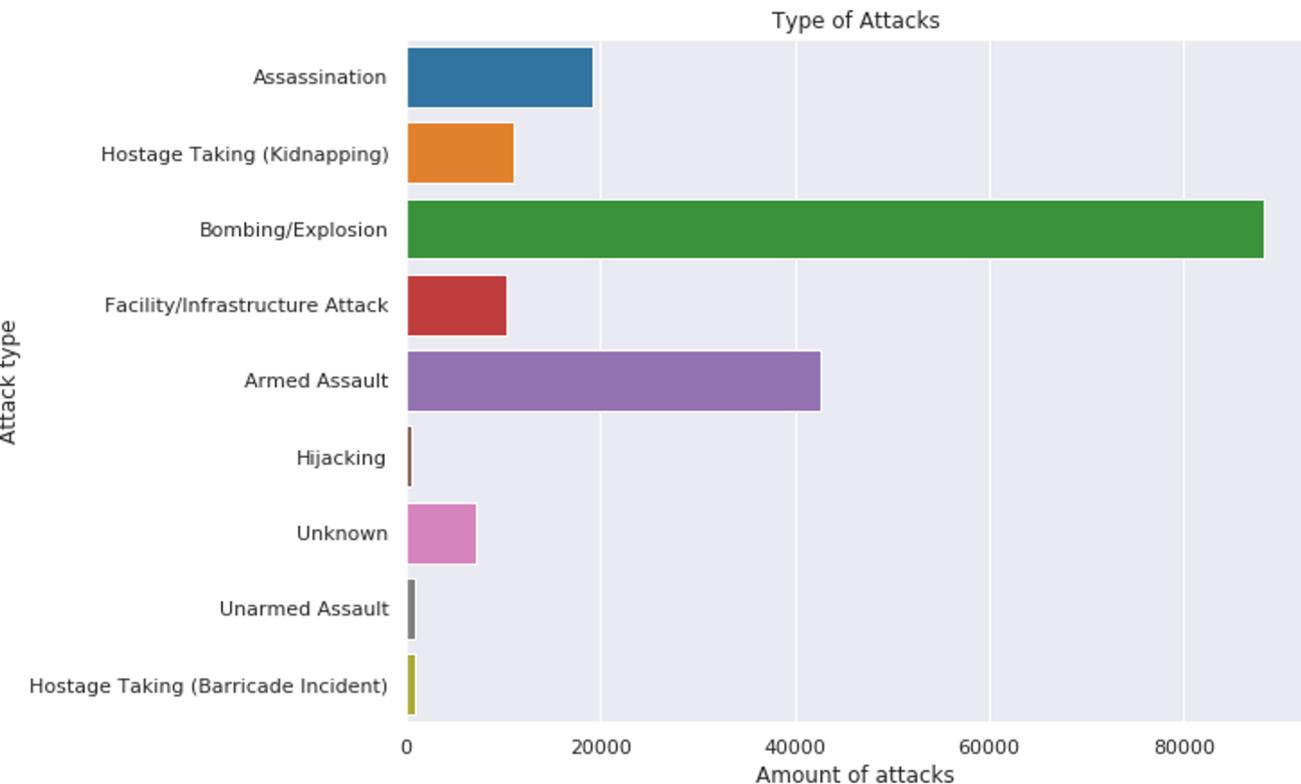
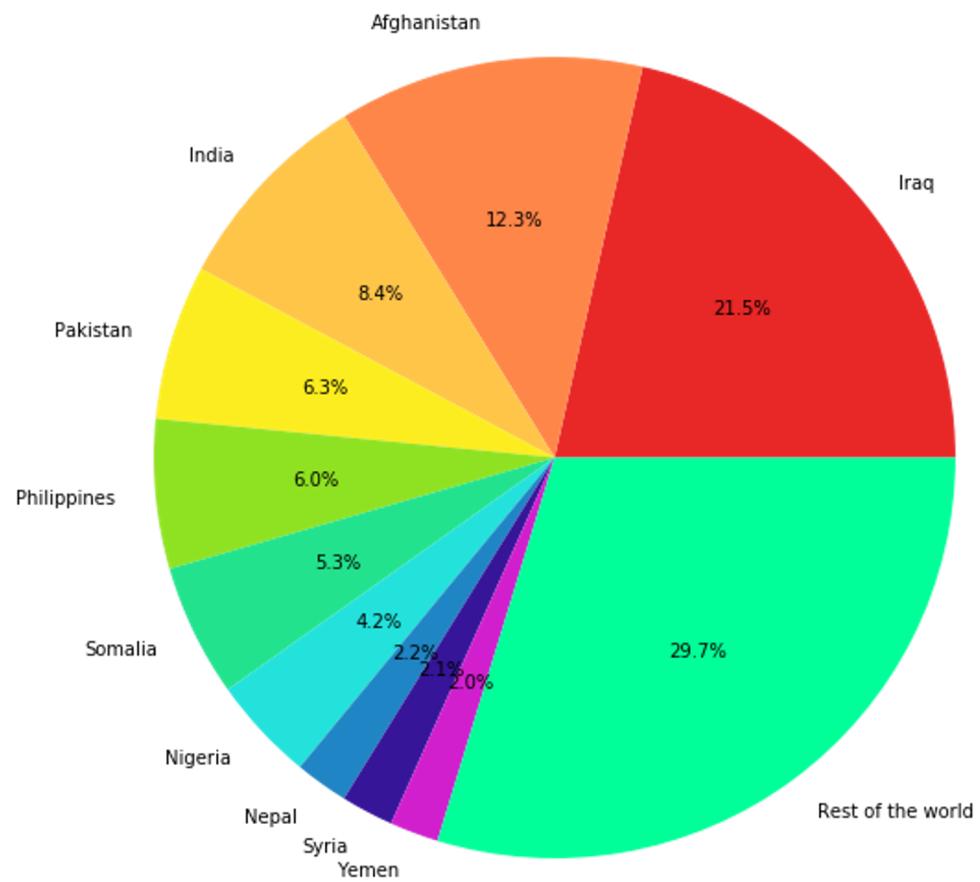
1. Data collection is objective
2. Data from 1994 and on is more relevant to the current and future
3. Definition of terrorism is inclusive
4. Transparency of coding system and criterias



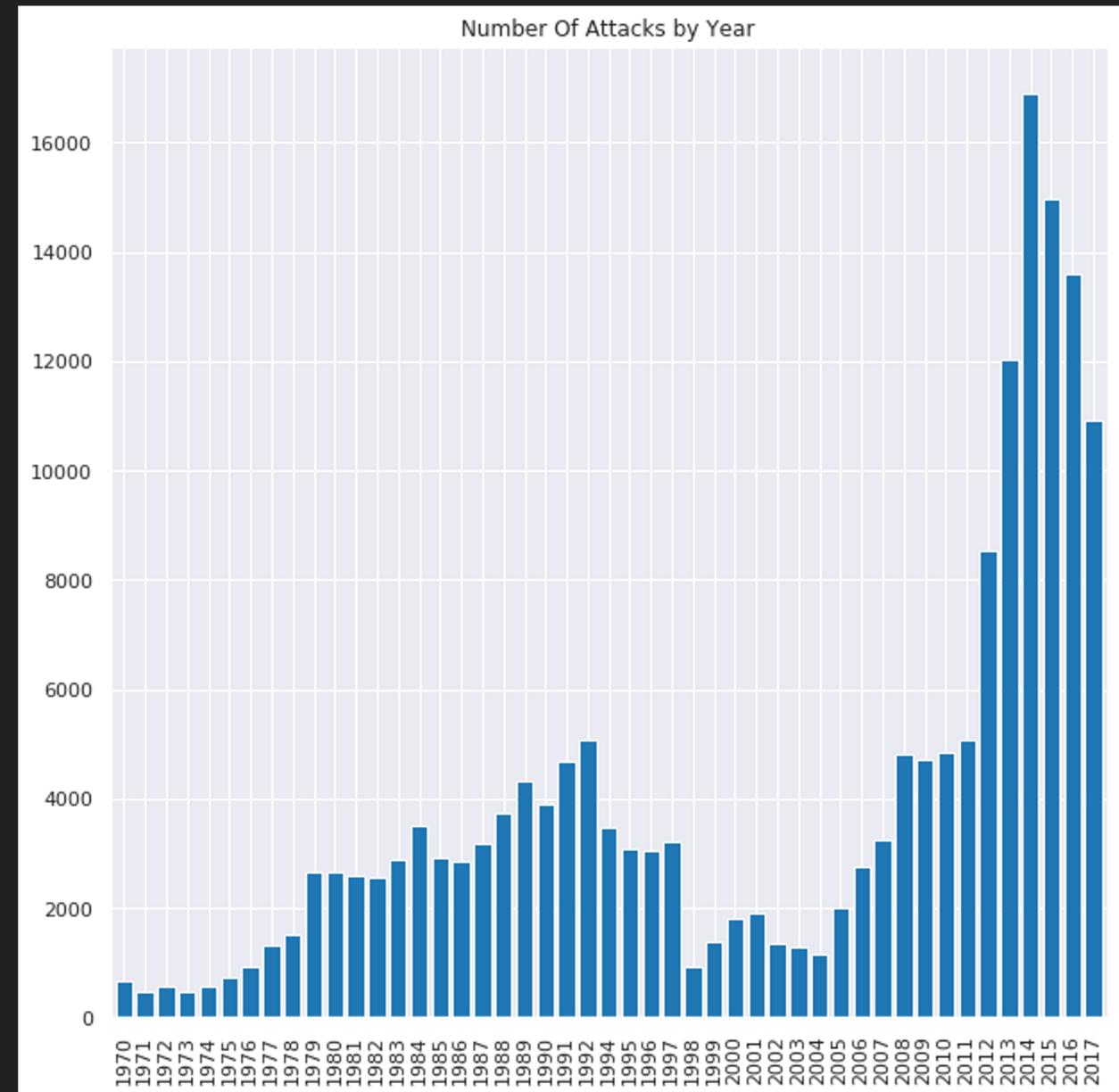
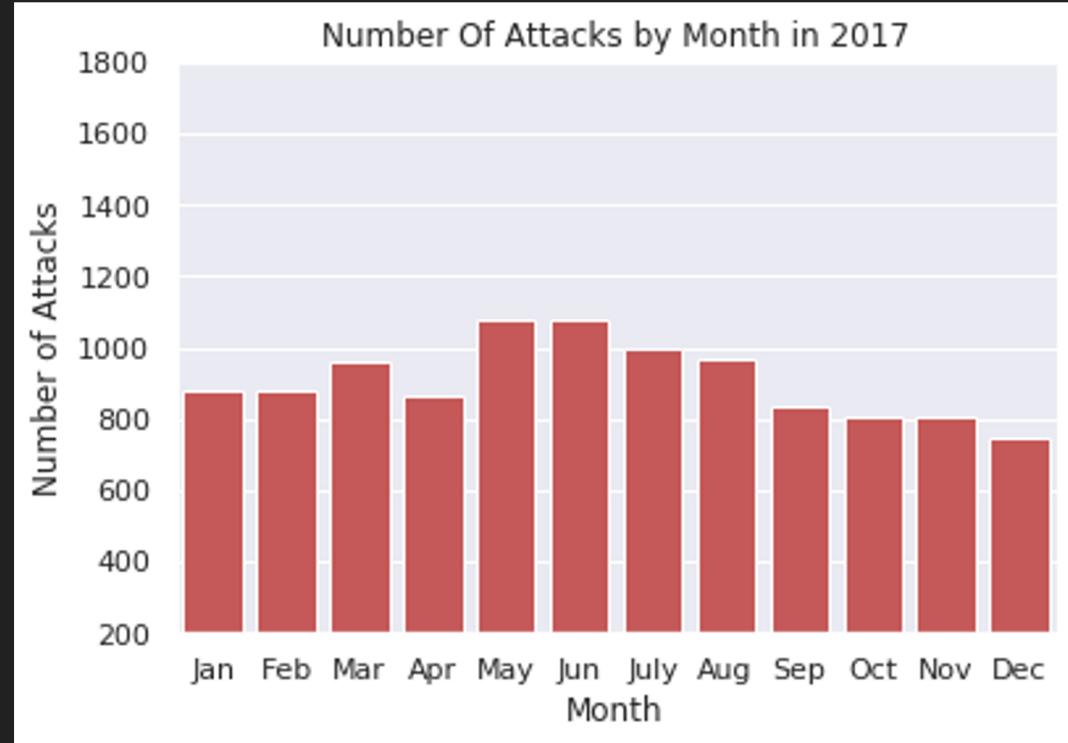
Exploratory Data Analysis

EDA (International)

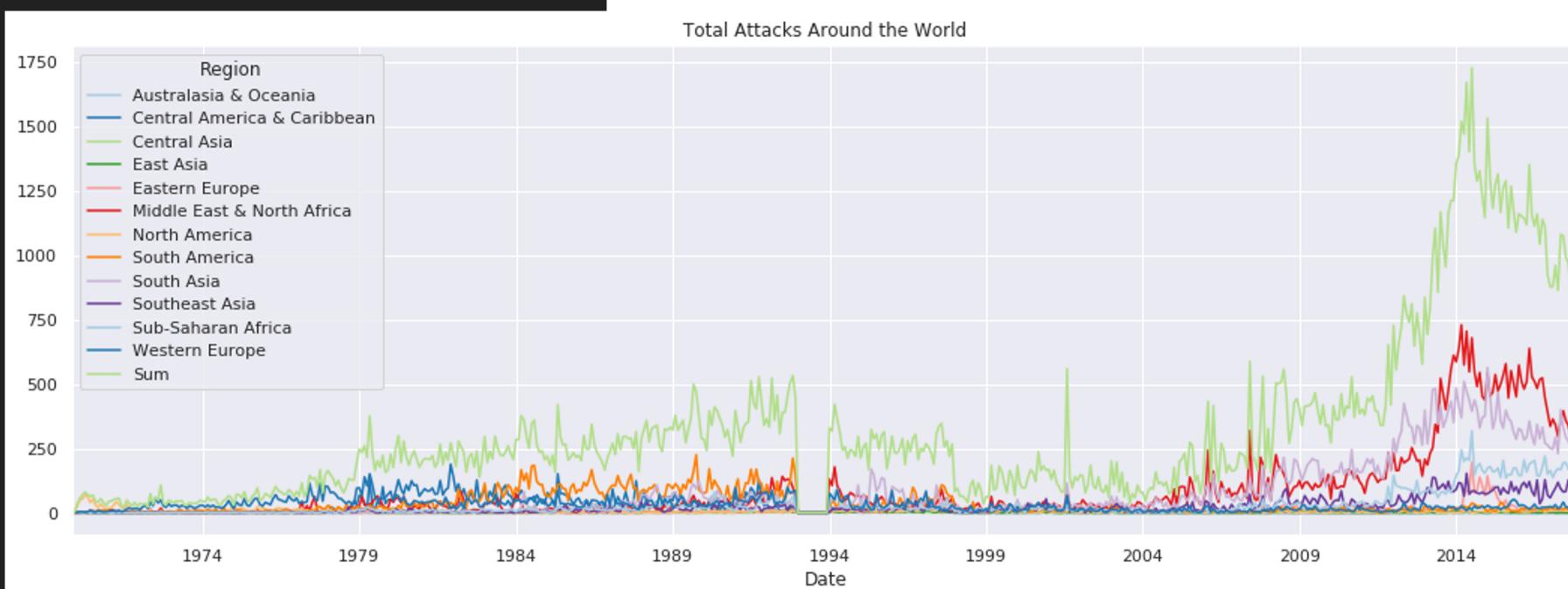
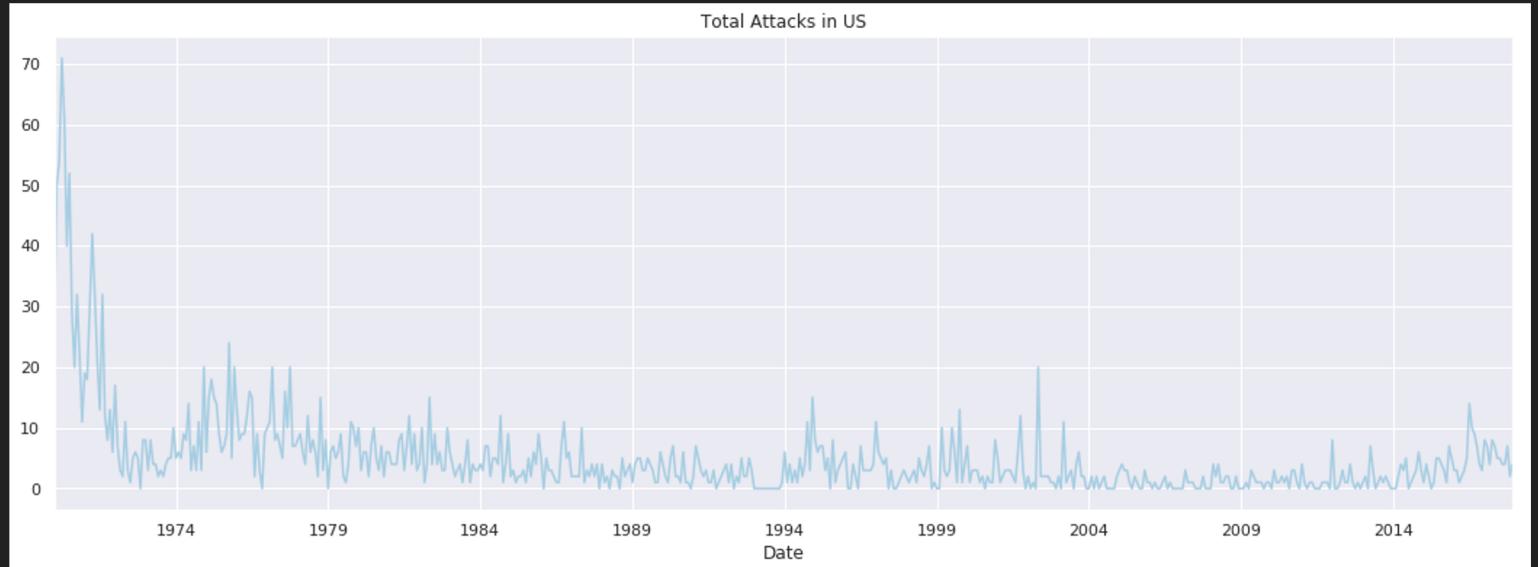
Spread of Attacks Cross Countries in 2017



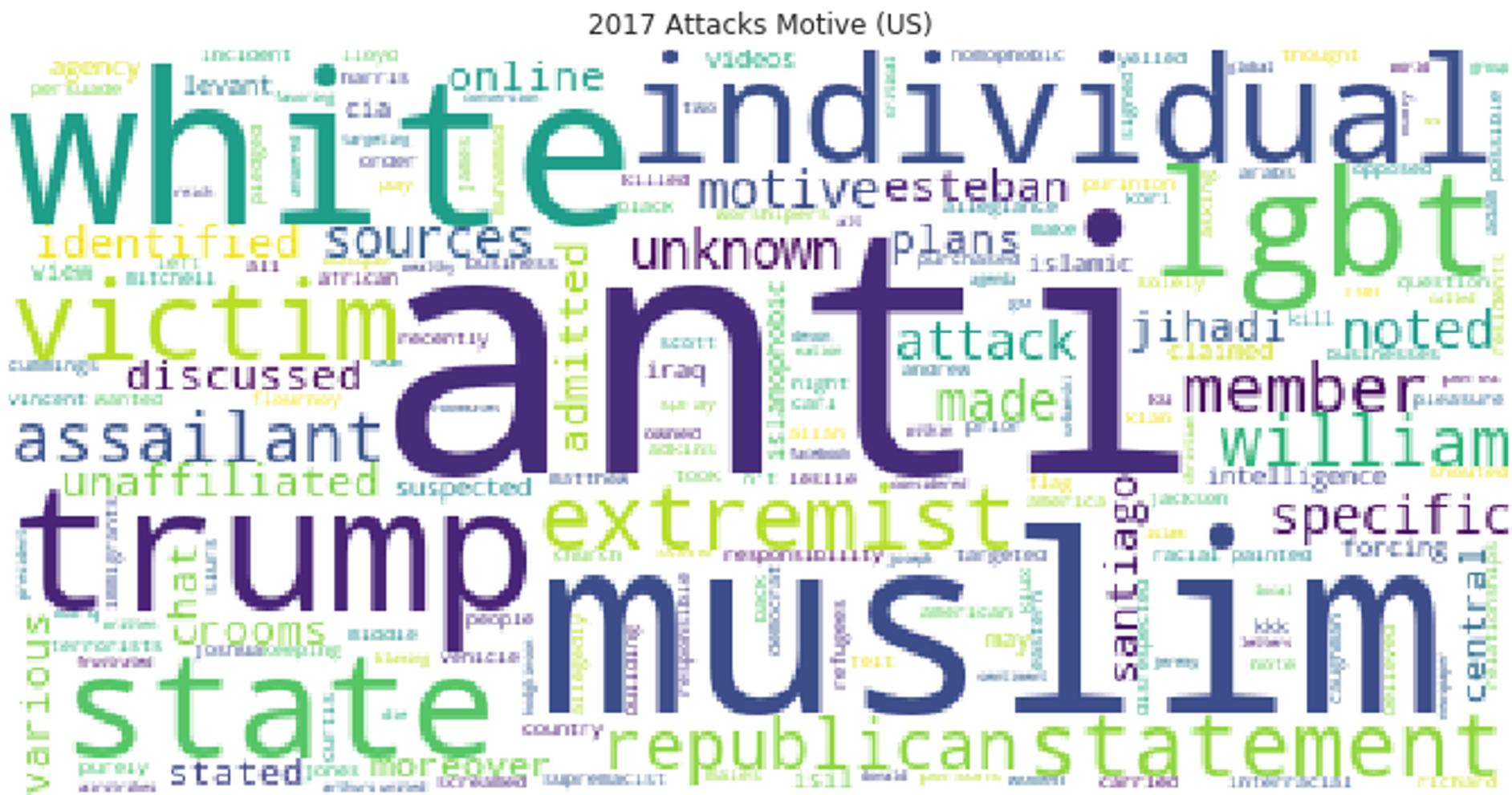
EDA (International)



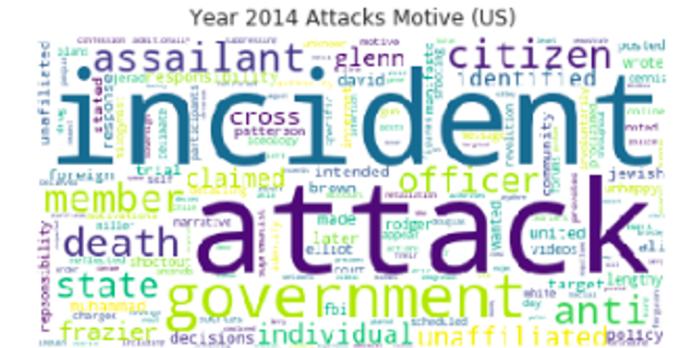
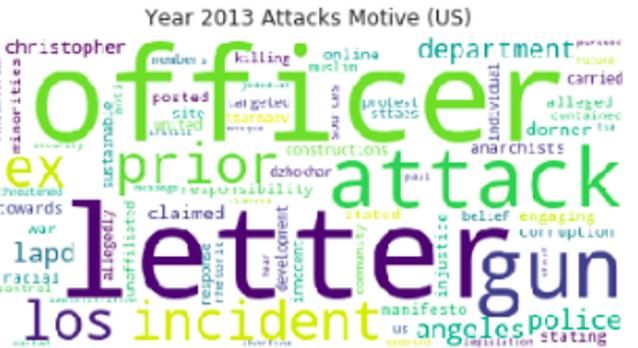
EDA (US & World)



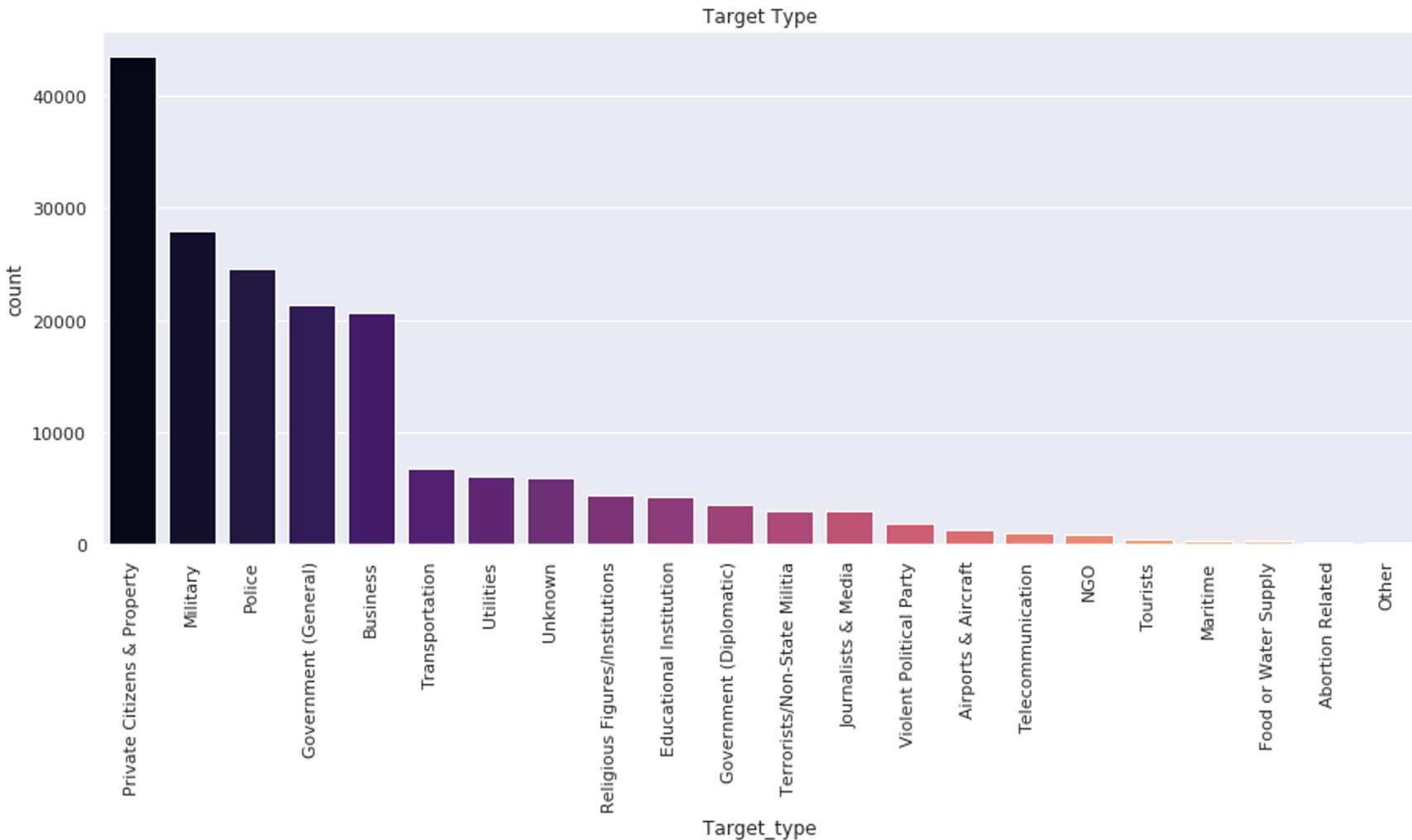
EDA (US)



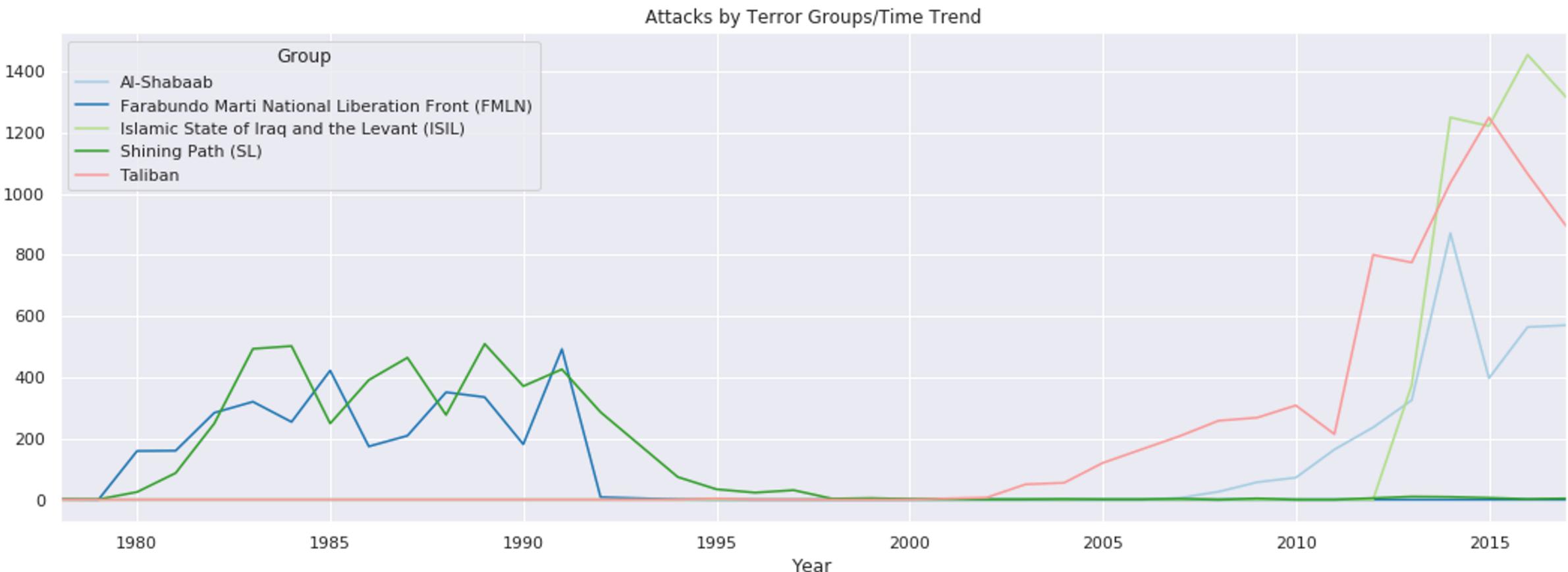
EDA (US)



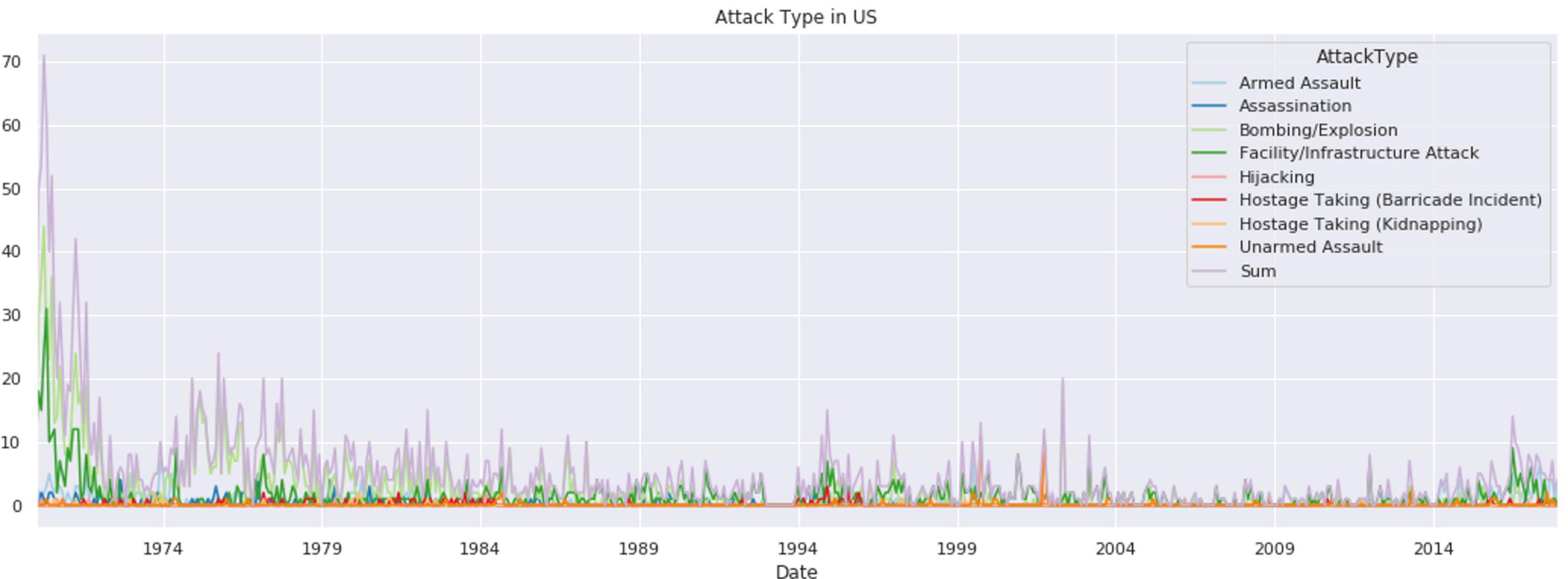
EDA (International)



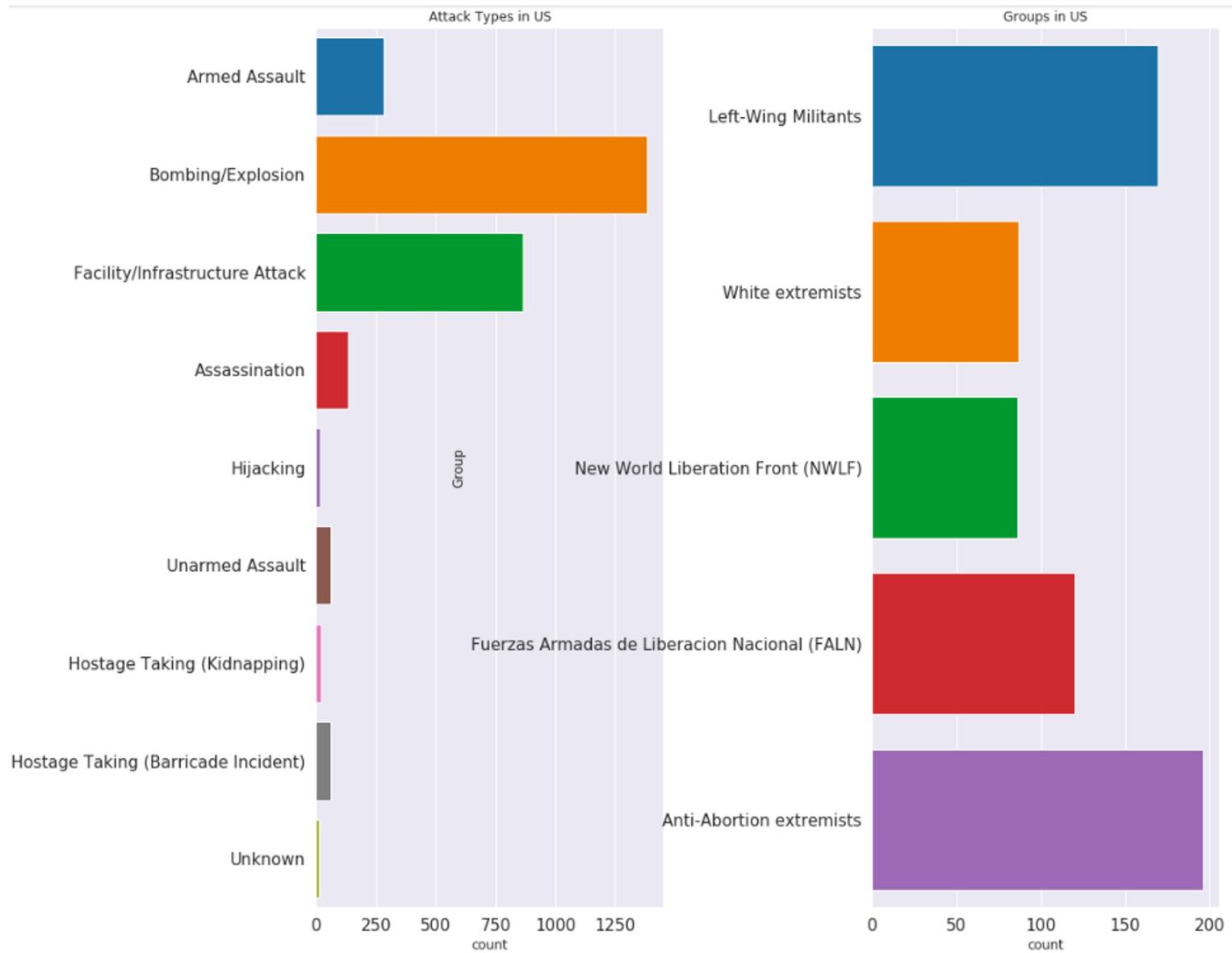
EDA (International)



EDA (US)



EDA (US)



EDA (US Bomb)

Goal:

Predict future number of BOMBING attacks in US in 2020?

Process Outline

Considering data that are more relevant to today's global political and economic situation, we used the monthly bombing data between 2007 and 2017 for time series analysis:

- (1) build time-series models to fit the bombing cases between January 2007 and October 2014;
- (2) make forecasts between November 2014 and December 2016, fit the residuals between the predicted and observed data;
- (3) validate the new model by forecasting the bombing in 2017
- (4) compare to forecasted 2017's bombing with actual 2017's bombing cases

EDA (US Bomb)

	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>
AR	3355.802	3442.896	3355.802	23.82223	23.82223
MA	1902.134	2023.076	1902.134	13.12163	13.12163
ARIMA	-9927.500	10406.981	9927.500	-70.76023	70.76023
TBATS	-3489.970	3557.332	3489.970	-24.73696	24.73696

4 rows

```
```{r}
Selection of the most accurate function
index <- c("AR" = min(abs(AR)),
 "MA" = min(abs(MA)),
 "ARIMA" = min(abs(ARIMA)),
 "TBATS" = min(abs(TBATS)))

#comparison of best models
as.data.frame(index) %>% rownames_to_column(var = "method") %>% arrange(index)

best_model <- names(index[index == min(index)]) # MA
best_model

...```

```

data.frame  
4 x 2

R Console

[1] "MA"

Time series models experimented here include:

autoregressive-moving average (ARMA) model (AR & MA)  
Autoregressive Integrated Moving Average (ARIMA)  
model  
TBATS model (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, Seasonality)

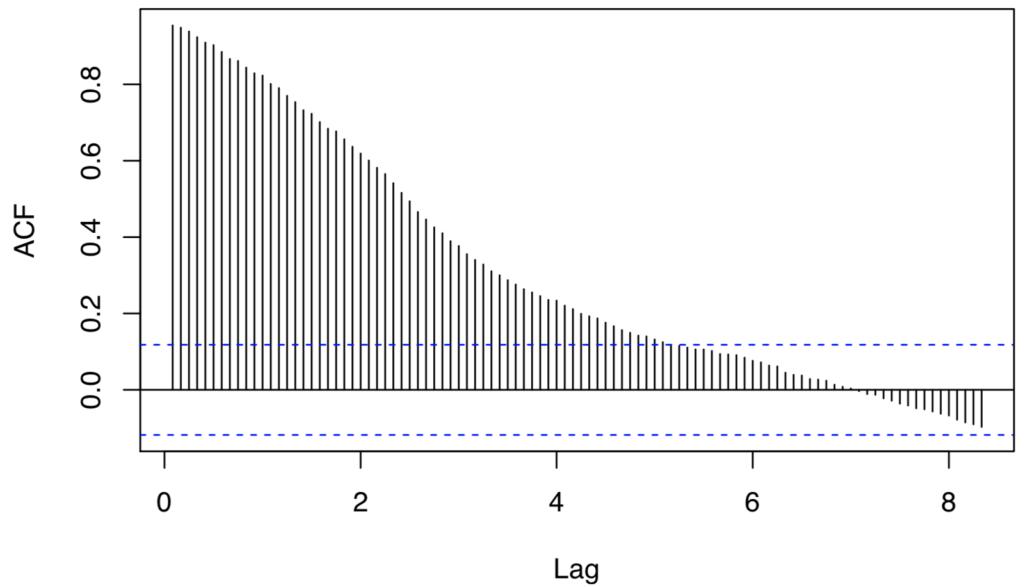


# Modeling

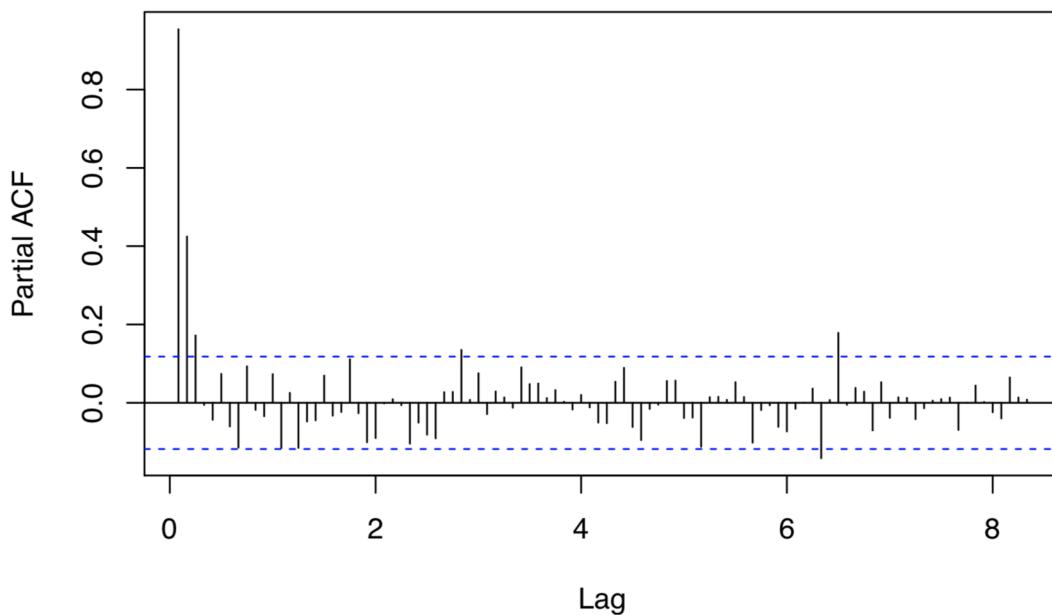
# Modeling

- Time-series used:
  - Count of terrorist attacks by month by world regions
  - Count of terrorist attacks by month world total
- 2 data sets were created:
  - One included data from 1994 - 2017
  - Second included data from November 1969 - December 2017 with missing value imputation
    - Imputation used: outlier analysis and imputing suggested values for total attacks
    - Missing value interpolation using stine: used a loop to replace in data grouped by world regions
- 3 Models were attempted with varying parameters
  - Arima model
    - arima (3,1,0)(1,0,0)[12] with box cox lambda and drift
    - total attacks starting from 1994 dataset
  - Direct recursive model
    - on top of arima (3,1,0)(1,0,0)[12] with box cox lambda and drift
    - total attacks starting from 1994 dataset
  - Hierarchical Model
    - hts with bottom up method and arima for forecasting method

**Series train**

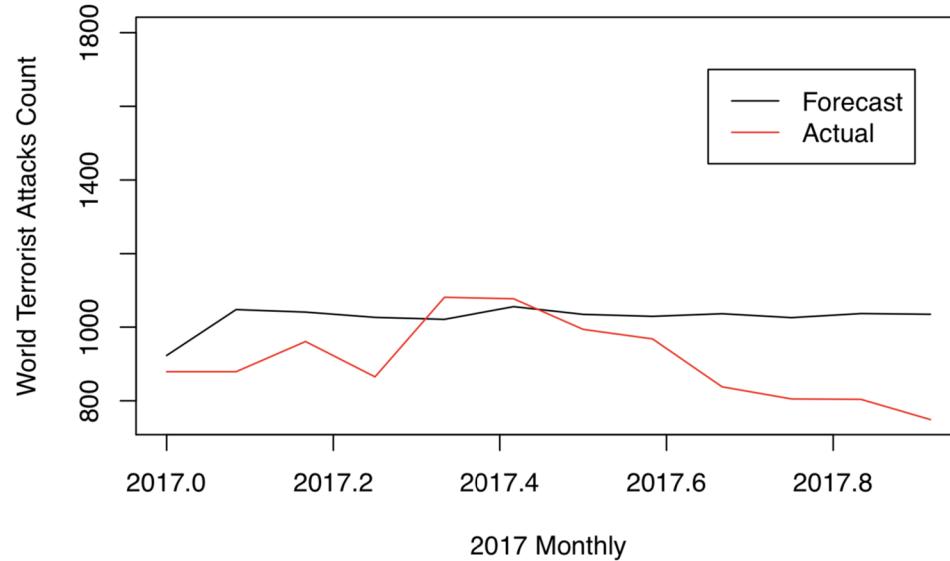


**Series train**

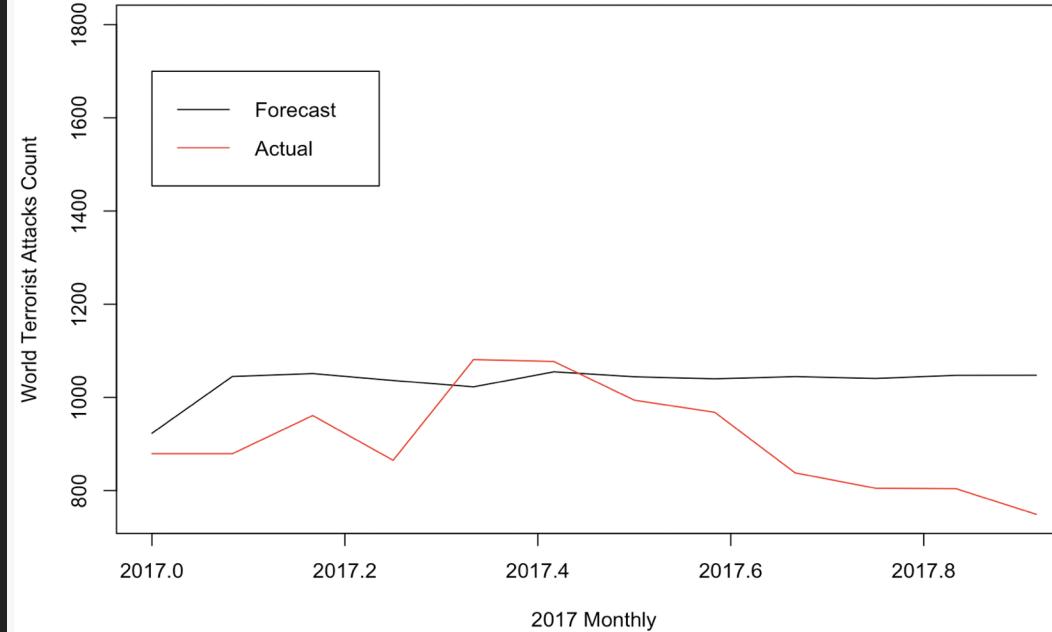


Data TimeFrame	Model	With Drift	With Lamba	AIC	BIC	MAPE	2017 Forecast Accuracy MAPE Score
1994-2016	auto arima (0,1,2)(0,0,1)			3301	3315	26.6	
1994-2016	auto arima (0,1,1)(0,0,1)		X	2103	2114	26.0	
1994-2016	arima (3,1,0)(1,0,0)	X	X	2109	2131	26.2	17.09
1994-2016	direct recursive (3,1,0)(1,0,0)	X	X	2109	2131	26.2	15.6
1969-2016	auto arima (0,1,1)(0,0,2)			6507	6525	23.7	
1969-2016	auto arima (0,1,1)(1,0,0)		X	2774	2792	23.4	
1969-2016	arima (3,1,0)(1,0,0)	X	X	2789	2815	23.4	17.7
1969-2016	arima (0,1,1)(0,1,1)		X	2725	2738	23.01	25.75
1969-2016	direct recursive (3,1,0)(1,0,0)	X	X	2789	2815	23.4	16.42
1994-2016	hts, bu, arima						16.13
1969-2016	hts, bu, arima						15.87

### Direct Recursive Model Forecast with 1994–2016 Time Series

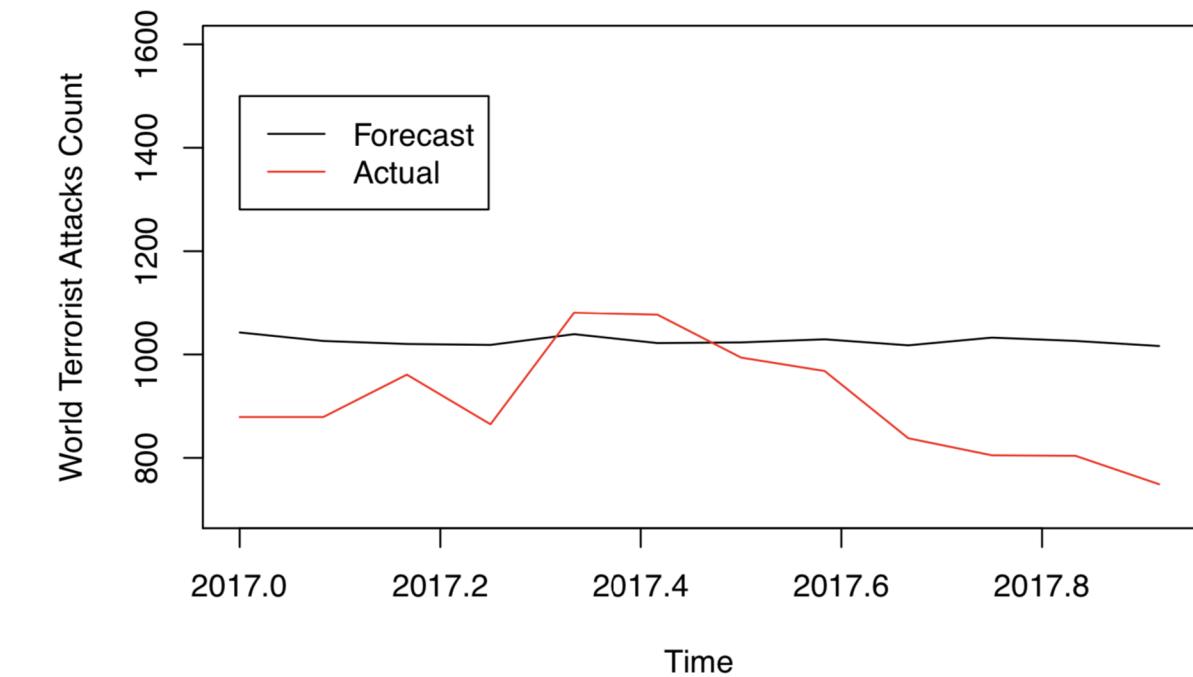


### Direct Recursive Model Forecast from 1969–2016 Time Series



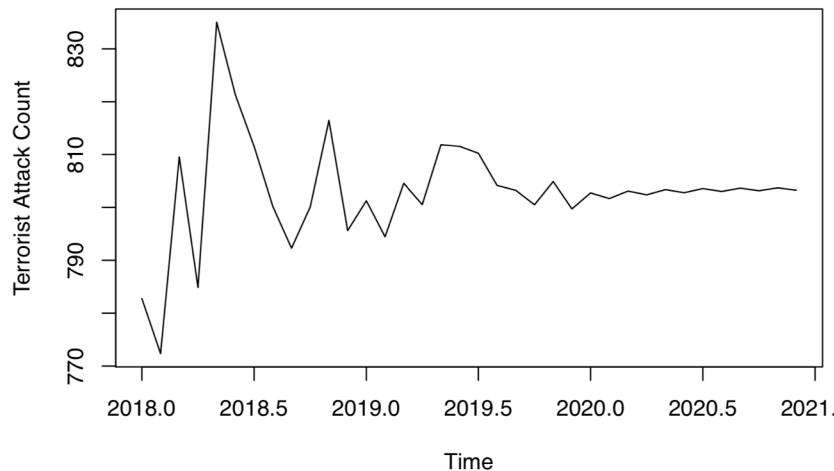
# 3 Top Models Forecast for 2017

### Hierarchical Model Forecast on 1969–2016 Time Series

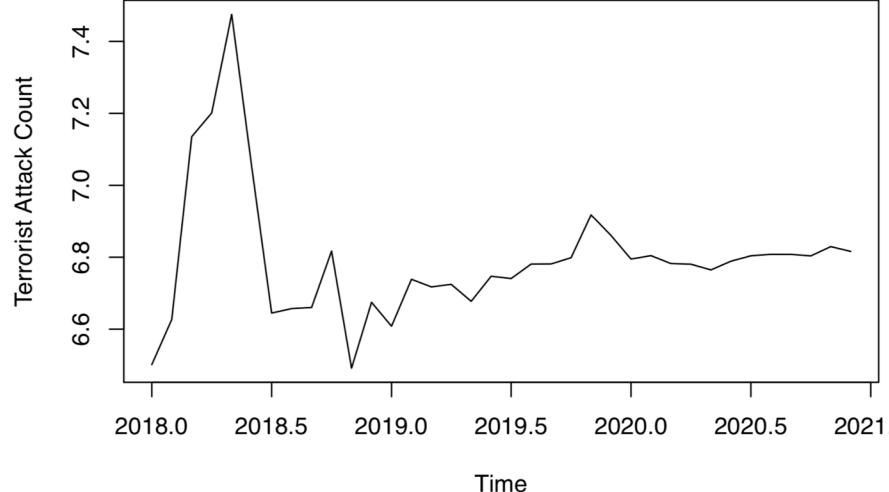


# Forecast for 2018-2020 Hierarchical Model

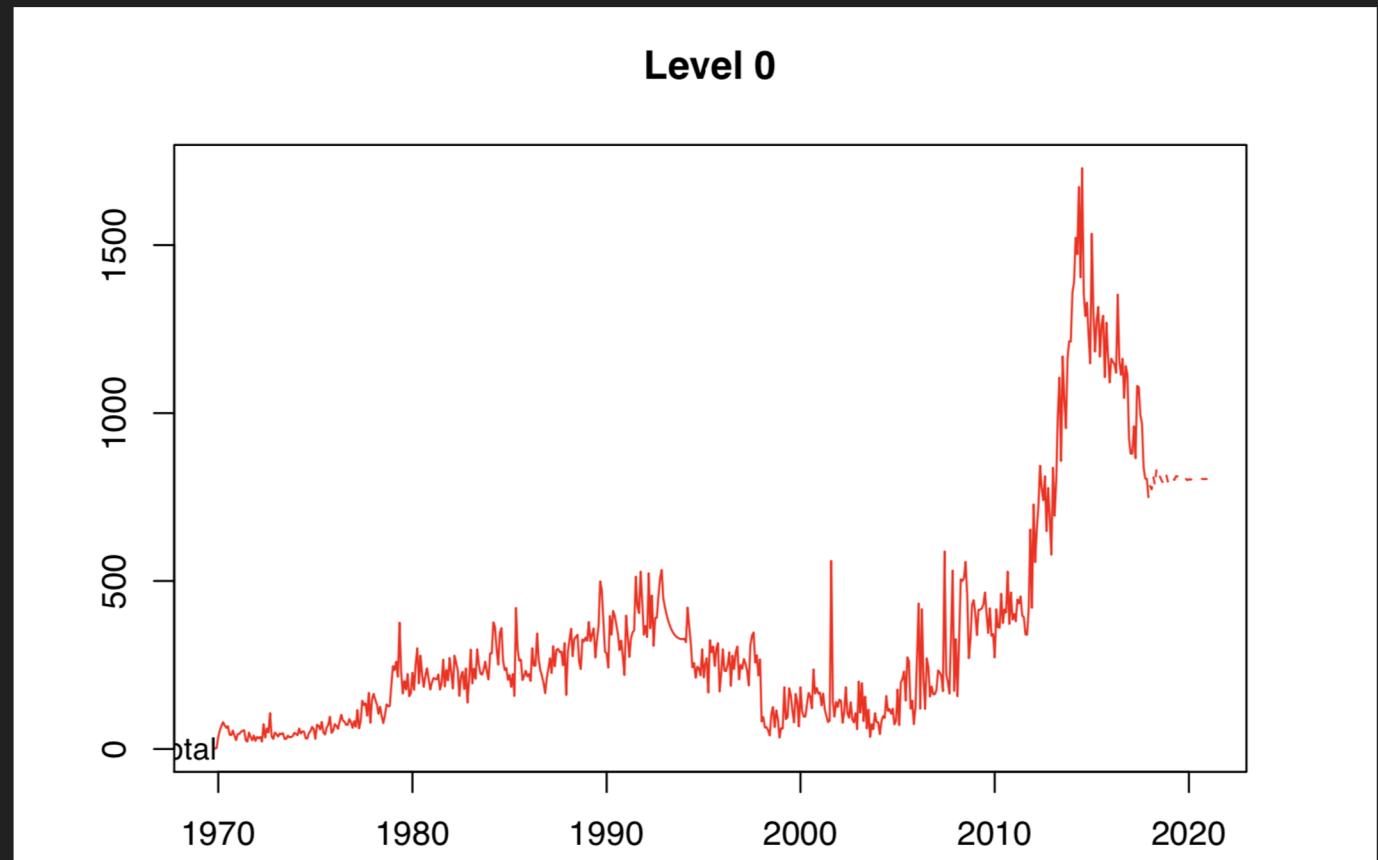
Hierarchical Model Forecast for 2018–2020 World Terrorist Attacks



Hierarchical Model Forecast for 2018–2020 US Terrorist Attacks



Level 0



# Future Work

1. Intervention Analysis (Pulse)
1. TBATS model (time series model for time series exhibiting multiple complex seasonalities)
1. More layers in the hierarchical model to improve accuracy
1. More ways to deal with the missing 1993 data