Global Terrorism

Team 08

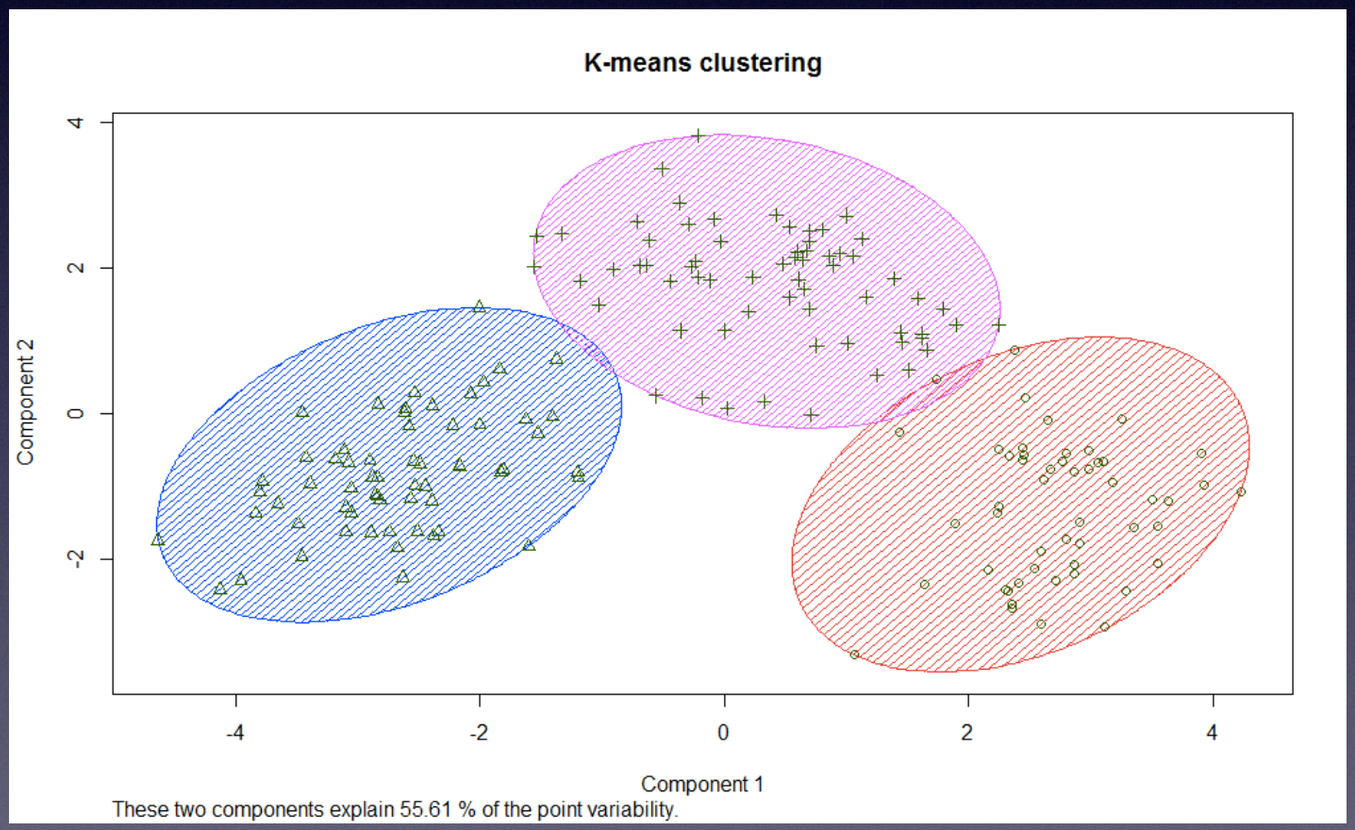
201411098 이덕상
201411160 허건

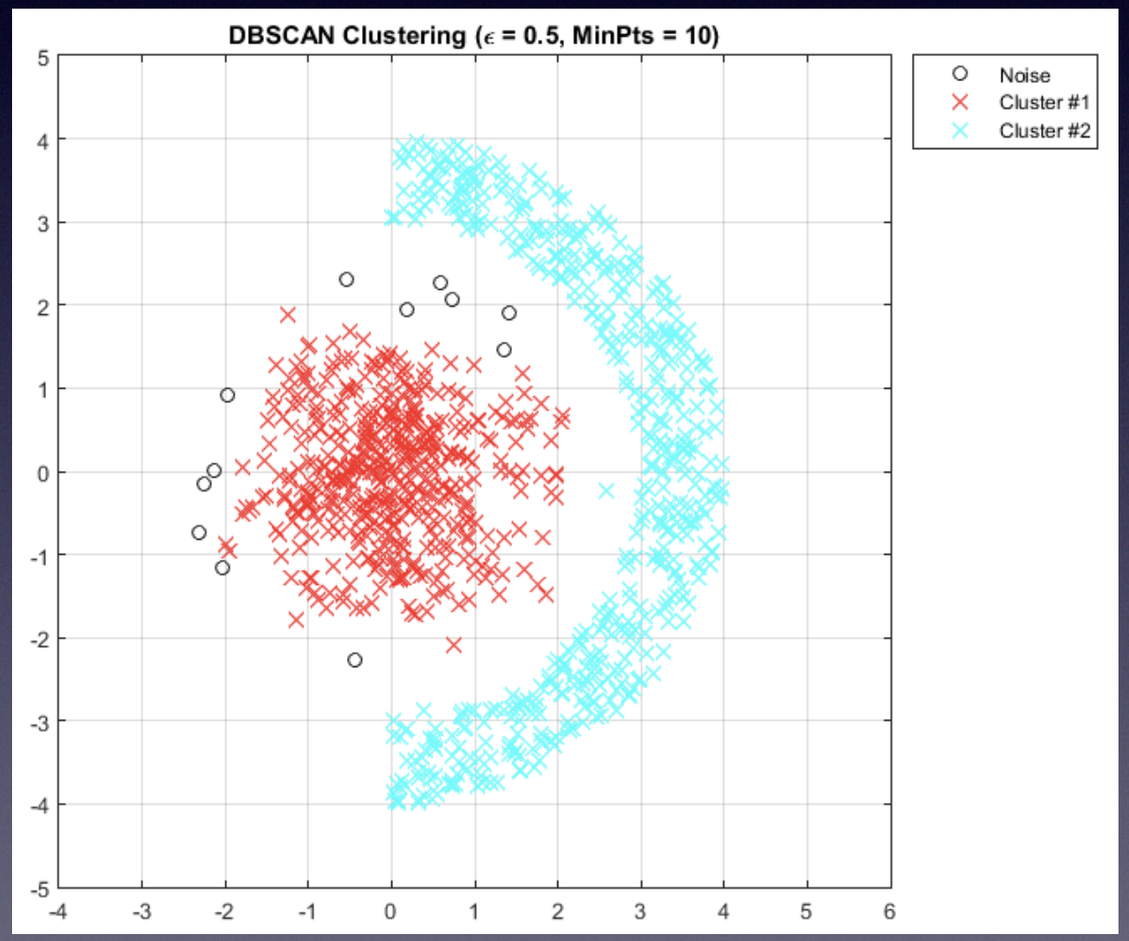
Data sets-Global Terrorism Database

suicide, attacktype1, targtype1, targsubtype1, natlty1,
individual,nperps,nperpcap,weaptype1,weapsubtype1,weaptype2,weapsubtype2,

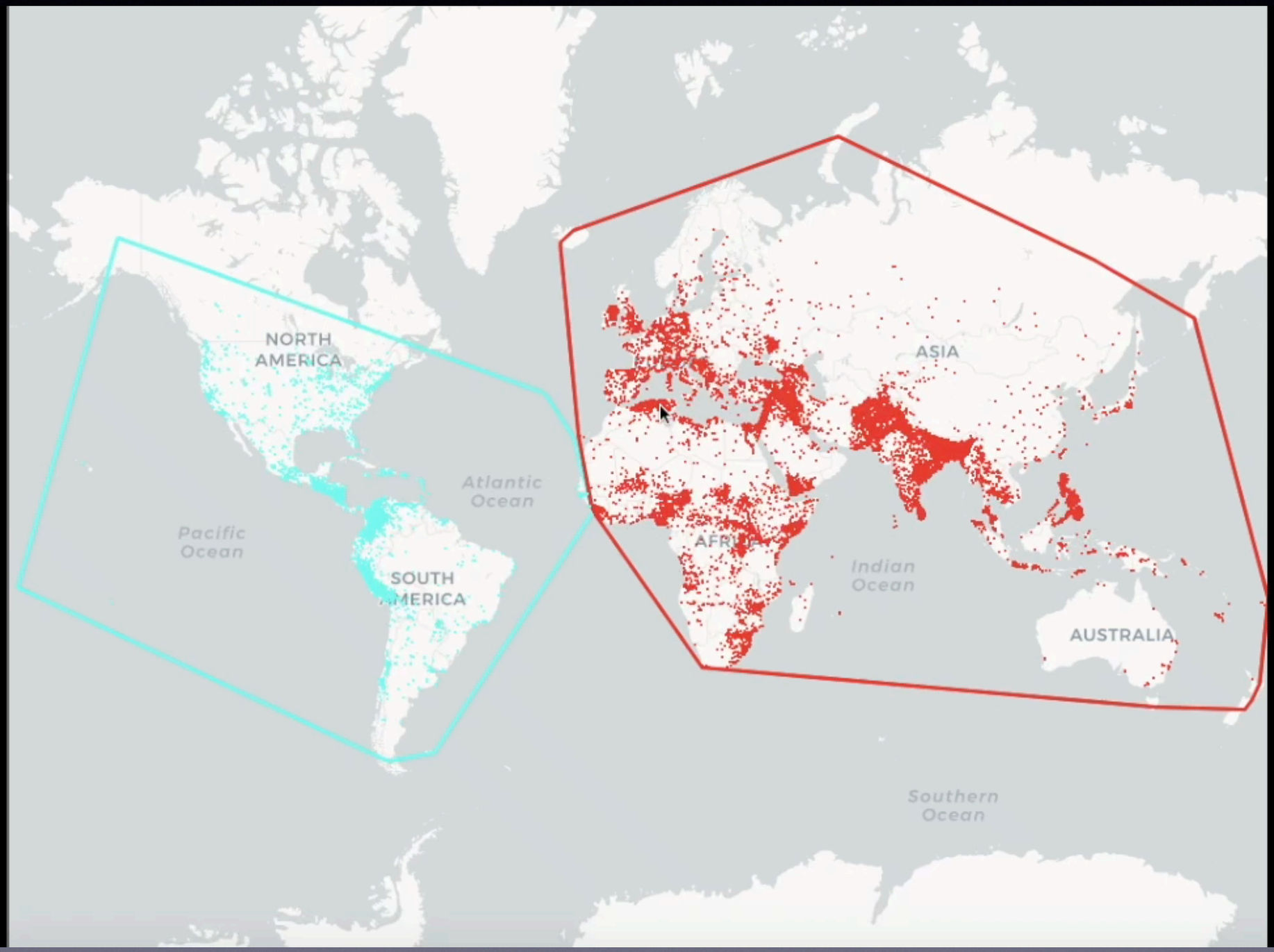
Time period : 1970-2016
Quantity : 170,000+ Terrorist Attacks
Variables : >100 variables
(location, tactics, perpetrators, targets, and outcomes)
Sources : Unclassified media articles

Result 1 : Clustering



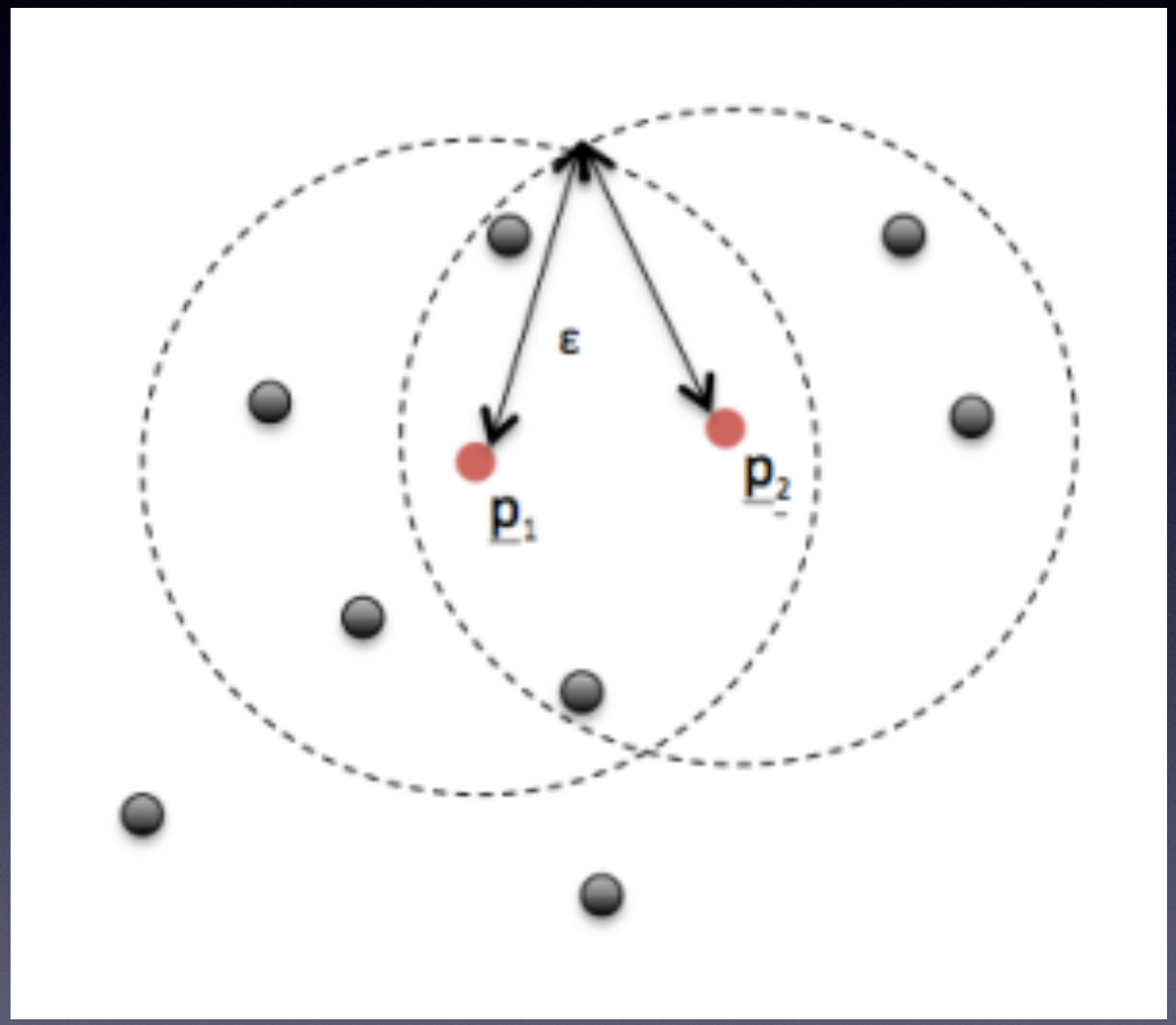


Result 1 : Visualization



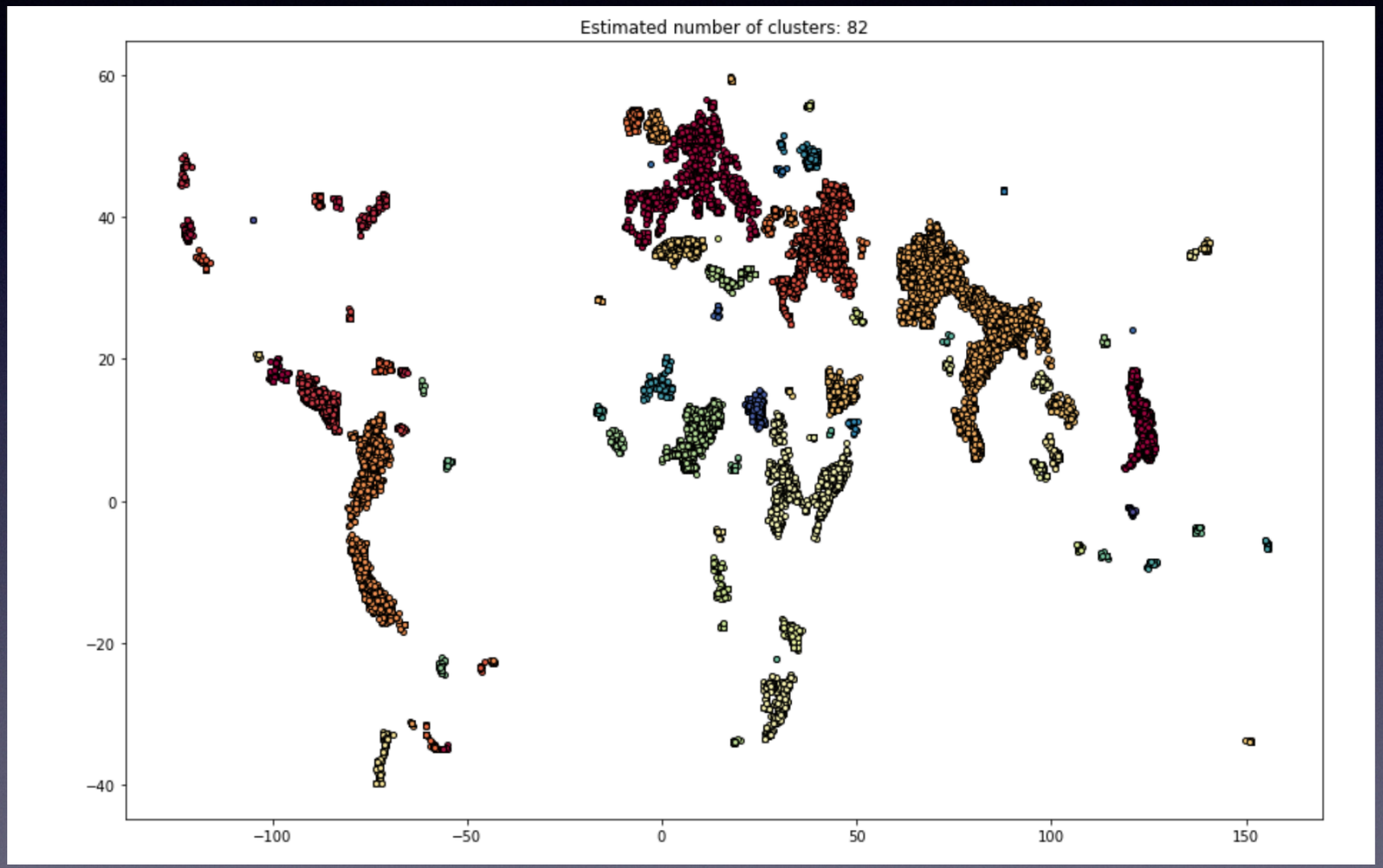
K-means clustering

Result 1 : DBSCAN



Parameter : Epsilon, Minimum Points

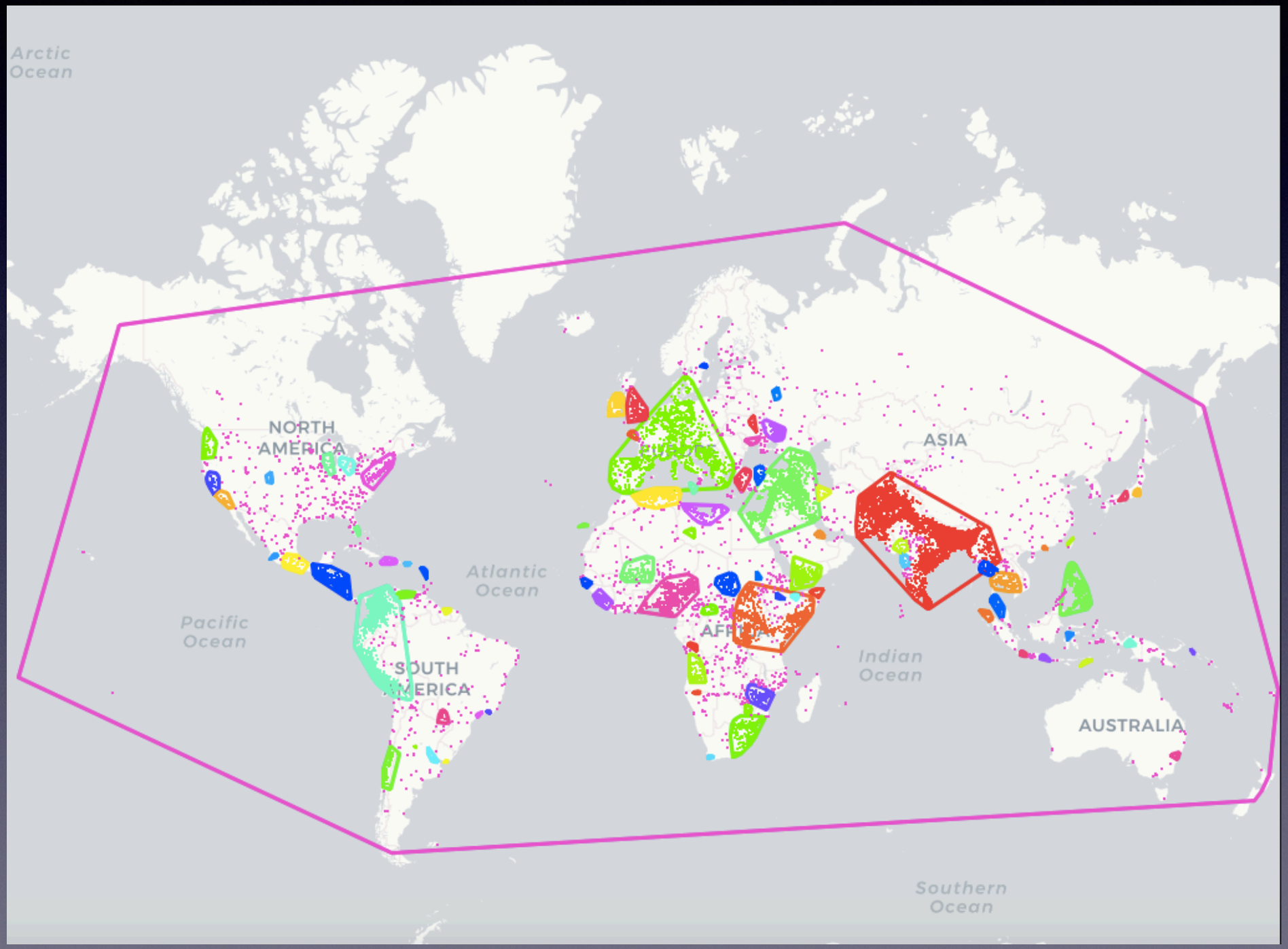
Result 1 : Visualization



Eps : 1.5
Min : 40
Cluster : 82

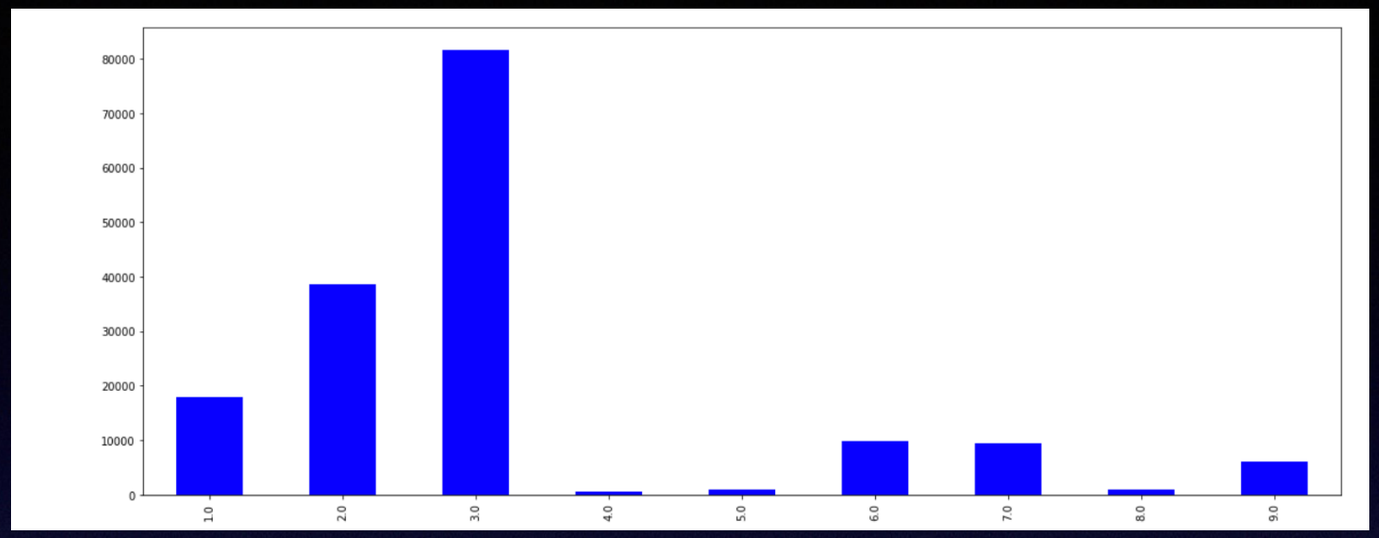
Density base clustering - Simple Plot

Result 1 : Visualization



Density based clustering

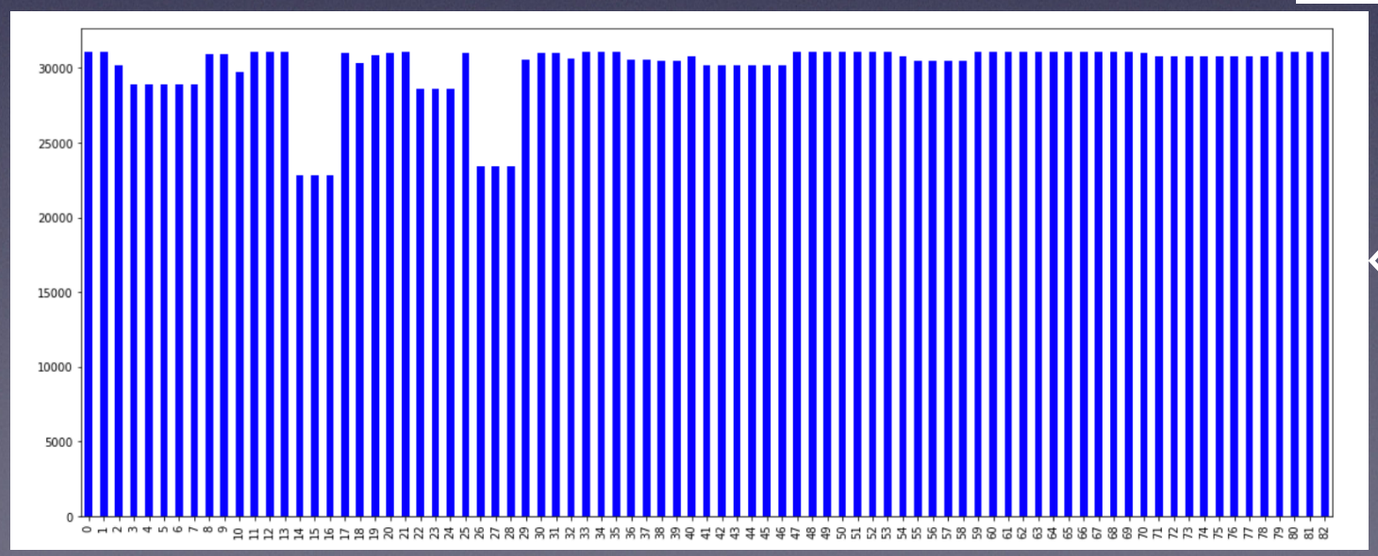
Result 2: 데이터 분석

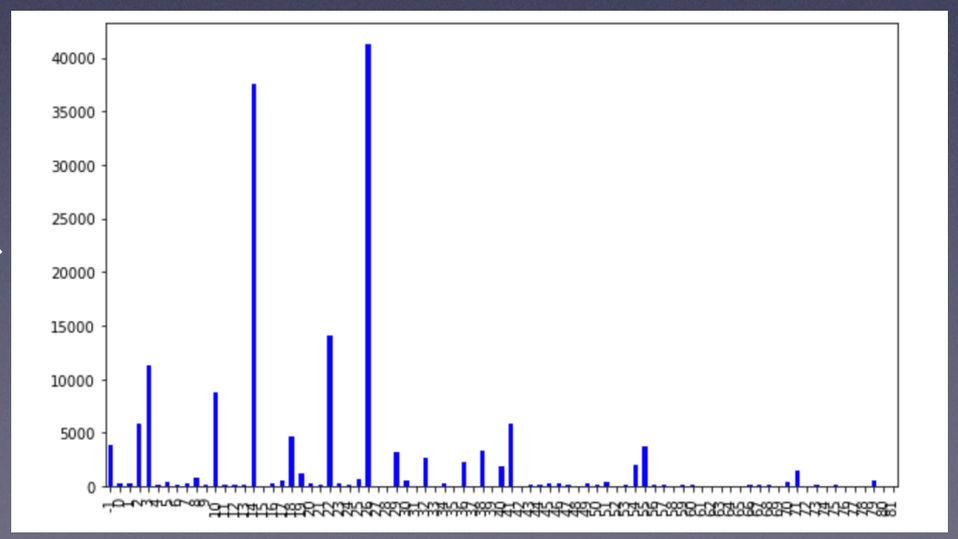


Attack type에 따른 테러발생률 : 3(bombing / explosion), 2(armed assault)

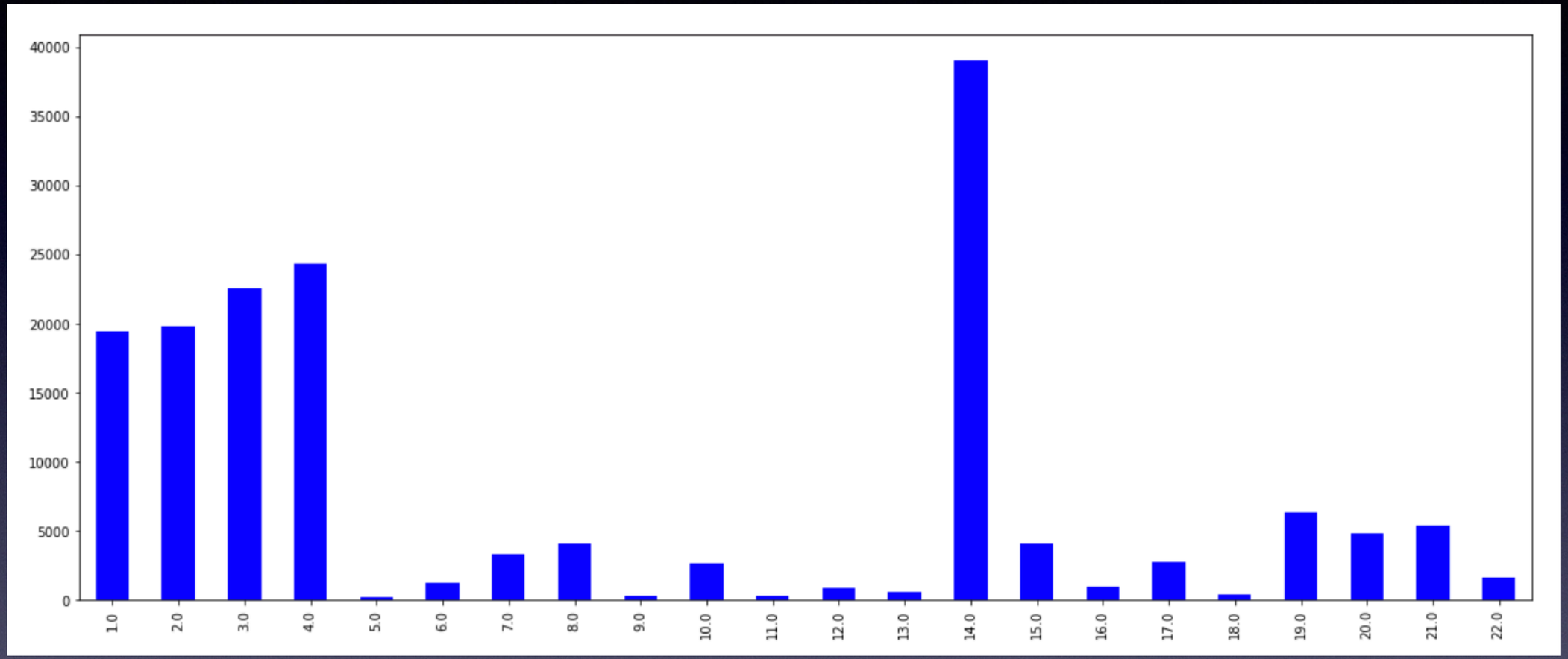
이 분포에 큰 영향을 끼친 cluster가 무엇이 있는가?

$$\frac{1}{N}\sum_{i=0}^{k}(x_i - y_i)^2$$



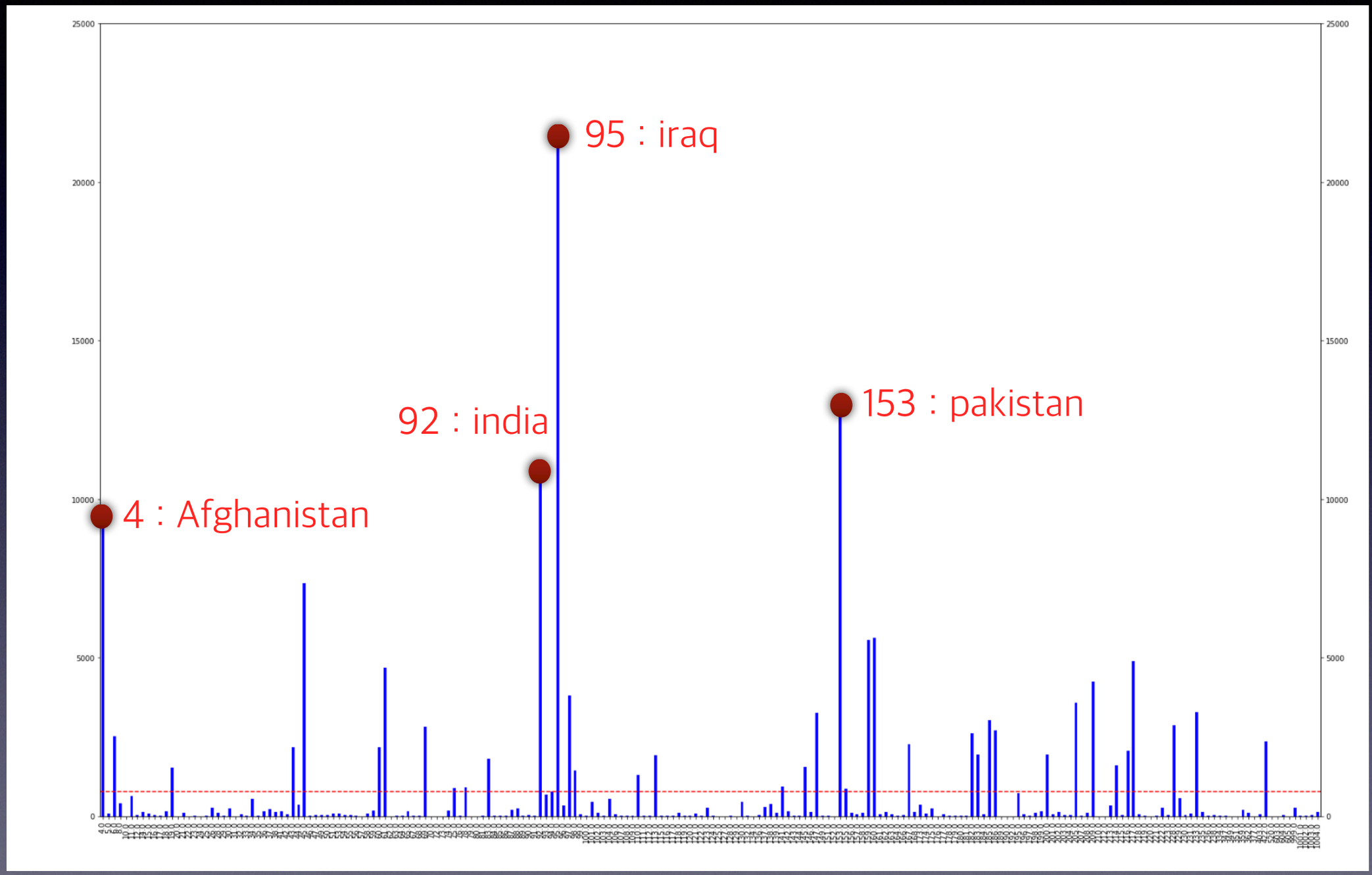


Result 2 : 데이터 분석



* target type에 따른 테러 발생률 : 14(private citizen, property)에서 가장 높음
* 그 외 1(business), 2(general government), 3(police), 4(military)에서 두드러짐

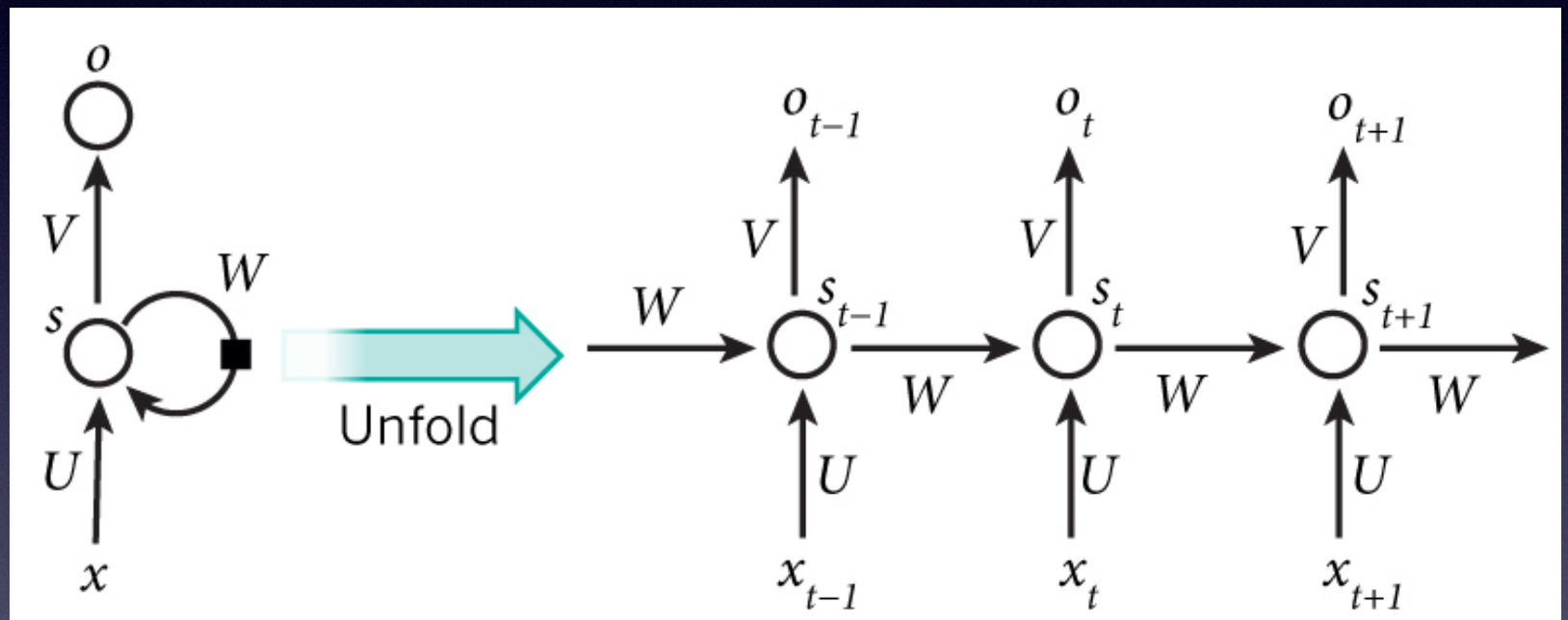
Result 2 : 데이터 분석

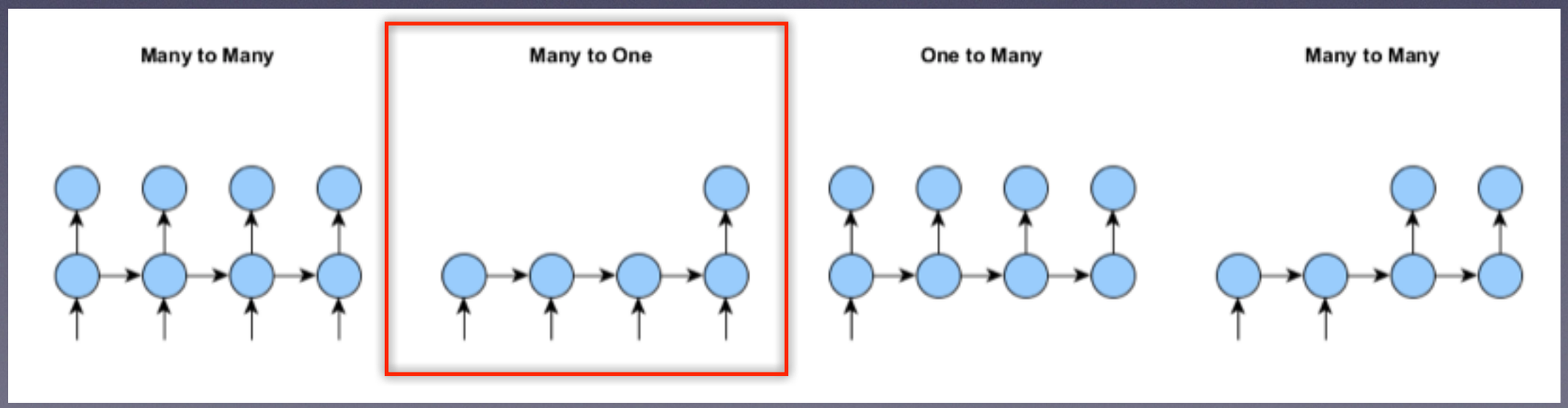


피해를 받은 기관이 소속되는 나라에 따른 테러 발생률

Result 3: RNN을 통한 예측

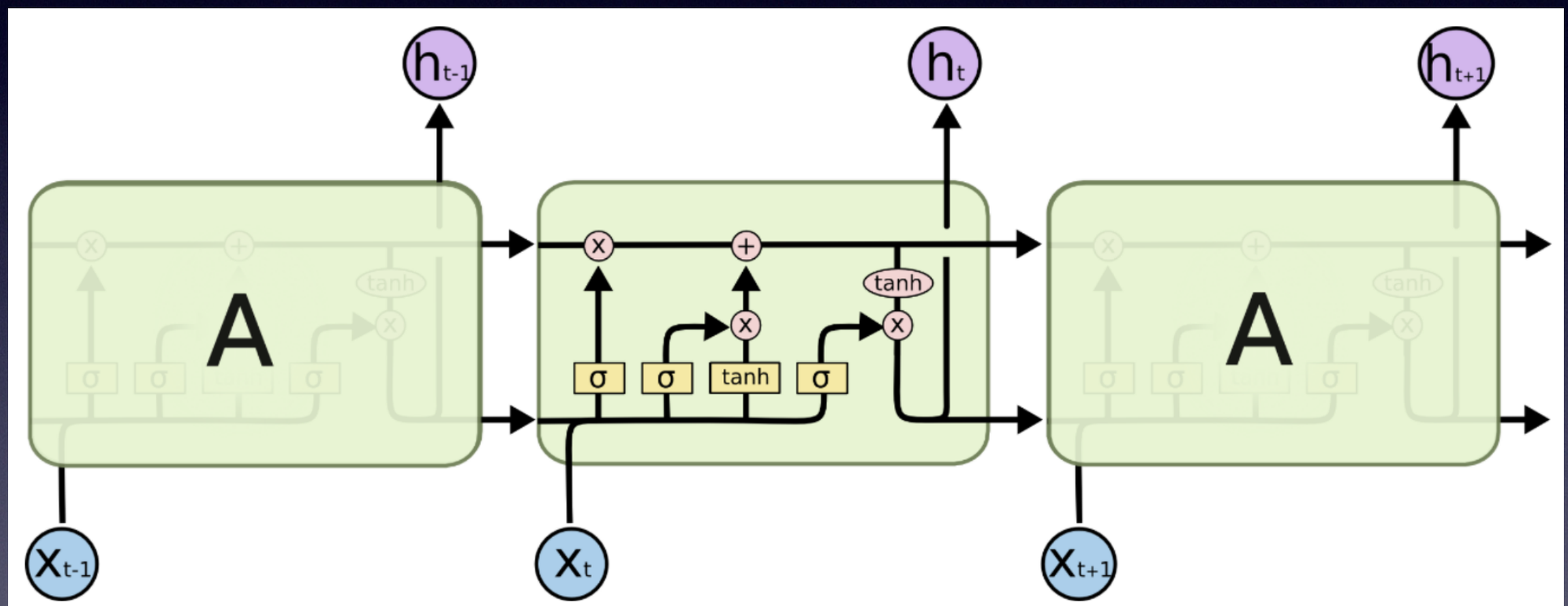
RNN(recurrent neural network)





Result 3: RNN을 통한 예측

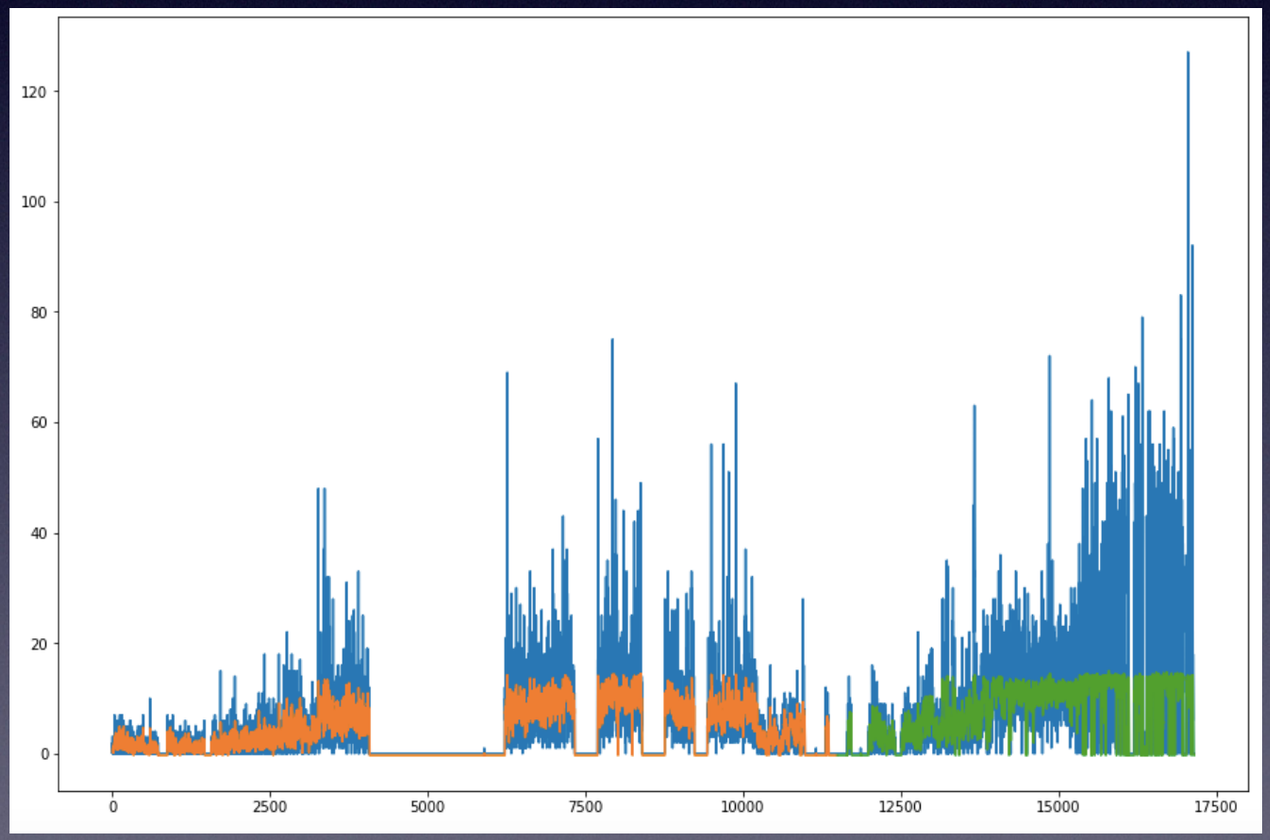
RNN – LSTM (Long Short Term Memory)



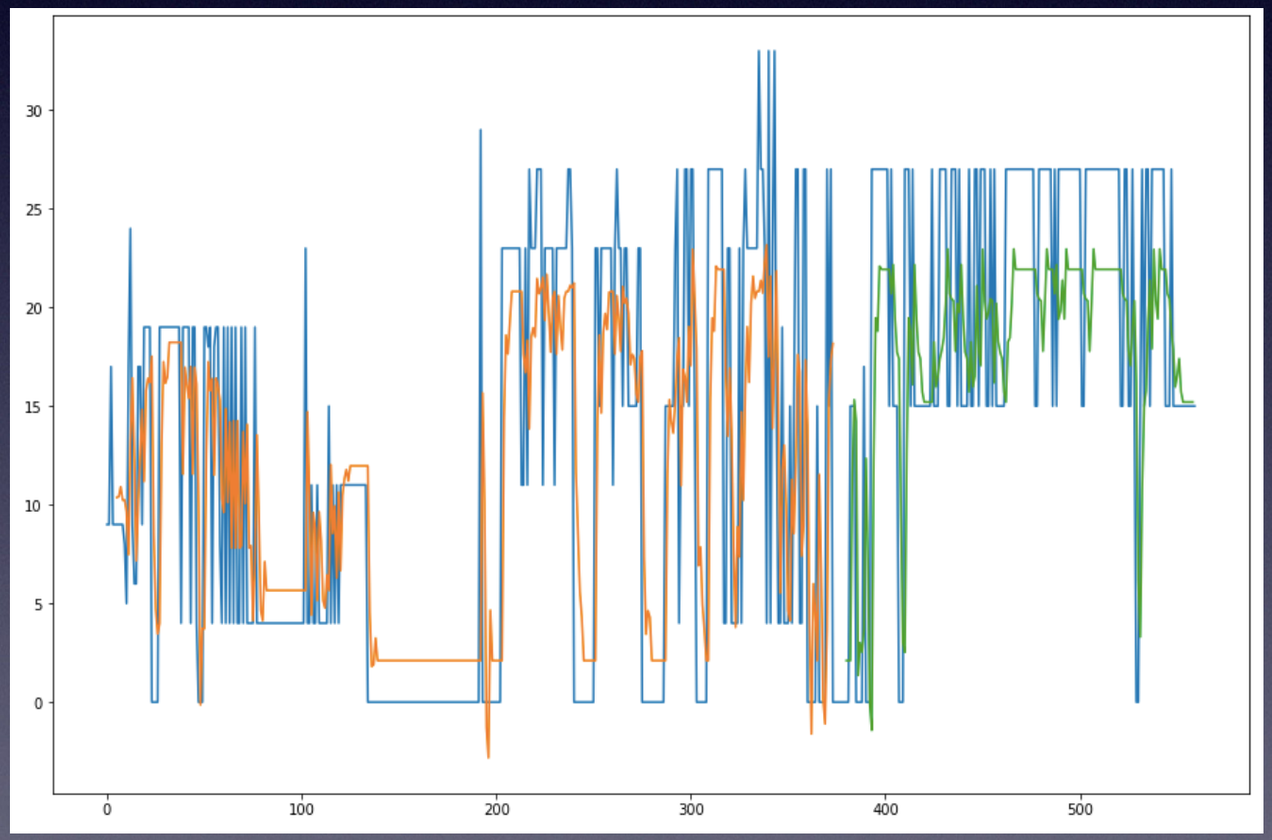
The repeating module in an LSTM contains four interacting layers.

Result 3: RNN을 통한 예측

RNN – LSTM (Long Short Term Memory)



일별 테러 발생 횟수 학습

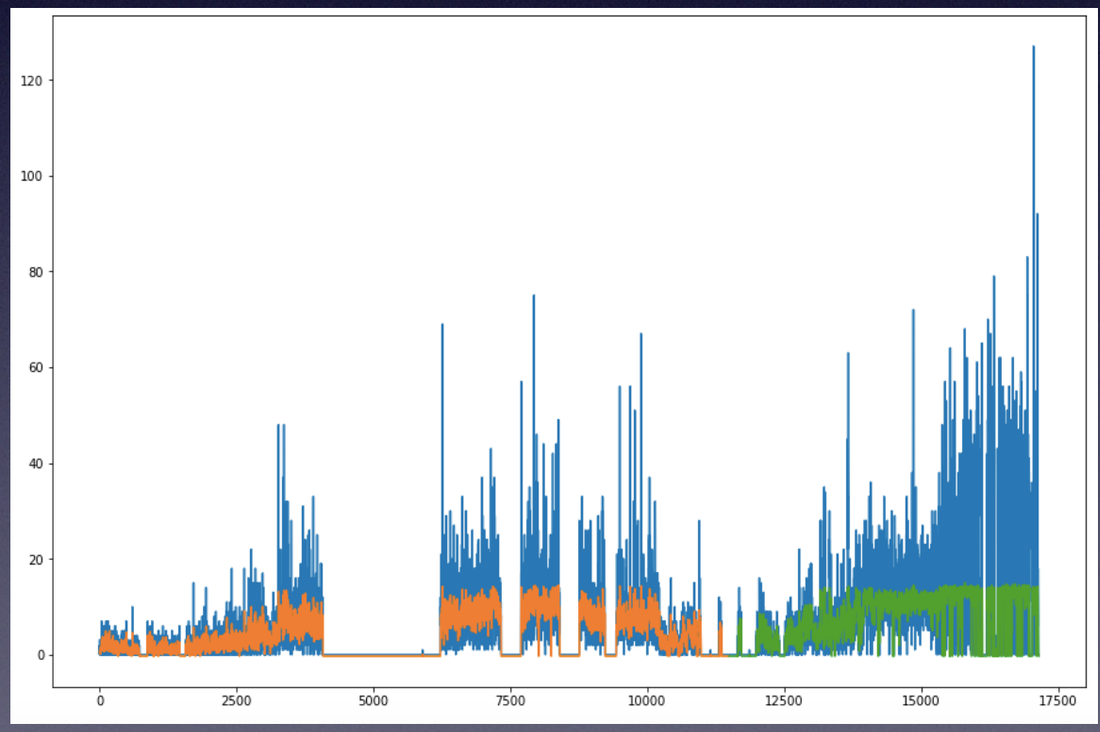


월별 테러 빈도가 가장 높은 지역

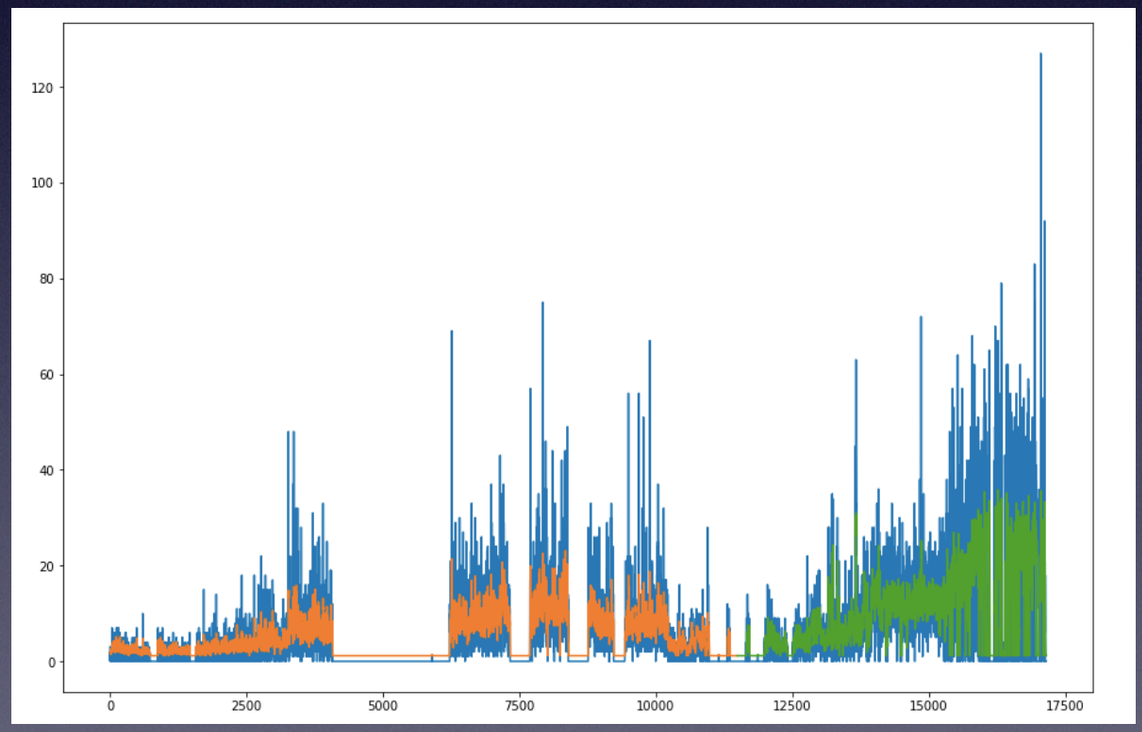
Result 3: RNN을 통한 예측

Limitation

데이터 수가 많고, 불규칙적인 데이터 이기 때문에 더 많은 학습과 layer가 필요



Epoch : 5
Train Error : 3.79
Test Error : 10.51



Epoch : 20
Train Error : 3.80
Test Error : 8.93

Conclusion & Summary

1. 국가, 도시 기준이 아닌 밀집 지역을 기준으로 분석하기 위해 클러스터링 이용

2. 테러가 활발하게 일어나는 지역들을 묘사하기 위해서는 보편적으로 이용되는 K-means보다 밀도 중심의 DBSCAN가 더 적합하다.

3. 테러 발생 예측에는 테러를 일으키는 사람들의 심리와 매우 밀첩하게 관련되어있기 때문에 rnn 내부의 마르코프 모델(markov transition matrix)를 변형하여 더 좋은 결과를 얻을 수 있다.

4. 데이터 분석을 통해 각 항목별 테러 발생 횟수의 distribution에 영향을 크게 미치는 cluster들의 특징들을 알아보는 방법론을 생각해보았다.