

Terrorist Attacks

Joël M. Fonseca, Nelson Antunes,
Enguerrand Granoux and Hedi Driss

EE-558 A Network Tour of Data Science
École Polytechnique Fédérale de Lausanne (EPFL)

January 25, 2018

Outline

① Motivation

② Data Acquisition

- Examine the raw data
- Fill in the gaps

③ Data Exploration

- Visualization
- Basic Facts

④ Data Exploitation

- Clustering

⑤ Conclusion

Outline

1 Motivation

2 Data Acquisition

- Examine the raw data
- Fill in the gaps

3 Data Exploration

- Visualization
- Basic Facts

4 Data Exploitation

- Clustering

5 Conclusion

Motivation

- Terrorism activity has increased drastically in recent years.
Understanding this phenomenon is paramount.
- A better comprehension allows to take measures to counteract and prevent such attacks.
- Use the Data Science process to explain and clarify the question marks we might have as much as possible.

Outline

1 Motivation

2 Data Acquisition

- Examine the raw data
- Fill in the gaps

3 Data Exploration

- Visualization
- Basic Facts

4 Data Exploitation

- Clustering

5 Conclusion

Data Acquisition



NATIONAL CONSORTIUM FOR THE
STUDY OF TERRORISM AND RESPONSES TO TERRORISM

- Global Terrorism Database (GTD) is an open-source database including information on terrorist events around the world.
- More than 170,000 terrorist attacks from 1970 to 2016.
- Annual updates planned for the future.
- GTD Codebook contains useful information.

Examine the raw data

Out[4]:

	eventid	iyear	imonth	iday	approxdate	extended	resolution	country	country_txt	region	...	addnotes	scite1
0	197000000001	1970	7	2	NaN	0	NaN	58	Dominican Republic	2	...	NaN	NaN
1	197000000002	1970	0	0	NaN	0	NaN	130	Mexico	1	...	NaN	NaN
2	197001000001	1970	1	0	NaN	0	NaN	160	Philippines	5	...	NaN	NaN
3	197001000002	1970	1	0	NaN	0	NaN	78	Greece	8	...	NaN	NaN
4	197001000003	1970	1	0	NaN	0	NaN	101	Japan	4	...	NaN	NaN

5 rows × 135 columns

- Lot of different fields. Some are subcategories of others.
- Need to select the more relevant and significant features for the analysis of the project.

Fill in the gaps

- Some values are missing.
- Need to verify which can be retrieved without making wrong assumptions and which is worth.

⇒ Recover the missing coordinates of the attacks where the city is known using the Google Maps Geocoding API.



Outline

1 Motivation

2 Data Acquisition

- Examine the raw data
- Fill in the gaps

3 Data Exploration

- Visualization
- Basic Facts

4 Data Exploitation

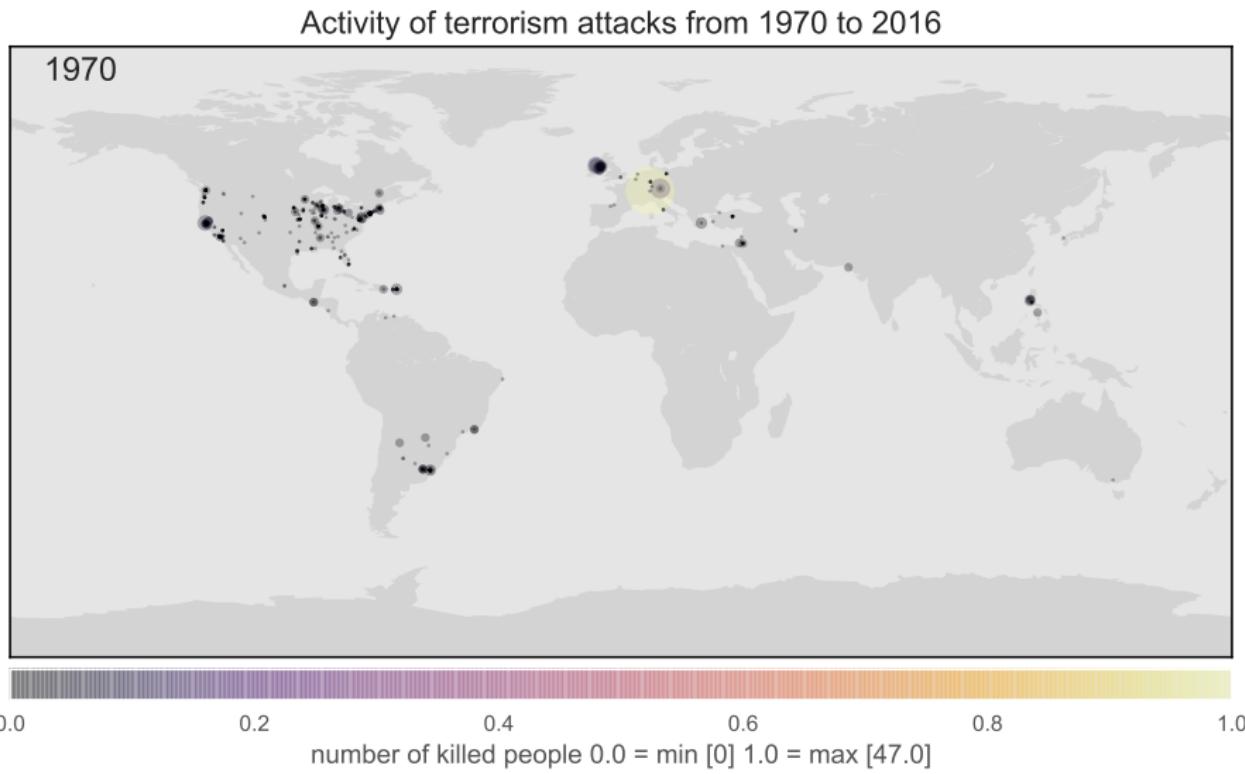
- Clustering

5 Conclusion

Javascript website

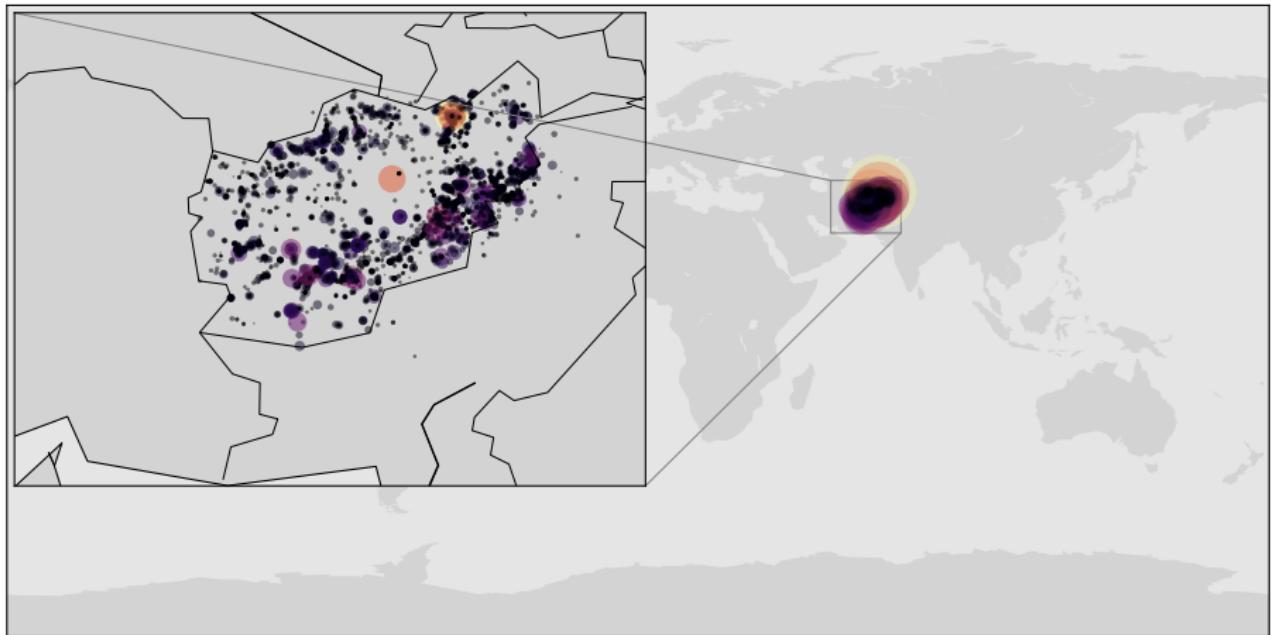
Have a try!

Animation of the attacks



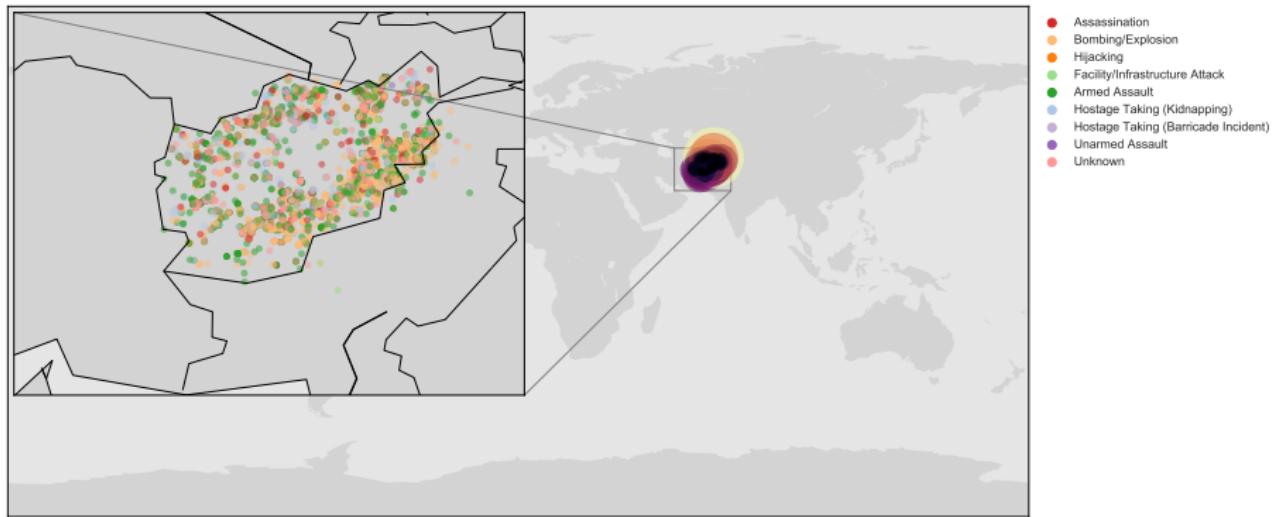
Local zoom

Intensity of attacks perpetrated by the Taliban group

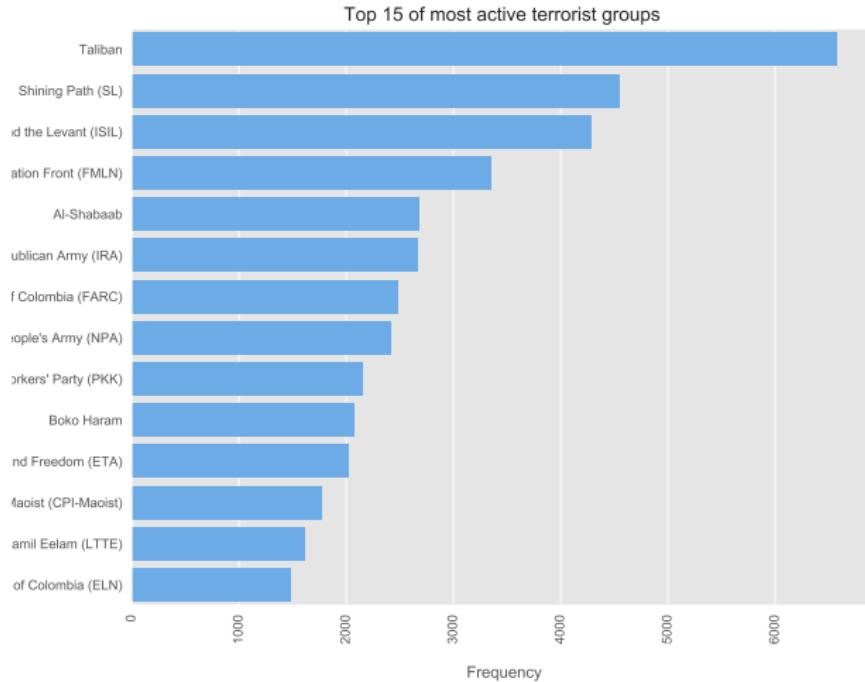


Attack types distribution

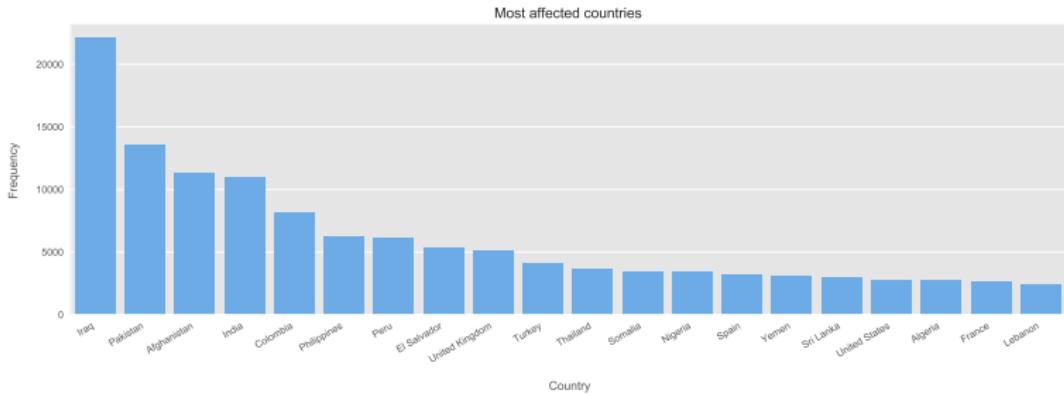
Attack types perpetrated by the Taliban group



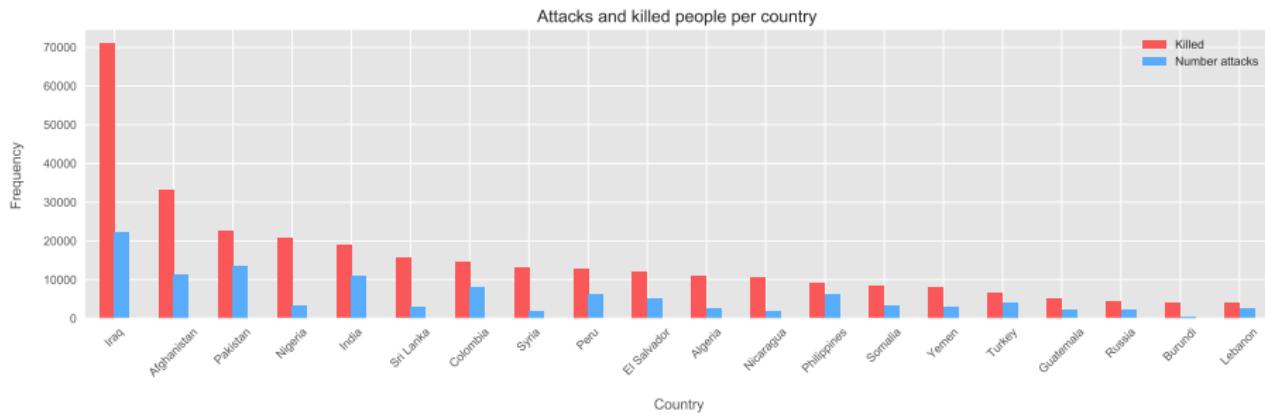
Most active terrorist groups



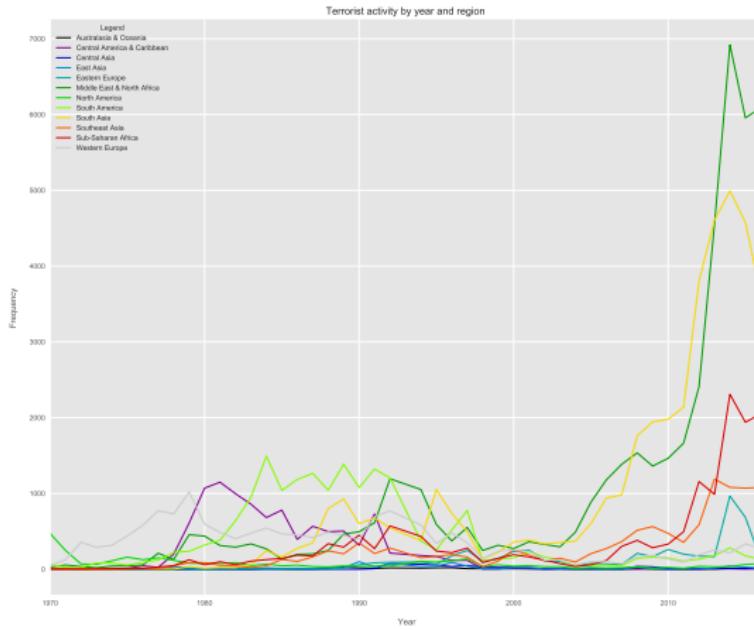
Most affected countries



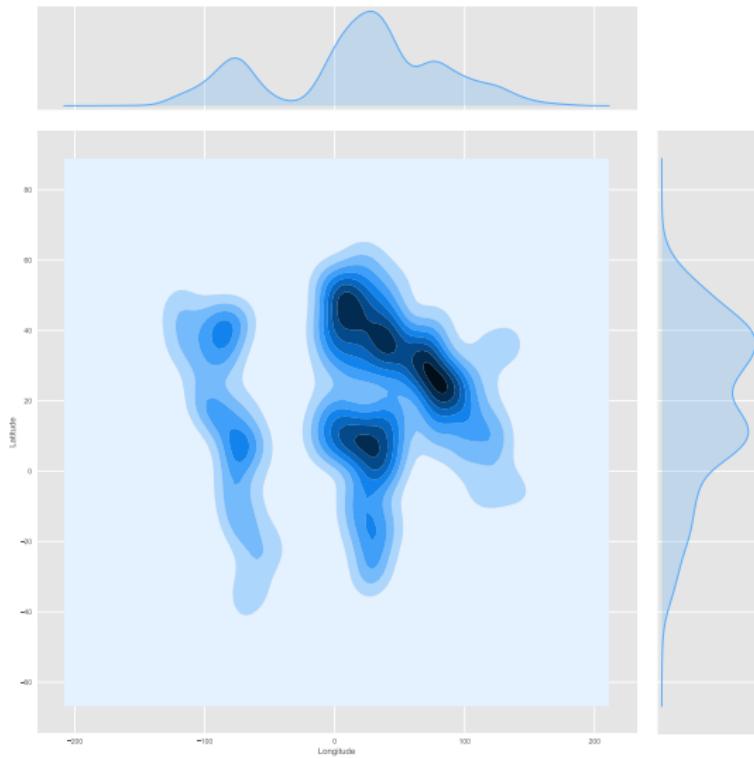
Most affected countries with number of killed people



Attacks by year and region



Joint plot for the coordinates



Outline

1 Motivation

2 Data Acquisition

- Examine the raw data
- Fill in the gaps

3 Data Exploration

- Visualization
- Basic Facts

4 Data Exploitation

- Clustering

5 Conclusion

First try

Group: Taliban

Features:

- number of killed people
- number of wounded people
- attack type
- weapon type
- target type

Techniques:

- Principal Component Analysis (PCA)
- Spectral Embedding using Laplacian Eigenmaps

How about the categorical features?

Categorical features can lead to misinterpretation of the results, if they are not dealt correctly.



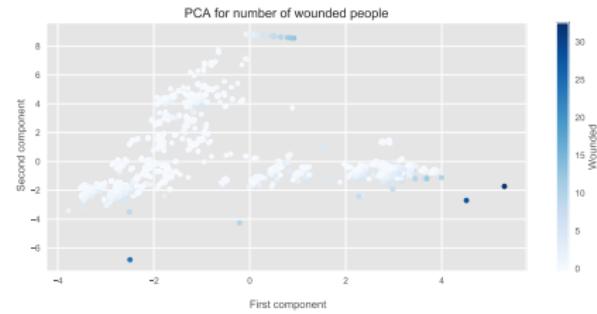
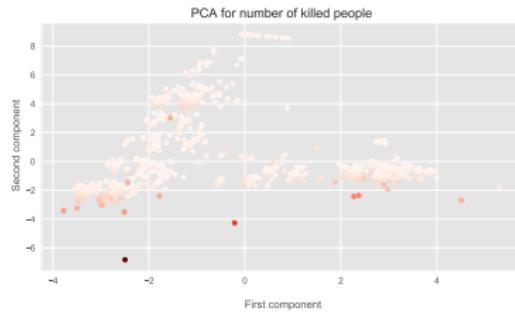
Represent the different categories as orthogonal vectors.



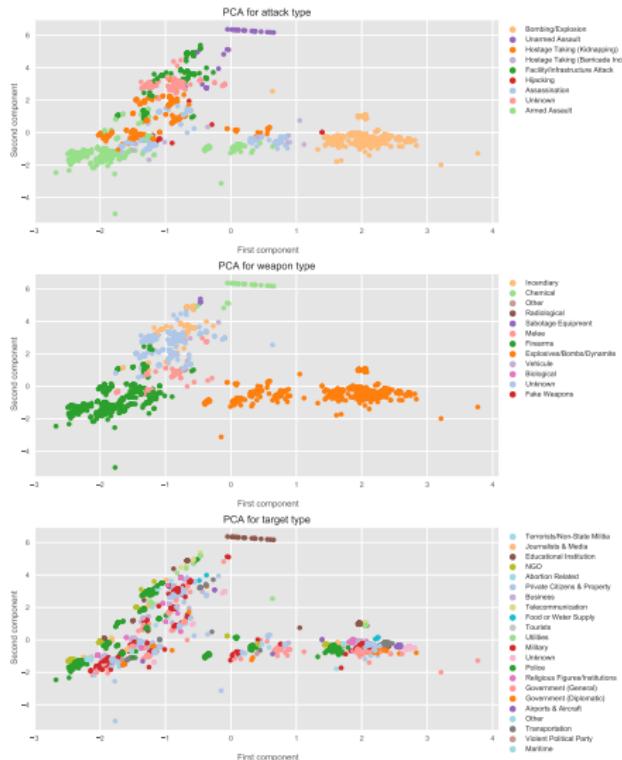
Avoid making wrong mathematical assumptions.



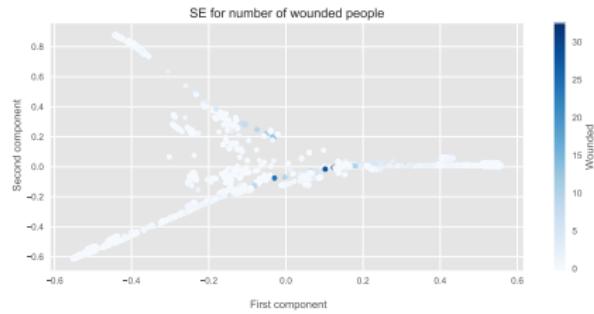
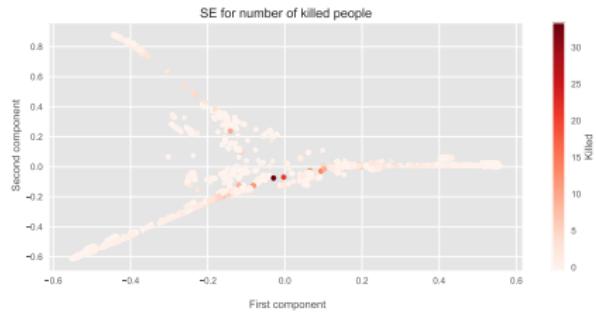
Principal Component Analysis (PCA)



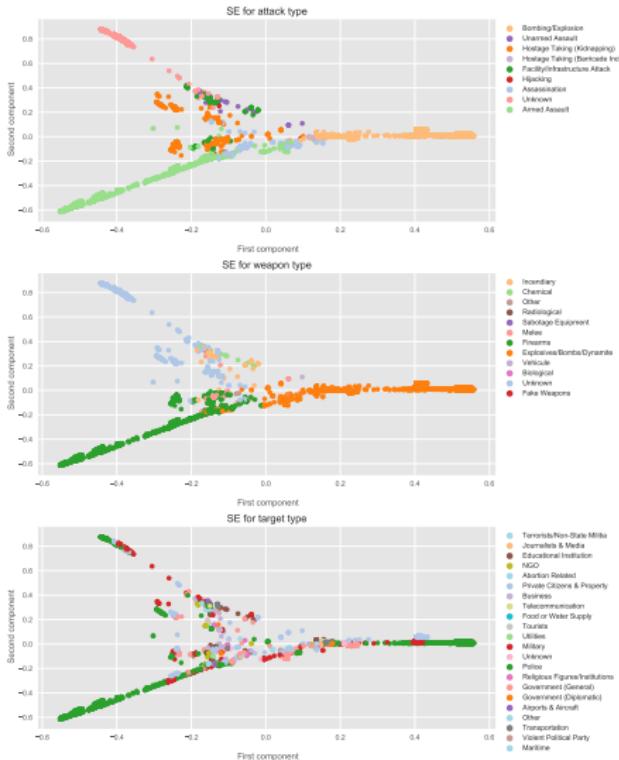
Principal Component Analysis (PCA)



Spectral Embedding using Laplacian Eigenmaps



Spectral Embedding using Laplacian Eigenmaps



Second try

Periods: Since 2010

Group: 5 most actives groups

Features:

- number of killed people
- number of wounded people
- region
- attack type
- weapon type
- target type
- capital

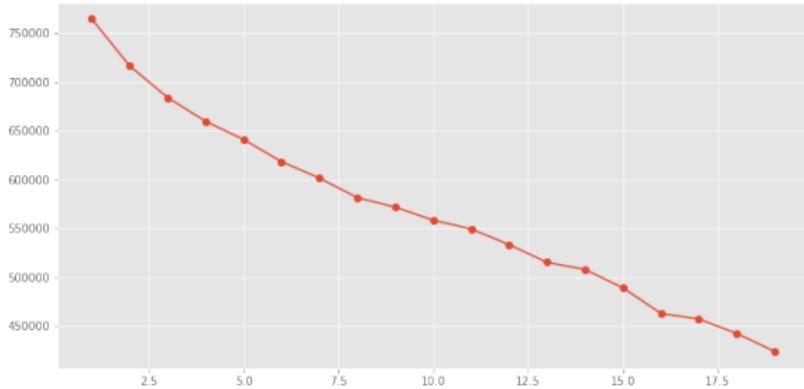
Techniques:

- K-Means Clustering
- Gausian Mixture Model Clustering

Goals

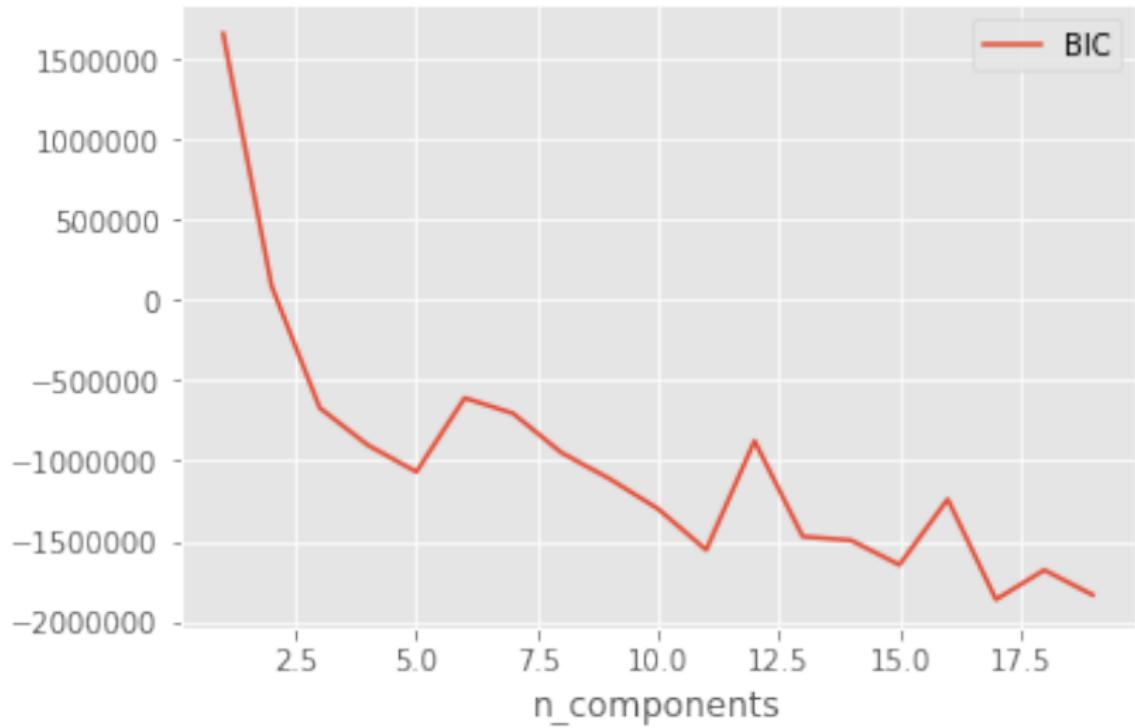
For this analyse, we remove groups labels and see if we can recover them with different clustering algorithms.

Kmeans - Find the best K

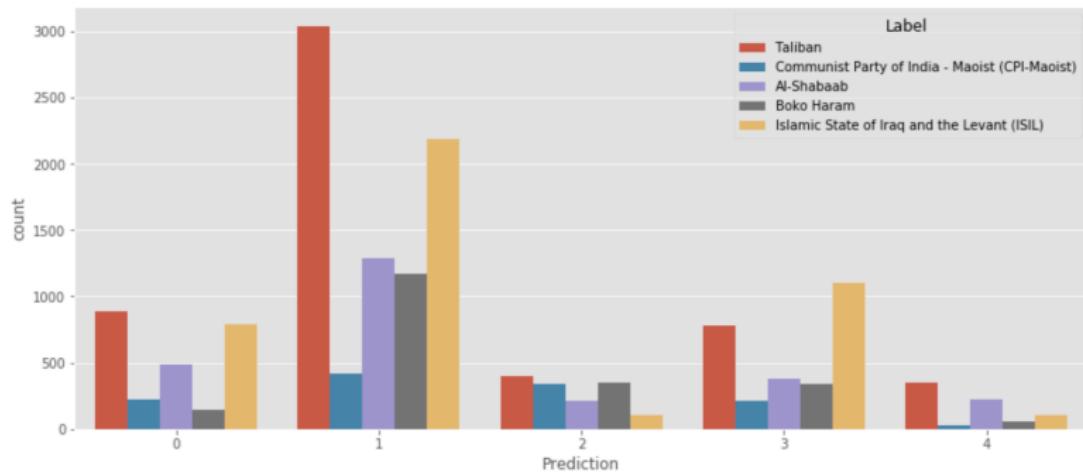


- The data may not be spherical
- Curse of dimensionality

GMM



Clustering of the data and comparison to our groups label



Outline

1 Motivation

2 Data Acquisition

- Examine the raw data
- Fill in the gaps

3 Data Exploration

- Visualization
- Basic Facts

4 Data Exploitation

- Clustering

5 Conclusion

Conclusion

- Visualization is great for data where lat/long coordinates are available. Possibilities are endless.
- Simple plots help understanding the context of terrorism.
- PCA allows to explain hidden correlations.
- Spectral embedding is well suited for clustering.
- Results of both techniques allow to draw interesting conclusions.

Thank you!