

~~Think of a more exciting title~~

Understanding Follow-Up Terrorist Attacks

Six weeks ago

Sun	Mon	Tue	Wed	Thu	Fri	Sat
26	27	28	29	30	Jul 1	2 <div>Istanbul Vacation</div>
3	4	5	6	7	8	9
Istanbul Vacation						
10	11	12	13	14	15	16
Istanbul Vacation						

Istanbul Ataturk airport attack: 41 dead and more than 230 hurt

© 29 June 2016 | Europe

Share



Turkey said early signs suggested so-called Islamic State was behind the attack

A gun and bomb attack on Istanbul's Ataturk airport has killed 41 people, at least 13 of them foreigners, and injured more than 230, officials say.

Today



July 2016

Day

Week

Month

4 Days

Agenda

More ▾



Sun

Mon

26

27

3

4

5

6

Thu

Fri

Sat

30

Jul 1

2

7

8

9

Istanbul Vacation

Istanbul Vacation

10

11

12

13

14

15

16

Istanbul Vacation

Istanbul Ataturk airport attack: 41 dead and more than 230 hurt

© 29 June 2016 | Europe

Share



Turkey said early signs suggested so-called Islamic State was behind the attack

A gun and bomb attack on Istanbul's Ataturk airport has killed 41 people, at least 13 of them foreigners, and injured more than 230, officials say.

Today



July 2016

Day

Week

Month

4 Days

Agenda

More ▾



Sun

Mon

26

27

3

4

5

6

Thu

Fri

Sat

30

Jul 1

2

Istanbul Vacation

7

8

9

Istanbul Vacation

10

11

12

13

14

15

16

Istanbul Vacation

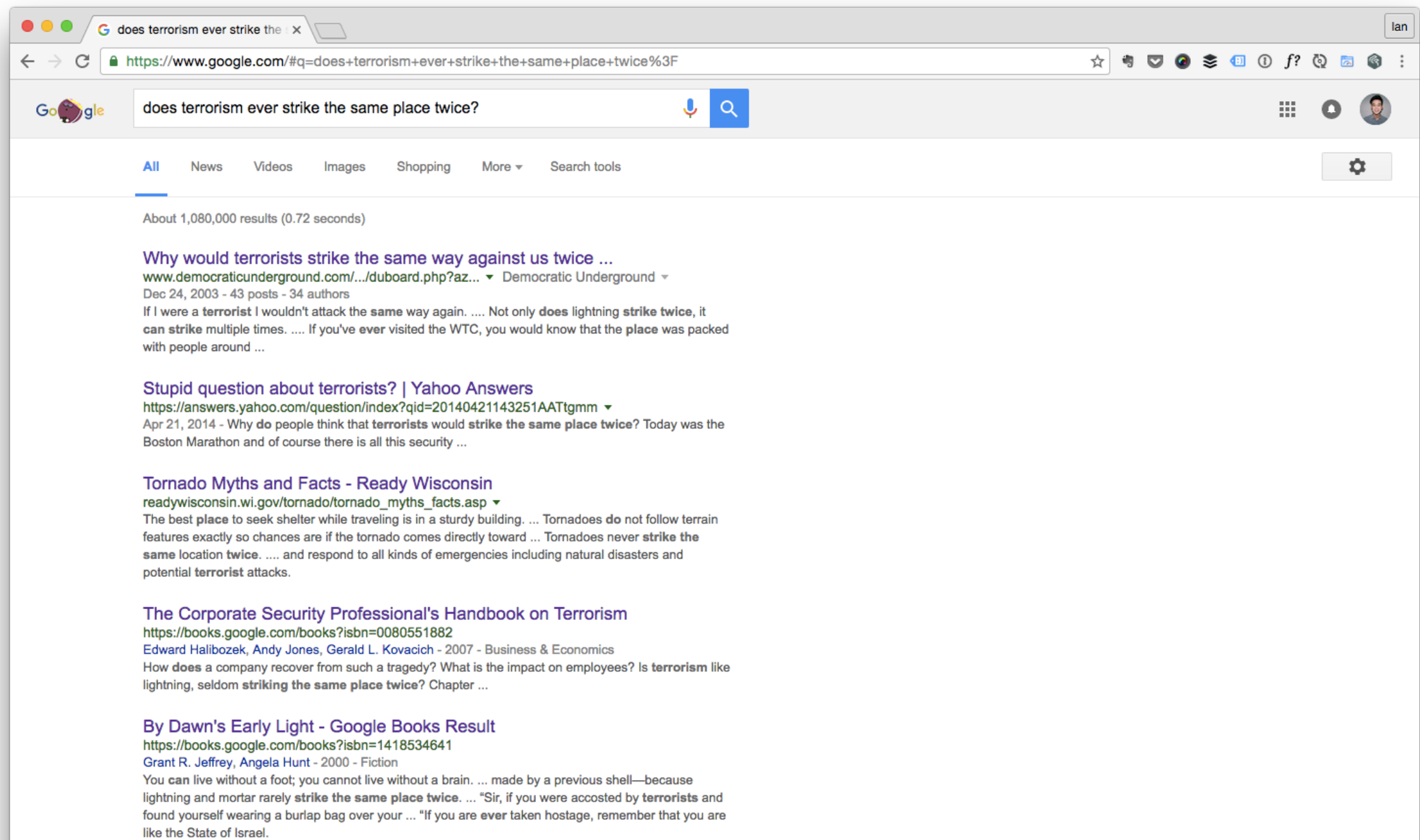
Turkey coup attempt: How a night of death and mayhem unfolded

By Chandrika Marayana, [Fahri Karimli](#), Farida Faruqi, and Shanna Pavlak CNN
Updated 3:39 PM ET, Sun July 17, 2016



“Does ~~lightning~~ terrorism ever
strike the same place twice?”

I don't care about your opinion, *Yahoo Answers*



data.shape

(156772, 137)

Global Terrorism Database

lan

https://www.start.umd.edu/gtd/

GTD

GLOBAL TERRORISM DATABASE

[ABOUT GTD](#) [USING GTD](#) [FAQ](#) [TERMS OF USE](#) [CONTACT](#) [START HOME PAGE](#)

Search the Database

SEARCH

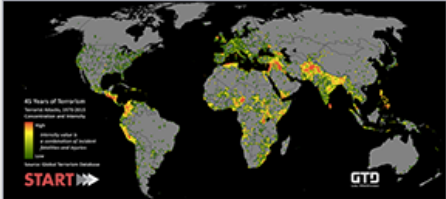
[I'm a New User](#) [ADVANCED SEARCH](#)

Browse by: Go

Information on more than 150,000 Terrorist Attacks

The Global Terrorism Database (GTD) is an open-source database including information on terrorist events around the world from 1970 through 2015 (with annual updates planned for the future). Unlike many other event databases, the GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 150,000 cases. [Learn more](#)

GTD DATA VISUALIZATIONS



Updated in 2016, the **GTD World Map: 45 Years of Terrorism** displays the concentration and intensity (combining fatalities and injuries) of terrorist attacks that occurred worldwide across 45 years of data.

[The GTD 2015 World Map is available here..](#)

[The GTD 2014 World Map is available here..](#)

[The GTD 2013 World Map is available here.](#)

[The GTD 2012 World Map is available here.](#)

THIS DATE IN TERRORISM

August 15

2014

Santiago, Chile

08/15/2014: An explosive device was discovered and safely defused at a Banco del Estado office in Santiago city, Santiago province, Chile. No group claimed responsibility for the incident.

[Learn more](#)

2015

Rafah, Egypt

08/15/2015: Security forces identified a group of eight assailants and killed them as they were attempting to attack a military checkpoint south of Rafah, North Sinai governorate, Egypt. No group claimed responsibility for the incident; however, sources attributed the attempted attack to the Sinai Province of the Islamic State.

[Learn more](#)

FEATURED

Additional START Datasets Now Available

Utilizing the Dataverse Network Project, START has created its own repository of datasets and databases on terrorism, conflict, and preparedness. This collection includes research funded by START as well as research for which START has been given permission to release. Users can read over detailed information about each dataset regarding its time period, geographic coverage, and sampling procedure.

[Read more](#)

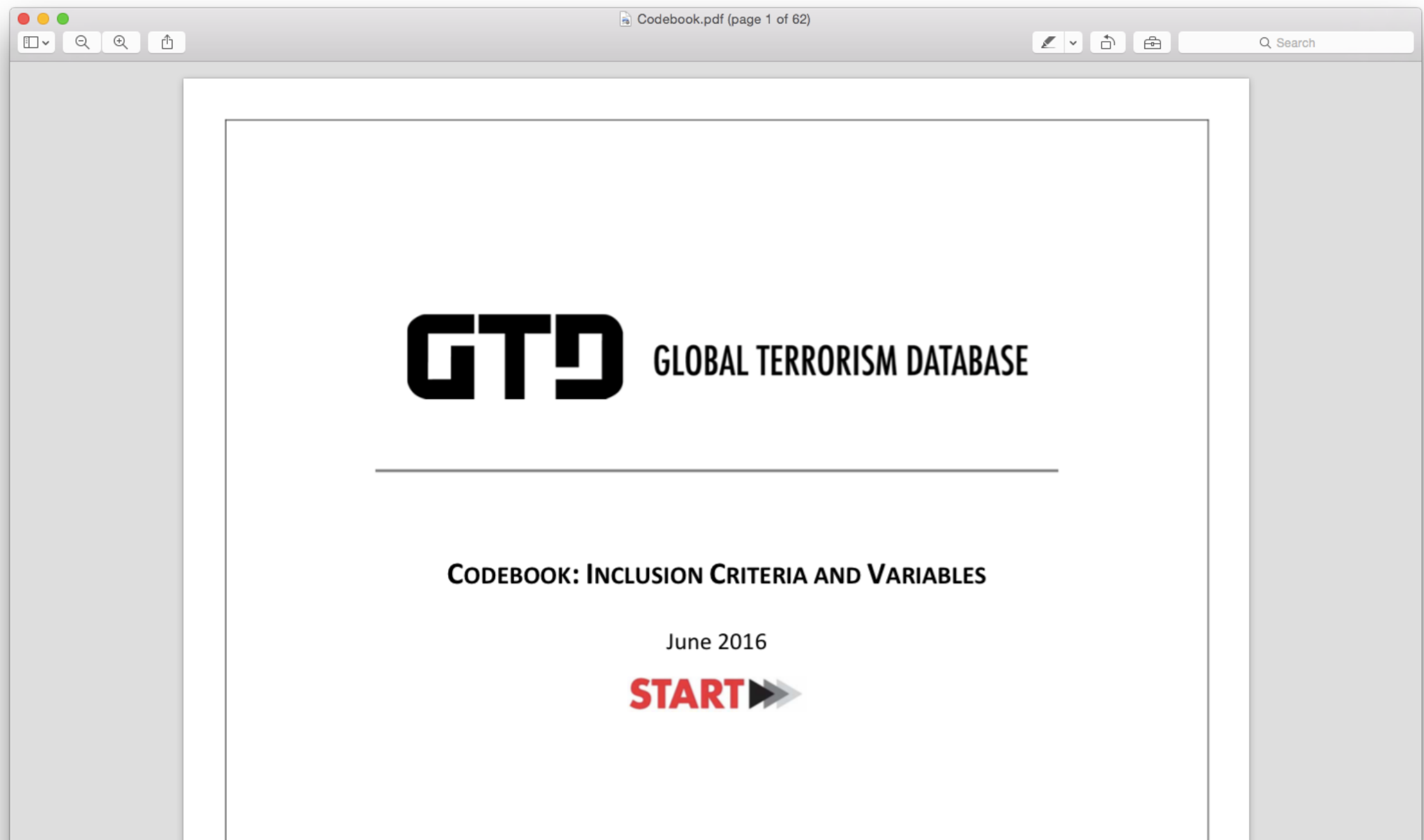
1

2

3

4

62-page codebook of mostly
categorical variables (1970-2015)



What is a terrorist act?

3/3:

The incident must be **intentional** – the result of a conscious calculation on the part of a perpetrator.

The incident must entail some level of **violence or immediate threat of violence** - including property violence, as well as violence against people.

The perpetrators of the incidents must be **sub-national actors**. The database does not include acts of state terrorism.

What is a terrorist act?

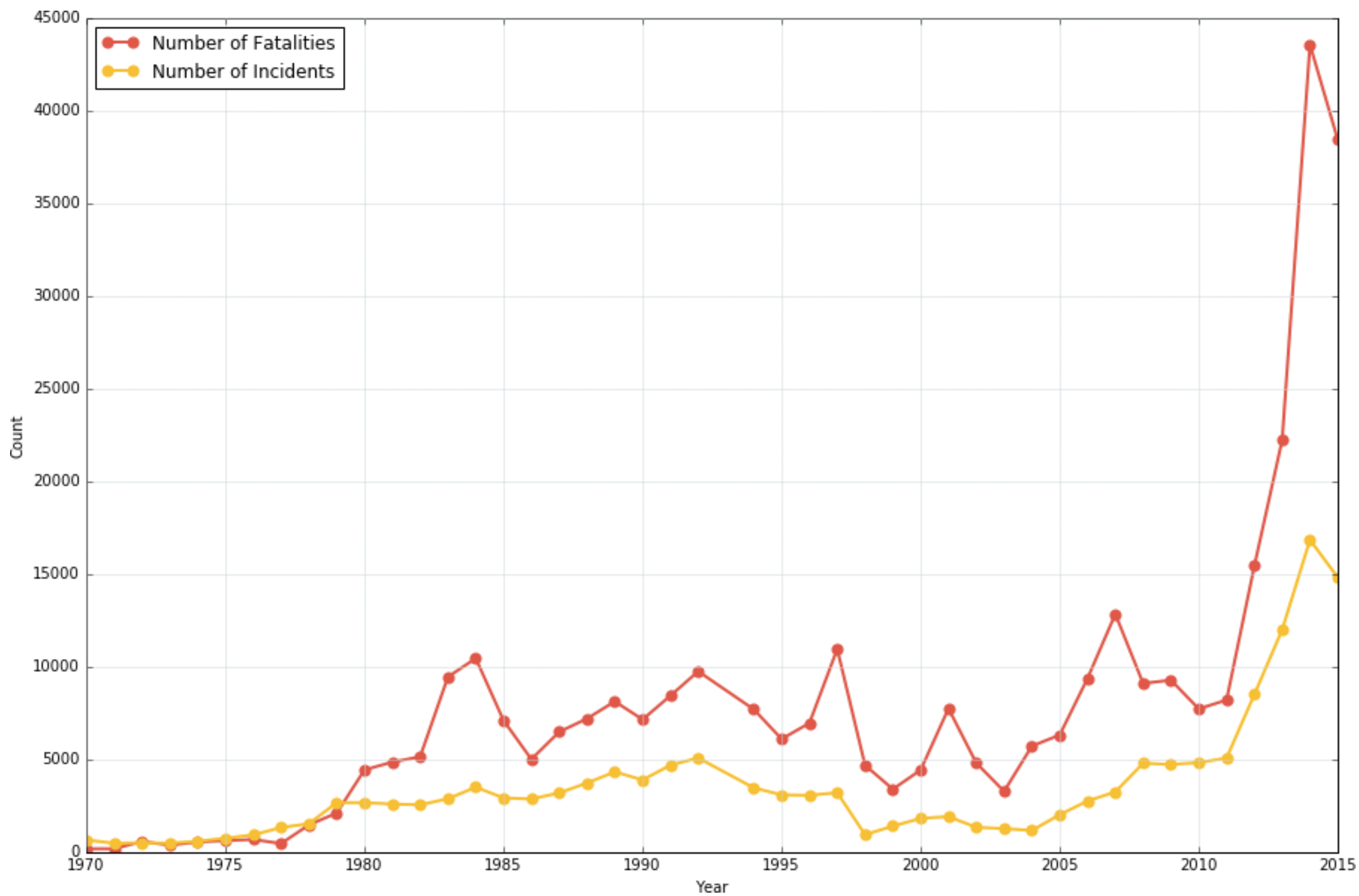
2/3:

The act must be aimed at attaining a **political, economic, religious, or social goal**. In terms of economic goals, the exclusive pursuit of profit does not satisfy this criterion. It must involve the pursuit of more profound, systemic economic change.

There must be evidence of an intention to **coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims**. It is the act taken as a totality that is considered, irrespective if every individual involved in carrying out the act was aware of this intention. As long as any of the planners or decision-makers behind the attack intended to coerce, intimidate or publicize, the intentionality criterion is met.

The action must be **outside the context of legitimate warfare activities**. That is, the act must be outside the parameters permitted by international humanitarian law (particularly the prohibition against deliberately targeting civilians or non-combatants).

First impressions



Calculating 'NextAttack'

AKA The Most Important Thing
I Learned This Semester

Country	City	Date	NextAttack
United States	Gotham	01-07-2015	8 days
United Kingdom	Hogsmeade	01-07-2015	0 days
United Kingdom	Hogsmeade	01-07-2015	NaN
United States	Metropolis	01-09-2015	NaN
United States	Gotham	01-15-2015	NaN

The Dumb Way

Country	City	Date	NextAttack
United States	Gotham	01-07-2015	8 days
United Kingdom	Hogsmeade	01-07-2015	0 days
United Kingdom	Hogsmeade	01-07-2015	NaN
United States	Metropolis	01-09-2015	NaN
United States	Gotham	01-15-2015	NaN

The Dumb Way

Country	City	Date	NextAttack
United States	Gotham	01-07-2015	8 days
United Kingdom	Hogsmeade	01-07-2015	0 days
United Kingdom	Hogsmeade	01-07-2015	NaN
United States	Metropolis	01-09-2015	NaN
United States	Gotham	01-15-2015	NaN

... takes 60 hours

The Less Dumb Way

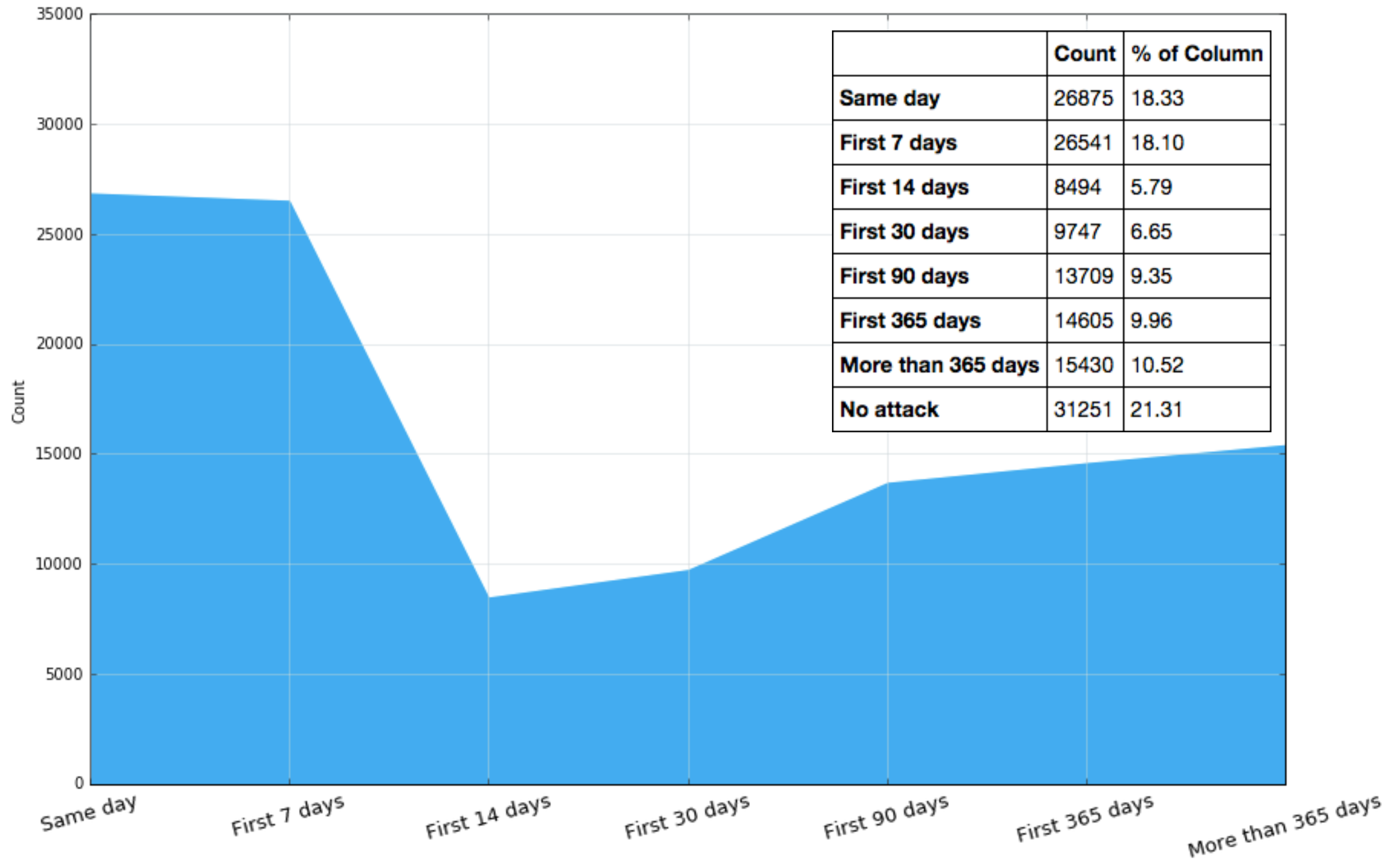
Country	City	Date	NextAttack
United States	Gotham	01-07-2015	8 days
United States	Gotham	01-15-2015	NaN
United States	Metropolis	01-09-2015	NaN
United Kingdom	Hogsmeade	01-07-2015	0 days
United Kingdom	Hogsmeade	01-07-2015	NaN

The Less Dumb Way

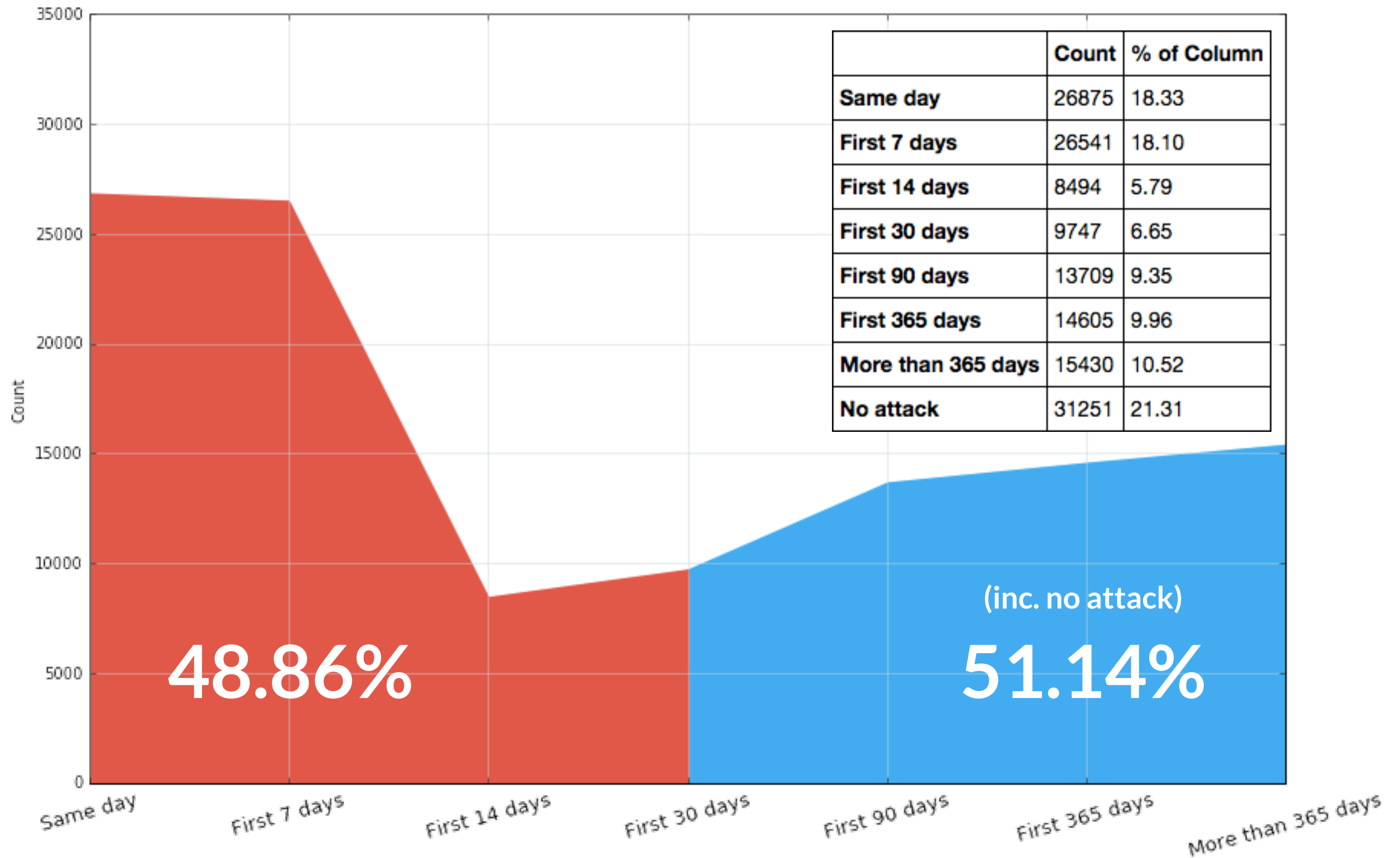
Country	City	Date	NextAttack
United States	Gotham	01-07-2015	8 days
United States	Gotham	01-15-2015	NaN
United States	Metropolis	01-09-2015	NaN
United Kingdom	Hogsmeade	01-07-2015	0 days
United Kingdom	Hogsmeade	01-07-2015	NaN

... takes 20 minutes

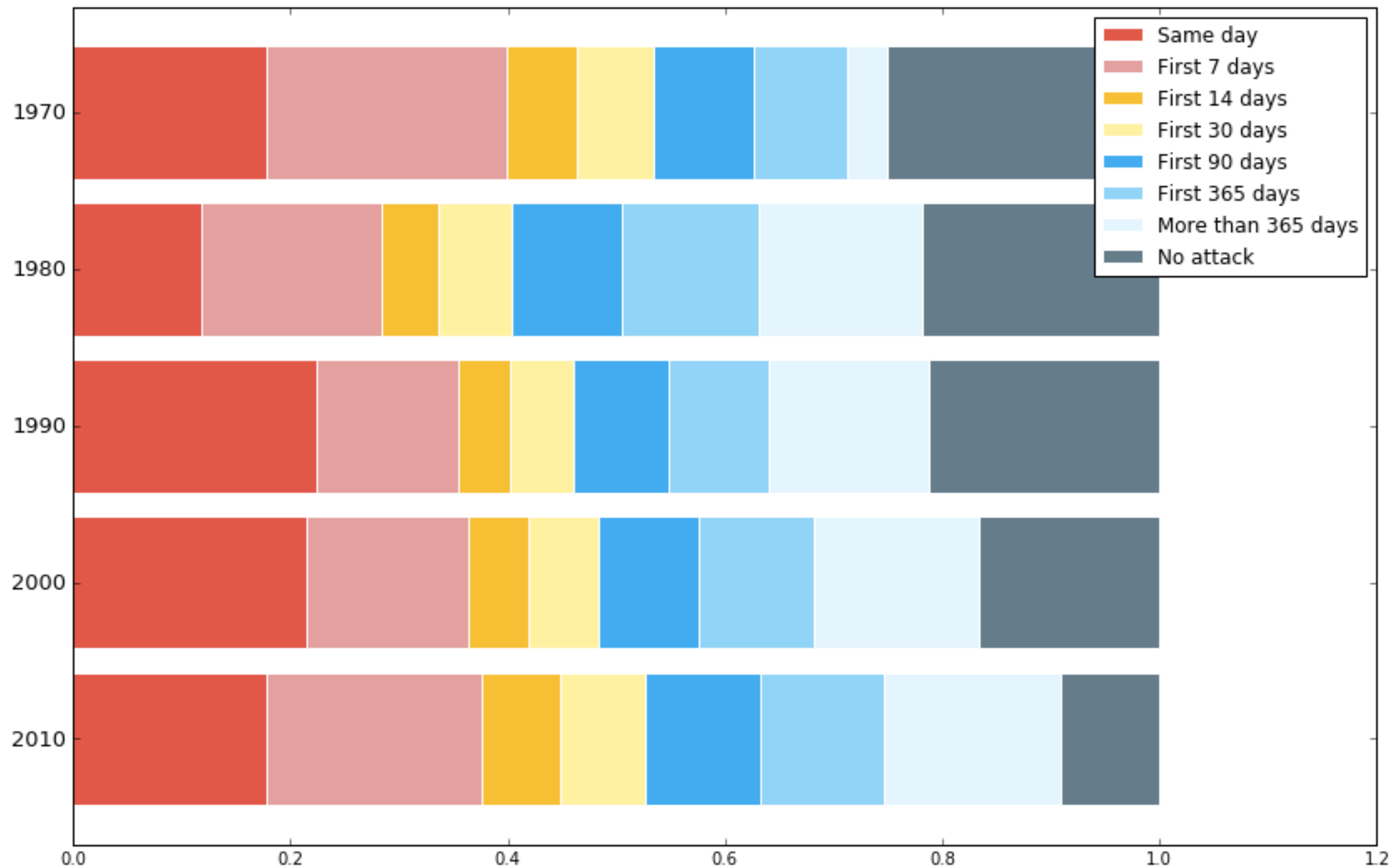
Distribution of 'NextAttack'



Binary outcomes - 'AttackIn30Days'



Sanity check - 'NextAttack' by decade



What factors could predict the chance of a city getting attacked again in 30 days?

The choice of weapon?

Explosives/Bombs/Dynamite
Sabotage Equipment
Incendiary
Chemical
Unknown
Firearms
Vehicle
Melee
Other

The country?

United Kingdom
Afghanistan
El Salvador
Philippines
Colombia
Pakistan
India
Peru
Iraq

The type of attack?

Hostage Taking (Barricade Incident)
Hostage Taking (Kidnapping)
Facility/Infrastructure Attack
Bombing/Explosion
Unarmed Assault
Armed Assault
Assassination
Unknown
Hijacking

What factors could predict the chance of a city getting attacked again in 30 days?

The target of the attack?

Religious Figures/Institutions
Private Citizens & Property
Government (General)
Educational Institution
Transportation
Business
Military
Utilities
Police

The group responsible?

Farabundo Marti National Liberation Front (FMLN)
Revolutionary Armed Forces of Colombia (FARC)
Islamic State of Iraq and the Levant (ISIL)
Basque Fatherland and Freedom (ETA)
Kurdistan Workers' Party (PKK)
Irish Republican Army (IRA)
New People's Army (NPA)
Shining Path (SL)
Taliban

```
data['imonth'].replace(to_replace=0, value=np.NaN, inplace=True)
data['iday'].replace(to_replace=0, value=np.NaN, inplace=True)
data.dropna(how='any', inplace=True)
```

data.shape
(146652, 6)

Dropped 10,120 observations or 6.45% of data

country_txt	attacktype1_txt	weaptype1_txt	targettype1_txt	gname	AttackIn30Days
Thailand	Bombing/Explosion	Explosives/Bombs/Dynamite	Educational Institution	Unknown	0
Egypt	Bombing/Explosion	Explosives/Bombs/Dynamite	Private Citizens & Property	Unknown	1
Colombia	Hostage Taking (Kidnapping)	Unknown	Government (General)	National Liberation Army of Colombia (ELN)	1
United Kingdom	Bombing/Explosion	Explosives/Bombs/Dynamite	Police	Dissident Republicans	1
Egypt	Bombing/Explosion	Explosives/Bombs/Dynamite	Unknown	Sinai Province of the Islamic State	1

All our data is categorical!

And if we were to dummify those categorical variables...

```
data.shape (146652, 3412)
```

A little overwhelming - this will definitely slow down our models

So let's start by taking a sample

```
data_sample = data.sample(len(data)/100)
```

```
data_sample.shape (1466, 394)
```

Perfect!

Relative values for dummified variables:

```
country_txt = Afghanistan, attacktype1_txt = Armed Assault,  
weaptype1_txt = Biological, targtype1_txt = Abortion Related, gname = 1 May
```

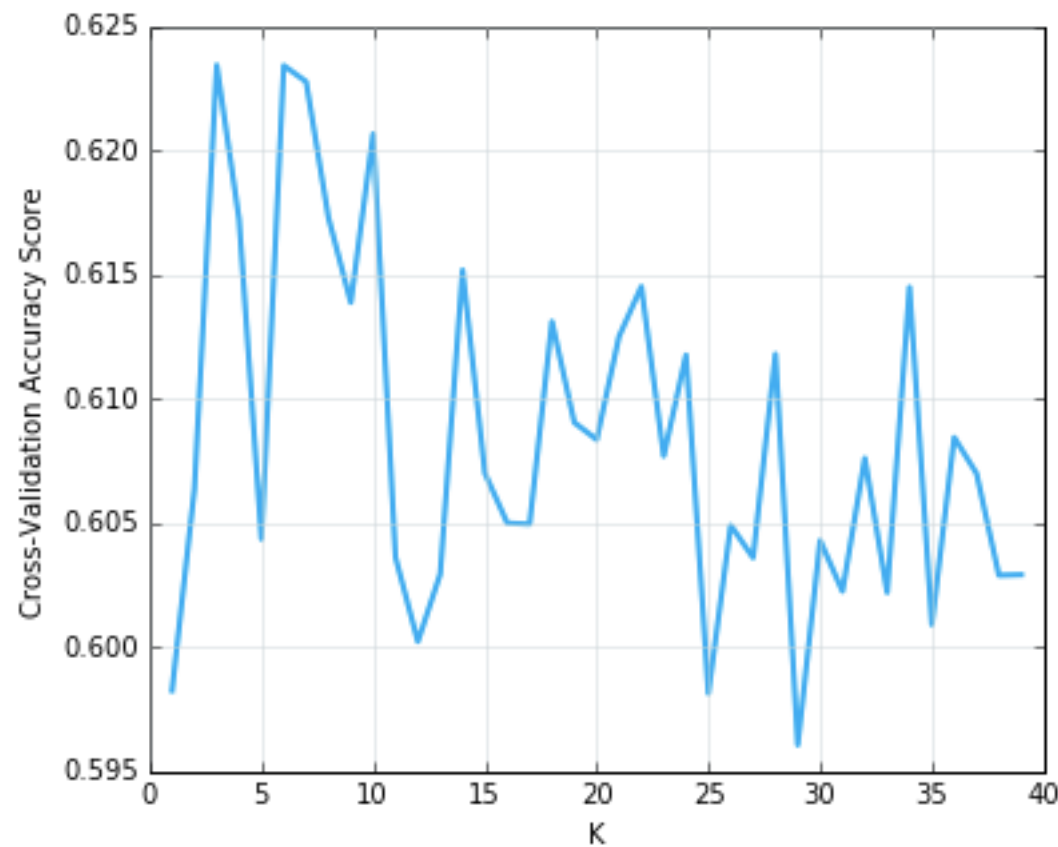
Can we see some
models? Please?

Current Best = 0.5114

Bachelor #1

K-Nearest Neighbors

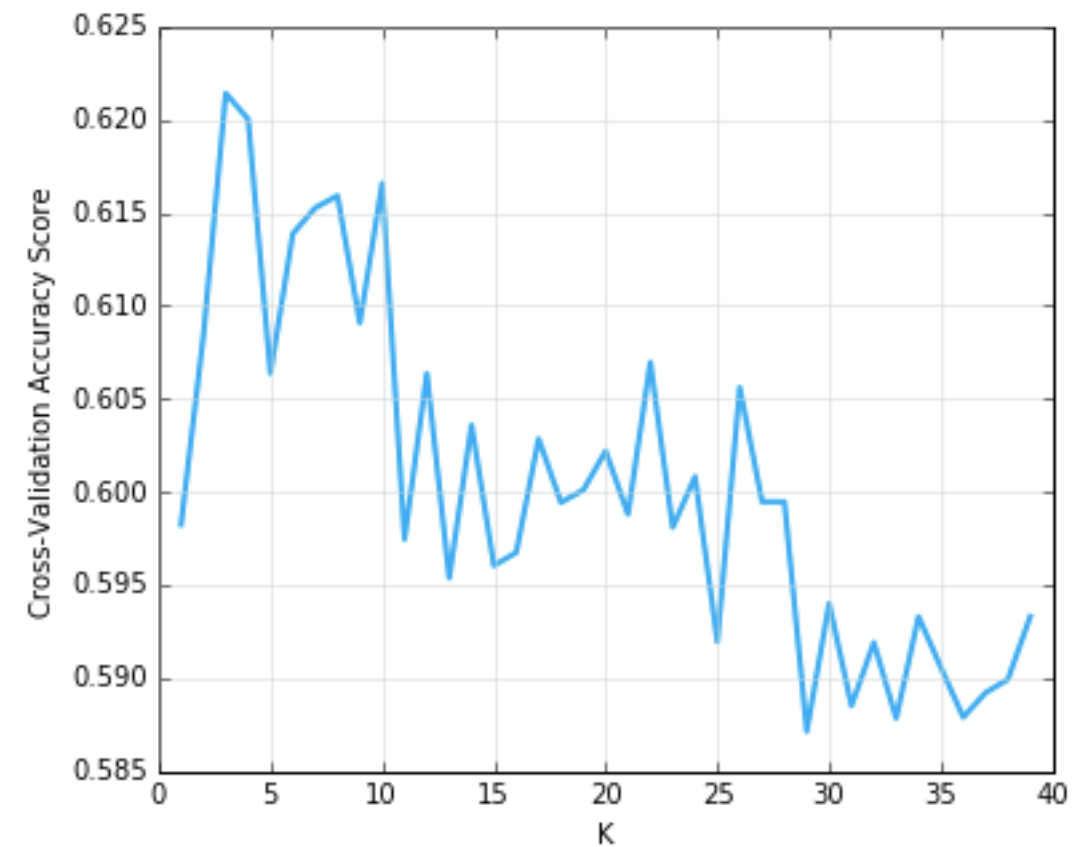
weights="uniform"



K = 3

CV Score = 0.6234

weights="distance"



K = 3

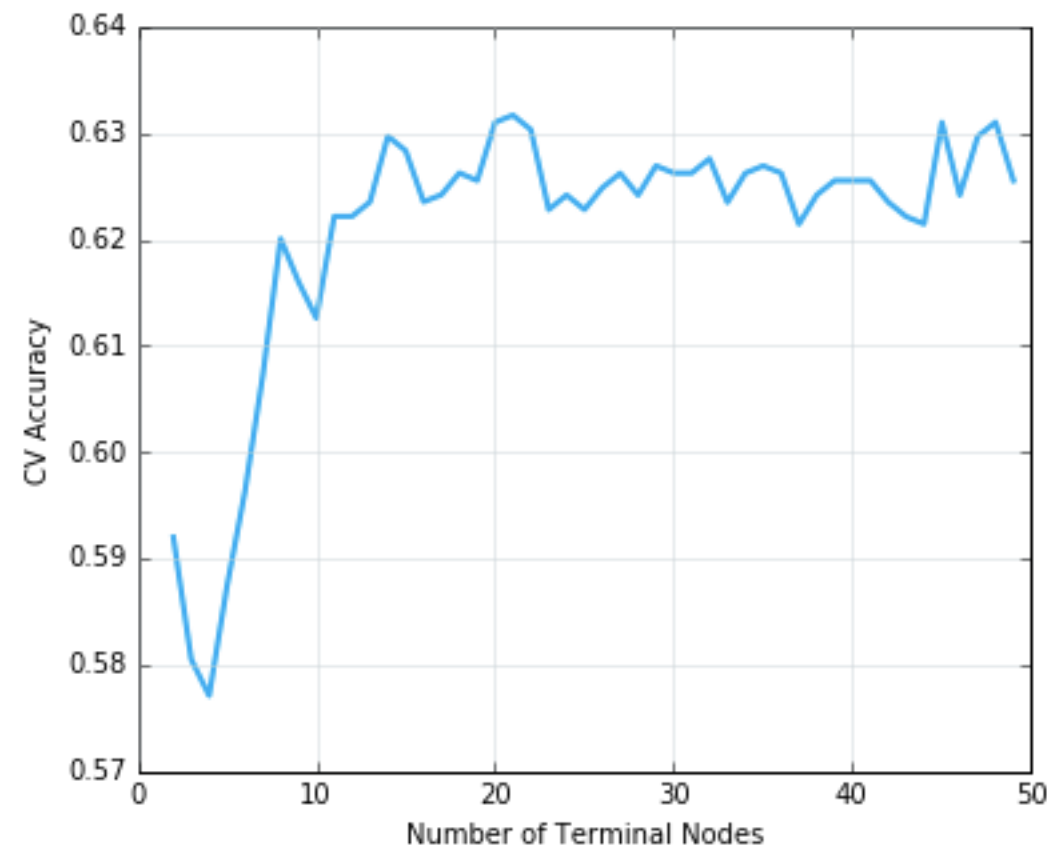
CV Score = 0.6214

Current Best = 0.6234

Bachelor #2

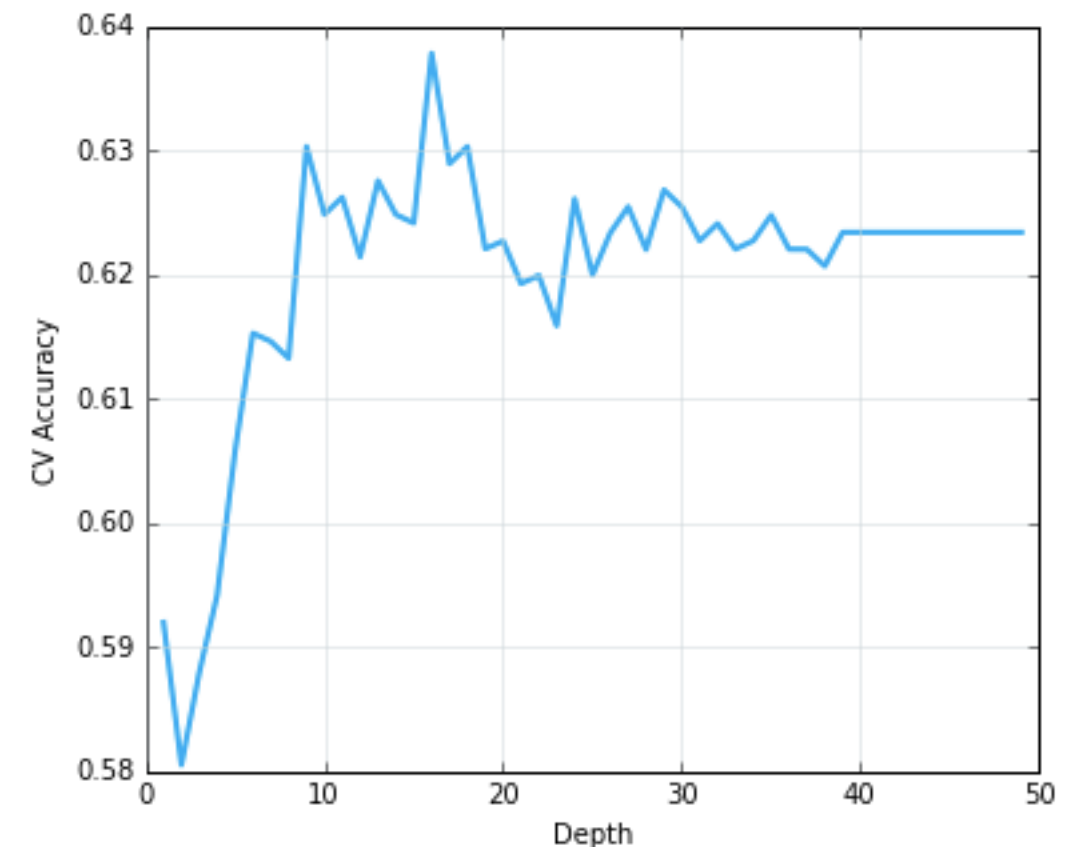
Simple Decision Tree

Tuning w/Terminal Nodes



Maximum Terminal Nodes = 21
Accuracy = 0.6317

Tuning w/Max Depth



Maximum Depth = 16
Accuracy = 0.6378

Current Best = 0.6378

Bachelor #2

Simple Decision Tree

Key Features sorted by importance

	feature	importance
42	country_txt_Iraq	0.293064
63	country_txt_Pakistan	0.049382
39	country_txt_India	0.046795
104	attacktype1_txt_Bombing/Explosion	0.042415
94	country_txt_United Kingdom	0.040878
29	country_txt_El Salvador	0.032511
103	attacktype1_txt_Assassination	0.030496
49	country_txt_Lebanon	0.027128
99	country_txt_Yemen	0.026787
19	country_txt_Chile	0.025378
372	gname_Unknown	0.023085
123	targtype1_txt_Government (General)	0.022575
87	country_txt_Syria	0.022386
57	country_txt_Myanmar	0.021935
44	country_txt_Israel	0.020052
124	targtype1_txt_Journalists & Media	0.019151

What factors could predict the chance of a city getting attacked again in 30 days?

The choice of weapon?

The country?

The type of attack?

The target of the attack?

The group responsible?

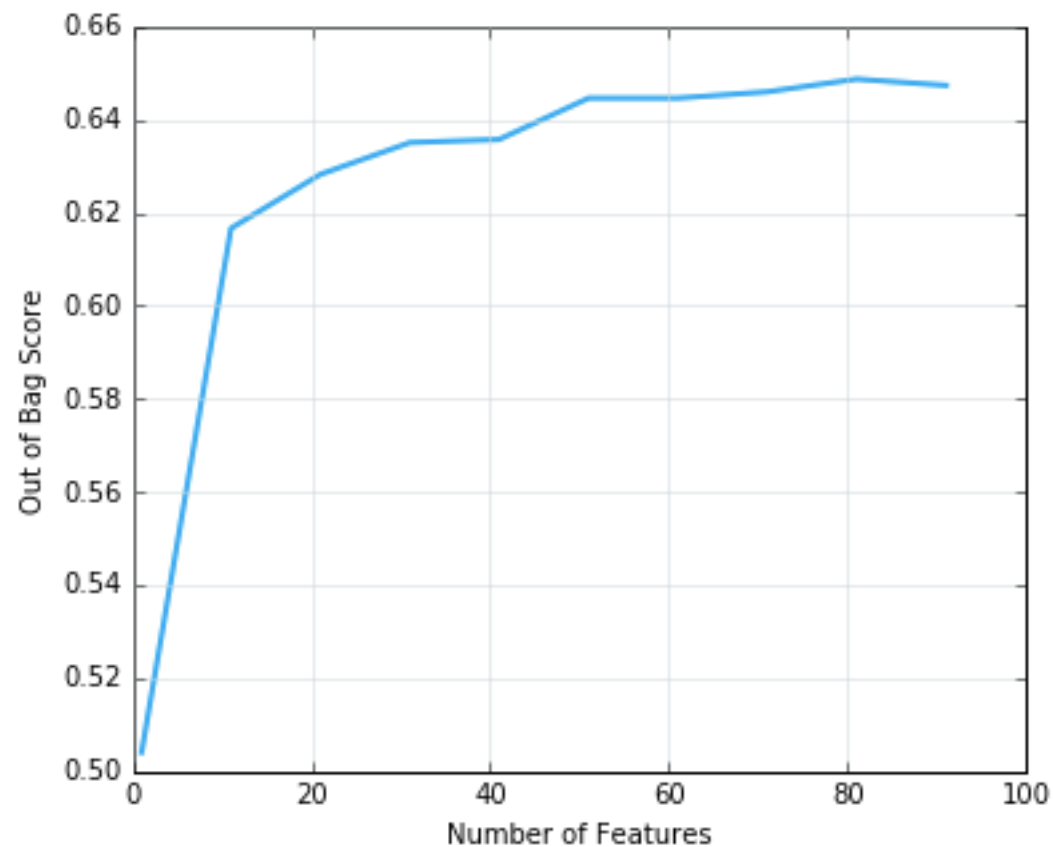


Current Best = 0.6378

Bachelor #3

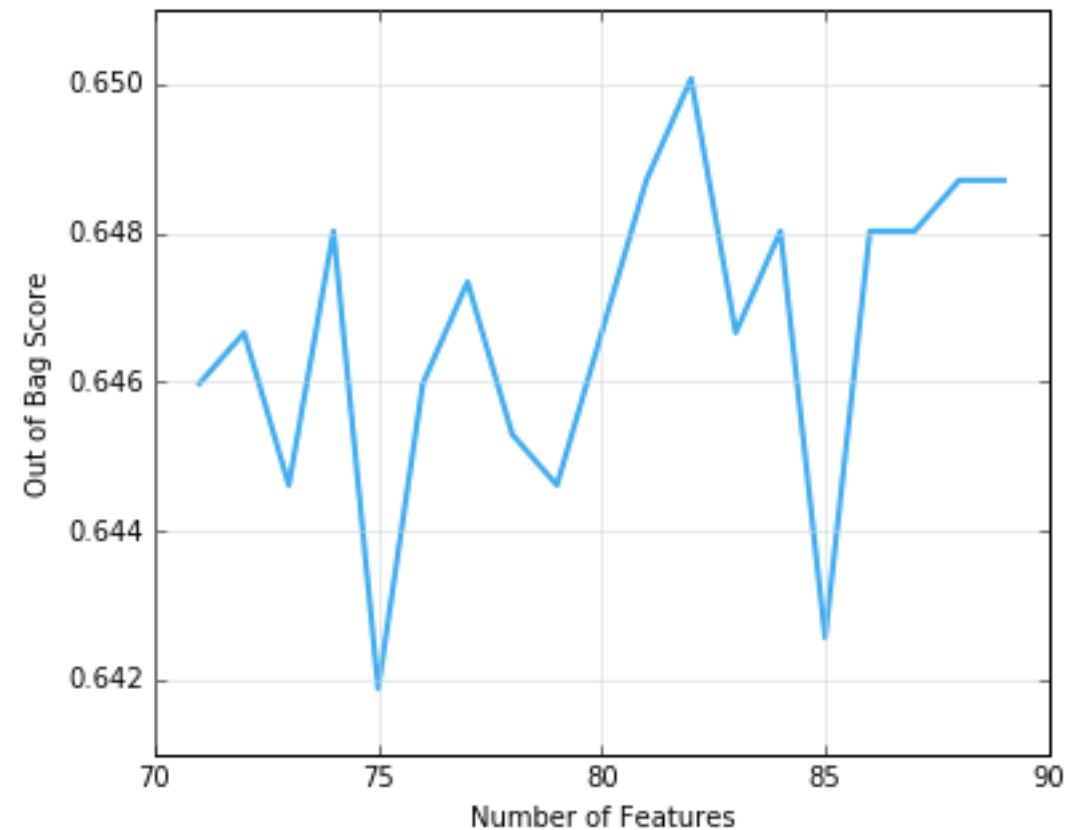
Random Forest

Tuning Max Features w/range(1, 100, 10)



Max Features = 81
OOB Score = 0.6487

Tuning Max Features w/range(75, 90)



Max Features = 82
OOB Score = 0.6501

Current Best = 0.6501

Bachelor #3

Random Forest

Key Features sorted by importance

	feature	importance
42	country_txt_Iraq	0.221031
39	country_txt_India	0.044121
372	gname_Unknown	0.041501
104	attacktype1_txt_Bombing/Explosion	0.040698
63	country_txt_Pakistan	0.036581
111	weaptype1_txt_Explosives/Bombs/Dynamite	0.029253
94	country_txt_United Kingdom	0.028956
119	targtype1_txt_Business	0.025708
129	targtype1_txt_Police	0.024205
66	country_txt_Philippines	0.024016
103	attacktype1_txt_Assassination	0.022104
29	country_txt_El Salvador	0.020522
123	targtype1_txt_Government (General)	0.019743
126	targtype1_txt_Military	0.019025
99	country_txt_Yemen	0.017848
130	targtype1_txt_Private Citizens & Property	0.017656

What factors could predict the chance of a city getting attacked again in 30 days?

The choice of weapon?



The country?



The type of attack?



The target of the attack?



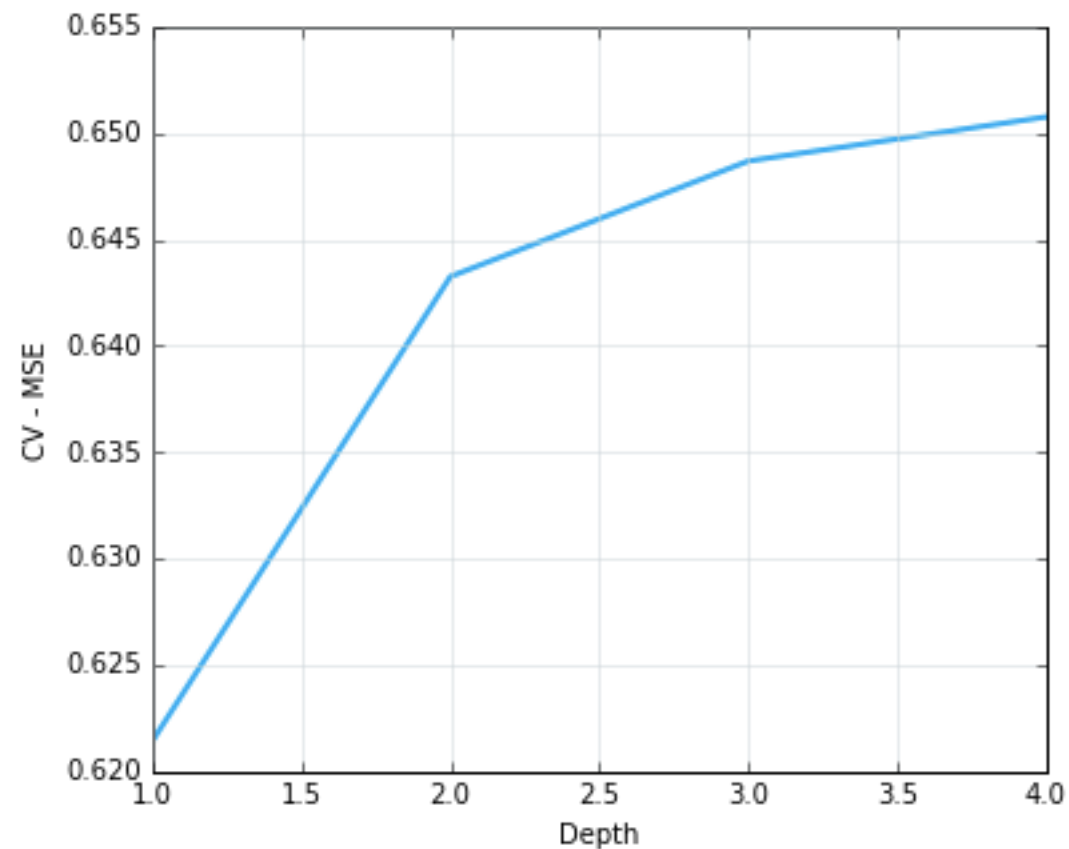
The group responsible?

Current Best = 0.6501

Bachelor #4

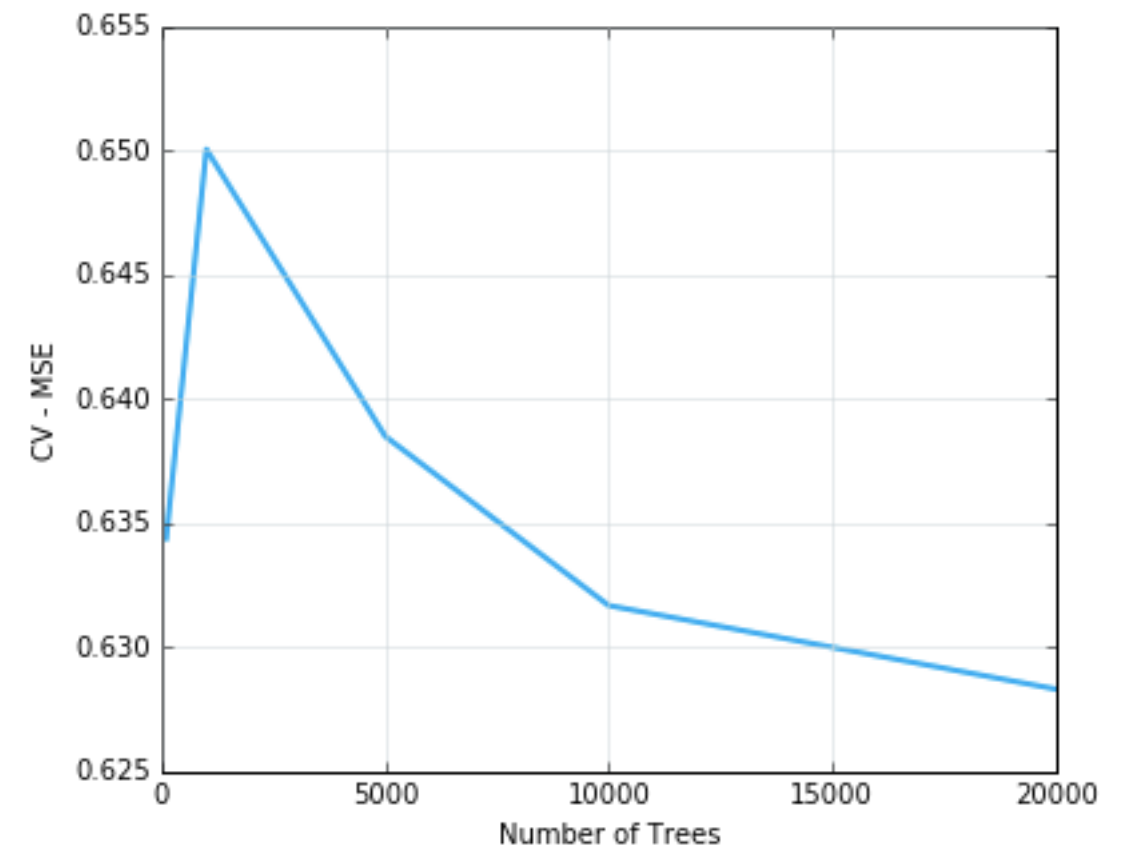
Boosted Decision Trees

Tuning Max Depth



Max Depth = 4
CV Score = 0.6508

Tuning NumberOfTrees



NumberOfTrees = 1000
CV Score = 0.6501

Current Best = 0.6508

Bachelor #4

Boosted Decision Trees

Key Features sorted by importance

	feature	importance
42	country_txt_Iraq	0.094209
65	country_txt_Peru	0.036883
372	gname_Unknown	0.034890
111	weaptype1_txt_Explosives/Bombs/Dynamite	0.033968
63	country_txt_Pakistan	0.033108
119	targtype1_txt_Business	0.030629
9	country_txt_Bangladesh	0.029909
19	country_txt_Chile	0.028017
103	attacktype1_txt_Assassination	0.027456
124	targtype1_txt_Journalists & Media	0.027214
49	country_txt_Lebanon	0.027184
78	country_txt_South Africa	0.026201
51	country_txt_Libya	0.025487
99	country_txt_Yemen	0.025143
94	country_txt_United Kingdom	0.025053
29	country_txt_El Salvador	0.023472

What factors could predict the chance of a city getting attacked again in 30 days?

The choice of weapon?



The country?



The type of attack?



The target of the attack?



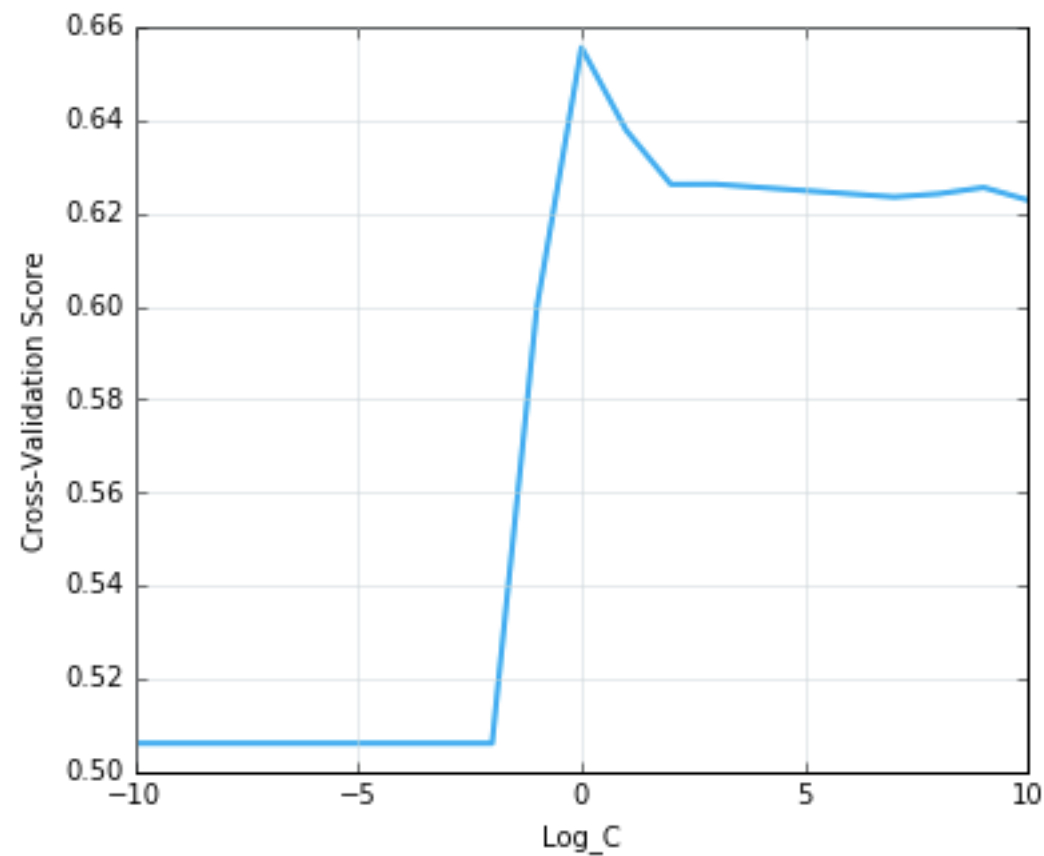
The group responsible?

Current Best = 0.6508

Bachelor #5

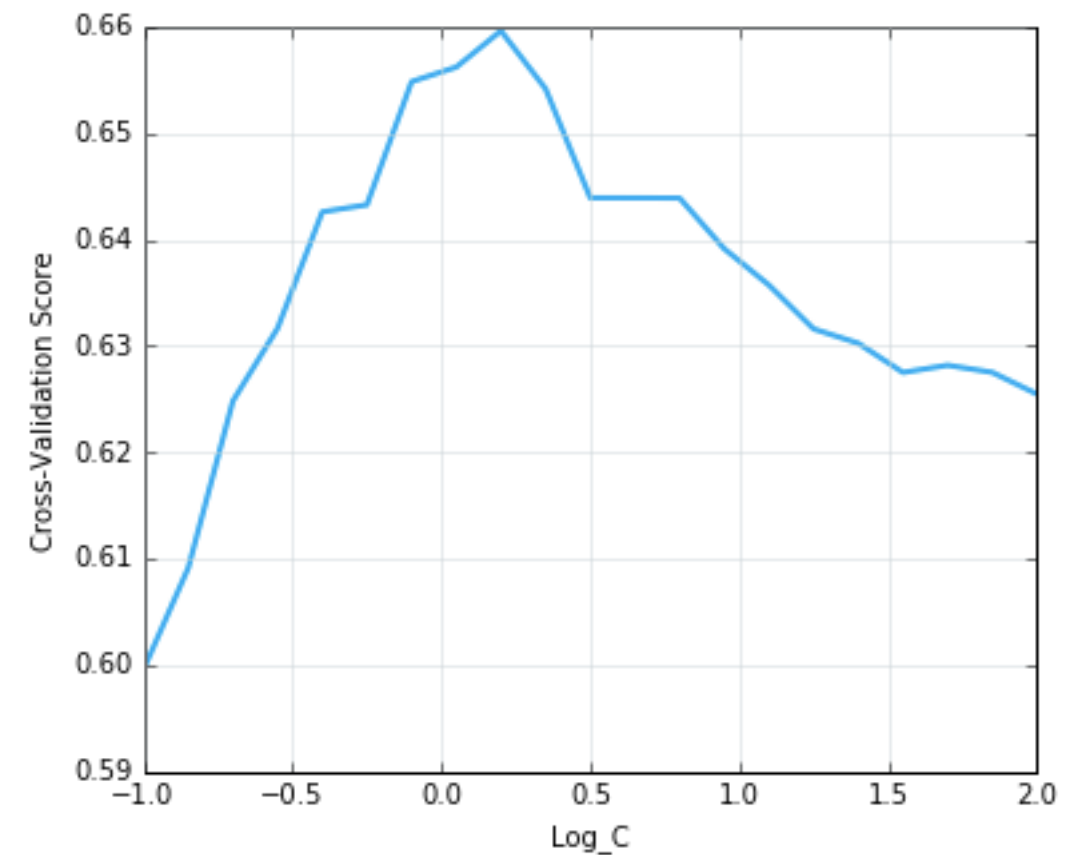
Logistic Regression

Tuning C w/range log(-10, 10, 21)



C = 1
Accuracy = 0.6555

Tuning C w/range log(-1, 2, 21)



C = 1.5849
Accuracy = 0.6596

Current Best = 0.6596

Bachelor #5

Logistic Regression

The model gives us a list of important variables, but we can improve the list

```
var_coef = zip(lm.coef_[0, :] ** 2, X.columns.values)
sorted(var_coef, reverse=True)
```

```
3.7952 'country_txt_Iraq',
2.4830 'country_txt_Myanmar',
2.2109 'gname_Dev Sol',
1.9578 'country_txt_Chile',
[...]
```

Let's focus on the 30 coefficients with the highest absolute values

```
vars_to_focus_on = [x[1] for x in sorted(var_coef)[len(var_coef) - 30:]]
X_updated = data[vars_to_focus_on]
```


Current Best = 0.6596

Bachelor #5

Logistic Regression

The updated coefficients

2.0813 'country_txt_Iraq',
1.7988 'gname_Dev Sol',
1.4421 'country_txt_Chile',
1.1056 'country_txt_Libya',
1.0717 'country_txt_El Salvador',
0.9721 'country_txt_South Africa',
0.9694 'country_txt_United Kingdom',
[...]
-0.765 'targettype1_txt_NGO',
-0.819 'targettype1_txt_Maritime',
-0.886 'gname_Kurdistan Workers' Party (PKK)',
-0.995 'gname_Other',
-1.099 'gname_Gunmen',
-1.455 'country_txt_Myanmar',
-1.658 'gname_African National Congress (South Africa)',

What factors could predict the chance of a city getting attacked again in 30 days?

The choice of weapon?



The country?



The type of attack?



The target of the attack?



The group responsible?



The choice of weapon?



The country?



The type of attack?



The target of the attack?



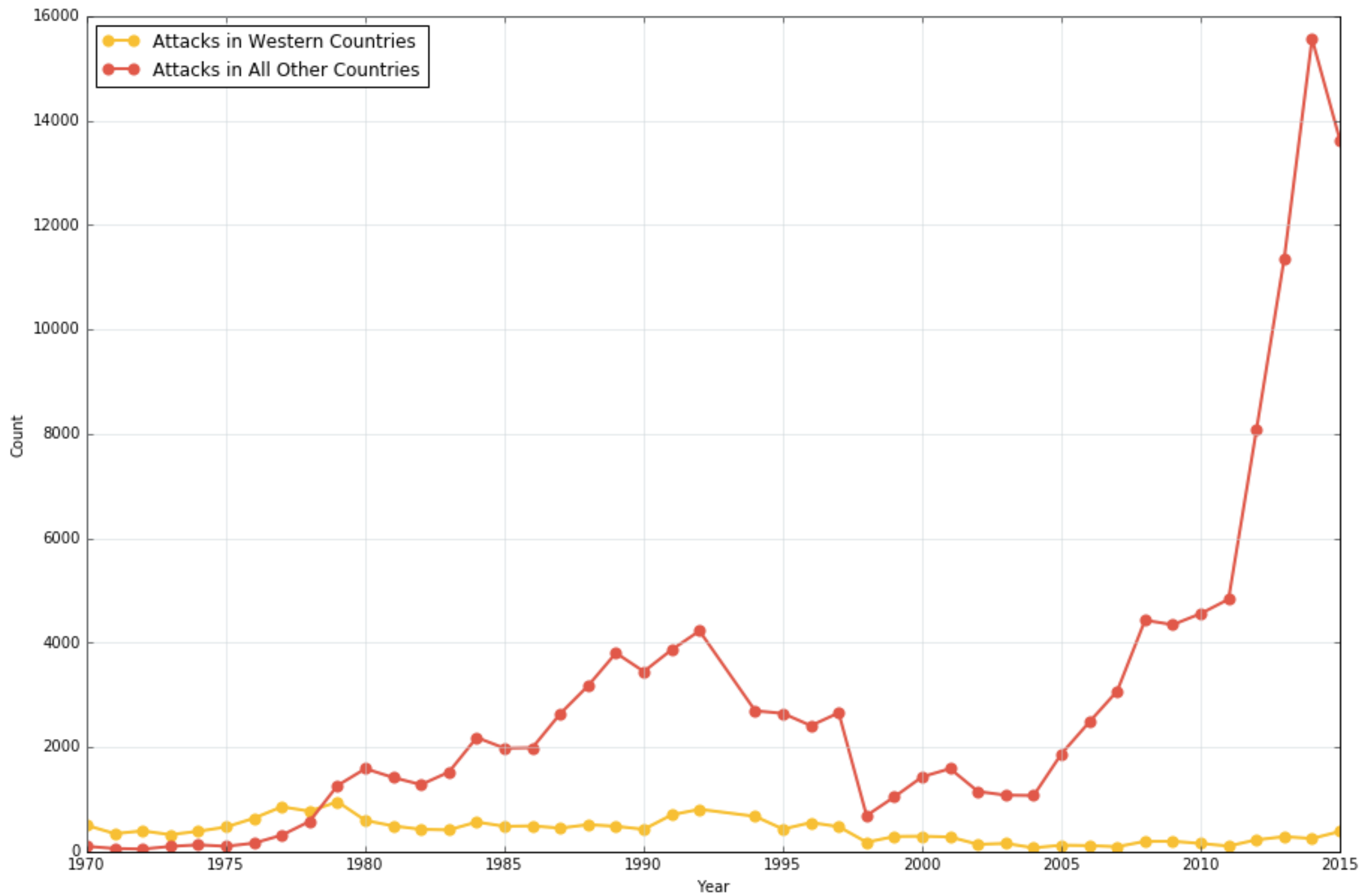
The group responsible?



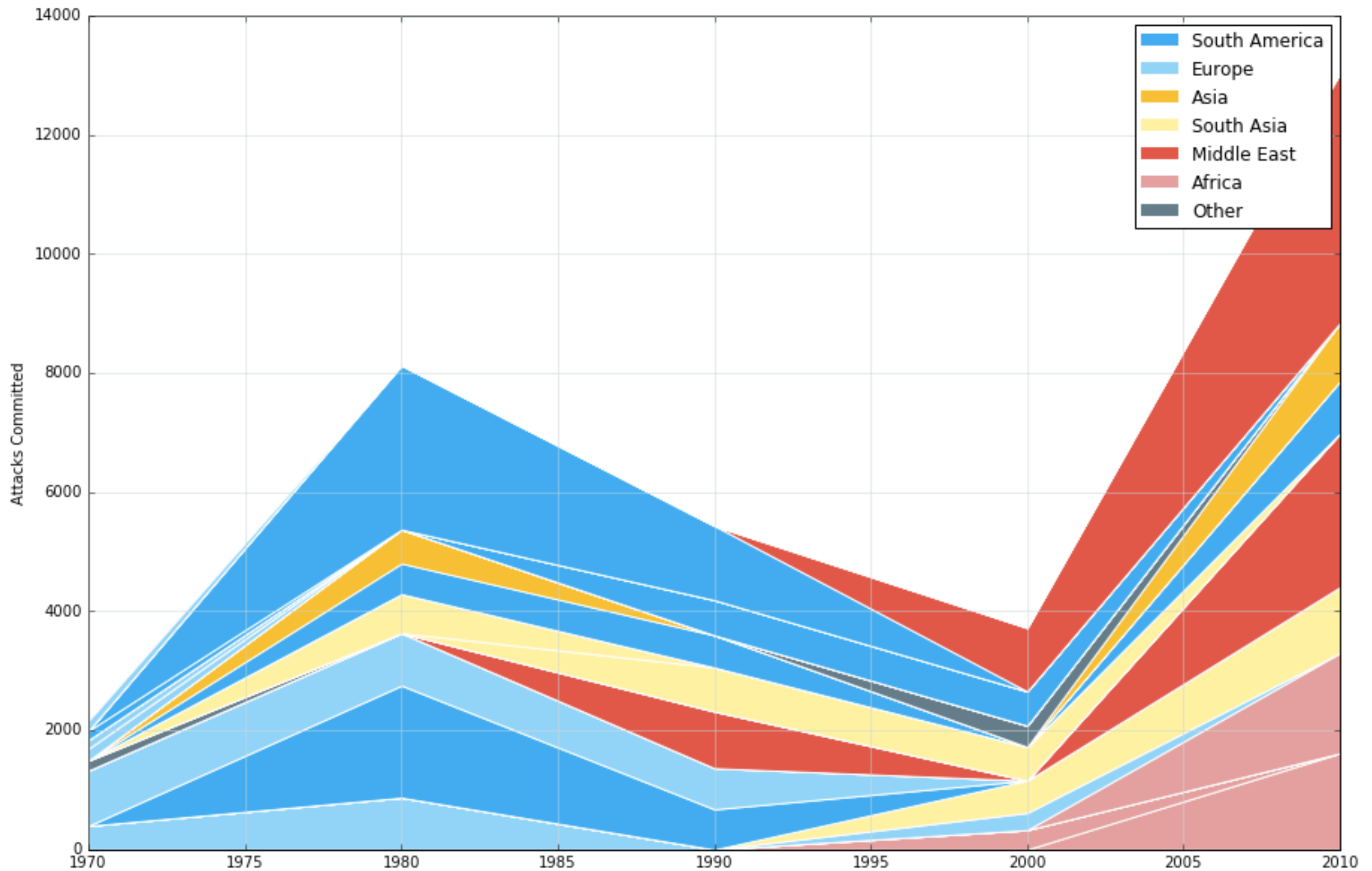
Location matters...

A lot

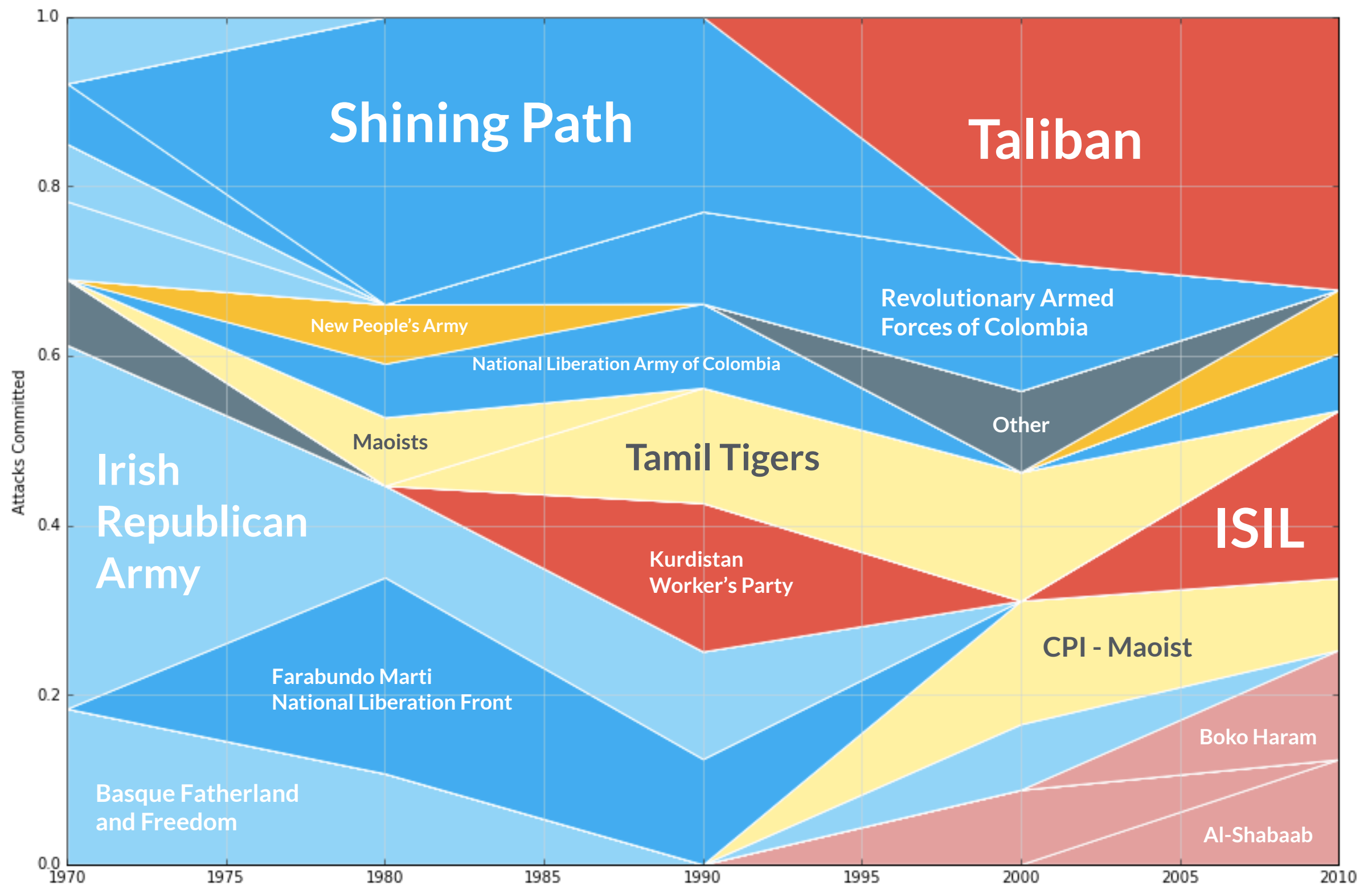
Are attacks in 'Western' countries common?



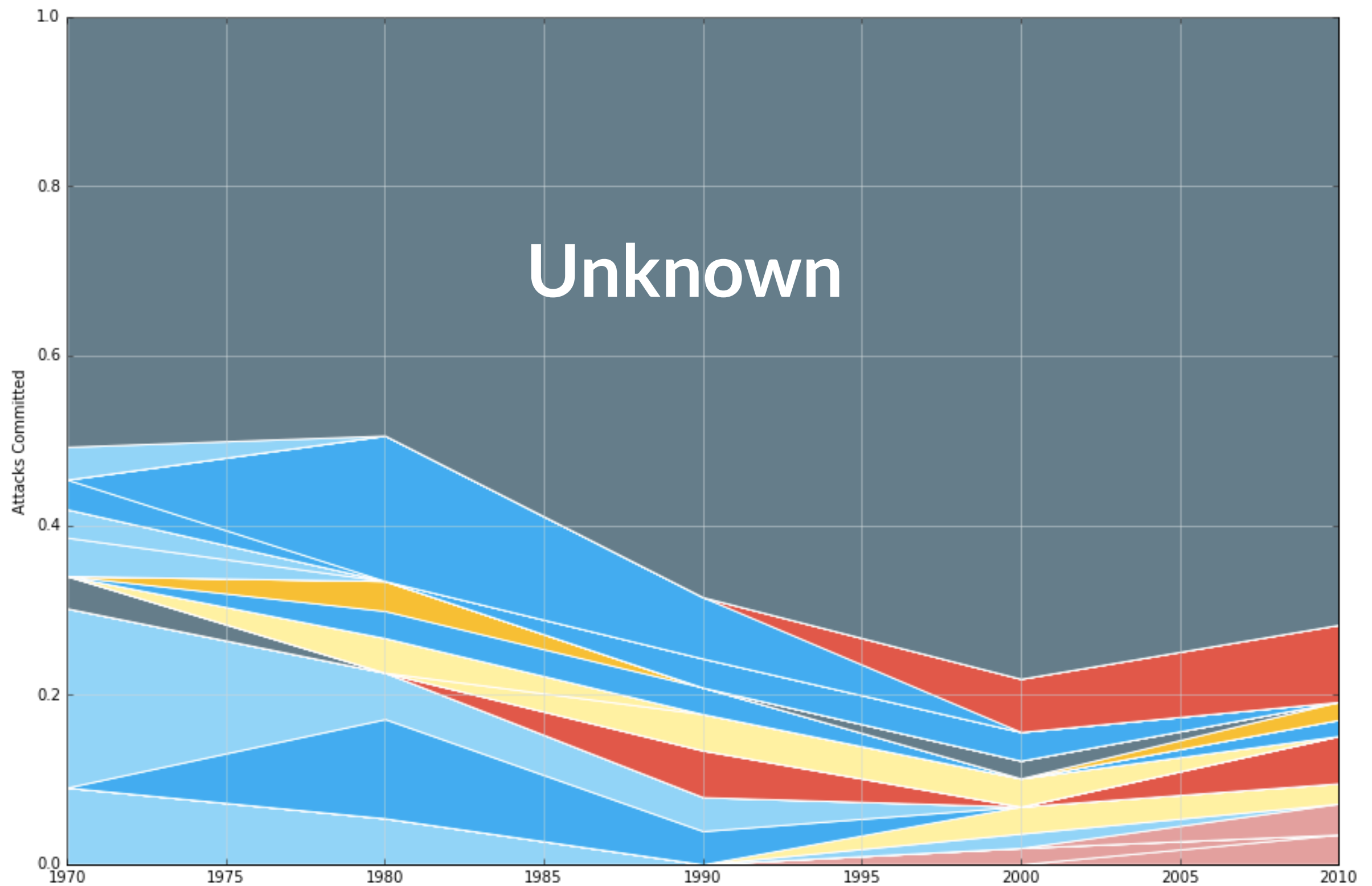
The most active terrorist groups by decade



The most active terrorist groups by decade



The most active terrorist groups by decade (inc. unknown)



What comes next?

So many possibilities!

Examine the dataset through various filters!

Only attacks resulting in **one or more fatalities**

In the ten countries with the **highest attack frequency**

Only attacks by the **thirty most active groups** of the last decade

Combine the dataset with additional political/sociological information!

The **population density** of the city

The city's **female to male ratio**

The society's **Power Distance Index**

But maybe stumbling blindly through the data isn't the best approach.

My biggest mistake was not having **stronger hypotheses** about the behavior I was interested in

I had only a surface understanding of these incidents and actors, making it difficult to decide **what to include and what to exclude** in the analysis.

My future projects will include a healthy amount of **prior research**.

Thank you!



theianchan.com/investigations