Terror Data Project

Evan Pennebaker, Cayden Dunn, Ezekiel Murphy

11/8/2018

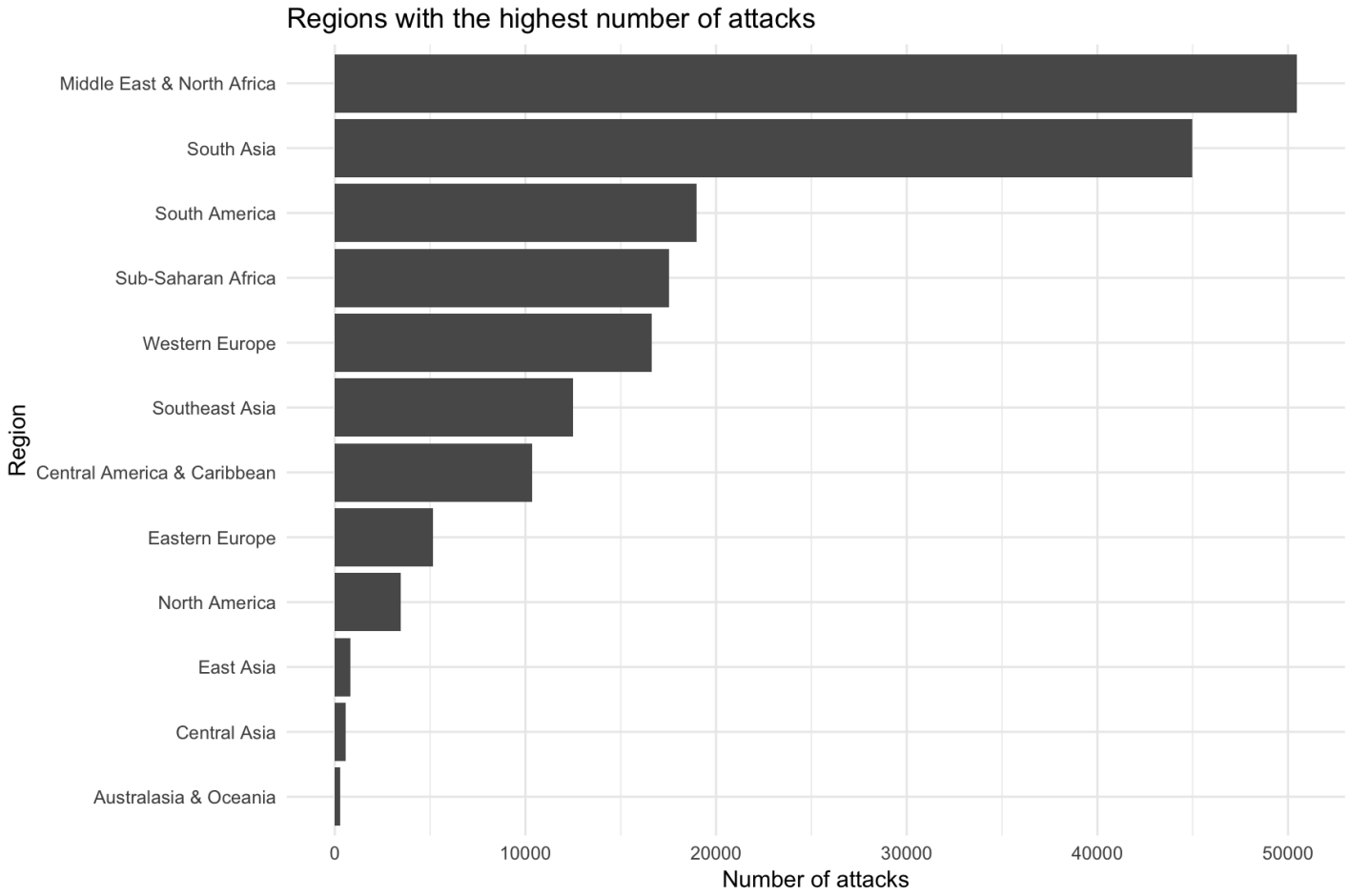
INTRO:

This markdown attempts to provide useful information for predictions on potential upcoming attacks, be it where they are or who is most likely to commit terrorist acts, this markdown can help these departments in preparing and deploying countermeasures. Its intended for anyone working for a government organization regarding national defense or overseas intelligence (DoD, CIA, FBI, Homeland Security, etc.). The Global Terrorism Database is an open-source database. The Global Terrorism Database includes systematic data on domestic as well as international terrorist incidents. Using this we attempt to answer these questions.

- Is terrorism increasing? What region is attacked the most? How are terrorist attacking us?
- How deadly is a certain type of attack? Use different features from the data set to determine the deadliness or effectiveness of attacks to help prescribe policy making
- Which terrorist group has the performed the highest amount of terrorist acts since 1970? Which of these terrorist groups is responsible for the most deaths since 1970?

Is terrorism increasing? What region is attacked the most? How are terrorist attacking us?

Which region gets attcked the most



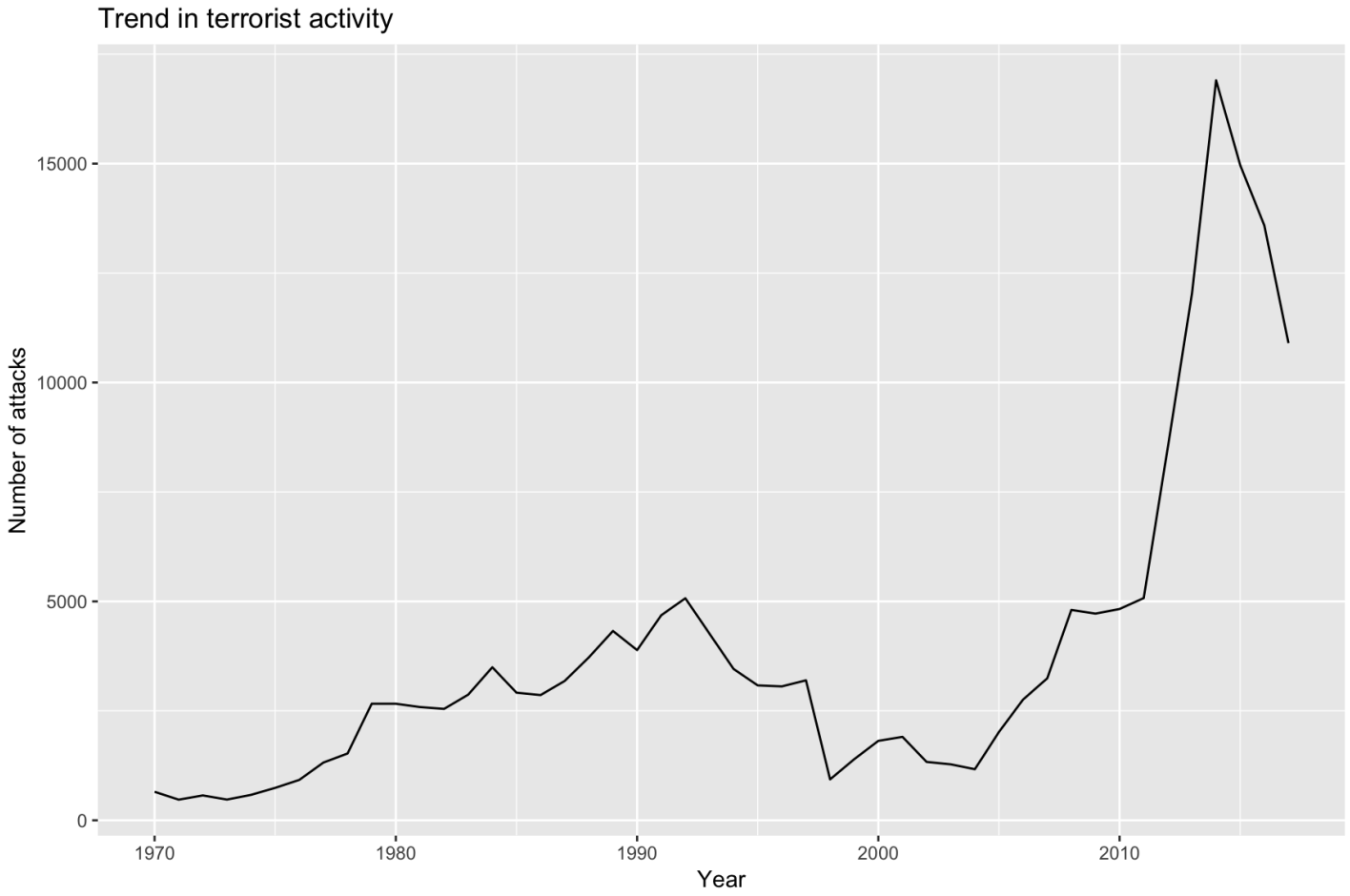
Take Away:

This graphic displays each region in the data set and orders these regions based off the number of attacks that's that have taken place since 1970 in the said region. After looking at this graphic we can conclude that the region The Middle East and North Africa is a hotbed for terrorist activity as it has more than 10,000 more attacks then the second most terrorised region of South Africa. This graphic shows which regions experience the highest concentration of terrorist activity and therefore which need the most help from the global community to rid it of this excessive terrorist activity.

Description:

1. First a data frame d1 was created out of the vector region_txt that consisted of two columns a frequency column and a region column. Both were constructed using the the table of d$region_txt with frequency column holding the frequency of the table and the region column holding the region name.
2. then using ggplot a bar graph was created that displays the highest number of attacks per region. The only technical part of this step beside basic ggplot knowledge was reordering the regions to display from greatest to least which was done with a simple reorder of region by frequency.
3. the graphic ends up displays and communicating to the viewer the frequency of attacks per region.
4. Otherwise there are only some simple ggplot features that were used to make the bar plot look nice and more effectively communicate the information displayed.

Has terrorism increased over the years



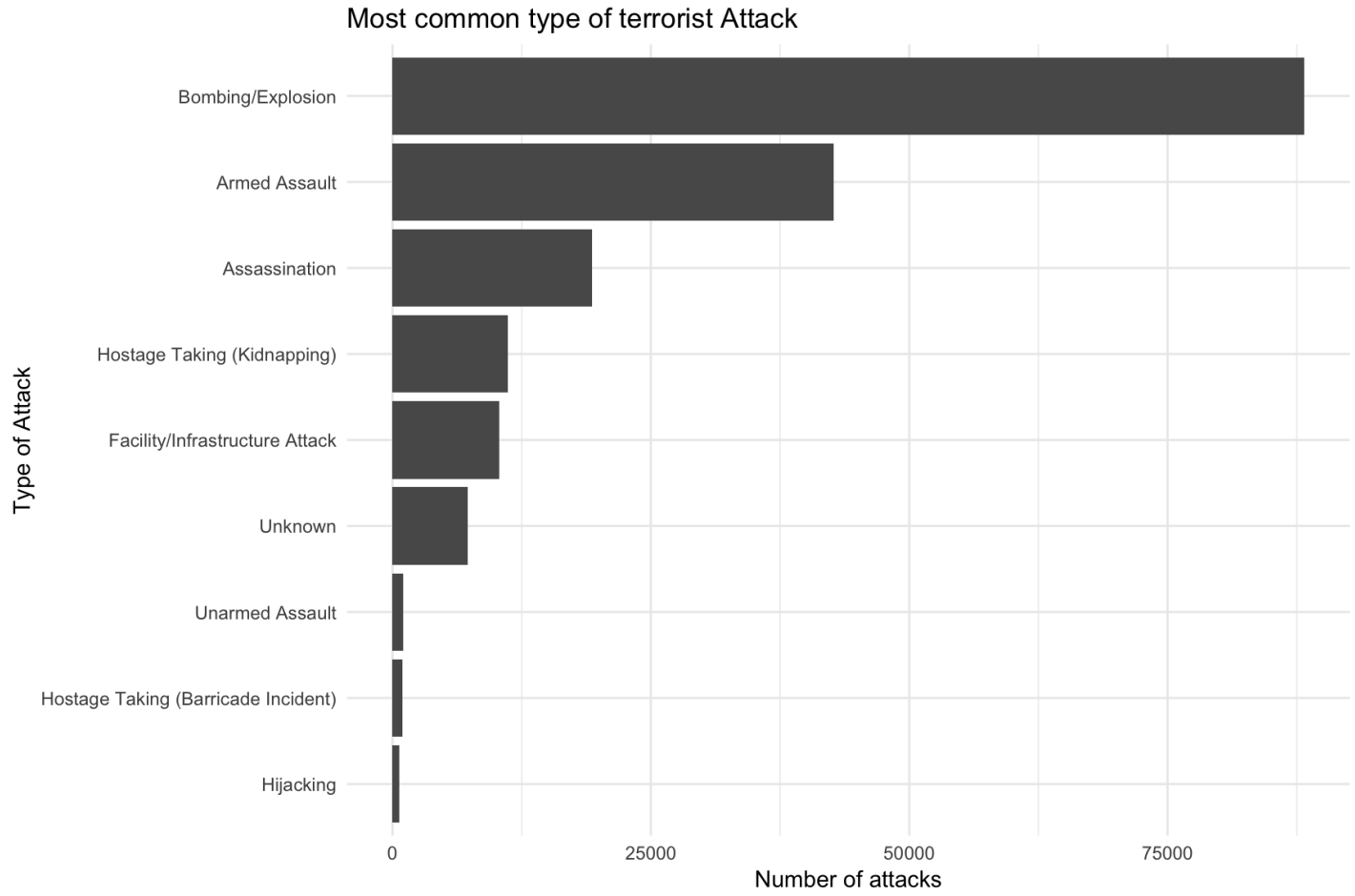
Take Away:

This graphic displays the frequency of terrorist attacks that happen per year ever since 1970. This graphic clearly displays the fact that after the year 2011 there was a significant spike in terrorist attacks. In fact after 2011 terrorist attacks nearly triple before seeing a quick decline. This graphic displays a clear spike in terrorist activity and can serve as a benchmark for how prevalent terrorist is global and how terror rates change over time as the model and data set are kept up to data as time goes on.

Description:

1. First a data frame d1 was created out of the vector iyear that consisted of two columns a frequency column and a years column. Both were constructed using the the table of d$iyear with frequency column holding the frequency of the table and the years column holding the year in which the attacks took place.
2. R was treating the years as numbers and throwing off ggplot so i treated d$iyear as numeric and this solved the problem but created another it made my tic marks on my graphic numeric so i added custom tick marks every 10 years to make the graphic communicate better to its intended audience
3. Then using ggplot i plotted using a line plot the frequency of attacks performed by terrorism per year.
4. the graphic ends up displays and communicating to the viewer the frequency of attacks per year.
5. Otherwise there are only some simple ggplot features that were used to make the bar plot look nice and more effectively communicate the information displayed.

What type of Terroist Attack is most common



Take Away:

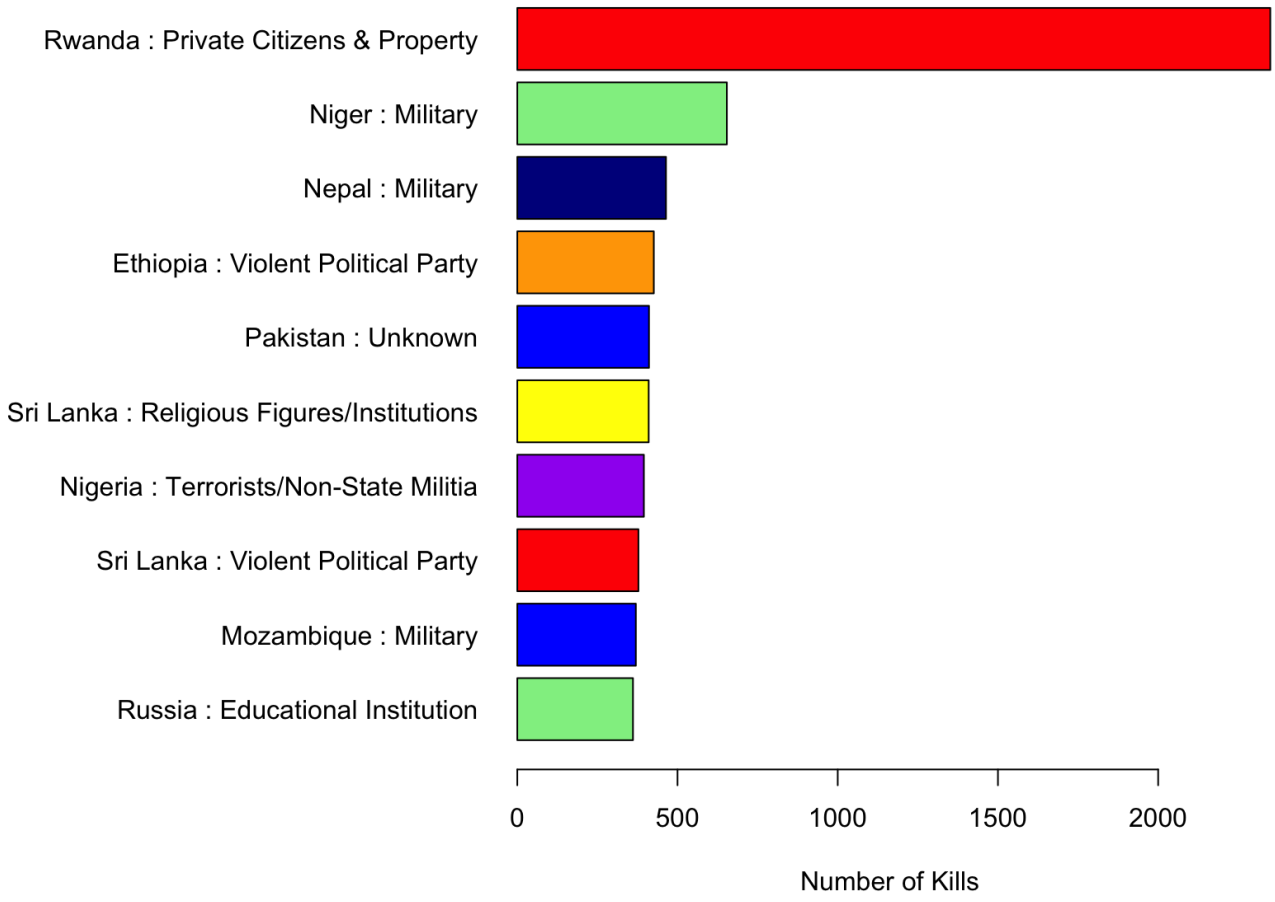
This graphic displays the frequency of a given type of terrorist attack carried out by terrorist since 1970. The graphic clearly displays that bombings and explosive attacks are the most prominent and most commonly used tool for terrorist to carry out attacks. in fact it's twice as likely than than a terrorist using any other means to carry out an attack. This graphic can assist in deterring tering terrorist attacks because if we know which method is most common for the terrorist to use to attack us we can better prepare and fortify for the said type of attack.

Description:

1. First a data frame d1 was created out of the vector attacktype1_txt that consisted of two columns a frequency column and a typ_of_atk column. Both were constructed using the the table of d$attacktype1_txt with frequency column holding the frequency of the table and the typ_of_atk column holding the typ_of_atk name.
2. then using ggplot a bar graph was created that displays the highest number of attacks per method of attacking. The only technical part of this step beside basic ggplot knowledge was reordering the regions to display from greatest to least which was done with a simple reorder of type of attack by frequency.
3. the graphic ends up displays and communicating to the viewer the frequency of attacks per method of attacking.
4. Otherwise there are only some simple ggplot features that were used to make the bar plot look nice and more effectively communicate the information displayed.

How deadly is a certain type of attack? Use different features from the data set to determine the deadliness or effectiveness of attacks to help prescribe policy making

Who is targeted the most in terrorist attacks?



Take Away:

This barplot shows the top 10 targets and which country they reside in for all terrorist attacks from 1970-2017. After looking at the graphic, we can conclude that Private Citizens & Property in Rwanda have been the highest targeted group since 1970. Also, the militaries, and institutions of the world have been higly targeted.

Description:

1. First ddply took the data frame d, and created a new data frame only looking at each country and targets, summing up the number of dead targets.
2. Then the data frame was sorted in increasing order.
3. Next, a new column was created, in the format of (“country : target”).
4. Par was used to format the barplot nicely. All NAs were omitted.
5. Then a barplot printed out the top 10 most targeted groups in the world.

Where are the clusters of attacks based off of the number of kills?

```
## [1] "The Centers for our 8 Clusters"
```

```
##         nkill    latitude     longitude
## 1 -0.14745726 -0.01402903   0.002613141
## 2  0.04552302  0.51026741   0.002540433
## 3 -0.08501721 -2.41645309   0.002122951
## 4 95.99691661  0.07907759 -81.815827012
## 5  0.02374731 -0.85237117   0.002378924
## 6 -0.17464826  0.84911701   0.002404344
## 7 -0.15865814  1.54544296   0.002326191
## 8  9.05238882 -0.21732321   0.002427452
```

```
## [1] "The 5 Attackts of Cluster 4 our Outlier Cluster"
```

```
## [1] tour of El Oregano

## [2] Group of Tutsi Refugees inside Catholic Church

## [3] Passengers and crew members on American Airlines Flight 11 and the people w
orking in the North Tower of the World Trade Center in New York City
## [4] Passengers and crew members on United Airlines Flight 175 and the people wo
rking in the South Tower of the World Trade Center in New York City
## [5] Soldiers

## 86008 Levels:  ... Zuwetina Oil Port
```

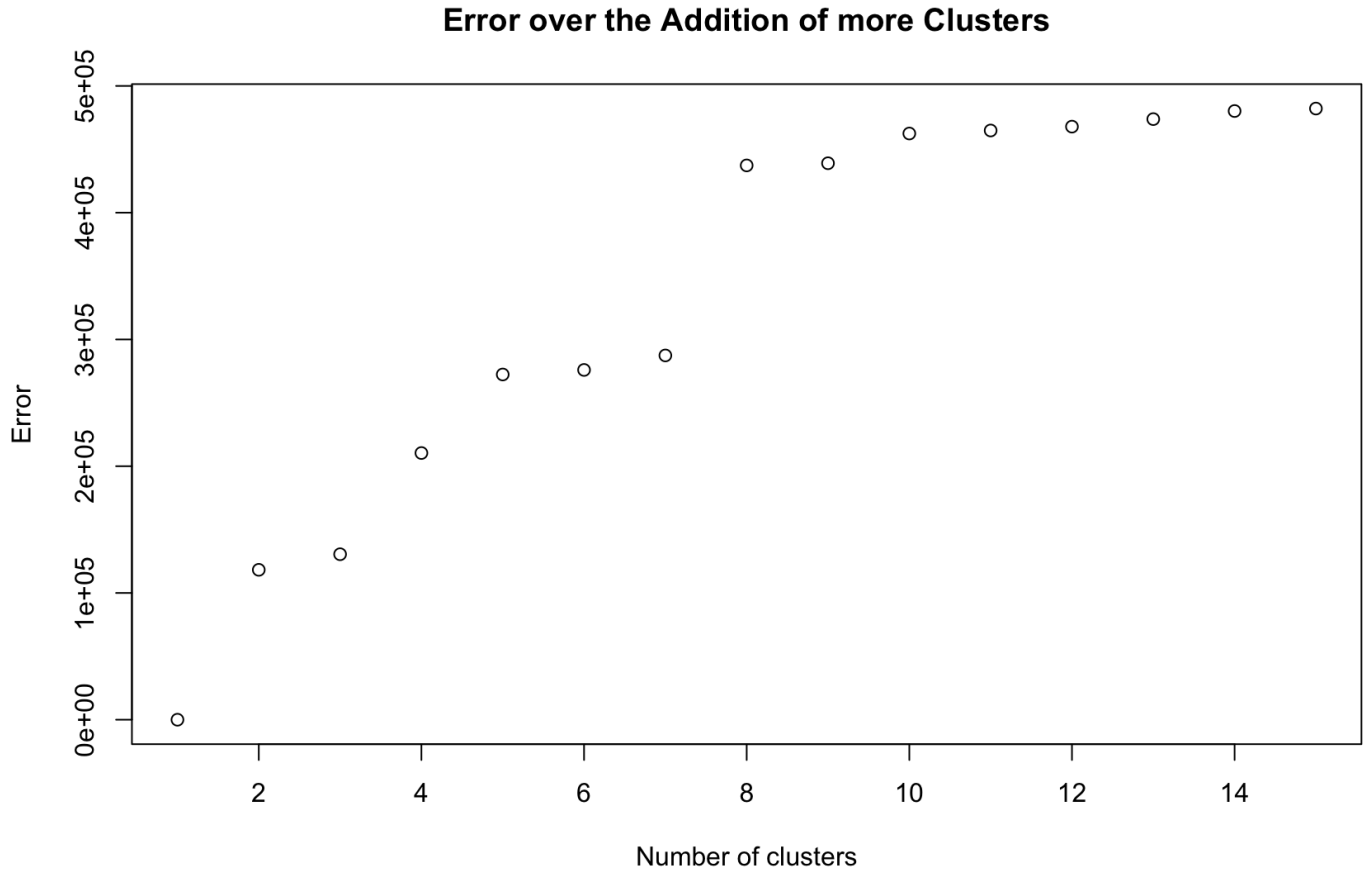
Take Away:

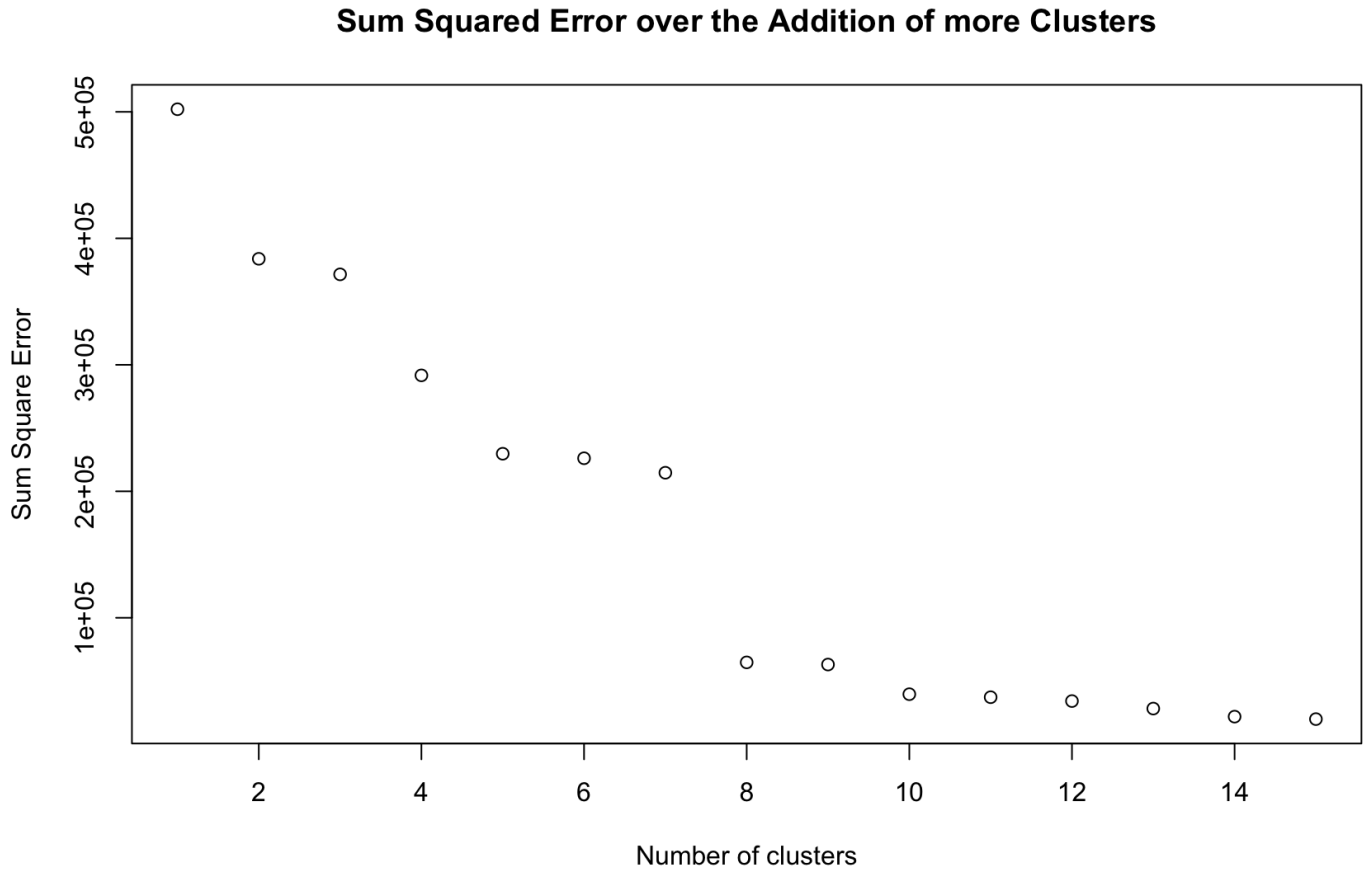
This kmeans cluster model shows that if 8 clusters were generated, these latitude and longitude points are the centers of these clusters. Also the fourth cluster only is an outliar because it only consists of 5 attacks. These attacks (2 of which are 9-11) had such a large number of deaths, that they became their own cluster.

Description:

1. First the seed was set to a specific number, 2012 in order to stop different clusters being produced each time.
2. A vector was created called cols that looked at the number of kills, latitude, and longitude for each attack.
3. Then all the NAs were omitted.
4. The kmeans cluster was created and stored as fit.
5. e represents a cluster, in this case the fourth one to show that its cluster was really small because the 9-11 attacks killed thousands.

Optimization of Model





Description of Optimization of Model:

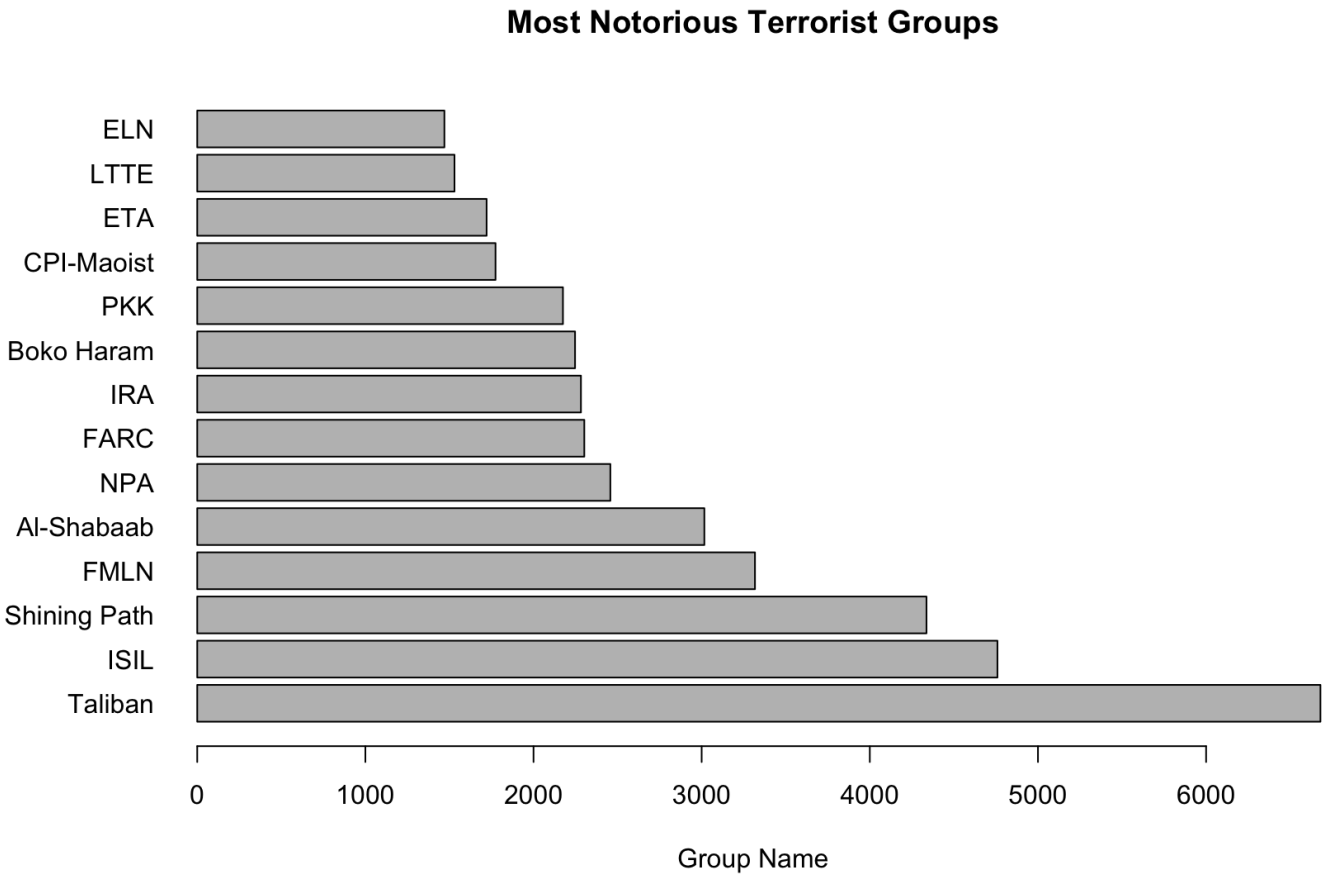
To find the optimal number of clusters to use in our kmeans model we created a for loop that ran our cluster model with an additional cluster each time and recorded the Error and Sum Square Error after the addition of a new cluster to the model. This was then plotted to show graphicly when the addition of a new cluster had little to no change on the error and Sum Squared Error. It's clear from the two graphics that after 8 clusters the addition of any more has little to no change to the Sum Squared Error and the Error numbers therefore 8 clusters is the optimal number for this kmeans model.

Which terrorist group has the performed the highest amount of terrorist acts since 1970? Which of these terrorist groups is responsible for the most deaths since 1970?

Most Notorious Terrorist Groups

Which terrorist Group has performed the most attacks since the 1970s?

According to the data, the Taliban has performed the most terrorist attacks by far since the 1970s. At the time of this dataset, the Taliban were responsible for at least 6,680 successful terrorist attacks. Right after them is the Islamic State of Iraq and the Levant (ISIL), coming in at 4,759 successful terrorist attacks, almost 2,000 less successful attacks than the Taliban. This dataset shows that the Taliban are by far the most notorious terrorist group within the last half century, with ISIL taking second place and Sendero Luminoso (Shining Path) taking 3rd place with 4,337 successful terrorist attacks.



Description:

Bar plot containing top 14 terrorist groups according to the number of successful terrorist attacks performed from 1970 to 2017.

Summary

This markdown provided useful information on current terrorism activities as well priceless knowledge on where to allocate the deployment of countermeasures. Its could vair well aid anyone working for a government organization regarding national defense or overseas intelligence (DoD, CIA, FBI, Homeland Security, etc.). The Global Terrorism Database is an open-source database and it will be updated as time progresses and these models can be fitted with the newdata set and can be used as time goes on to further aid the fight against terrorism down the road. Overall the markdown provides easily digestible answers to is pressed questions, and the fact that it's not a one time analysis makes its utility skyrocket.