Terrorist Attack Analysis in Afghanistan 2002 - 2017

Christopher Cahill

December 2, 2018

Abstract

This report was created to analyze terrorist attacks in Afghanistan starting in 2002 after United States involvement to remove the Taliban from power and hurt Al-Qaeda cells operating in the country. What began as a simple task with the military strength of the United States and its NATO allies lingered on as a growing insurgency flourished in Afghanistan. In examining terrorist attacks in the country for fifteen years, it was clear that the Taliban is more active than ever. An approach to examining the attacks in a seasonal effect also highlights how there are spikes of insurgent attacks in the summer months. While this report is exploratory in nature, it opens ideas for future studies.

The Data

The data in this report comes from The University of Maryland's Global Terrorism Database which is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The full dataset ranges from 1970 through 2017 and includes terrorist attacks throughout the world. There are a total of 93 recorded variables, including a binary response variable for successful attacks. Carefully cleaning the variables and properly coding them can allow for a logistic model to be made to find the odds ratios and what makes attacks successful.

Attacks by Terrorist Groups

In examining the terrorist attacks in Afghanistan, it was important to find the total attempted and successful attacks for each of the 27 groups to have attempted terrorist attacks during the given timeframe. Analysis was completed in R and the code chunk is provided below.

The first task was to only include relevant variables. Of the 135 variables in the original dataset, many were not needed, such as weapon details. It was also important to recode some of the days as some exact dates were unknown and left as 0 for the numerical day. For the analysis in this report, only month and year matter, but a day of zero could not be read in a date format, so all cases were changed from 0 to 1. These string were then combined to create full dates.

```r
attacks_TS<-select(attacks_AF,iyear,imonth,iday,provstate,city,gname,success)
attacks_TS$attempts<-rep(1,nrow(attacks_TS))
for(i in 1:nrow(attacks_TS)){
  if(attacks_TS$iday[i]==0){
    attacks_TS$iday[i]<-1
  }else{
    attacks_TS$iday[i]<-attacks_TS$iday[i]
  }
}
attacks_TS$date<-as.Date(paste(attacks_TS$iyear,attacks_TS$imonth,attacks_TS$iday,
                         sep='/'),'%Y/%m/%d')
attacks_TS$date<-substr(attacks_TS$date,1,7)
```

Now that dates were created for years and month, the data was aggregated over group name and date. A dataframe was created with columns being the terrorist groups and the rows being the dates. The inital dataframe was full of zeros and was then filled with total attempts and successes for each group.

Specifically, the original format of the data, shown below:

	iyear	imonth	iday	provstate	city	gname	success	attempts
73566	2002	1	1	Kabul	Kabul	Unknown	1	1
73656	2002	1	17	Kabul	Kabul	Unknown	1	1
73951	2002	4	8	Nangarhar	Jalalabad	Al-Qaida	0	1
73963	2002	4	12	Kabul	Kabul	Unknown	0	1
73972	2002	4	13	Unknown	Unknown	Taliban	0	1
73973	2002	4	13	Khost	Khost	Unknown	1	1

Was reformated to what is shown below. Only the first few columns are shown, but there were a total of 28 for each of the data frames for successful and attempted attacks.

| month	Unknown	Al-Qaida	Taliban	Hizb-I-Islami	Anti-United States extremists	Saif-ul-Muslimeen
2002-01	2	0	0	0	0	0
2002-02	0	0	0	0	0	0
2002-03	0	0	0	0	0	0
2002-04	1	0	0	0	0	0
2002-05	2	0	1	0	0	0
2002-06	0	1	0	0	0	0

After reorganizing the data into groups, the totals can be plotted. The plot below shows the timeline of all years from 2002 through 2017. What is alarming is the seemingly increased attacks, especially by Taliban forces.

Terrorist Attacks by Group
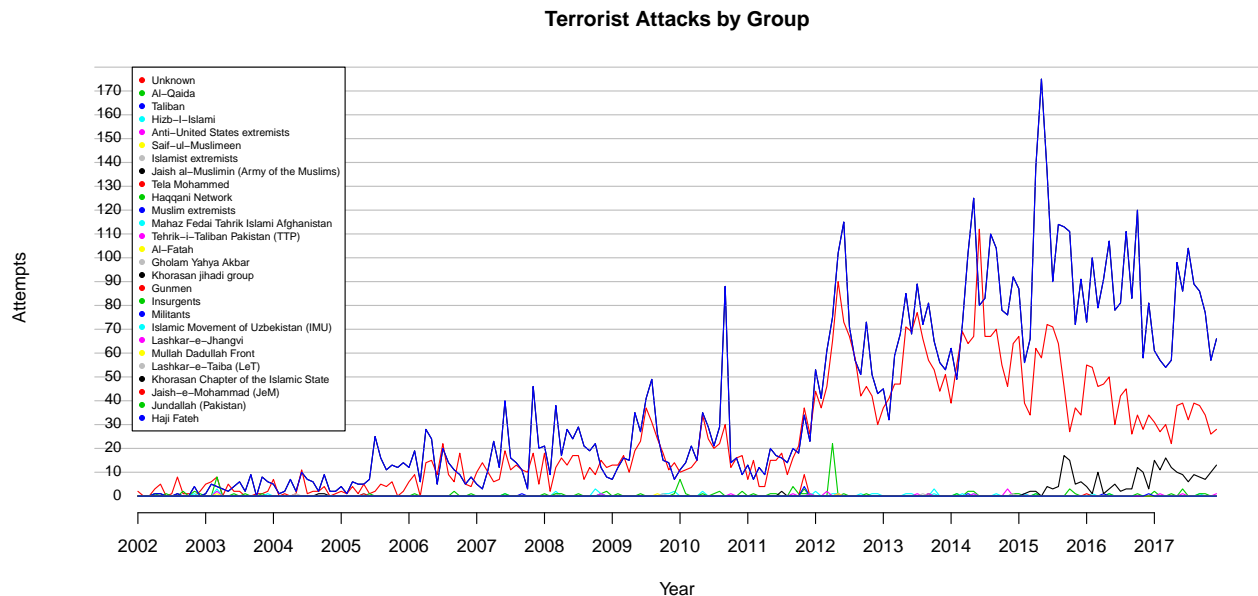
The second largest group is Unknown, but despite a spike in 2013-2014, attacks by Unknown groups have gone down. Whether this is because of smaller insurgencies being put down or the attackers being identified better is not known.

Attacks and NATO Troop Levels

In looking at terrorist attacks, it is also important to note the differences between successful attacks and failed attempts. In the plot below, all groups were combined and it can be seen that in recent years, while there have been more attempts and more successful attacks, there is a wider gap between successful attacks and attempts, indicating that there has been an improvement in preventing successful terrorist attacks.

**Total Terrorist Attacks 2002 – 2017**
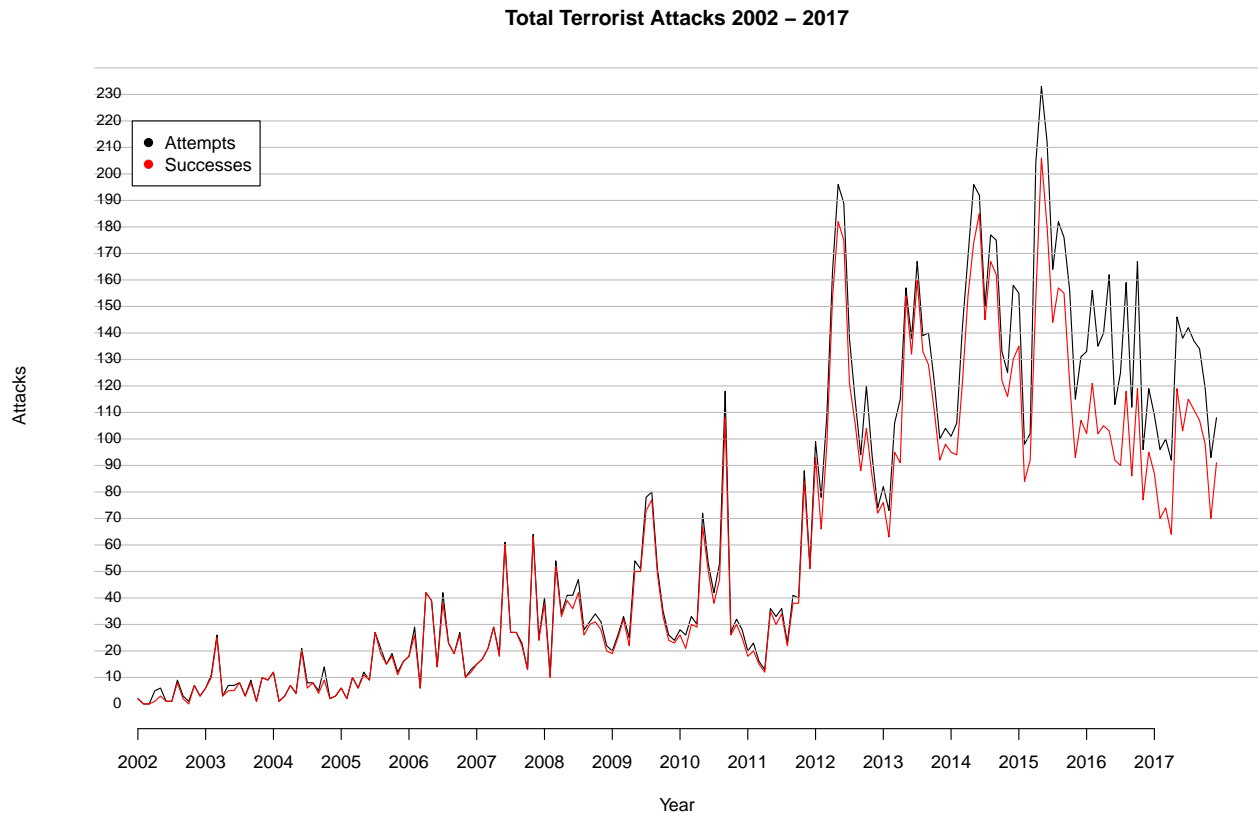
The lowest number of terrorist attacks appears to be in 2011, when NATO troop levels reached their highest at about 100,000 troops. Beginning in 2012, the slow withdrawal of combat forces began, and this seems to mirror the increased amount of terrorist attacks. This seems to indicate that increased troop levels were able to prevent attempts, but were those attacks that went through more likely to succeed?

Time Series Forecasting

In examining the terrorist attacks, it would be beneficial to note any patterns or seasonal effects. Due to the large spike in 2012, only the years 2013-2017 were examined. When examining seasonal data, it is common to separately examine pieces that appear to be naturally separated.

Total Attack Attempts 2013–2017
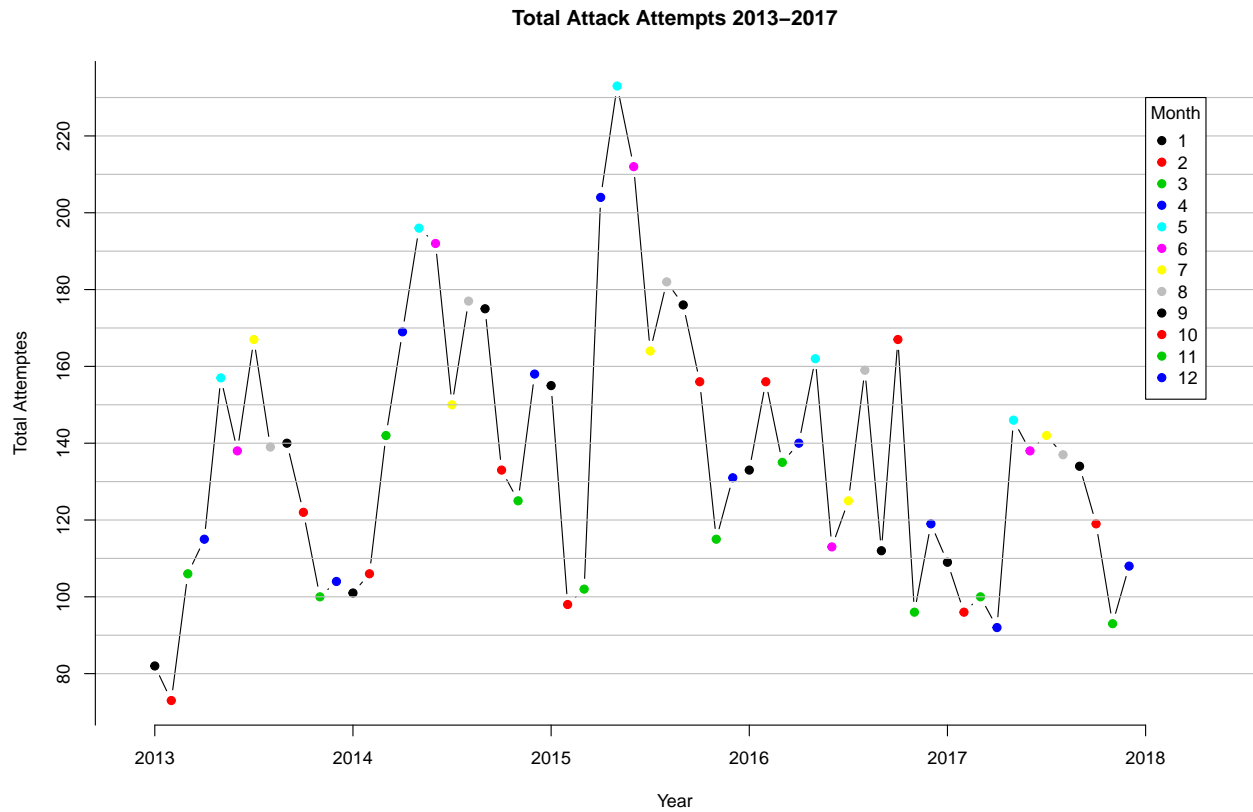
By coloring coding the months, it appears that during the summer months, there is a spike in the number of attempted attacks. A seasonal ARIMA model is evaluated and an ARIMA(1,0,0) is found to be the best model. When forecasting using the same model below, note how there is a predicted spike in the summer months. At the time of this report, the 2018 attack data was not published.
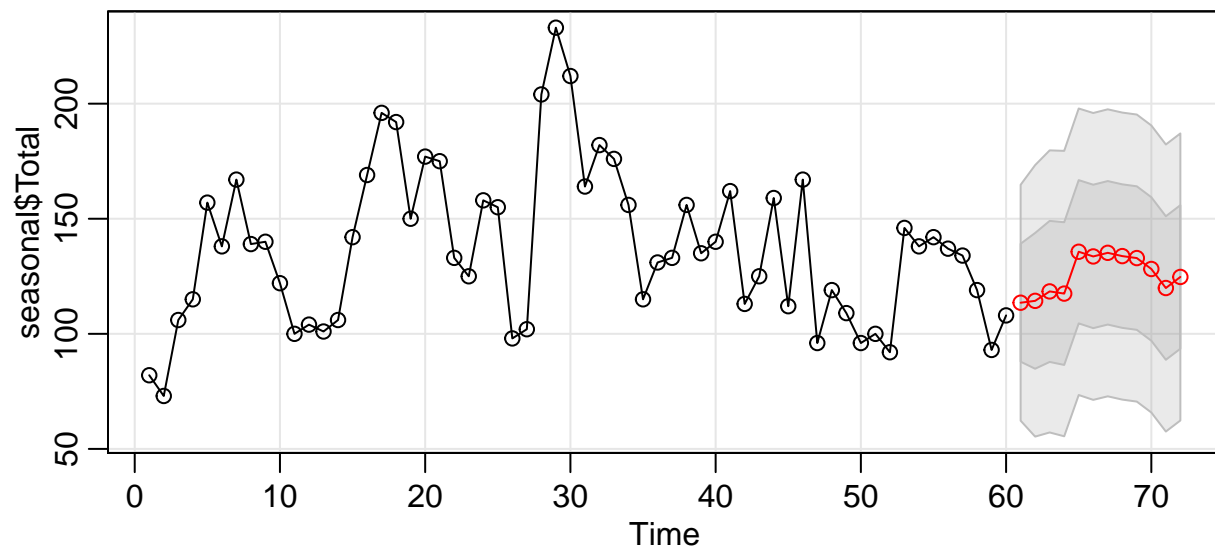
Conclusion

Since the start of the war in Afghanistan, terrorist attack attempts have risen, closely mirroring successful attacks. After a troop surge in 2011, there was a drastic reduction in terrorist attacks, but after a steady

troop redrawal, we have seen a drastic increase in terrorist attacks, most notably with the Taliban gaining more power over time.

When considering time, it can be seen that there is typically a peak in attacks during the summer months. When fitted with an ARIMA(1,0,0) model, the model shows significant values emphazising a good fit for the data. This can be used to predict future consistent attacks in the next year with a spike during the summer months.

It should be noted that since the start of this report, it has been announced that the United States will again be decreasing troop levels in Afghanistan. While these predictions may no longer be valid, it can be predicted that with this troop withdrawal, there will be increased terrorist attacks as was seen following the withdrawal in 2012. What this data can be useful in is carefully examining the variables in the data set more closely and fitting a logistic model to determine the odds of success for an attack. With decreased resources in Afghanistan, it will be more important than ever to find common factors in attacks to thwart potential threats in the future.