Analyzing and Predicting Terror Attacks in Sub-Saharan Africa:

Before diving into the dataset, I first decided to exclude columns that have little bearing on predicting a terror event, such as feature about whether there was a ransom involved. Going through the entirety of the Global Terrorism Database (GTD), I first observed the most frequent types of attacks in each of the 12 regions listed. I honed in on two—the Middle East & North Africa (MENA) and the Sub-Saharan Africa regions, which, like most of the regions, had bombings and armed assault as their highest terror occurrences. Interestingly, Sub-Saharan Africa was one of the only regions in the world from 1970 to 2015 to have armed assaults overtake bombings (the Caribbean was another region that displayed this, but I decided to observe regions with similar geopolitical dynamics, like Sub-Saharan Africa and the MENA regions).

The dataset counted incidents only each day they occurred, so, to plot their occurrences over time, I created new dataframes that had the summed occurrences of armed assaults and bombings in each region for every year from 1970-2015, and simply plotted them and the number of people killed in each region over that time. From these plots I noticed certain spikes and trends—for example, bombings in Sub-Saharan Africa spiked in the late 1980s , and armed assaults spiked in the MENA region in the early 1990s. The former was likely due to conflicts like the Lord’s Resistance Army insurgency that began in 1987, and latter being due to the first Gulf War.

Using Bayesian probability tools, I also compared two populations within Sub-Saharan Africa (Nigeria and Somalia) and saw how they compared to our prior assumption of when an armed conflict would take place in the entire region. I chose this prior as armed assault was the most common form of terrorism in the region. As expected, both Nigerian and Somalian armed assaults followed a standard distribution, but both their standard deviations demonstrated a left skew. The difference of means was also significant, conveying that our populations are significantly different. More research will be required to see whey this is.

The GTD has one major flaw, however—it has no recorded values from 1993. To solve this issue, we simply have to average the values recorded for 1992 and 1994, and assign them to our labels. I was able to group the values by month and country in Sub-Saharan Africa for 1992 and 1994, and average those values into new arrays.