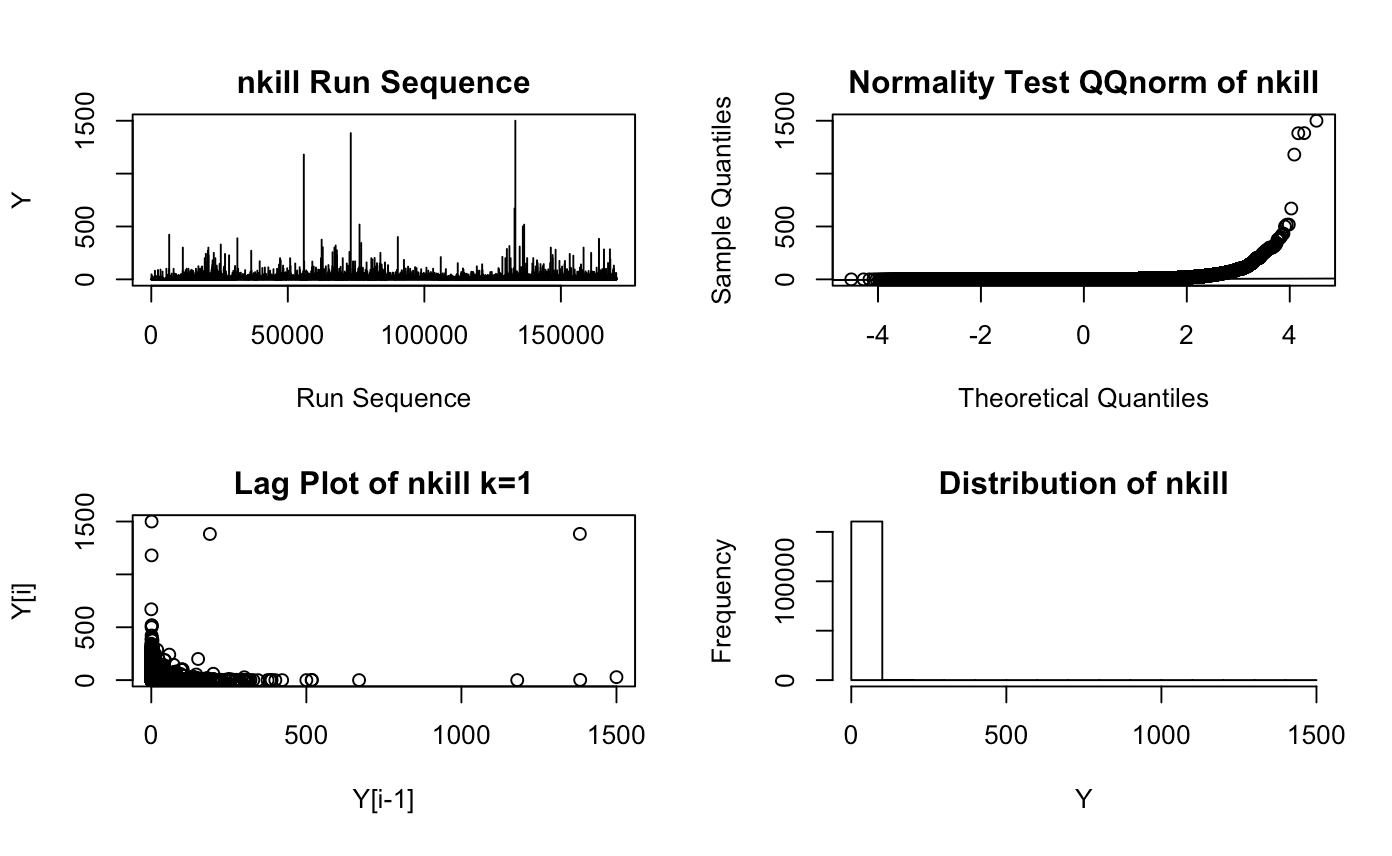
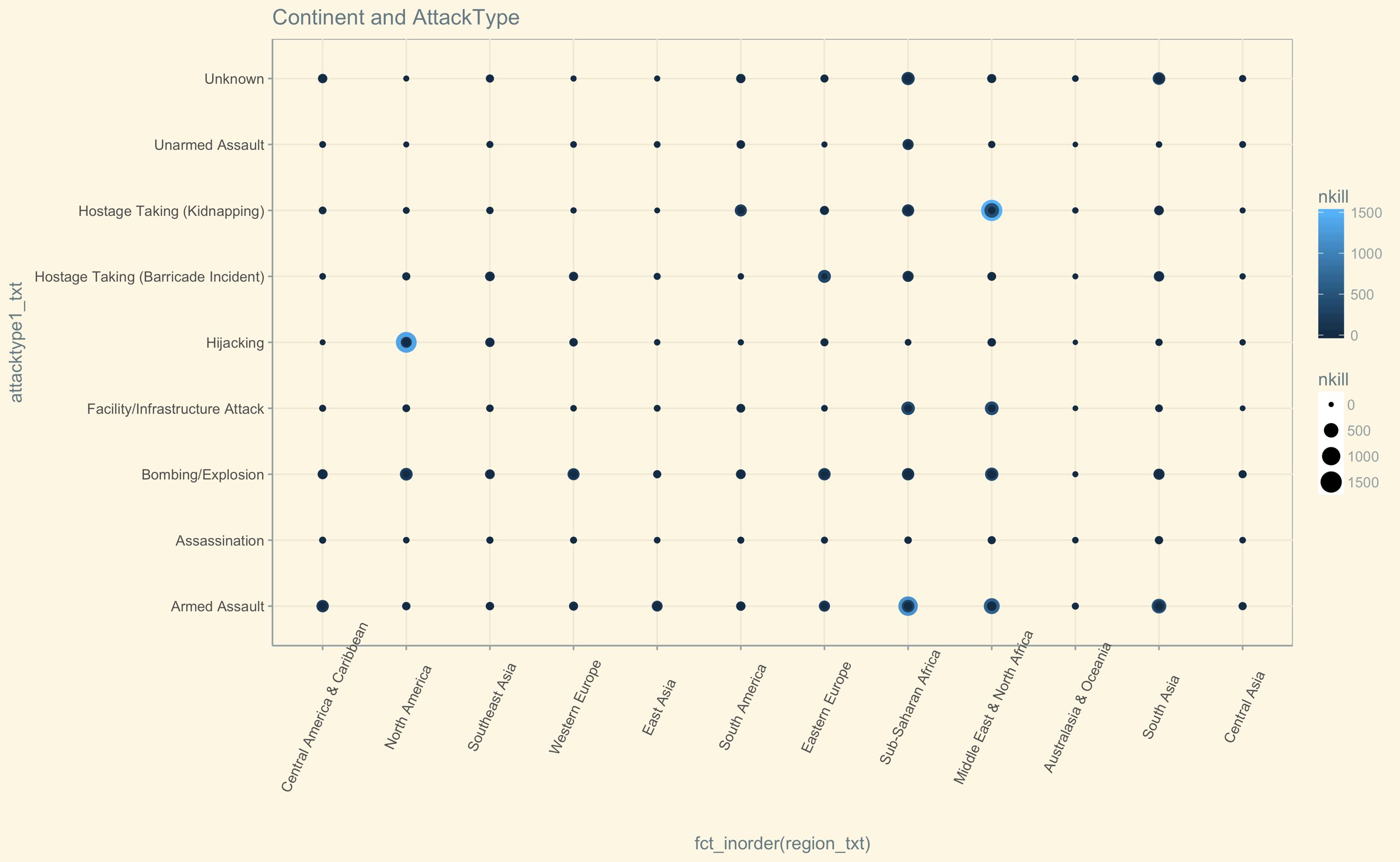
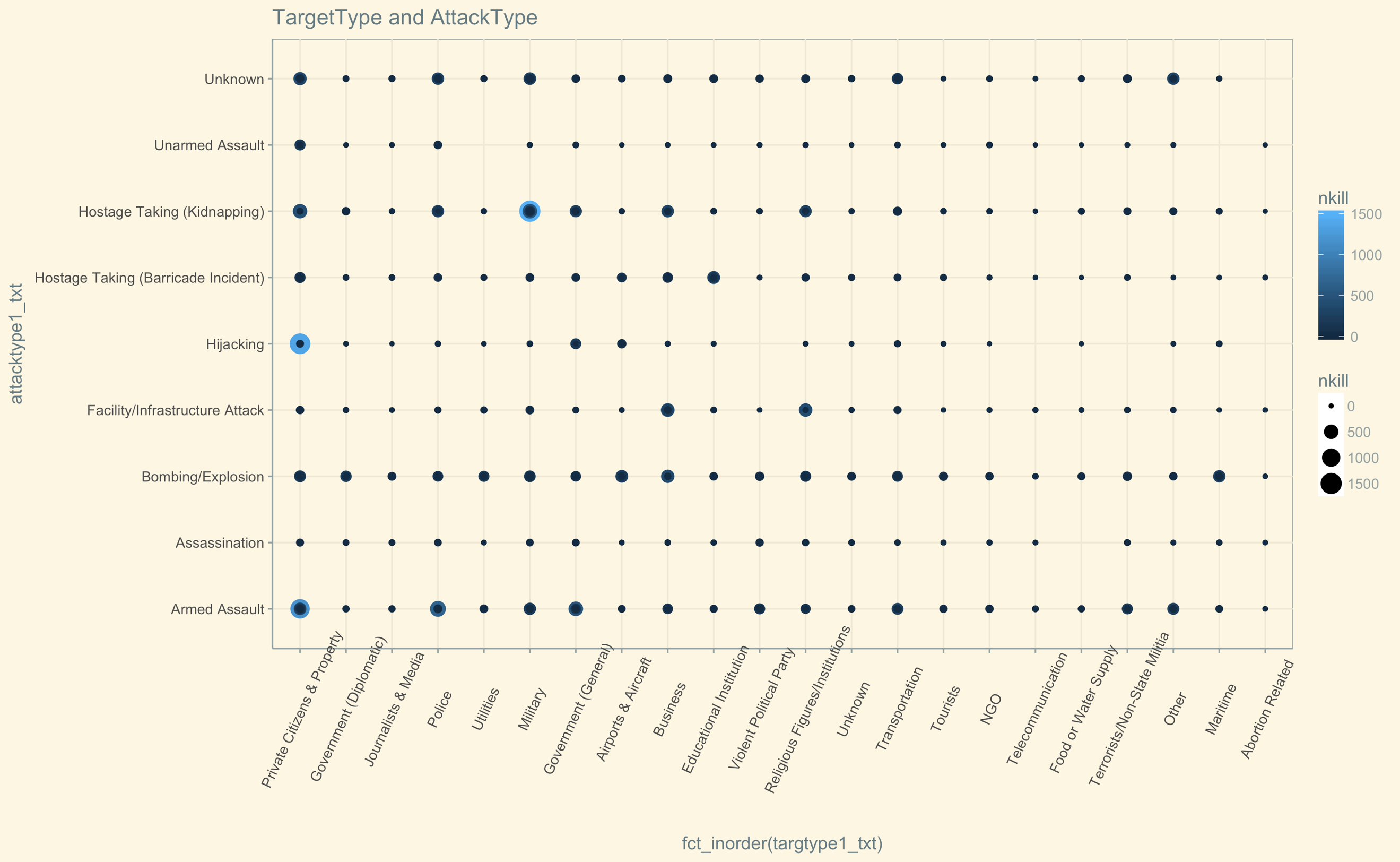
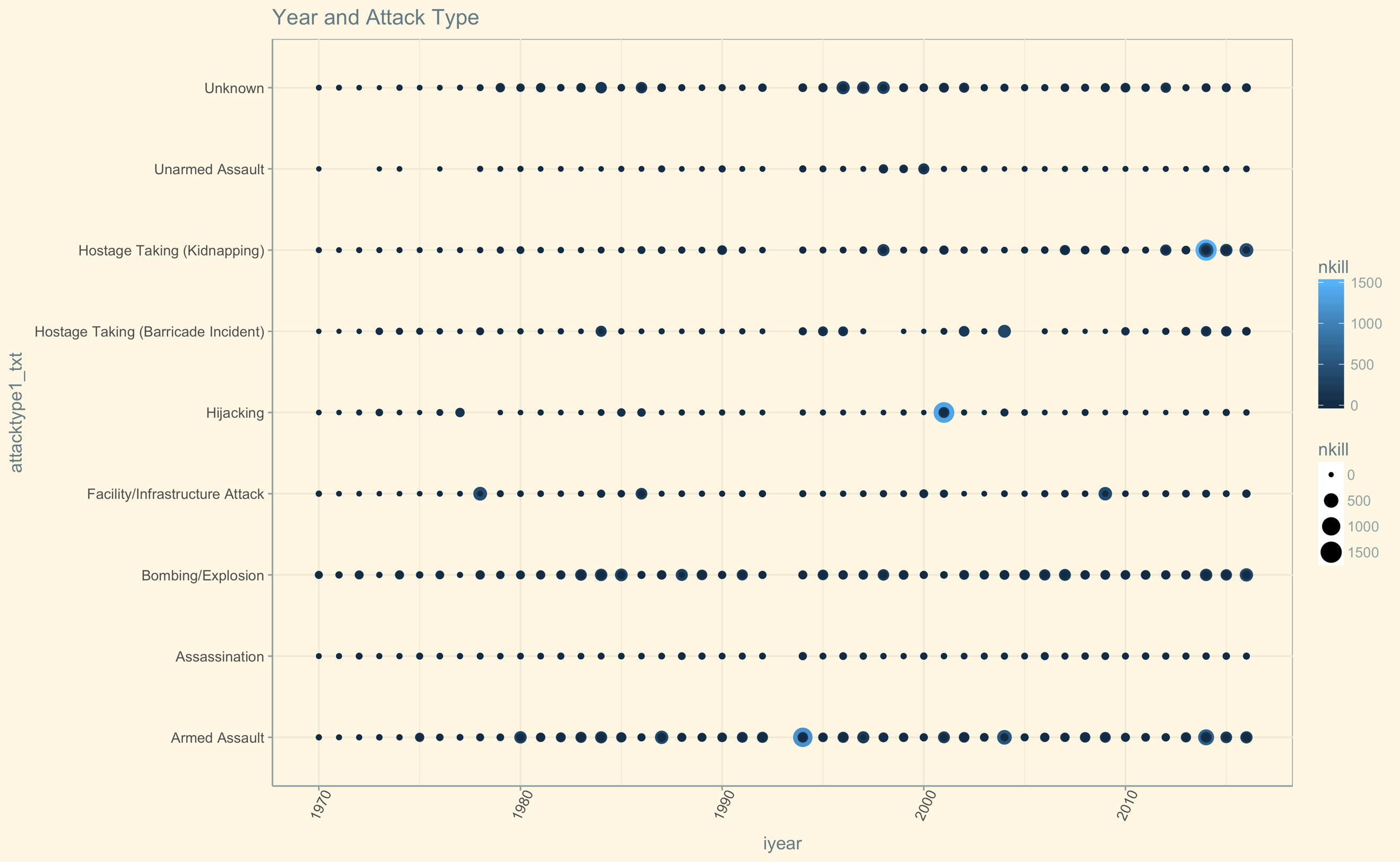
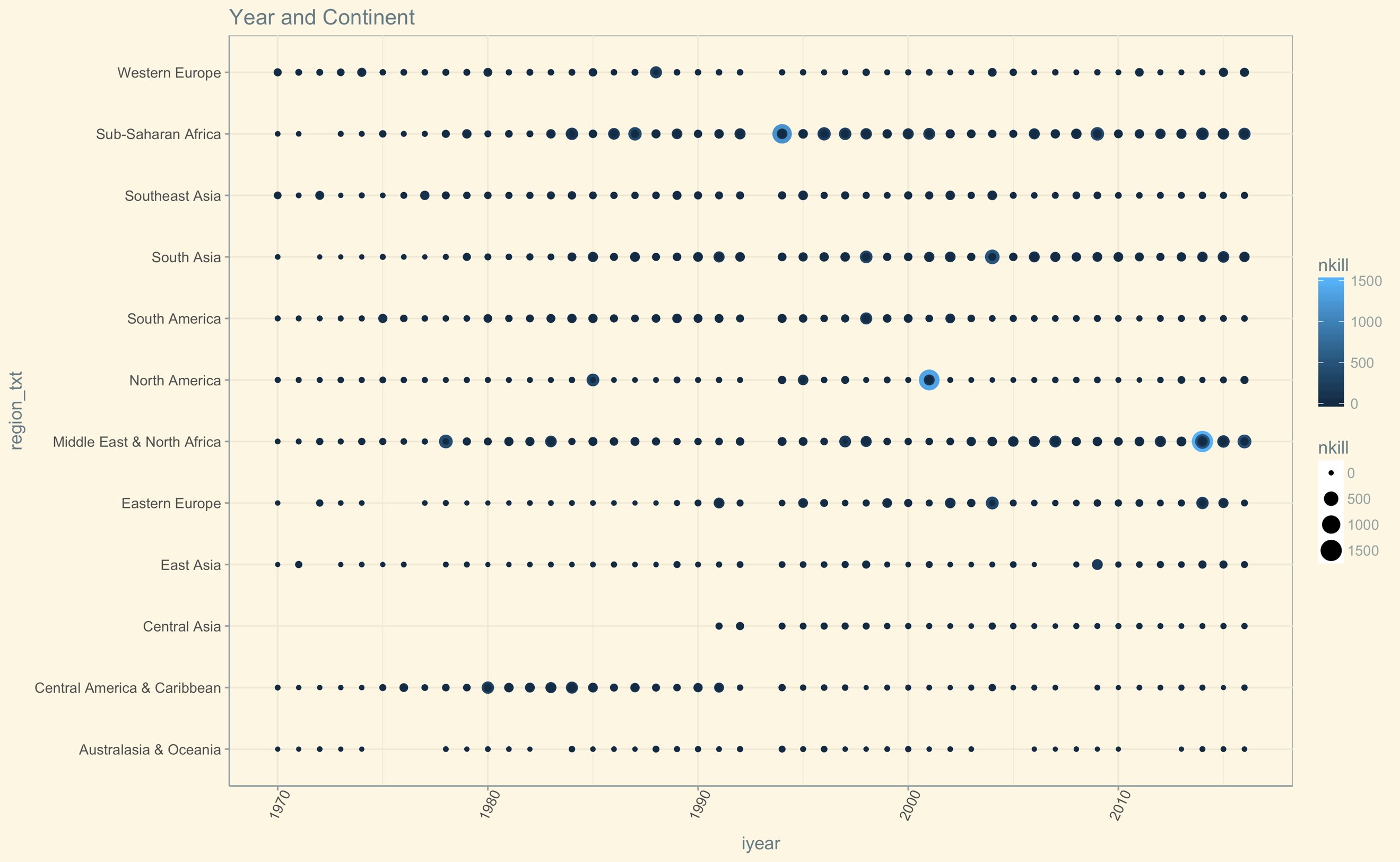
About the dataset: <http://start.umd.edu/gtd/downloads/Codebook.pdf>

Data origin: <https://www.kaggle.com/START-UMD/gtd>



EDA analysis:

This dataset of terrorism attack of the world has been collected from 1970 to 2016 and it is still updating. The dataset consists of 170350 rows and 135 rows of data. I used R with ggplot2 and investigated the data’s distribution. After ran the summary of data in R, the data owns too many attributes to investigate, so I decided to look at the area, time, attack types base and the death count.

The first graph is done by looking through the 46 years with the distribution along all the continents, and the size and color indicates the death counts (show in the legend). There appears to be three years that had serious terror attack, which is 1994 in Sub-Saharan Africa, 2001 in North America, and 2014 in Middle east & North Africa. Also, the plot shows that there is a gap year of 1993 with no data been documented.

The second graph is done by looking at the year by the attack types with the death count show as bubble size and color. The major attack types are Bombing and Armed Assault, both of these two type keeps consistent along the years. Hijacking is the type with serious fatalities which is 9/11 in the US. The top three death count in a year are Hostage Taking, Hijacking and Armed Assault.

The third graph Is taking a deeper look with the attack types and the targets with the death count as size and color. From the above plots I can take a closer look at the target types in these attacks. Hostage Taking often cause a lot of death in private property, police and military. Hijacking has a very serious incident on the private poverty. Lastly, Armed Assault cause serious fatalities on private property, Government and police.

The fourth plot take a deeper look on the continents and the attack types. With the serious fatalities on the North America, Sub-Saharan Africa, and Middle east a& North Africa.

The last graph take a look at the numeric attribute death counts, the data of nkill appear to be random. Data not normally distributed, the qqnorm plot trend is show as nolinear linear, and the histogram is distributed on the lower numbers. The lag plot is more focus on the lower counts, and the run sequence shows that there are several peaks of the death count.