Terrorist Attacks
and
the happiness of nations

By:

Juan Diaz Sada, Melissa Gonzalez, Arturo Garza, Raul Valadez

Abstract

The purpose of this paper is to investigate the relationship between terrorist attacks and countries' happiness. For this study we used data for both the Global Terrorism Database and the World Happiness Report. The Global Terrorism Database includes a list of terrorist attacks recorded since 1970. The WHR includes a list of countries and their corresponding happiness score as well as the happiness that correspond to happiness such as health, family, GDP, etc. To visualize the data better, we created a map that interactively shows terrorists attacks across the world as dots. We aimed to find the correlation between the variables that influence the happiness score of a country and the number of terrorist attacks that the country had during the same time period. We found and analyzed the correlation coefficient between all the countries at a global level, and in different clusters. Our clusters included income levels of countries according the World Bank income classification scheme, and regions of the world. After analyzing our data we came to the conclusion that there is indeed some correlation between terrorist attacks and the variables that influence the happiness score of the citizens in most regions of the world. The correlations we found were weak to moderate, so we concluded that other factors have a greater influence on the happiness scores of countries.

Data Preprocessing

The World Happiness Report is a landmark survey of the state of global happiness. The World Happiness Report 2018, which ranks 156 countries by their happiness levels. Happiness is considered to be the proper measure of social progress and the goal of public social. The happiness of each country is based on answers to the main life evaluation question asked in the poll. An example of a question asked is to rate their own current lives on that 0 to 10 scale. The sample size of 2,000 to 3,000 is large enough to give a fairly good estimate at national level this is confirmed by a 95% confidence intervals. The Global Terrorism Database is an open-source database including information of terrorist events around the world from 1970 through 2017 with updates planned for the future.

We downloaded the Global Terrorism Database and the World Happiness Report and used a dataframe in Jupyter Notebook to represent all of the data. We found the amount of NaN variables in each column and used that in our decision for choosing the columns with the most relevant information. The data set included a vast amount of data columns that are not necessarily useful for our research, after careful consideration and analysis of our various data variables we have selected the following columns as the most relevant to our research:

Global Terrorism Database:

1. Country: Country where the terrorist attack happened.
2. Region: Region of the world where the attack happened.

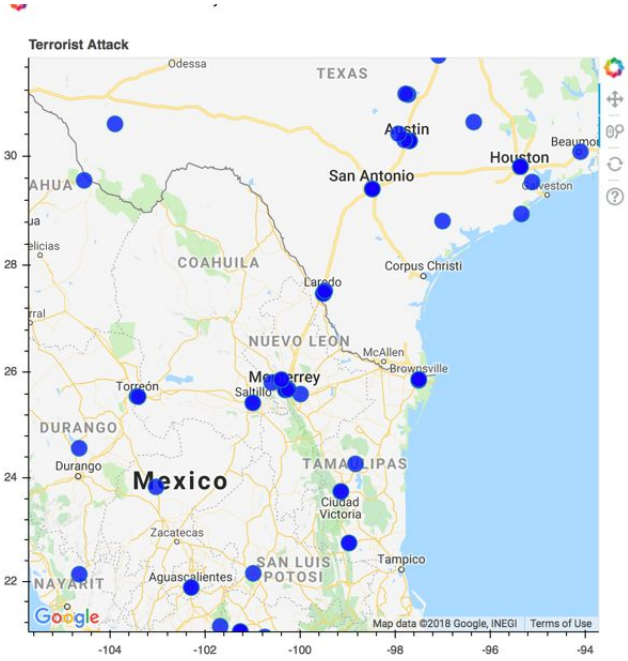
3. N. wounds: Number of wounded people in the terrorism attack

World happiness report:

1. Country: Name of the country.
2. Region: Region the country belongs to
3. Economy: The extent to which GDP contributes to the calculation of the Happiness Score
4. Freedom: The extent to which Freedom contributed to the calculation of the Happiness Score.
5. Health: The extent to which Life expectancy contributed to the calculation of the Happiness Score
6. Trust (Government Corruption): The extent to which Perception of Corruption contributes to Happiness Score.
7. Generosity: The extent to which Generosity contributed to the calculation of the Happiness Score.
8. Happiness Score: A metric measured by asking the sampled people the question: "How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest."
9. Happiness Rank: Rank of the country based on the Happiness Score.
10. Family: The extent to which Family contributes to the calculation of the Happiness Score
11. Dystopia Residual: The extent to which Dystopia Residual contributed to the calculation of the Happiness Score.

Visualization: Interactive Map of Terrorist Attacks

In order to better visualize our data, we used the latitude and longitude columns from the terrorism data set and integrated them into a map of the world using a Google Maps API. Our map interactively shows terrorists attacks across the world as dots. Here is an example of how the map looks when zoomed in to the South Texas/Northern Mexico area:



Terrorism vs Happiness in the world

The first correlation we aimed to find was between the variables in the happiness report which contribute to a country's overall happiness score and the number of terrorist attacks recorded in that country. Because our terrorist dataset contains each individual attack, we had to count the number of attacks in each country and extract them into a separate data frame. We then concatenated our information with the happiness score of each country that comes from the happiness data set using the country as the index of our new data frame. Since the terrorism database contains values for years 1970s to present and the Happiness Report is only spans the 2013-2017 we had to use only the relevant years in the terrorism dataset(excluding values before 2013).

We then plotted our data to see if we could visually see any linear correlation between the happiness score and the number of attacks recorded in the country. We found a pearson correlation of -0.2 between the number of terrorist attacks and happiness in countries (in the time period covered by the happiness score database).

The following heatmap represents the Pearson correlation coefficient between the number terrorist attacks in a country and all the other variables in the the happiness database.

Notably the number of attacks in countries had weak negative correlations with “Family” (-0.3), Happiness Score(-0.2), “Freedom” (-0.2). This is for all countries in the world.

After analyzing our findings, we decided that using our entire dataset including every country might be limiting us from finding other correlations that might exist.

We first starting sampling our data into random countries to try to see we could work with our data more easily that way. After varying the sizes of our samples, we realized all of those correlations were too weak to talk about. This is what brought us to thinking about stratified sampling. We used the countries’ regions in order to stratify our data into homogenous groups. More specifically, we began to consider different factors to stratify our data such as income. The reason we did not use Stratified Random Sampling is because there weren’t enough countries in each region or income to choose randomly from.

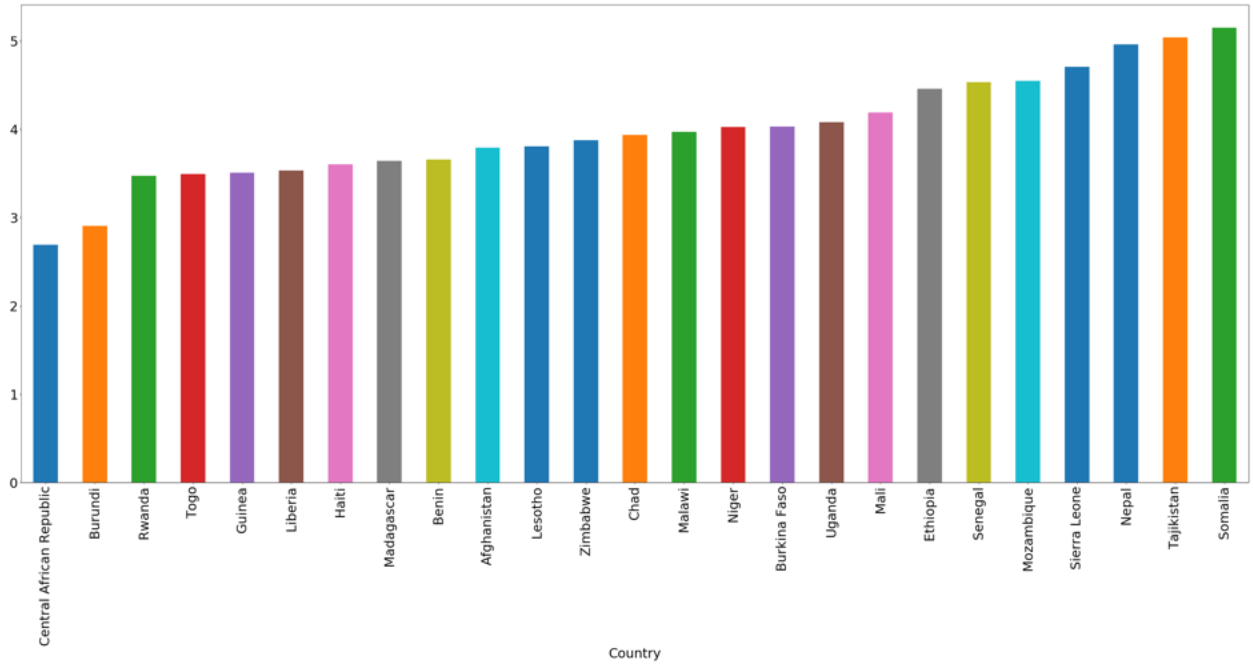
Terrorism vs Happiness and income classes

We decided to cluster our data in groups and see if we can find more specific findings. We decided to separate our data into groups of countries according to each country specific income

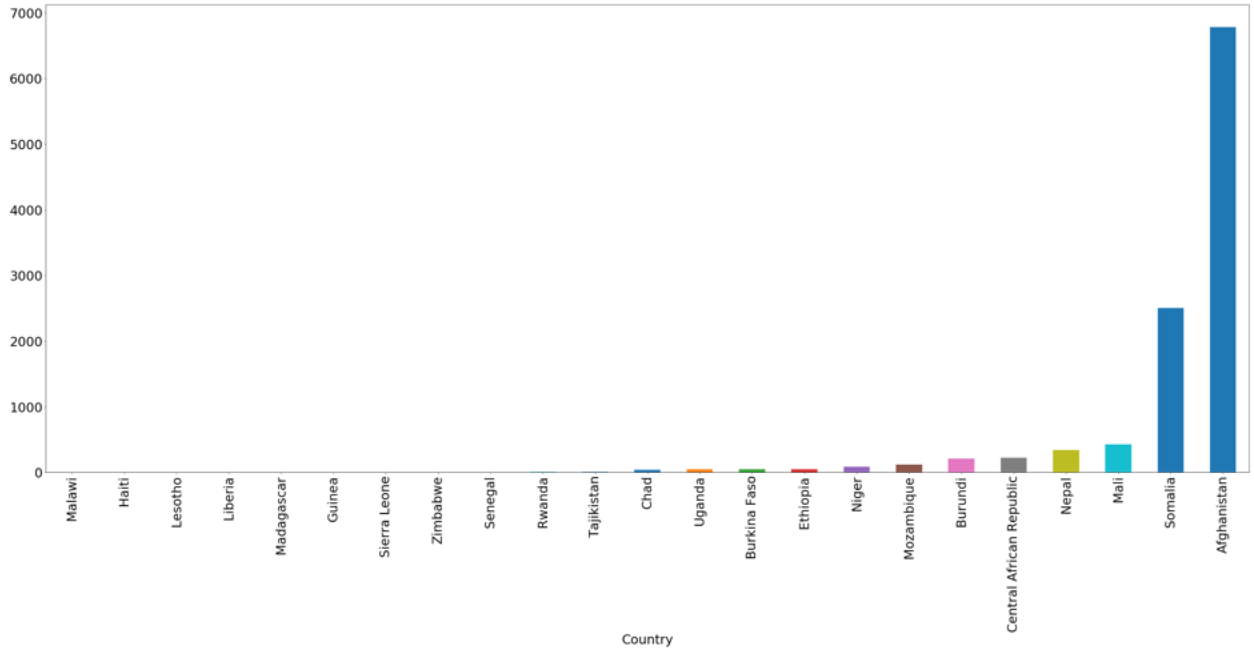
per capita. We used data from Undata.com website which lists each country's GDP per capita. We used the world bank standard for dividing countries into low-income (Less than $1,035 per capita), lower middle income(between $1,036 and $4,085 per capita), upper middle income(between $4,085 and $12,615 per capita), and high-income countries(greater than $12,615 per capita). Our reasoning behind this categorical scheme is that it would allow us to compare countries with similar development levels which are not necessarily in the same geographical region.

We matched the countries in the Terrorism database with their GDP from UNData data set. Unfortunately, some of the countries which are included in the terrorism database are not included in the happiness report so we had to omit those when we combined the data. We decided to include the graphs which represent this data in order to help visualize the analysis.

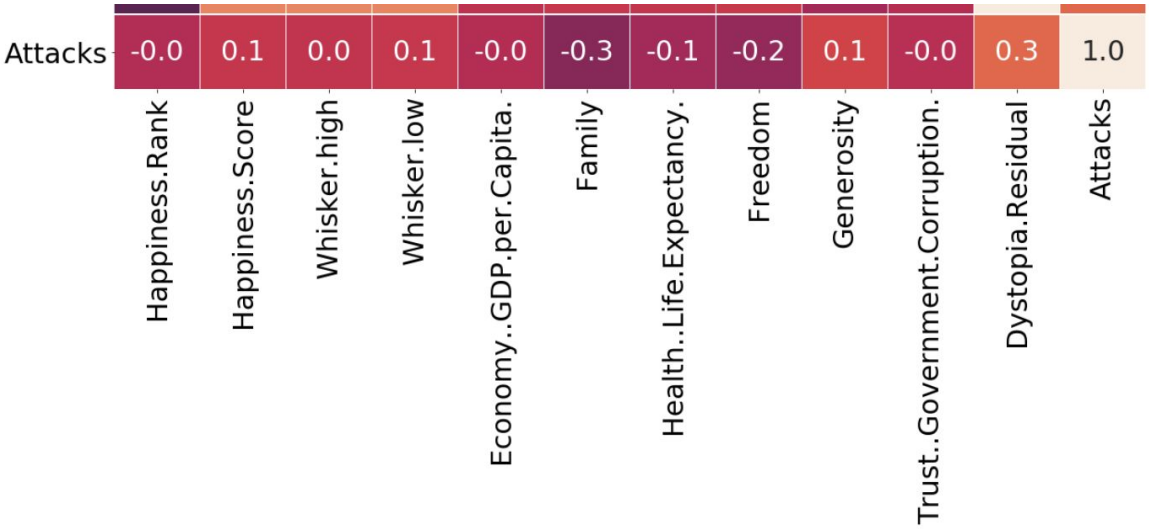
The following graph represents the happiness score of low-income countries:



The following graph represents the terrorist attacks of low-income countries:



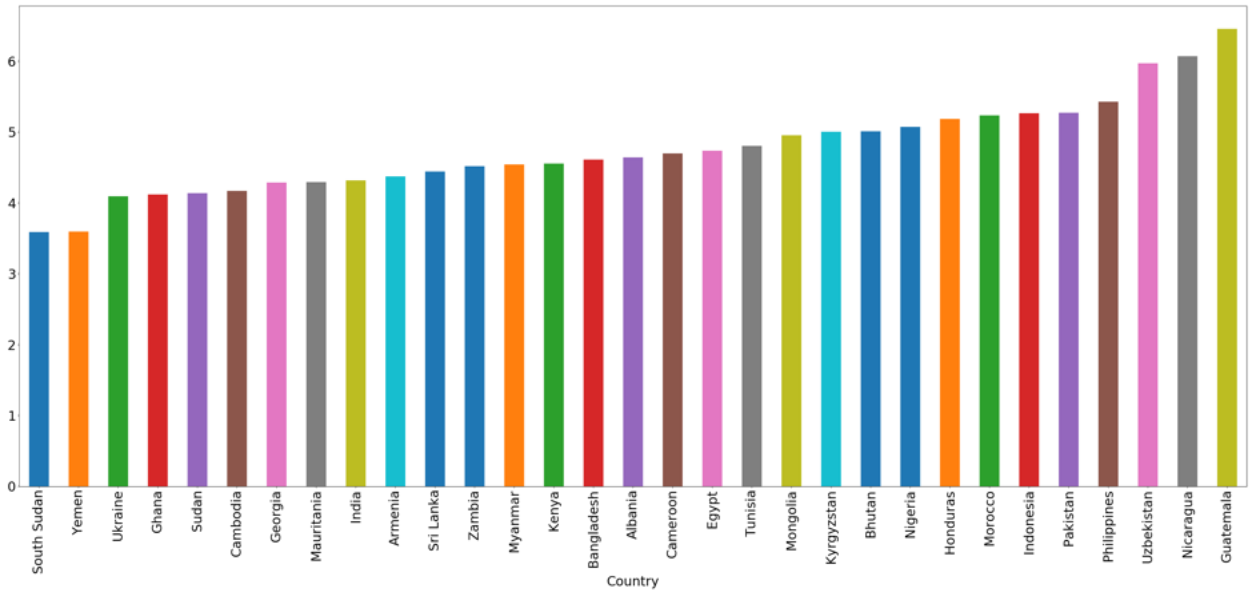
The following heat map represents the correlation between the number of attacks in low-income countries and each of the variables in the happiness score database:



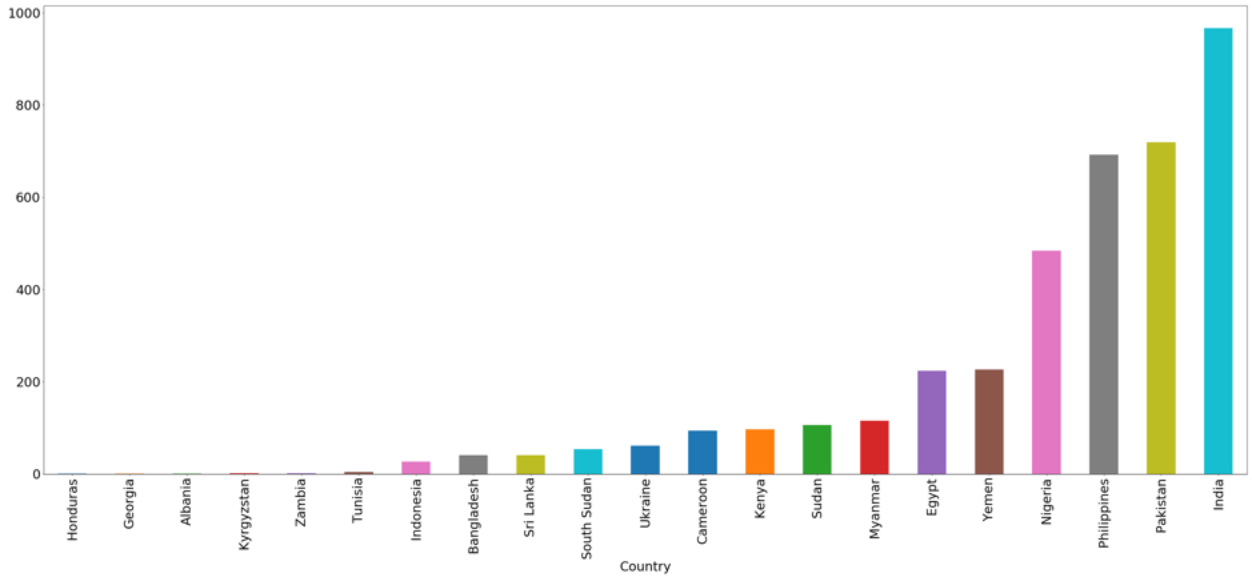
Notably the strongest correlations were “Family”(-0.3), Freedom”(-0.2)”, and “Dystopia Residual” (0.3)

The happiness score did not seem to have any statistical correlation with attacks. This points to the possibility that other factor play a larger role in the happiness of low-income countries.

The following graph represents the happiness score of lower middle income countries:

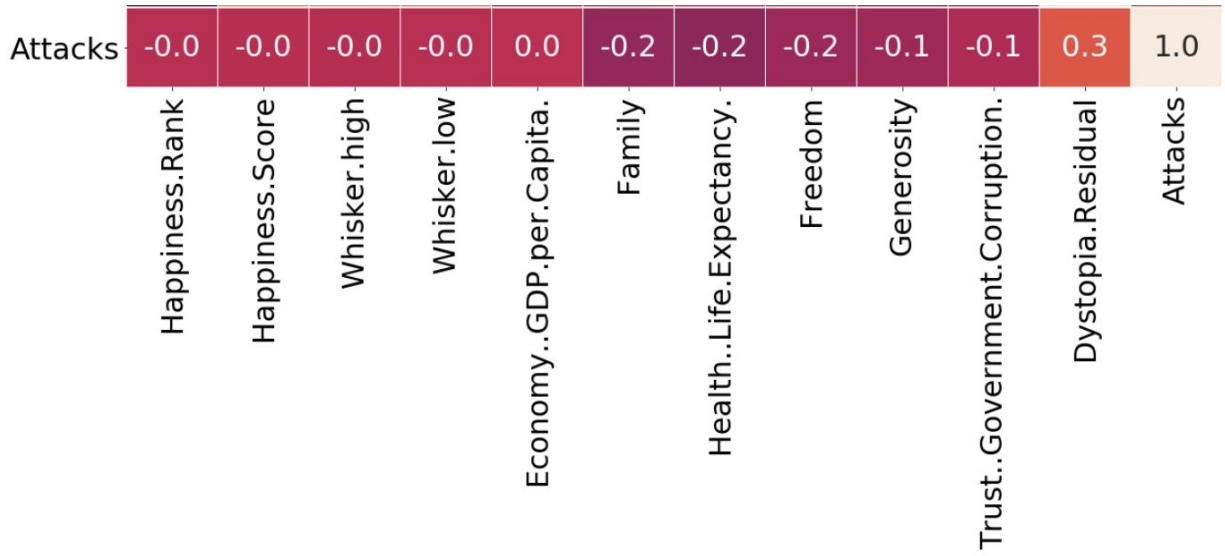


The following graph represent the number of terrorist attacks in lower-middle income countries:



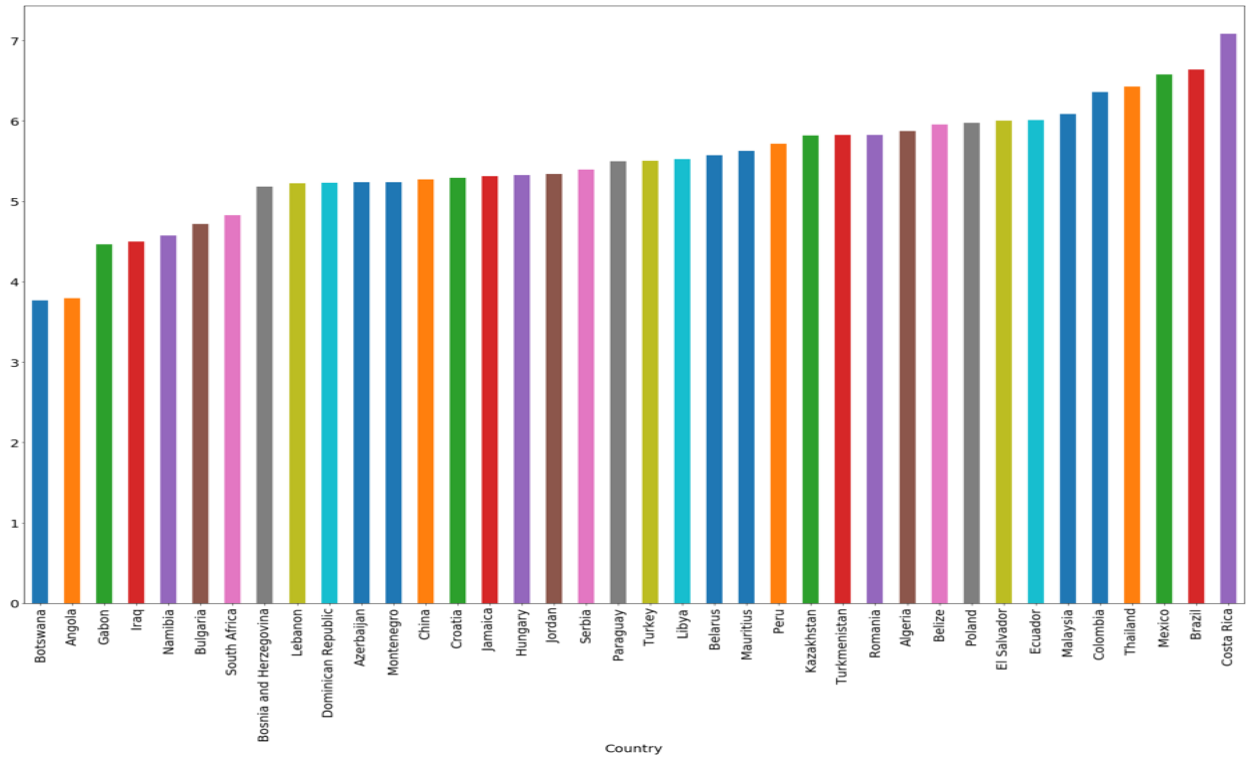
The following heat map represents the correlation between the number of attacks in lower middle

income countries and each of the variables in the happiness score database: :

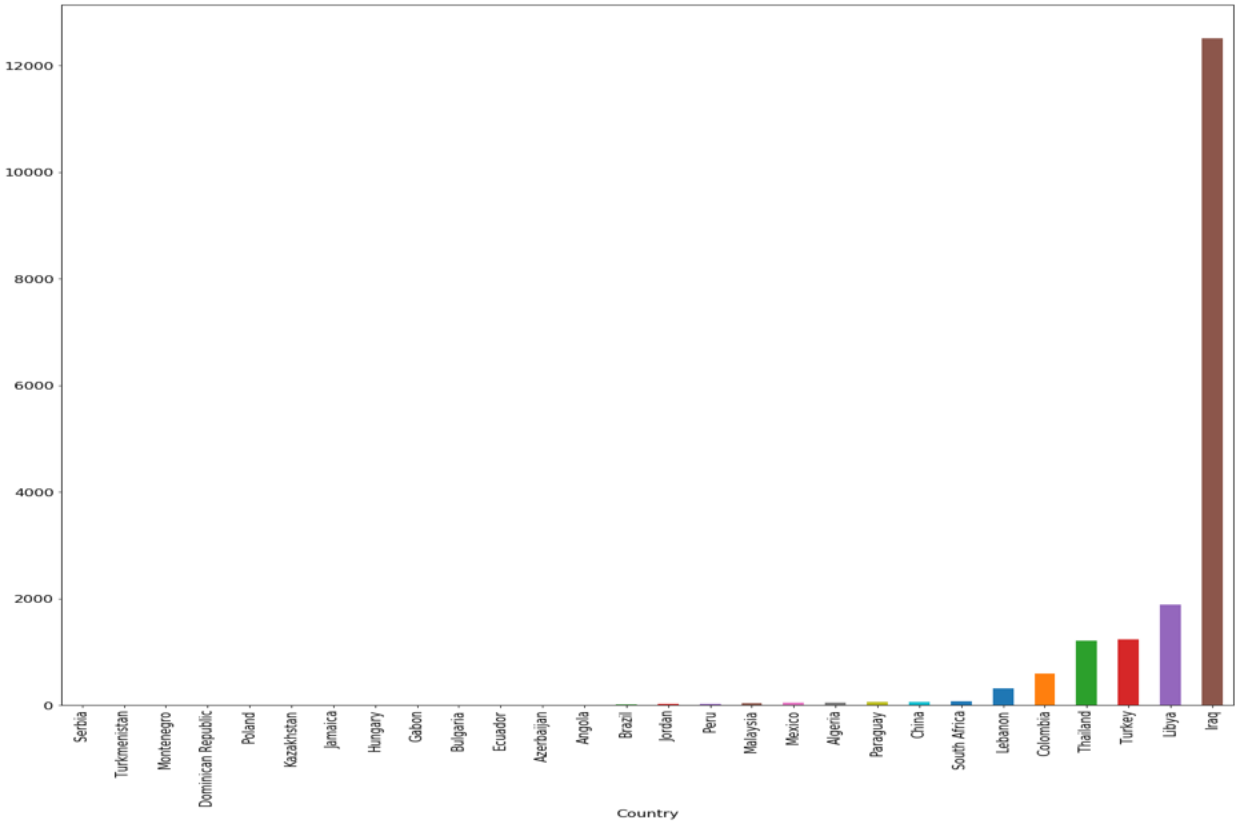


Notable correlations are Family, Health & Life Expectancy, Freedom, and Dystopia.
The happiness score was not correlated to the number of attacks for this group of lower middle income countries which also points to the possibility that terrorist attacks do not substantially contribute to the happiness of this group of countries, other factors play a greater role at least for this particular group of countries.

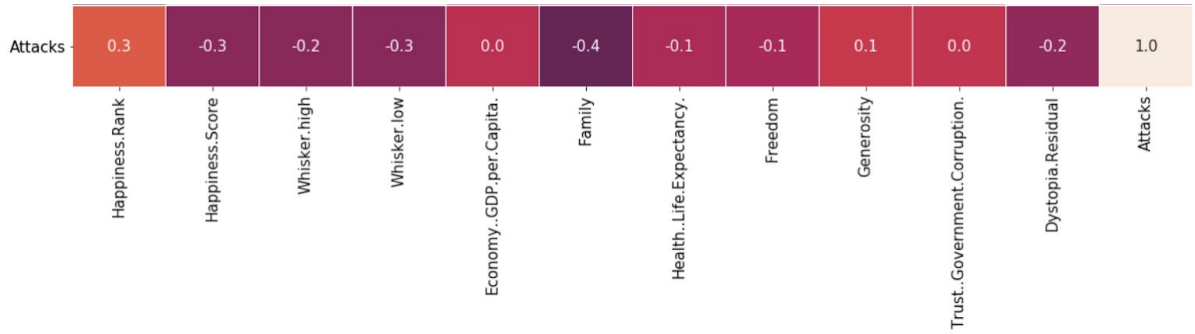
The following graph represents the happiness score of upper middle income countries:



The following graph represents the number of terrorist attacks in upper middle income countries:

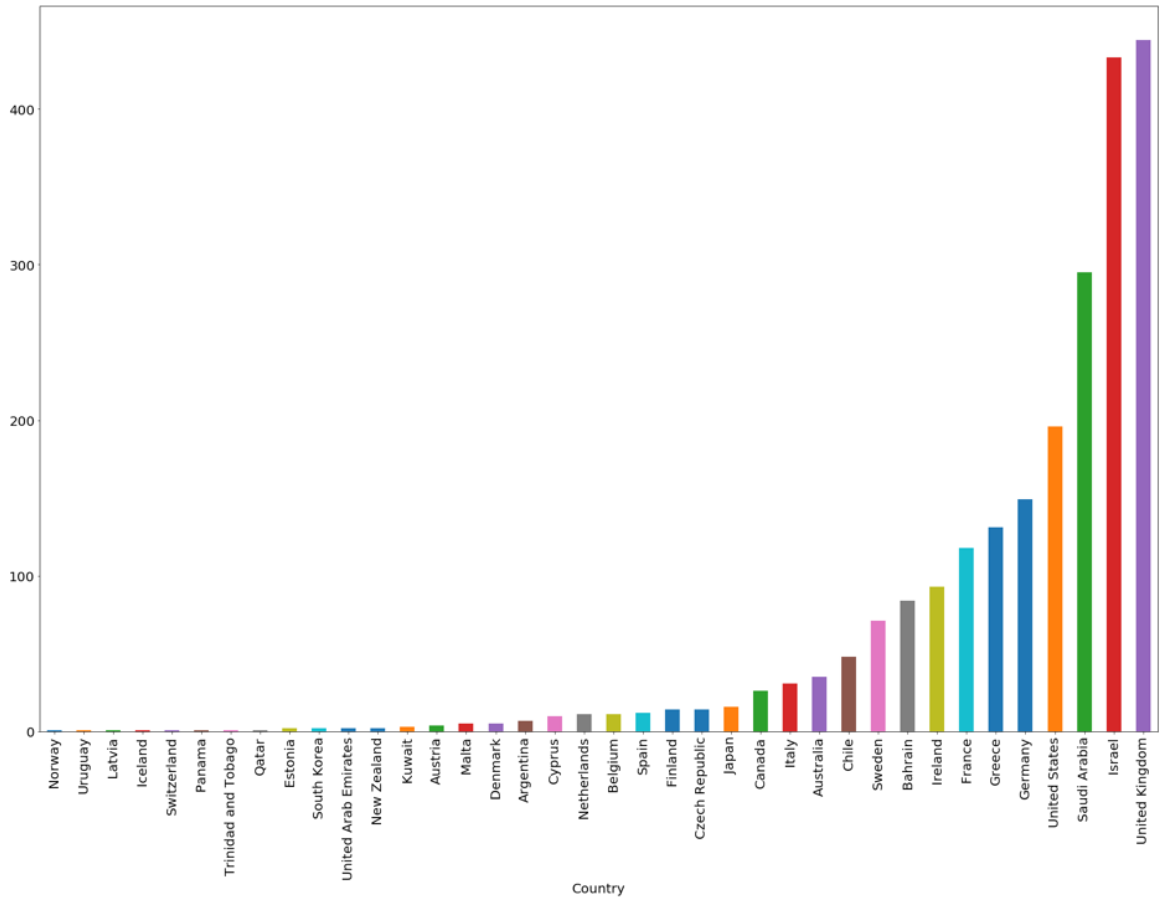


The following heat map represents the correlation between the number of attacks in upper middle income countries and each of the variables in the happiness score database:

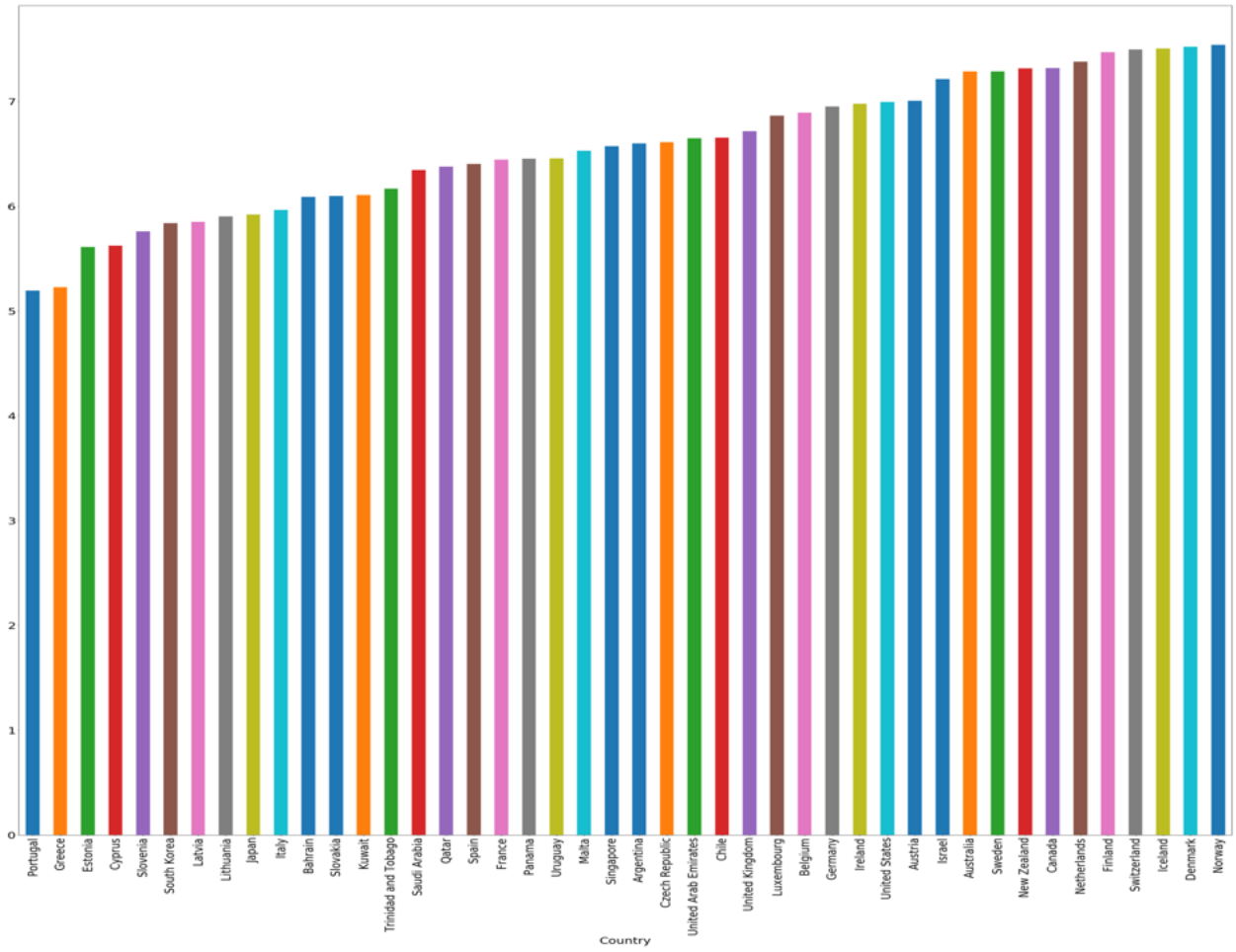


Interestingly, it seems that the happiness variables of upper-middle income have a higher correlation with the number of terrorist attacks than the previous clusters. Notably, the variables family(-0.4), happiness score(-0.3) and dystopia residual(-0.2) show that happiness is correlated to happiness score for upper middle-income countries.

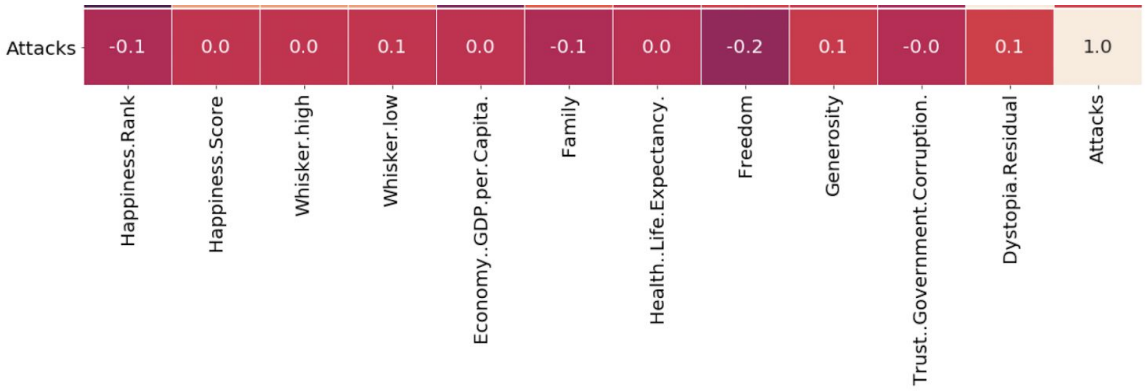
The following graph represents the number of attacks in high-income countries:



The following graph represents the happiness score of high income countries:



The following heatmap represents the correlation between the number of terrorist attacks in high income countries and the variables in the happiness data set:

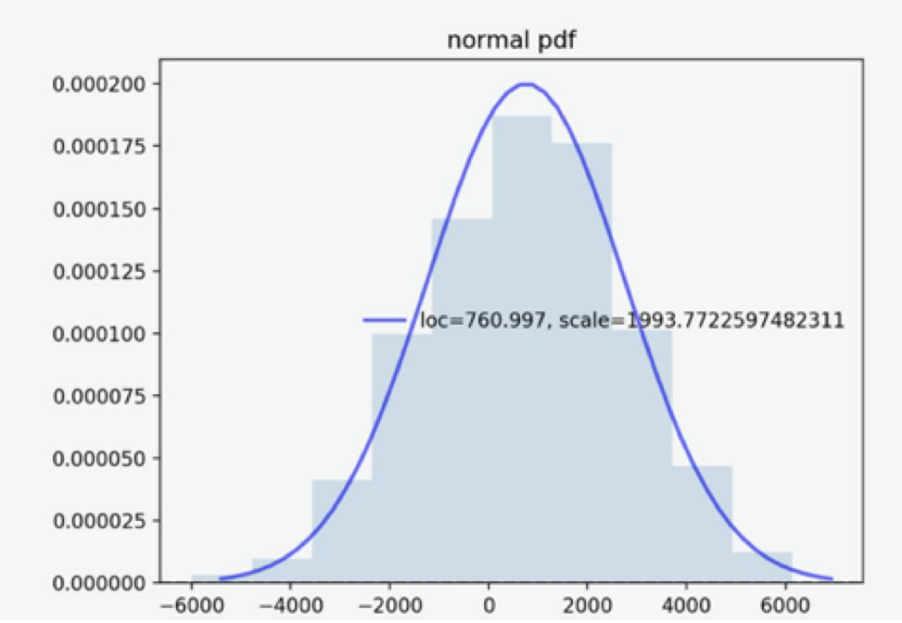


For this cluster, the correlations between number of terrorist attacks and happiness variables does is not as clear as the rest. It seem that high income countries are less affected by terrorist attacks or at least there doesn't seem to be substantial correlation between the number of terrorist attacks

and the factors that contribute to happiness score. The only weak correlation we found was "Freedom" (-0.2).

Distributions

For our distributions, we focused on the number of terrorist attacks throughout the world and what probabilities we can determine from our global terrorism dataset. In order to come up with the most accurate standard deviation and mean of our dataset, we extracted 20 different samples from our dataset of size 50. The reason we did not take into account all the combinations of all the samples we could possibly make of every country is because it would take too long. Therefore, we decided to use 20 simple random samples. Because the Central Limit Theorem states that when our sample size becomes sufficiently large, the sampling distribution can be well approximated by a normal curve even when the population distribution is not itself normal, we have used a normal curve to represent our sampling. A normal distribution will display negative values, however, such have no meaning in our experiment for there is no such thing as a negative terrorist attack. The following graph depicts our findings and averages of all the means and standard deviations found in our 20 random samples.
Because we have used the python function df.sample() our graph may vary every time we run this part of our project.



Because there are so many meaningless values in our normal distribution, we decided to look into different distributions.

We realized the poisson distribution might more accurately fit our needs, for we are dealing with discrete data. Using a poisson distribution, we obtained the following findings:

USA:
The probability there will be exactly 200 attacks in the US within the next five-year period:
0.008836103705157198
The probability there will be more than 200 attacks in the US within the next five-year period:
0.06685510229484393

Iran:
The probability there will be exactly 900 attacks in Iran within the next five-year period:
0.0046439888630927113
The probability there will be more than 2000 attacks in Iran within the next five-year period:
0.9725611492330197

Iraq:
The probability there will be exactly 2100 attacks in Iraq within the next five-year period:
0.002001735073177817
The probability there will be more than 2000 attacks in Iraq within the next five-year period:
0.99995349841246

This sort of prediction and testing can be further developed and used for every country. However, the predictions do not take into consideration the political status of the country, trade or policies, therefore our predictions will not be as accurate as we'd like them to be. These are solely based on the previous 40 years of terrorist attacks recorded on this specific dataset.

Stratified sampling by region

To have a better understanding about whether the geography of a country plays a part on the terrorism attacks or in the happiness of it, we decided to take a look to sampling by region and this is what we found.

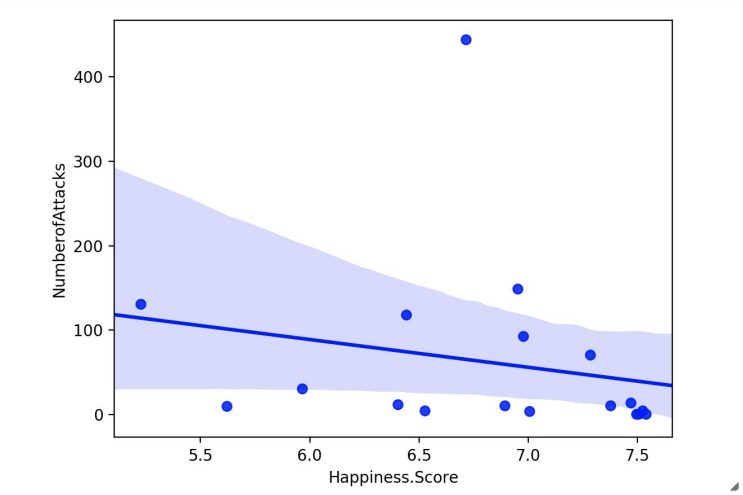
These are the regions with the most terrorist attacks: (ranked)

Middle East & North Africa	50474
South Asia	44974
South America	18978

Sub-Saharan Africa 17550
Western Europe 16639
Name: Region, dtype: int64

The results to our stratified samples can be seen below.
Western Europe contains the following countries from our dataset:
Norway, Denmark, Iceland, Switzerland, Finland, Netherlands, Sweden, Austria, Ireland, Germany, Belgium, United Kingdom, Malta, France, Spain, Italy, Cyprus, Greece

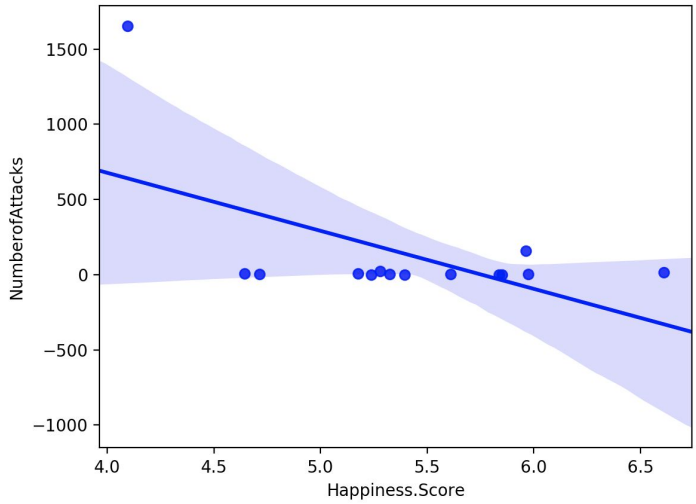
Western Europe




```
Slope is -32
Intercept is 285
R value is -0.210182
P value is 0.402526
Std Error is 38
```

Eastern Europe contains the following countries in our dataset:
Czech Republic, Poland, Russia, Latvia, Moldova, Estonia, Serbia, Hungary, Kosovo, Montenegro, Macedonia, Bulgaria, Albania, Ukraine

Eastern Europe



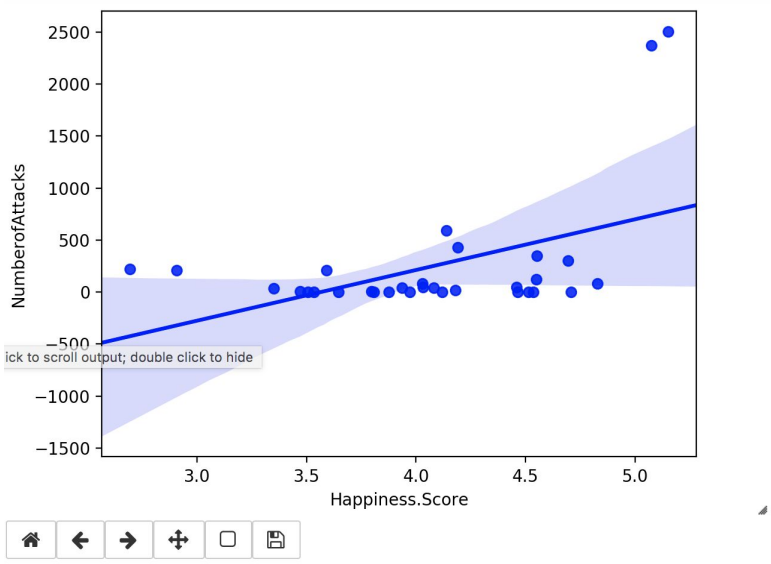

```
Slope is -385
Intercept is 2220
R value is -0.565177
P value is 0.035195
Std Error is 162
```

Southeast Asia

So far, we found some results that supported our original hypothesis that stated: as the number of terrorist attacks increase, the happiness score will decrease. However, we found a something surprising in the Sub-Saharan African region that showed the exact opposite.

Sub-Saharan Africa contains the following countries in our dataset:
Somailia, Nigeria, South Africa, Sierra Leone, Cameroon, Kenya, Mozambique, Senegal, Zambia, Gabon, Ethiopia, Mali, Ivory Coast, Sudan, Ghana, Uganda, Burkina Faso, Niger, Malawi, Chad, Zimbabwe, Lesotho, Angola, Madagascar, South Sudan, Liberia, Guinea, Rwanda, Tanzania, Burundi, Central African Republic

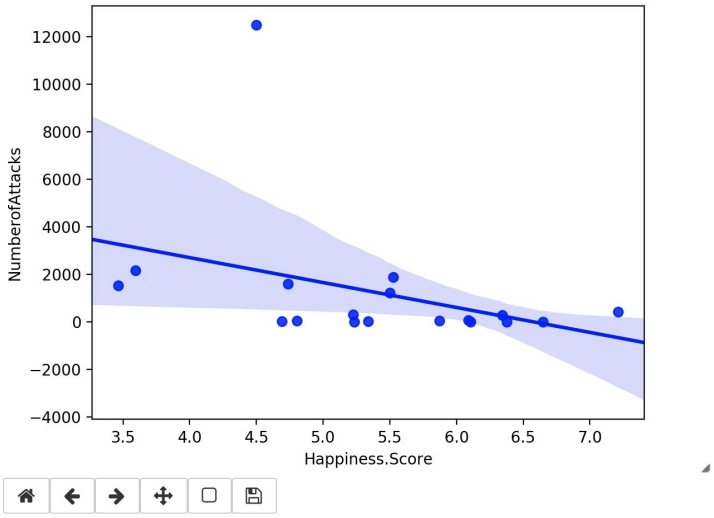
Sub-Saharan African region




```
Slope is 487
Intercept is -1735
R value is 0.469961
P value is 0.007638
Std Error is 169
```

The middle east consists of the following countries in our dataset:
Israel, United Arab Emirates, Qatar, Saudi Arabia, Kuwait, Bahrain, Algeria, Libya, Turkey, Jordan, Morocco, Lebanon, Tunisia, Egypt, Iran, Iraq, Yemen, Syria

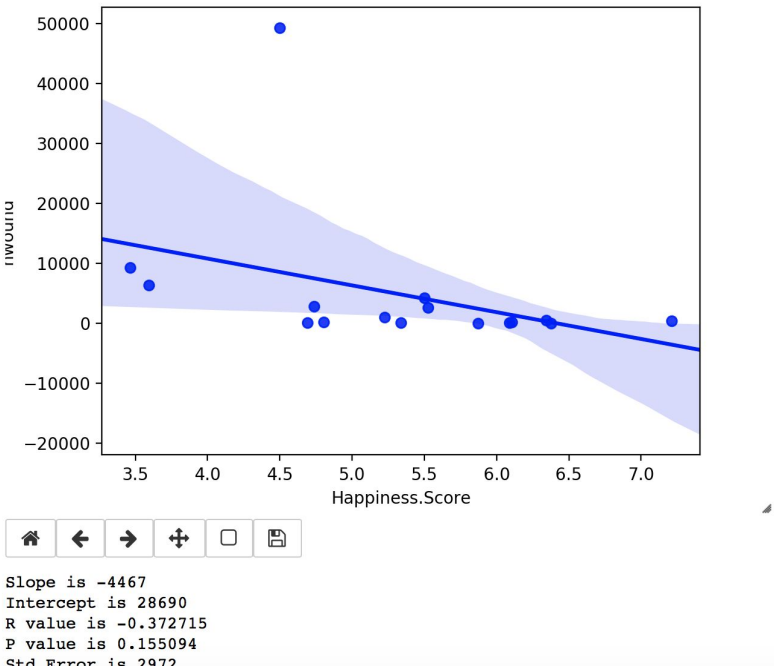
Middle East




```
Slope is -1049
Intercept is 6905
R value is -0.361460
P value is 0.140529
Std Error is 676
```

We also decided to find any correlations between the number of wounded and each region's happiness score. We used the middle east first because that sample has the most terrorist attacks. In the middle east, we found a weak negative correlation between the number of wounded people and the happiness score of the country. We believe that there might be some bias in this sampling when using number of wounds because a specific region, like the middle east, has the most terrorist attacks in the world. However, the reason we took the same approach with the number of wounds is because the total data of the whole world might have too drastic differences, unable to give us concrete conclusions.

Middle east: Number of wounds vs Happiness Score



Conclusion

After analyzing all our data we came to the conclusion the best fit distribution for our data was the poisson distribution due to its discrete properties as well as our data's. We also determined that using stratified sampling in this project was the best approach, for it would be hard to find any strong correlations using the entire dataset or random, very distinct, values for the number of terrorist attacks. We determined that that the variables that influence happiness are indeed negatively correlated to terrorist attacks but that there are other factors which have a greater impact on the happiness score in spite of a high number of terrorist attacks.

There is a significant difference between correlations in the different income brackets. The highest income countries showed the least correlation between terrorist attacks and their happiness scores. Middle income countries showed the greatest correlation between happiness score and terrorist attacks. Surprisingly, there are cases where some countries with high terrorist attack have a high happiness and don't exhibit a substantial correlation between the two.
Some regions like Sub-Saharan African region display a direct relationship between its happiness score and the number of terrorist attacks in the country. This could be due to outliers skewing the data, or other factors not provided in our dataset. Further analysis is needed.

Overall There are many other factors that can affect a country happiness that we are not aware off but terrorism does play a part in it.

Sources:
https://www.kaggle.com/unsdsn/world-happiness
https://www.kaggle.com/START-UMD/gtd
http://data.un.org/
https://blogs.worldbank.org/opendata/new-country-classifications-income-level-2018-2019

Comments on experience:

It was a challenging dataset to analyze but it allowed us to become completely familiar with the numpy, pandas, seaborn, and statsmodels packages. Regarding our teamwork, we could have furthered our analysis on these two datasets if every team member had contributed equally to this project.

Juan Diaz Sada: Data analysis , Income brackets clustering and analysis , data visualization and google maps API , Power point, reports.
Melissa Gonzalez: Data analysis ,stratified sampling for region, linear regression analysis, report, distribution analysis
Arturo Garza: Abstract, helped with report
Raul Valadez: copy/pasted our code into one jupyter notebook, resized images in powerpoint