Aryan Anerao
ID# 801129736
P 2: Project Report
DTSC 3602
12/16/2021

Predicting Terrorist Targets in South America

Problem Statement and Background

Global warming due to climate changes continue to affect the world, since the original observation of this phenomenon in the early 20th century. The global average temperature has increased by one degree Celsius since the pre-industrial period and many scientists theorize that this increase is primarily due to human activity. Theoretical models suggest that continuation of climate change trends could lead to increases in severity of the following: global land and ocean temperature; rising sea levels; melting of ice at Earth's poles and in mountain glaciers; frequency of extreme weather such as hurricanes, heatwaves, wildfires, droughts, and floods[i] (NASA). These effects could drastically impair the economy and infrastructure of many countries, especially in regions near the ocean, leading to economic and political instability.

Climate changes could make South America more prone to political and economic turmoil, which could lead to a rise in terrorism. Ten of the twelve countries in South America have coastline and the west coast of South America lies directly adjacent to the Nazca tectonic plate. The seismic hazard map of South America, shown in Figure 1, depicts how the fault lines by the Nazca Plate are a hotspot of seismic activity. The seismic volatility of the west coast of South America makes this region especially susceptible to earthquakes. I suspected that the proximity to natural disasters may make cities more vulnerable to terrorism due to the resulting damage to the local infrastructure and economy.

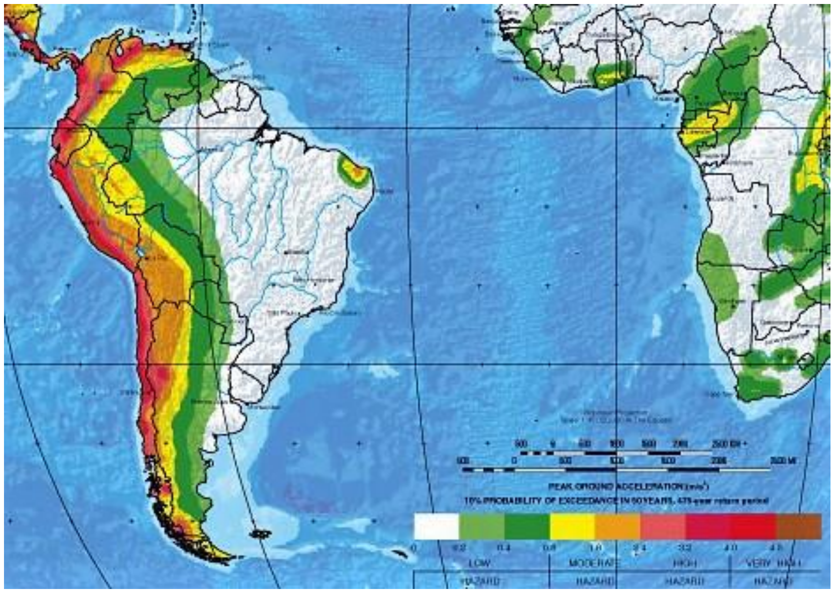

Figure 1: Seismic Hazard Map of South America (Institute of Geophysics)

During the cold war the United States and Soviet Union endorsed organizations in South America that employed terrorist tactics. In the early 1980s, many of the countries regained their democracy after the fall of military dictatorship regimes. However, the political instability led to groups voicing their disapproval with the newly elected leaders through terrorism. The ensuing political and economic stability from the transition of power from 1985 to 1995 led to a significant increase in terrorism. The occurrences of extreme weather events due to climate change could lead to another period of instability resulting in the rise of terrorism in this area.

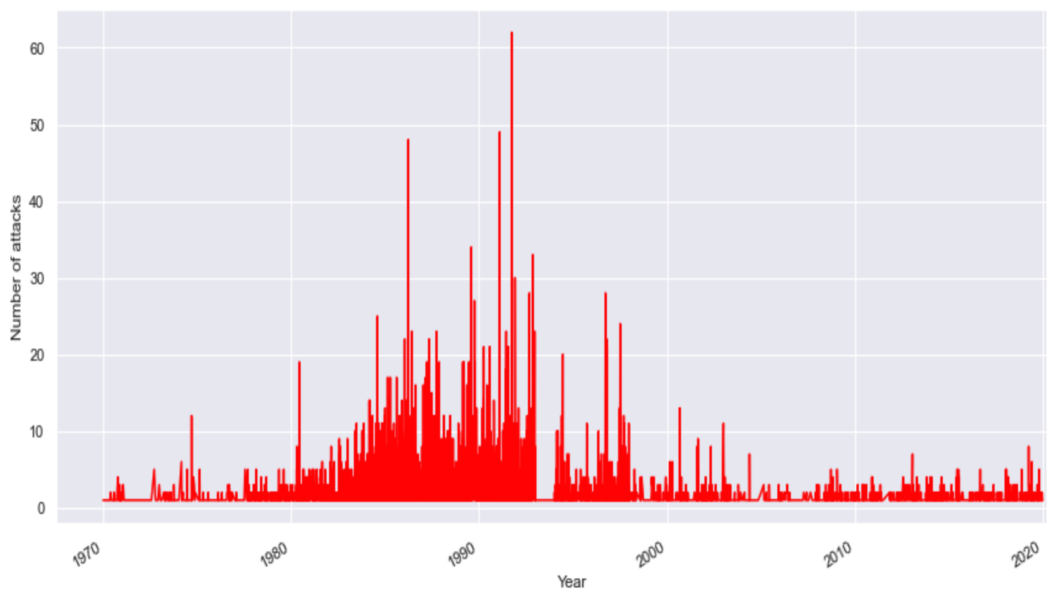

Figure 2: Number of Terrorist Attacks in South America over the Last 50 Years

In the project, I developed several predictive models using machine learning techniques to predict the target city of a terrorist attack based on the attack type, target type and distance to the nearest earthquake hotspot as predictors. I evaluate each model in terms of its accuracy of predictions. I determined the accuracy using a function that calculated the number of correctly predicted cities divided by the total number of attacks. I also measured the cross-validation scores of each model.

Data

For my analysis, I used data from the following sources: The Global Terrorism Database (GTD), The USGS Earthquake catalog (USGS), Global Seismic Hazard Map (GSHAP), as well as a climate map of south America. The GTD is an open-source dataset, where each row contains information on domestic and international terrorist attacks around the world from 1970 to 2019. This data base defines acts of terrorism as “The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.”(GTD) The GTD dataset includes information on the time, location, and details of each terrorist attack. Initial exploration of the number of terrorist attacks in each country supports our initial hypothesis that countries along the western coast of South America have higher numbers of terrorist attacks, as seen in Figure 3.

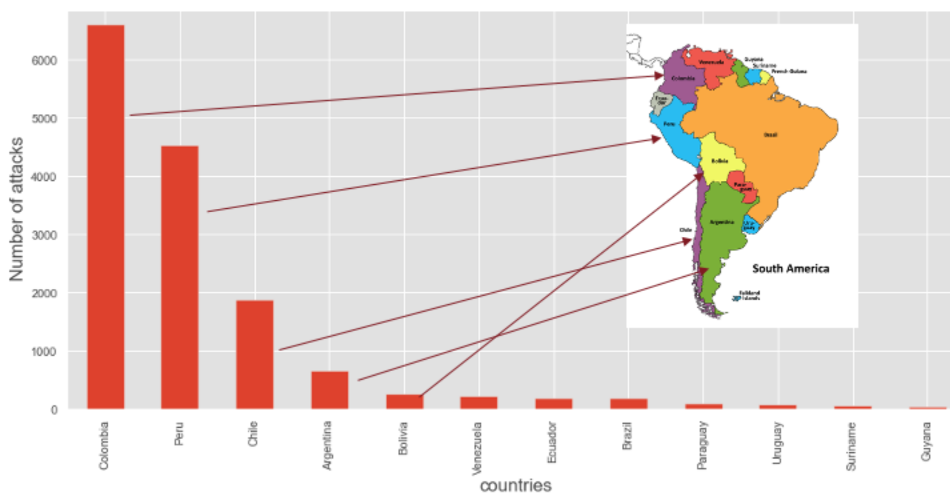

Figure 3: Number of Terrorist attacks per Country in the last 50 years

The USGS Earthquake Hazards Program is a part of the U.S. Geological Survey and is led by the National Institute of Standards and Technology. The earthquake catalog contains earthquake information about earthquake around the world from 1940 and I applied clustering techniques on this data to determine Earthquake hotspots in South America to then derive a predictor variable for modeling, as seen in Figure 5 below.

The data I collected had enough values present to study. I selected columns containing information about the date, location, attack, and weapon type, as well as the target of the attacks from the GTD dataset. The columns that I used for my models did not have any data sparsity issues as there were information for most events. However, there were some terror events that did not have a known city. The cities that were unknown were labeled as "Unknown". I resolved this issue by creating a column that concatenates the city and country together being separated by a comma. For example, an unknown city in Colombia would be displayed as "Unknown, Colombia". This ensures that the model does not classify all unknown cities as equal. The longitude and latitude of each event was present in the data, this also ensures that when the model uses the calculated distance predictor it would use the different distances as a factor to predict the unknown cities as well.

In order to predict the target city of each terrorist attack, I assigned the predictor variables as attack type and target type directly from the GTD. The third predictor being closest distance to an earthquake hotspot, was as calculated column using the longitudinal data of the cities from the GTD and the longitudinal data of the earthquakes from the USGS catalog.

Methods & Results

I first imported the following python packages:

- Pandas
- Numpy
- Literal_eval
- Matplotlib.pyplot
- Scipy
- Geopy
- Various sklean packages

These packages allow for data frame alterations, calculations, modeling, and graphing.

Prior to modeling, I first prepared the raw data. After loading the terrorist attacks from the GTD database into a Pandas data frame, I selected all the attacks that occurred in locations in South America and the date, time, location, attack details for each. Initial exploration of the distribution of attacks revealed that certain cities and countries did not have enough data points for modeling. I removed any rows where that was the only attack occurring for the corresponding city. Additionally, I removed attacks from French Guiana and the Falkland Islands because they had less than ten occurrences.

I also dropped any rows with missing values for attack type and target type. I then consolidated all date and time columns into a single column as well as consolidated city and country columns into a single column separated by a comma. This would ensure that each city is unique to its country in case there are cities with the same name in different countries. I formatted the longitude and latitude columns into a separate column having the data separated by a comma and gave it the name: coordinate. This allows me to use the Geopy python package to measure the distance between the location of the terror event and the nearest earthquake hotspot. The final GTD data set had 13453 rows, where each row represented a distinct terrorist attack in South America.

Next, I loaded the USGS earthquake data into another Pandas data frame selected earthquakes in South America with a magnitude of greater that 5.0 from 1970 to 2019. From this dataset, I selected date, time, location, and magnitude columns and formatted each one to match the GTD data frame formatting. I postured this data frame by first dividing the continental land of South America into 3 regions: upper, middle, and lower based on the longitude and latitude. I then dropped all earthquakes that do not fall into these regions as they would be too far from the land mass.

I created the predictor “closest distance to an earthquake hotspot” by first performing K-means clustering on the imported coordinates of the earthquakes. The “Elbow plot” in Figure 4 shows that 5 clusters was the optimal number of clusters for the Earthquake clusters. I then plotted the coordinate clusters along with a X displaying the centroid of each cluster, as seen in Figure 5. The five centroids of the represents the five earthquake hotspots. Using the Geopy package I calculated the distance of each city in the GTD to the nearest hotspot. I then created a column in the data frame with the terror events for the nearest distance, which would be a predictor variable in subsequent modeling.

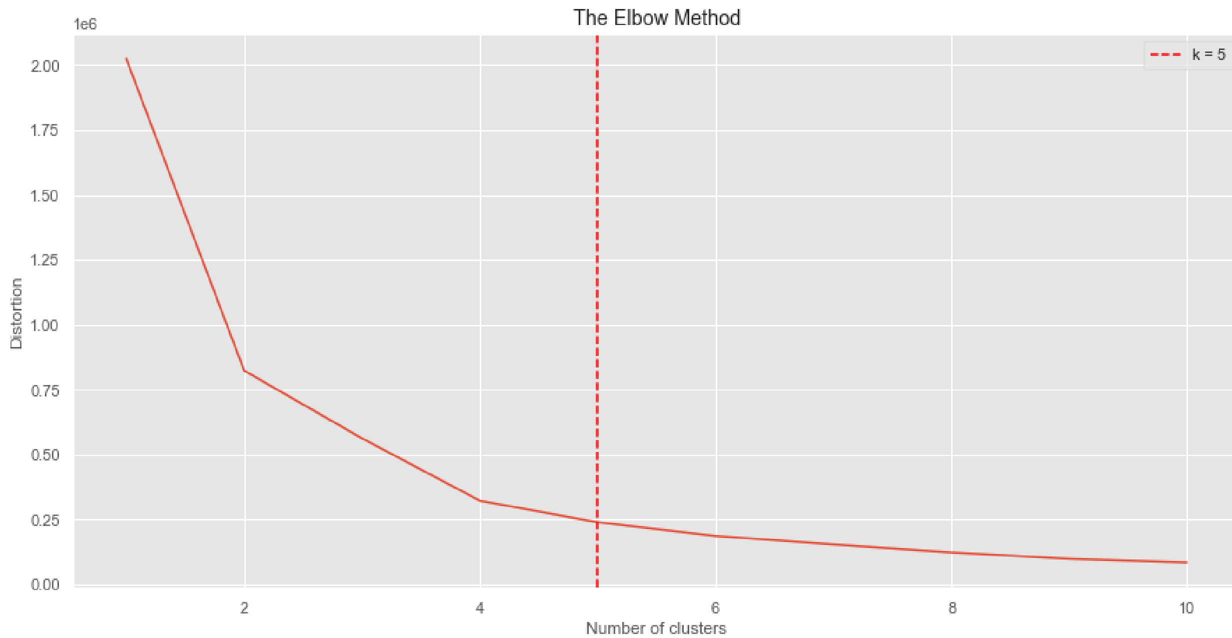

Figure 4: Elbow Plot for K-Means Clustering

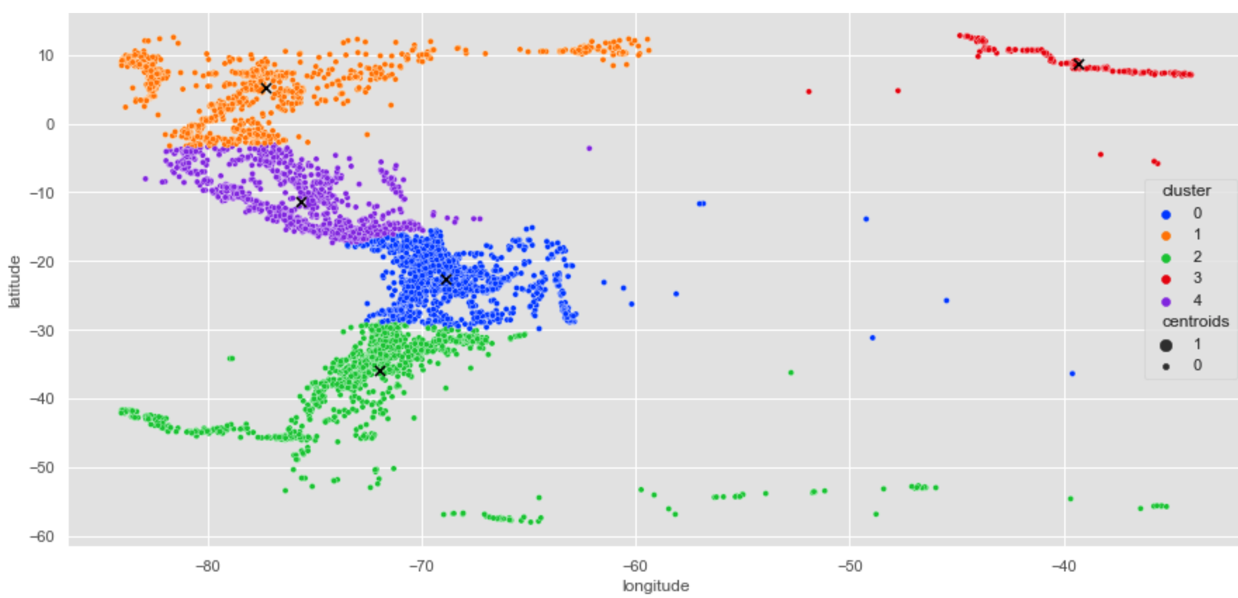

Figure 5: K-Means Earthquake Hotspot Cluster Map

To predict the target city for each terrorist attack, I developed the following supervised machine learning algorithms: CART, C50.0, Random Forest, KNN, and SVM. They would predict which cities will experience terrorist attack based on attack type, target type and distance to earthquake hotspot by classifying each attack into a city based on the predictors. To train these models I randomly split the terrorist attack data using SKLearn such that 70% of the data would be used to train the models, and the remaining 30% would be used to test the model performance. Additionally, I used one-hot encoding method to convert categorical variables to numeric for modeling.

The C50 and CART models were tuned using an impurity criterion of entropy to produce more accurate analysis at the expense of additional run time compared to Gini. I set the max leaf nodes as 100 because when using applying a one hot encoding technique the number of features are increased to be over 100. The Random Forrest model was tuned to have 10 as the number of estimators. Although more estimators increase model accuracy, the cost of computation time for learning these additional estimators requires a practical cutoff, in this case 10. To slightly increase the accuracy of the Random Forrest model, I used Gini as the impurity metric. Based on the accuracy plot shown in Figure 6, the optimal number of nearest neighbors for the KNN plot was 1.

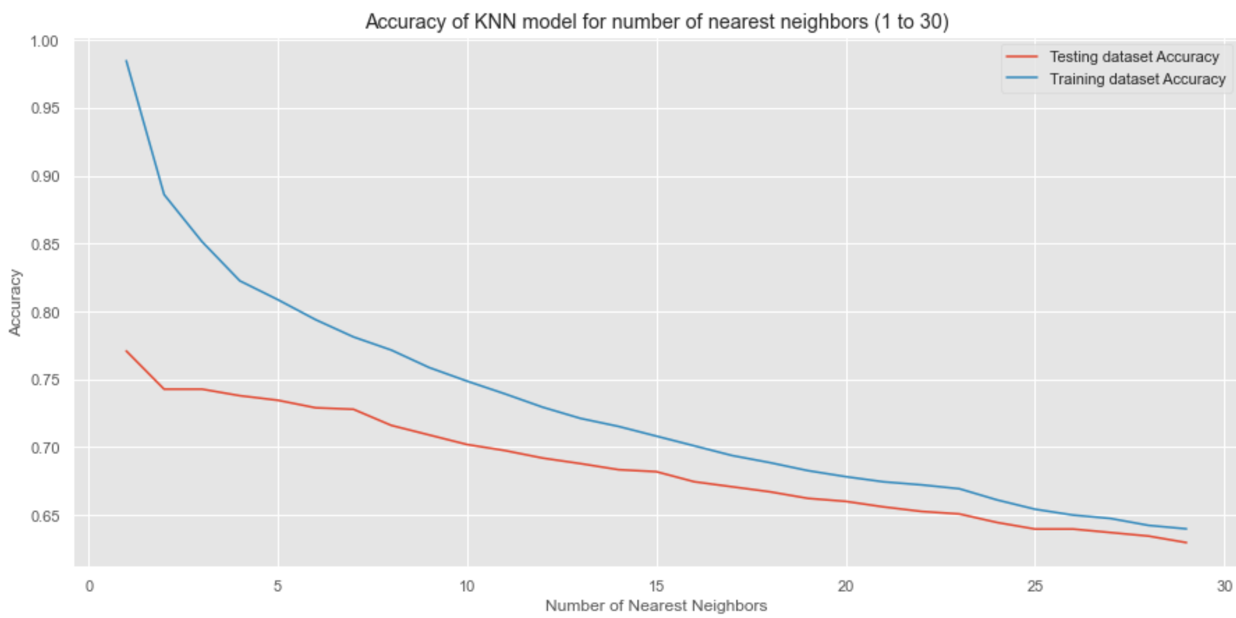

Figure 6: K-Nearest Neighbors Accuracy Plot

Results

Using the models to predict the target city, I evaluated the average cross-validation score and the accuracy of the model. To measure the accuracy, I created a function that counts the number of correctly predicted cities and then divides it by the total number of cities being predicted. The accuracy is measured from the test set. I used the same prediction featured and training/testing split for all the models. The accuracy and average cross validation scores for each model are shown in Table 1.

Table 1: Model Accuracy and Cross-Validation Scores

Model	Average Training Cross-Validation Score (K=10)	Accuracy
KNN	75.65%	98.70%
Random Forest	57.14%	95.59%
Cart	62.11%	68.98%
C5.0	62.54%	68.94%
SVM	45.43%	46.26%

The Random Forest and The KNN models were the most accurate models for predicting the target city with an accuracy of 98.7 and 95.6% correct predictions respectively. I believe that these models performed the best because they are better suited for classify categorical variables for a data set of this size. When looking at the variation between the cross validation and the accuracy score, the SVM had the least variation followed by the Cart and C5.0 models. Based on the low variation we can conclude that in these models there is minimal overfitting, or generalization. I believe that the KNN had the best performance over all because it had the highest accuracy score, and average cross validation score. The variation between the two is about 23% different.

My goal in this study was to determine a predictive model could accurately predict cities with a history of terrorism in south America by using the distance the cities are to earthquake hotspots, the attack and target types of the cities. The accuracy of the models suggests that given these parameters, the model can successfully classify cities.

Ethical implications

Exploratory data analysis revealed that most cities had enough data points for modeling. Terrorist attacks from nine cities from the Falkland Islands and French Guiana were removed due to insufficient sample size. To eliminate bias through my analysis I limited the number of predictors used to only three variables and only used ones that were related to the question I aimed to study. I also created a correlation matrix of the predictors and found that there were no columns that had a strong correlation, indicating that the data did not have any multicollinearity issues. I tuned the parameters of each model to optimize for accuracy and run time. For the KNN model. I used a single nearest neighbor of the because there were 3 predictors. The random Forrest had 10 as the number of estimators. The Cart and C50.0 models had 100 as the maximum number of leaf nodes because when I use one hot encoding on the data for the models the number of data columns exceed 100.

Lessons Learned

This project sought to develop supervised learning models to predict the target city of a terrorist attack based on the attack type, target type and distance to the nearest earthquake hotspot as predictors. The KNN and Random Forest models correctly predicted the target cities with a 98.7 and 95.6% accuracy respectively. I aimed to decern if economic instability from climate changes effects could make South America more prone to political and economic turmoil, by performing various models using terrorist events in the area. The high accuracy of the models validates the distance to earthquake hotspots as a strong predictor of the location of a terrorism event. When visually comparing the earthquake hotspot map and the location of the terror events in South America, there appears to be a correlation. The Top 5 countries with most terrorist attacks located are all located on or near earthquake hotspots. The targets of the events also help support the relationship between economic instability and terrorism as business and infrastructure were most common terrorism targets.

Sources of Information

1. Adrian Garlati (2021) Climate Change and Extreme Weather Events in Latin America: An Exposure Index | Publications. Retrieved December 20, 2021, from https://publications.iadb.org/publications/english/document/Climate-Change-and-Extreme-Weather-Events-in-Latin-America-An-Exposure-Index.pdf
2. Báez, A. A., Sztajnkrycer, M. D., Zane, R., & Giráldez, E. (2008). Twenty-five years of violence: the epidemiology of terrorism in South America. *Prehospital and disaster medicine*, *23*(2), 128-132.
3. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/B97C49EF9DCB376AF0F512E62EB61726/S1049023X00005732a.pdf/twenty-five-years-of-violence-the-epidemiology-of-terrorism-in-south-america.pdf
4. Feldmann, A. E., & Perälä, M. (2004). Reassessing the Causes of Nongovernmental Terrorism in Latin America. Latin American Politics and Society, 46(2), 101–132. Retrieved from https://www.jstor.org/stable/3177176?seq=1#metadata_info_tab_contents
5. GTD | Global Terrorism Database. (2021, December 20). Retrieved from https://www.start.umd.edu/gtd
6. Halkides, M. (1995). How not to study terrorism. *Peace Review*, *7*(3-4), 253-260. https://www.tandfonline.com/doi/abs/10.1080/10402659508425888
7. Hernandez, D. H. B., & Church, A. L. (2003). Terrorism in Peru. *Prehospital and disaster medicine*, *18*(2), 123-126. https://www.cambridge.org/core/journals/prehospital-and-disaster-medicine/article/terrorism-in-peru/BBCC971A24DF0AAE6DDFCC2942CFDB5D
8. Javier Yanes. (2021, October 08). Retrieved from https://www.bbvaopenmind.com/en/science/environment/climate-change-trigger-earthquakes-tsunamis-volcanic-eruptions
9. Kluch, S. P., & Vaux, A. (2016). The non-random nature of terrorism: An exploration of where and how global trends of terrorism have developed over 40 years. *Studies in Conflict & Terrorism*, *39*(12), 1031-1049.https://www.tandfonline.com/doi/full/10.1080/1057610X.2016.1159070
10. NASA. (2021) Risk perceptions of nuclear energy, climate change, and earthquake: How are they correlated and differentiated by ideologies? - ScienceDirect. Retrieved December 20, 2021, from https://www.sciencedirect.com/science/article/pii/S2212096321000267
11. Python, A., Bender, A., Nandi, A. K., Hancock, P. A., Arambepola, R., Brandsch, J., & Lucas, T. C. (2021). Predicting non-state terrorism worldwide. *Science advances*, *7*(31), eabg4778. https://www.science.org/doi/full/10.1126/sciadv.abg4778
12. Smith, P. J. (2007). Climate change, weak states and the" War on Terrorism" in South and Southeast Asia. *Contemporary Southeast Asia*, 264-285. https://www.sciencedirect.com/science/article/pii/S0962629807000601
13. Sullivan, M. (2016, December 14) Latin America: Terrorism Issues. congressional Research Service. Retrieved from https://sgp.fas.org/crs/terror/RS21049.pdf
14. Suarez, M. A. G. (2012). Terrorism and international politics: an approach to South America. *Contexto Internacional*, *34*(2), 363-396. https://www.scielo.br/j/cint/a/Kb7LKYVvSPgGzxLX737BJdy/?lang=pt
15. USGS Earthquake Catalog. (2021, December 20). Retrieved from https://earthquake.usgs.gov/earthquakes/search
16. Voirin, J. L. C. (2018). State Terrorism: South America during the 1970s. A Case of State Terrorism Transnationalization. *História Unicap*, *5*(9), 270-283. https://dialnet.unirioja.es/servlet/articulo?codigo=7832395
17. Weiffen, B., Wehner, L., & Nolte, D. (2013). Overlapping regional security institutions in South America: The case of OAS and UNASUR. *International Area Studies Review*, *16*(4), 370-389. https://journals.sagepub.com/doi/full/10.1177/2233865913503466

---

i