Database Management

Are Rising Temperatures Increasing Terrorism?

December 15, 2021

Giorgio Cassata, John Murphy, Daming Wang

Colorado School of Mines

Introduction

For Project 9 – Create, we decided to find correlations between two datasets and analyze the results. We used global land temperatures and terrorism data from the Global Terrorism Database (GTD) from the University of Maryland. Average land temperatures all over the world are famously rising due to global warming and media perception of terrorism and unrest has also seen an increase in recent years. The GTD contains 135 pieces of information linked to each entry. We used information for events linked to terrorism such as the number of deaths linked to it, property damage occurrence, ransom occurrence, and whether the attack was successful. We linked our data by date and country. Our analysis point was to compare countries overall by summing and averaging all of their data over the years.

Data Used

We obtained both datasets used from Kaggle on kaggle.com. Both datasets had a .csv download which was easy to work with and upload. However, the Global Terrorism Database (GTD) we found on Kaggle has specific license restrictions. We acknowledge that all information obtained from the dataset is Copyright University of Maryland 2018. Any conclusions drawn from the modification of this data, or subjective interpretations made, are our own. We do not have any express permission to publicly display any of the GTD data, nor use it commercially. The dataset is cited as per the licensing agreement, including the source site even though the file was obtained elsewhere. [1] Despite the licensing terms we decided to move forward with this dataset since our use case is within the licensing agreement. We also desired to use the GTD for its' number of entries and number of columns to relate with the other dataset. The GTD is "Currently the most comprehensive unclassified database on terrorist attacks in the world" according to their website. [1]

The other dataset used was a publicly available global land temperatures dataset. This was also obtained through Kaggle, but exists in the public domain. [2] We wanted to use this dataset because it is very extensive, complete, and contains common columns to relate other datasets to. We were impressed that the date entries started as early as 1743, and 243 unique countries listed. These two attributes, date and country, were what we used to relate the two datasets to one another.

We uploaded the .csv files for the cleaned versions of these datasets to the Mines' Postgres server using Daming Wang's account: damingwang. Appendix B

Processing

Before we uploaded and started working with the data, we had to do some preprocessing with Python and Pandas. For the global land temperatures dataset, we removed the AverageTemperatureUncertainty column since there were no useful comparisons or conclusions to be drawn, and only kept the data from 1970 to 2013 since those are the only years the two datasets overlap. The Global Terrorism Database required much more preprocessing and was a necessary step before uploading the data. First, we extracted only the columns that we were interested in since the dataset contains 135 columns, many of which are less complete than others. Secondly, we combined the year, month, and day columns into one date column that would be accepted by psql. Finally, we capped the lower bounds of some of the columns that had improper, or otherwise invalid values which caused issues when trying to upload the data. Please see Appendix A for the script used for all of our preprocessing.

To organize and prepare our data for analysis we also did some work in SQL. We merged the tables for both datasets into one table. To do this we used subqueries to select data by year from both datasets, and then inner joined them by year and country. This gave us a table where each country had one entry per year with the average temperature and sum of various terror information. An example would be the sum of successful terror attacks for a country within a given year.

Results

First, we looked at the information by country. We did this by averaging all the temperature data for each country and taking the sum of each metric related to terrorism that we were interested in. The purpose of this was to look at the potential influence that temperature might have on terrorism. We plotted the number of successful terror attacks vs. overall average temperatures (fig. 1) and the number of deaths due to terror attacks vs. overall average temperatures. (fig. 2) In both figures we saw that most countries had relatively small terror metrics, but the more extreme values tended to be in warmer countries.

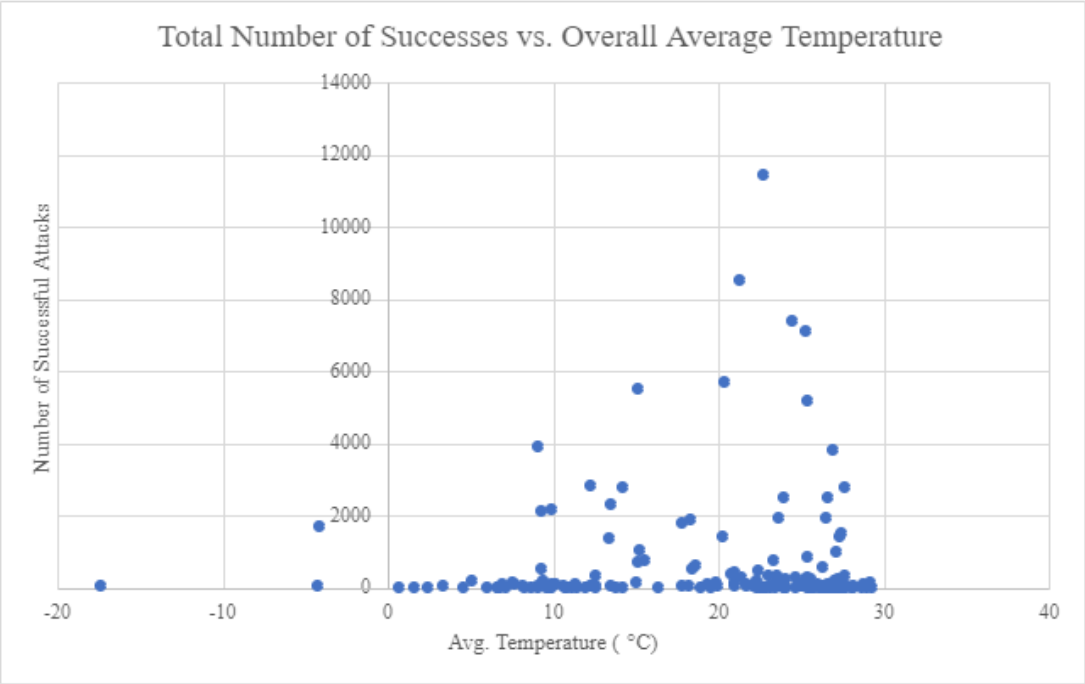

Figure 1. Total Number of Successful Terror Attacks vs. Overall Average Temperature by Country

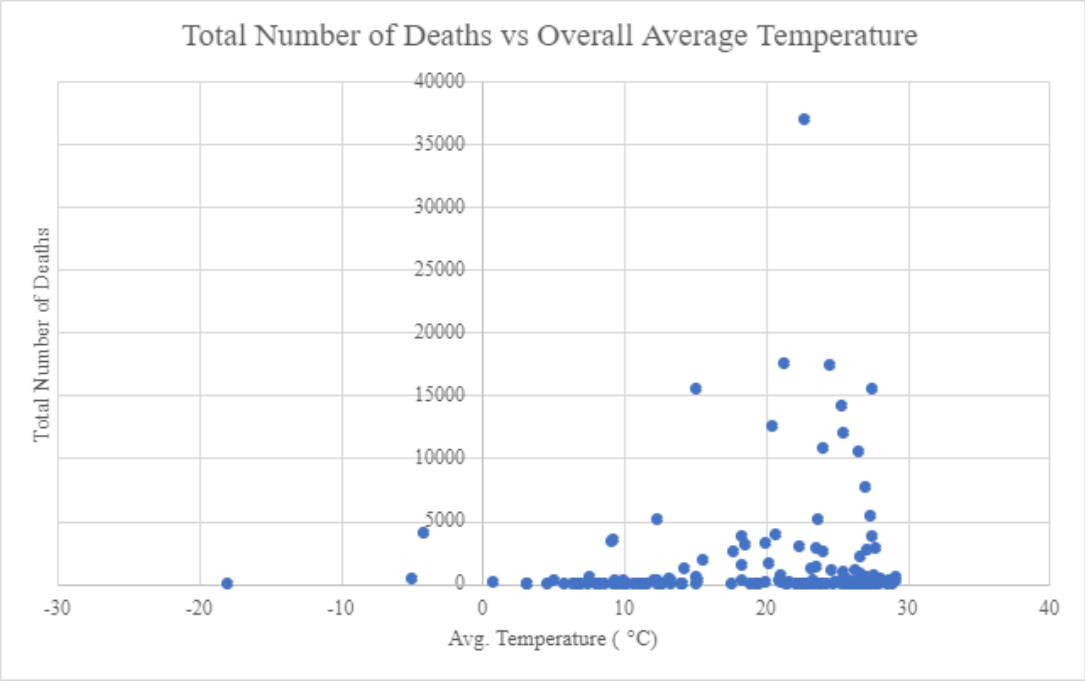

Figure 2. Total Number of Deaths Due to Terror Attacks vs. Overall Average Temperature by Country

Next, we looked at an individual country: The United States. For this we plotted the number of successful terror attacks for a certain year against the average temperature for the same year. (fig. 3) For this we noted that cooler years had a higher count of successful terror events, the opposite of the previous comparisons.

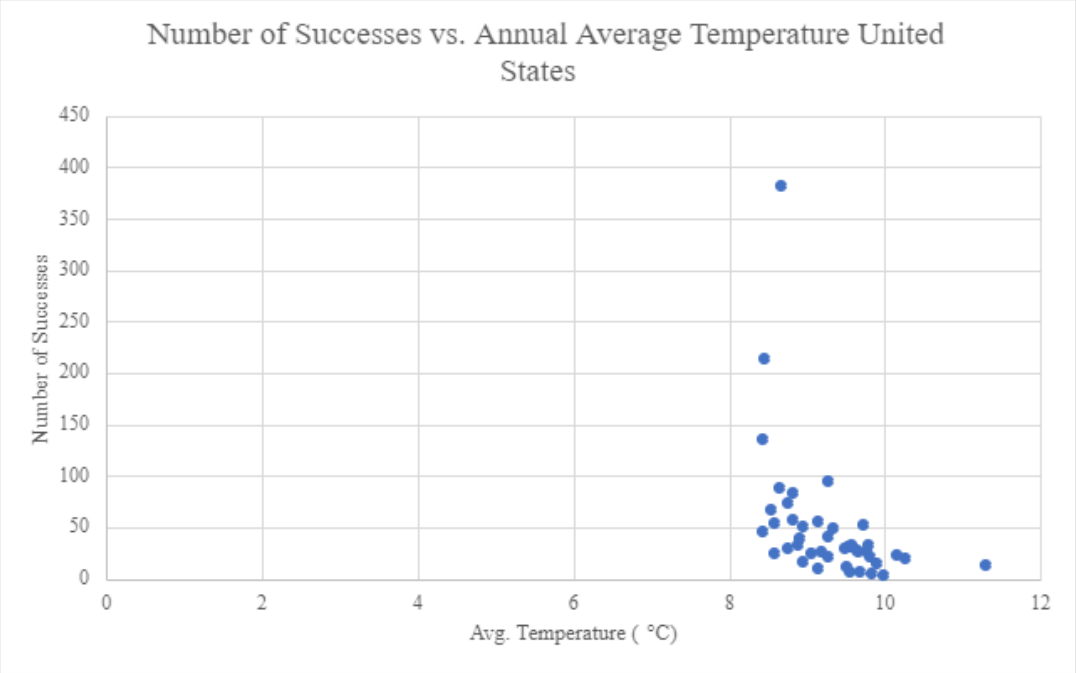

Figure 3. Annual Number of Successful Terror Attacks vs. Annual Average Temperature for the United States

Challenges

Our largest initial challenge was finding data that was large enough, interesting enough, and available enough for us to use and analyze. The Global Terrorism Database actually gave us some trouble because of its' size. Aside from the sheer number of columns, there were several columns whose values extended past the possible range of expected values. Initially this showed up as negative values in Boolean datatype columns which prevented us from uploading the data. In the end we decided to remove these columns to prevent us from running into this issue again. Additionally, we removed rows with negative values in columns where only positive values were expected.

Another part of the GTD that gave us trouble was the date information since it was into separate columns. We used Python and Pandas to combine these into a single date datatype column that would be accepted by PSQL. This was a larger than expected roadblock since PSQL didn't offer very much feedback on the formatting issues that we encountered. One issue we encountered with this was that the month and day entries for the day were required to have a leading zero if only one

digit was present. E.g. 2021-9-2 must be 2021-09-02. A more serious issue encountered was that some date entries had either the month, day, or both set to 0. This was not due to indexing that would be solved via an offset, but instead just bad data. We dropped all of the rows with either of these errors.

Conclusions

In the end, our results were not awfully exciting. Any of the patterns that we observed were likely due more to correlation than any form of causation. It would be somewhat expected that more developed nations would have a decreased level of terror instances. Currently more developed nations tend to be cooler which could explain the correlation observed in figures 1 & 2. The United States has become more developed over the years which could explain the decrease observed in figure 3. Global warming has also caused an increase in ground temperatures over the years which could also explain the trend seen. Because of this and the multitude of other factors at play, we would be very hesitant to make any conclusive remarks concerning our results. There is also the possibility other analysis could be affected by major terror events such as 9/11 would have if we were to observe number of deaths for the US rather than number of successes. However, it was interesting to see the data from the Global Terrorism Database and visualize said data. Data from the GTD is not as available or presented as often as the global temperature data is since global warming is a pressing issue with definitive causes that can be easily addressed on the individual level in addition to systematically.

Citations

[1] National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland. (2018). The Global Terrorism Database (GTD) [globalterrorismdb_0718dist.csv]. Retrieved from https://www.start.umd.edu/gtd Accessed: December 8, 2021

[2] Global Land Temperatures by Country [GlobalLandTemperaturesByCountry.csv] Retrieved from https://www.kaggle.com/vijayvvenkitesh/global-land-temperatures-by-country Accessed: December 8, 2021

APPENDIX A: Python code

```
import pandas as pd
import numpy as np
import os, sys


def main():
    path_in_terror  = os.path.join(sys.path[0], "globalterrorismdb.csv")
    path_out_terror = os.path.join(sys.path[0], "global_terrorism_cleaned.csv")
    path_in_temps  = os.path.join(sys.path[0],
"GlobalLandTemperaturesByCountry.csv")
    path_out_temps = os.path.join(sys.path[0], "global_temps_cleaned.csv")

    '''
        Global Terrorism Database
    '''
    data = pd.read_csv(path_in_terror, low_memory=False)
    df = pd.DataFrame(data, columns=['iyear', 'imonth', 'iday', 'country_txt',
'success', 'suicide', 'nkill', 'property', 'ransom'])

    #fix dates
    invalidDays = df[df['iday'] == 0].index
    df.drop(invalidDays, inplace=True)
    invalidMonths = df[df['iday'] == 0].index
    df.drop(invalidMonths, inplace=True)

    df['iday'] = df['iday'].apply(lambda x: '{0:0>2}'.format(x))
    df['imonth'] = df['imonth'].apply(lambda x: '{0:0>2}'.format(x))
    dates = df['iyear'].astype(str) + '-' + (df['imonth']).astype(str)  + '-' +
(df['iday']).astype(str)
    df.insert(0, 'date', dates)


    df = df.drop(['iyear', 'imonth', 'iday'], axis=1)

    df['nkill'] = df['nkill'].astype("Int64")
    df['ransom'] = df['ransom'].astype("Int64")

    df.loc[df['property'] <= 0, 'property'] = np.NaN
    df.loc[df['ransom'] <= 0, 'ransom'] = np.NaN
    df.loc[df['nkill'] <= 0, 'nkill'] = np.NaN

    df.dropna() #remove rows with invalid values

    df.to_csv(path_out_terror, index = False, header=True)  # save file
    print("saved cleaned terrorism data")

    '''
        Global Land Temperatures Dataset
    '''
    data = pd.read_csv(path_in_temps, low_memory=False)
```

```
    df = pd.DataFrame(data, columns=['dt', 'AverageTemperature', 'Country'])

    df.to_csv(path_out_temps, index = False, header=True)  # save file
    print("saved cleaned temperature data")
    print("done")



if __name__ == "__main__":
    main()
```

APPENDIX B: PSQL Queries

```
-- load data onto server
DROP TABLE IF EXISTS temp CASCADE;
DROP TABLE IF EXISTS terrorism CASCADE;

CREATE TABLE temp (
    date DATE,
    avg_temp NUMERIC,
    country TEXT);

\COPY temp FROM
'C:\Users\wangd\OneDrive\Desktop\CSCI403\final\global_temps_cleaned.csv' WITH
(FORMAT csv);


CREATE TABLE terrorism (
    date DATE,
    country TEXT,
    success BOOLEAN,
    suicide BOOLEAN,
    people_killed INTEGER,
    property BOOLEAN,
    ransom BOOLEAN);

\COPY terrorism FROM
'C:\Users\wangd\OneDrive\Desktop\CSCI403\final\global_terrorism_cleaned.csv' WITH
(FORMAT csv);



-- create combination table for the two datasets, grouped by year and country
DROP TABLE IF EXISTS combo_table;

CREATE TABLE combo_table
    AS SELECT temp_year, temp_country, temp_avg, terror_success, terror_nkills,
terror_suicide, terror_property, terror_ransom

    FROM
        (SELECT date_trunc('year', date) AS temp_year, country AS temp_country,
avg(avg_temp) AS temp_avg FROM temp GROUP BY temp_year, temp_country) AS
new_temp,

        (SELECT date_trunc('year', date) as terror_year, country AS
terror_country, count(CASE WHEN success THEN 1 END) AS terror_success,
sum(people_killed) AS terror_nkills, count(CASE WHEN suicide THEN 1 END) AS
terror_suicide, count(CASE WHEN property THEN 1 END) AS terror_property,
count(CASE WHEN ransom THEN 1 END) AS terror_ransom FROM terrorism GROUP BY
terror_year, terror_country) AS new_terror

    WHERE temp_country = terror_country AND temp_year = terror_year;
```

```
-- create table byKilled which includes average temperatures and total people
killed grouped by countries from the combinaiton table
CREATE TABLE byKilled AS SELECT temp_country, AVG(temp_avg), SUM(terror_nkills)
FROM combo_table WHERE terror_nkills IS NOT NULL GROUP BY temp_country ORDER BY
SUM(terror_nkills) DESC;

-- export the byKilled table to a local csv file
\COPY byKilled TO 'C:\Users\wangd\OneDrive\Desktop\CSCI403\final\byKilled.csv'
CSV;

-- create table bySuccess which includes avergage temperatures and total
successful terrorism attacks grouped by countries from the combination table
CREATE TABLE bySuccess AS SELECT temp_country, AVG(temp_avg), SUM(terror_success)
FROM combo_table WHERE terror_success IS NOT NULL GROUP BY temp_country ORDER BY
SUM(terror_success) DESC;

-- export the bySuccess table to a local csv file
\COPY bySuccess TO 'C:\Users\wangd\OneDrive\Desktop\CSCI403\final\bySuccess.csv'
CSV;

-- create table US which includes US's average temperatures and terrorism data
from all recorded years
SELECT * from combo_table WHERE temp_country = 'United States' ORDER BY
temp_year;

-- export the US table to a local csv file
\COPY US to 'C:\Users\wangd\OneDrive\Desktop\CSCI403\final\US.csv' CSV;
```