Natural Disasters Over Time - Final Project Report

by Haley Egan, Maggie Houck, William Johnson, and Diana Morris

UVA School of Data Science, DS 5100

Introduction

Climate Change is a major crisis facing humanity. Communities are facing unprecedented challenges such as drought, wildfires, flooding, extreme temperatures, and more. Scientists have communicated that as Climate Change continues, these natural disasters will increase in frequency and severity.

In order to gain more insight into natural disasters and their trends over time, we examined two datasets, one on natural disasters and another on the economic impacts of these natural disasters.

We wanted to answer the following questions:

- Are natural disasters increasing over time?
- Are natural disasters increasing in severity over time?
- Are natural disasters and their severity linked to Climate Change?

Our work can be found on GitHub, at https://github.com/HaleyEgan/Climate-Change-Natural-Disasters.

Data

Our data set is drawn from the International Disaster Database (EM-DAT), curated by the Centre for research on the Epidemiology of Disasters, at the Catholic University of Louvain. This larger set is updated continuously, with qualifying disasters added as they are received, though the public version is updated to reflect the changes once monthly. All disasters are independently verified before being added to this list. The data is highly reliable, but because of the verification process, there is a lag between the disaster and its addition to the data set.

The original data set is far larger in that it includes all "disasters", including technological and hazardous waste disasters, like chemical spills, and several other details such as location, exact date(s), magnitude of disaster by relevant measure (e.g. wind speed), etc. Our subset is restricted to natural disasters and economic damages. Moreover, it is a snapshot from 2018, consolidated on Kaggle. For this reason, we eventually chose to restrict our analysis to the years up to 2015 because many qualifying disasters were not visible in the snapshot because of the reporting lag time.

Defining a Natural Disaster

The requirements for a qualifying disaster are clearly established and the variables are carefully described in the EM-DAT guidelines, per this website: https://public.emdat.be/about.

The primary criteria for a disaster to be included in the data set is one of more of the following:

- At least 10 deaths
- At least 100 people affected (including injured or homeless)
- The declaration of a state of emergency or an appeal for international assistance.

In the event that exact numbers were missing, secondary criteria were also considered, e.g. if a disaster was recognized as the worst in that category for the given country.

Defining Economic Damage

The figures for total economic damage are derived from the following:

- The total estimated damages (i.e. economic losses from damaged property, infrastructure, environmental, etc.)
- Reconstruction costs (the cost of *replacing* lost assets, which differs from the previous in that it involves the present cost of goods, and any costs associated with prevention or mitigation measures regarding future disasters)
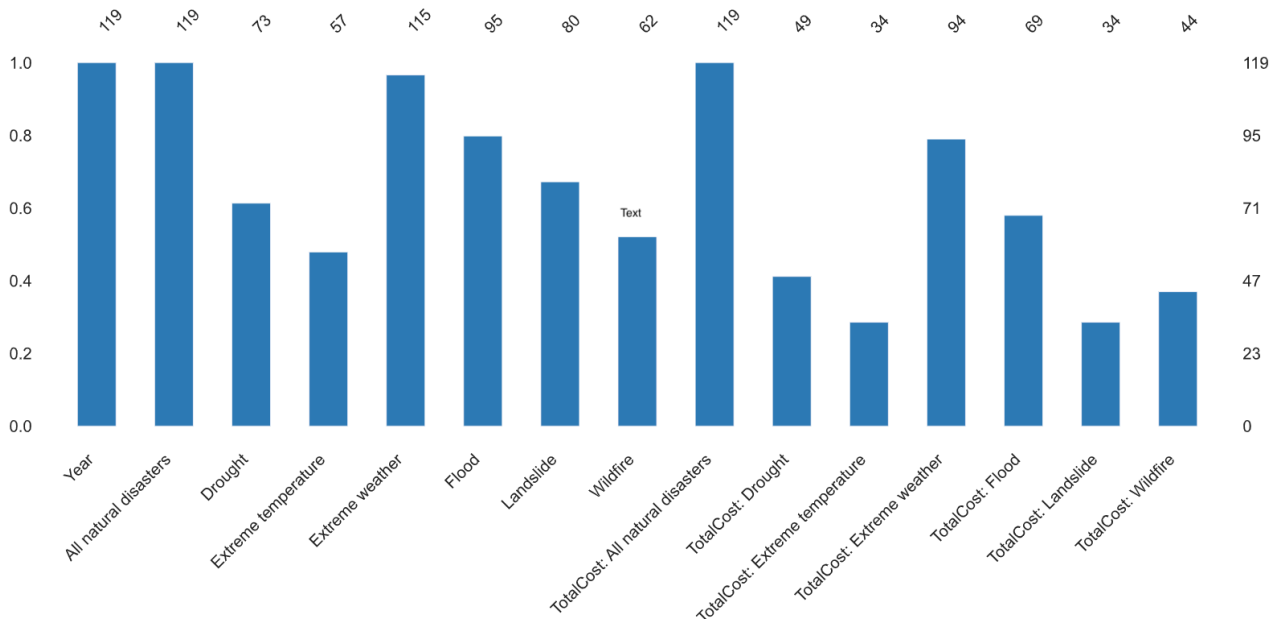- Insured losses (economic damages covered by insurance companies).

All amounts are given in thousands of U.S. dollars, as valued in 2019.

Data Cleaning

While the data we obtained had already been cleaned and organized, it was necessary to format and categorize the data to our specific needs.

The types of natural disasters included in the original dataset were: drought, earthquake, extreme temperature, extreme weather, flood, impact, landslide, mass movement (dry), volcanic activity, wildfire, and information on all these natural disasters combined. Since we were interested in natural disasters in relation to Climate Change, it was not necessary for our analysis to keep all of the variables. We removed earthquakes, volcanoes, mass movement (dry), and impact, as these variables are largely independent from Climate Change. We also changed some of the variable names to be more clear, such as renaming 'Entity' as 'Disaster'. After removing several variables, we updated the 'All Natural Disasters' variable to only include the selected variables for our dataset. We also removed the 'code' column, and organized the data by date, and then disaster type, instead of by the original disaster type. We combined the economic data set with the natural disaster data set for easier analysis and comparisons among natural disaster types and trends over time.

Missing Data

Year 119
All natural disasters 119
Drought 73
Extreme temperature 57
Extreme weather 115
Flood 95
Landslide 80
Wildfire 62
TotalCost: All natural disasters 119
TotalCost: Drought 49
TotalCost: Extreme temperature 34
TotalCost: Extreme weather 94
TotalCost: Flood 69
TotalCost: Landslide 34
TotalCost: Wildfire 44

Dataset statistics

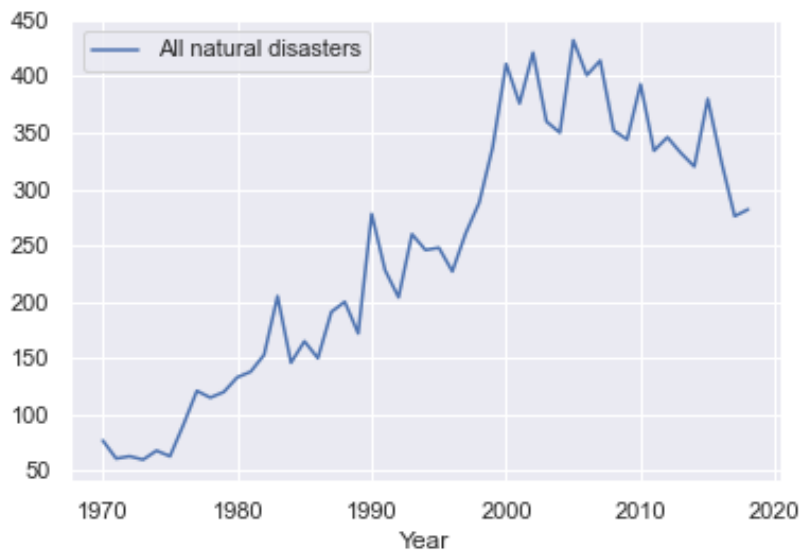|	
Number of variables	15
Number of observations	119
Missing cells	622
Missing cells (%)	34.8%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	14.1 KiB
Average record size in memory	121.1 B

Variable types

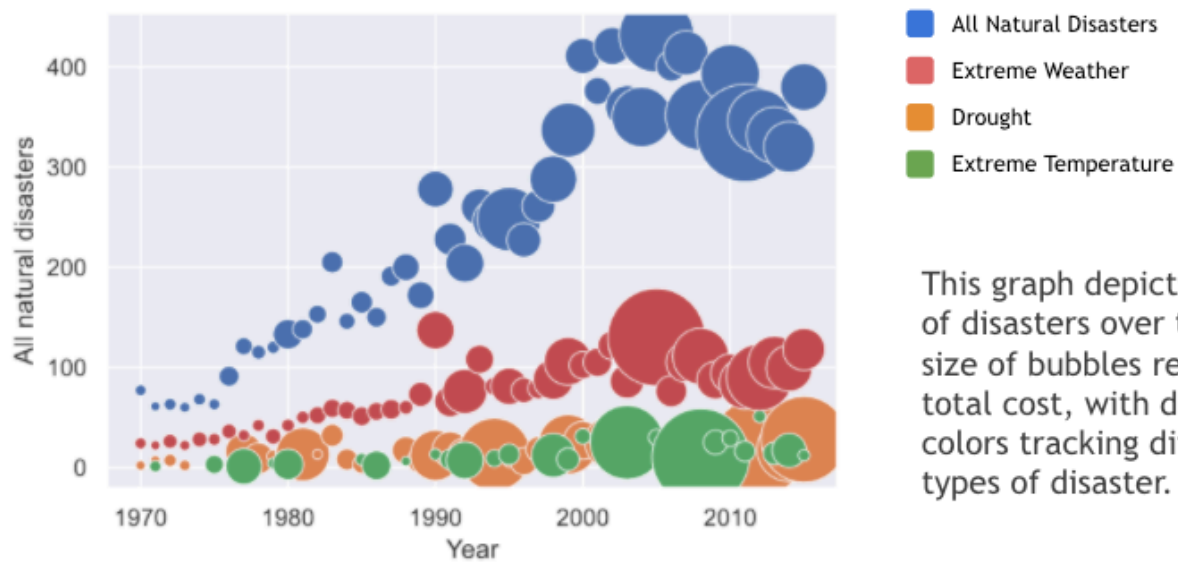|	
Numeric	15

Exploratory Data Analysis
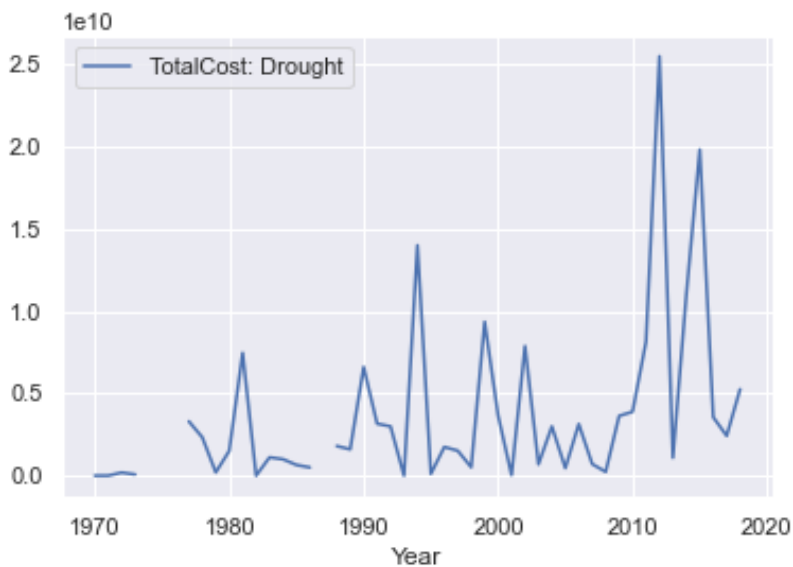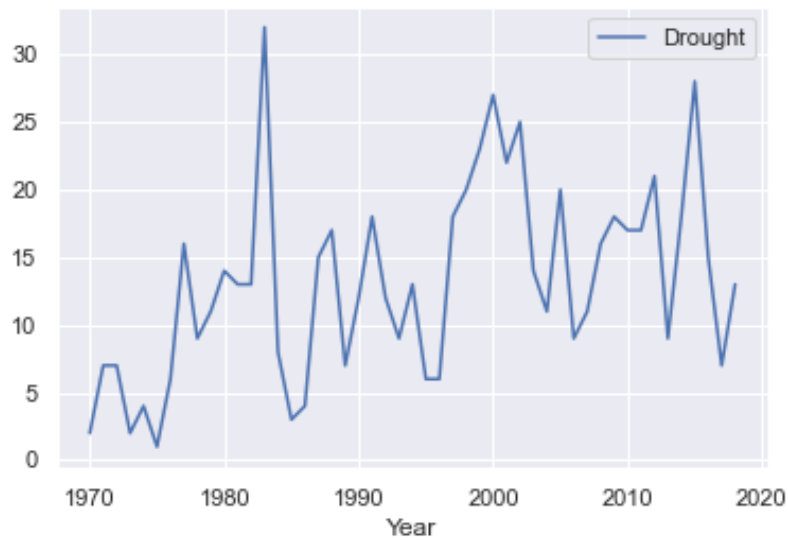
Correlation Matrix
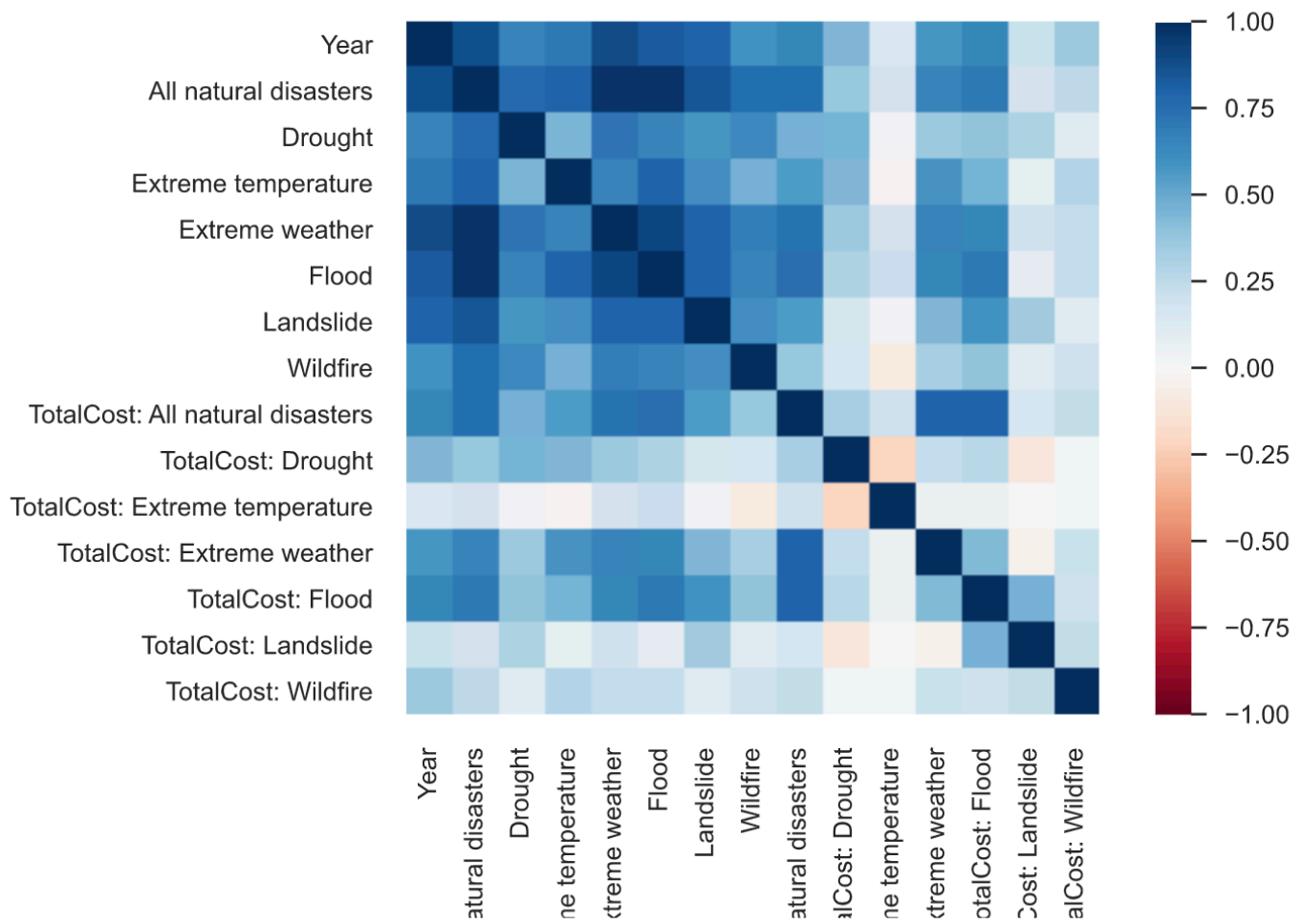
All Natural Disasters
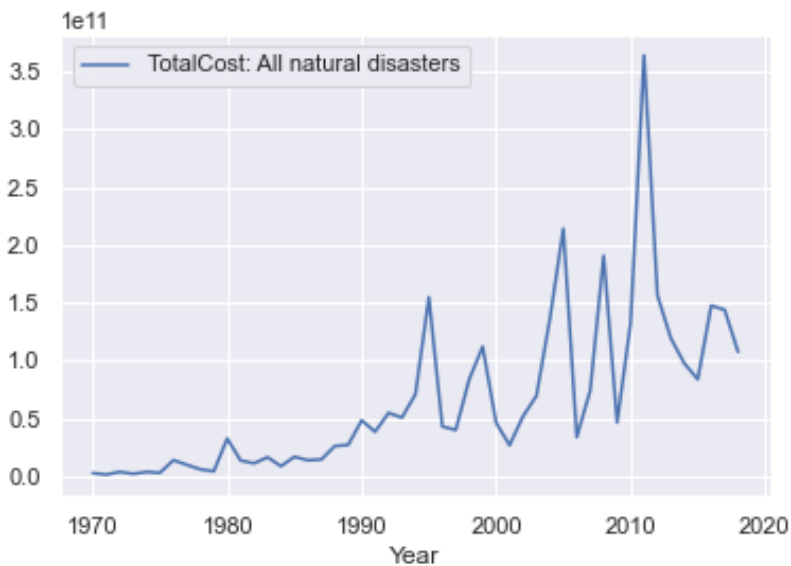
From some basic line graphs, we can see the trend of all global natural disasters recorded since 1970. Since 1970, there is an exponential upward trend until the early 2000s. There is a decrease in natural disasters after 2010, but the number of disasters is still high. The second graph visualizes the total cost of natural disasters since 1970. The cost of natural disasters follows the similar growth trend, with a peak around 2010.

This graph depicts the trend of disasters over time. The size of bubbles represents the total cost, with different colors tracking different types of disaster.

Drought

These line graphs examine the trend of droughts around the world since 1970, and the total economic impact of the droughts. There have been several droughts since 1970, with spikes around 1984, 2000, and 2015. While there have been many droughts the level of natural disasters, the economic impact of the droughts appears to be increasing. After 2010, the total cost of droughts is significantly higher than any other year.

Extreme Weather
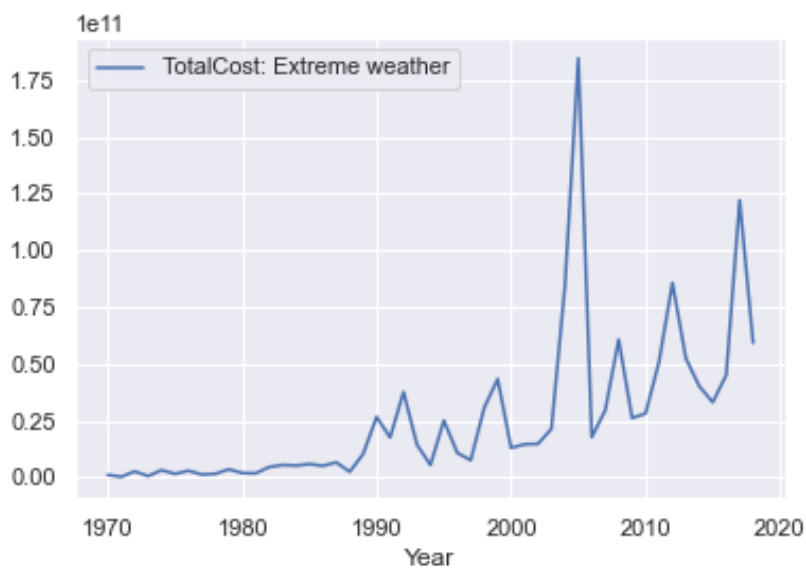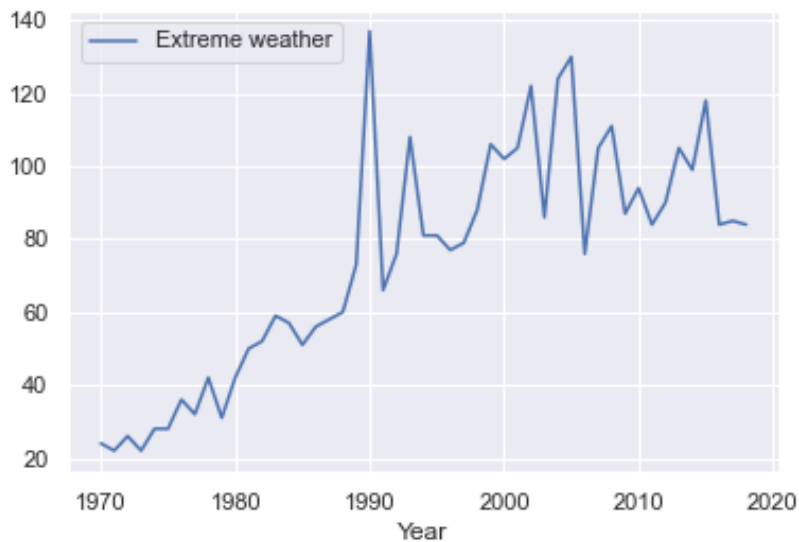
Extreme weather is an aggregate of a few natural disaster types: typhoons, tornadoes, and hurricanes. Based on the left line graph, there appears to be a steady increase in extreme weather events since 1970. The right graph also communicates the increase in economic impact of extreme weather events over time.

Extreme Temperature
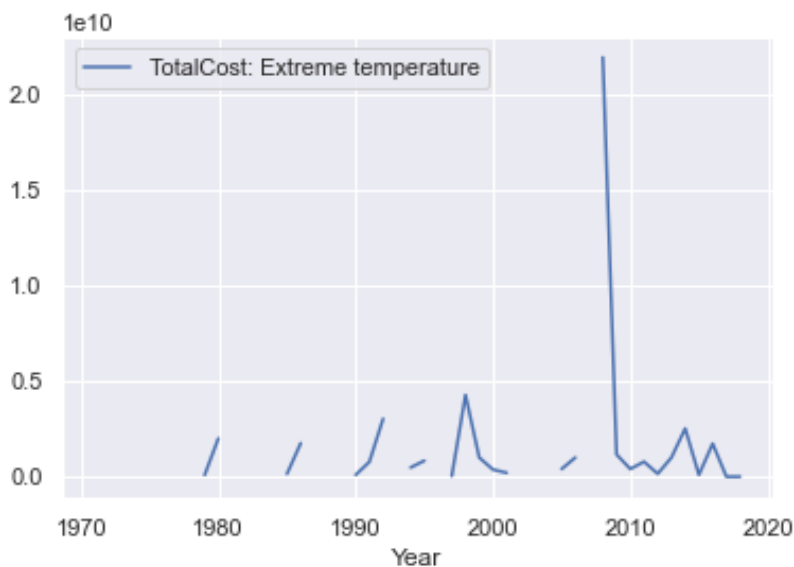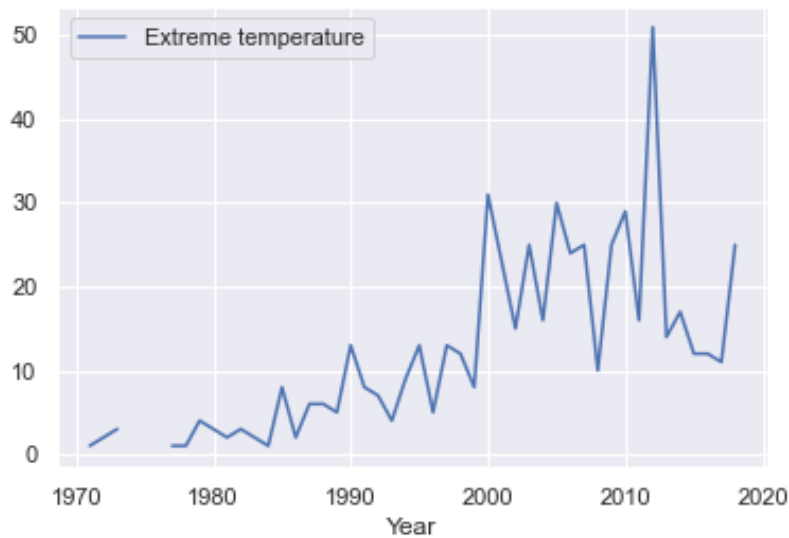
Extreme temperature appears to be increasing steadily since 1970, with a spike around 2010, that is consistent with the sharp increase in economic cost of extreme temperature around 2010. Extreme temperature seems to fluctuate from year to year, but the graph shows a consistent upward trend over time. Overall, the economic impact of extreme temperature does not appear to have the same economic impact as other natural disasters.

Unittesting

In order to support reproducibility and future work with this dataset and code, we created unit tests to test expected inputs and outputs for the dataset.

Conclusion

From our initial visual analysis, there appears to be an increase in the number of natural disasters over time

From our initial visual analysis, there appears to be an increase in the number of natural disasters over time, for all natural disaster types, as well as an increase in economic impact. The increase in economic impact may indicate the increase in severity of the natural disasters over time. From the visual data analysis of six types of natural disasters (drought, extreme weather, extreme temperature, flood, landslide, and wildfire), there was a clear upward trend for all disasters. This suggests that natural disasters are increasing over time, and are increasing in severity. With our current dataset, we are unable to conclude that these increases are directly tied to Climate Change, something that will take further analysis with a larger scope.

Future Work

Our work with this dataset is just the beginning. While there does appear to be an increase in natural disasters over time, and their severity is increasing over time due to increasing economic impact, there are many questions left unanswered. The direct link between natural disasters and Climate Change cannot be made with our current dataset, because the dataset only examines the increase in natural disasters and costs. In order to compare the trend of natural disasters with known Climate Change contributors, we recommend adding data on indicators such as greenhouse gases and emissions. It could be interesting to compare the trend of emissions/greenhouse gases over time with the frequency of natural disasters to see if there is a correlation.

We also noticed that there is a trend throughout the data showing a decrease in all types of natural disasters after 2010. It is unclear if this dip is accurate, or an error with the collection of the data. Our team contacted the managers of the dataset to learn more about the data collection, and discovered that there can be a several month or year delay in adding some data to the dataset, due to an extensive verification process. However, it is not clear if this is the sole reason for the dip in data.

The next step with our current dataset on natural disasters and economic impact is to perform statistical analysis such as linear regression modeling and predictive modeling. It will be interesting to model the trend of data over time for each type of natural disaster, and predict the trend in the future. People can use our data to predict the severity and cost of future natural disasters. Our data can also be used to examine trends of natural disasters around the world, and to predict which areas and communities are most likely to be impacted in the future.

References

OFDA/CRED International Disaster Database, Université catholique de Louvain. EMDAT, 2019. Brussels, Belgium. https://ourworldindata.org/natural-disastershttps://ourworldindata.org/natural-disasters.

Natural Disaster Data: Occurrence and Economic Impact, 2018. Kaggle. https://www.kaggle.com/dataenergy/natural-disaster-data?select=economic-damage-from-natural-disasters.csv.