

# Research on US Natural Disasters Data

## 1. Introduction

Natural disasters are a problem we have faced for a long time, and their occurrence can have a huge impact on people's lives and property. Our data comes from Kaggle and is recorded by FEMA (Federal Emergency Management Agency). The data recorded detailed information on disasters that have occurred on US soil since 1953. We hope that through this data, we can find out which factors are associated with the occurrence of various types of natural disasters, such as geographic location, climate, environment, and so on.

## 2. Variable Selection

Since we are focusing on the types of disasters and locations the disasters happen. We will choose the following variables as our dataset.

**disaster\_number**: Sequential number used to designate an event or incident declared as a disaster.

**state**: US state, district, or territory.

**declaration\_type**: "DR": major disaster, "EM": emergency management, or "FM": "fire management"

**fy\_declared**: Fiscal year in which the disaster was declared.

**incident\_type**: Type of incident such as "Fire", "Flood", or "Hurricane". The incident type will affect the types of assistance available.

After selecting variables, our data frame has 63029 observations and 5 variables with no NaNs.

## 3. Analysis

**Question 1: Which disaster occurs most frequently and in which state? What are the possible reasons this kind of disaster happen so often in that state?**

[illegible]

After sorting the data, we have a data frame showing the numbers of the most frequent disasters in each state. Then we find the most disasters in each state.

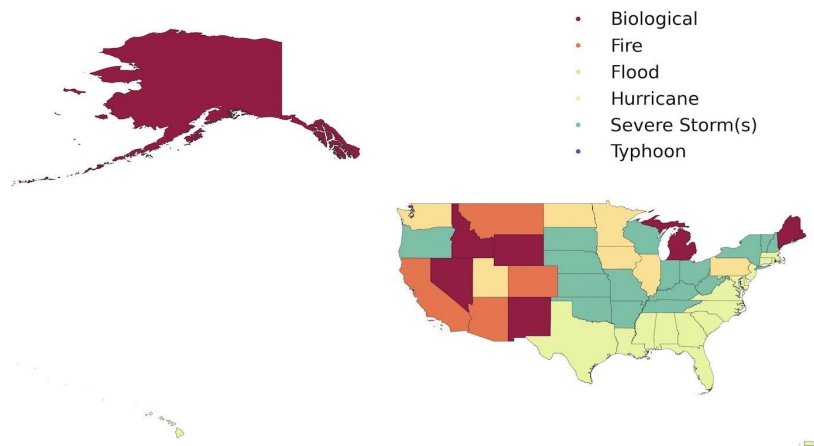


Figure 2 - Most Frequently Disasters in the US

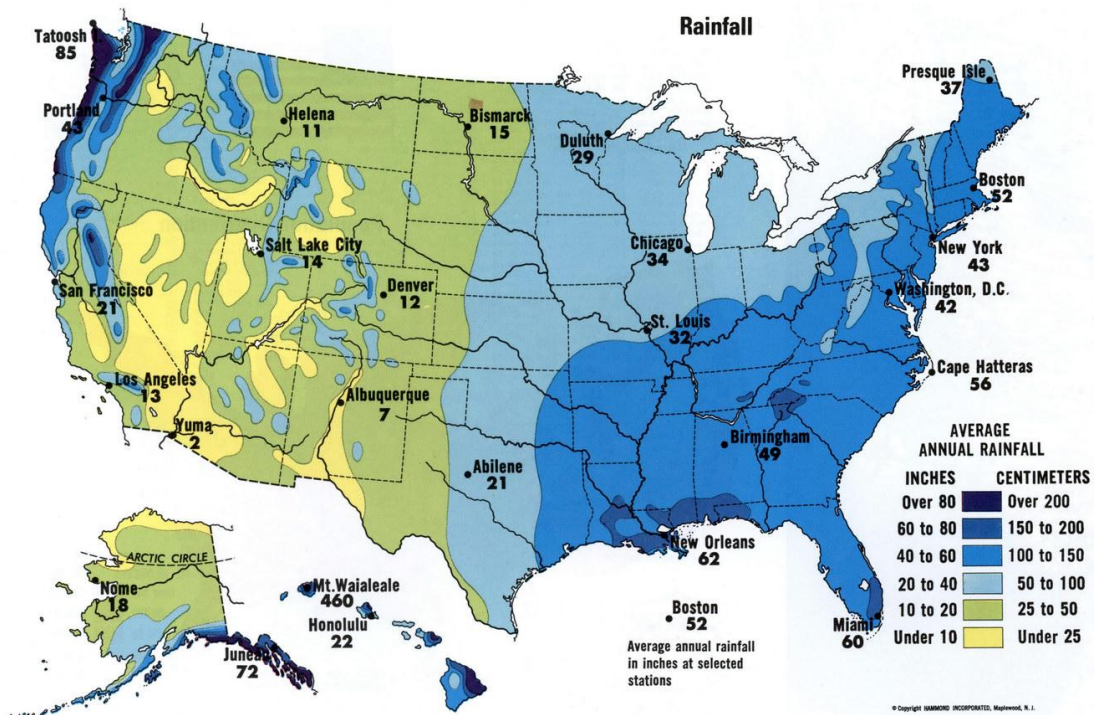


Figure 3 - US Rainfall Map. Source: <https://us-canad.com/rainfall-usa-map.html>

**Question 2: Find the trend of numbers of disasters recorded each year. Which year has the most disasters and the least disasters? What are these disasters?**

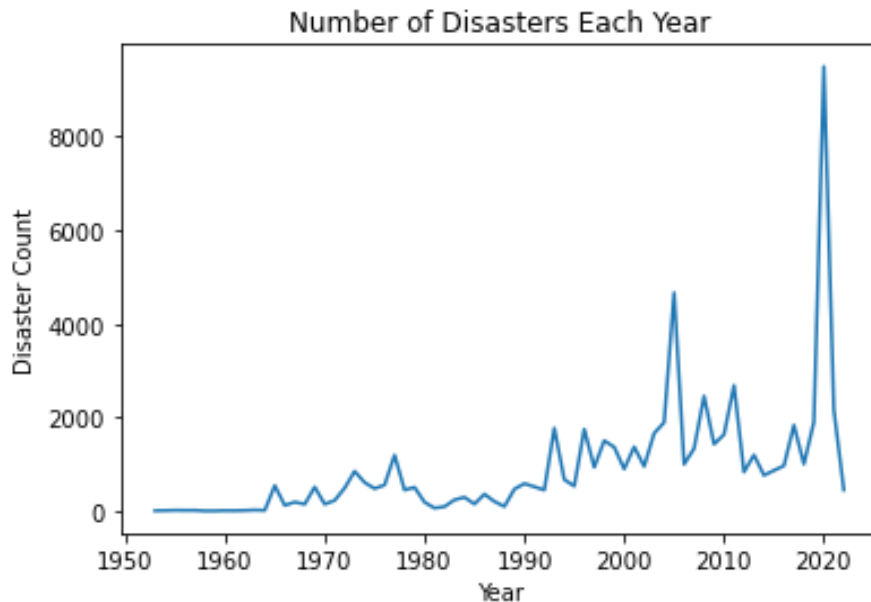


Figure 4 - Line Chart: Number of Disasters Each Year

```

fy_declared
2020    9490
2005    4661
2011    2684
2008    2456
2021    2166
...
1960     13
1961     11
1953     10
1959      7
1958      6
Name: count, Length: 70, dtype: int64

```

Figure 5 - Sorted Number of Disasters Each Year

According to the line chart and the output values. We can find that the year 2020 has the most disasters. There are 9490 disasters in 2020. However, the year 1958 has the least disasters. There are 6 disasters this year. From 1953 to 1963, we can find that the total number of disasters in these years was relatively small. This may be caused by incomplete statistics.

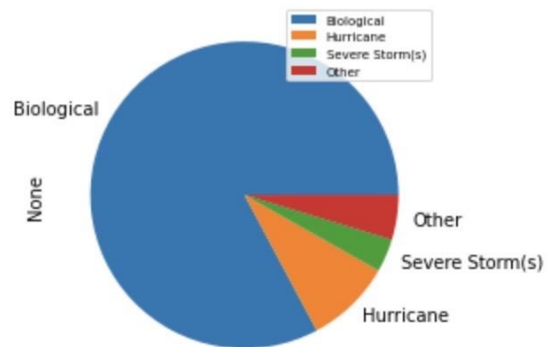
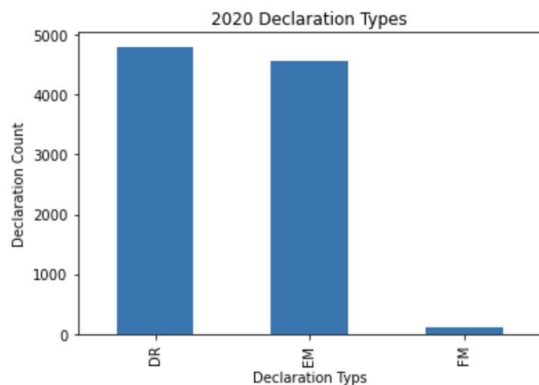


Figure 6 - Bar plot: Number of declaration types Figure 7-Pie chart: Proportion of disasters in 2020

In 2020, the total number of disasters is 9490. From the pie chart, we can find that biological disaster has the largest proportion which approximately equals to 83% and from the bar plot, we can see that almost half of the disaster types belong to emergency, which corresponds to the 2020 covid-19 pandemic and various emergencies caused by covid. Meanwhile, typhoons, one of the disasters this year, have the least proportion, which is 0.84%.

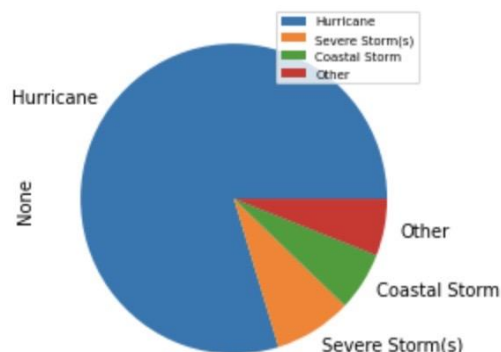
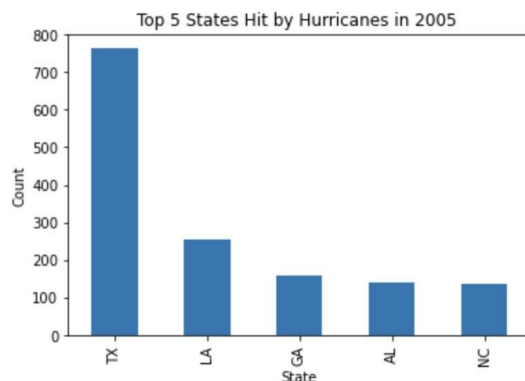


Figure 8- Bar plot: Top 5 states hit by hurricanes in 2005 Figure 9- Pie chart: proportion of disasters in 2005

According to the line chart and output value, another spike from 1953 to 2021 took place in the year 2005, in which 4661 disasters were recorded. From the pie chart, we can identify that the major disaster to blame was the hurricane, which accounts for 79% percent of the whole year.

The second and third places were severe storms and coastal storms. This finding corresponds to major hurricanes that happened in 2005, such as Hurricane Katrina and Hurricane Rita caused tremendous damage to Louisiana and Texas. According to the bar plot, we can identify the top 5 states hit by hurricanes in 2005 were Texas, Louisiana, Georgia, Alabama,, and North Carolina. Among them,, Texas was the highest with 763 disasters reported, which is more than twice as high as the second-highest.

## 4. Discussion

The first and most difficult part of this analysis is the huge size of the data. We have data from over fifty states and nearly a century of data. Parsing through the data to find anomalies or errors is challenging and time-consuming. However, this is a critical first step in conducting the analysis. If the cleaned data does not reflect the real world, then the corresponding summaries and explorations will never capture any aspect of reality, leading to seriously flawed data and inferences in the future. Our main focus in this study is on reality-based data. The Kaggle website has already cleaned its data so we can use them directly.

We have many research questions to discuss but due to skill and space limitations, we can only discuss a few of them that we think are meaningful and implementable.

We used data visualization to allow us to visually notice obvious patterns and phenomena. Since there are a certain number of variables in the dataset, we can adjust or omit those that are needed depending on the question we are discussing.

In addition, measurement errors or recording errors in the original data can seriously distort the conclusions in the response reality. However, we did not have access to the original data recording and entry so our current study is also based on the fact that the data are all true and accurate.

## 5. Conclusion

By compiling our data, we first identified what the most frequent disasters were in each state and correlated them with factors of the natural environment such as geography. We concluded that coastal states are more prone to hurricanes. And western states with less precipitation are more likely to have fires. And states adjacent to the Mississippi River are more prone to flooding. According to the pie chart and line graph, we find that the year 2020 and 2005 have the most disasters. The year 1958 has the least disasters. In the period from 1953 to 1963, we can find that the total number of disasters in these years is relatively low. In 2020, biological disasters have the largest proportion, and typhoons have the smallest proportion. So we need to be more concerned about the growth of biological disasters in the future and the warning of disasters by location.

## 6. References

- Tails, H. or. (2022, May 20). *US Natural Disaster Declarations*. Kaggle. Retrieved May 24, 2022, from <https://www.kaggle.com/datasets/headsortails/us-natural-disaster-declarations>
- United States Census Bureau.(2021, October 8). *Cartographic Boundary Files - Shapefile*. Census.gov. <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>