Extreme Weather Has Affected One in Three Americans: Analyzing Extreme Weather Events Trends since 2000

W200 project 2, 11/30/2022

Mahmoud Ghanem, Reed Evans, Ryan Chen

GitHub: https://github.com/UC-Berkeley-I-School/Project2_Chen_Evans_Ghanem

Primary dataset: https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/

Introduction

According to the article: Extreme Weather Has Affected One in Three Americans and a Gallup poll released in April. 2022, One in three Americans say they have personally been affected by an extreme weather event in the last two years, reflecting a period that included a paralyzing ice storm in Texas, one of the largest wildfires in California's history, and a powerful storm that killed dozens of people in the New York region. The survey was based on interviews conducted last month with about 1,000 adults living in all 50 states and Washington, D.C. Furthermore, the survey also revealed that extreme-weather victims are more concerned about climate change.

In this project, unlike conducting a survey in the article, we are trying to view the story from the perspective of extreme weather events data from NOAA, describe the impact of extreme weather across the country in general scope, and dive into the most impressive extreme weather events, and explicitly demonstrate them through analysis of data.

Americans' Experiences with Extreme Weather

In the past two years, have you been personally affected by an extreme weather event where you live, or not? What type of extreme weather were you affected by?

	U.S. Adults
	%
Yes, affected	33
(Extreme cold)	(7)
(Hurricane)	(6)
(Snow/Ice storm/Blizzard)	(5)
(Extreme heat)	(5)
(Flood)	(4)
(Tornado)	(4)
(Fire or wildfire)	(4)
(Drought)	(3)
(High winds)	(3)
(Extreme thunderstorms)	(2)
(Extreme rain/Hail)	(1)
(Earthquake)	(1)
(Other)	(1)
No, not affected by extreme weather	67

The sum of the weather events exceeds 33% because some respondents named multiple weather events.

GALLUP, MARCH 1-18, 2022

Global Warming Attitudes Among Those Affected Versus Not Affected by Extreme Weather Events

Table shows responses among those who have personally been affected by an extreme weather event where they live in the past two years versus those who have not been affected by such an event.

	Affected by Extreme Weather	**Not Affected by Extreme Weather**	**Difference**
	%	%	pct. pts.
Worry "a great deal" about global warming	63	33	+30
Believe global warming effects "have already begun to happen"	78	51	+27
Believe seriousness of global warming "generally underestimated" in the news	59	30	+29
Say global warming will pose "a serious threat" to their way of life	64	36	+28
Say they understand global warming issue "very well"	33	23	+10
Say government doing "too little" to protect environment	67	48	+19
Prioritize environmental protection over economic growth	67	47	+20

GALLUP, MARCH 1-18, 2022

Data source:https://news.gallup.com/poll/391508/extreme-weather-affected-one-three-americans.aspx

Questions that we are looking to explore:

1. The impact of extreme weather events since 2000, What are the trends of these weather events?
2. What are states with the most property loss/death caused by extreme weather events
3. What was happening to specific events such as the 1999 Heatwave, Katrina, and 2011 Tornadoes

Data Structure

NOAA database has been tracking tornado data since 1950. In 1992,they began tracking all storm events in database. For the storm_event_details there are about 1.7 million rows and 51 columns originally downloaded from NOAA and through the EDA process, the data set is narrowed down to 1259,884 rows and 8 columns. The EDA process will be introduced in the sanity check session explicitly. For a detailed analysis of fatalities, we will use the storm_fatalities file provided by NOAA in the same repository. We join on the event_id key, however, we need to group by on either side of the joining table to ensure uniqueness and avoid many too many scenarios.

storm_event_details

- **date (datetime64):** When events occurred, with format %y-%m%d.
- **year(int64):** The year when events occurred.
- **month(int64):** The month when events occurred, in numerical format, 1 means January.
- **day(int64):** The day in a month when events occurred.
- **state(object):** States where events occurred in the U.S., in the full state name
- **event_id(object):** ID assigned by NWS for each individual storm event contained within a storm episode; links the record with the same event in the storm_fatalities dataset
- **event_Type(object):** Type of extreme weather events
- **ev_type_grp(object):** Manually categorized groups of extreme events
- **injuries_direct(int64):** Direct injuries caused by events, in numerical format
- **injuries_indirect(int64):** Indirect injuries caused by events, in numerical format
- **deaths_direct(int64):** Direct deaths caused by events, in numerical format
- **deaths_indirect(int64):** Indirect injuries caused by events, in numerical format
- **damge_property(float64):** Property loss caused by events, in numerical format, adjusted by the inflation rate according to the U.S Bureau of Labor Statistics.

storm_fatalities (merge on event_id)

- **cz_fips(int64):** The county FIPS number is a unique number assigned to the county by the National Institute for Standards and Technology (NIST)
- **state_fips(int64):** The state FIPS is a unique number assigned to the state by the National Institute for Standards and Technology (NIST)
- **fatality_age(int64):** The age in years of the fatality (sometimes ‘null’ if unknown)
- **fatality_sex(object):** The gender of the fatality (sometimes ‘null’ if unknown)
- **fatality_location(object):** the location of the fatality, e.g. Permanent Home, Mobile/Trailer Home, Boating

Sanity Check

The pristine data of extreme weather events from NOAA contains 51 columns and since we want to find out the impact of these events since 2000 based on the fact of climate change, we then narrowed down the columns to columns having necessary information and columns that can help us identify duplicates such as event id, time, states, event types, injuries, deaths, and property loss.

	EVENT_ID	YEAR	MONTH_NAME	BEGIN_DAY	BEGIN_TIME	END_TIME	STATE	EVENT_TYPE	INJURIES_DIRECT	INJURIES_INDIRECT	DEATHS_DIRECT	DEATHS_INDIRECT	DAMAGE_PROPERTY
0	5165377	2000	12	31	600	900	FLORIDA	Extreme Cold/Wind Chill	0	0	0	0	NaN
1	5165378	2000	12	31	600	900	FLORIDA	Extreme Cold/Wind Chill	0	0	0	0	NaN
2	5165379	2000	12	31	700	800	FLORIDA	Extreme Cold/Wind Chill	0	0	0	0	NaN
3	5165449	2000	12	13	2200	400	WEST VIRGINIA	Winter Storm	0	0	0	0	NaN
4	5172568	2000	8	3	1410	1410	MISSISSIPPI	Thunderstorm Wind	0	0	0	0	2K

Referring to the data set above, we assign event_id as the primary key, which means the data are input separately. The top two entries have different event_id but they happened at the same time (same year, month, day, and specific time). Originally, we consider these top two entries as duplicates. However, along with data exploration, we find out that owing to the reasons that we picked 8 out of 51 columns and a single event might be reported separately with different information. So if we drop the duplicate entries, we may lose some important information. As a result, we decide to keep the duplicates.

	EVENT_ID	YEAR	MONTH_NAME	BEGIN_DAY	BEGIN_TIME	END_TIME	STATE	EVENT_TYPE	INJURIES_DIRECT	INJURIES_INDIRECT	DEATHS_DIRECT	DEATHS_INDIRECT	DAMAGE_PROPERTY
0	5165377	2000	12	31	600	900	FLORIDA	Extreme Cold/Wind Chill	0	0	0	0	NaN
1	5165378	2000	12	31	600	900	FLORIDA	Extreme Cold/Wind Chill	0	0	0	0	NaN
2	5165379	2000	12	31	700	800	FLORIDA	Extreme Cold/Wind Chill	0	0	0	0	NaN
3	5165449	2000	12	13	2200	400	WEST VIRGINIA	Winter Storm	0	0	0	0	NaN
4	5172568	2000	8	3	1410	1410	MISSISSIPPI	Thunderstorm Wind	0	0	0	0	2K

```
# US States: #'FLORIDA', 'WEST VIRGINIA', 'MISSISSIPPI', 'MAINE', 'CONNECTICUT','GEORGIA', 'NORTH CAROLINA', 'ARIZONA', 'TEXAS', 'TENNESSEE', 'MARYLAND', 'NEW YORK', 'OREGON', 'NEW JERSEY',
            # 'KANSAS', 'MICHIGAN', 'OKLAHOMA', 'PENNSYLVANIA', 'LOUISIANA', 'DELAWARE', 'CALIFORNIA', 'COLORADO', 'INDIANA', 'IOWA', 'ALASKA', 'MONTANA', 'OHIO', 'ILLINOIS', 'ARKANSAS',
            # 'HAWAII', 'WASHINGTON', 'MINNESOTA', 'IDAHO', 'WYOMING', 'SOUTH DAKOTA', 'ALABAMA','VERMONT', 'NEW HAMPSHIRE', 'NORTH DAKOTA', 'MISSOURI', 'DISTRICT OF COLUMBIA', 'NEVADA',
            # 'RHODE ISLAND', 'NEBRASKA', 'NEW MEXICO', 'UTAH', 'KENTUCKY', 'MASSACHUSETTS',, 'VIRGINIA', 'WISCONSIN', 'SOUTH CAROLINA'

# Not US: # 'PUERTO RICO','AMERICAN SAMOA','GUAM', 'LAKE ERIE', 'LAKE HURON', 'GULF OF MEXICO','LAKE ST CLAIR', 'E PACIFIC', 'ATLANTIC SOUTH', 'HAWAII WATERS','LAKE MICHIGAN',
         # 'ATLANTIC NORTH', 'LAKE ONTARIO', 'ST LAWRENCE R','LAKE SUPERIOR''GULF OF ALASKA', 'VIRGIN ISLANDS', nan
```

Not only there are duplicate entries in the data set, but also entries that don't happen in the U.S. Since our research target is extreme weather events in the U.S. We decided to drop them. Entries counts drop from 1304,860 to 1259,884 after removing events not happened in the U.S. Furthermore, We also noticed that the format of damage_property is presented as 2K, we would like to translate it to 2000 instead. Besides, we also need to consider the effect of inflation during these years. As a result, we adjust the damage_property by an inflation rate of 1.75 according to the U.S Bureau of Labor Statics.

In order to better express the finding through figures, we categorized event_type into several groups. For example, flash floods, floods, and heavy rain are categorized as rain. We also tagged them on possibly linked to climate or not and drop entries that have no connection to climate. Furthermore, fill the empty entries with 0 in the columns of injuries, deaths, and property loss

event_type	ev_type_grp	possibly linked to climate
Dense Smoke	Fire	x
Wildfire	Fire	x
Dense Fog	Fog	x
Hail	Hail	x
Heat	Heat	x
Drought	Heat	x
Excessive Heat	Heat	x
Avalanche	Heat	x
Marine Hurricane/Typhoon	Hurricane	x
Coastal Flood	Hurricane	x
Storm Surge/Tide	Hurricane	x
Hurricane (Typhoon)	Hurricane	x
High Surf	Hurricane	x
Hurricane	Hurricane	x
Tropical Storm	Hurricane	x
Tropical Depression	Hurricane	x
Lightning	Lightning	x
Heavy Rain	Rain/Floods	x
Debris Flow	Rain/Floods	x
Flood	Rain/Floods	x
Lakeshore Flood	Rain/Floods	x
Flash Flood	Rain/Floods	x
Dust Devil	Strong Wind	x
Marine Thunderstorm Wind	Strong Wind	x
Marine High Wind	Strong Wind	x
Marine Strong Wind	Strong Wind	x
Thunderstorm Wind	Strong Wind	x
High Wind	Strong Wind	x
Strong Wind	Strong Wind	x
Dust Storm	Strong Wind	x
Waterspout	Tornado	x
Tornado	Tornado	x
Tsunami	Tsunami	x
Extreme Cold/Wind Chill	Winter	x
Freezing Fog	Winter	x
Winter Storm	Winter	x
Ice Storm	Winter	x
Winter Weather	Winter	x
Frost/Freeze	Winter	x
Lake-Effect Snow	Winter	x
Sleet	Winter	x
Cold/Wind Chill	Winter	x
Heavy Snow	Winter	x
Blizzard	Winter	x
Seiche	NaN	NaN
Sneakerwave	NaN	NaN
Rip Current	NaN	NaN
Funnel Cloud	NaN	x
Marine Dense Fog	NaN	NaN
Marine Hail	NaN	NaN

After the initial EDA process, we have the data set with the desired format below and are ready to execute the analysis.

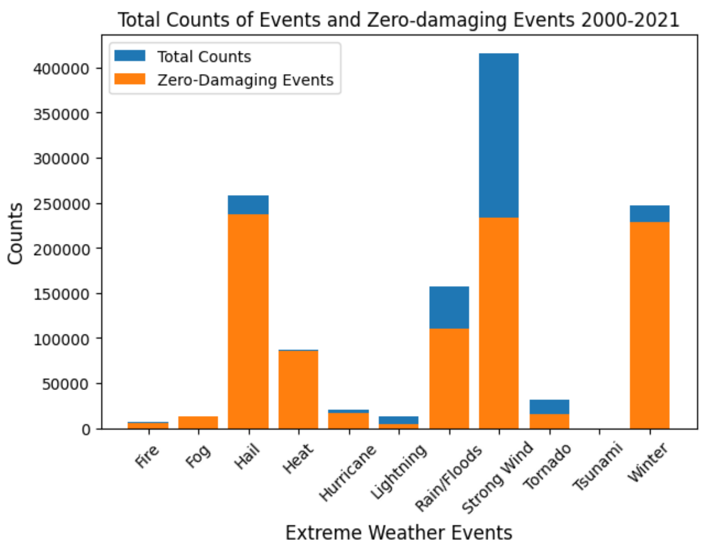
date	year	month	day	state	event_type	ev_type_grp	injuries_direct	injuries_indirect	deaths_direct	deaths_indirect	damage_property
2000-12-31	2000	12	31	FLORIDA	Extreme Cold/Wind Chill	Winter	0	0	0	0	0.0
2000-12-31	2000	12	31	FLORIDA	Extreme Cold/Wind Chill	Winter	0	0	0	0	0.0
2000-12-31	2000	12	31	FLORIDA	Extreme Cold/Wind Chill	Winter	0	0	0	0	0.0
2000-12-13	2000	12	13	WEST VIRGINIA	Winter Storm	Winter	0	0	0	0	0.0
2000-08-03	2000	8	3	MISSISSIPPI	Thunderstorm Wind	Strong Wind	0	0	0	0	3500.0
2000-08-09	2000	8	9	MISSISSIPPI	Thunderstorm Wind	Strong Wind	0	0	0	0	3500.0
2000-08-09	2000	8	9	MISSISSIPPI	Thunderstorm Wind	Strong Wind	0	0	0	0	1750.0
2000-01-21	2000	1	21	MAINE	Blizzard	Winter	0	0	0	0	0.0
2000-01-21	2000	1	21	MAINE	Blizzard	Winter	0	0	0	0	0.0

```
date                 datetime64[ns]
year                          int64
month                         int64
day                           int64
state                        object
event_type                   object
ev_type_grp                  object
injuries_direct               int64
injuries_indirect             int64
deaths_direct                 int64
deaths_indirect               int64
damage_property             float64
dtype: object
```

Initial Data Exploration

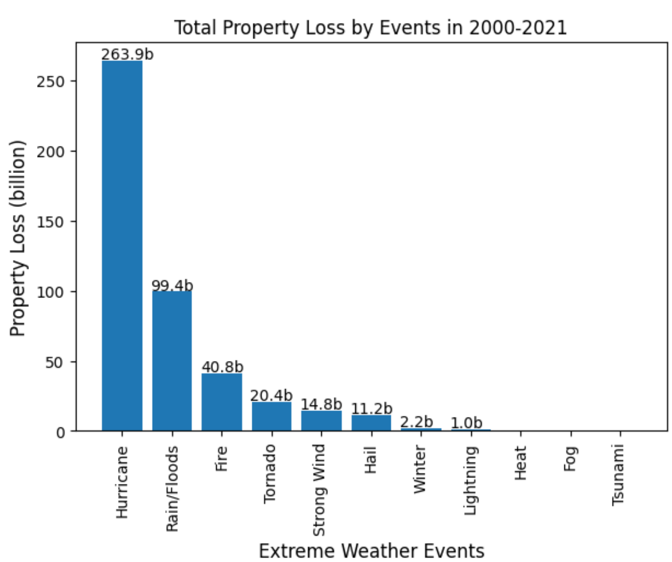
Insights of property loss caused by extreme weather events since 2000.

Not all extreme weather events would cause damage. They can land in nowhere, or too mild to cause any damage, or people are aware of it and have prevention against it. As a result, in the first move of data exploration, we decided to separate total events and events that cause no damage. In the following figure, we notice that most of the events don't actually deal any damage except rain, strong wind, and tornadoes. About twenty to thirty percent of them do cause damage.



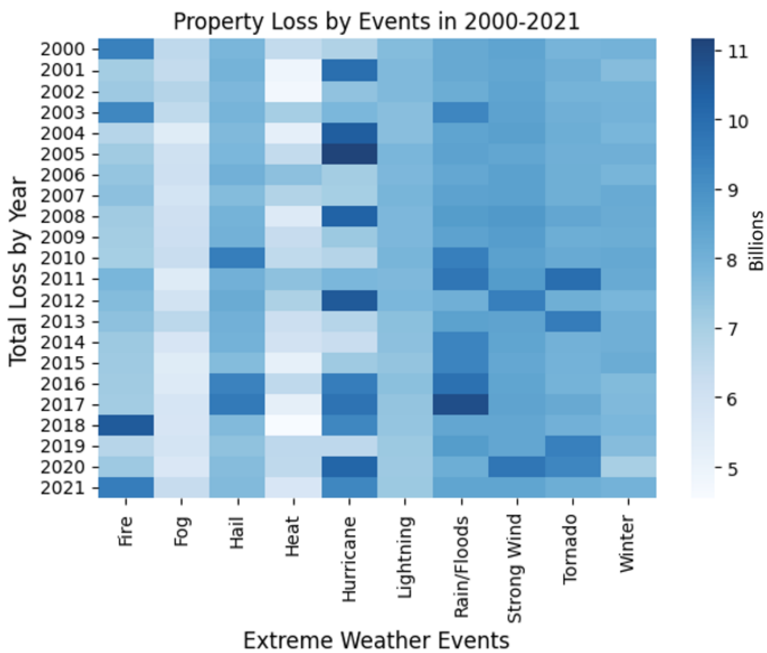
ev_type_grp	damage_property
Fire	7441
Fog	13337
Hail	258563
Heat	86495
Hurricane	20704
Lightning	13586
Rain/Floods	157127
Strong Wind	415357
Tornado	31025
Tsunami	22
Winter	246684

We know that among all extreme weather events, rain, strong winds, and tornados have the highest probability to deal damage. However, hurricanes are actually the most destructive. Referring to the graph below, we can see hurricanes dealt with more than $250 billion dollars since 2000, and category 5 hurricane Katrina itself, caused over 1,800 fatalities and $125 billion in property loss.



ev_type_grp	damage_property
Hurricane	2.639191e+11
Rain/Floods	9.939835e+10
Fire	4.078525e+10
Tornado	2.039375e+10
Strong Wind	1.483721e+10
Hail	1.122764e+10
Winter	2.159000e+09
Lightning	9.706247e+08
Heat	1.101314e+08
Fog	3.161812e+07
Tsunami	2.493750e+06

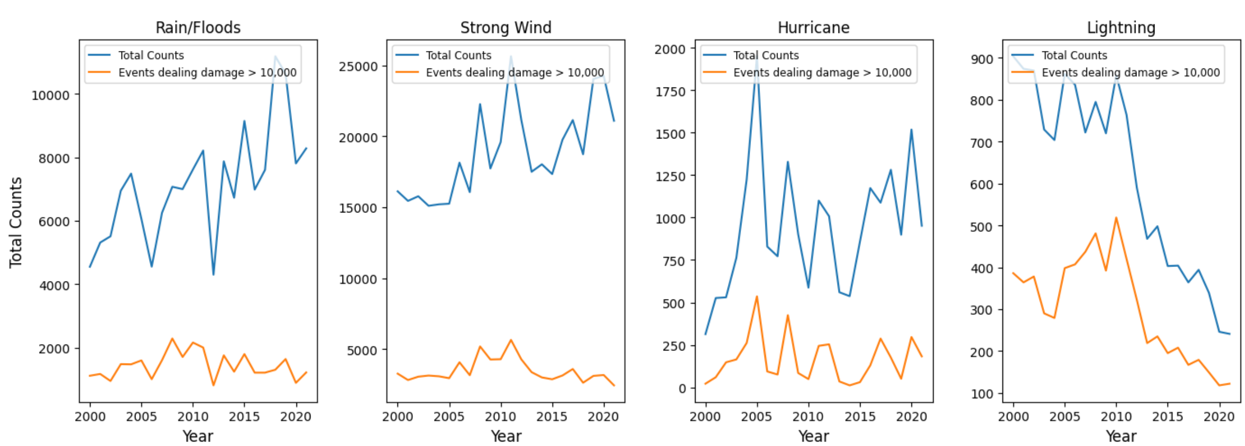
As you can see in the heatmap and data set down below, not all events cause evenly damage every year. Severe fire events happened in 2018, according to history, The 2018 wildfire season was the deadliest and most destructive wildfire season in California history including $25.4 billion in property damage. As for hurricanes, the Notorious hurricane Katrina hit Florida in 2005 leaving traumatized memory for Americans.



ev_type_grp	Fire	Hurricane	Rain/Floods
year			
2000	9.423758	6.817127	8.269496
2001	7.085797	9.945287	8.278997
2002	7.194147	7.422711	8.155282
2003	9.269227	7.800981	9.310833
2004	6.704612	10.448377	8.429301
2005	7.131302	11.152031	8.477260
2006	7.351955	7.097295	8.337815
2007	7.497168	7.022350	8.476530
2008	7.129591	10.289902	8.664565
2009	7.095646	7.238936	8.480261
2010	7.035477	6.732842	9.482697
2011	7.851637	7.804058	9.755421
2012	7.642592	10.545597	8.053641
2013	7.473275	6.706408	8.500566
2014	7.218293	6.231686	9.331642
2015	7.174598	7.170588	9.339584
2016	7.148996	9.549834	9.874214
2017	7.124594	9.828600	10.862948
2018	10.510469	9.278613	8.342423
2019	6.672467	6.524481	8.659139
2020	7.191157	10.207095	8.127517
2021	9.546303	9.293331	8.408932

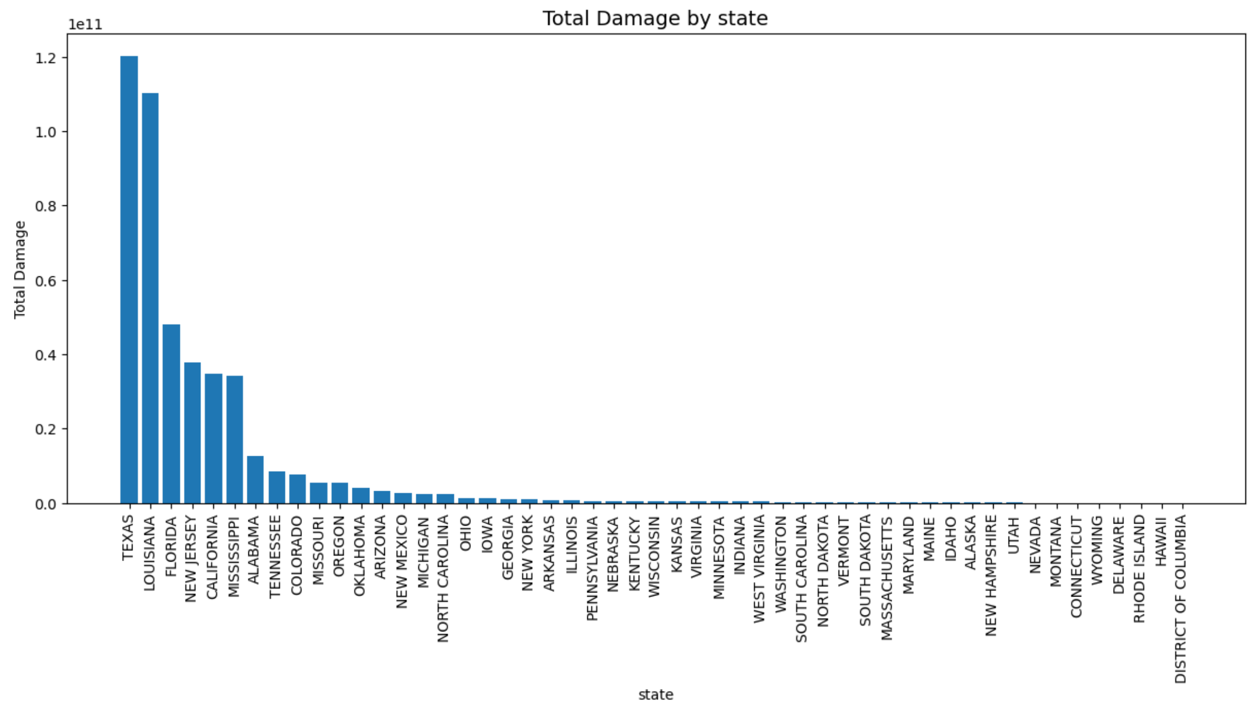
As for the frequency of extreme weather events, Owing to rising temperatures, we can see there are upward trends in rains, strong winds, and hurricanes. However, the number of severe events causing more than 10,000 dollars somehow remains steady since 2000. In sum, the frequency of extreme weather events is increasing but not for high-cost disasters.

On the other hand, we are surprised that lightning is declining throughout these years. As we dig into the root, we found out that even though the frequency of lightning across the country is decreasing on average, but not for lightning events in California. According to the article: **Lightning activity in the U.S. was below average in 2020, with notable exceptions.** The increasing frequency of lightning-ignited wildfires is threatening residents' safety in CA.

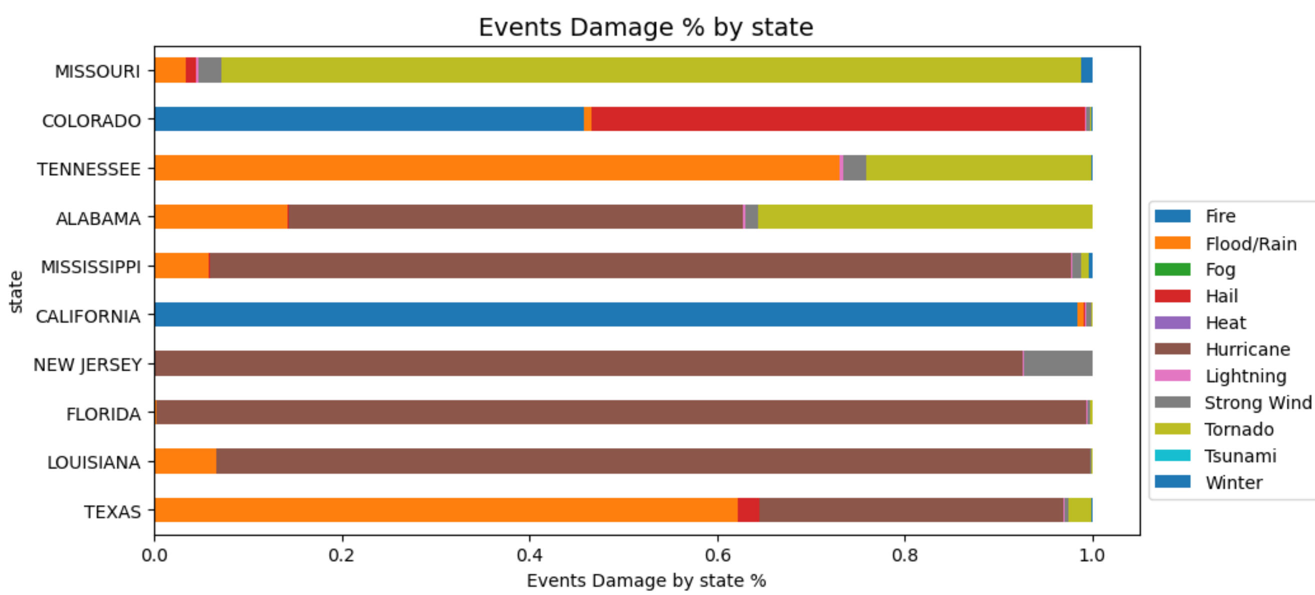


Insights of extreme weather events by states

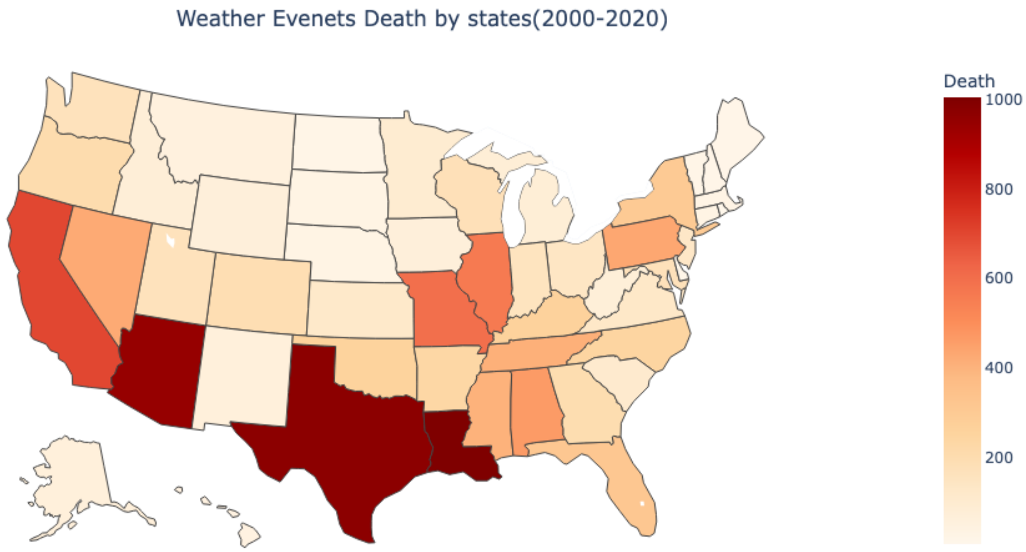
We wanted to look into the fatality and damage of weather events in all states to see which states were impacted the most by climate change related weather events during the period from 2000-2020. We created a bar chart looking into total damage and death for each state. We find that **Texas**, **Louisiana** & **Florida** were impacted with the most damage in the US caused by the weather events with **Texas**: *$120,263.7 M* - **Louisiana** *$110,249.9 M* - **Florida**: *$48,033.85 M*



When we look deeper into weather events (2000-2020) that cause the biggest loss in the highest 10 states damage. We can see that **Flood/Rain** caused more the 50% of Texas & Tennessee damage, **Hurricane** cause almost all the damage of Louisiana, Florida, New Jersey, Mississippi and more than 50% of Alabama, **Fire** are the highest cause of damage in California and **Tornado** in Missouri by looking at the following chart.

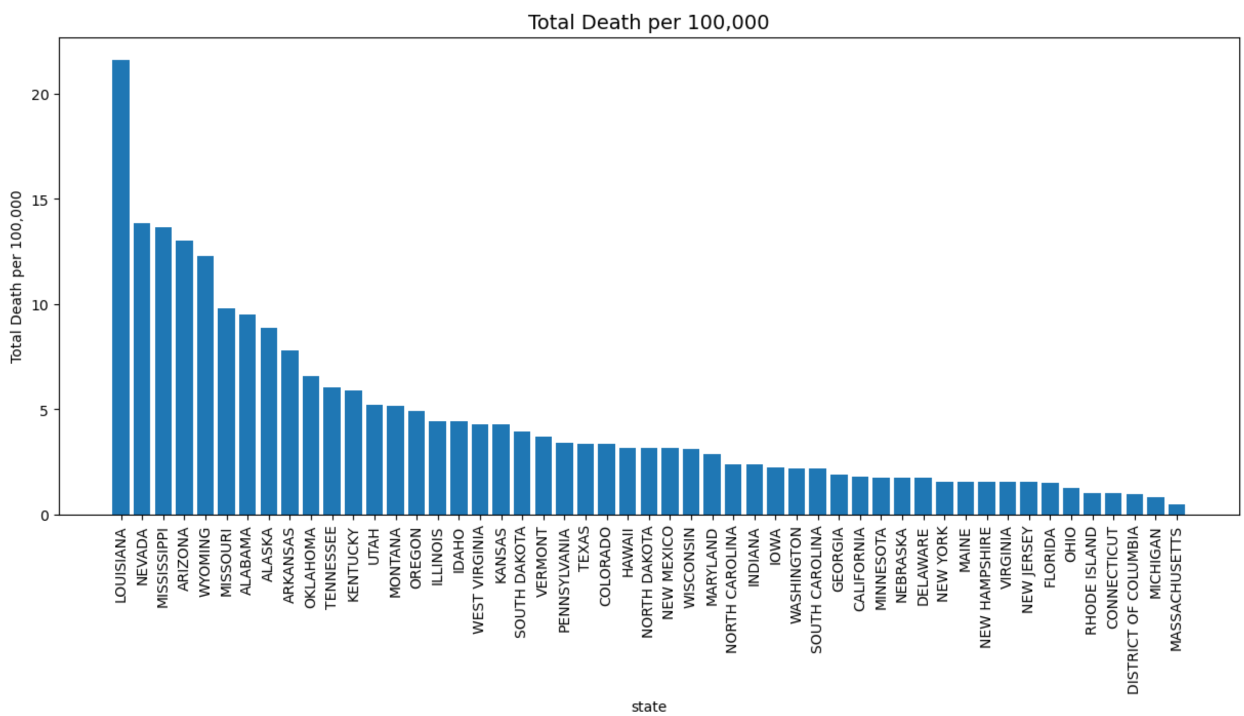


The weather events cause death cases every year in the US and we can see from the map that **Louisiana**, **Texas**, and **Arizona** are the most impacted states with the total number of death from weather events with **Louisiana:** *1005* **- Texas**: *975* **- Arizona**: *949* **- California**: *701*.



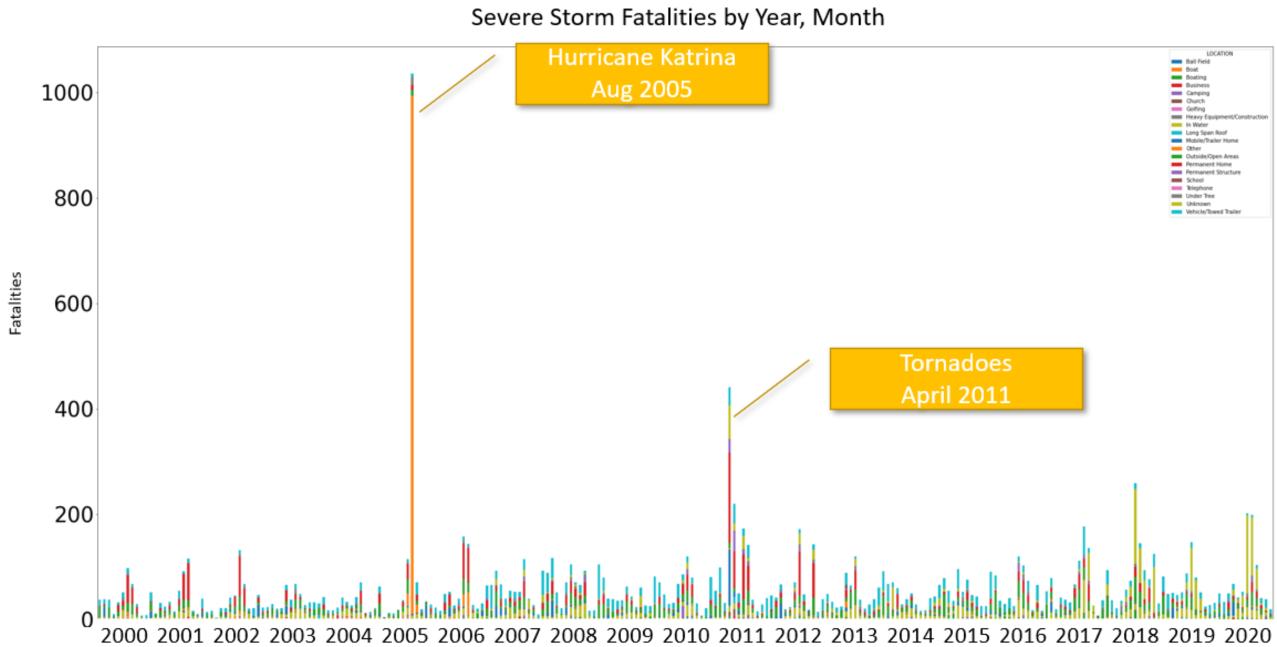
EV_TYPE_GRP state	Fire	Flood/Rain	Fog	Hail	Heat	Hurricane	Lightning	Strong Wind	Tornado	Tsunami	Winter	total_death
LOUISIANA	0.0	36.0	1.0	0.0	39.0	840.0	21.0	28.0	30.0	0.0	10.0	1005
TEXAS	30.0	345.0	0.0	1.0	241.0	49.0	51.0	56.0	58.0	0.0	144.0	975
ARIZONA	19.0	80.0	0.0	1.0	787.0	0.0	20.0	13.0	0.0	0.0	29.0	949
CALIFORNIA	213.0	111.0	10.0	1.0	187.0	69.0	14.0	60.0	0.0	1.0	35.0	701
MISSOURI	0.0	105.0	0.0	0.0	180.0	0.0	23.0	34.0	247.0	0.0	11.0	600

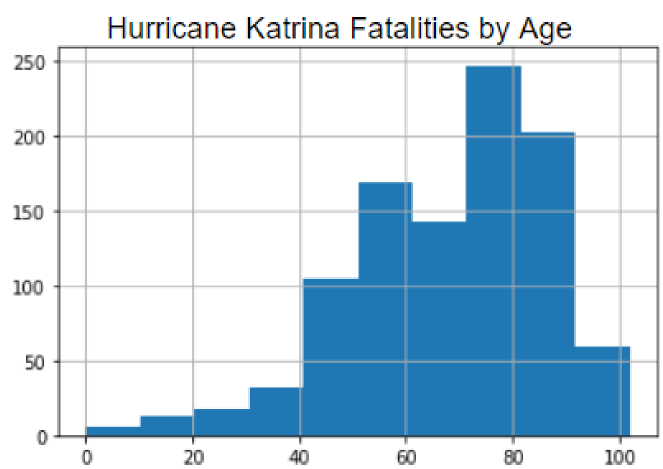
We realized that the population number could play an effect on the order of states impacted by the weather events so we decided to look into the total number of **Death per *100,000*** in the states to eliminate the population factor. We notice that after dismissing the population factor from the analysis that **Louisiana**, **Nevada**, and **Mississippi** are the most impacted states by the weather events death numbers during the period from **2000-2020**.

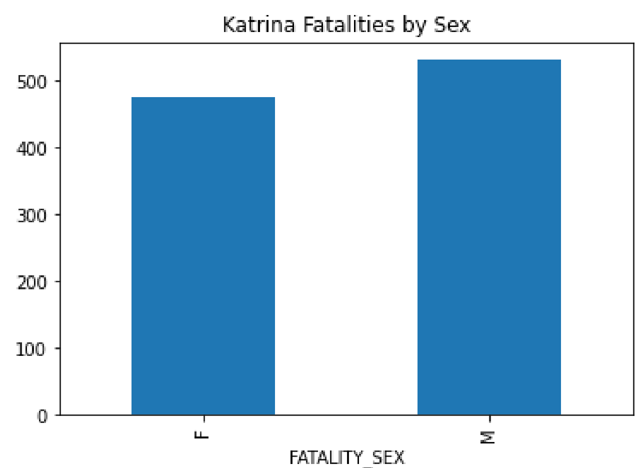


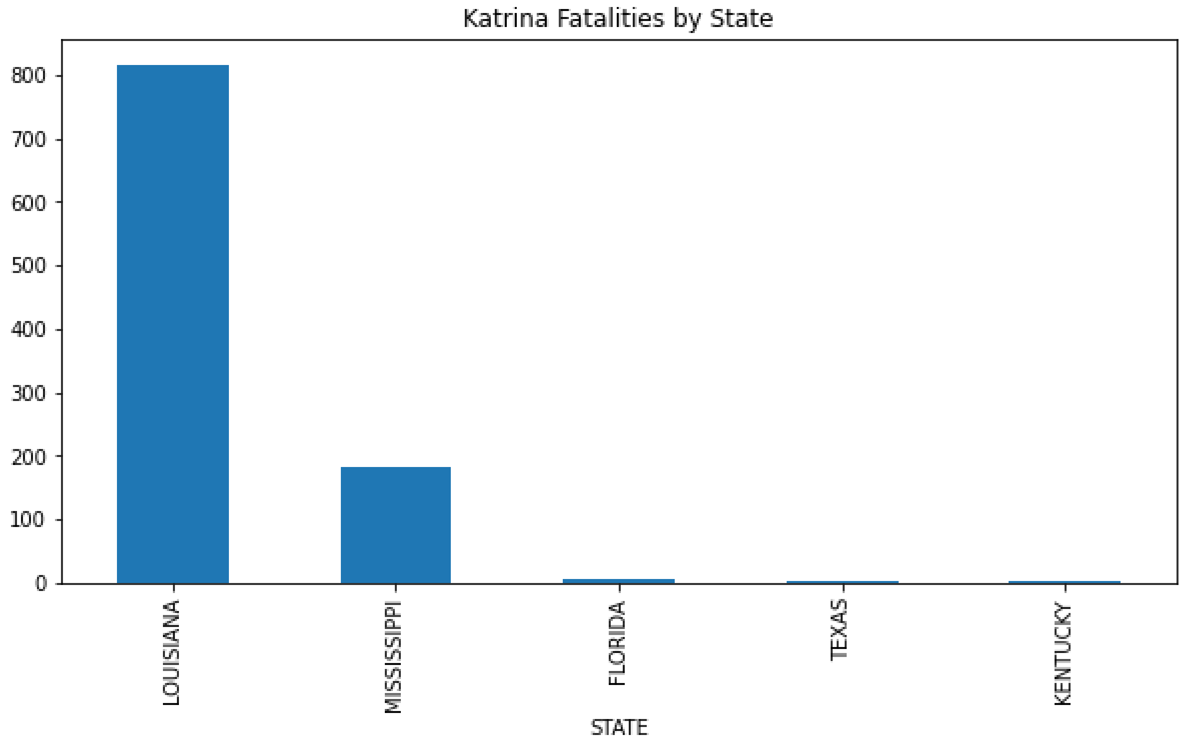
Details of Fatalities and Specific Events (Katrina, 2011 Tornadoes)

Doing a deeper dive into specific events showed the scale of some of the most deadly events in U.S. history. The data grouped by month and year shows the extreme loss of life due to Hurricane Katrina and April 2011 Tornado outbreak that swept the South Eastern US.

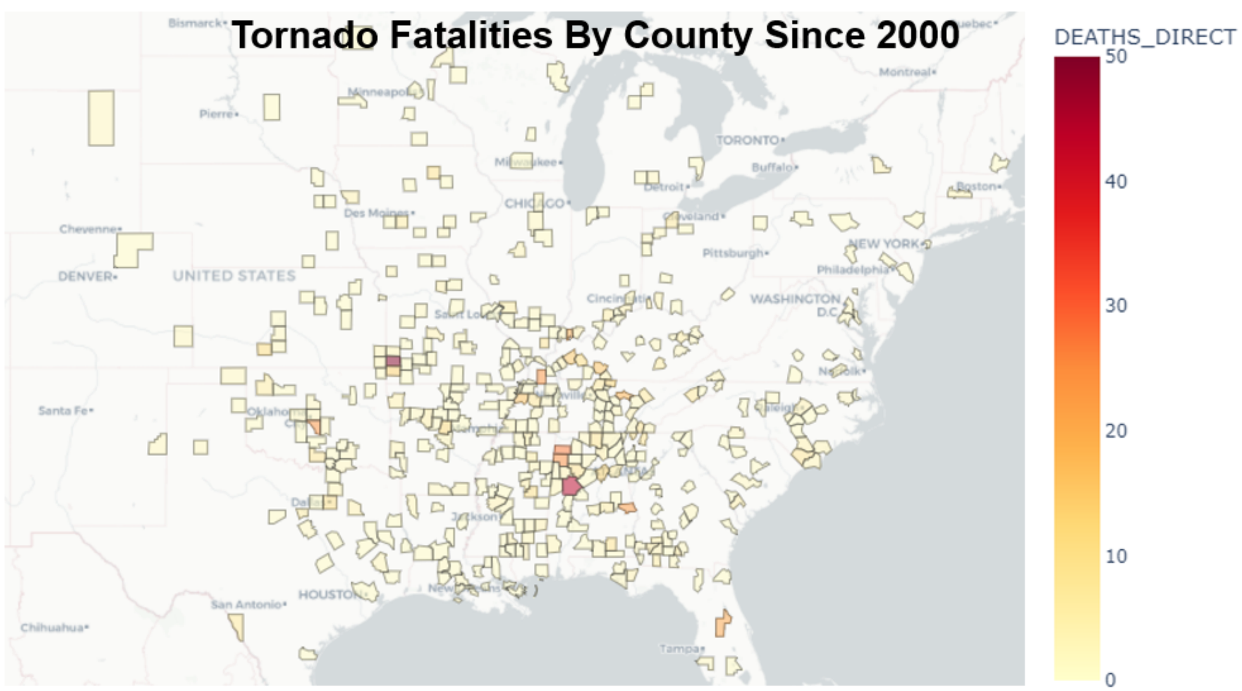








To obtain Event Type/Specific event along with data about the fatalities such as, Age, Gender, and Location (House, Trailer, Vehicle), we have to join on EVENT_ID in both data sets.



Conclusion

- In the end, we avoided drawing a statistical conclusion on the relationship between these severe events and Climate change, per instructions but we think this data could be useful to determine that relationship.
- The impact to communities and cost of Severe Weather in the US is clear
- Hurricanes cause most damage, Louisiana is most affected, Katrina being the most devastating event
- Overall, thoughts on NOAA Data Set
    - Needed some cleaning, i.e. Property Loss need to translate to proper format.
    - EventID is not unique and can span rows, i.e A single event can be reported differently.
    - Other interesting data not explored
        - Tornado Data since 1950 with Start/End Lat/Long
        - Tornado F Scale
        - Crop Damage
        - Episode/Event narratives detailing each event