Storm Analysis

By
Kathryn Keck
Alister Gan
Ji Xian Fu
Simeon Bundt
Grant Springman

Outline

- Introduction
- Datasets
- Analysis and Visualization of data
- Conclusion

Introduction

With the current changing weather patterns and more extreme weather and storms throughout the country in recent times, we want to look at multiple datasets on different storm types, location of storms, and fatalities to better understand which part of the country is most affected and our hope is that our analysis can increase awareness to the community and respective authorities to take the necessary actions.

Datasets

We gathered our data from **National Centers for Environmental Information**

1. "StormEvents_details-ftp_v1.0_d2023_c20241121.csv.gz"
   - Details on each storm including the event type, information on property damage, injuries, and magnitude.
2. "StormEvents_fatalities-ftp_v1.0_d2023_c20241121.csv.gz"
   - Information on fatalities and the sex, location of death, and date the death occurred.
3. "StormEvents_locations-ftp_v1.0_d2023_c20241121.csv.gz".
   - Information on locations of the storms, longitude and latitude, and the range of the storm.

Types of Storm

Questions to answer:

- Which storm types result in the highest number of casualties?
- Which storm types are the deadliest overall?
- How often do different storm types occur, how does their frequency relate to severity, and which storm type has the highest average casualties per event?
- Which storm types dominate during peak times?
- How does storm intensity vary by geography and time?

Data Cleaning

```{r}
# Import datasets
fatalities <- read.csv("Storm_Fatalities.csv")
locations <- read.csv("Storm_locations.csv")
details <- read.csv("Storm_Details.csv")
```

Data Import: Loaded three datasets (`Storm_Fatalities`, `Storm_Locations`, `Storm_Details`) using `read.csv`.

```{r}
# Merge all datasets by EVENT_ID
storms_event <- details %>%
  inner_join(locations, by = "EVENT_ID") %>%
  inner_join(fatalities, by = "EVENT_ID")

# Remove duplicate column
storms_event <- storms_event %>% select(-FAT_YEARMONTH)

# Convert damage columns to numeric
storms_event <- storms_event %>%
  mutate(
    DAMAGE_PROPERTY = as.numeric(gsub("[KkMm]", "", DAMAGE_PROPERTY)) *
                      ifelse(grepl("K", DAMAGE_PROPERTY, ignore.case = TRUE), 1e3,
                             ifelse(grepl("M", DAMAGE_PROPERTY, ignore.case = TRUE), 1e6,
1)),
    DAMAGE_CROPS = as.numeric(gsub("[KkMm]", "", DAMAGE_CROPS)) *
                   ifelse(grepl("K", DAMAGE_CROPS, ignore.case = TRUE), 1e3,
                          ifelse(grepl("M", DAMAGE_CROPS, ignore.case = TRUE), 1e6, 1))
  )
```

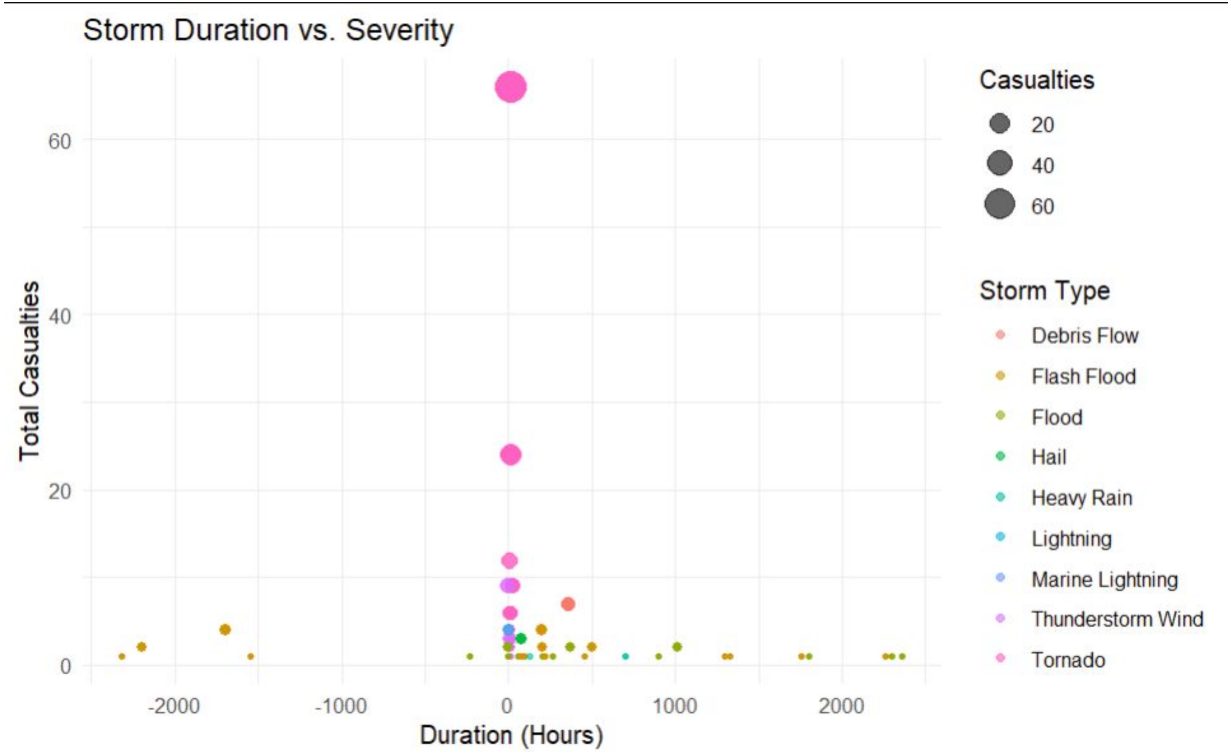
<merged on the common column `EVENT_ID`

Converted `DAMAGE_PROPERTY` and `DAMAGE_CROPS` columns to numeric by removing unit

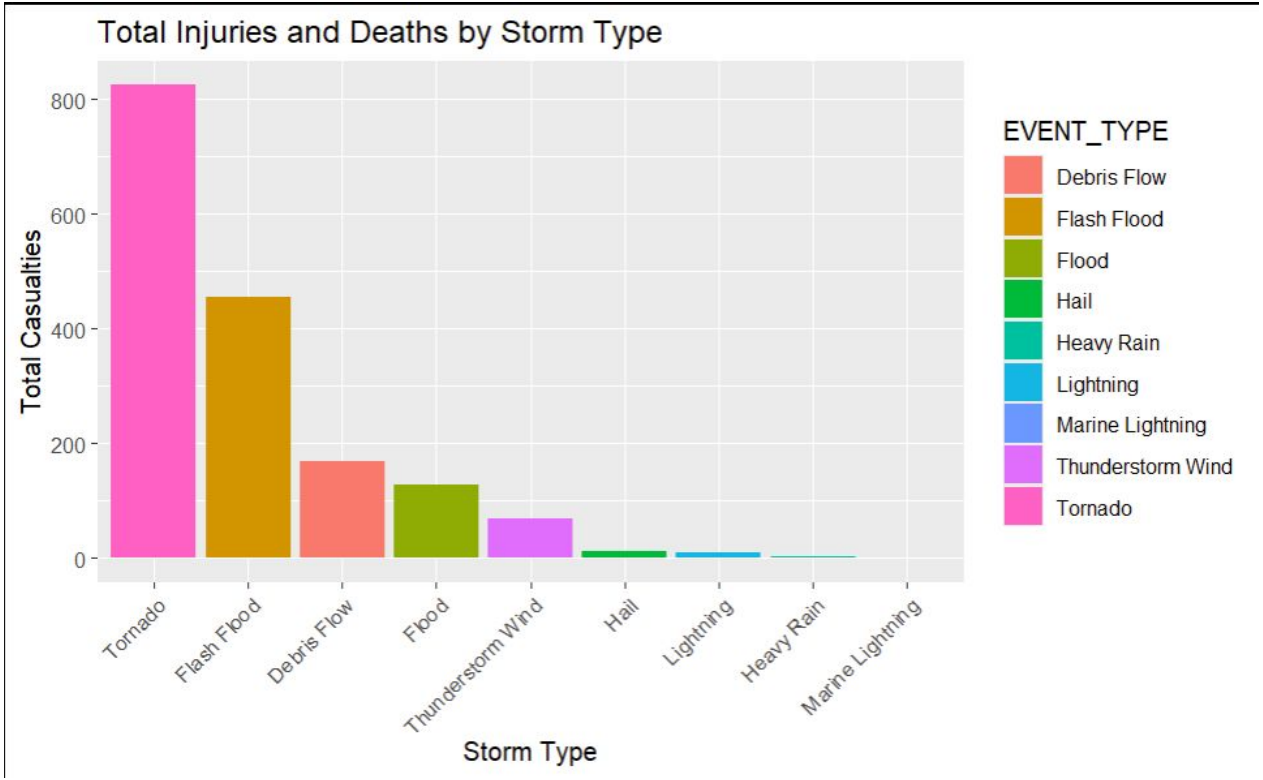
Which storm types result in the highest number of casualties?

- `Duration` is computed as the difference between `END_TIME` and `BEGIN_TIME`.
- Casualties are summed up from various injury and death columns.
- The size of bubbles represents `Total_Casualties`, and color differentiates `EVENT_TYPE`.

Key Findings: Longer storms don't always correlate with higher casualties, but certain storm types show distinct patterns.



Which storm types are the deadliest overall?



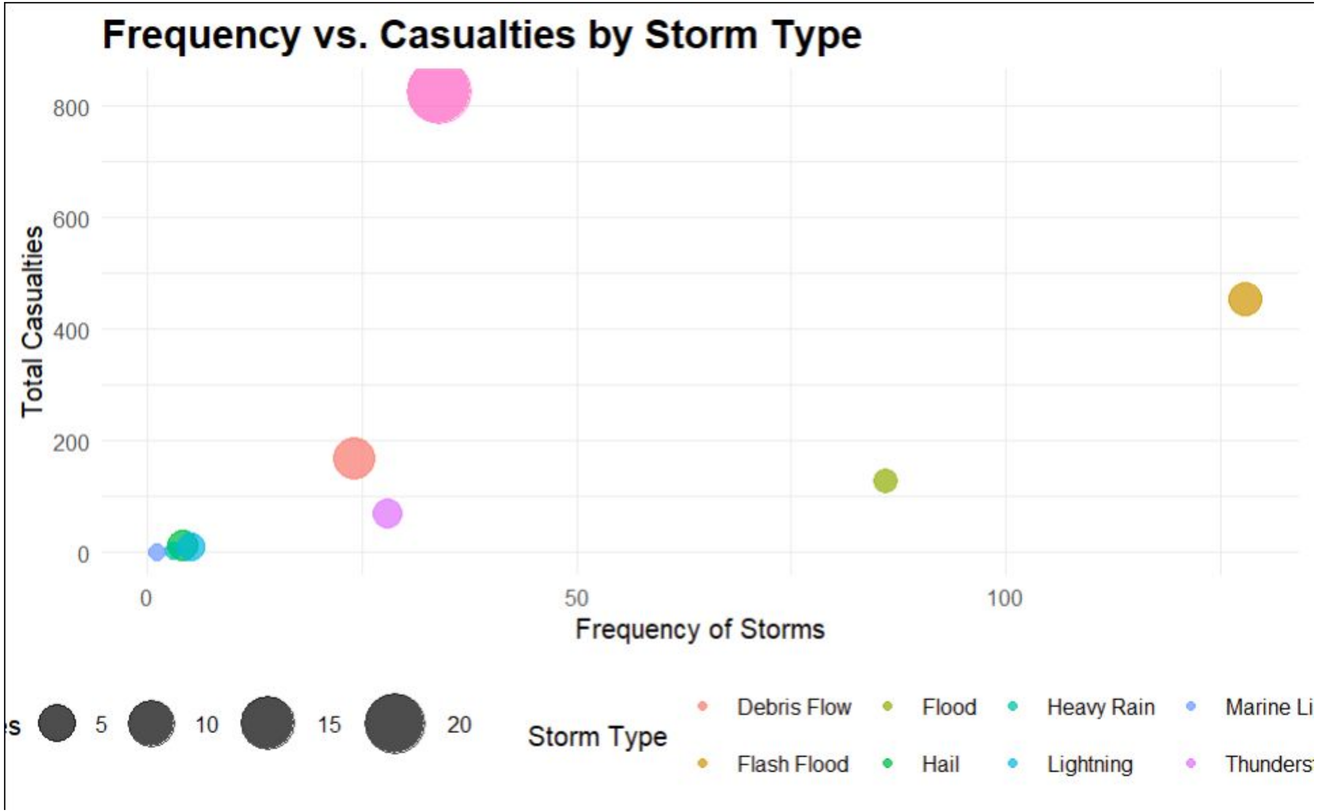
- Data is grouped by `EVENT_TYPE`.
- Total injuries and deaths are calculated and displayed as a stacked bar chart.

Key Findings: Certain storm types, such as tornadoes and hurricanes, cause disproportionately high casualties.

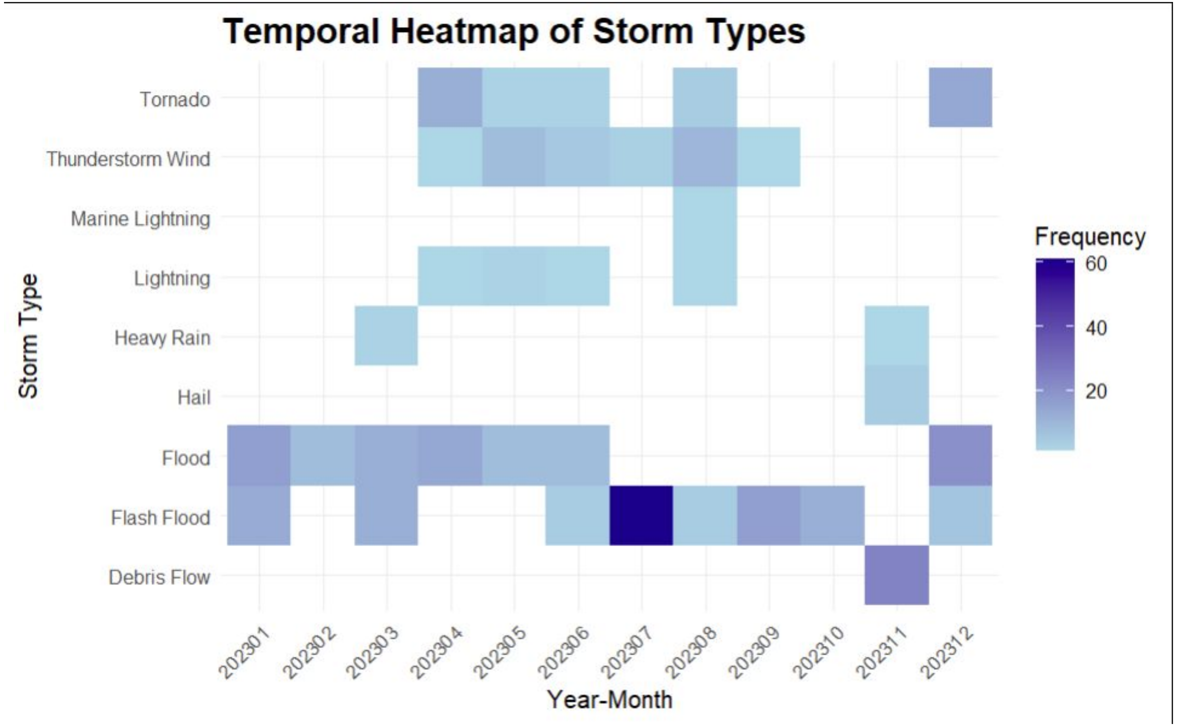
How often do different storm types occur, how does their frequency relate to severity, and which storm type has the highest average casualties per event?

- Frequency (`Count`) and casualties are calculated per storm type.
- A bubble chart visualizes the relationship between storm frequency and total casualties, with bubble size indicating `Average_Casualties`.

Key Findings:
High-frequency storms may have lower average casualties, while rare events like hurricanes cause severe impacts.



Which storm types dominate during peak times?



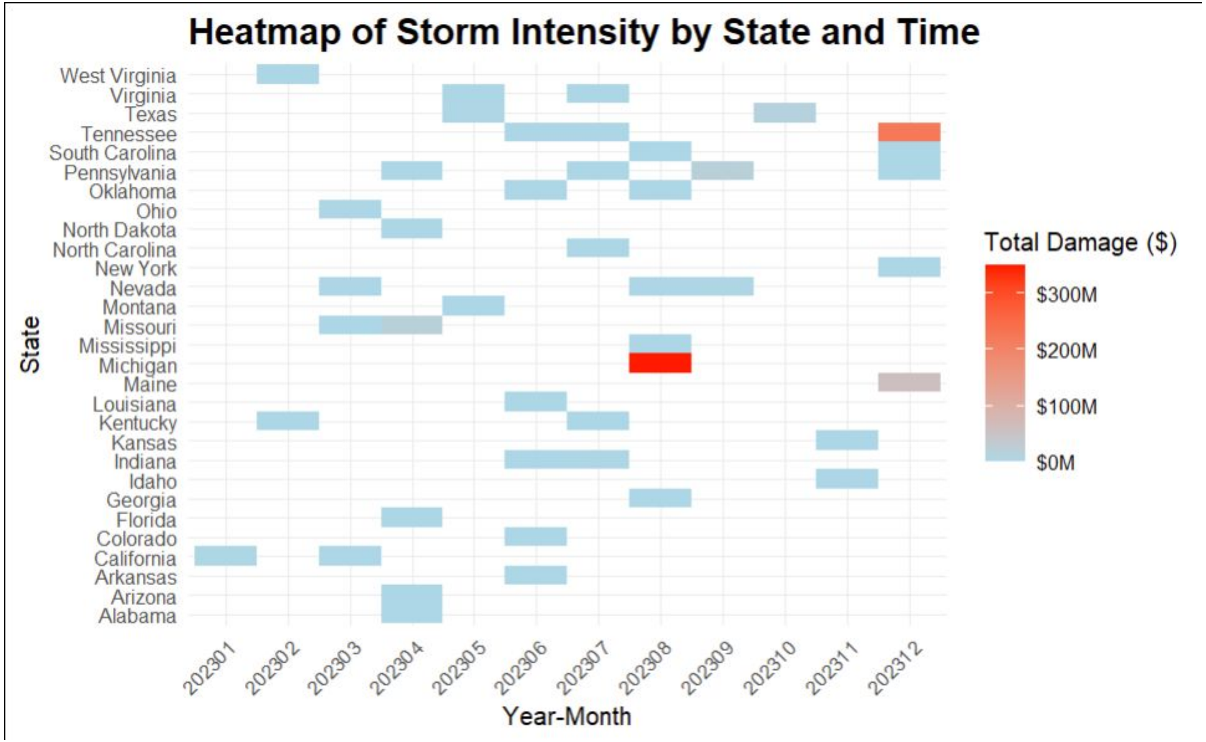
- Data is grouped by `BEGIN_YEARMONTH` and `EVENT_TYPE` to calculate frequency.
- A heatmap visualizes this temporal distribution, where darker colors represent higher frequencies.

Key Findings: Certain storm types exhibit seasonality or increasing/decreasing trends over time.

How does storm intensity vary by geography and time?

- Total damages (property + crops) are summed and grouped by state (`STATE_FIPS`) and time (`BEGIN_YEARMONTH`).
- State names are added using the `maps` package, and a heatmap is generated.

Key Findings: States in storm-prone regions (e.g., Gulf Coast, Midwest) show higher damages during specific months.



Storm Locations

Questions to answer:

- What are the locations of storms across the U.S?
- How does the range of a storm affect its magnitude?
- Which locations are most prone to storms?
- What storm types occur most frequently in each place/city?

Storm Location Dataset - Data Cleaning

```
locations <- read.csv("Storm_locations.csv")
colnames(locations)
```

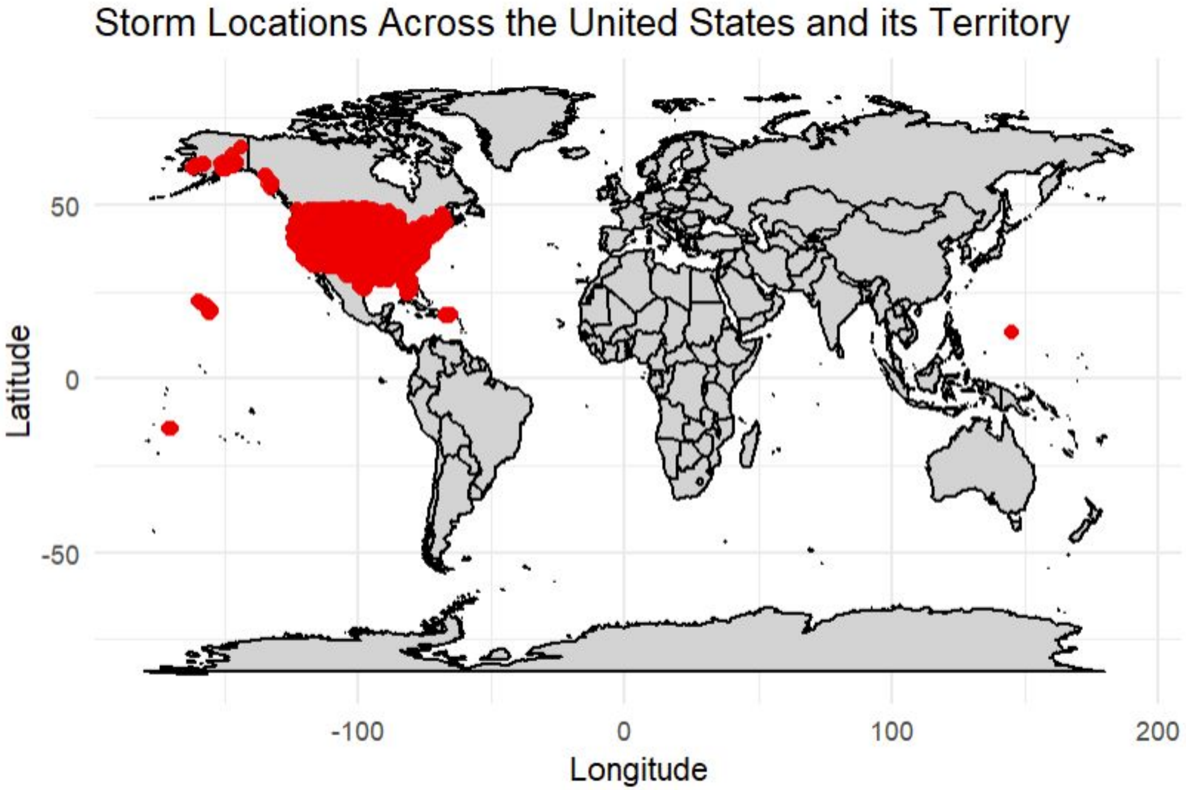
- Read the storm_locations dataset using read_csv function
- Check all the variables in the dataset

What are the locations of storms across the U.S?

```r
library(ggplot2)
library(sf)
library(maps)
library(tidyverse)
```

```r
# Ensure `locations` is defined and clean
locations <- locations %>%
  filter(!is.na(LATITUDE), !is.na(LONGITUDE))  # Remove missing values
```

What are the locations of storms across the U.S?



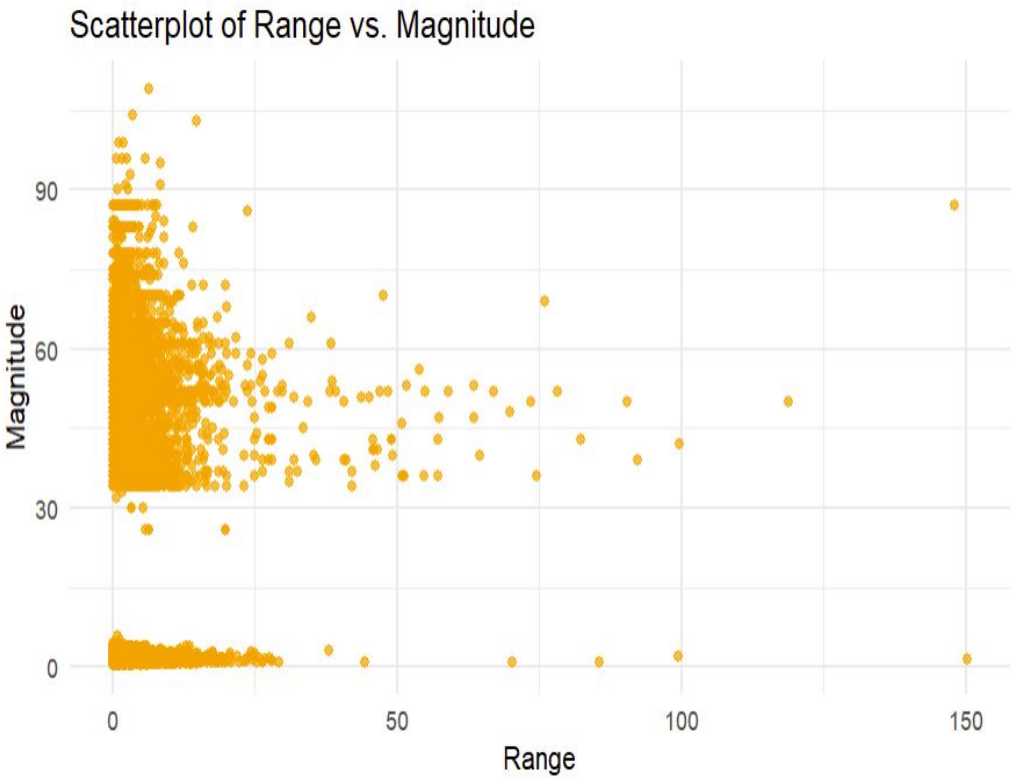
How does the range of a storm affect its magnitude?

```r
# Merge locations and details
storm_magnitude <- merge(locations, details, by = "EVENT_ID", all.y = TRUE)
```

```r
# Drop any rows where the magnitude is null
storm_magnitude <- storm_magnitude %>%
  filter(!is.na(MAGNITUDE) & !is.na(RANGE))
```

- Merge the locations and details datasets by the “EVENT_ID”
- EVENT_ID column in common
- Remove any rows that do not have data for the Magnitude or Range

How does the range of a storm affect its magnitude?



Findings:

- Significant number of storms occur at a distance close to the geographical center of a city/village
- Several outliers where the storms have abnormally high magnitude or high range values
- While the overall relationship between the range and magnitude of storms isn't strongly defined, storms with higher magnitudes tend to occur at both short and median ranges

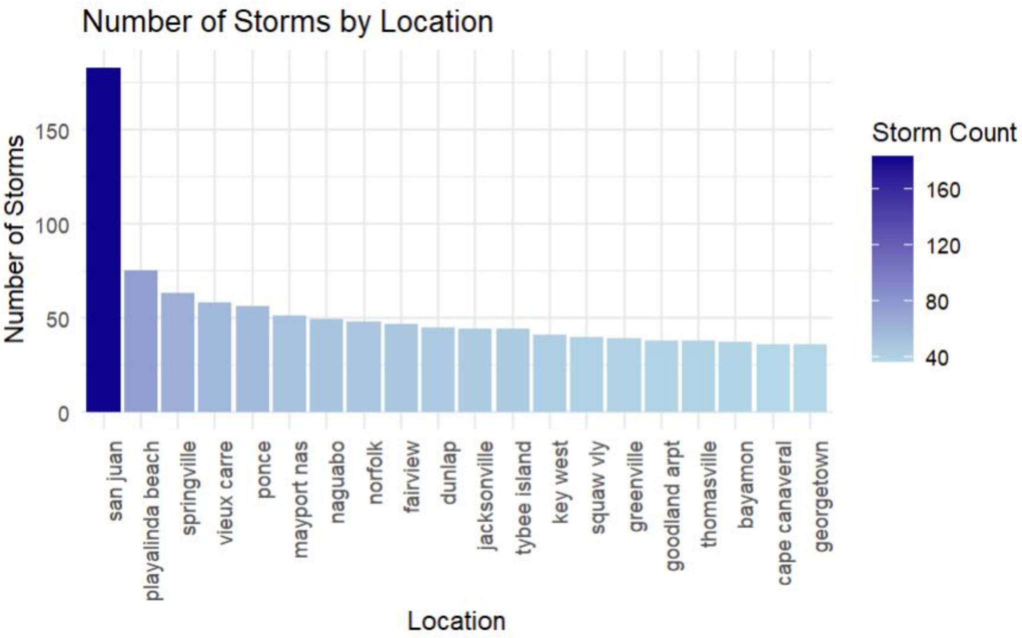
Which locations are most prone to Storms occurrence?

```
# Clean the 'location' column: remove spaces and convert to lowercase
locations <- locations %>%
  mutate(LOCATION = tolower(trimws(LOCATION)))  # trimws() removes
leading/trailing spaces

# Now you can group by Location
location_counts <- locations %>%
  group_by(LOCATION) %>%
  summarise(StormCount = n())

# Consider only the top 10 locations with highest number of storms
top_locations <- location_counts %>%
  arrange(desc(StormCount)) %>%
  head(20)
```

Which locations are most prone to Storms occurrence?



Findings:

- San Juan has the highest number of storms, exceeding 150, significantly outpacing other locations.
- The second highest is Playalinda Beach, though its storm count is much lower than San Juan's.
- The rest of the locations have relatively similar count of storms occurrence.

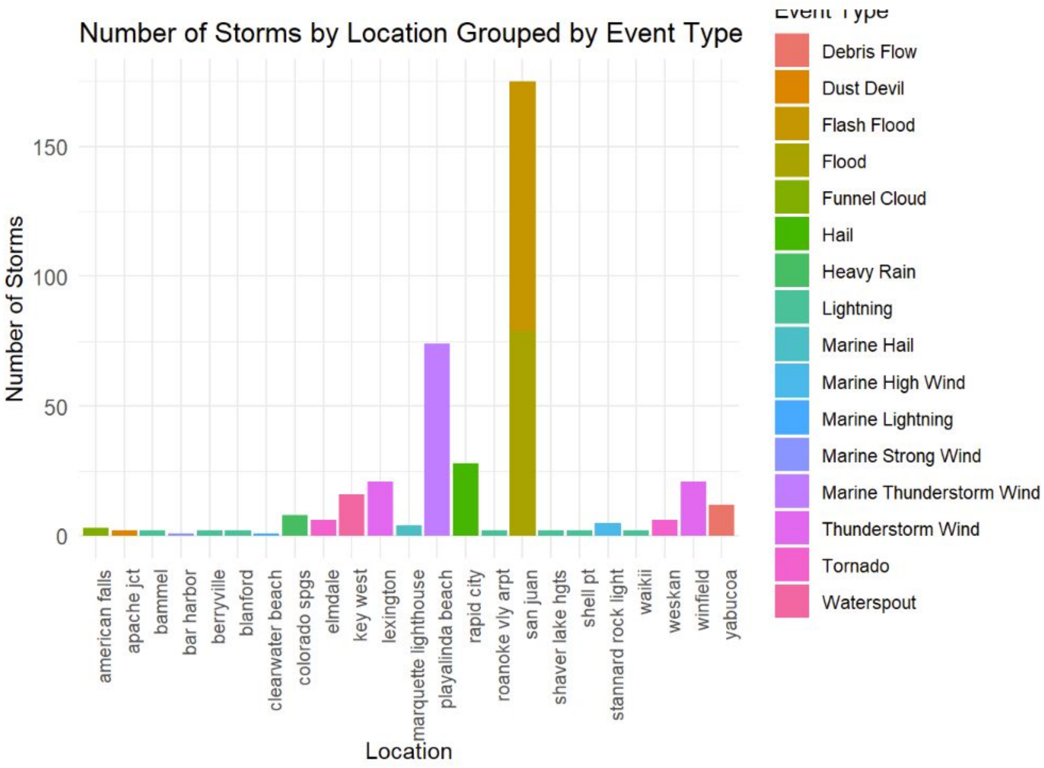
What storm types occur most frequently in each place/city?

```r
storms_events <- merge(locations, details, by = "EVENT_ID", all.y = TRUE)


# Drop rows with missing LOCATION or EVENT_TYPE
storms_events <- storms_events %>%
  filter(!is.na(LOCATION) & !is.na(EVENT_TYPE))
```

We will then group the data by their locations and storms type before calculating the number of occurrences for each storms for each locations. We will then apply the filter function to find the highest count for each storms type

What storm types occur most frequently in each place/city?



Types of Fatalities

Questions to answer:

- What storms caused the most direct and indirect fatalities?
- When throughout the year do fatalities for each storm occur?
- What age group and gender have the most fatalities?

Fatality Types Data Cleaning

- Need to merge fatalities and details datasets together
- EVENT_ID column in common

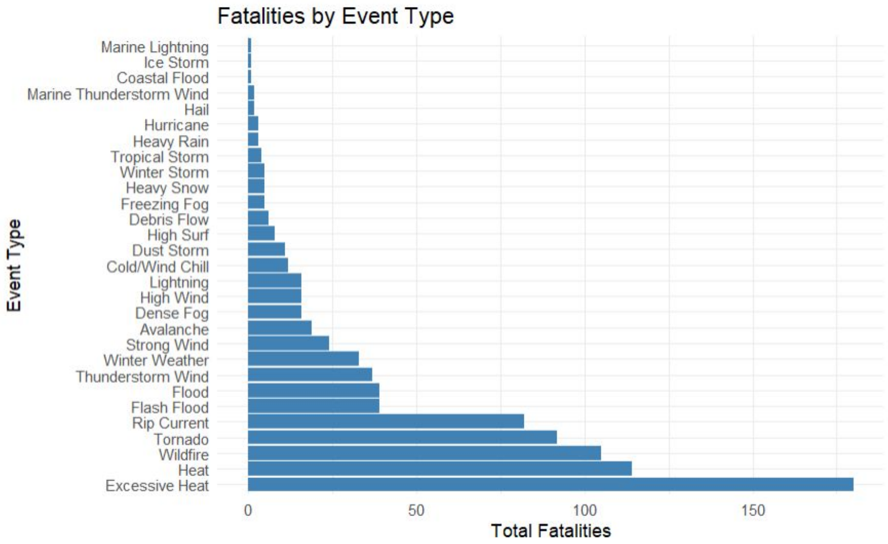
~~~r
```{r}
fatalities_details <- merge(fatalities, details, by = "EVENT_ID", all.x = TRUE)
head(fatalities_details)
```
~~~

Fatalities by Event Type

- Totaled the number of fatalities for each event type
- Most fatalities Excessive Heat, Heat, Wildfire, and Tornado
- Only one fatality: Marine Lightning*, Ice Storm, and Coastal Flood

Extreme heat is a period of unusually hot weather, while heat is simply a high temperature - National Center for Environmental Information

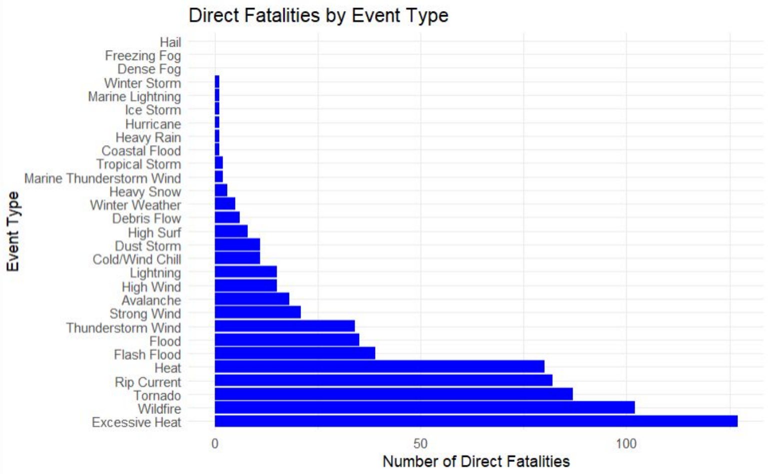
*lightning that strikes the ocean

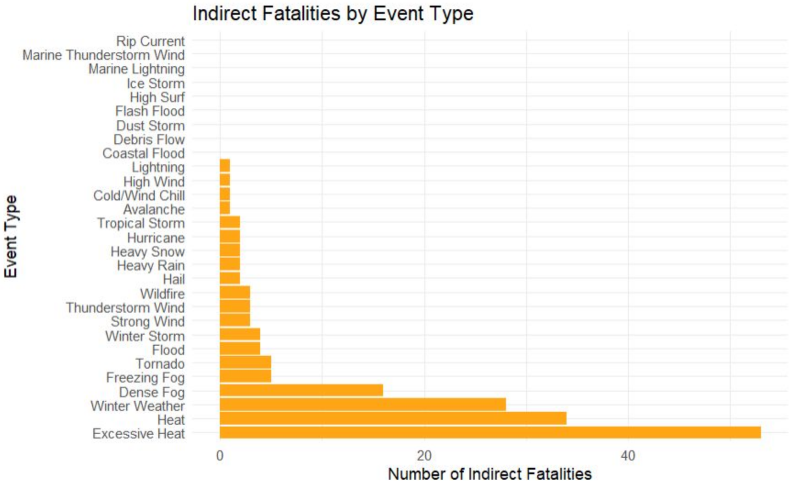


Direct and Indirect Fatalities

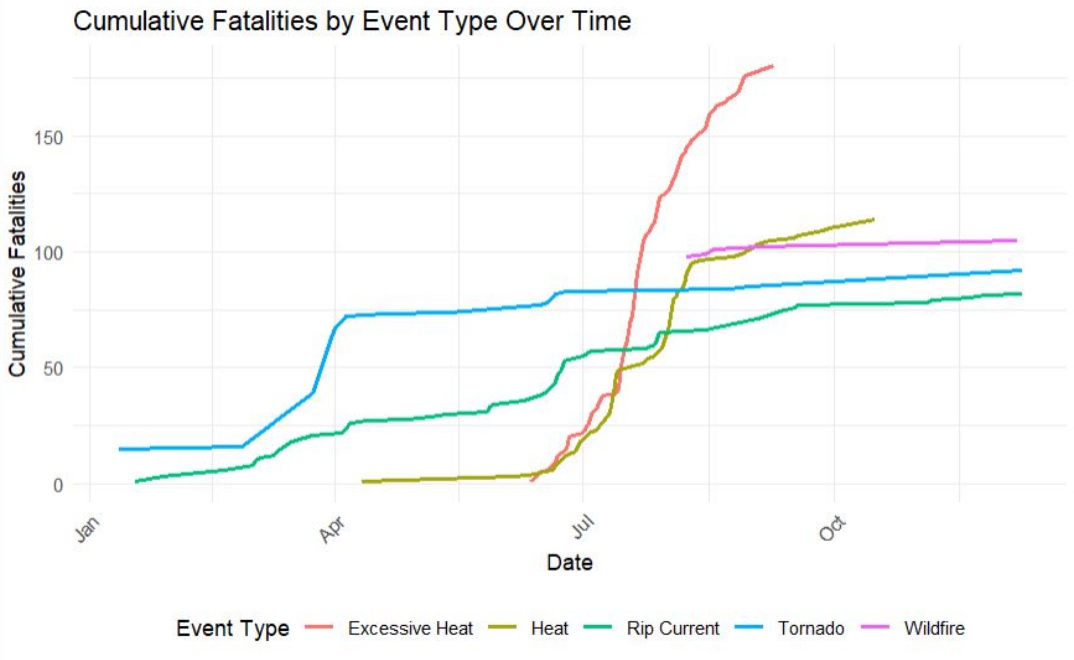
- More direct fatalities recorded
- Excessive heat, wildfire, tornado, and rip current* highest direct
- Excessive heat, heat, winter weather, and dense fog highest indirect
- Many event types have no indirect fatalities

*"a relatively strong, narrow current flowing outward from the beach through the surf zone and presenting a hazard to swimmers" - Oxford dictionary



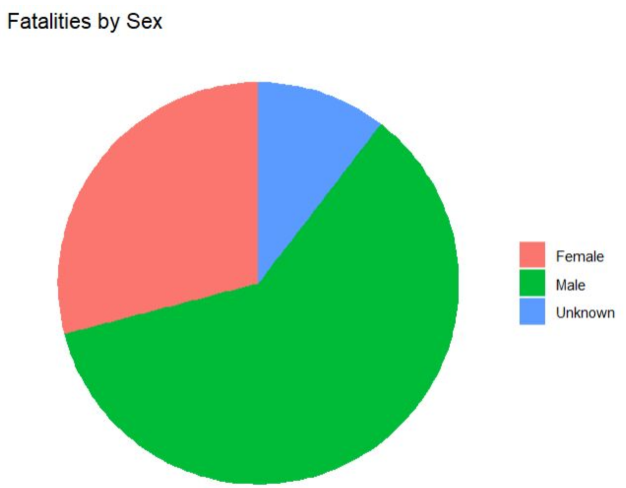


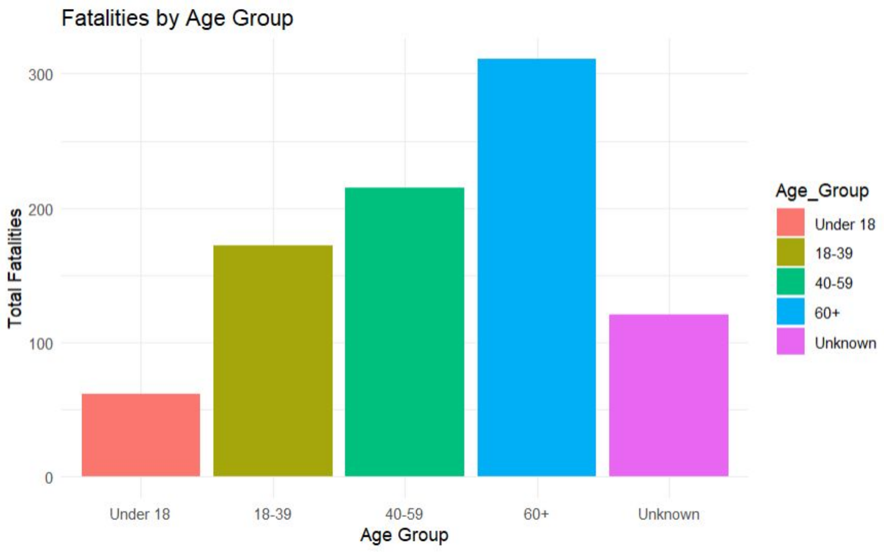
Fatalities Over Time



- Top 5 Events

Fatalities Demographics





Fatalities and Location

Questions to answer:

- What type of fatality was the greatest?
- Where in the US has the most fatalities due to storms?
- What age group have the most fatalities in certain areas?

Cleaning the Fatalities Dataset

- Renamed all the column names to UPPERCASE
- Removed extra spaces from the text data
- Kept only the rows were EVENT_ID and FATALITIY_TYPE are not empty
- Only included rows that had either D (direct) or I (indirect) fatality type
- Removed duplicate rows based on EVENT_ID and FATALITY_TYPE
- Fixed inconsistent data formats

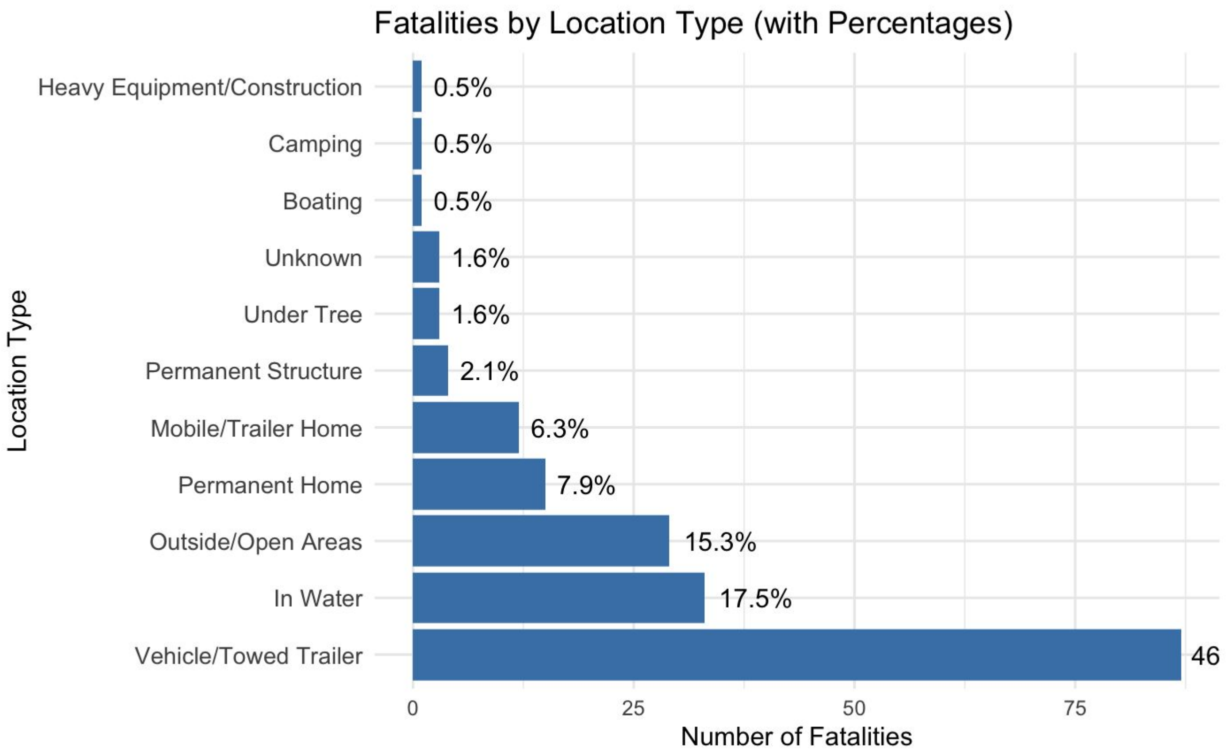
Cleaning the Locations Dataset

- Renamed all columns to uppercase
- Removed extra spaces from text data
- Kept rows where EVENT_ID, LATITUDE, and LONGITUDE are not empty
- Made sure latitude and longitude are valid:
  - Latitude between -90 and 90
  - Longitude between -180 and 180
- Removed duplicate rows based on EVENT_ID, LATITUDE, and LONGITUDE
- Fixed inconsistent date formats

Merge the Data

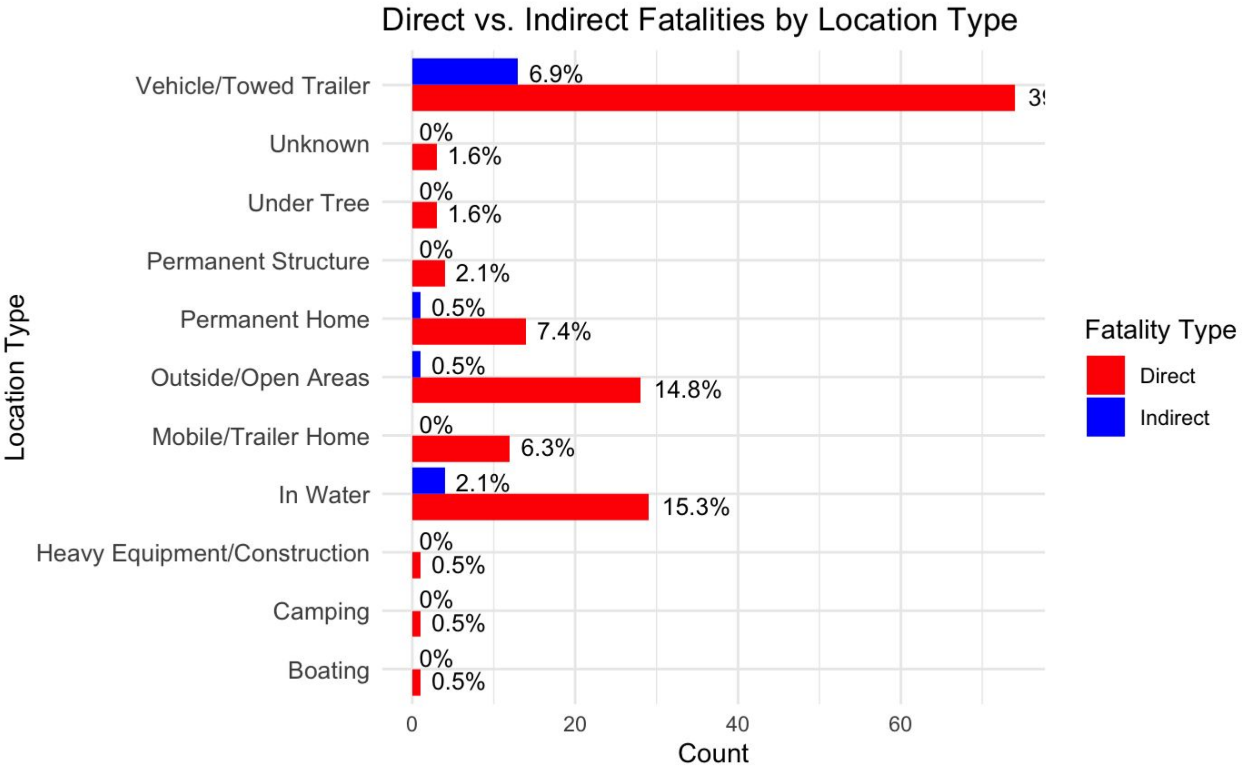
- Combined fatalities and locations so we can connect events to places
- Matched rows from both datasets using EVENT_ID
- Kept only the rows that appear in both datasets.

Fatalities by Location Type



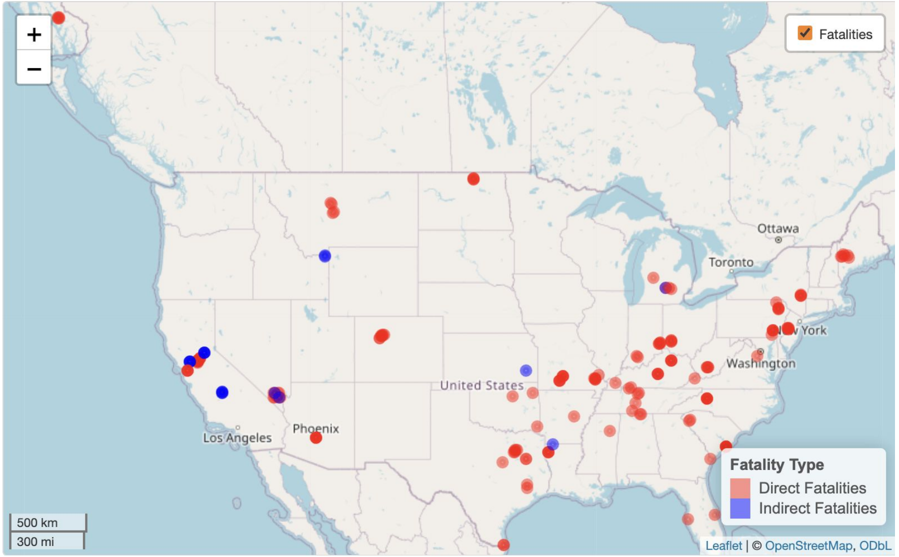
Direct vs. Indirect Fatalities by Location Type

- Direct deaths occur directly by the event itself
- Indirect deaths occur as a result of the event, but not from its immediate effects
- Direct deaths are a measure of the immediate danger of an event.
- Indirect deaths show the broader impact on human life.



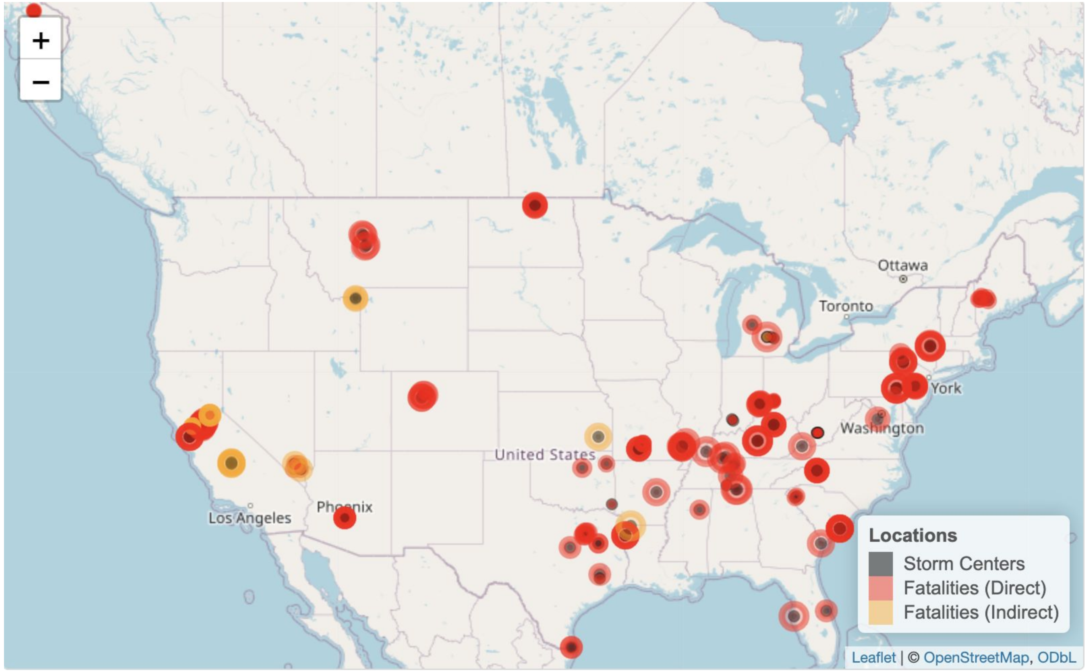
Spatial Distribution of Fatalities and Storm Centers

- The size of each marker is proportional to the total number of fatalities, making it easy to identify areas with high impact.
- Pop Ups provide detailed information for each marker, such as event type, total fatalities, and location.
- Allows us to analyze the map for clusters where there are more deaths and potential hotspots for storms, which can help to see where there may be more safety measures needed
- The red and the blue show us where there may be longer lasting effects from storms compared to just during the storm



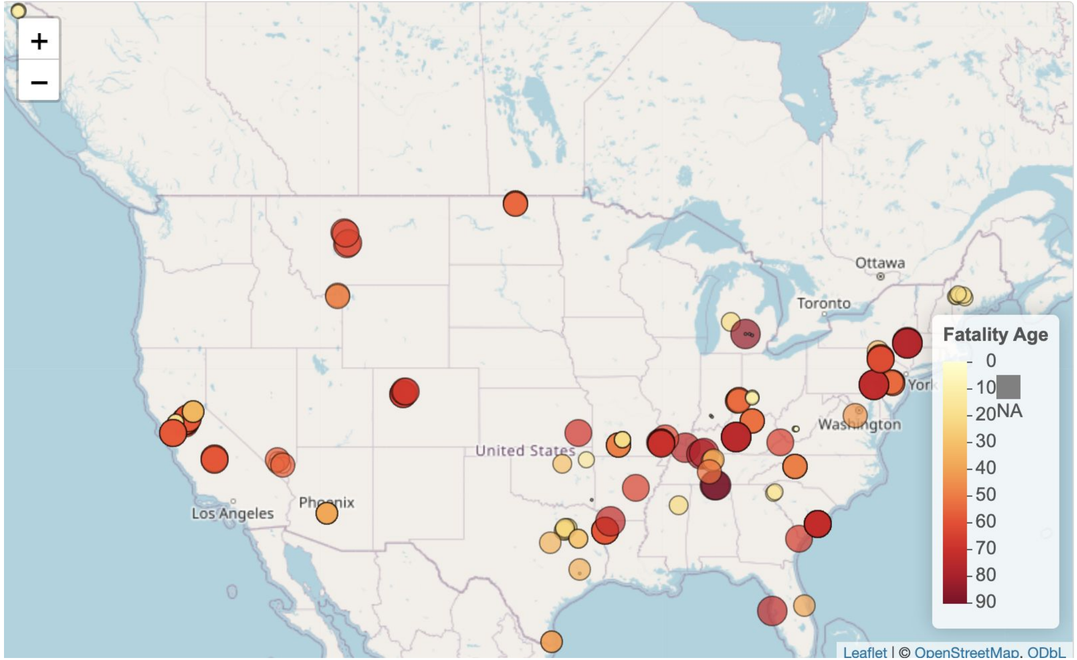
Overlay of Fatalities and Storm Locations

- The size and color of fatality markers reflect the number of fatalities:
  - Red markers = higher fatalities
  - Orange markers = lower fatalities
- Pop Ups provide more information about the deaths and storm type.
- Shows the the relationship between the storm centers and the distances to the deaths
- Spatial proximity of fatalities to storm centers may suggest areas with poor infrastructure or delayed evacuation responses.



Interactive Fatality Map with Severity Levels

- Markers represent fatalities, with sizes scaled based on the age of fatalities
- Colors are determined by the severity of the fatalities (e.g., older individuals may be represented with different colors)
- Popups offer detailed information about the fatality, such as age and type of fatality
- Shows the relationship between the age of fatalities and geographic location.
- Understanding fatality age or other severity indicators can inform policy decisions and aid in identifying vulnerable populations or places/villages



Property / Crop Damage caused by Storms

Questions we want to answer:

- How much damage occurred from storms in 2023?
- Where did the damage occur and what was the severity?
- Do certain types of storms cause more damage?
- Do storms cause more damage during certain times of the year?

Data Cleaning for Storm Damage

- Combine data tables for “Storm_locations” and “Storm_Details” by “EVENT_ID”
- Filter out any storm with N/A values for both crop and property damage or no damage at all
- Add “LATITUDE” and “LONGITUDE” to the dataset
- Change the damage values from chr type to a usable int value EX. 10.00K to 10,000
- Filtered the location to only include storms on the mainland U.S.

```
onlyDamage <- details %>%
  filter(!(is.na(DAMAGE_PROPERTY) | DAMAGE_PROPERTY == "") | !(is.na(DAMAGE_CROPS) | DAMAGE_CROPS == ""))

damageWLoc <- onlyDamage %>%
  semi_join(locations, by = "EVENT_ID")


locationsID <- locations %>%
  select(EVENT_ID, LATITUDE, LONGITUDE)

damageWLoc <- left_join(damageWLoc, locationsID, by = "EVENT_ID")

damageWLoc <- damageWLoc %>%
  filter(damageWLoc$LATITUDE > 25 & damageWLoc$LATITUDE < 50)

damageWLoc$DAMAGE_PROPERTY <- ifelse(
  grepl("K$", damageWLoc$DAMAGE_PROPERTY),
  as.numeric(gsub("K", "", damageWLoc$DAMAGE_PROPERTY)) * 1000,
  ifelse(
    grepl("M$", damageWLoc$DAMAGE_PROPERTY),
    as.numeric(gsub("M", "", damageWLoc$DAMAGE_PROPERTY)) * 1000000,
    0
  )
```

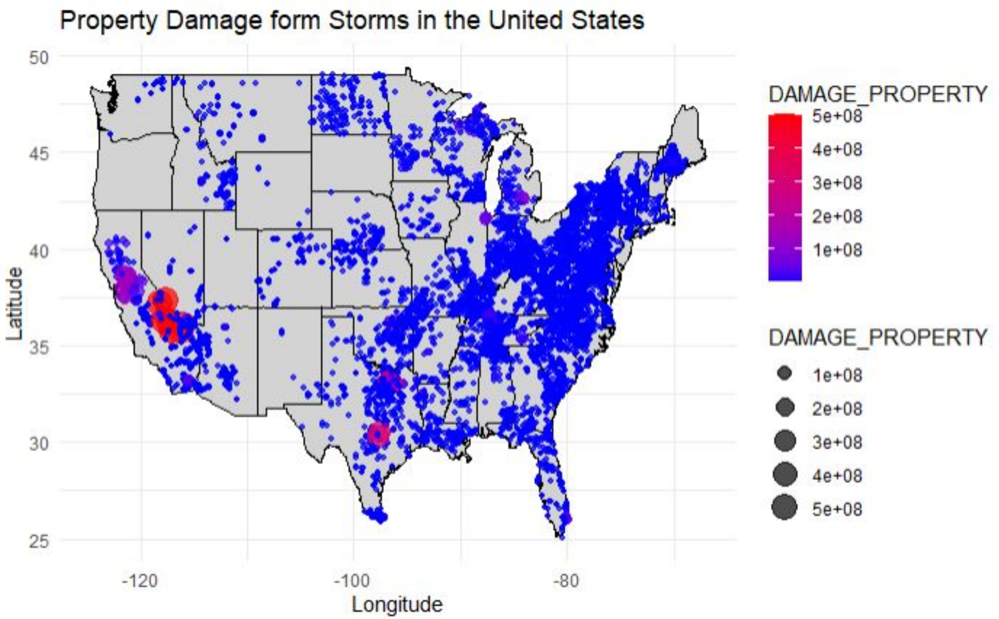
Total Storm Damage in 2023

- Total of 38,360 storms in the combined dataset with 11,750 causing damage
- About 32.9% of storms caused property or crop damage
- Total of $12,564,324,980 of property damage, $187,665,561 of crop damage
- Average damage of about $1,070,000 per storm

```
# Shows that 32.9% of all storms recorded caused at least either crop or property damage
damPercent <- nrow(stormsWDamage) / nrow(damageWLoc)
# $4,151,324,010 of property damage occurred during 2023
total_dam_prop <- sum(damageWLoc$DAMAGE_PROPERTY)
# $99,082,870 of crop damage occurred during 2023
total_dam_crop <- sum(damageWLoc$DAMAGE_CROPS)
```

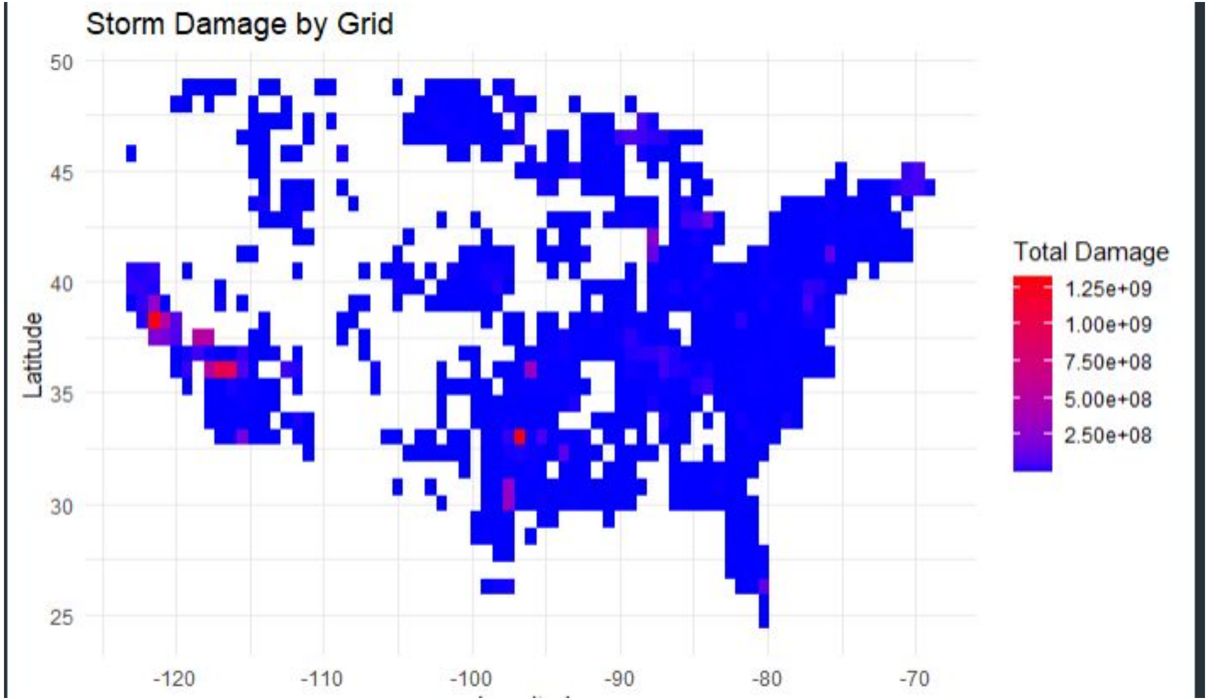
Where did Property Damage Occur and How Severe was it

- 11,457 storms caused property damage
- Severe damage occurred in major cities like Dallas, Austin, San Francisco
- Damage was more frequent on the Eastern half of the United States.



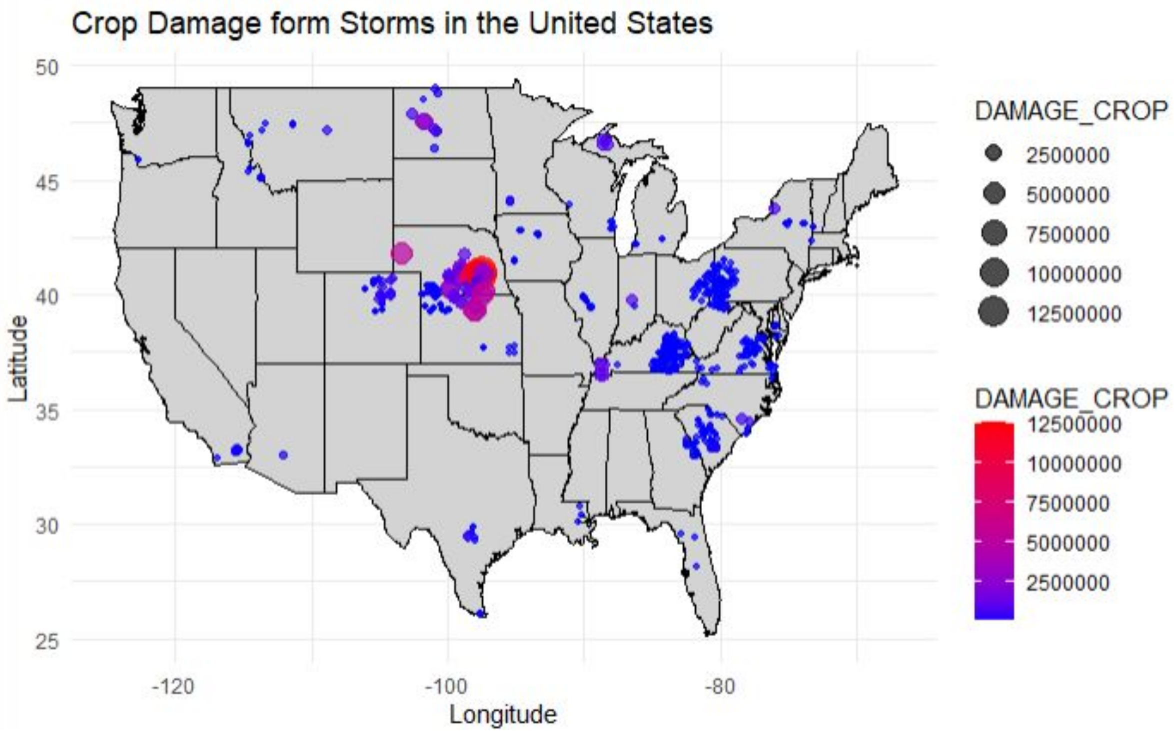
Where did Property Damage Occur and How Severe was it

- 11,457 storms caused property damage
- Severe damage occurred in major cities like Dallas, Austin, San Francisco
- Damage was more frequent on the Eastern half of the United States.



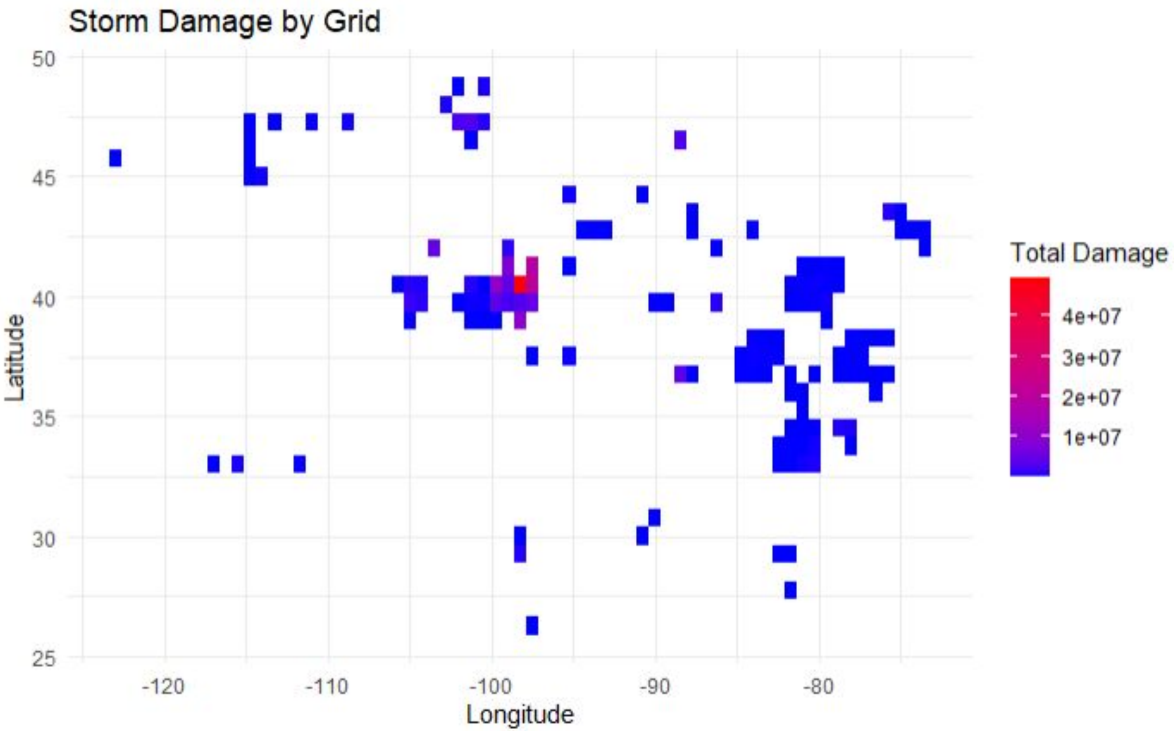
Where Did the Crop Damage Occur and How Severe was it

- 867 storms caused crop damage
- Severe amounts of damage occurred in Nebraska and Kansas
- Clusters of damage from hurricane Ophelia on East coast
- Damage along rivers in North Dakota from flooding



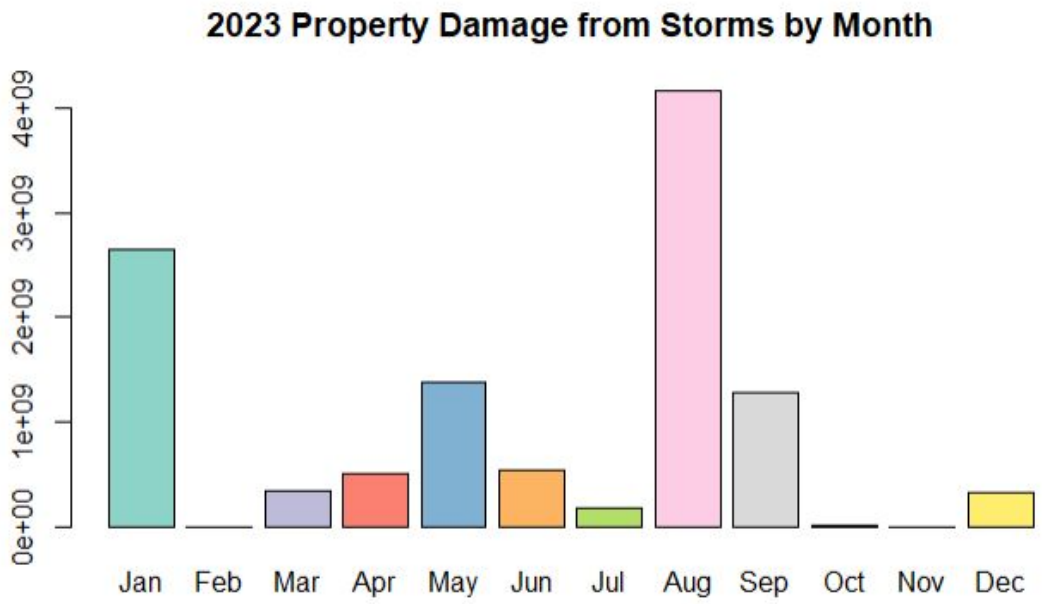
Where Did the Crop Damage Occur and How Severe was it

- 867 storms caused crop damage
- Severe amounts of damage occurred in Nebraska and Kansas
- Clusters of damage from hurricane Ophelia on East coast
- Damage along rivers in North Dakota from flooding



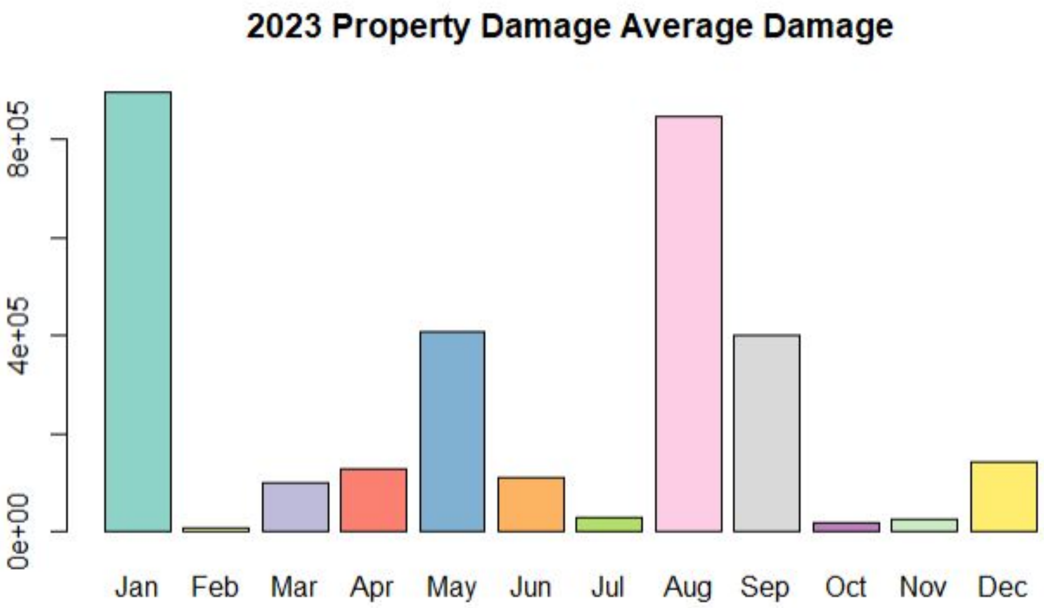
When does the Most Property Damage Occur

- August had the most amount of property damage
- Peak tornado season is between May and early June
- Peak hurricane season is from August to mid October
- January has a lot of damage for the amount of storms during the month



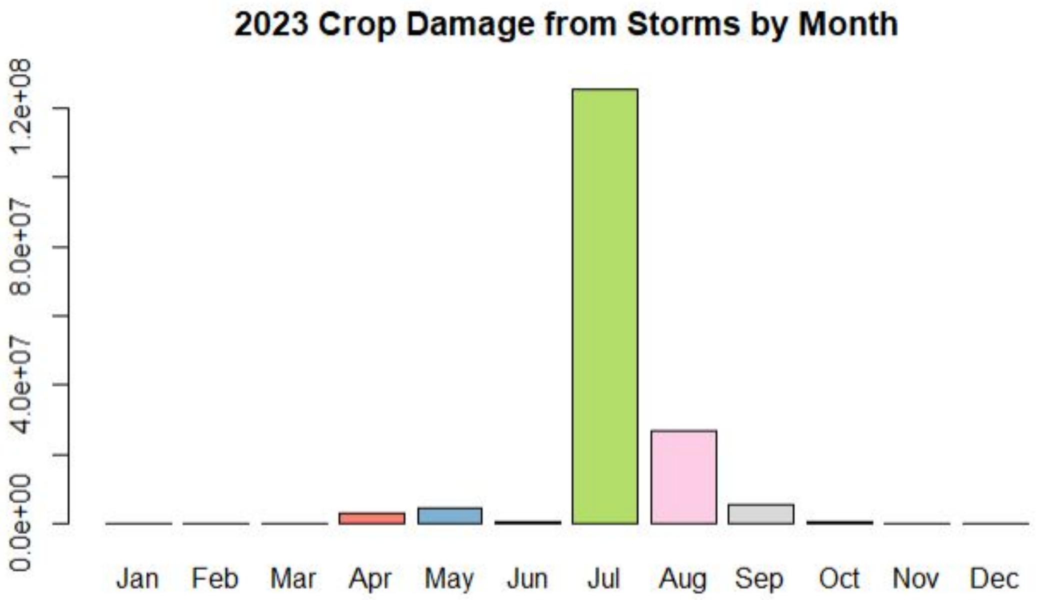
Which Month is the Most Severe for Damage

- Graph shows the average amount of property damage from each storm during the month
- January and August are the most severe with around $850,000 of damage per storm
- February and October are the least severe with $10,000 of damage per storm



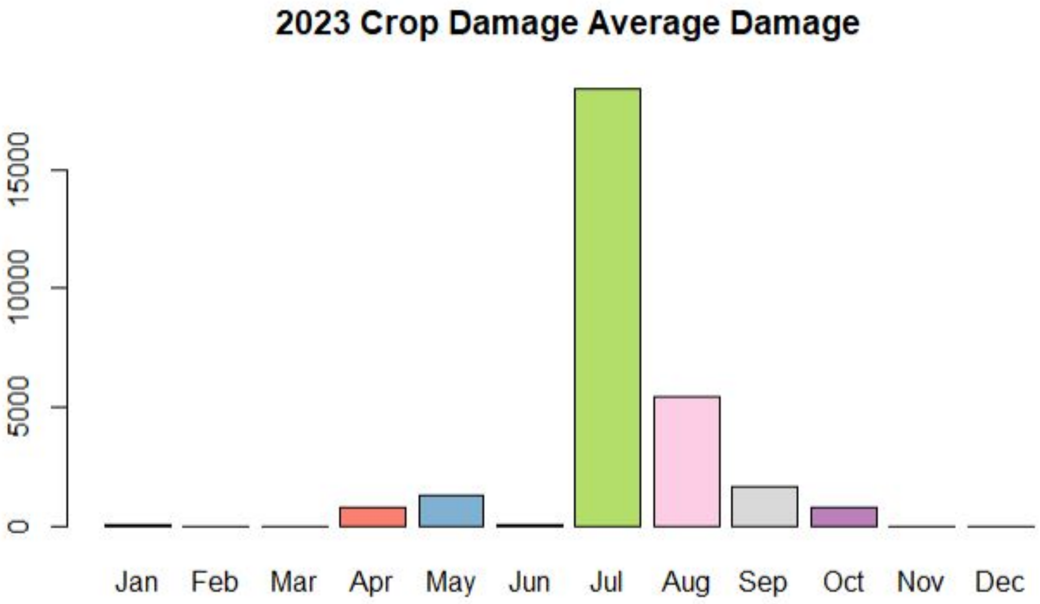
When Does the Most Crop Damage Occur

- The majority of crop damage occured in July ($120,000,000)
- August, September, and May also saw some crop damage
- These months are the peak life cycle for crops



What Month is the Most Severe for Crop Damage

- July is the most severe month for crop damage with around $15,000 of damage per storm



What Type of Storms Cause the Most Damage

Top 3 Causes of Property Damage:
- Flash Floods / Floods ($8,080,000,000)

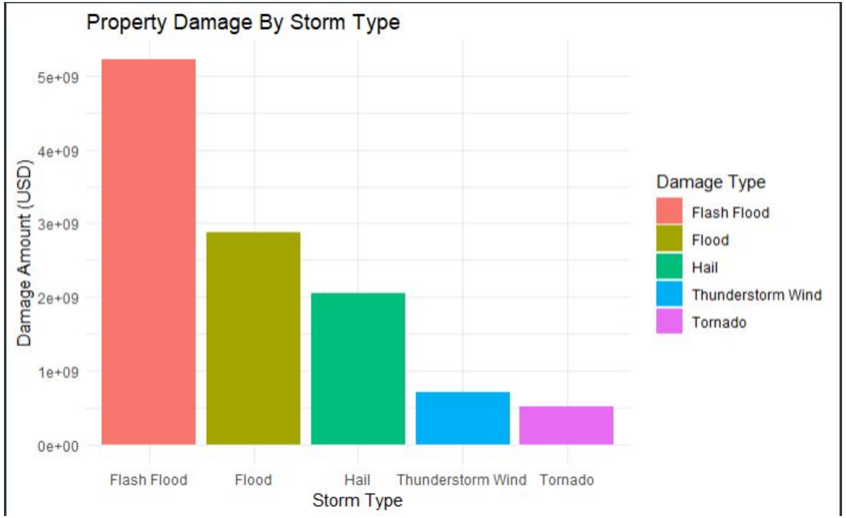
- Hail ($2,049,000,000)

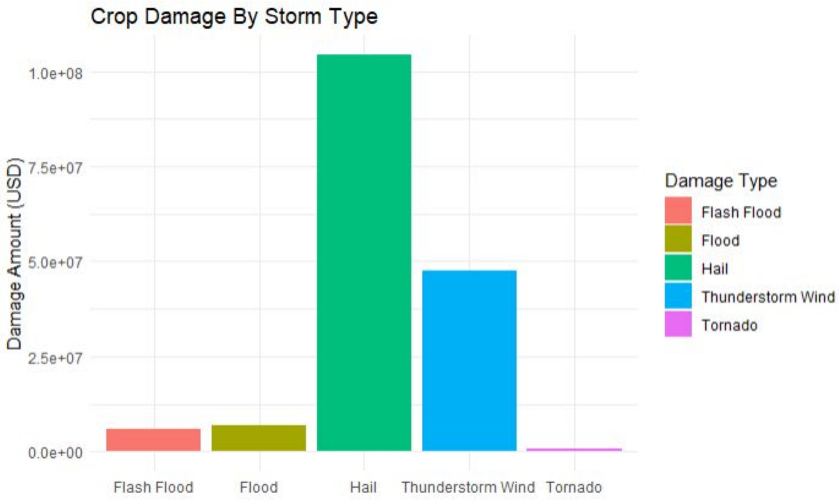
- Thunderstorm Wind ($710,969,800)

Top 3 Causes of Crop Damage:
- Hail ($104,261,000)

- Thunderstorm Wind ($47,482,000)

- Flood ($6,889,260)





What Type of Storms are the Most Severe

Top 3 Most Severe Property Damage Causes:
- Flash Floods / Floods ($582,897)

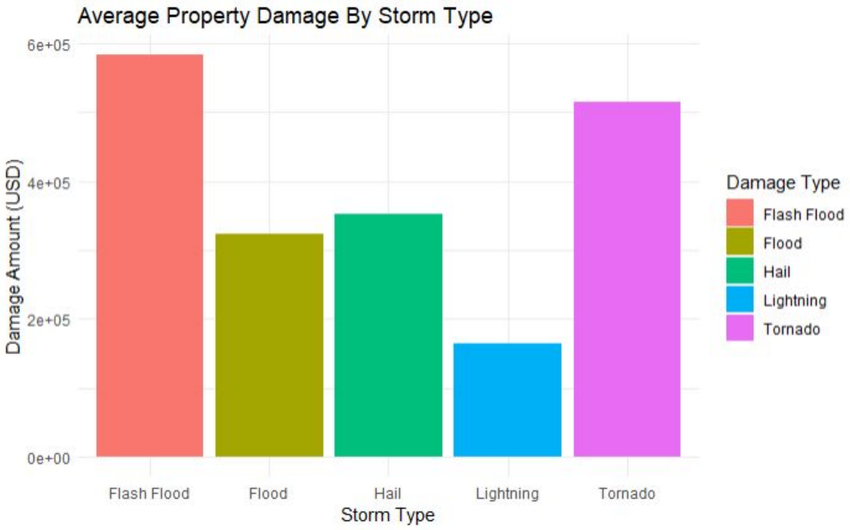
- Tornado($515,298)

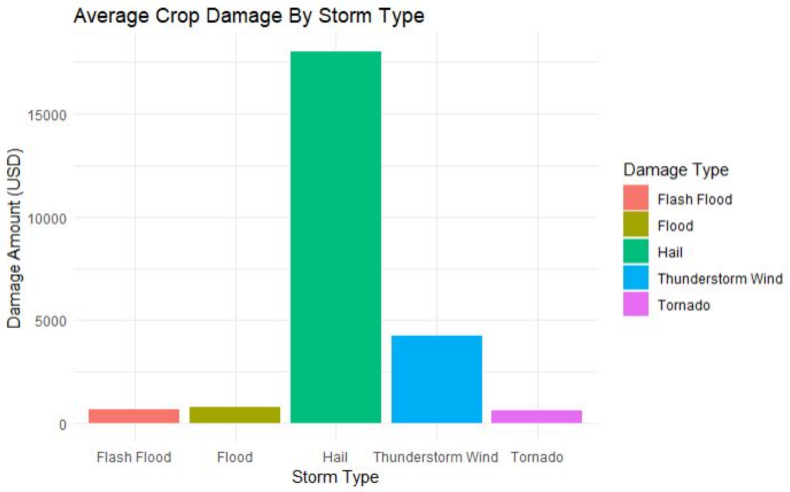
- Hail ($353,278)

Top 3 Most Severe Crop Damage Causes:
- Hail ($17,976)

- Lightning($5,734)

- Thunderstorm Wind ($4,221)






Thank
you