

Exploring the NOAA Storm Database

Abdallah Mohammed

2024-10-04

Introduction

This project focuses on exploring the U.S. National Oceanic and Atmospheric Administration (NOAA) storm database, which provides detailed information on significant storms and weather events across the United States. The database records the location and timing of these events, as well as their impact in terms of fatalities, injuries, and property damage.

Severe weather events, such as storms, can lead to major public health and economic challenges for communities and local governments. Many of these events result in significant loss of life, injury, and damage to property, making it crucial to mitigate these impacts as much as possible.

The objective of this analysis is to investigate the NOAA Storm Database and answer key questions about severe weather events:

1. Which types of events are most harmful to population health across the United States?
2. Which types of events have the most significant economic consequences nationwide?

Data was analyzed using R, with a focus on identifying the types of events that are most harmful to public health and those that result in the greatest economic losses. Graphs and tables were used to illustrate the findings

Data Processing

Reading data and calculating damage from storms and floods

```
StormData <- read.csv("StormData.csv")

# Convert PROPDGMGEXP values
StormData$PROPDGMGEXP <- toupper(StormData$PROPDGMGEXP) # Convert all symbols to uppercase
convert_prop <- function(exp) {
  if (exp == "K") return(1e3)
  else if (exp == "M") return(1e6)
  else if (exp == "B") return(1e9)
  else if (exp == "H") return(1e2)
  else if (exp %in% 0:8) return(10^as.numeric(exp))
  else return(1) # dealing with empty values and unknown values
```

```

}

StormData$PROPDMGEXP_numeric <- sapply(StormData$PROPDMGEXP, convert_prop)

# Convert units to numbers
StormData$CROPDMGEXP <- toupper(StormData$CROPDMGEXP) # Convert all symbols to uppercase
convert_crop <- function(exp) {
  if (exp == "K") return(1e3)
  else if (exp == "M") return(1e6)
  else if (exp == "B") return(1e9)
  else if (exp %in% 0:8) return(10^as.numeric(exp))
  else return(1) # dealing with empty values and unknown values
}

StormData$CROPDMGEXP_numeric <- sapply(StormData$CROPDMGEXP, convert_crop)

# Calculate final damage values
StormData$Total_PROPDMG <- StormData$PROPDMG * StormData$PROPDMGEXP_numeric
StormData$Total_CROPDMG <- StormData$CROPDMG * StormData$CROPDMGEXP_numeric

```

Analyze the data

The questions

1. What types of events are most detrimental to the health of the population across the United States?

The goal is to create a subset of the number of injuries and deaths for each type of event (such as hurricanes or floods) and then plot this data to see which events cause the most injuries or deaths.

Answer Steps: Create a subset of data: It collects the data related to injuries and deaths (INJURIES and FATALITIES) for each type of event (EVTYPE), such as hurricanes or floods.

Data Analysis: Using the summarise function from the dplyr library, it will collect the total injuries and deaths for each type of event.

Visualization: The ggplot2 library is used to plot the data. The plot will show which type of event caused the most injuries or deaths.

Why is this important? Knowing which events are most damaging to a population helps direct resources towards early warnings and preventive measures. As Benjamin Franklin said, “An ounce of prevention is worth a pound of cure.”

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

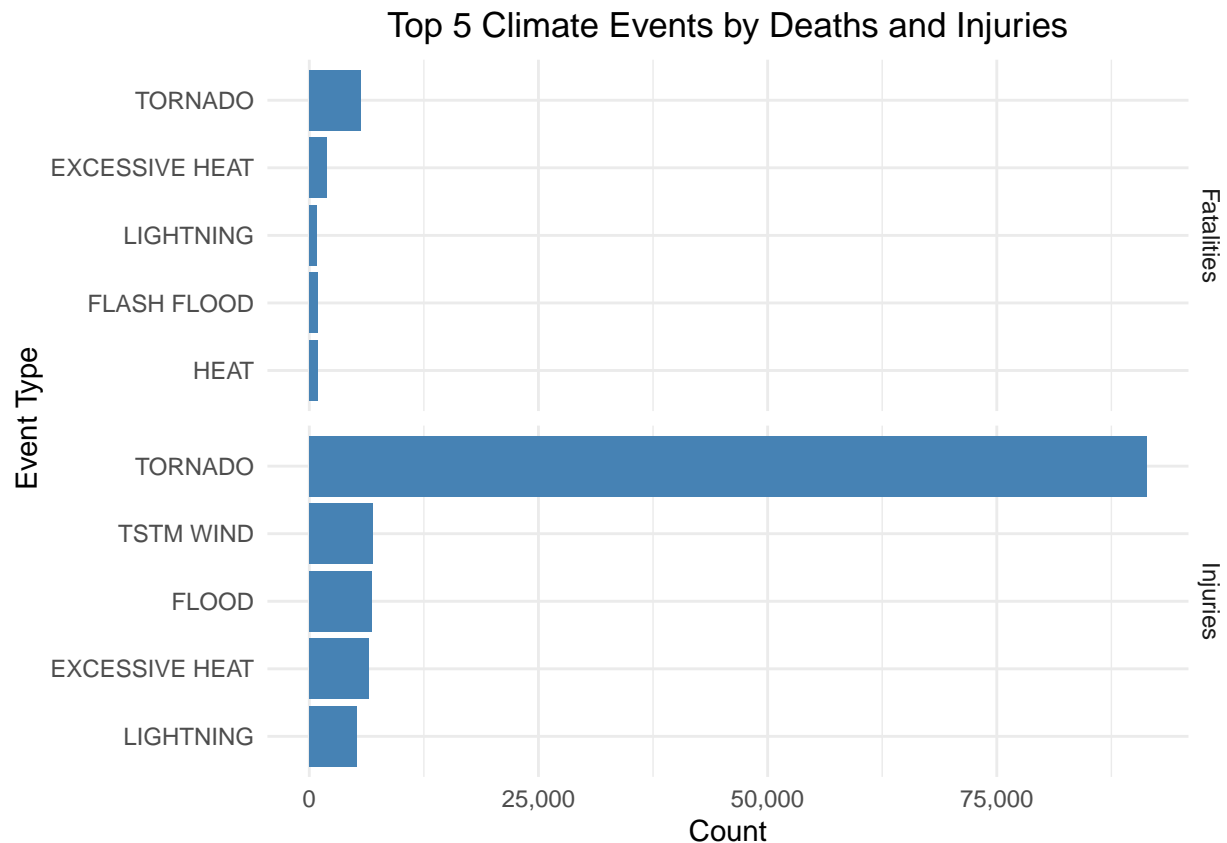
```
library(ggplot2)
library(reshape2)

Injuries_and_deaths_per_event <- StormData %>% group_by(EVTYPE) %>%
  summarise(Injuries = sum(INJURIES), deaths = sum(FATALITIES)) %>%
  melt(.,id = c("EVTYPE"),measure.vars = c("Injuries","deaths")) %>%
  arrange(desc(value))

max_Injuries <- Injuries_and_deaths_per_event %>%
  filter(variable == "Injuries") %>%
  filter(value %in% sort(value,decreasing=T)[1:5])
max_deaths <- Injuries_and_deaths_per_event %>%
  filter(variable == "deaths") %>%
  filter(value %in% sort(value,decreasing=T)[1:5])

combined_Health_Impact_data <- rbind(
  data.frame(EVTYPE = max_deaths$EVTYPE, value = max_deaths$value, variable = "Fatalities"),
  data.frame(EVTYPE = max_Injuries$EVTYPE, value = max_Injuries$value, variable = "Injuries")
)

ggplot(data = combined_Health_Impact_data, aes(x = reorder(EVTYPE, value), y = value)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  facet_grid(variable ~ ., scales = "free_y") +
  coord_flip() +
  labs(title = "Top 5 Climate Events by Deaths and Injuries",
       x = "Event Type",
       y = "Count") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



2. What types of events have the most significant economic consequences nationwide?

The goal is to create a subset of data that contains the total property and crop damage for each type of event (such as hurricanes or floods) and then plot this data to see which events cause the most economic damage.

Answer Steps: Create a subset of data: Collects data on property and crop damage (PROPDMG and CROPDMG) for each type of event (EVTYPE), such as hurricanes or floods.

Data analysis: Using the summarise function from the dplyr library, you will sum the total property and crop damage for each type of event.

Visualization: The ggplot2 library is used to plot the data. The plot will show which type of event caused the most economic damage to property and crops.

Why is this important? Knowing which events cause the most economic damage helps direct efforts to protect resources and property. Proactive measures can be taken to mitigate this damage, whether it is through insurance or strengthening infrastructure to cope with these events.

```
library(tidytext)
PROPDMG_and_CROPDMG_per_event <- StormData %>% group_by(EVTYPE) %>%
  summarise(PROPDMG = sum(Total_PROPDMG), CROPDMG = sum(Total_CROPDMG))%>%
  melt(.,id = c("EVTYPE"),measure.vars = c("PROPDMG", "CROPDMG")) %>%
  arrange(desc(value))
```

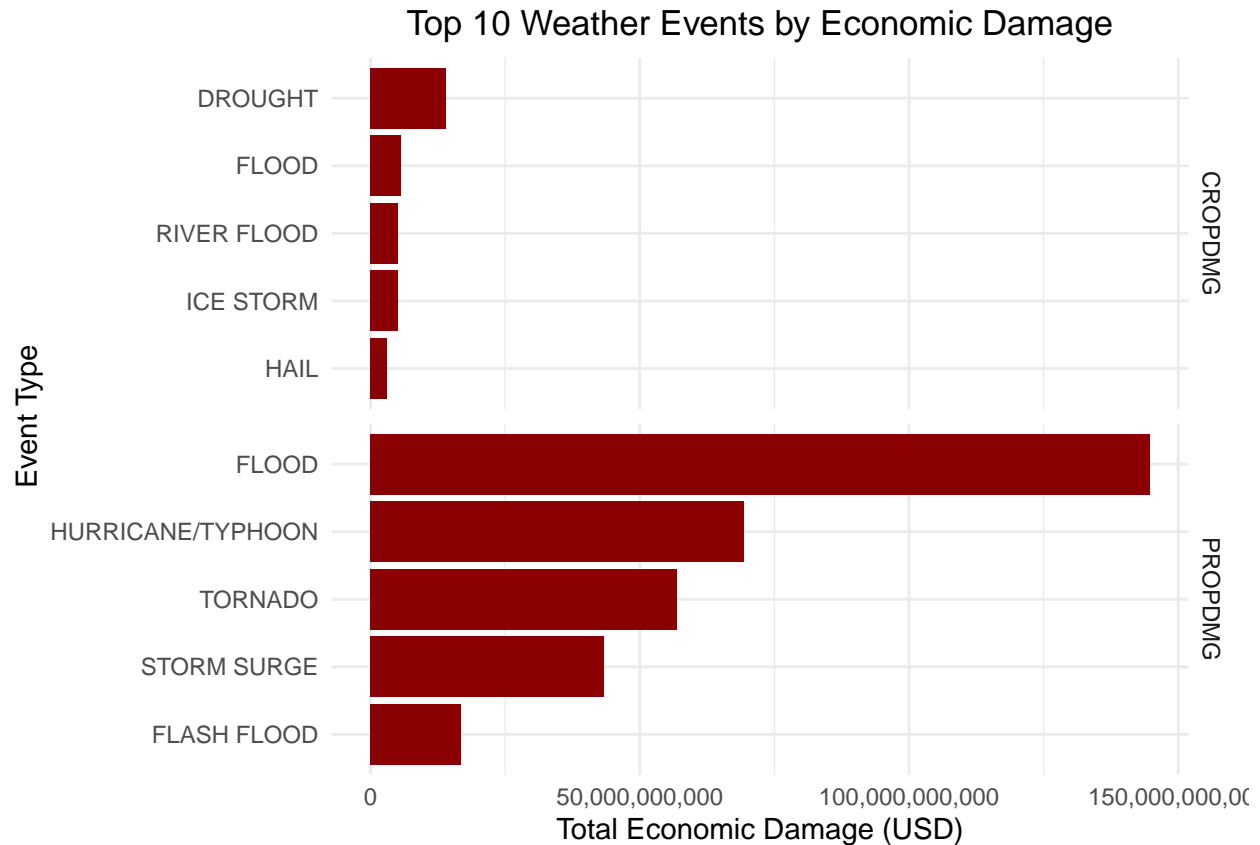
```

max_PROPDMG <- PROPDMG_and_CROPDMG_per_event %>%
  filter(variable == "PROPDMG") %>%
  filter(value %in% sort(value,decreasing=T)[1:5])
max_CROPDMG <- PROPDMG_and_CROPDMG_per_event %>%
  filter(variable == "CROPDMG") %>%
  filter(value %in% sort(value,decreasing=T)[1:5])

combined_Economic_Damage_data <- rbind(
  data.frame(EVTYPE = max_PROPDMG$EVTYPE, value = max_PROPDMG$value, variable = "PROPDMG"),
  data.frame(EVTYPE = max_CROPDMG$EVTYPE, value = max_CROPDMG$value, variable = "CROPDMG")
)

ggplot(data = combined_Economic_Damage_data,
  aes(x = reorder_within(EVTYPE, value, variable), y = value)) +
  geom_bar(stat = "identity", fill = "darkred") +
  facet_grid(variable ~ ., scales = "free_y") +
  coord_flip() +
  labs(title = "Top 10 Weather Events by Economic Damage",
    x = "Event Type",
    y = "Total Economic Damage (USD)") +
  scale_y_continuous(labels = scales::comma) +
  scale_x_reordered() +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))

```



Results

1. Events with the greatest public health impact:

- The results showed that TORNADO was the most damaging to the population, with the highest rates of injury and death. For example, TORNADO caused more than 9.1346×10^4 of injuries and 5633 of deaths, making it among the deadliest events.

2. Events with the greatest economic impact:

- The data showed that FLOOD and DROUGHT were the most damaging to the economy, with the highest rates of property and crop damage. For example, the FLOOD hurricane/flood caused more than 1.4465771×10^{11} in property damage, while the DROUGHT hurricane/flood caused 1.3972566×10^{10} in crop damage.