

# Most Severe Weather Events in the United States and Their Economic Consequences

## Synopsis

This document analyzes storm data from the National Oceanic and Atmospheric Administration (NOAA) Storm Database. The document seeks to answer which types of natural events were most harmful to the United States' population from 1950-1991. It will investigate which type of natural event had the most economic consequences during the time. To answer these questions, data collected from previous years will be utilized. Important factors we will look at include storm locations, the time events took place, amount property damage caused, the number of injuries, and the number of fatalities.

Using this data, it was deduced that tornadoes were the most frequent natural disaster in the United States which caused the most injuries and fatalities. It was also observed that tornadoes caused the biggest economic impact in terms of damage to property.

## Data Processing

The data used for this assignment was collected from April 1950 to April 1991. We see a surge in the number of recorded events after 1973 in comparison to the years previous. This is likely due to improved technology and storm monitoring. The storm data was downloaded from the NOAA Storm Database: <https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>

After downloading the NOAA data, the .csv file is read into RStudio.

```
CSVfile <- read.csv("repdata_data_StormData.csv")
```

## Cleaning the Data

The first step in cleaning the data is to choose the variables of interest for analysis.

```
cleanData <- CSVfile[, c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG")]
cleanData$EVTYPE <- gsub("^HEAT$", "EXCESSIVE HEAT", cleanData$EVTYPE)
cleanData$EVTYPE <- gsub("^TSTM WIND$", "THUNDERSTORM WIND", cleanData$EVTYPE)
cleanData$EVTYPE <- gsub("^THUNDERSTORM WIND$", "THUNDERSTORM WINDS", cleanData$EVTYPE)
```

Aggregate the data to find the natural event type that caused the most injuries.

```
aggregatedInjuries <- aggregate(cleanData$INJURIES, by = list(cleanData$EVTYPE),
                                FUN = sum, na.rm=TRUE)
colnames(aggregatedInjuries) = c("event.type", "injury.total")
catInjuries <- aggregatedInjuries[order(-aggregatedInjuries$injury.total) , ]
mostInjuries <- catInjuries[1:15 , ]
mostInjuries$event.type <- factor(mostInjuries$event.type,
                                levels = mostInjuries$event.type,
                                ordered = TRUE)
```

Aggregate the data to find the natural event type that caused the most fatalities.

```
aggregatedFatalities <- aggregate(cleanData$FATALITIES, by = list(cleanData$EVTYPE),
                                FUN = sum, na.rm = TRUE)
```

```
colnames(aggregatedFatalities) = c("event.type", "fatality.total")
catFatalities <- aggregatedFatalities[order(-aggregatedFatalities$fatality.total) , ]
mostFatalities <- catFatalities[1:15 , ]
mostFatalities$event.type <- factor(mostFatalities$event.type,
                                   levels = mostFatalities$event.type,
                                   ordered = TRUE)
```

Aggregate the data to find the natural event type that caused the most damage to property.

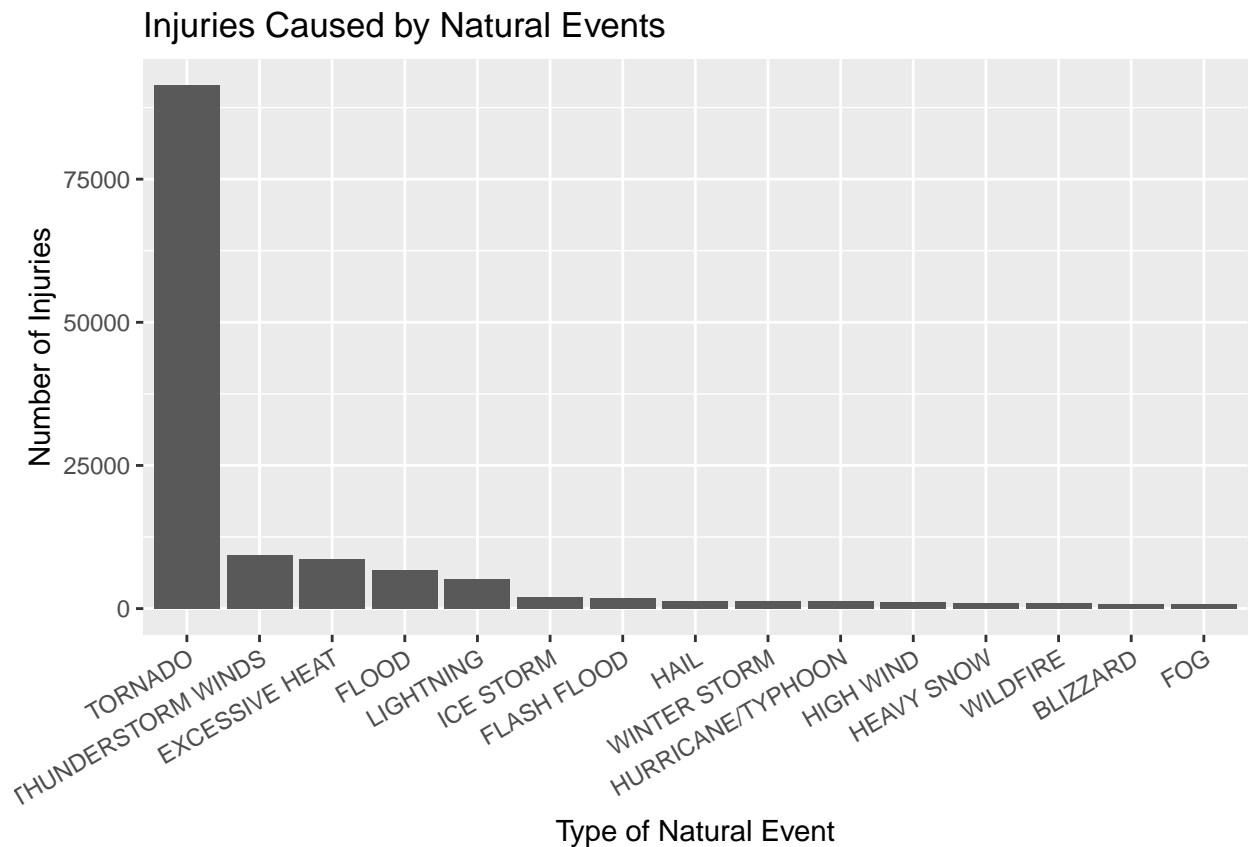
```
aggregatedPropertyD <- aggregate(cleanData$PROPDMG, by = list(cleanData$EVTYPE),
                                FUN = sum, na.rm = TRUE)
colnames(aggregatedPropertyD) = c("event.type", "prop.dmg.total")
catPropertyD <- aggregatedPropertyD[order(-aggregatedPropertyD$prop.dmg.total) , ]
mostPropertyD <- catPropertyD[1:15 , ]
mostPropertyD$event.type <- factor(mostPropertyD$event.type,
                                   levels = mostPropertyD$event.type,
                                   ordered = TRUE)
```

## Results

Once the data has been cleaned, the results are graphed for analysis.

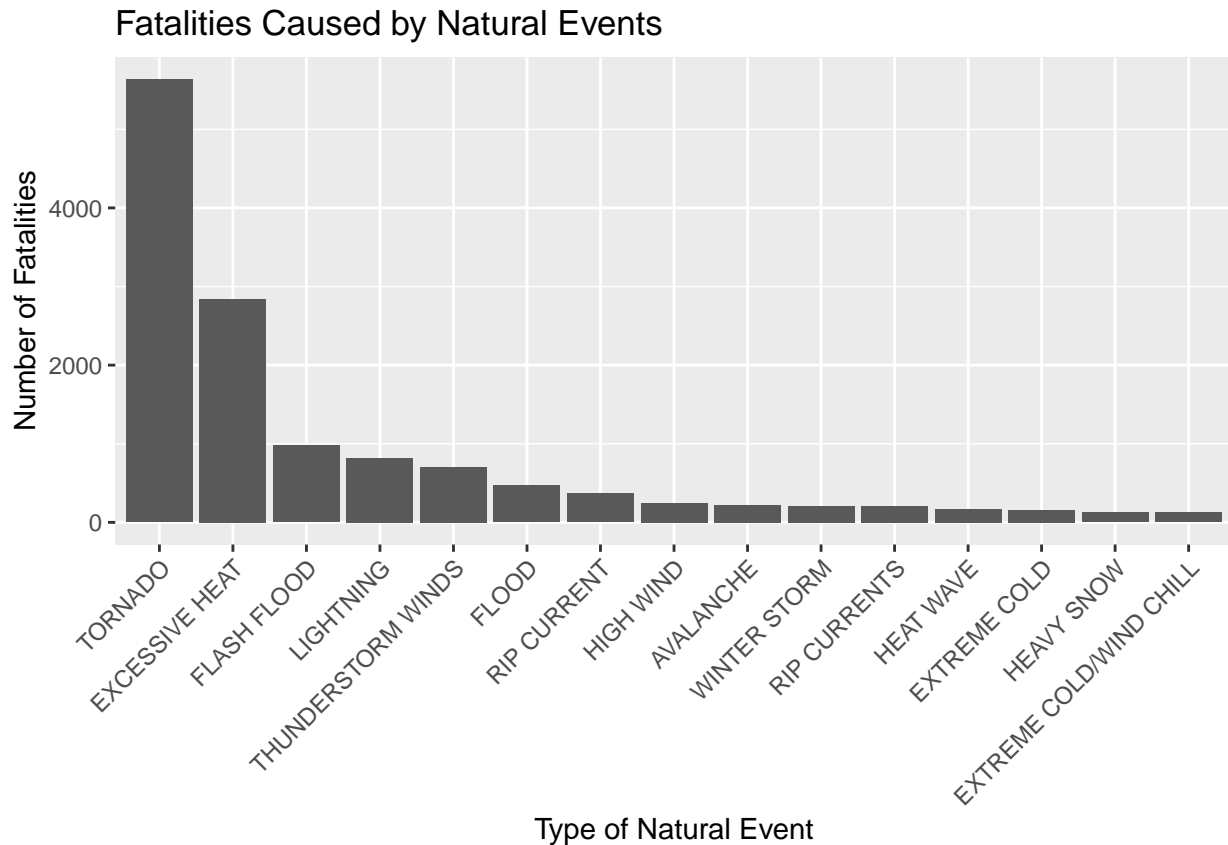
Here is the graph showing the number of injuries caused by natural events in the United States from 1951 to 1991. As we can see, tornadoes caused the most injuries.

```
library(ggplot2)
ggplot(data = mostInjuries, aes(x = event.type, y = injury.total)) +
  ggtitle("Injuries Caused by Natural Events") +
  geom_bar(stat = "identity") + xlab("Type of Natural Event") +
  ylab("Number of Injuries") +
  theme(axis.text.x = element_text(angle = 30, hjust = 1))
```



Here is the graph showing the number of fatalities caused by natural events in the United States from 1951 to 1991. As we can see, tornadoes caused the most fatalities.

```
library(ggplot2)
ggplot(data = mostFatalities, aes(x = event.type, y = fatality.total)) +
  ggtitle("Fatalities Caused by Natural Events") +
  geom_bar(stat = "identity") + xlab("Type of Natural Event") +
  ylab("Number of Fatalities") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Here is the graph showing the number of properties that sustained damage caused by natural events in the United States from 1951 to 1991. As we can see, tornadoes caused the most property damage.

```
library(ggplot2)
ggplot(data = mostPropertyD, aes(x = event.type, y = prop.dmg.total)) +
  ggtitle("Property Damage Caused by Natural Events") +
  geom_bar(stat = "identity") + xlab("Type of Natural Event") +
  ylab("Number of Properties Damaged") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

