# Reproducible Research Project 2

Minh Tri 04/04/2022

## Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. The basic goal of this report is to explore the NOAA Storm Database by using R and answer some basic questions about severe weather events.

## Data Processing

The data for the analysis was downloaded from the NOAA storm database. After the data is downloaded from the website, it is uncompressed and read into R environment

```
setwd("D:/Statistics/R/R data/Reproducible Research Project 2")
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile = "r
data <- read.csv(bzfile("repdata_data_StormData.csv.bz2"))
```

## What causes the most injuries?

To answer the 1st question, we create a bar chart illustrate the relation between Event type(x axis) and the number of injuries(y axis)

At the beggining, let summerize the number of injuries according to Event type. After that, reorder the event and drawing a bar graph.

```
injuries <- aggregate(data$INJURIES, by = list(EVENT= data$EVTYPE), sum)
injuries <- injuries[order(injuries$x, decreasing = TRUE), ]
head(injuries)
```

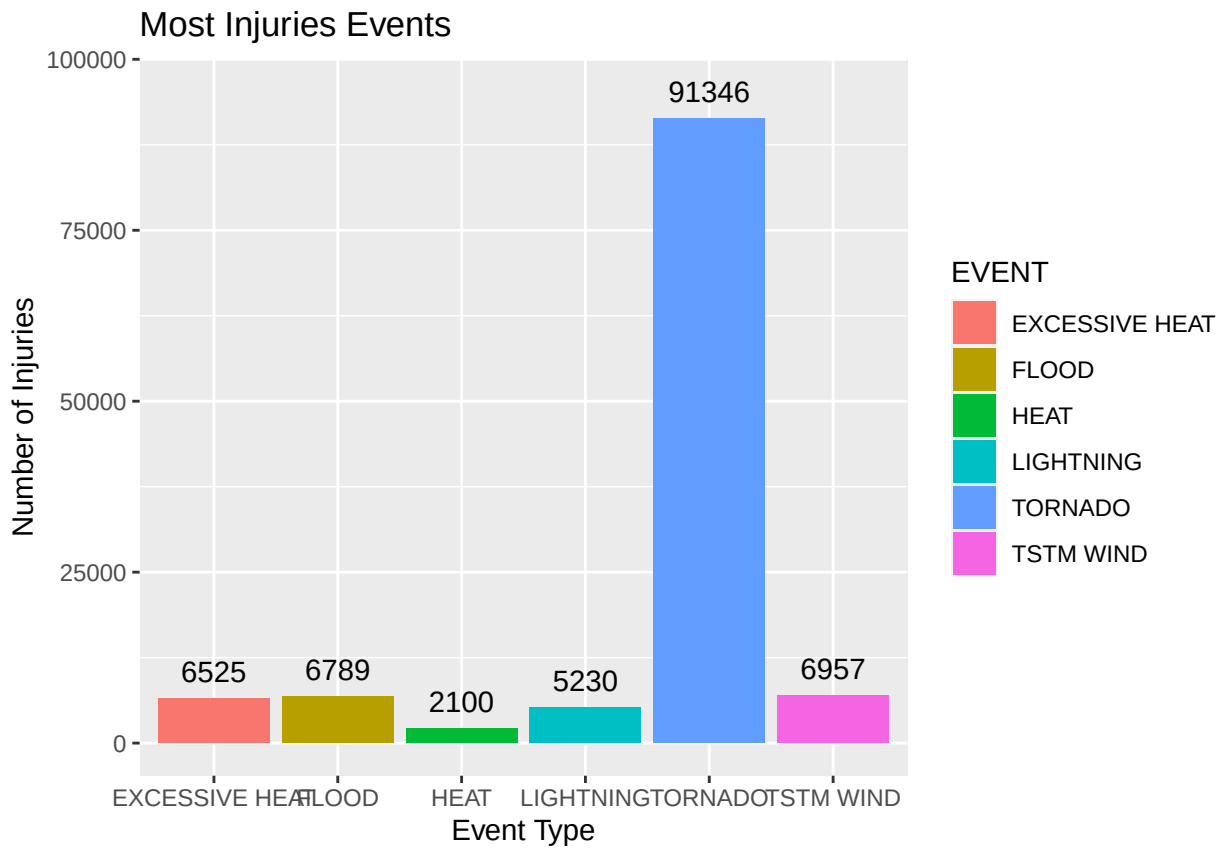
```
##              EVENT     x
## 834        TORNADO 91346
## 856      TSTM WIND  6957
## 170          FLOOD  6789
## 130 EXCESSIVE HEAT  6525
## 464      LIGHTNING  5230
## 275           HEAT  2100
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
bar <- ggplot(injuries[1:6, ], aes(EVENT,  x, fill = EVENT, label = x))
bar + stat_summary(geom = "bar") + labs(x = "Event Type", y = "Number of Injuries") + geom_text(nudge_y
```

```
## No summary function supplied, defaulting to 'mean_se()'
```




Looking at the graph, we can see that TORNADO is responsible for most of injuries (91346 events occurred)

### What causes the most fatalities?

To answer this question, we create a bar chart illustrate the relation between Event type(x axis) and the number of fatalitis(y axis)

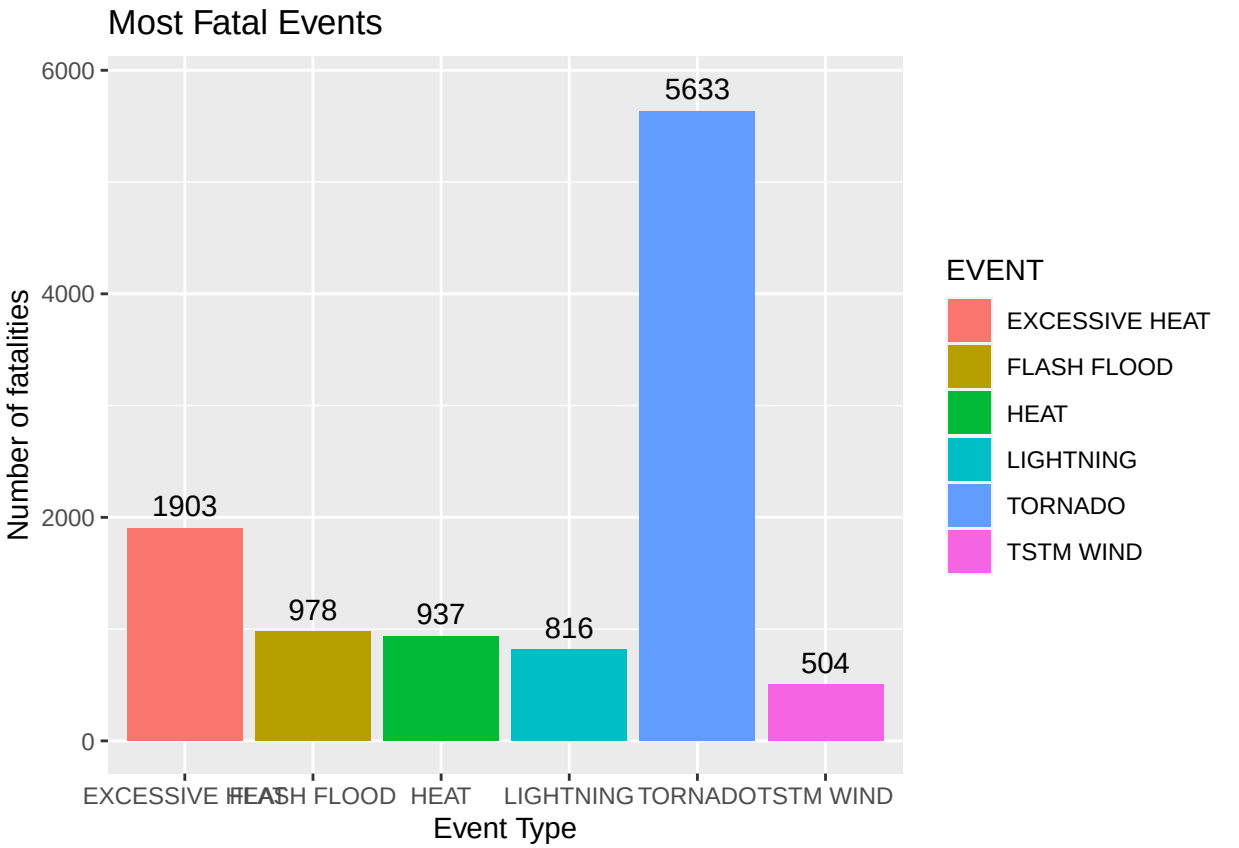
At the beggining, let summerize the number of fatalities according to Event type. After that, reorder the event and drawing a bar graph.

```
fatalities <- aggregate(data$FATALITIES, by = list(EVENT= data$EVTYPE), sum)
fatalities <- fatalities[order(fatalities$x, decreasing = TRUE), ]
head(fatalities)
```

```
##              EVENT    x
## 834        TORNADO 5633
## 130 EXCESSIVE HEAT 1903
## 153    FLASH FLOOD  978
## 275           HEAT  937
## 464      LIGHTNING  816
## 856      TSTM WIND  504
```

```
library(ggplot2)
bar1 <- ggplot(fatalities[1:6, ], aes(EVENT,  x, fill = EVENT, label = x))
bar1 + stat_summary(geom = "bar") + labs(x = "Event Type", y = "Number of fatalities") + geom_text(nudg
```

```
## No summary function supplied, defaulting to ‘mean_se()‘
```




Looking at the graph, we can see that TORNADO is responsible for most of fatalities (5633 events occurred)

## Across the United States, which types of events have the greatest economic consequences?

Checking the all the characters of PRO/CROPDMGEXP variables

```
unique(data$PROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
```

```
unique(data$CROPDMGEXP)
```

```
## [1] ""  "M" "K" "m" "B" "?" "0" "k" "2"
```

Changing these characters to upper case

```
data$PROPDMGEXP <- toupper(data$PROPDMGEXP)
data$CROPDMGEXP <- toupper(data$CROPDMGEXP)
unique(data$PROPDMGEXP); unique(data$CROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "+" "0" "5" "6" "?" "4" "2" "3" "H" "7" "-" "1" "8"
```

```
## [1] ""  "M" "K" "B" "?" "0" "2"
```

Assigning numeric value according to the characters: Billion (9), Hundred (2), Kilo (3), and Million (6)

```
data[data$PROPDMGEXP == "B", "PROPDMGEXP"] <- 9
data[data$PROPDMGEXP == "M", "PROPDMGEXP"] <- 6
data[data$PROPDMGEXP == "K", "PROPDMGEXP"] <- 3
data[data$PROPDMGEXP == "H", "PROPDMGEXP"] <- 2
data[data$PROPDMGEXP %in% c("", "+", "-", "?"), "PROPDMGEXP"] <- "0"
data[data$CROPDMGEXP %in% c("", "+", "-", "?"), "CROPDMGEXP"] <- "0"
data[data$CROPDMGEXP == "B", "CROPDMGEXP"] <- 9
data[data$CROPDMGEXP == "M", "CROPDMGEXP"] <- 6
data[data$CROPDMGEXP == "K", "CROPDMGEXP"] <- 3
data[data$CROPDMGEXP == "H", "CROPDMGEXP"] <- 2
unique(c(data$PROPDMGEXP, data$CROPDMGEXP))
```

```
##  [1] "3" "6" "0" "9" "5" "4" "2" "7" "1" "8"
```

Assign the PDMGEXP value

```
data$PROPDMGEXP <- 10^(as.numeric(data$PROPDMGEXP))
data$CROPDMGEXP <- 10^(as.numeric(data$CROPDMGEXP))
```

Calculate the total damage

```
data$DMGTOTAL <- data$PROPDMGEXP * data$PROPDMG + data$CROPDMGEXP * data$CROPDMG
```

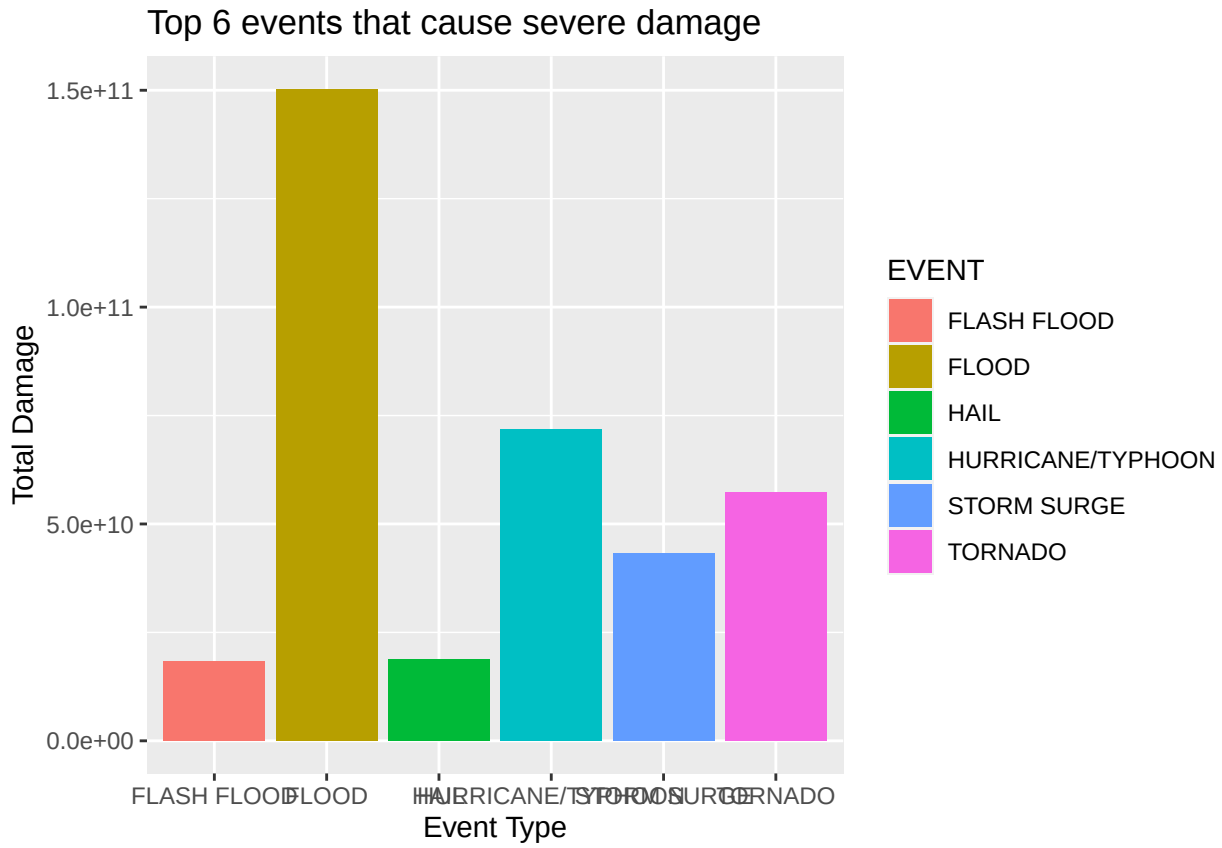
Extract the value (DMGToTAL, EVTYPE)

```
DamageByType  <- aggregate(data$DMGTOTAL, by = list(EVENT= data$EVTYPE), sum)
DamageByType <- DamageByType[order(DamageByType$x, decreasing = TRUE), ]
head(DamageByType)
```

```
##                 EVENT           x
## 170             FLOOD 150319678257
## 411 HURRICANE/TYPHOON  71913712800
## 834           TORNADO  57362333947
## 670       STORM SURGE  43323541000
## 244              HAIL  18761221986
## 153       FLASH FLOOD  18243991079
```

Drawing plot

```
library(ggplot2)
bar2 <- ggplot(DamageByType[1:6, ], aes(EVENT,  x, fill = EVENT))
bar2 + geom_bar(stat = "identity") + labs(x = "Event Type", y = "Total Damage") +  ggtitle("Top 6 events
```




Looking at the graph, we can see that FLOOD is responsible for most of the damage occurred.

## Conclusion

Based on the graph obtained, while FLOOD is responsible for most of the damage, TORNADO is the major cause of human injuries and fatalities.