

Top Storm Event Index

Sachin B.

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities.

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.

This data analysis address the following questions :

1. Across the United States, which types of events are most harmful with respect to population health ?
2. Across the United States, which types of events have the greatest economic consequences ?

This analysis shows by aggregating the data by storm events type :

1. Tornado is the harmful event with respect to population health
2. Flood is the event which have the greatest economic consequences.

1. Data Processing

1.1 Libraries

```
# Loading all required libraries  
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(reshape2)  
library(ggplot2)
```

1.2 Download data from Website

```
# download dataset from website
if(!file.exists("./data")){dir.create("./data")}
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",destfile = "./data/StormData.csv.bz2")
```

1.3 Loading and Preprocessing

```
# Read required column from dataset in dataframe
df = read.csv(file="./data/StormData.csv.bz2", sep=",", header = TRUE)[,c("EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CROPDMG","CROPDMGEXP")]

# Remove incomplete observation
df=df[complete.cases(df), ]

# dataset view
head(df)
```

```
##      EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0        15    25.0           K          0
## 2 TORNADO          0         0     2.5           K          0
## 3 TORNADO          0         2    25.0           K          0
## 4 TORNADO          0         2     2.5           K          0
## 5 TORNADO          0         2     2.5           K          0
## 6 TORNADO          0         6     2.5           K          0
```

```
# datatype of columns
str(df)
```

```
## 'data.frame':    902297 obs. of  7 variables:
## $ EVTYPE      : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
## $ INJURIES   : num  15 0 2 2 2 6 1 0 14 0 ...
## $ PROPDMG    : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
## $ PROPDMGEXP: chr   "K" "K" "K" "K" ...
## $ CROPDMG    : num  0 0 0 0 0 0 0 0 0 0 ...
## $ CROPDMGEXP: chr   "" "" "" "" ...
```

```
# Transforming Event Type as factor
df$EVTYPE <- factor(df$EVTYPE)
```

1.4 Health Impact Analysis

```
# Calculating sum of fatalities and injuries as per Event Type
df_casualties <- df %>%
  group_by(EVTYPE) %>%
  summarize(fatalities = sum(FATALITIES),
            injuries = sum(INJURIES))
```

```

# Calculating mean of fatalities and injuries
df_casualties <- mutate(df_casualties, mean_of_two = (fatalities+injuries)/2)

# selecting top 10 rows depicting maximum damage
df_health <- head(arrange(df_casualties, desc(mean_of_two)), 10)

# for creating 'fatalities' and 'injuries' a categorical variables
df_health_melt <- melt(df_health, id.vars = "EVTYPE", measure.vars = c("fatalities", "injuries"))

```

1.5 Economic Impact Analysis

```

# Selecting rows where damage happened and column required for property damage analysis
df_economy <- df[, c("EVTYPE", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")]

```

```

# character units to number translator table
NumScale <- data.frame(
  Name= c("h", "H", "k", "K", "m", "M", "b", "B"),
  Num = c(10^2, 10^2, 10^3, 10^3, 10^6, 10^6, 10^9, 10^9)
)

```

```

# checking for meaningful labels
unique(df_economy$PROPDMGEXP)

```

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

```

# convert character units to number and calculate property damage
df_economy$PROPMUL <- 10^0

for (i in 1:nrow(NumScale)) {
  df_economy$PROPMUL[df_economy$PROPDMGEXP == NumScale[i, "Name"]] <- NumScale[i, "Num"]
}

df_economy$PROPTOTAL <- df_economy$PROPDMG * df_economy$PROPMUL

```

```

# checking for meaningful labels
unique(df_economy$CROPDMGEXP)

```

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

```

# convert character units to number and calculate property damage
df_economy$CROPDMGEXP <- 10^0

for (i in 1:nrow(NumScale)) {
  df_economy$CROPDMGEXP[df_economy$CROPDMGEXP == NumScale[i, "Name"]] <- NumScale[i, "Num"]
}

df_economy$CROPTOTAL <- df_economy$CROPDMG * df_economy$CROPDMGEXP

```

```

# Calculating sum of property and crop as per Event Type
df_eco_total <- df_economy %>%
  group_by(EVTYPE) %>%
  summarize(property = sum(PROPTOTAL),
             crop = sum(CROPTOTAL)
            )

# selecting top 10 rows depicting maximum damage
df_property <- head(arrange(df_eco_total, desc(property)), 10)
df_crop <- head(arrange(df_eco_total, desc(crop)), 10)

```

2. Result

2.1 Health Impact Analysis

Question 1: Across the United States, which types of events are most harmful with respect to population health ?

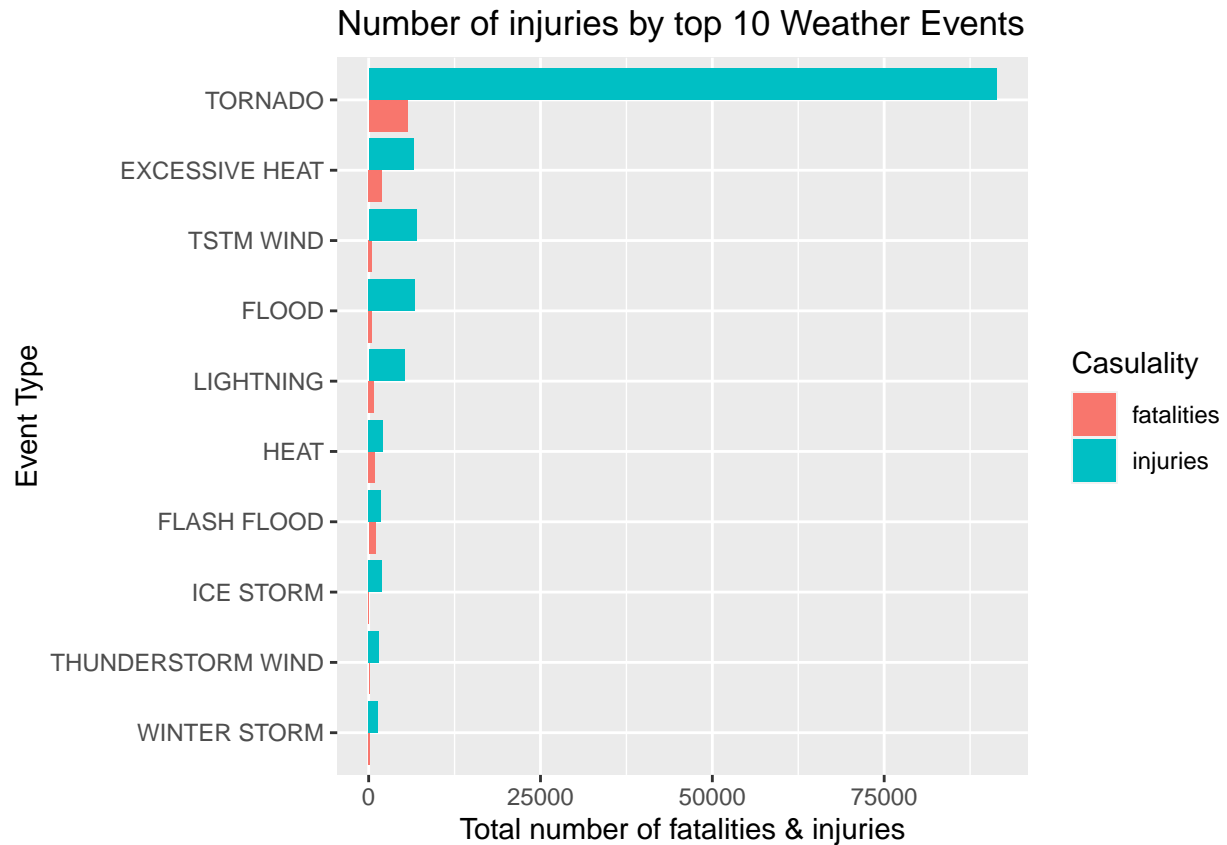
```

# Plot: Number of injuries with the most harmful event type

g <- ggplot(data = df_health_melt, aes(x=reorder(EVTYPE,value), y=value, fill = variable))

g +
  geom_bar(stat = "identity", position = position_dodge()) +
  coord_flip() +
  xlab("Event Type") +
  ylab("Total number of fatalities & injuries") +
  labs(fill = "Casuality") +
  ggtitle("Number of injuries by top 10 Weather Events")

```



Conclusion: The weather event that causes the most harm to public health is Tornadoes. They have shown in the graphs above to be the largest cause of fatalities and injuries due to weather events in the United States.

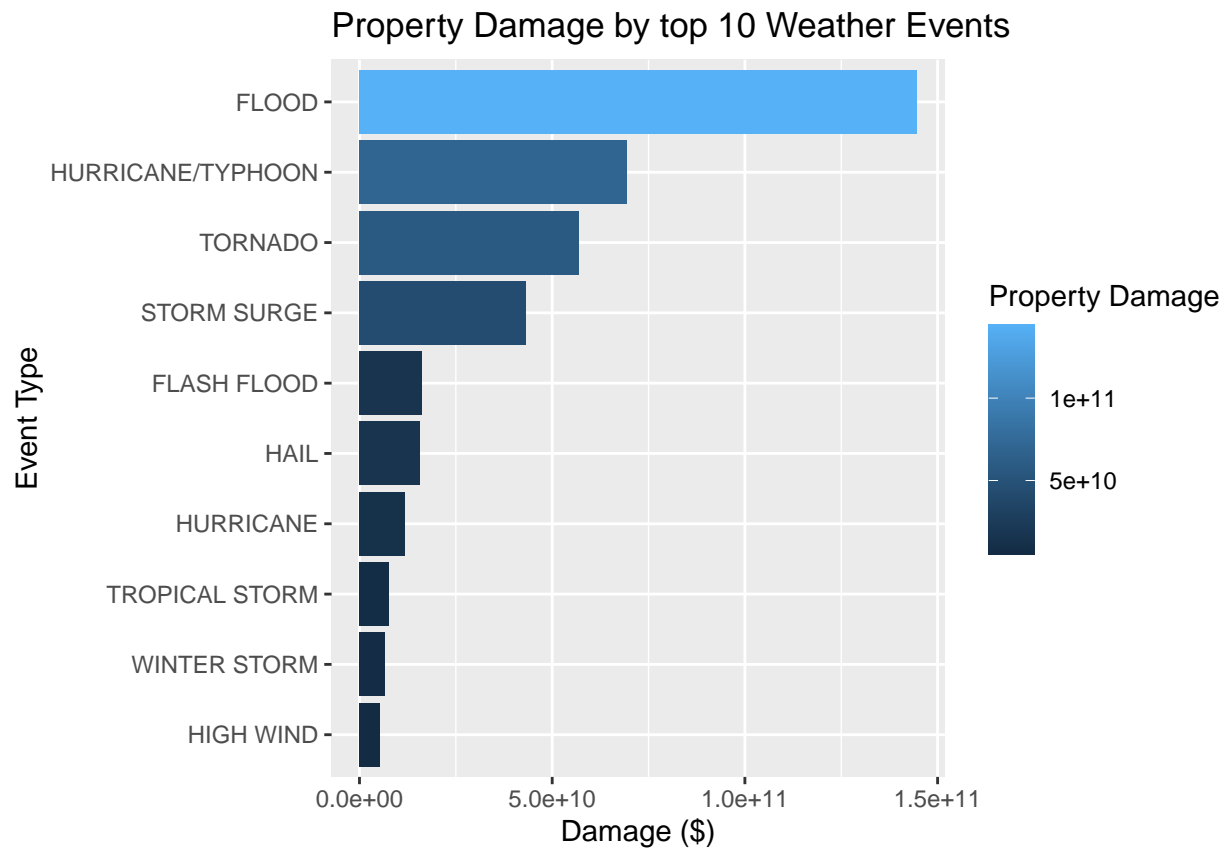
2.2 Economic Impact Analysis

Question 2: Across the United States, which types of events have the greatest economic consequences?

```
# Plot: Number of damages with the most harmful event type

g <- ggplot(data = df_property, aes(x=reorder(EVTYPE,property), y=property, fill=property))

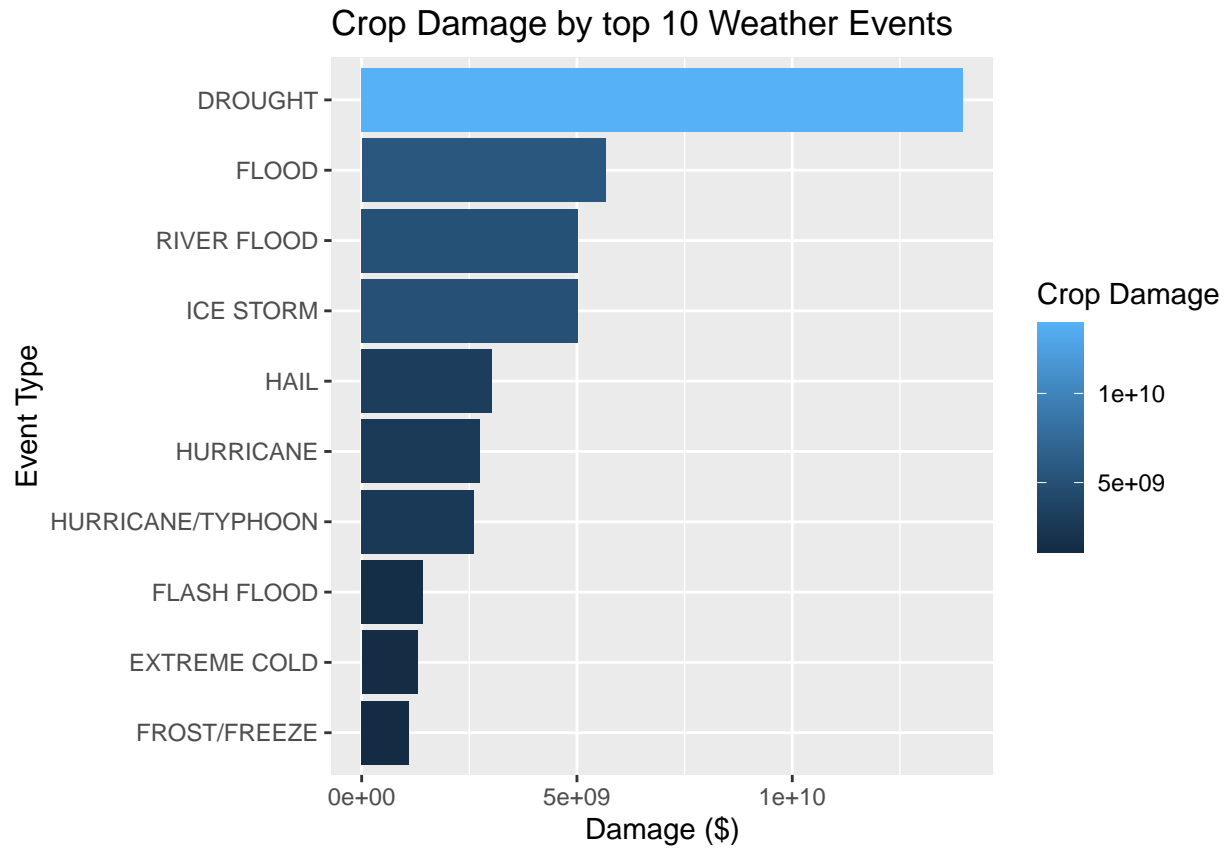
g +
  geom_bar(stat = "identity") +
  coord_flip() +
  xlab("Event Type") +
  ylab("Damage ($)") +
  labs(fill = "Property Damage") +
  ggtitle("Property Damage by top 10 Weather Events")
```



```
# Plot: Number of damages with the most harmful event type

g <- ggplot(data = df_crop, aes(x=reorder(EVTYPE,crop), y=crop, fill=crop))

g +
  geom_bar(stat = "identity") +
  coord_flip() +
  xlab("Event Type") +
  ylab("Damage ($)") +
  labs(fill = "Crop Damage") +
  ggtitle("Crop Damage by top 10 Weather Events")
```



Conclusion: 'Drought' is the event which causes maximum impact on crops but 'Flood' is the event which have the greatest economic consequences altogether.