# StormAnalysis

Robinsson Deantonio

15/5/2020

# US: Impact Of Natural Disasters On The Economy And The Public Health

Natural and anthropic disasters generate damage to the integrity of people and their property, it is important to know the magnitude of the impact they cause in order to make decisions that allow governments and communities to be more prepared to avoid the greatest number of deaths and less damage to the local, regional and national economy. 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.

## Synopsis

### This analysis must 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?

### The analysis revealed that:

1. Tornado is the most hazordous climate event with more than 5600 deaths and 91400 injuries.
2. Floods have caused the most significant economic damage - more than 157 billion USD.

## Basic Settings

```
# import library
library(grid)
require(gridExtra)
library(dplyr)
```

## Data Processing

The analysis was performed on Storm Events Database, provided by National Climatic Data Center. The data is from a comma-separated-value file available here.
There is also some documentation of the data available here.

1. Then, we read the generated csv file

2. The next step is to extract rows corresponding to the event from the documentation, I will also choose the columns which are relevant to our analysis:

- EVTYPE -> Type of event
- FATALITIES -> Number of fatalities
- INJURIES -> Number of injuries
- PROPDMG -> Amount of property damage in orders of magnitude
- PROPDMGEXP -> Order of magnitude for property damage (e.g. K for thousands)
- CROPDMG -> Amount of crop damage in orders of magnitude
- PROPDMGEXP -> Order of magnitude for crop damage (e.g. M for millions)

```
StormData <- select(StormData, EVTYPE, FATALITIES, INJURIES, PROPDMG,
                        PROPDMGEXP, CROPDMG, CROPDMGEXP)
```

3. Extract data corresponding to the events as described in the documentation (Event Types).

```
events <- c("Astronomical Low Tide", "Avalanche|landslide", "Blizzard", "Coastal Flood", "Cold|Wind Chill", "Debris Flow", "Dense Fog|FOG", "Dense Smoke", "Drought", "Dust Devil", "Dust Storm", "Excessive Heat|HIGH TEMPERATURE", "Extreme cold Chill|Extreme WindChill", "Flash Flood", "Flood", "Freezing", "Frost|Freeze|ice", "Funnel Cloud", "Hail", "Heat", "Heavy Rain|RAINFALL", "Heavy Snow|SNOW", "High Surf", "High Wind", "Hurricane|Typhoon", "Ice Storm", "Lakeshore Flood", "Lake-Effect Snow", "Lightning", "Marine Hail|MARINE MISHAP", "Marine High Wind", "Marine Strong Wind", "MARINE TSTM WIND|Marine Thunderstorm Wind|TSTM WIND", "Rip Current", "Seiche", "Sleet", "Storm Tide|TIDE", "Strong Wind|GUSTY WIND|WINDS|MICROBURST WINDS", "Thunderstorm Wind|DRY MICROBURST", "Tornado", "Tropical Depression", "Tropical Storm", "Tsunami", "Volcanic Ash", "Waterspout", "Wildfire|WILD FIRES", "Winter Storm", "Winter Weather")
```

4. Using keywords and including them within the available events, a dataset was generated where all events will be categorized by class

```
options(scipen = 999)  # force fixed notation of numbers instead of scientific
cleandata <- data.frame(EVTYPE = character(0), FATALITIES = numeric(0),
                        INJURIES = numeric(0), PROPDMG = numeric(0),
                        PROPDMGEXP = character(0), CROPDMG = numeric(0),
                        CROPDMGEXP = character(0))
control <- StormData
for (i in 1:length(events)) {
    rows <- control[grep(events[i], ignore.case=T, control$EVTYPE), ]
    control <- control[grep(events[i], ignore.case=T, invert=T, control$EVTYPE), ]
    CLEANNAME <- c(rep(events[i], nrow(rows)))
    rows <- cbind(rows, CLEANNAME)
    cleandata <- rbind(cleandata, rows)
}
```

5. Take into account the order of magnitude of property and crop damage (H = hundreds, K = thousands, M = millions, B= billions)

```
# convert letter exponents to integers
cleandata[cleandata$PROPDMGEXP %in% c("K","k"), ]$PROPDMGEXP <- 3
cleandata[cleandata$PROPDMGEXP %in% c("M","m"), ]$PROPDMGEXP <- 6
cleandata[cleandata$PROPDMGEXP %in% c("B","b"), ]$PROPDMGEXP <- 9

cleandata[cleandata$CROPDMGEXP %in% c("K","k"), ]$CROPDMGEXP <- 3
cleandata[cleandata$CROPDMGEXP %in% c("M","m"), ]$CROPDMGEXP <- 6
cleandata[cleandata$CROPDMGEXP %in% c("B","b"), ]$CROPDMGEXP <- 9
```

6. Generate numerical values based on orders of magnitude

```
suppressWarnings(cleandata$PROPDMG <- cleandata$PROPDMG * 10^as.numeric(cleandata$PROPDMGEXP))
suppressWarnings(cleandata$CROPDMG <- cleandata$CROPDMG * 10^as.numeric(cleandata$CROPDMGEXP))
```

7. Convert NA values to numeric value of 0 to allow additions between columns without error

```
cleandata$PROPDMG[is.na(cleandata$PROPDMG)] <- 0
cleandata$CROPDMG[is.na(cleandata$CROPDMG)] <- 0
```

8. Compute combined economic damage and clean data

```
suppressWarnings(TOTECODMG <- cleandata$PROPDMG + cleandata$CROPDMG)
cleandata <- cbind(cleandata, TOTECODMG)

cleandata <- cleandata[, colnames(cleandata)[c(1,2,3,4,6,8)]]
```

# Results

## Question 01 : Across the United States, which types of events are most harmful with respect to population health?

Fatalities and Injuries

As for the impact on public health, we have got two sorted lists of severe weather events below by the number of people badly affected.

- Aggregate Data for Fatalities

```
fatalities <- aggregate(FATALITIES ~ CLEANNAME, data = cleandata, FUN = sum)
fatalities <- fatalities[order(fatalities$FATALITIES, decreasing = TRUE), ]

MaxFatalities <- fatalities[1:10, ]
print(MaxFatalities)
```

```
##                                              CLEANNAME FATALITIES
## 35                                             Tornado       5636
## 11                      Excessive Heat|HIGH TEMPERATURE       1920
## 19                                                Heat       1212
## 13                                         Flash Flood       1035
## 25                                           Lightning        817
## 29                                         Rip Current        577
## 28 MARINE TSTM WIND|Marine Thunderstorm Wind|TSTM WIND        524
## 14                                               Flood        484
## 5                                      Cold|Wind Chill        451
## 23                                           High Wind        294
```

- Aggregate Data for Injuries

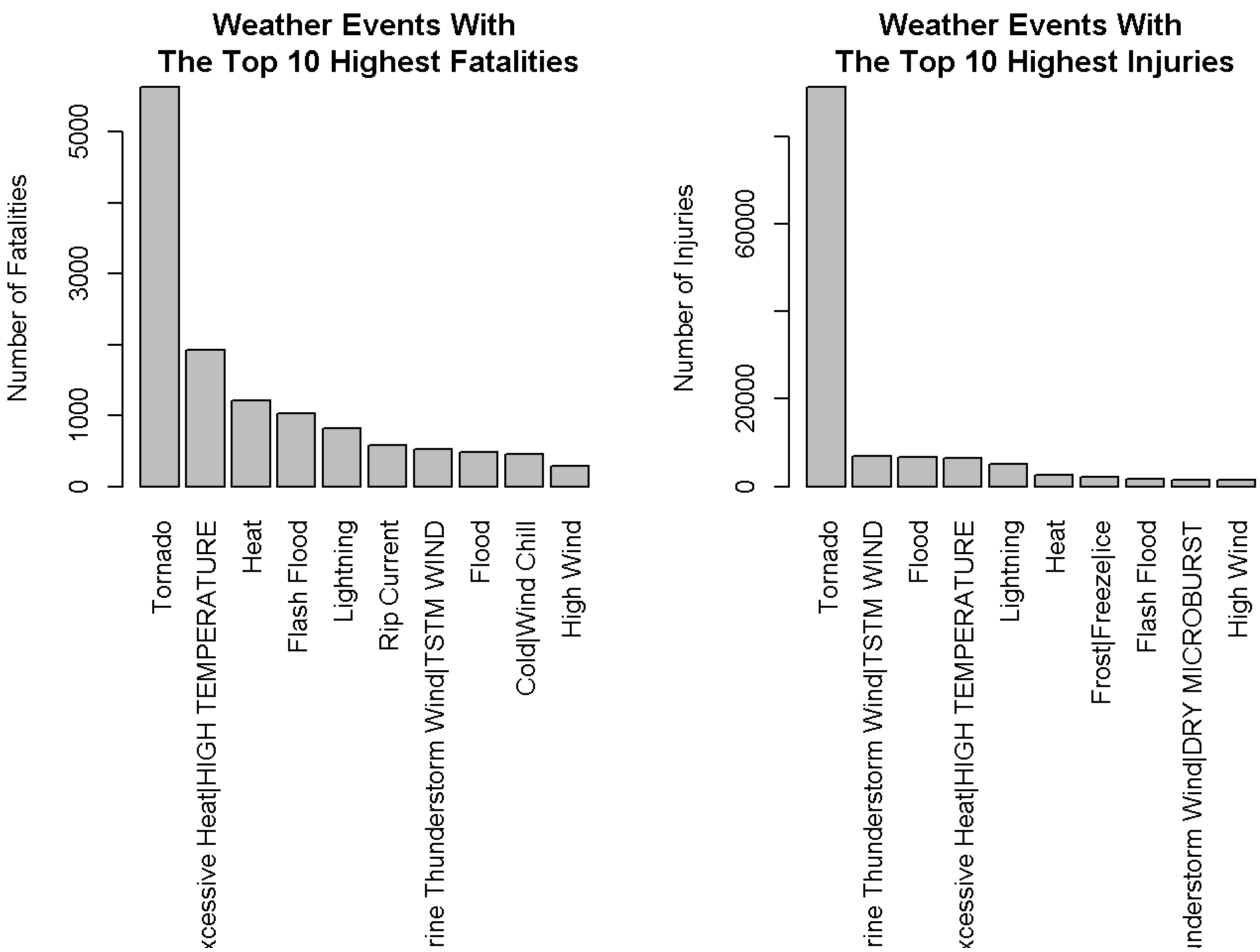
```
injuries <- aggregate(INJURIES ~ CLEANNAME, data = cleandata, FUN = sum)
injuries <- injuries[order(injuries$INJURIES, decreasing = TRUE), ]

MaxInjuries <- injuries[1:10, ]
print(MaxInjuries)
```

```
##                                              CLEANNAME INJURIES
## 35                                             Tornado    91407
## 28 MARINE TSTM WIND|Marine Thunderstorm Wind|TSTM WIND     6996
## 14                                               Flood     6795
## 11                      Excessive Heat|HIGH TEMPERATURE     6525
## 25                                           Lightning     5232
## 19                                                Heat     2684
## 16                                    Frost|Freeze|ice     2167
## 13                                         Flash Flood     1802
## 34                   Thunderstorm Wind|DRY MICROBURST     1516
## 23                                           High Wind     1476
```

And the following is a pair of graphs of Total Fatalities and Total Injuries caused by these Severe Weather Events.

```
par(mfrow = c(1, 2), mar = c(15, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(MaxFatalities$FATALITIES, las = 3, names.arg = MaxFatalities$CLEANNAME, main = "Weather Ev
ents With\n The Top 10 Highest Fatalities", ylab = "Number of Fatalities", col = "grey")
barplot(MaxInjuries$INJURIES, las = 3, names.arg = MaxInjuries$CLEANNAME, main = "Weather Events W
ith\n The Top 10 Highest Injuries", ylab = "Number of Injuries", col = "grey")
```




Based on the above histograms, we find that **Tornado** and **Heat** had caused most fatalities and **Tornado** had caused most injuries in the United States from 1995 to 2011.

## Question 02 : Across the United States, which types of events have the greatest economic consequences?

### Property and Crops combined Economic Damage

As for the impact on economy, we have got two sorted lists below by the amount of money cost by damages.

- Aggregate Data for Property Damage.

```
propdmg <- aggregate(PROPDMG ~ CLEANNAME, data = cleandata, FUN = sum)
propdmg <- propdmg[order(propdmg$PROPDMG, decreasing = TRUE), ]
# 5 most harmful causes of injuries
propdmgMax <- propdmg[1:10, ]
print(propdmgMax)
```

```
##                CLEANNAME       PROPDMG
## 14                 Flood 150205807650
## 24     Hurricane|Typhoon   85256410010
## 35               Tornado   57003317814
## 18                  Hail   17622990956
```

```
## 13          Flash Flood  17588740879
## 37       Tropical Storm   7714390550
## 42         Winter Storm   6688997251
## 23            High Wind   6041395000
## 41 Wildfire|WILD FIRES   5489714000
## 32      Storm Tide|TIDE   4650613150
```

- Aggregate Data for Crop Damage

```
cropdmg <- aggregate(CROPDMG ~ CLEANNAME, data = cleandata, FUN = sum)
cropdmg <- cropdmg[order(cropdmg$CROPDMG, decreasing = TRUE), ]
# 5 most harmful causes of injuries
cropdmgMax <- cropdmg[1:10, ]
print(cropdmgMax)
```

```
##                CLEANNAME     CROPDMG
## 8                Drought 13972621780
## 14                 Flood 10847855950
## 16     Frost|Freeze|ice  7019175300
## 24    Hurricane|Typhoon  5506117800
## 18                  Hail  3114212870
## 13           Flash Flood  1532197150
## 5        Cold|Wind Chill  1416765550
## 20 Heavy Rain|RAINFALL   795409800
## 37       Tropical Storm   694896000
## 23            High Wind   694291900
```

- Aggregate Total Economic Damage

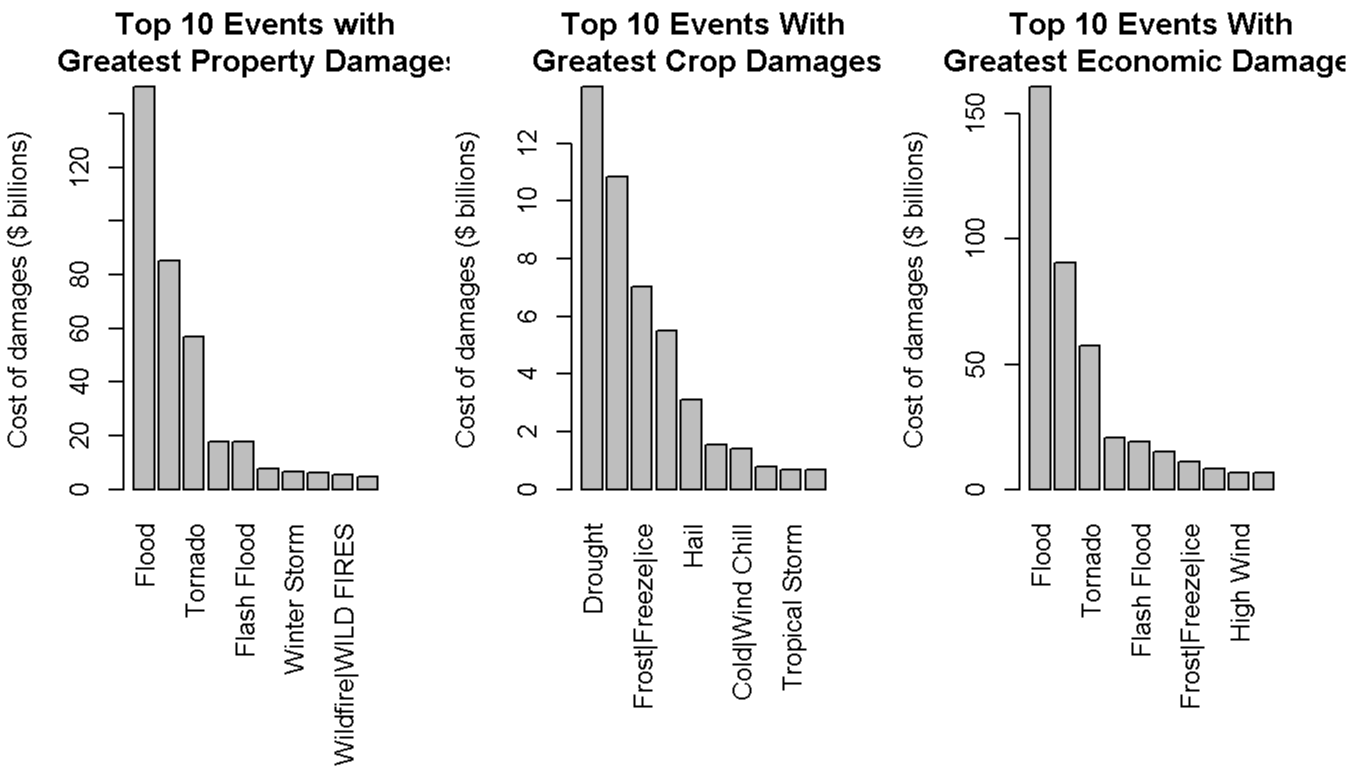
```
ecodmg <- aggregate(TOTECODMG ~ CLEANNAME, data = cleandata, FUN = sum)
ecodmg <- ecodmg[order(ecodmg$TOTECODMG, decreasing = TRUE), ]
# 5 most harmful causes of property damage
ecodmgMax <- ecodmg[1:10, ]
print(ecodmgMax)
```

```
##              CLEANNAME    TOTECODMG
## 14               Flood 161053663600
## 24 Hurricane|Typhoon  90762527810
## 35             Tornado  57418279284
## 18                Hail  20737203826
## 13         Flash Flood  19120938029
## 8              Drought  15018927780
## 16   Frost|Freeze|ice  11006669710
## 37     Tropical Storm   8409286550
## 23          High Wind   6735686900
## 42       Winter Storm   6716441251
```

And the following are graphs of Total Property Damages, Total Crop Damages and Total Economic Damages caused by these Severe Weather Events.

```
par(mfrow = c(1, 3), mar = c(15, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(propdmgMax$PROPDMG/(10^9), las = 3, names.arg = propdmgMax$CLEANNAME, main = "Top 10 Event
s with\n Greatest Property Damages", ylab = "Cost of damages ($ billions)", col = "grey")
barplot(cropdmgMax$CROPDMG/(10^9), las = 3, names.arg = cropdmgMax$CLEANNAME, main = "Top 10 Event
s With\n Greatest Crop Damages", ylab = "Cost of damages ($ billions)", col = "grey")
```

```
barplot(ecodmgMax$TOTECODMG/(10^9), las = 3, names.arg = ecodmgMax$CLEANNAME, main = "Top 10 Events With\n Greatest Economic Damages", ylab = "Cost of damages ($ billions)", col = "grey")
```




The weather events have the Greatest Economic Consequences are: **Flood**, **Drought**, **Tornado** and **Typhoon**.
Across the United States, **Flood**, **Tornado** and **Typhoon** have caused the Greatest Damage to Properties.
**Drought** and **Flood** had been the causes for the Greatest Damage to Crops.

# Conclusion

From these data, we found that **Excessive Heat** and **Tornado** are most harmful with respect to Population Health, while **Flood** and **Hurricane/Typhoon** have the greatest Economic Consequences.