Analysis of Severe Weather Events on USA based on NOAA Data

Diptam Dutta

November 17, 2017

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.

Source Data

- The storm data can be downloaded from this link
- Documentation on the data can be found from this link
- National Climatic Data Center Storm Events FAQ

Data Processing

To process the data we need to download the data from the above mentioned link. let's start by downloading the data

Download Data

```
if (!"StormData.csv.bz2" %in% dir("./")) {
        download.file(
        "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
        destfile = "StormData.csv.bz2"
        )
}
```

Now Let's read the downloaded data

```
if (!"stormdata" %in% ls()) {
    stormdata <- read.csv(bzfile("StormData.csv.bz2"), sep = ",", header = TRUE, stringsAsFactors = FALS
}
dim(stormdata)
```

```
## [1] 902297     37
```

Extact Storm Event Type

```
Stormevents <-
        c(
        "Astronomical Low Tide",
        "Avalanche",
        "Blizzard",
        "Coastal Flood",
        "Cold/Wind Chill",
        "Debris Flow",
        "Dense Fog",
        "Dense Smoke",
        "Drought",
        "Dust Devil",
        "Dust Storm",
        "Excessive Heat",
        "Extreme cold/Wind Chill",
        "Flash Flood",
        "Flood",
        "Freezing",
        "Frost/Freeze",
        "Funnel Cloud",
        "Hail",
        "Heat",
        "Heavy Rain",
        "Heavy Snow",
        "High Surf",
        "High Wind",
        "Hurricane/Typhoon",
        "Ice Storm",
        "Lakeshore Flood",
        "Lake-Effect Snow",
        "Lightning",
        "Marine Hail",
        "Marine High Wind",
        "Marine Strong Wind",
        "Marine Thunderstorm Wind",
        "Rip Current",
        "Seiche",
        "Sleet",
        "Storm Tide",
        "Strong Wind",
        "Thunderstorm Wind",
        "Tornado",
        "Tropical Depression",
        "Tropical Storm",
        "Tsunami",
        "Volcanic Ash",
        "Waterspout",
        "Wildfire",
        "Winter Storm",
        "Winter Weather"
        )
```

Some events are having charecters ie. '\', so we will use regex to identify that

```
Stormevents_regex <-
        c(
        "Astronomical Low Tide|Low Tide",
        "Avalanche",
        "Blizzard",
        "Coastal Flood",
        "Cold/Wind Chill",
        "Debris Flow",
        "Dense Fog",
        "Dense Smoke",
        "Drought",
        "Dust Devil",
        "Dust Storm",
        "Excessive Heat",
        "Extreme cold/Wind Chill|Extreme Cold|Wind Chill",
        "Flash Flood",
        "Flood",
        "Freezing",
        "Frost/Freeze|Frost|Freeze",
        "Funnel Cloud",
        "Hail",
        "Heat",
        "Heavy Rain",
        "Heavy Snow",
        "High Surf",
        "High Wind",
        "Hurricane/Typhoon|Hurricane|Typhoon",
        "Ice Storm",
        "Lakeshore Flood",
        "Lake-Effect Snow",
        "Lightning",
        "Marine Hail",
        "Marine High Wind",
        "Marine Strong Wind",
        "Marine Thunderstorm Wind|Marine tstm Wind",
        "Rip Current",
        "Seiche",
        "Sleet",
        "Storm Tide",
        "Strong Wind",
        "Thunderstorm Wind|tstm wind",
        "Tornado",
        "Tropical Depression",
        "Tropical Storm",
        "Tsunami",
        "Volcanic Ash",
        "Waterspout",
        "Wildfire",
        "Winter Storm",
        "Winter Weather"
        )
```

We keep only the fields needed to our analysis:

- EVTYPE: the type of event
- FATALITIES: number of fatalities
- INJURIES: number of injuries
- PROPDMG: damage to properties in USD
- PROPDMGEXP: magnitude for properties damage (K for thousands, M for millions, B for billions)
- CROPDMG: damage to crops in USD
- CROPDMGEXP: magnitude for crops damage (K for thousands, M for millions, B for billions)

```
options(scipen = 999)
cleandata <- data.frame(EVTYPE = character(0), FATALITIES = numeric(0), INJURIES = numeric(0), PROPDMG =
for (i in 1:length(Stormevents)) {
        rows <-
                stormdata[grep(Stormevents_regex[i], ignore.case = TRUE, stormdata$EVTYPE),]
        rows <-
                rows[, c(
                        "EVTYPE",
                        "FATALITIES",
                        "INJURIES",
                        "PROPDMG",
                        "PROPDMGEXP",
                        "CROPDMG",
                        "CROPDMGEXP"
                )]
        CLEANNAME <- c(rep(Stormevents[i], nrow(rows)))
        rows <- cbind(rows, CLEANNAME)
        cleandata <- rbind(cleandata, rows)
}
```

Adjusting the valid values are K for thousands(3 exp), M for millions(6 exp) and B for billions(9 exp)

```
cleandata[(cleandata$PROPDMGEXP == "K" | cleandata$PROPDMGEXP == "k"), ]$PROPDMGEXP <- 3
cleandata[(cleandata$PROPDMGEXP == "M" | cleandata$PROPDMGEXP == "m"), ]$PROPDMGEXP <- 6
cleandata[(cleandata$PROPDMGEXP == "B" | cleandata$PROPDMGEXP == "b"), ]$PROPDMGEXP <- 9
cleandata[(cleandata$CROPDMGEXP == "K" | cleandata$CROPDMGEXP == "k"), ]$CROPDMGEXP <- 3
cleandata[(cleandata$CROPDMGEXP == "M" | cleandata$CROPDMGEXP == "m"), ]$CROPDMGEXP <- 6
cleandata[(cleandata$CROPDMGEXP == "B" | cleandata$CROPDMGEXP == "b"), ]$CROPDMGEXP <- 9
```

Calculate the properties and crop damages

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

```
## Warning: NAs introduced by coercion
```

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

```
## Warning: NAs introduced by coercion
```

Calculate the total damages

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

Thus concludes the initial prepearation of data.

Q.1: Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

Calulating the number of FATALITIES

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

Top 10 Events causing the most FATALITIES

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

```
##                  CLEANNAME FATALITIES
## 38                 Tornado       5661
## 19                    Heat       3138
## 11          Excessive Heat       1922
## 14                   Flood       1525
## 13             Flash Flood       1035
## 28               Lightning        817
## 37       Thunderstorm Wind        753
## 33             Rip Current        577
## 12 Extreme cold/Wind Chill        382
## 23               High Wind        299
```

Calulating the number of INJURIES

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

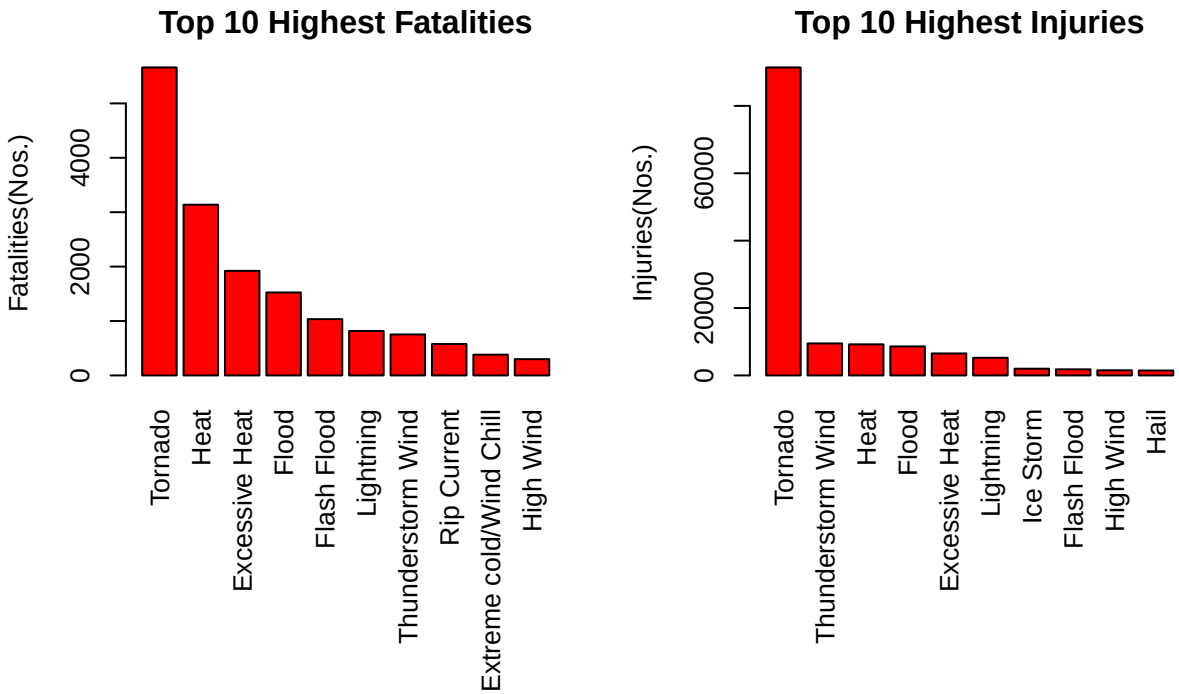
Top 10 Events causing the most INJURIES

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

```
##            CLEANNAME INJURIES
## 38           Tornado    91407
## 37 Thunderstorm Wind     9493
## 19              Heat     9224
## 14             Flood     8604
## 11    Excessive Heat     6525
## 28         Lightning     5232
## 25         Ice Storm     1992
## 13       Flash Flood     1802
## 23         High Wind     1523
## 18              Hail     1467
```

Plotting the above information as a pair of graphs of total FATALITIES & INJURIES

```
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 = "Top 10 Highest Fatalities",
        ylab = "Fatalities(Nos.)",
        col = "RED"
)
barplot(
        MaxInjuries$INJURIES,
        las = 3,
        names.arg = MaxInjuries$CLEANNAME,
        main = "Top 10 Highest Injuries",
        ylab = "Injuries(Nos.)",
        col = "RED"
)
```



ANS for Qus 1

As you can see with respect to population health

*** Fatalities : 1. Tornado & 2. Heat**

*** Injuries : Tornado**

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

Calculating the total property damage

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

Top 10 Events causing the most property damage

```
propdmgMax <- propdmg[1:10, ]
print(propdmgMax)
```

```
##            CLEANNAME      PROPDMG
## 14             Flood 168212215589
## 24 Hurricane/Typhoon  85356410010
## 38           Tornado  58603317864
## 18              Hail  17622990956
## 13       Flash Flood  17588791879
## 37 Thunderstorm Wind  11575228673
## 40    Tropical Storm   7714390550
## 45      Winter Storm   6749997251
## 23         High Wind   6166300000
## 44          Wildfire   4865614000
```

Calculating the total crop damage

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

Top 10 Events causing the most crop damage

```
cropdmgMax <- cropdmg[1:10, ]
print(cropdmgMax)
```

```
##                 CLEANNAME     CROPDMG
## 8                 Drought 13972621780
## 14                  Flood 12380109100
## 24       Hurricane/Typhoon  5516117800
## 25               Ice Storm  5022113500
## 18                    Hail  3114212870
## 16           Frost/Freeze  1997061000
## 13            Flash Flood  1532197150
```

```
## 12 Extreme cold/Wind Chill  1313623000
## 37       Thunderstorm Wind  1255947980
## 19                    Heat   904469280
```

Calculating the total economic damage

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

Top 10 Events causing the most economic damage

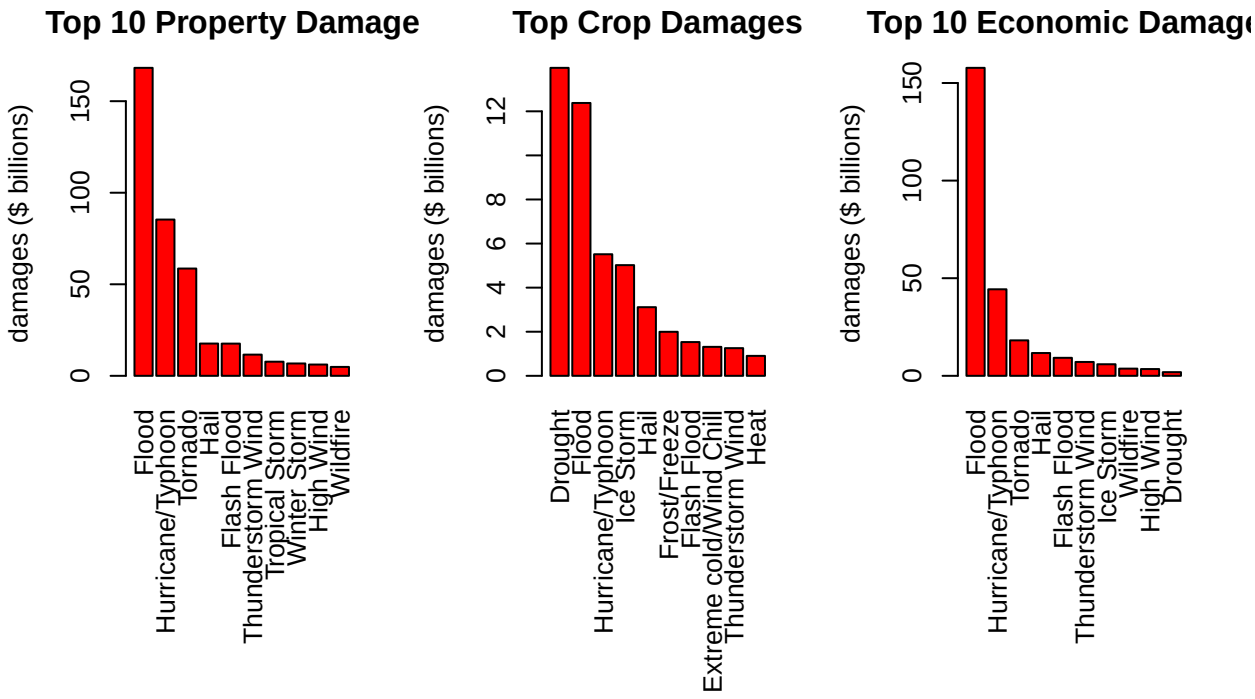
```
ecodmgMax <- ecodmg[1:10, ]
print(ecodmgMax)
```

```
##            CLEANNAME    TOTECODMG
## 14             Flood 157764680787
## 24 Hurricane/Typhoon  44330000800
## 38           Tornado  18172843863
## 18              Hail  11681050140
## 13       Flash Flood   9224527227
## 37 Thunderstorm Wind   7098296330
## 25         Ice Storm   5925150850
## 44          Wildfire   3685468370
## 23         High Wind   3472442200
## 8            Drought   1886667000
```

Plotting the above information as a pair of graphs of total property, crop & economic damage

```
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 Property Damages",
        ylab = "damages ($ billions)",
        col = "RED"
)
barplot(
        cropdmgMax$CROPDMG / (10 ^ 9),
        las = 3,
        names.arg = cropdmgMax$CLEANNAME,
        main = "Top Crop Damages",
        ylab = "damages ($ billions)",
        col = "RED"
)
barplot(
        ecodmgMax$TOTECODMG / (10 ^ 9),
```

```
        las = 3,
        names.arg = ecodmgMax$CLEANNAME,
        main = "Top 10 Economic Damages",
        ylab = "damages ($ billions)",
        col = "RED"
)
```



ANS for Qus 2

As you can see with respect to economic consequences causing damage

*** Property : Flood**

*** Crop : Drought and Flood**

*** Economic : Flood followed by Hurricane/Typhoon**

Conclusion

From the above analysis we can see that Tornado and Heat are most harmful with respect to Population Health, while Flood, Drought and Hurricane/Typhoon have the greatest Economic Consequences.

Project Codes have been uploaded to Github