Analysis of Effects of Severe Weather Events on Population Health and Economic Consequences using NOAA Storm Database

Jonathan Lok-Chuen Lo

October 24, 2015

Synopsis

This project aims to investigate which type of severe weather events are the most harmful to public health and causing the most economic damages, by 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 data was collected during the period from 1950 and November 2011.

Using this dataset, tornado is found to be most harmful to public health, with the highest number of both injuries and deaths. In terms of economic damages, flooding caused the most damage in property and crop combined. However, for crop only, drought is causing the most damages, while flooding is still the top event for property damage.

Data Processing

Below are the packages required for the data processing.

```
library(dplyr)
library(reshape2)
library(ggplot2)
```

Download and Load the Raw Data

The raw data is stored in the course website here. Date come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size.

Download the data file to the local drive and load the data into object `rawData`.

```
file.url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

if (!file.exists("./repdata-data-StormData.csv.bz2")){
    download.file(
        file.url,
        destfile = "./repdata-data-StormData.csv.bz2")
}

rawData <- read.csv(bzfile("./repdata-data-StormData.csv.bz2"))
```

Clean up the Raw Data

Units of the property and crop damages are encoded either 1) using alphabetical characters to signify magnitude, e.g. "H" for hundreds, "K" for thousands, "M" for millions, and "B" for billions or 2) using numbers to indicate the Significant digits to be added.

Create variables `PropertyDamage` and `CropDamage` by converting the damages according to these rules above. In addition, covert eventy types `EVTYPE` to upper case for consistent formating.

```
convertSize <- function(size, unit) {
    unit <- as.character(unit)
    size * switch(
        unit,
        '0' = 1,
        '1' = 10,
        '2' = 100,
        '3' = 1000,
        '4' = 10000,
        '5' = 100000,
        '6' = 1000000,
        '7' = 10000000,
        '8' = 100000000,
        'h' = , 'H' = 100,
        'k' = , 'K' = 1000,
        'm' = , 'M' = 1000000,
        'b' = , 'B' = 1000000000,
        0)
}

rawData <- rawData %>%
    mutate(EVTYPE = factor(toupper(EVTYPE)),
           PropertyDamage = mapply(convertSize, PROPDMG, PROPDMGEXP),
           CropDamage = mapply(convertSize, CROPDMG, CROPDMGEXP))
```

Process the Data

For each severe weather event types, compute the total fatalities, injuries, property damages and crop damages.

```
byEventData <- rawData %>%
    group_by(EVTYPE) %>%
    summarise(
        TotalFatalities = sum(FATALITIES) / 1000,
        TotalInjuries = sum(INJURIES) / 1000,
        TotalHealthImpact = TotalFatalities + TotalInjuries,
        TotalPropertyDamage = sum(PropertyDamage) / 1e9,
        TotalCropDamage = sum(CropDamage) / 1e9,
        TotalEconomicDamage = TotalPropertyDamage + TotalCropDamage)
```

To assess the health impact, retrieve the top 10 events causing the highest number of total fatalities and injuries. The data frame is melted to a long form for plotting purposes.

```
healthDamageData <- byEventData %>%
    arrange(desc(TotalHealthImpact)) %>%
    slice(1:10) %>%
    select(
        EVTYPE,
        Injuries = TotalInjuries,
        Fatalities = TotalFatalities) %>%
    mutate(EVTYPE = factor(EVTYPE, as.character(EVTYPE)))

# Convert the factor EVTYPE to include only those contributed to the top 10
healthDamageData$EVTYPE = factor(
    healthDamageData$EVTYPE,
    levels = unique(healthDamageData$EVTYPE))

healthDamageData <- melt(
    healthDamageData,
    id = "EVTYPE",
    variable.name = "HealthEffect",
    value.name = "Count")
```

Similarly, to assess the economic impact, retrieve the top 10 events causing the highest cost from total property and crop damages.

```
economicDamageData <- byEventData %>%
    arrange(desc(TotalEconomicDamage)) %>%
    slice(1:10) %>%
    select(
        EVTYPE,
        PropertyDamage = TotalPropertyDamage,
        CropDamage = TotalCropDamage) %>%
    mutate(EVTYPE = as.character(EVTYPE))

# Convert the factor EVTYPE to include only those contributed to the top 10
economicDamageData$EVTYPE = factor(
    economicDamageData$EVTYPE,
    levels = unique(economicDamageData$EVTYPE))

economicDamageData <- melt(
    economicDamageData,
    id = "EVTYPE",
    variable.name = "EconomicEffect",
    value.name = "Count")
```

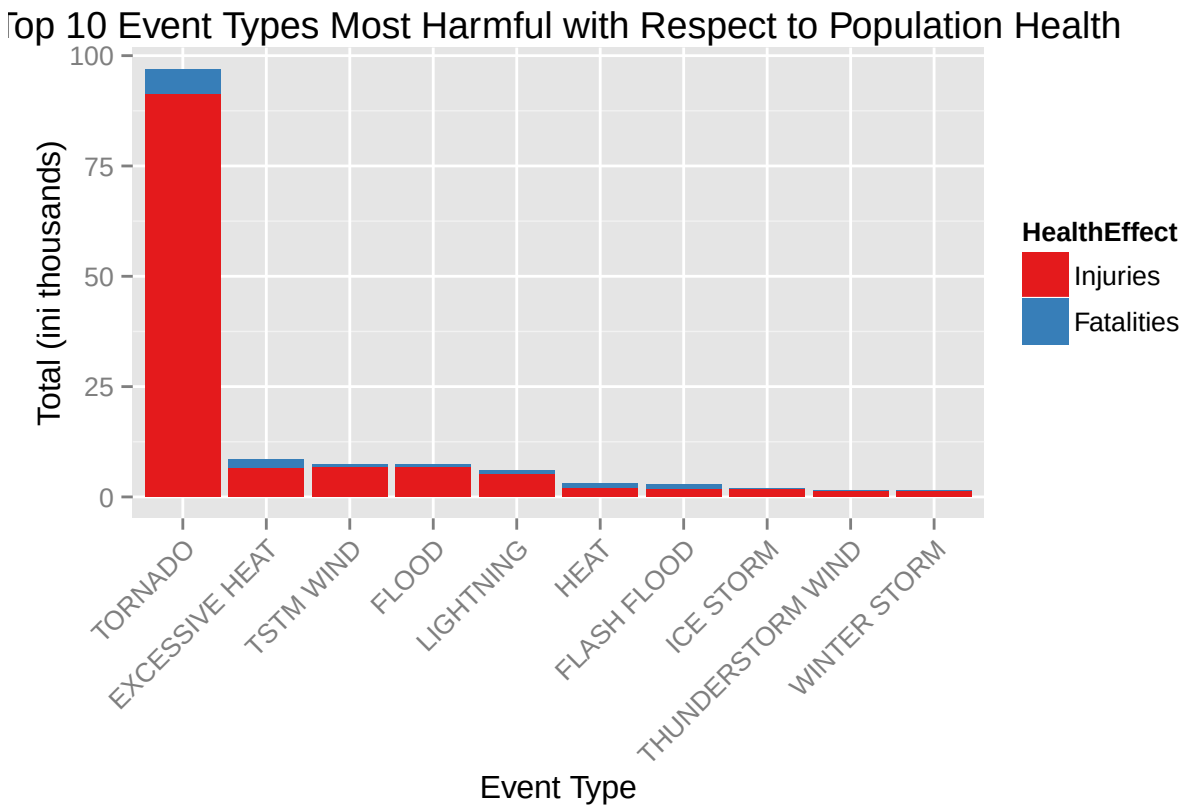
Results

Across the United States, which types of events are most harmful with respect to population health? Below is a plot showing the top 10 event types most harmful with respect to population health. Tornado is causing the most combined injuries and fatalities.

```
g <- ggplot(healthDamageData, aes(x = EVTYPE, y = Count, fill = HealthEffect))

g + geom_bar(stat = "identity") +
```

```
    labs(
        x = "Event Type",
        y = "Total (ini thousands)",
        title = "Top 10 Event Types Most Harmful with Respect to Population Health") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    scale_fill_brewer(palette = "Set1")
```



We also look at which events is causing the most injuries and fatalities separately.

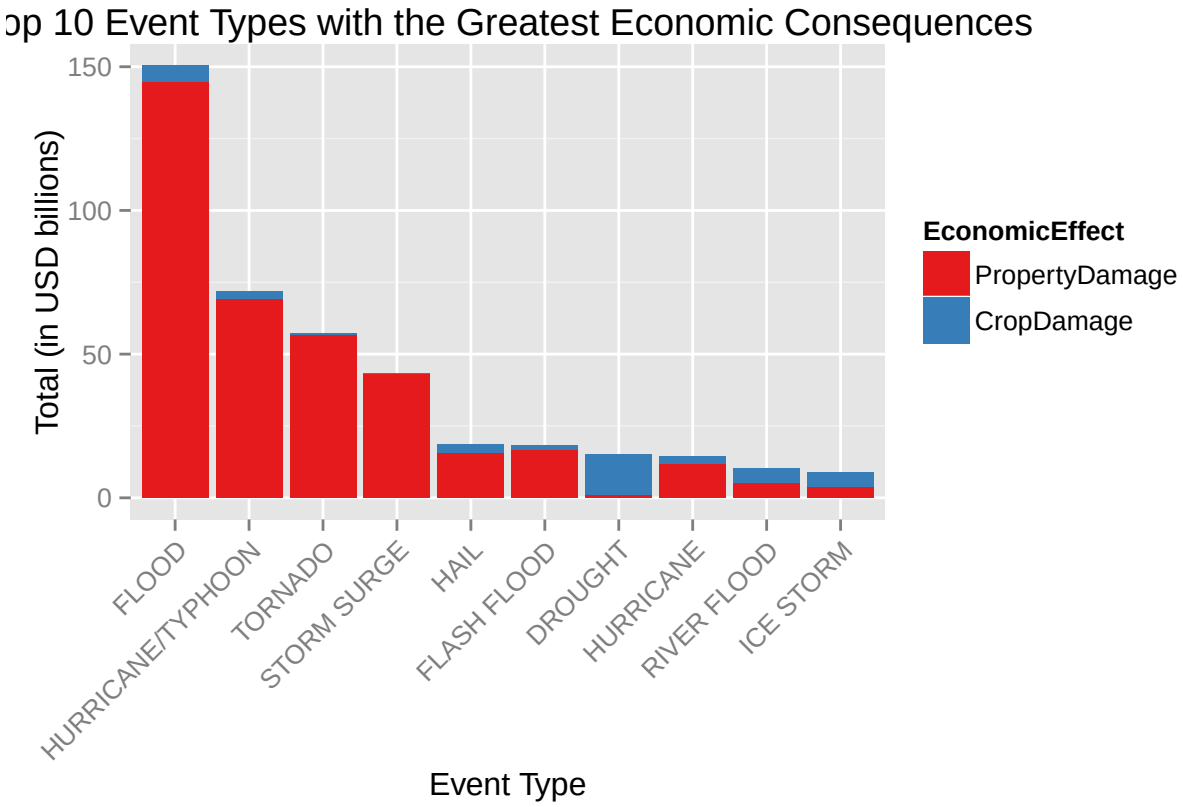
```
rbind(
    Most_Fatalities = as.character(
        byEventData$EVTYPE[which.max(byEventData$TotalFatalities)]),
    Most_Injuries = as.character(
        byEventData$EVTYPE[which.max(byEventData$TotalInjuries)])
)
```

```
##                 [,1]
## Most_Fatalities "TORNADO"
## Most_Injuries   "TORNADO"
```

Across the United States, which types of events have the greatest economic consequences? Below is a plot showing the top 10 event types having the greatest economic consequences. Flood is causing the largest damange for property and crop combined.

```r
g <- ggplot(economicDamageData, aes(x = EVTYPE, y = Count, fill = EconomicEffect))

g + geom_bar(stat = "identity") +
    labs(
        x = "Event Type",
        y = "Total (in USD billions)",
        title = "Top 10 Event Types with the Greatest Economic Consequences") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    scale_fill_brewer(palette = "Set1")
```



Again, We look at which events is causing the most property and crop damages separately.

```r
rbind(
    Most_Property_Damage = as.character(
        byEventData$EVTYPE[which.max(byEventData$TotalPropertyDamage)]),
    Most_Crop_Damage = as.character(
        byEventData$EVTYPE[which.max(byEventData$TotalCropDamage)])
)
```

```
##                       [,1]
## Most_Property_Damage "FLOOD"
## Most_Crop_Damage     "DROUGHT"
```