Health and economy impact of Severe Weather Events: An outlook of the most harmful event type in the US between 1950 to 2011

Ivan Romero

20/6/2020

1. Introduction

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 assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events: The data analysis must address the following questions:

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

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

2. Synopsis

The present document reports the exploratory analysis of the NOAA Storm Database. The analysis is focused on the health and economic impacts of severe weather events across the United States between 1950 and 2011, establishing the weather event type that have the highest impact on several indicators such as fatalities, injuries, property damage costs and crop damage costs. Tornados are the most harmful weather events according to the health related event index, followed by Excessive Heat, TSTM Wind, Floods, and Lighting. Floods are the Weather events with the highest economic impact estimated as the total damage from property and crops, followed by Hurricane/Typhoons, Tornado, Storm Surge and Hails.

3. Data processing

Raw data for this report can be downloaded from the course website: Storm Database. Supplementary information about the assembly of the Storm Database can be looked up on Storm Data Documentation. In a preparatory step previous to getting adn cleaning data for the analysis, folder directories for data, figures and R code are created in the following chunk of code:

```r
## Creating directories inside the project folder
if(!dir.exists("data")) {dir.create("data")}
if(!dir.exists("figures")) {dir.create("figures")}
if(!dir.exists("Rcode")) {dir.create("Rcode")}
```

As previously stated, raw data can be obtained from the course website. The .csv file is compressed as a .bz2 file that can be extracted by using the **bunzip.2** function from the **R.utils** package. This process is executed in the following lines of code, including the reading and storing of the raw data file into the "storm_rawdata" dataframe. In order to get a sense about the variables incorporated into this data frame, a summary of the data frame is presented

```r
## Download file
rdatafilename <- "stormdata.zip"
rdatafileloc <- paste0("./",rdatafilename)
if(!file.exists(rdatafileloc)) {
        fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
        download.file(fileUrl, rdatafileloc, method = "curl")
}

## Decompressing the zip file and verifying
if(!file.exists("./data/storm.csv")) {
        library(R.utils)
        bunzip2(rdatafileloc, "./data/storm.csv", remove = TRUE)
}

## Reading the data
storm_rawdata <- read.csv("./data/storm.csv")
summary(storm_rawdata)
```

```
##     STATE__       BGN_DATE           BGN_TIME          TIME_ZONE
##  Min.   : 1.0   Length:902297      Length:902297      Length:902297
##  1st Qu.:19.0   Class :character   Class :character   Class :character
##  Median :30.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :31.2
##  3rd Qu.:45.0
##  Max.   :95.0
##
##      COUNTY       COUNTYNAME           STATE              EVTYPE
##  Min.   :  0.0   Length:902297      Length:902297      Length:902297
##  1st Qu.: 31.0   Class :character   Class :character   Class :character
##  Median : 75.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :100.6
##  3rd Qu.:131.0
##  Max.   :873.0
##
##    BGN_RANGE          BGN_AZI           BGN_LOCATI          END_DATE
##  Min.   :   0.000   Length:902297      Length:902297      Length:902297
##  1st Qu.:   0.000   Class :character   Class :character   Class :character
##  Median :   0.000   Mode  :character   Mode  :character   Mode  :character
##  Mean   :   1.484
##  3rd Qu.:   1.000
##  Max.   :3749.000
##
##    END_TIME           COUNTY_END COUNTYENDN       END_RANGE
##  Length:902297      Min.   :0    Mode:logical   Min.   :  0.0000
##  Class :character   1st Qu.:0    NA's:902297    1st Qu.:  0.0000
##  Mode  :character   Median :0                   Median :  0.0000
##                     Mean   :0                   Mean   :  0.9862
##                     3rd Qu.:0                   3rd Qu.:  0.0000
```

```
##                    Max.   :0                   Max.   :925.0000
##
##    END_AZI           END_LOCATI           LENGTH              WIDTH
##  Length:902297      Length:902297      Min.   :   0.0000   Min.   :   0.000
##  Class :character   Class :character   1st Qu.:   0.0000   1st Qu.:   0.000
##  Mode  :character   Mode  :character   Median :   0.0000   Median :   0.000
##                                        Mean   :   0.2301   Mean   :   7.503
##                                        3rd Qu.:   0.0000   3rd Qu.:   0.000
##                                        Max.   :2315.0000   Max.   :4400.000
##
##        F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG         PROPDMGEXP            CROPDMG         CROPDMGEXP
##  Min.   :   0.00   Length:902297      Min.   :  0.000   Length:902297
##  1st Qu.:   0.00   Class :character   1st Qu.:  0.000   Class :character
##  Median :   0.00   Mode  :character   Median :  0.000   Mode  :character
##  Mean   :  12.06                      Mean   :  1.527
##  3rd Qu.:   0.50                      3rd Qu.:  0.000
##  Max.   :5000.00                      Max.   :990.000
##
##      WFO             STATEOFFIC          ZONENAMES            LATITUDE
##  Length:902297      Length:902297      Length:902297      Min.   :   0
##  Class :character   Class :character   Class :character   1st Qu.:2802
##  Mode  :character   Mode  :character   Mode  :character   Median :3540
##                                                           Mean   :2875
##                                                           3rd Qu.:4019
##                                                           Max.   :9706
##                                                           NA's   :47
##    LONGITUDE        LATITUDE_E     LONGITUDE_        REMARKS
##  Min.   :-14451   Min.   :   0   Min.   :-14455   Length:902297
##  1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0   Class :character
##  Median :  8707   Median :   0   Median :     0   Mode  :character
##  Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   : 17124   Max.   :9706   Max.   :106220
##                   NA's   :40
##      REFNUM
##  Min.   :     1
##  1st Qu.:225575
##  Median :451149
##  Mean   :451149
##  3rd Qu.:676723
##  Max.   :902297
##
```

As expressed in the introductory lines of this report, the aim of the analysis is to explore the health and economic effects of severe weather events. Thus, only variables such as *FATALITIES*, and *INJURIES* for health impact; and *PROPDMG*, and *CROPDMG* will be considered. It is worth mentioning that *PROPDMGEXP* and *CROPDMGEXP* economic will also be considered for economic impact evaluation, because this variable represents the exponent of the Property and Crop damage value estimation. These variables are alphabetical characters used to signify damage magnitude, expressed as following: "K" for thousands, "M" for millions, and "B" for billions. The extraction process for these variables, and the specific values for exponent are shown in the next chunk of code as a data cleaning stage previous to the initial analysis:

```
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
```

```
storm_rawdata <- storm_rawdata %>% select(STATE, BGN_DATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDM
storm_rawdata <- storm_rawdata[storm_rawdata$PROPDMGEXP %in% c("", "K", "M", "B") & storm_rawdata$CROPDM
str(storm_rawdata)
```

```
## 'data.frame':    901921 obs. of  9 variables:
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
##  $ 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  "" "" "" "" ...
```

For estimating the health impact of severe weather events, the functions **group_by**, **summarize**, and **arrange** from the **dplyr** package with the pipeline operator will be use. Initially, the raw data frame is grouped by event type, and summarized by the total number of injuries and fatalities in separated data frames. Next, the data frames are merged and the total number of health related events that can be expressed as the sum of the injuries and fatalities is calculated, including a danger inidicator that is estimated as the percentage of fatalities from the health related events. The data frame is ordered in a descendent fashion according the number of health related events, and the top 10 is stored in the data frame **healthimpact_evtype**. The previously described process is executed in the next lines of code:

```
## Calculating FATALITIES AND INJURIES per Event Type
INJURIES_EVTYPE <- storm_rawdata %>% group_by(EVTYPE) %>%
        summarize(TOTAL_INJURIES = sum(INJURIES, na.rm = TRUE)) %>%
        arrange(desc(TOTAL_INJURIES))

FATALITIES_EVTYPE <- storm_rawdata %>% group_by(EVTYPE) %>%
        summarize(TOTAL_FATALITIES = sum(FATALITIES, na.rm = TRUE)) %>%
        arrange(desc(TOTAL_FATALITIES))
```

```
## Creating Health impact Data frame
healthimpact_evtype <- merge(INJURIES_EVTYPE, FATALITIES_EVTYPE, by = "EVTYPE")
healthimpact_evtype <- healthimpact_evtype %>% mutate(EVENTS = (TOTAL_FATALITIES+TOTAL_INJURIES), DANGE
                                                arrange(desc(EVENTS))
# Subsetting top 10 event types per number of health related events
healthimpact_evtype <- healthimpact_evtype[1:10,]
```

For calculating the property and crop damages cost estimation in dollars, the multivariate loop function **mapply** is used. A function that identifies the alphabetical character is introduced as argument to the mapply function, and applied to the DMG and EXP values from property and crop damage cost estimation. New columns for property and crop damage cost estimation in dollars are created

```
## Calculating Crop and Property damages in US
storm_rawdata$PROPDMG_inUS <- mapply(function(DMGval, DMGexp) {if (DMGexp == "K" | DMGexp == "k") {DMGva
                            else if (DMGexp == "M" | DMGexp == "m") {DMGval * 1e+06}
                            else if (DMGexp == "B") {DMGval * 1e+09}
                            else if (DMGexp == "") {DMGval*0}},
                            storm_rawdata$PROPDMG, storm_rawdata$PROPDMGEXP)


storm_rawdata$CROPDMG_inUS <- mapply(function(DMGval, DMGexp) {if (DMGexp == "K" | DMGexp == "k") {DMGva
                                 else if (DMGexp == "M" | DMGexp == "m") {DMGval * 1e+06}
                                 else if (DMGexp == "B") {DMGval * 1e+09}
                                 else if (DMGexp == "") {DMGval*0}},
                                 storm_rawdata$CROPDMG, storm_rawdata$CROPDMGEXP)
```

Similarly to health impact estimation process, the functions **group by**, **summarize**, and **arrange** are executed with the pipeline operator for grouping and summarizing the total crop and peroperty damage cost per weather event type. These data frames are merged, the total damage cost is calculated, and the top 10 most harmful weather events on the economy are extracted. This process is shown in the following code chunk:

```
library(dplyr)
## Calculating Crop and Property Damages economic impact per event type
PROPDMG_EVTYPE <- storm_rawdata %>% group_by(EVTYPE) %>%
        summarize(TOTAL_PROPDMG = sum(PROPDMG_inUS, na.rm = TRUE)/(1e+06)) %>%
        arrange(desc(TOTAL_PROPDMG))

CROPDMG_EVTYPE <- storm_rawdata %>% group_by(EVTYPE) %>%
        summarize(TOTAL_CROPDMG = sum(CROPDMG_inUS, na.rm = TRUE)/(1e+06)) %>%
        arrange(desc(TOTAL_CROPDMG))

economicimpact_evtype <- merge(CROPDMG_EVTYPE, PROPDMG_EVTYPE, by = "EVTYPE")
economicimpact_evtype <- economicimpact_evtype %>% mutate(TOTAL_DMG = (TOTAL_CROPDMG+TOTAL_PROPDMG)) %>
        arrange(desc(TOTAL_DMG))

# Subsetting top 10 economic impact per event types
economicimpact_evtype <- economicimpact_evtype[1:10,]
```

4. Results

4.1 Health impact of severe weather events in United States

Results from the health impact of weather events were stored in the **healthimpact_evtype** data frame. The next table shows the top 10 of severe weather event on the number of fatalities+injuries (or health related events) in the US. Also, the danger indicator is shown.

```
healthimpact_evtype
```

```
##              EVTYPE TOTAL_INJURIES TOTAL_FATALITIES EVENTS     DANGER
## 1           TORNADO          91285             5630  96915 0.05809214
## 2    EXCESSIVE HEAT           6525             1903   8428 0.22579497
## 3         TSTM WIND           6957              504   7461 0.06755127
## 4             FLOOD           6789              470   7259 0.06474721
## 5         LIGHTNING           5230              816   6046 0.13496527
## 6              HEAT           2100              937   3037 0.30852815
## 7       FLASH FLOOD           1777              978   2755 0.35499093
## 8         ICE STORM           1975               89   2064 0.04312016
## 9  THUNDERSTORM WIND          1488              133   1621 0.08204812
## 10     WINTER STORM           1321              206   1527 0.13490504
```

The next figure shows the total number of health related events in the use. The severity indicator as the proportion of fatalities to total events is used to fill the bars in the barplot.

```
library(ggplot2)
p <- ggplot(healthimpact_evtype, aes(x=reorder(EVTYPE, EVENTS), y=EVENTS, fill = DANGER)) +
  geom_bar(stat="identity") + theme_bw() + coord_flip() +
      labs(x = "Weather event types") +
      labs(y = "Number of health related events") +
      labs(title = "Health impact of severe weather events") +
      labs(colour = "Severity")
p
```

4.2 Economic impact of severe weather events in United States

The top 10 of the Weather events with the highest economic damages in Millions of US dollars is shown in the next table.

```
economicimpact_evtype
```

```
##              EVTYPE TOTAL_CROPDMG TOTAL_PROPDMG  TOTAL_DMG
## 1             FLOOD     5661.9685    144657.710 150319.678
## 2 HURRICANE/TYPHOON     2607.8728     69305.840  71913.713
## 3           TORNADO      364.9501     56925.485  57290.436
## 4       STORM SURGE        0.0050     43323.536  43323.541
## 5              HAIL     3000.5374     15727.166  18727.703
## 6       FLASH FLOOD     1420.7271     16140.812  17561.539
## 7           DROUGHT    13972.5660      1046.106  15018.672
## 8         HURRICANE     2741.9100     11868.319  14610.229
## 9       RIVER FLOOD     5029.4590      5118.945  10148.405
## 10        ICE STORM     5022.1100      3944.928   8967.038
```

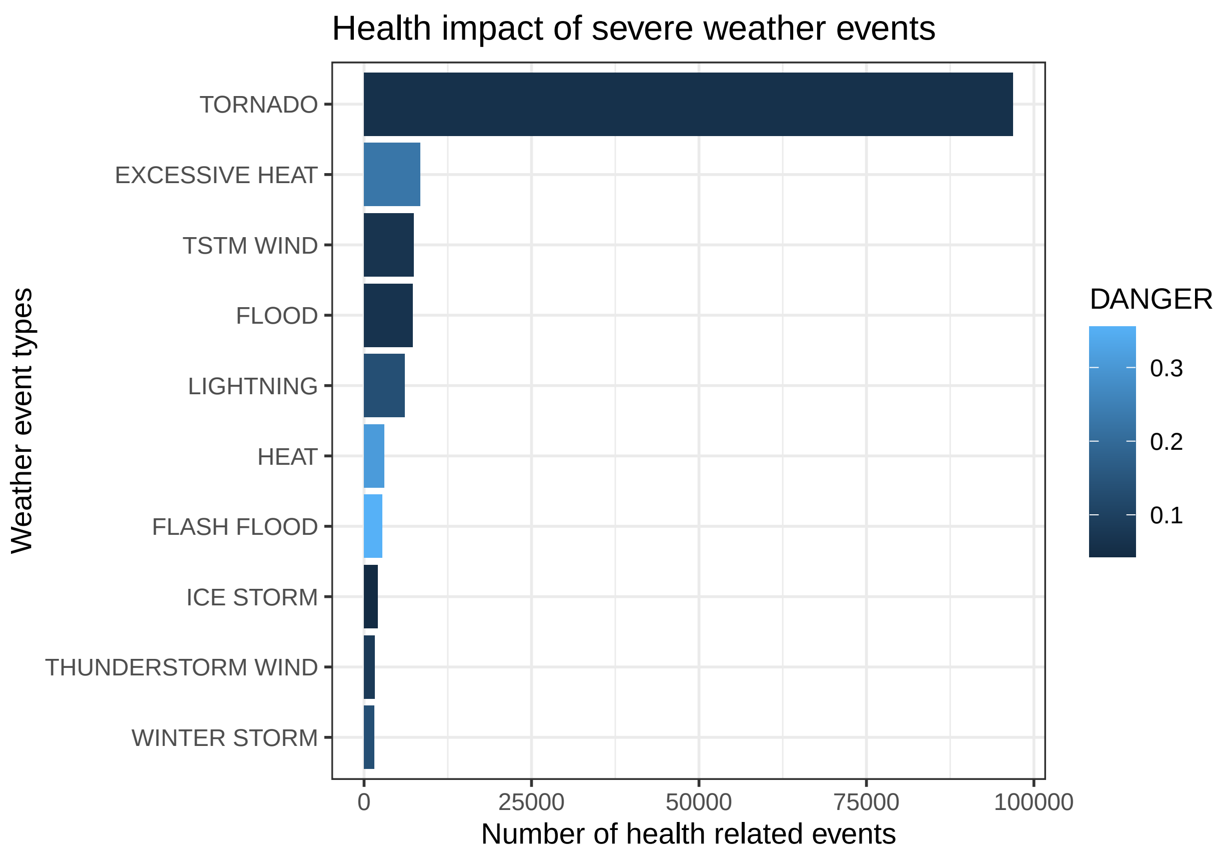

Figure 1: Figure 1. Health impact of severe weather events in US between 1950 to 2011

The next plot shows the total economic impact of Weather events in the US from 1950 to 2011.

```
library(ggplot2)
p <- ggplot(economicimpact_evtype , aes(x=reorder(EVTYPE, TOTAL_DMG), y=TOTAL_DMG)) +
  geom_bar(stat="identity") + theme_bw() + coord_flip() +
      labs(x = "Weather event types") +
      labs(y = "Total economic impact in Millions of US") +
      labs(title = "Economy impact of severe weather events")
p
```

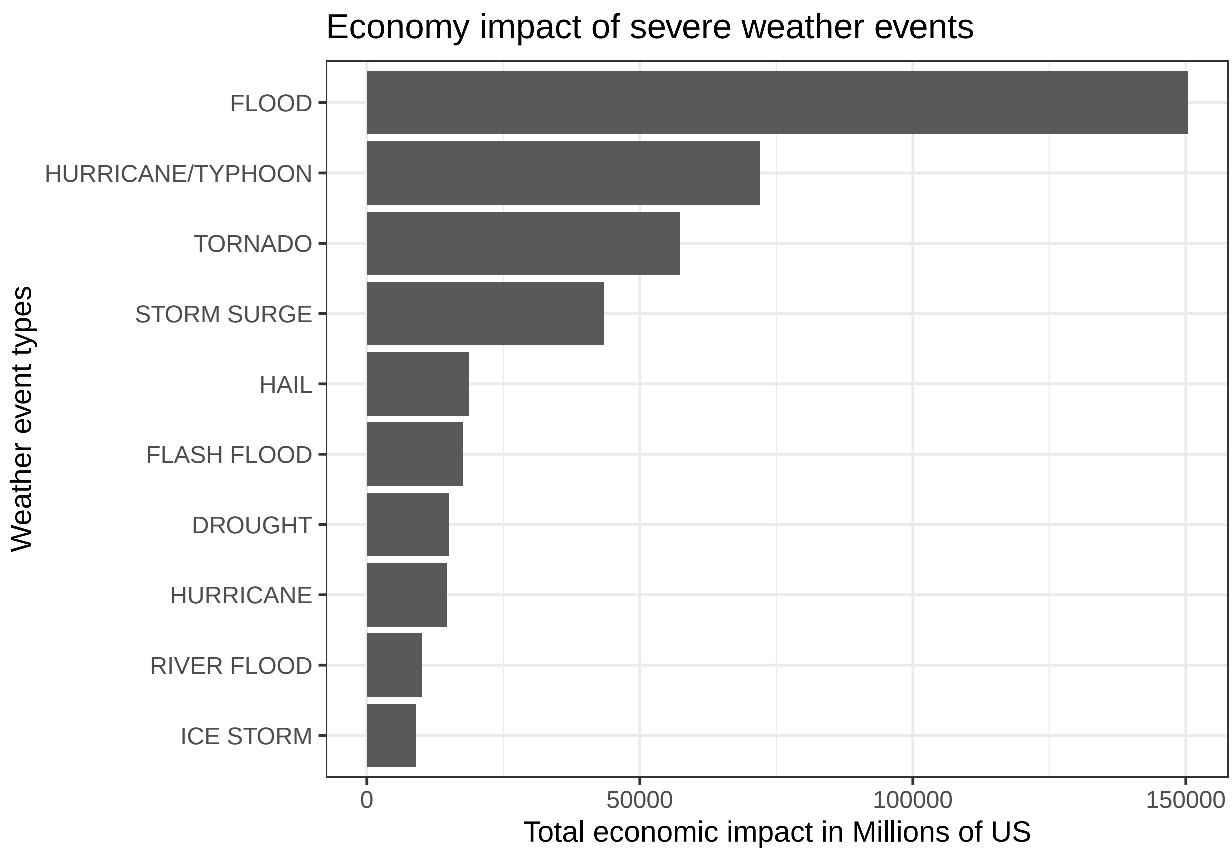

Figure 2: Figure 1. Economy impact of severe weather events in US between 1950 to 2011

Discussion

Clearly, Tornados are the most harmful weather events according to the health related event index, followed by Excessive Heat, TSTM Wind, Floods, and Lighting. Nevertheless, in terms of total number of fatalities, the rank changes starting from third place in which Flash Floods, and Heath occupies third and fourth place in the ranking maintaining Lighting in fifth place. Floods are the Weather events with the highest economic impact estimated as the total damage from property and crops, followed by Hurricane/Typhoons, Tornado, Storm Surge and Hails. While crop damage cost is important, property damage cost have a higher effect on the ranking results.