Major Wheater Events in the US: Period 1950 - 2011

Peer-graded Assignment: Course Project 2

Carlos Sanchez

15 February 2021

#Synopsis

The following study analyze data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database from different points of view. This database includes information about:

a. Storms and other significant weather phenomena that have provoked injuries or fatalities to people and crop or property damages
b. Rare weather phenomena
c. Other meteorological events that occur combined with another event

The current study includes the data collected between the years 1950 and 2011. Data from last years are more complete and with better quality.

#Data Processing

First we load the data:

```
zipFile <- "StormData.bz2"

if(!file.exists(zipFile)){
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(fileUrl, destfile = "StormData.bz2")
}

StormData <- read.csv("StormData.bz2")
```

Once loaded, we analyze the dataset dimensions and the different variables we have. We can see that we have 902297 rows and 37 variables. We can also see the first rows of the dataset:

```
dim(StormData)
```

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

```
head(StormData)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL TORNADO
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
```

```
## 4         0                                                    0         NA
## 5         0                                                    0         NA
## 6         0                                                    0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0                    14.0   100 3   0          0       15    25.0
## 2         0                     2.0   150 2   0          0        0     2.5
## 3         0                     0.1   123 2   0          0        2    25.0
## 4         0                     0.0   100 2   0          0        2     2.5
## 5         0                     0.0   150 2   0          0        2     2.5
## 6         0                     1.5   177 2   0          0        6     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
## 6          K       0                                         3450      8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
## 6          0          0              6
```

#Results

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

To answer this question, we will group the data by `EVTYPE` variables and will summarize the total number of Fatalities and Injuries that each event have provoked.

```
StormData_death_injured <- StormData %>% group_by(EVTYPE) %>% summarise(TotalFatalities = sum(FATALITIES
# StormData_death_injured <- subset(StormData_death_injured, TotalFatalities>100 | TotalInjuries > 1000

colors <- c("Total Injuries" = "blue", "Total Fatalities" = "red")
  # ggplot(StormData_death_injured) +
  # geom_point(aes(reorder(EVTYPE, -TotalFatalities), TotalFatalities, color = "Total Fatalities")) +
  #   geom_point(aes(EVTYPE, TotalInjuries, color = "Total Injuries")) +
  #   theme(axis.text.x = element_text(angle= 45, vjust = 1, hjust = 1, size = 6),
  #         legend.position = "right") +
  #   labs(x = "Event Type",
  #        y = "Total Events",
  #        color = "") +
  #   scale_color_manual(values = colors)
```

```
   ggplot(head(StormData_death_injured,10)) +
  geom_point(aes(reorder(EVTYPE, -TotalFatalities), TotalFatalities, color = "Total Fatalities")) +
    geom_point(aes(EVTYPE, TotalInjuries, color = "Total Injuries")) +
    theme(axis.text.x = element_text(angle= 45, vjust = 1, hjust = 1, size = 6),
          legend.position = "right") +
    labs(x = "Event Type",
         y = "Total Events",
         color = "") +
    scale_color_manual(values = colors)
```

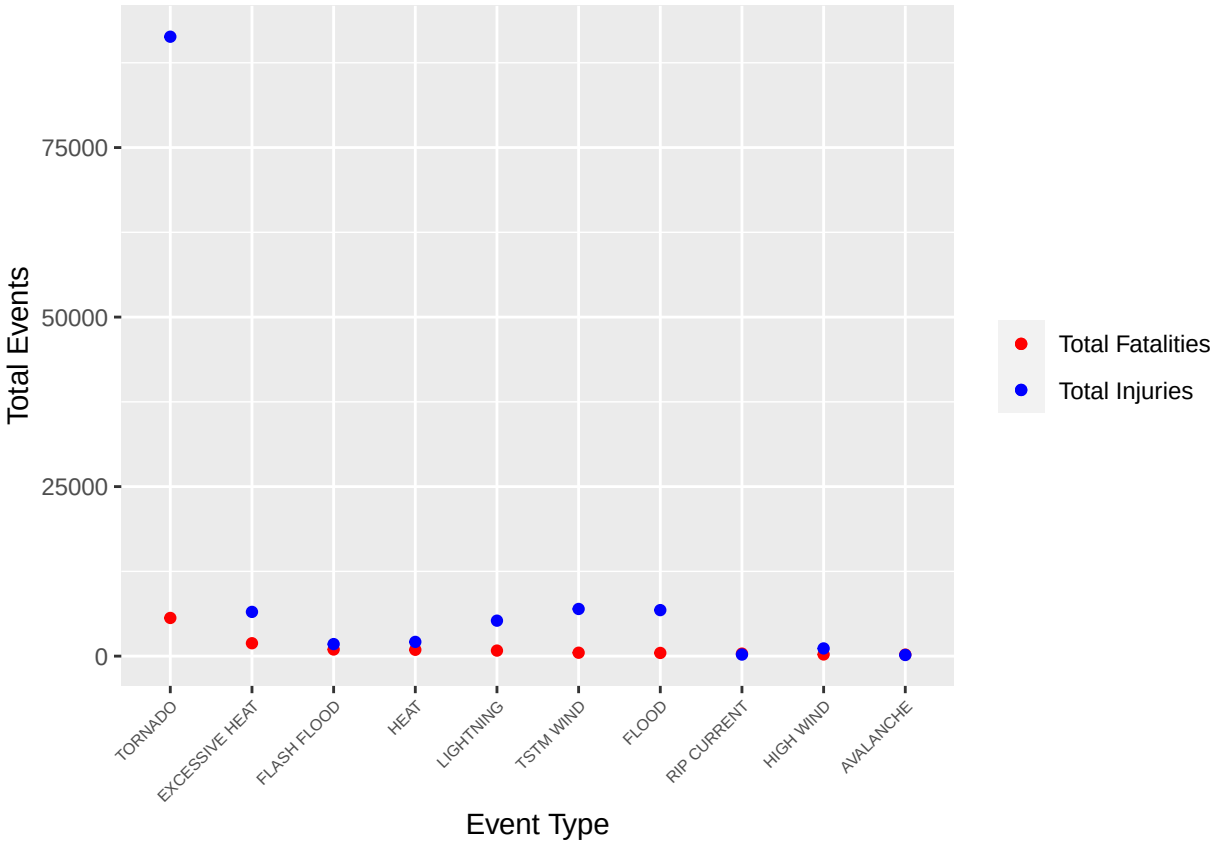

Figure 1: Total Fatalities and Injuries by Event

```
#   ggplot(StormData_death_injured) +
# geom_point(aes(reorder(EVTYPE, -TotalFatalities), log10(TotalFatalities), color = "Total Fatalities
#   geom_point(aes(EVTYPE, log10(TotalInjuries), color = "Total Injuries")) +
#   theme(axis.text.x = element_text(angle= 45, vjust = 1, hjust = 1, size = 6),
#         legend.position = "right") +
#   labs(x = "Event Type",
#        y = "Total Events",
#        color = "") +
#   scale_color_manual(values = colors)
```

If we display the top 10 Events in total Fatalities and Injuries in a table:

```
knitr::kable(head(StormData_death_injured, 10))
```

EVTYPE	TotalFatalities	TotalInjuries
TORNADO	5633	91346
EXCESSIVE HEAT	1903	6525
FLASH FLOOD	978	1777
HEAT	937	2100
LIGHTNING	816	5230
TSTM WIND	504	6957
FLOOD	470	6789
RIP CURRENT	368	232
HIGH WIND	248	1137
AVALANCHE	224	170

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

```
# StormData <- select(StormData, BGN_DATE, EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)

convert_letters <- function(x){ifelse(toupper(x)=="H", 100,
                                      ifelse(toupper(x)=="K", 1000,
                                             ifelse(toupper(x)=="M", 1e6,
                                                    ifelse(toupper(x)=="B", 1e9, 1))))}

StormData$PROPDMGEXP_num <- convert_letters(StormData$PROPDMGEXP)
StormData$CROPDMGEXP_num <- convert_letters(StormData$CROPDMGEXP)

StormData_Property_Damage <- StormData %>%
  filter(PROPDMG>0 | CROPDMG>0) %>%
  select(BGN_DATE, EVTYPE, PROPDMG, PROPDMGEXP_num, CROPDMG, CROPDMGEXP_num) %>%
  group_by(EVTYPE) %>%
  mutate(Total_Damage = PROPDMG*PROPDMGEXP_num + CROPDMG*CROPDMGEXP_num) %>%
  summarise(Total_Damage = sum(Total_Damage)) %>%
  arrange(desc(Total_Damage))

  ggplot(head(StormData_Property_Damage, 10)) +
   geom_col(aes(reorder(EVTYPE, -Total_Damage), Total_Damage, fill = Total_Damage)) +
    theme(axis.text.x = element_text(angle= 45, vjust = 1, hjust = 1, size = 6),
          legend.position = "right") +
    labs(x = "Event Type",
         y = "Total Damage",
         legend_title="",
```

```
        title = "Total Damage of Crop and Property by Cause")
```

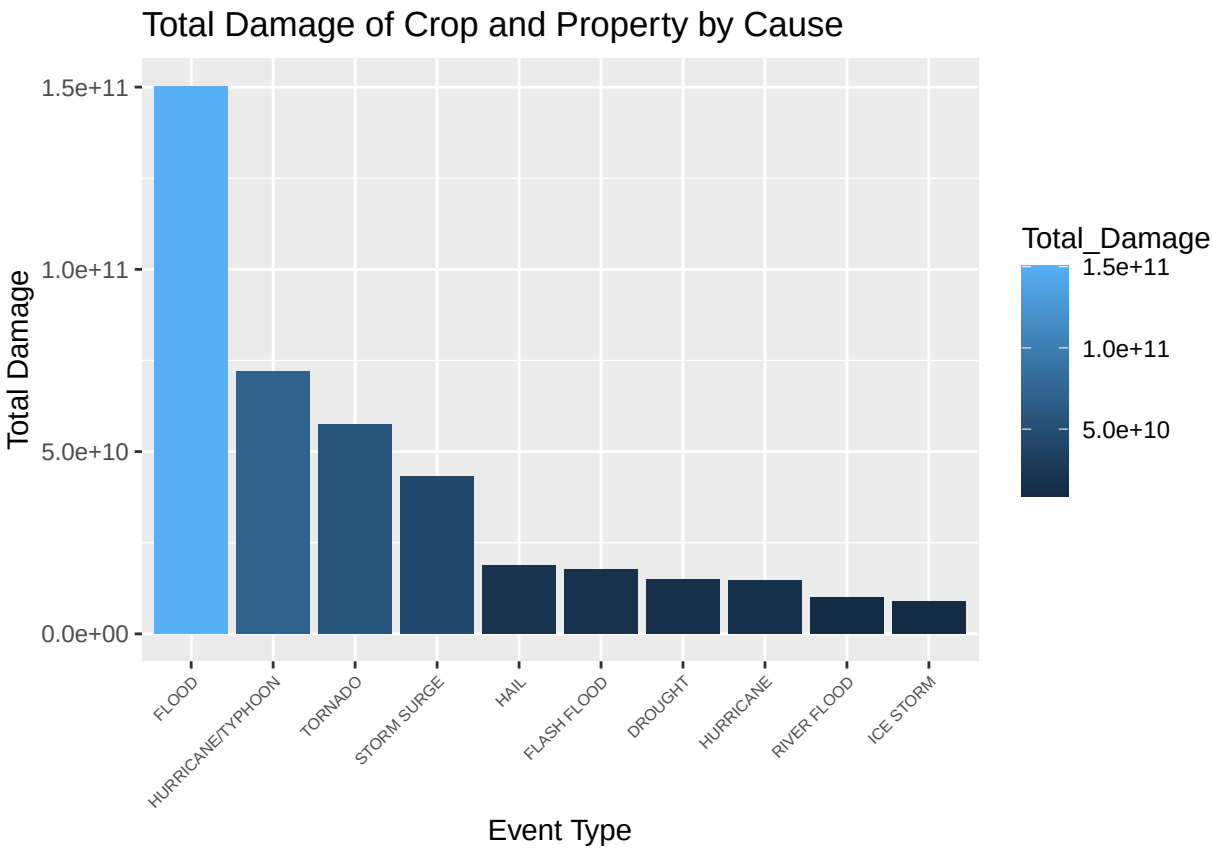

Figure 2: Total Damage of Crop and Property by Cause