Analysis of the impact of weather events in USA (1950-2011)

Nourdine Dehri

December 5, 2016

Synopsis

The basic goal of this assessment is to explore the U.S. National Oceanic and Atmospheric Administration's (NOAA) Storm Database and answer some basic questions about severe weather events.
We will use the database to answer the questions below and show you the code for the entire analysis.
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?

Instructions

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.

Data

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site:

- Storm Data [47Mb]

There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

- National Weather Service Storm Data Documentation
- National Climatic Data Center Storm Events FAQ

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

Data Processing

Load Libraries

```
library(R.utils)
library(data.table)
library(dplyr)
library(ggplot2)
library(gridExtra)
```

Load Data

First, we download the Storm Data and unzip it.

```
setwd("~/myR_scripts/coursera/Reproducible-Research/Assignment-week4")

if(!file.exists("Data/StormData.csv.bz2")){

        # Download the data
        download.file(url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
        destfile = "Data/StormData.csv.bz2"
        )

        # Unzip the file
        bunzip2("Data/StormData.csv.bz2", "Data/StormData.csv", remove = FALSE, skip = TRUE)

}
```

Then, we load the data into R. If the data already exists in the working environment, we do not need to load it again.

```
if (!"ds" %in% ls()) {
    ds <- fread("Data/StormData.csv")
}
```

```
##
Read 26.9% of 967216 rows
Read 49.6% of 967216 rows
Read 67.2% of 967216 rows
Read 78.6% of 967216 rows
Read 91.0% of 967216 rows
Read 902297 rows and 37 (of 37) columns from 0.523 GB file in 00:00:07
```

```
dim(ds)
```

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

The dataset contains 902297 **observations** and 37 **variables**.

Results

1. Across the United States, which types of events are most harmful with respect to population health?

We will focus on the top 3 of weather events that caused the most **fatalities** and **injuries**.

Fatalities

```
FATAL <- group_by(ds, EVTYPE) %>%
      summarize(TOT.FATALITIES = sum(FATALITIES)) %>%
      arrange(desc(TOT.FATALITIES))

FATAL <- as.data.frame(FATAL[1:3,])
head(FATAL)
```

```
##           EVTYPE TOT.FATALITIES
## 1        TORNADO           5633
## 2 EXCESSIVE HEAT           1903
## 3    FLASH FLOOD            978
```

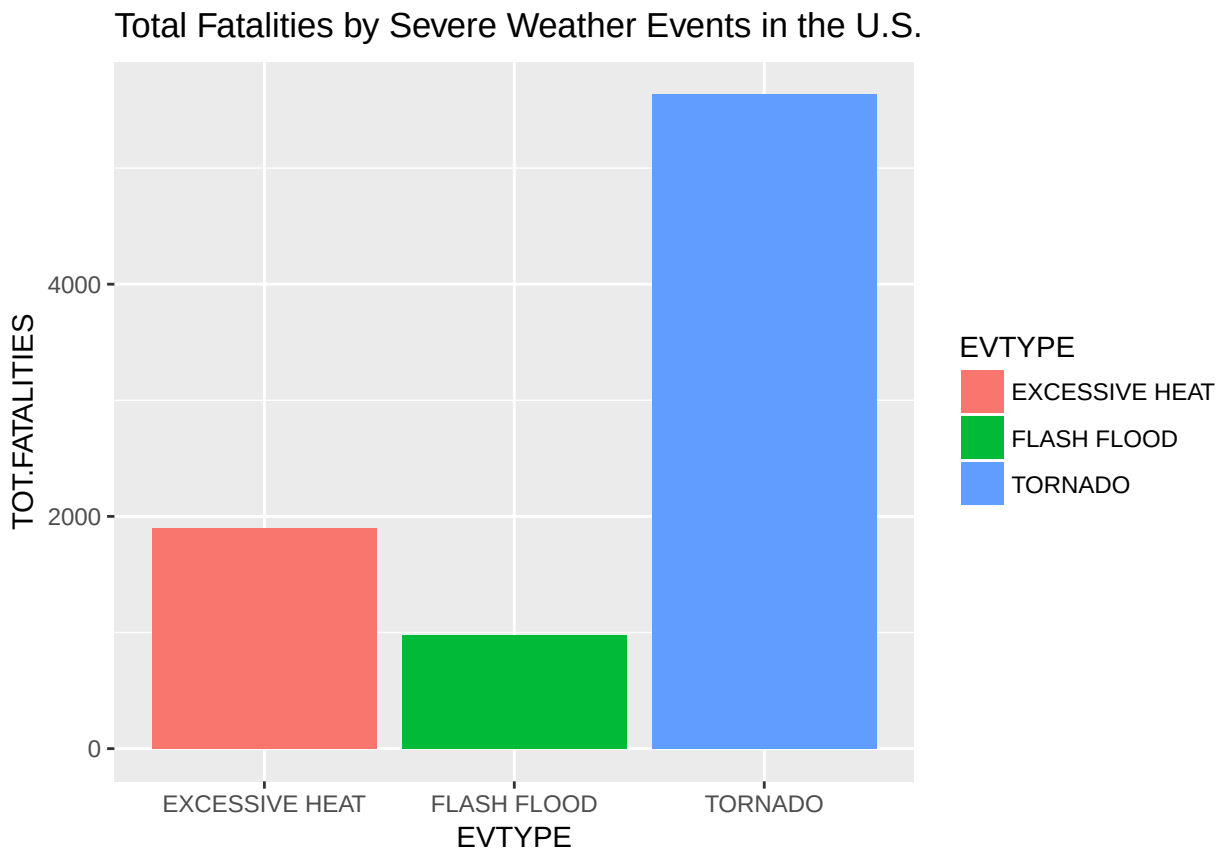
Injuries

```
INJURY <- group_by(ds, EVTYPE) %>%
      summarize(TOT.INJURIES = sum(INJURIES)) %>%
      arrange(desc(TOT.INJURIES))


INJURY <- as.data.frame(INJURY[1:3,])
head(INJURY)
```

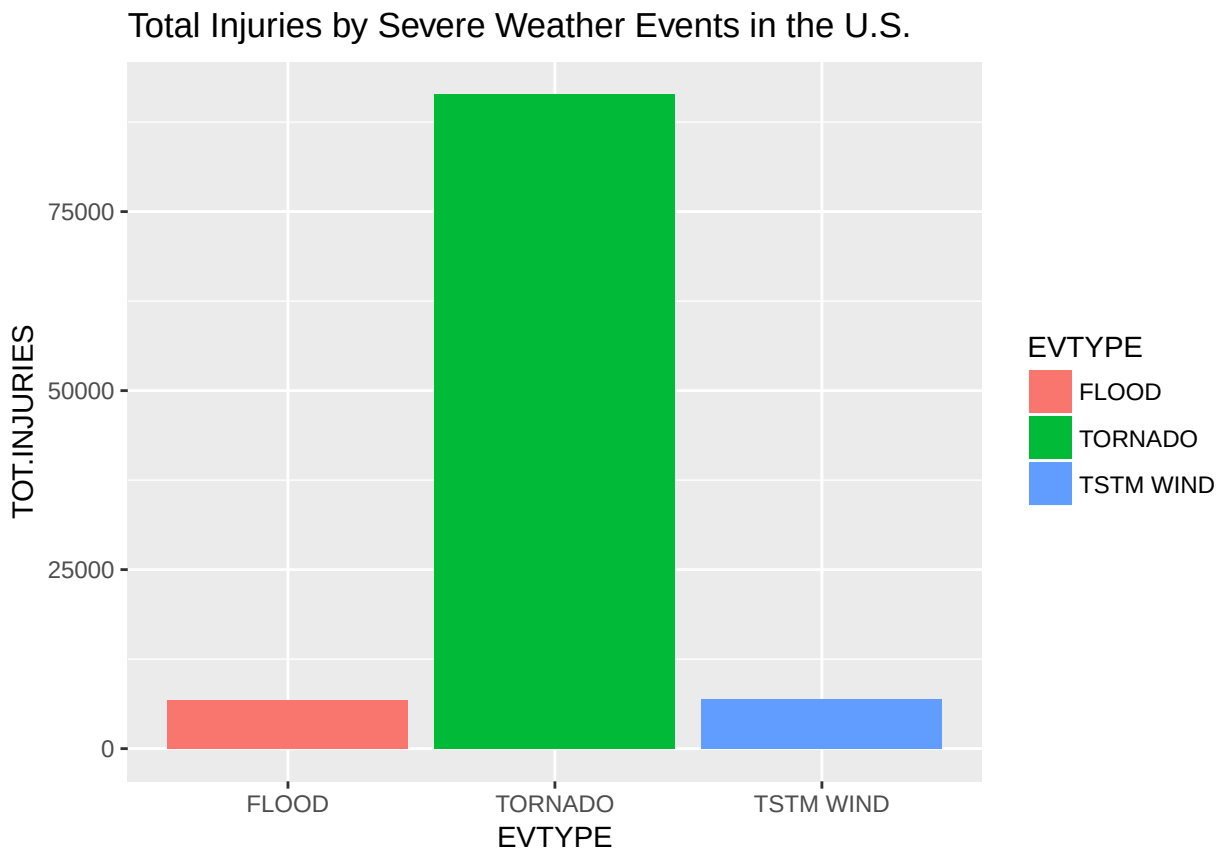
```
##      EVTYPE TOT.INJURIES
## 1   TORNADO        91346
## 2 TSTM WIND         6957
## 3     FLOOD         6789
```

Graphically,

```
# Fatalities plot
ggplot(data = FATAL, aes(x=EVTYPE, y=TOT.FATALITIES, fill=EVTYPE)) + geom_bar(stat = "identity") +
        ggtitle("Total Fatalities by Severe Weather Events in the U.S.")
```




```
# Injuries plot
ggplot(data = INJURY, aes(x=EVTYPE, y=TOT.INJURIES, fill=EVTYPE)) + geom_bar(stat = "identity") +
        ggtitle("Total Injuries by Severe Weather Events in the U.S.")
```



Based on the above bar plot, we find that **tornado** causes most fatalities and most injuries in the United States.

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

We will focus on the top 3 of weather events that have the greastest **economic impact** in **properties** and **crop**

Properties Damage

```
ds <- mutate(ds, PROPDMGEXP = toupper(PROPDMGEXP))

ds <- mutate(ds, PROPDMG = ifelse(PROPDMGEXP == "K", PROPDMG*1000,
                                  ifelse(PROPDMGEXP == "M", PROPDMG*100000,
                                         ifelse(PROPDMGEXP == "B", PROPDMG*100000000,0
                                         )
                                  )
)
)


PROP.damage <- group_by(ds, EVTYPE) %>%
               summarize(TOT.PROPDMG = sum(PROPDMG)) %>%
               arrange(desc(TOT.PROPDMG))
```

```
PROP.damage <- as.data.frame(PROP.damage[1:3,])
head(PROP.damage)
```

```
##              EVTYPE TOT.PROPDMG
## 1             FLOOD 15256447800
## 2           TORNADO  8540848480
## 3 HURRICANE/TYPHOON  6932357000
```

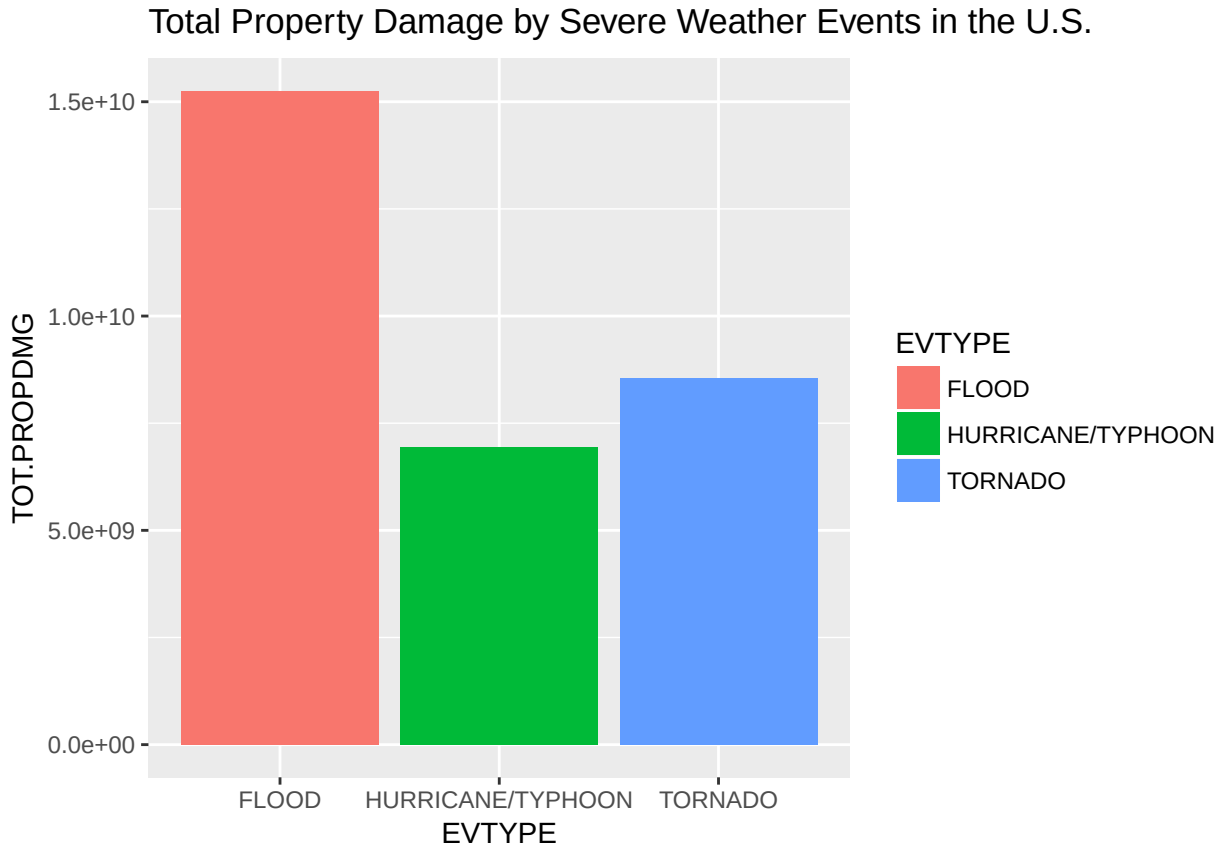
Crop Damage

```
ds <- mutate(ds, CROPDMGEXP = toupper(CROPDMGEXP))

ds <- mutate(ds, CROPDMG = ifelse(CROPDMGEXP == "K", CROPDMG*1000,
                                  ifelse(CROPDMGEXP == "M", CROPDMG*100000,
                                         ifelse(CROPDMGEXP == "B", CROPDMG*100000000,0
                                         )
                                  )
)
)

CROP.damage <- group_by(ds, EVTYPE) %>%
               summarize(TOT.CROPDMG = sum(CROPDMG)) %>%
               arrange(desc(TOT.CROPDMG))

CROP.damage <- as.data.frame(CROP.damage[1:3,])
head(CROP.damage)
```

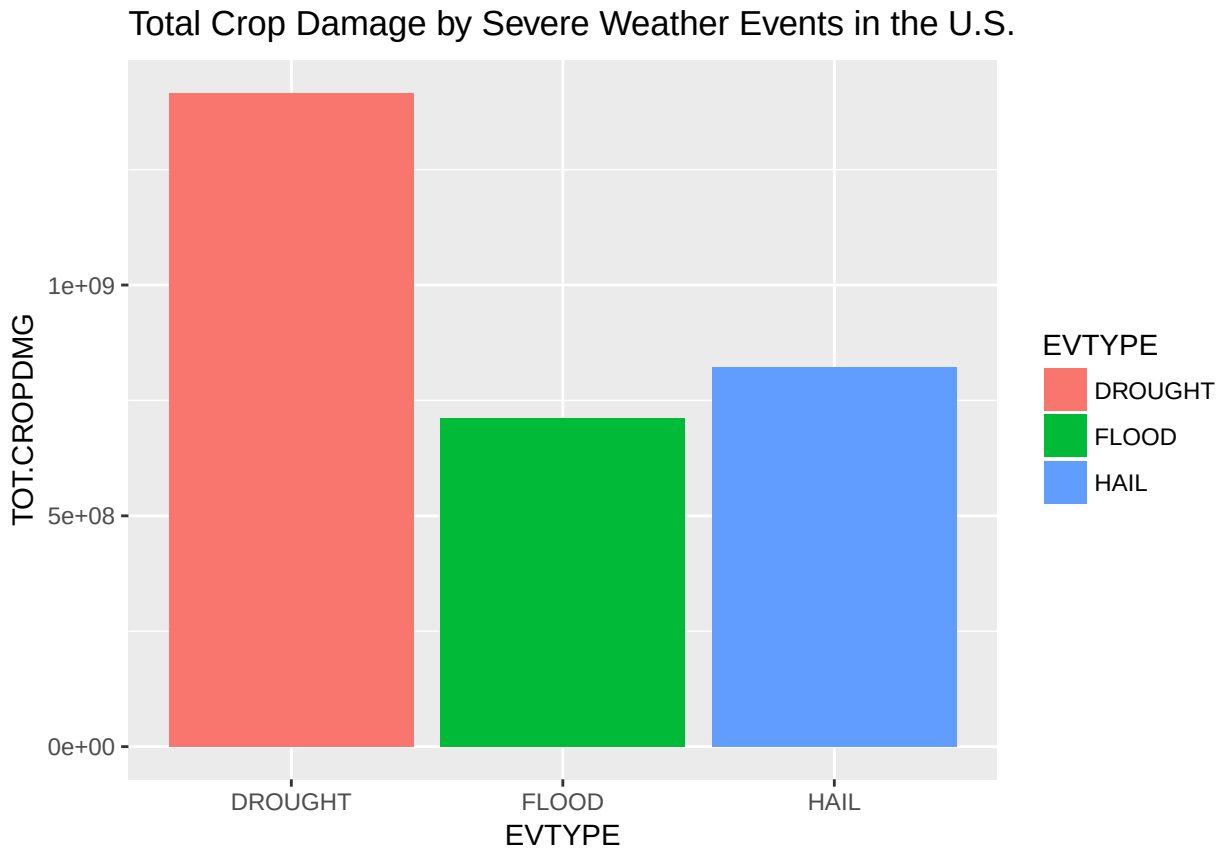
```
##    EVTYPE TOT.CROPDMG
## 1 DROUGHT  1416558000
## 2    HAIL   822007450
## 3   FLOOD   712481450
```

Graphically,

```
# Property damage plot
ggplot(data = PROP.damage, aes(x=EVTYPE, y=TOT.PROPDMG, fill=EVTYPE)) + geom_bar(stat = "identity") +
        ggtitle("Total Property Damage by Severe Weather Events in the U.S.")
```




```
# Crop damage plot
ggplot(data = CROP.damage, aes(x=EVTYPE, y=TOT.CROPDMG, fill=EVTYPE)) + geom_bar(stat = "identity") +
        ggtitle("Total Crop Damage by Severe Weather Events in the U.S.")
```



Based on the bar plot above, we find that :
- **flood** causes most property damage;
- **drought** causes most crop damage;
in the United States

Conclusion

From these data, we found that **tornado** causes most fatalities and most injuries in the United States; while **flood** and **drought** have the greatest economic impact.