How severe weather events impact public health and economy in the United States

Synopsis

In this analysis, we aim to analyze the impact of different severe weather events on public health and economy. It will be based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration's (NOAA) from 1950 - 2011. We will use the estimates of fatalities, injuries and property to decide which types of event are most harmful to the population health and economy. From these data, we found that excessive heat and tornado are most harmful with respect to population health, while flood and hurricane/typhoon have the greatest economic consequences.

Basic settings

We fixed a few basic settings to be used during this analysis.

```
setwd("~/Documents/ProgramsGitHub/RepData_PeerAssessment2")  # working directory
library(R.utils)       # to be able to unzip bzip2 files
library(ggplot2)       # to be able to graph
require(gridExtra)     # to be able join two graphs
```

Data processing

In this section we download the database that we are going to use and unzip it.

```
if ( !"stormdata.csv" %in% dir() ) {                        # existence of csv file
    download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                  destfile = "stormdata.csv.bz2")           # to download the bz2 file
    bunzip2("stormdata.csv.bz2", overwrite=F, remove=T)  # to unzip/delete bz2 file
}
```

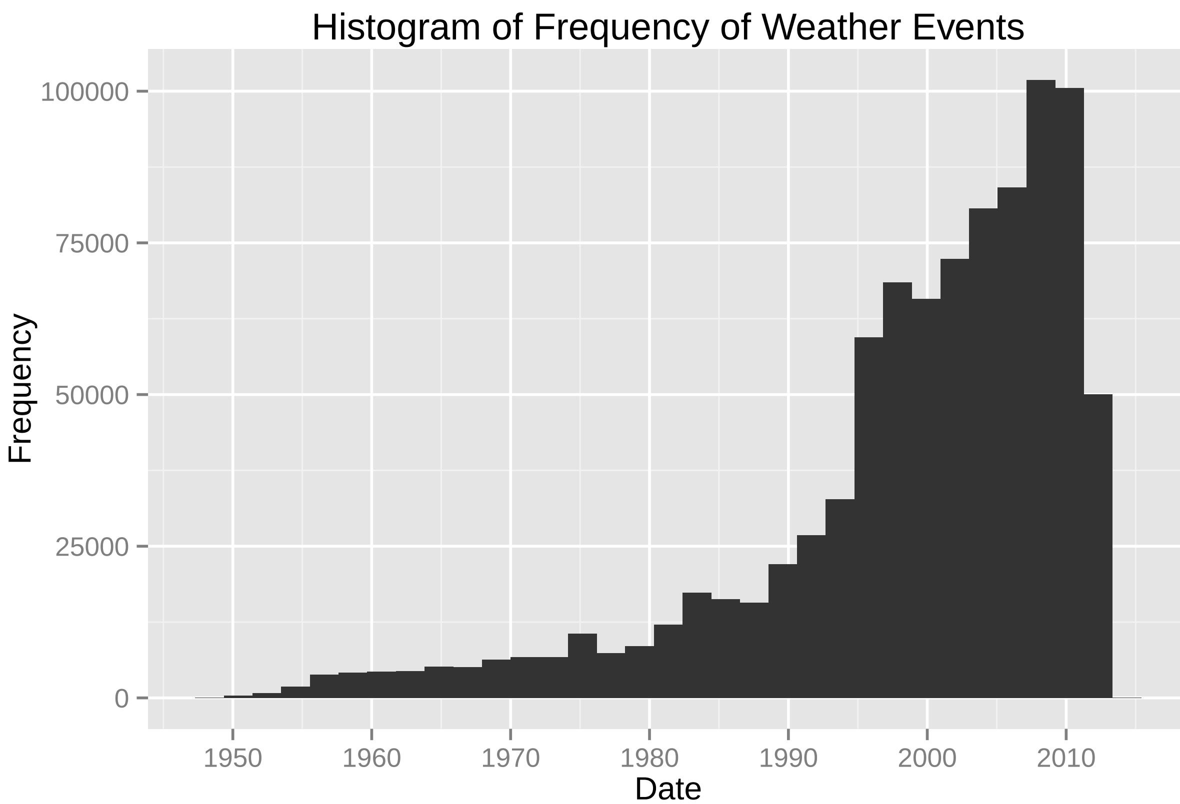
This is the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. The events in the database start in the year 1950 and end in November 2011. For more information about the data: Storm Data Documentation.

Next we load the data to the working enviroment.

```
if ( !"stormdata" %in% ls() ) {                        # existence of the data
    stormdata <- read.csv("stormdata.csv", sep = ",")  # to load the data
}
```

The second column of this database BGN_DATE documents the begin date of the severe weather event. Using this information we graph the frequency of the severe weather events during different years.

```
stormdata$BGN_DATE <- as.Date(stormdata$BGN_DATE,format = "%m/%d/%Y %H:%M:%S")
# to convert to Date Format
g <- ggplot(stormdata,aes(BGN_DATE))  # data
g <- g + geom_histogram()             # type of graph
g <- g + labs(title = "Histogram of Frequency of Weather Events",x="Date",y="Frequency")
g
```



We can see that in the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. Therefore, we are going to restrict our analysis to years between 1997-2011, this give us a time period of 15 years with a better record of events.

```
stormdata <- stormdata[stormdata$BGN_DATE >= as.Date("1997-01-01"), ]
# to select dates between 1997-2011
```

Fatalities

In this section we are going to collect the total number of fatalities that are caused by severe weather events and select the the first 10 most severe types of weather events with fatalities.

```
fatalities <- aggregate(stormdata$FATALITIES, list( type = stormdata$EVTYPE), sum)
# to collect the total number of fatalities by weather event
fatalities <- fatalities[order(fatalities$x,decreasing=TRUE), ]   # to order the data
fatalities <- fatalities[1:10,]                                   # to take first 10
```

Injuries

In this section we are going to collect the total number of injuries that are caused by severe weather events and select the the first 10 most severe types of weather events with injuries.

```
injuries <- aggregate(stormdata$INJURIES, list( type = stormdata$EVTYPE), sum)
# to collect the total number of injuries by weather event
injuries <- injuries[order(injuries$x,decreasing=TRUE), ]   # to order the data
injuries <- injuries[1:10,]                                 # to take first 10
```

Economy

Finally, we convert the property damage data into comparable numerical forms according to the meaning of units described in the code book. We colect the total property damage caused by severe weather events and select the the first 10 most severe types of weather events with property damage.

```
levels(stormdata$PROPDMGEXP)                                              # to see levels
```

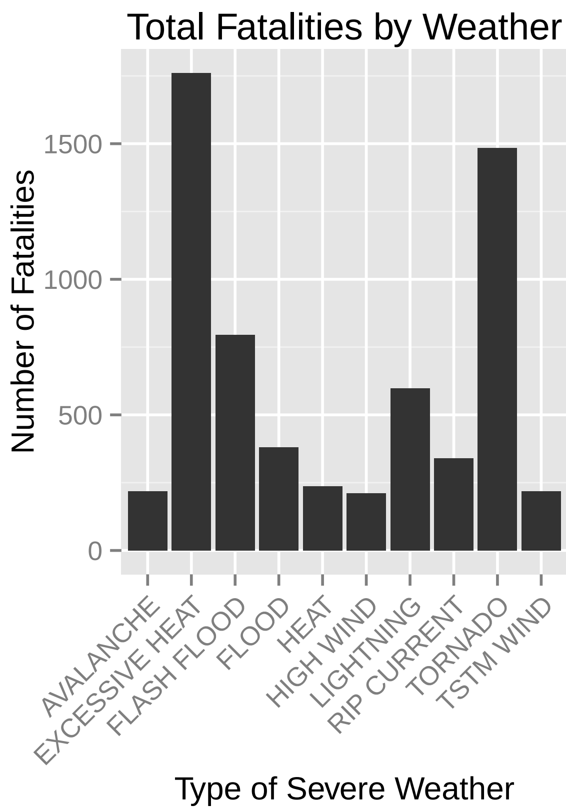
```
##  [1] ""  "-" "?" "+" "0" "1" "2" "3" "4" "5" "6" "7" "8" "B" "h" "H" "K"
## [18] "m" "M"
```

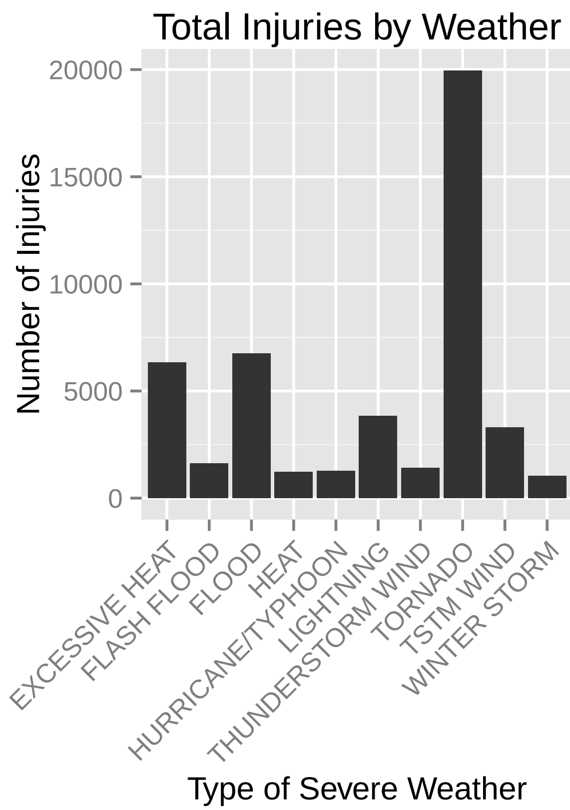
```
levels(stormdata$PROPDMGEXP) <- c("1",NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,
                                  "1000000000","100","100","1000","1000000","1000000")
# to rename levels
stormdata$PROPDMGEXP<-as.numeric(as.character(stormdata$PROPDMGEXP))
# to convert levels to numbers
stormdata$PROPDMG <- stormdata$PROPDMG*stormdata$PROPDMGEXP
#to get correct values of property damage
property <- aggregate(stormdata$PROPDMG, list( type = stormdata$EVTYPE), sum)
# to collect the total number of property damage by weather event
property <- property[order(property$x,decreasing=TRUE), ]    # to order the data
property <- property[1:10,]                                  # to take first 10
```

Results

Here we graph the data that we obtain about fatalities and injuries:

```
g1 <- ggplot(fatalities,aes(type,x))
g1 <- g1 + geom_histogram(stat="identity")
g1 <- g1 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
g1 <- g1 + scale_y_continuous()
g1 <- g1 + labs(title = "Total Fatalities by Weather")
g1 <- g1 + labs(x="Type of Severe Weather",y="Number of Fatalities")
g2 <- ggplot(injuries,aes(type,x))
g2 <- g2 + geom_histogram(stat="identity")
g2 <- g2 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
g2 <- g2 + scale_y_continuous()
g2 <- g2 + labs(title = "Total Injuries by Weather")
g2 <- g2 + labs(x="Type of Severe Weather",y="Number of Injuries")
grid.arrange(g1,g2,ncol=2)
```

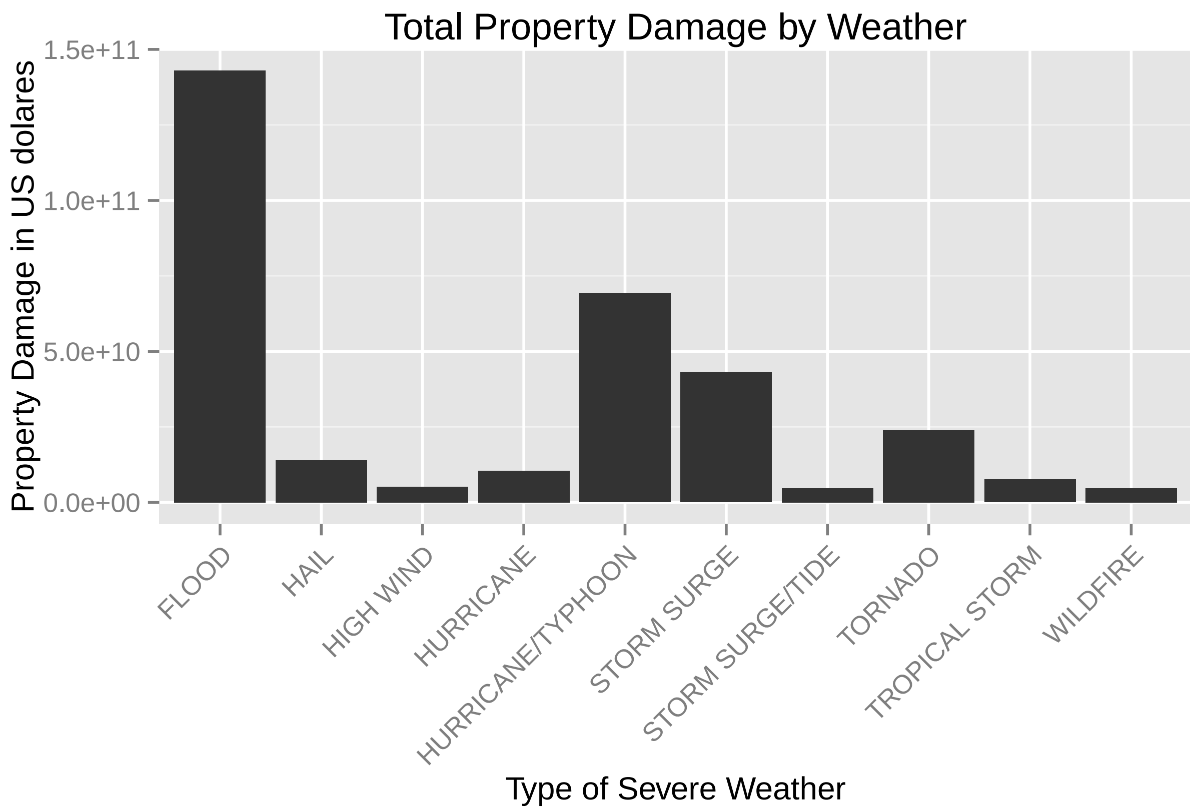




Based on the above graphs, we find that excessive heat and tornado cause most fatalities; tornato causes most injuries in the United States between 1997 to 2011.

Finally, we graph the information about property damage from the previous section.

```
g3 <- ggplot(property,aes(type,x))
g3 <- g3 + geom_histogram(stat="identity")
g3 <- g3 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
g3 <- g3 + scale_y_continuous()
g3 <- g3 + labs(title = "Total Property Damage by Weather")
g3 <- g3 + labs(x="Type of Severe Weather",y="Property Damage in US dolares")
g3
```



Based in the last graph, Flood and Hurricane/Typhoon cause most property damage in the United States between 1997 to 2011.