The health and economic impact of severe weather events in USA based on NOAA Storm Database

Djamila Azib

March 24, 2018

Synopsis

The goal of the assignment is to explore the NOAA Storm Database in order to found out the effects of severe weather events on both population and economy in USA. Specifically, the analysis aims to answer the two following questions :

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?

The NOAA Storm Database covers the time period between 1950 and November 2011.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.

Our anlysis shows that Tornadoes are the most devastating ,causes the most large fatalities and injuries.The flood have the largest total economic impact out of all severe weather events.

Data Processing

The data can be downloaded from the course website:

Storm Data

Documentation of the database is available at National Weather Service: ***Storm Data Documentation*** National Climatic Data Center Storm Events : ***FAQ***

required library

for this project we will use:

library(ggplot2)

library(dplyr)

Loading data

The data for this assignment come in the form of a comma-separated-value compressed via the bzip2 algorithm to reduce its size. The data is downlaoded from the course website by using download.file().

read.csv() can handle this compressed file automatically ,read the file in table format and create a data frame.

```
bz2File = "storm.csv.bz2"
#download the file is it does not exist
if (!file.exists(bz2File)){
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url,bz2File,method="curl")
```

```r
}

#read the  file in table format and create a data frame.
stormData<-read.csv( bz2File ,stringsAsFactors = FALSE)
str(stormData)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ 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" .
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ 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  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

The dataset contains 902297 obs. of 37 variables.

Extracting the required data

The variables required to analyze the impact of severe weather events on the public health and economy are :

1. EVTYPE: Weather event
2. INJURIES: Non-fatal injuries
3. FATALITIES: Fatal injuries

4. CROPDMG: Dollar amount of damage in crops, without level of magnitude
5. CROPDMGEXP: Level of Magnitude of the damage in crops
6. PROPDMG: Dollar amount of property damage, without level of magnitude
7. CROPDMGEXP: Level of Magnitude of property damage

select() (package dplyr) get the required columns from the data set.

```
requiredColomns<-c("EVTYPE", "INJURIES", "FATALITIES","CROPDMG", "CROPDMGEXP", "PROPDMG", "PROPDMGEXP")
stormDataSubset <- stormData %>% select( requiredColomns)
```

processing data for Health impact analysis

Sum number of fatalities by Event type and sort it in descending order

Sum number of injuries by Event and sort it in descending order.

```
#group by envent type et sum fatalities
fatalByEvtype <- stormDataSubset%>%group_by(EVTYPE )%>%
  summarise(fatalities = sum(FATALITIES),na.rm=TRUE)
#sort in descending order fatalities
fatalByEvtype<-fatalByEvtype%>%arrange(-fatalities)

#group by envent type et sum injuries
injuryByEvtype<-stormDataSubset%>%group_by(EVTYPE)%>%
  summarise(injuries = sum(INJURIES),na.rm=TRUE)
#sort in descending order injuries
injuryByEvtype<-injuryByEvtype%>%arrange(-injuries)
```

Processing data for economic consequences analysis

```
#get unique value for the crop exponent data CROPDMGEXP
print(cropEXp<-unique(toupper(stormDataSubset$CROPDMGEXP)))
```

```
## [1] ""  "M" "K" "B" "?" "0" "2"
```

```
#get unique value for  the property exponent PROPDMGEXP
print(propExp<-unique(toupper(stormDataSubset$PROPDMGEXP)))
```

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

```
valExp<-c(1000, 1e+06 ,1,  1e+09, 0, 1, 1e+05 ,1e+06 ,0 ,1e+04, 100, 1000,100, 1e+07,0, 10, 1e+08)

#create a data frame  with two columns x=propExp and val the corresponding numerical values
## with just propExp because cropEXp is include in propexp
print(mp<-data.frame(x=propExp,val= valExp))
```

```
##    x   val
## 1  K 1e+03
## 2  M 1e+06
## 3    1e+00
## 4  B 1e+09
## 5  + 0e+00
## 6  0 1e+00
## 7  5 1e+05
## 8  6 1e+06
```

```
## 9  ? 0e+00
## 10 4 1e+04
## 11 2 1e+02
## 12 3 1e+03
## 13 H 1e+02
## 14 7 1e+07
## 15 - 0e+00
## 16 1 1e+01
## 17 8 1e+08
```

```
#expNum return the corresponding numerical value of exp
expNum <- function(exp){

return(mp$val[mp$x==toupper(exp)])
}
#each element is the result of applying expNum to the corresponding element of stormDataSubset$PROPDMGE
propExpNum <- sapply(stormDataSubset$PROPDMGEXP, function(exp) expNum(exp))

# Same ,we get the corresponding numerical values to each element of stormDataSubset$CROPDMGEXP
cropExpNum<- sapply(stormDataSubset$CROPDMGEXP, function(exp) expNum(exp))

#properties damage  and crop damage in billion dollars
stormDataSubset$PropDMGVal <- (stormDataSubset$PROPDMG * propExpNum)/10^9
stormDataSubset$CropDMGVal <- (stormDataSubset$CROPDMG * cropExpNum)/10^9

# sum properties damage  and crop damage
stormDataSubset$totalDMGval <- stormDataSubset$CropDMGVal + stormDataSubset$PropDMGVal

#group by envent type and sum totalDMGval
damageByEvtype <- stormDataSubset%>%group_by(EVTYPE)%>% summarise(totalDamage = sum(totalDMGval ,na.rm=
#sort in descending order totalDamage
damageByEvtype<-damageByEvtype%>%arrange(-totalDamage)
```

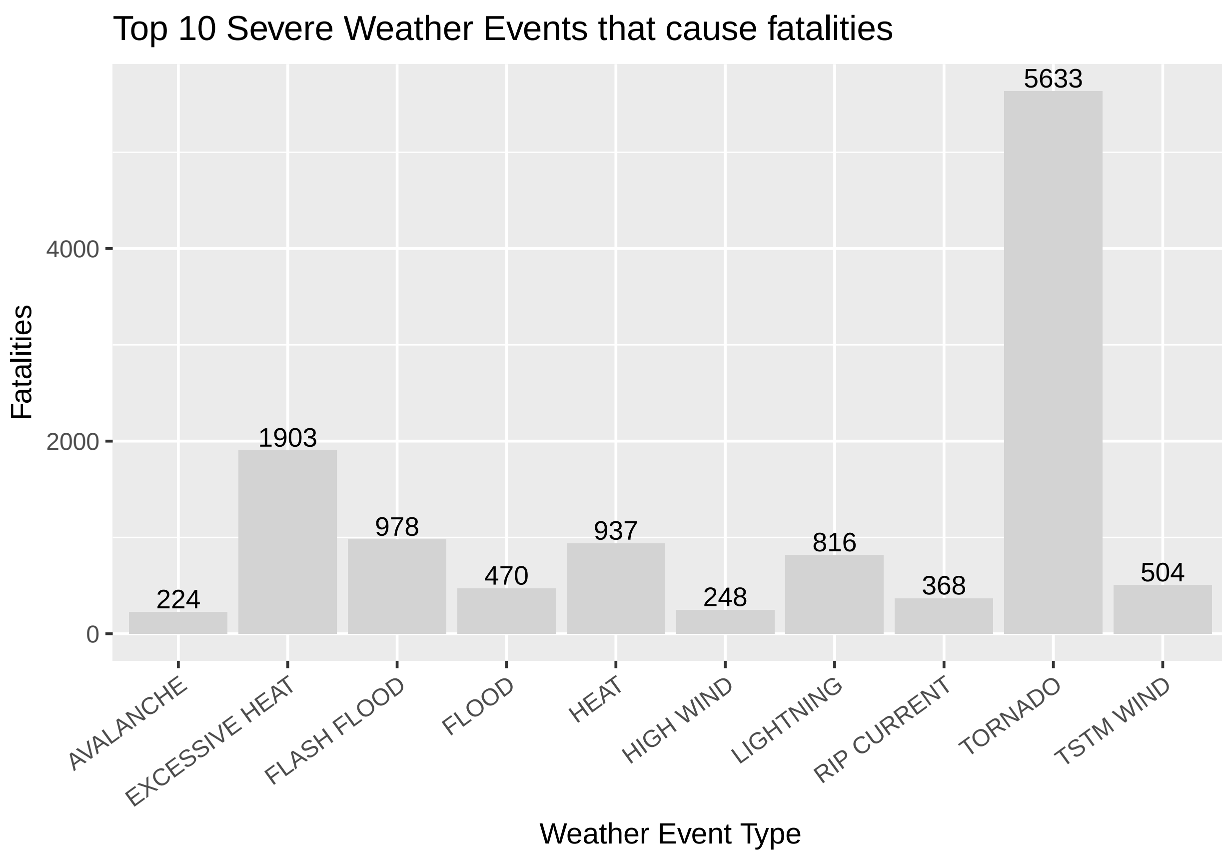
Results

Health impact of severe weather events

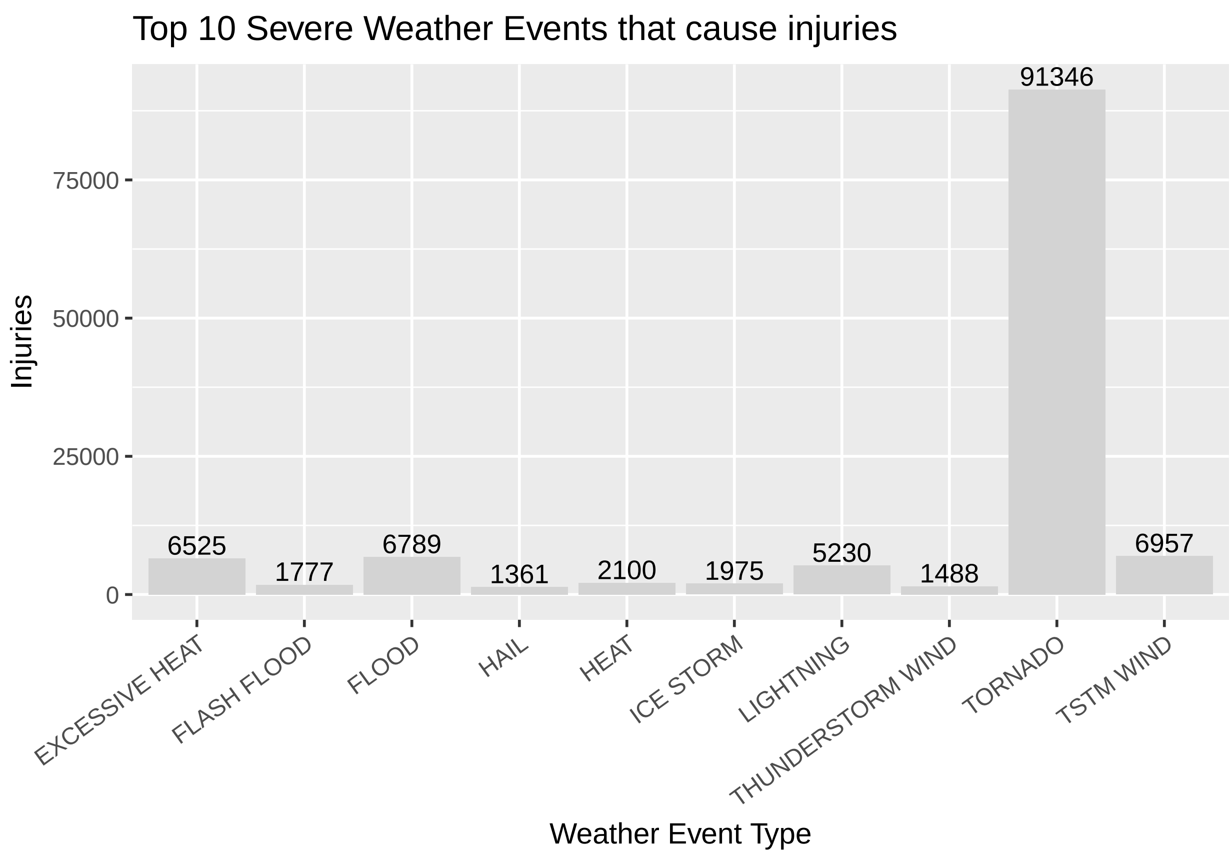
```
#top 10 event type that cause fatalities
topFatal<-fatalByEvtype%>%head(10)

#plot top 10 Severe Weather Events that cause fatalities

ggplot(data=topFatal, aes(x=EVTYPE, y=fatalities )) +
geom_bar(stat="identity",fill="lightgray") + xlab("Weather Event Type") + ylab("Fatalities") +
ggtitle("Top 10 Severe Weather Events that cause fatalities") + geom_text(aes(label=fatalities), vjust=
```



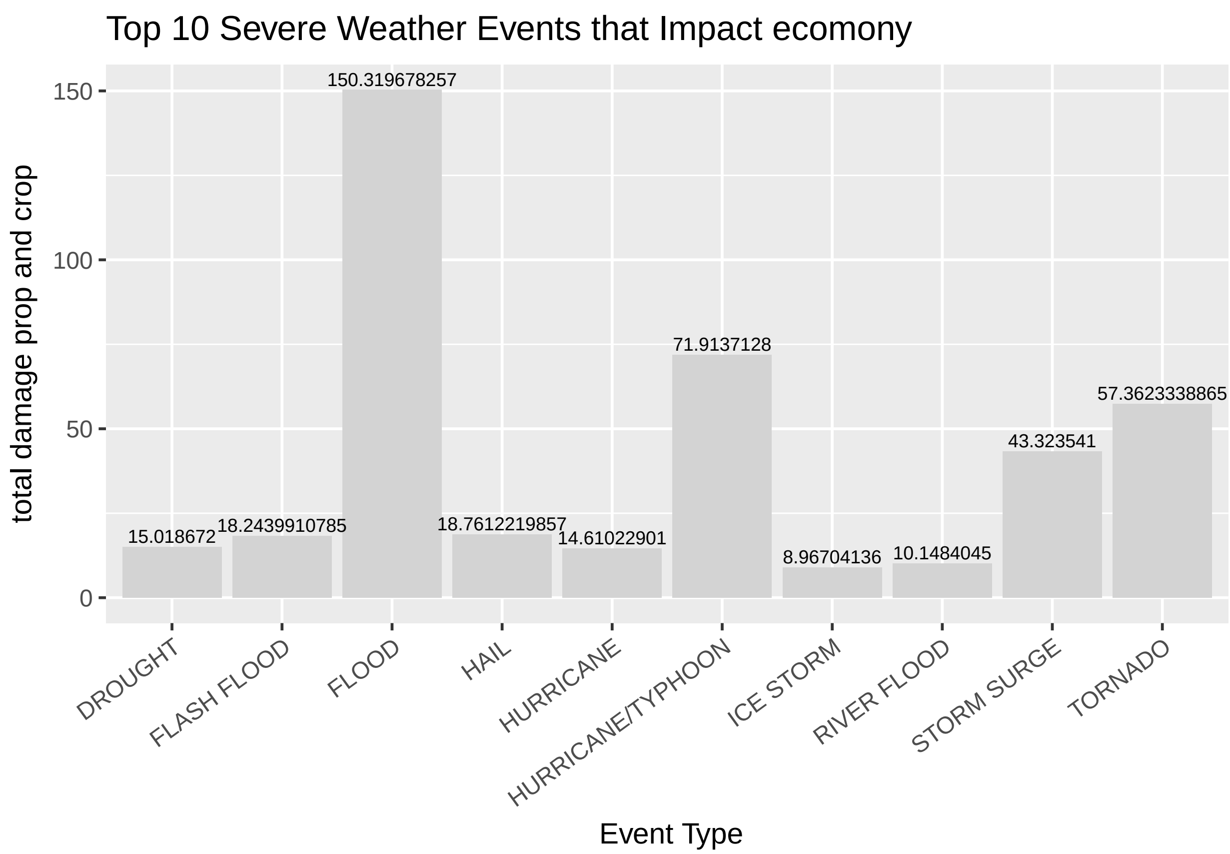

```
#top 10 event type that causes injuries
topInjury<-injuryByEvtype%>%head(10)
ggplot(data=topInjury, aes(x=EVTYPE, y=injuries)) +
geom_bar(stat="identity",fill="lightgray") + xlab("Weather Event Type") + ylab("Injuries") +
ggtitle("Top 10 Severe Weather Events that cause injuries") + geom_text(aes(label=injuries), vjust=-0.2
```



Economic impact of severe weather events

```
#top  event type that impact economy
topdamage<-damageByEvtype%>%head(10)

#plot Top 10 Severe Weather Events that Impact economy
ggplot(data=topdamage, aes(x=EVTYPE, y=totalDamage )) +
geom_bar(stat="identity",fill="lightgray") + xlab(" Event Type") + ylab("total damage prop and crop") +
ggtitle("Top 10 Severe Weather Events that Impact ecomony") + geom_text(aes(label=totalDamage), vjust=-(
```



Conclusion

Tornadoes are the most devastating.They cause the largest number of deaths(5633) and injuries(91346).

Flood events have the largest total economic impact.They costed 150.3 billion dolllars. hurricanes/typhoons costed 71.9 billions of dollars,tornadoes costed 57.4 billion of dollars.