# Risk analysis of the NOAA Storm Database

This is a analysis for the Coursera course "Reprodcible Research" (Peer Assessment 2).

## Synopsis

The basic goal of this assignment is to explore the National Oceanic and Atmospheric Administration's (NOAA) Storm Database and answer some basic questions about severe weather events. The data analysis wants to address two main questions: (i) Across the United States, which types of events are most harmful with respect to population health? (ii) Across the United States, which types of events have the greatest economic consequences? The data analysis demonstrates that Tornados cause most damage to humans and property, whereas hail is most detrimental for crops.

## Data processing

The U.S. NOAA Storm database is downloaded from the public server and read into a table. The column "REMARKS" is removed from further analysis, since it contains free text information that blows up table size and makes displaying cumbersome. The resulting object is stored for faster access.

```
setwd("~/workspace/coursera")
suppressPackageStartupMessages(library(data.table))
suppressPackageStartupMessages(library(gplots))

noaa_link= "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
noaa_file= "repData_peer2/noaa_stormdata.csv.bz2"
noaa_bak= "repData_peer2/noaa_table.RDS"

if(!file.exists(noaa_bak)){
    download.file(noaa_link, destfile=noaa_file)
    colidx_remarks= 36
    noaa_table= as.data.table(read.csv(noaa_file))[,setdiff(1:37,colidx_remarks), with=F]
    saveRDS(noaa_table, file=noaa_bak)
}else{
    noaa_table= readRDS(noaa_bak)
}
```

In this analysis I want to focus on human, property and crop damage, therefore the table is restricted to the following columns: EVTYPE (weather events), FATALITIES (number of fatalities), INJURIES (number of injuries), PROPDMG (property damage), CROPDMG (crop damage).

```
dmg_noaa_table= noaa_table[, c('EVTYPE', 'FATALITIES', 'INJURIES', 'PROPDMG', 'CROPDMG'), with=F]
dmg_noaa_table
```

```
##             EVTYPE FATALITIES INJURIES PROPDMG CROPDMG
##      1:    TORNADO          0       15    25.0       0
##      2:    TORNADO          0        0     2.5       0
##      3:    TORNADO          0        2    25.0       0
##      4:    TORNADO          0        2     2.5       0
```

```
##      5:    TORNADO          0        2     2.5       0
##     ---
## 902293:  HIGH WIND          0        0     0.0       0
## 902294:  HIGH WIND          0        0     0.0       0
## 902295:  HIGH WIND          0        0     0.0       0
## 902296:   BLIZZARD          0        0     0.0       0
## 902297: HEAVY SNOW          0        0     0.0       0
```

Generate basic statistics about the data. Which events are documented? How frequent are they? What is distribution of all damages, i.e. uniform or normally distributed?

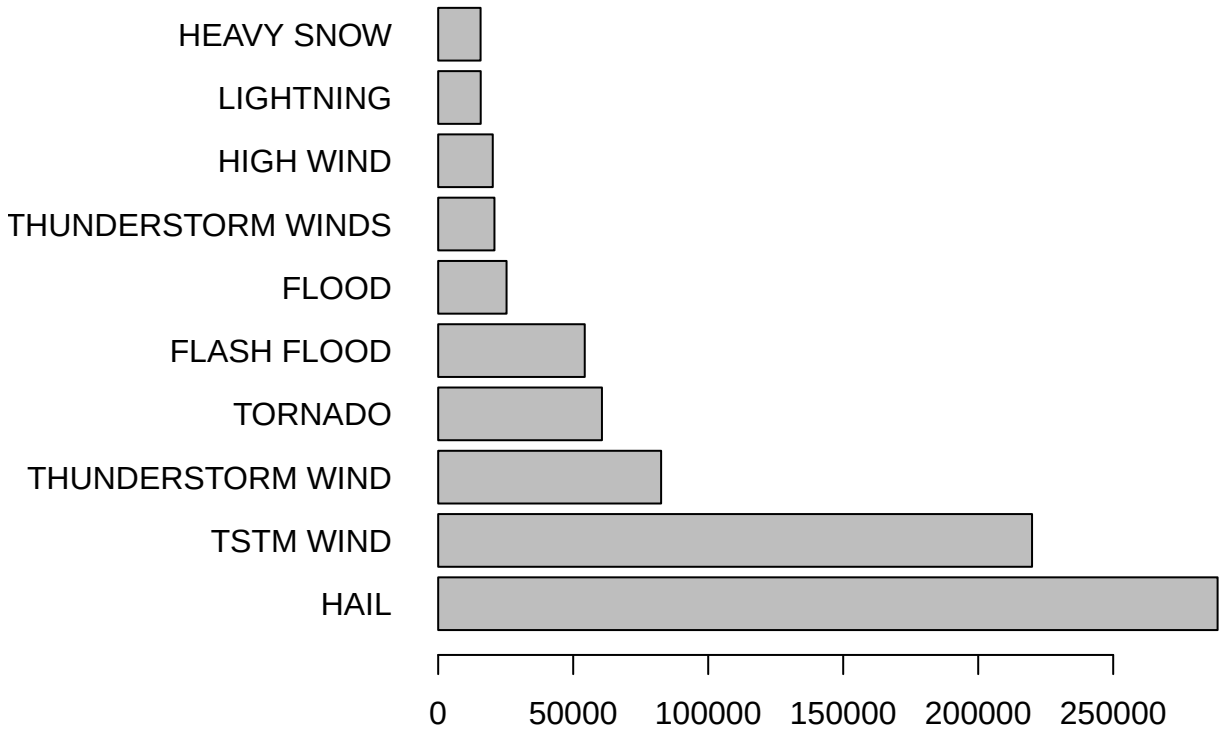
```
num_evtypes= length(unique(dmg_noaa_table$EVTYPE))
```

The table contains 985 different weather events.

## Results

What are the ten most frequent events?

```
evcounts= sort(table(dmg_noaa_table$EVTYPE), decreasing = T)
par(mar=c(4,11,1,1))
barplot2(evcounts[1:10], horiz=T, las=1)
```



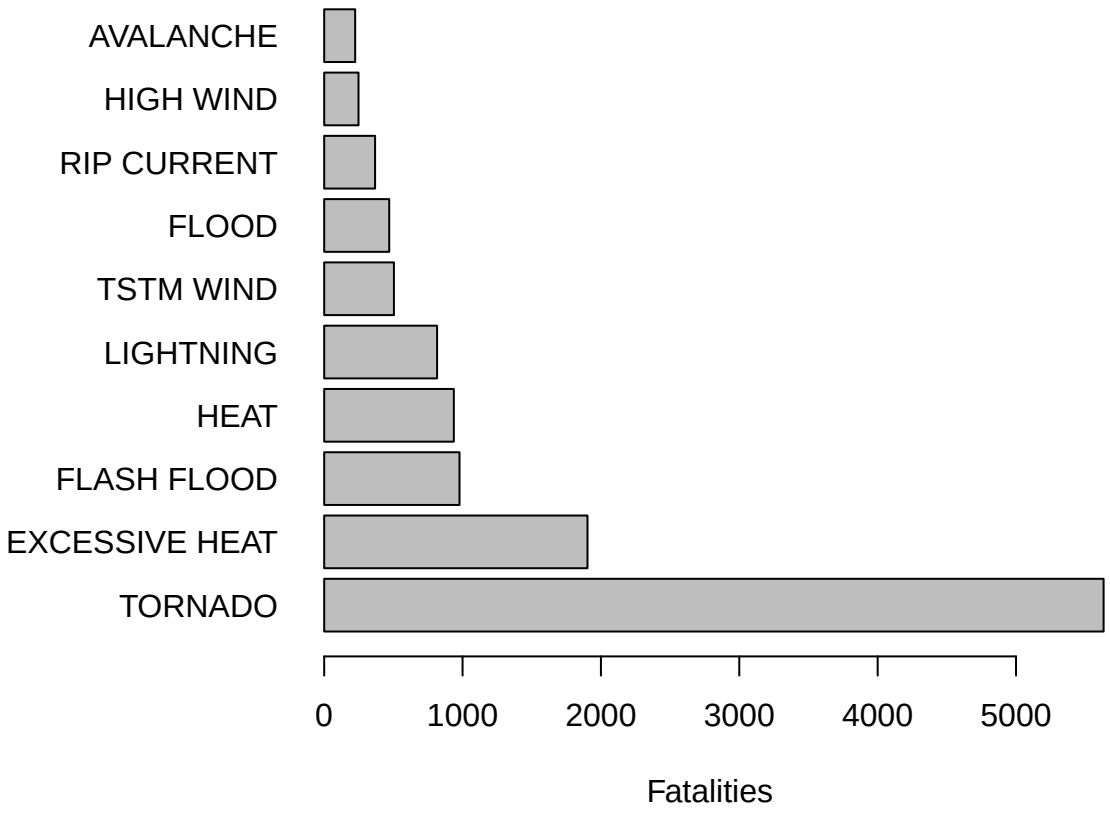
**Across the United States, which types of events are most harmful with respect to population health?**

```
dmg_noaa_table[,c('sum_fat', 'sum_inj'):=list(sum(FATALITIES), sum(INJURIES)), by=EVTYPE]
```

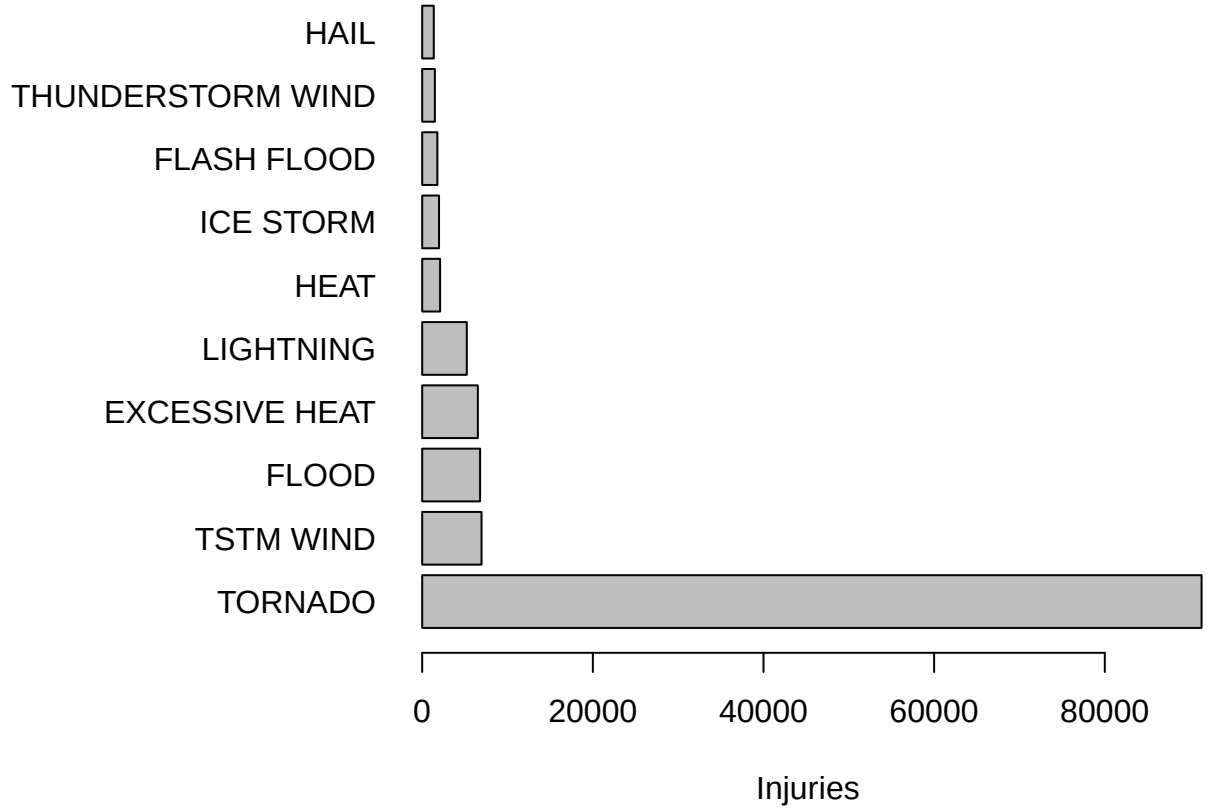
```
##             EVTYPE FATALITIES INJURIES PROPDMG CROPDMG sum_fat sum_inj
##      1:    TORNADO          0       15    25.0       0    5633   91346
##      2:    TORNADO          0        0     2.5       0    5633   91346
##      3:    TORNADO          0        2    25.0       0    5633   91346
##      4:    TORNADO          0        2     2.5       0    5633   91346
##      5:    TORNADO          0        2     2.5       0    5633   91346
##     ---
## 902293:  HIGH WIND          0        0     0.0       0     248    1137
## 902294:  HIGH WIND          0        0     0.0       0     248    1137
## 902295:  HIGH WIND          0        0     0.0       0     248    1137
## 902296:   BLIZZARD          0        0     0.0       0     101     805
## 902297: HEAVY SNOW          0        0     0.0       0     127    1021
```

```
human_dmg= unique(dmg_noaa_table[,list(EVTYPE, sum_fat, sum_inj)])

top_fat= human_dmg[order(-sum_fat)][1:10]
par(mar=c(4,11,1,1))
barplot2( top_fat[['sum_fat']], names.arg = top_fat[['EVTYPE']],
          horiz=T, las=1, xlab='Fatalities')
```



```
top_inj= human_dmg[order(-sum_inj)][1:10]
par(mar=c(4,11,1,1))
barplot2( top_inj[['sum_inj']], names.arg = top_inj[['EVTYPE']],
          horiz=T, las=1, xlab='Injuries')
```




During Tornados most people dieand get injured compared to the other weather events in the database.

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

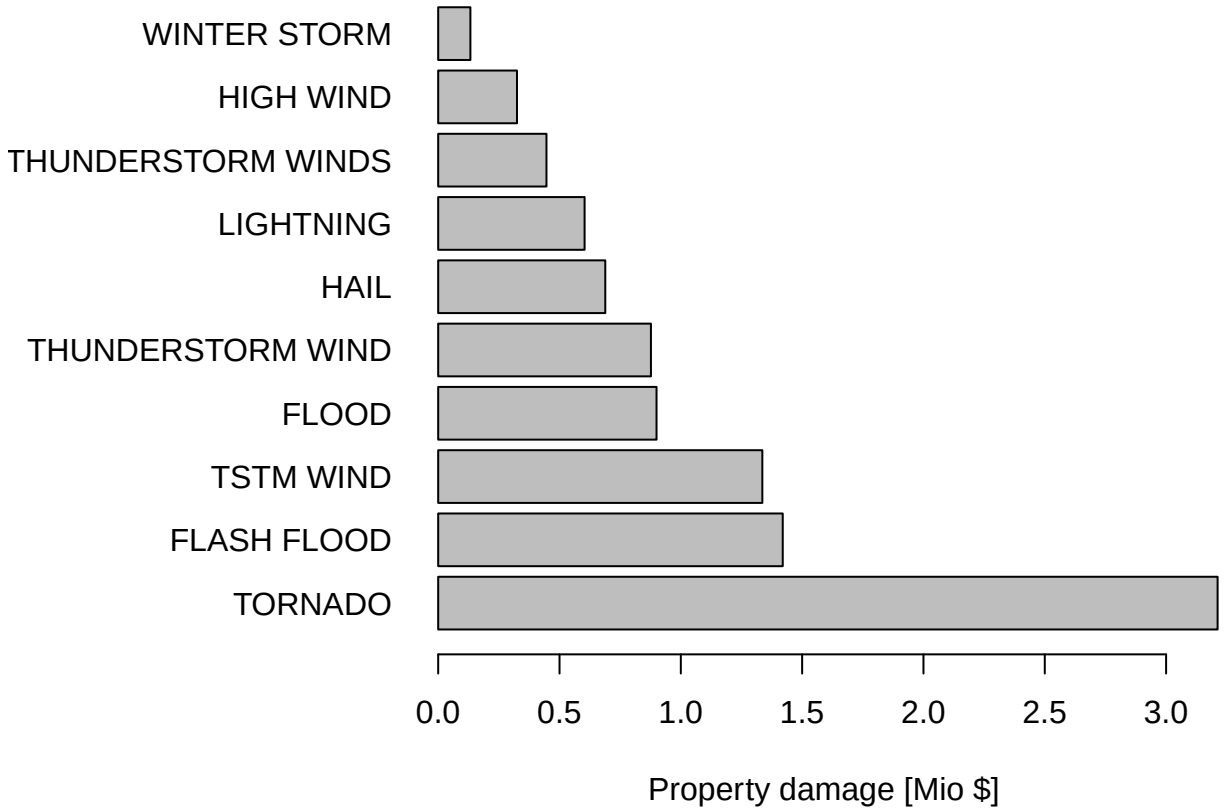
```
dmg_noaa_table[,c('sum_prop', 'sum_crop'):=list(sum(PROPDMG), sum(CROPDMG)), by=EVTYPE]
```

```
##             EVTYPE FATALITIES INJURIES PROPDMG CROPDMG sum_fat sum_inj
##      1:    TORNADO          0       15    25.0       0    5633   91346
##      2:    TORNADO          0        0     2.5       0    5633   91346
##      3:    TORNADO          0        2    25.0       0    5633   91346
##      4:    TORNADO          0        2     2.5       0    5633   91346
##      5:    TORNADO          0        2     2.5       0    5633   91346
##     ---
## 902293:  HIGH WIND          0        0     0.0       0     248    1137
## 902294:  HIGH WIND          0        0     0.0       0     248    1137
## 902295:  HIGH WIND          0        0     0.0       0     248    1137
## 902296:   BLIZZARD          0        0     0.0       0     101     805
```

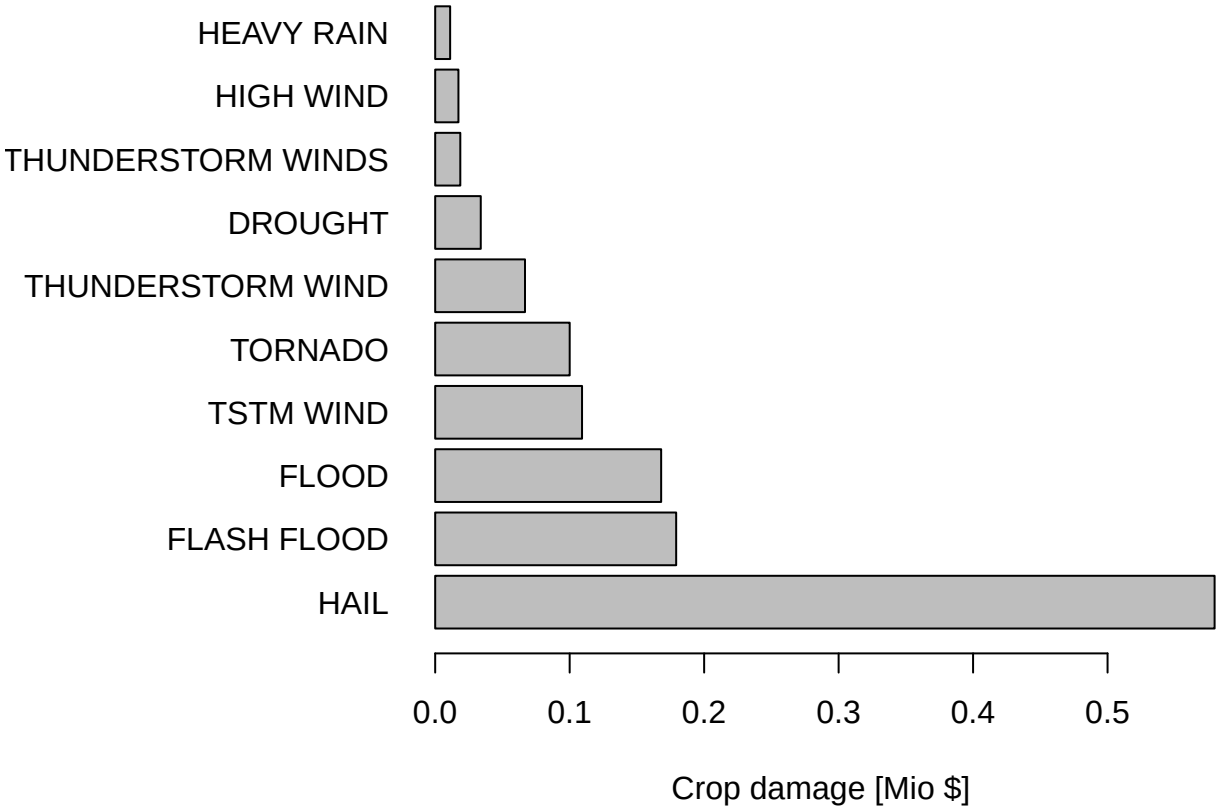
```
## 902297: HEAVY SNOW          0        0     0.0       0     127    1021
##           sum_prop  sum_crop
##      1: 3212258.16 100018.52
##      2: 3212258.16 100018.52
##      3: 3212258.16 100018.52
##      4: 3212258.16 100018.52
##      5: 3212258.16 100018.52
##     ---
## 902293:  324731.56  17283.21
## 902294:  324731.56  17283.21
## 902295:  324731.56  17283.21
## 902296:   25318.48    172.00
## 902297:  122251.99   2165.72
```

```
econ_dmg= unique(dmg_noaa_table[,list(EVTYPE, sum_prop, sum_crop)])

top_prop= econ_dmg[order(-sum_prop)][1:10]
par(mar=c(4,11,1,1))
barplot2( top_prop[['sum_prop']]/1e6, names.arg = top_prop[['EVTYPE']],
          horiz=T, las=1, xlab='Property damage [Mio $]')
```




```
top_crop= econ_dmg[order(-sum_crop)][1:10]
par(mar=c(4,11,1,1))
barplot2( top_crop[['sum_crop']]/1e6, names.arg = top_crop[['EVTYPE']],
          horiz=T, las=1, xlab='Crop damage [Mio $]')
```




During Tornados the highest property damage is caused compared to the other weather events in the database. During Hail the highest crop damage is caused compared to the other weather events in the database.

## Discussion

To get more fine grained statistics you could normalize the damages for each event by the frequency of the event. This could result in events to prepare for with the highest impact on damage reduction.