Damage analysis based on 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, whereas drought is most detrimental for crops and floods cause most damage at property.

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. Guidelines how to prepare storm data are available here: https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf .

```
setwd("~/workspace/coursera/repData_peer2/")
suppressPackageStartupMessages(library(data.table))
suppressPackageStartupMessages(library(gplots))

noaa_link= "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
noaa_file= "noaa_stormdata.csv.bz2"
noaa_bak= "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), PROPDMGEXP (property damage exponent), CROPDMGEXP (crop damage expponent). Alphabetical characters used to signify magnitude include "K" for thousands, "M" for millions, and "B" for billions (page 12, section 2.7 Damage from the Storm data preparation guidelines). I will compute new columns representing the absolute dollar amount of damage. The symbol h/H should correspond to 10^2. For numeric exponents base 10 is assumed. For -/+/? no exponent is assumed. The translation from "symbol" to "exp" was encoded by hand.

```
dmg_noaa_table= noaa_table[, c('EVTYPE',
                               'FATALITIES', 'INJURIES',
                               'PROPDMG', 'CROPDMG', 'PROPDMGEXP', 'CROPDMGEXP'), with=F]
# get all exponents
dmg_exp= data.table(symbol=sort(with(dmg_noaa_table,
                                     union( unique(CROPDMGEXP), unique(PROPDMGEXP) )
                                     )),
                    exp=c(rep(0,4), 0:8, 9, 2,2, 3,3,6,6)
                    )
dmg_exp
```

```
##     symbol exp
##  1:          0
##  2:      -   0
##  3:      ?   0
##  4:      +   0
##  5:      0   0
##  6:      1   1
##  7:      2   2
##  8:      3   3
##  9:      4   4
## 10:      5   5
## 11:      6   6
## 12:      7   7
## 13:      8   8
## 14:      B   9
## 15:      h   2
## 16:      H   2
## 17:      k   3
## 18:      K   3
## 19:      m   6
## 20:      M   6
```

```
# compute absolute dollar damage
dmg_noaa_table[, c('prop_dollar', 'crop_dollar'):=
                    list(PROPDMG * 10 ^ dmg_exp[match(PROPDMGEXP, symbol), exp],
                         CROPDMG * 10 ^ dmg_exp[match(CROPDMGEXP, symbol), exp]
                         )
               ]
```

```
##             EVTYPE FATALITIES INJURIES PROPDMG CROPDMG PROPDMGEXP
##      1:    TORNADO          0       15    25.0       0          K
##      2:    TORNADO          0        0     2.5       0          K
##      3:    TORNADO          0        2    25.0       0          K
##      4:    TORNADO          0        2     2.5       0          K
##      5:    TORNADO          0        2     2.5       0          K
##     ---
## 902293:  HIGH WIND          0        0     0.0       0          K
## 902294:  HIGH WIND          0        0     0.0       0          K
## 902295:  HIGH WIND          0        0     0.0       0          K
## 902296:   BLIZZARD          0        0     0.0       0          K
## 902297: HEAVY SNOW          0        0     0.0       0          K
##         CROPDMGEXP prop_dollar crop_dollar
```

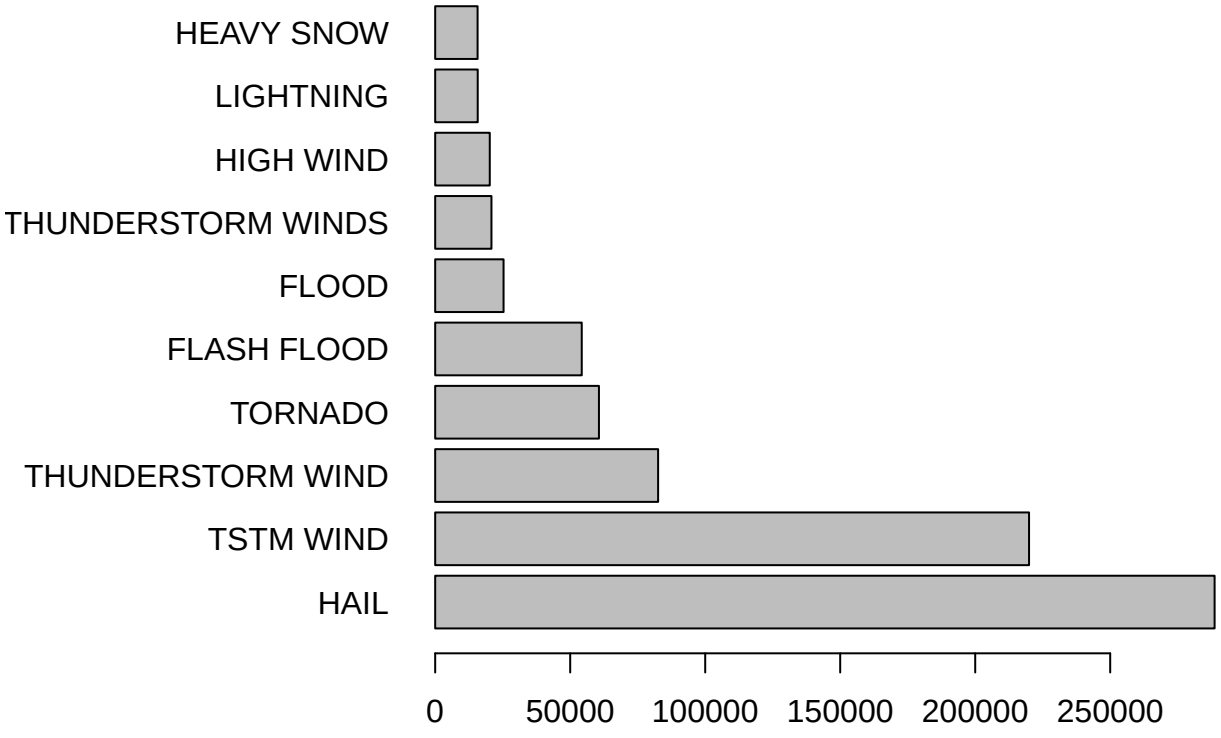
```
##      1:                  25000           0
##      2:                   2500           0
##      3:                  25000           0
##      4:                   2500           0
##      5:                   2500           0
##     ---
## 902293:          K           0           0
## 902294:          K           0           0
## 902295:          K           0           0
## 902296:          K           0           0
## 902297:          K           0           0
```

Results

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

The table contains 985 different weather events. 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)
```



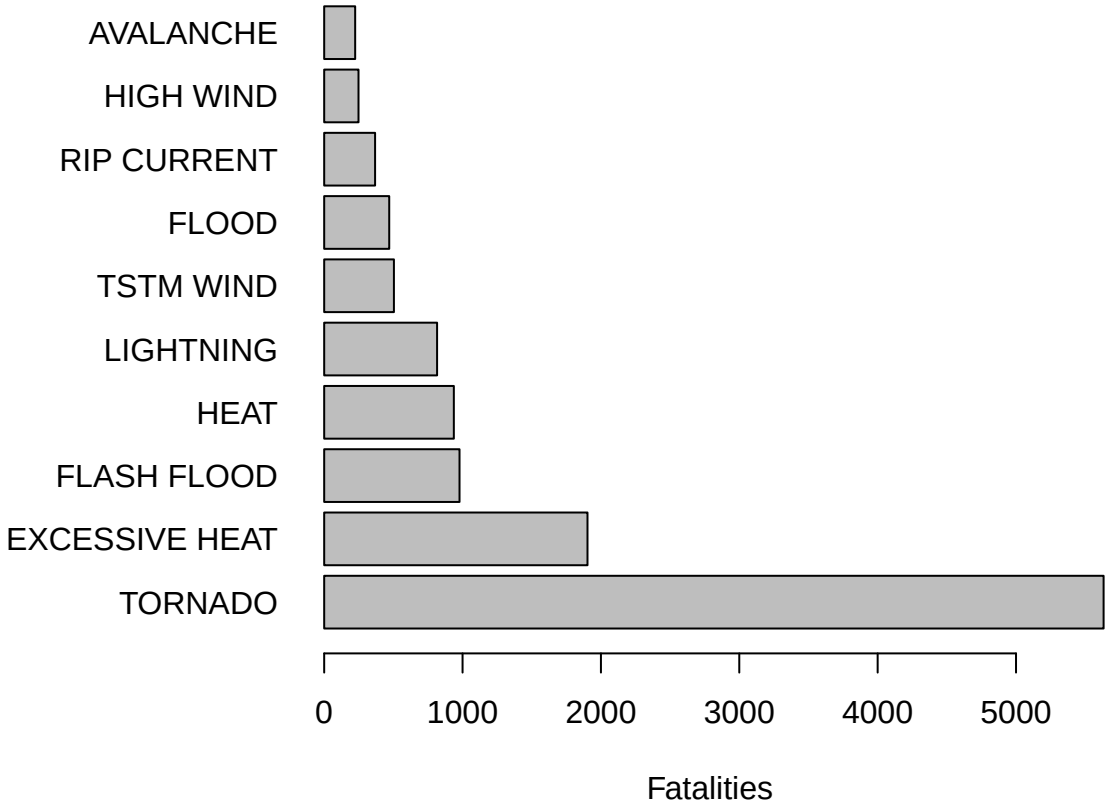
Across the United States, which types of events are most harmful with respect to population health?

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

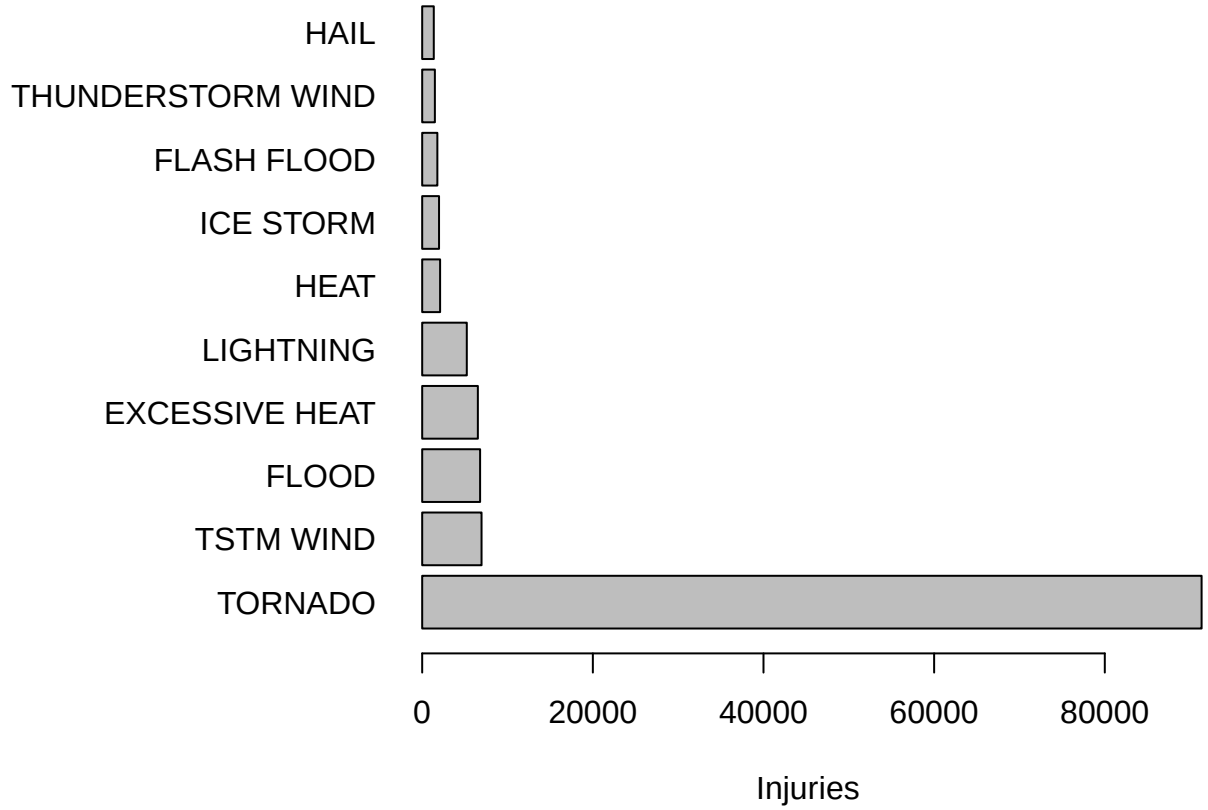
```
##             EVTYPE FATALITIES INJURIES PROPDMG CROPDMG PROPDMGEXP
##      1:    TORNADO          0       15    25.0       0          K
##      2:    TORNADO          0        0     2.5       0          K
##      3:    TORNADO          0        2    25.0       0          K
##      4:    TORNADO          0        2     2.5       0          K
##      5:    TORNADO          0        2     2.5       0          K
##     ---
## 902293:  HIGH WIND          0        0     0.0       0          K
## 902294:  HIGH WIND          0        0     0.0       0          K
## 902295:  HIGH WIND          0        0     0.0       0          K
## 902296:   BLIZZARD          0        0     0.0       0          K
## 902297: HEAVY SNOW          0        0     0.0       0          K
##         CROPDMGEXP prop_dollar crop_dollar sum_fat sum_inj
##      1:                 25000           0    5633   91346
##      2:                  2500           0    5633   91346
##      3:                 25000           0    5633   91346
##      4:                  2500           0    5633   91346
##      5:                  2500           0    5633   91346
##     ---
## 902293:          K           0           0     248    1137
## 902294:          K           0           0     248    1137
## 902295:          K           0           0     248    1137
## 902296:          K           0           0     101     805
## 902297:          K           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 die and get injured compared to the other weather events in the database.

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

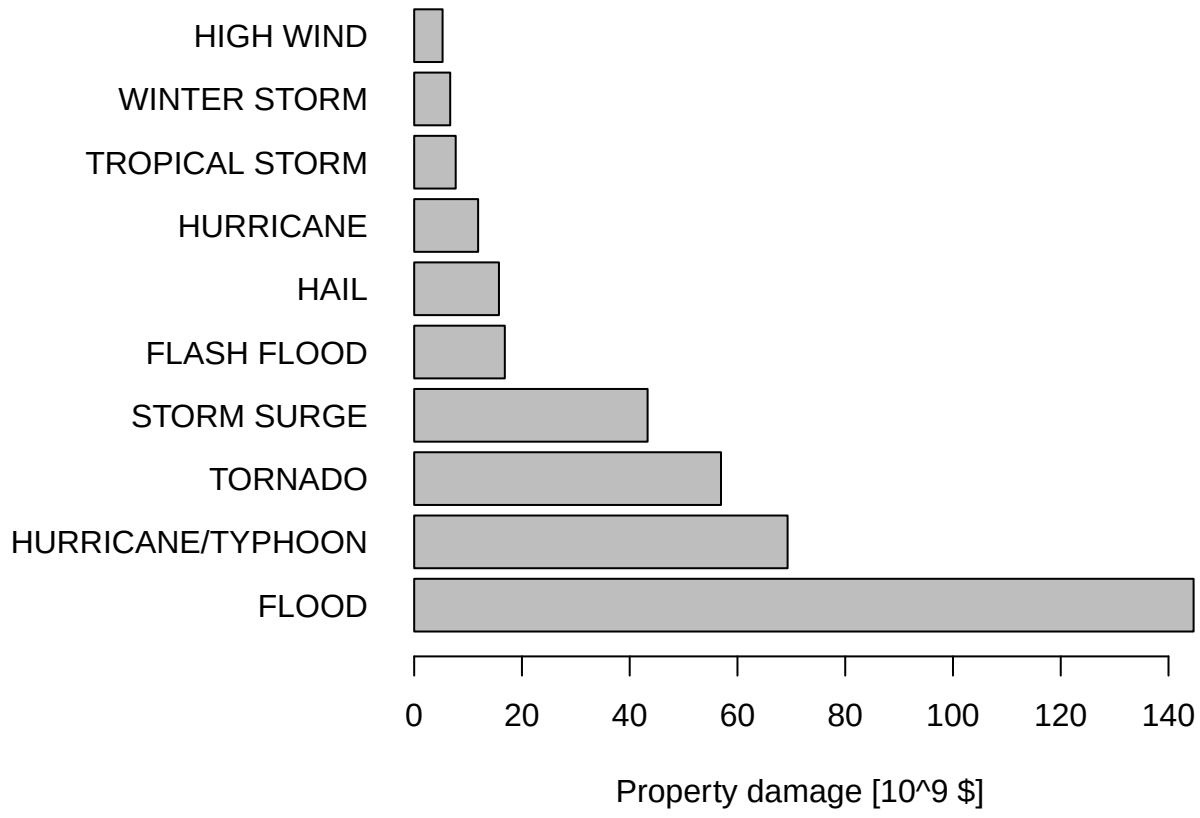
```
dmg_noaa_table[,c('sum_prop', 'sum_crop'):=list(sum(prop_dollar), sum(crop_dollar)), by=EVTYPE]
```

```
##             EVTYPE FATALITIES INJURIES PROPDMG CROPDMG PROPDMGEXP
##      1:    TORNADO          0       15    25.0       0          K
##      2:    TORNADO          0        0     2.5       0          K
##      3:    TORNADO          0        2    25.0       0          K
##      4:    TORNADO          0        2     2.5       0          K
##      5:    TORNADO          0        2     2.5       0          K
##     ---
## 902293:  HIGH WIND          0        0     0.0       0          K
## 902294:  HIGH WIND          0        0     0.0       0          K
## 902295:  HIGH WIND          0        0     0.0       0          K
## 902296:   BLIZZARD          0        0     0.0       0          K
## 902297: HEAVY SNOW          0        0     0.0       0          K
##         CROPDMGEXP prop_dollar crop_dollar sum_fat sum_inj    sum_prop
##      1:                 25000           0    5633   91346 56947380676
##      2:                  2500           0    5633   91346 56947380676
##      3:                 25000           0    5633   91346 56947380676
##      4:                  2500           0    5633   91346 56947380676
```

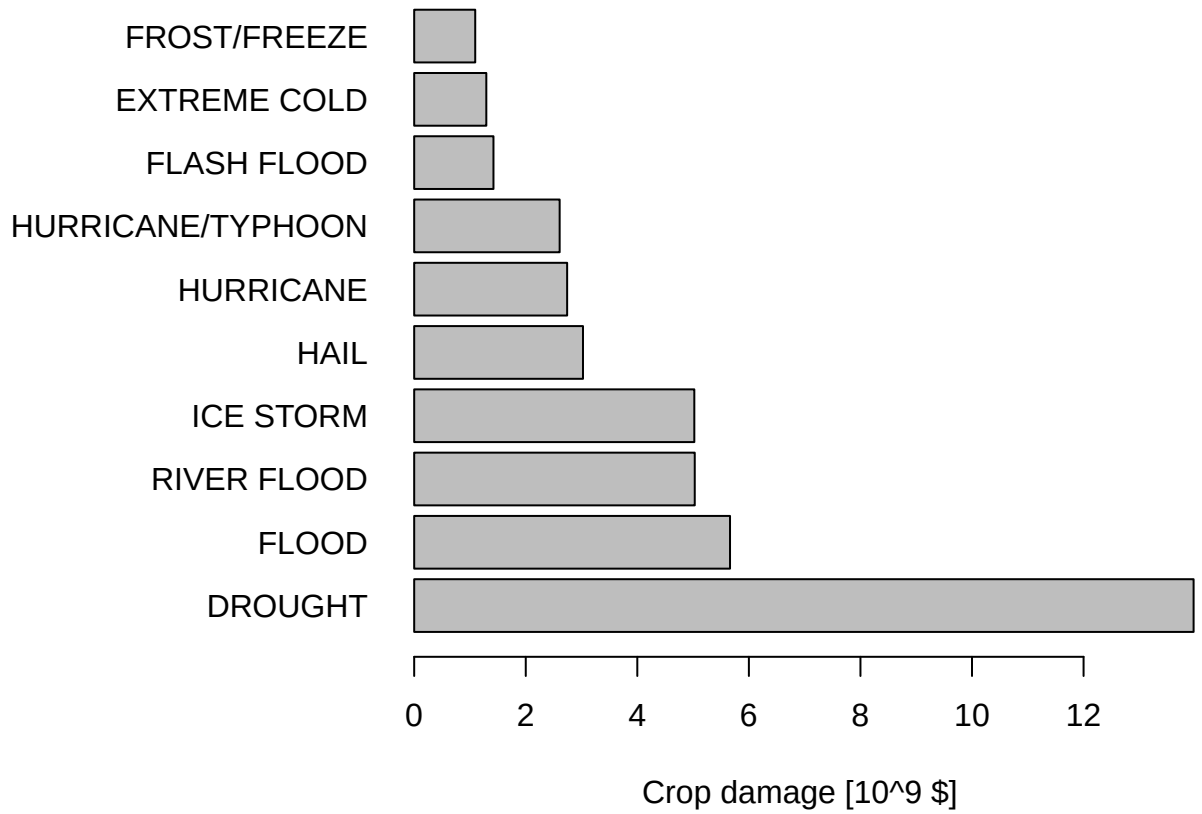
```
##      5:                  2500           0     5633    91346 56947380676
##     ---
## 902293:          K           0           0      248     1137  5270046295
## 902294:          K           0           0      248     1137  5270046295
## 902295:          K           0           0      248     1137  5270046295
## 902296:          K           0           0      101      805   659213950
## 902297:          K           0           0      127     1021   932759140
##          sum_crop
##      1: 414953270
##      2: 414953270
##      3: 414953270
##      4: 414953270
##      5: 414953270
##     ---
## 902293: 638571300
## 902294: 638571300
## 902295: 638571300
## 902296: 112060000
## 902297: 134653100
```

```
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']]/1e9, names.arg = top_prop[['EVTYPE']],
          horiz=T, las=1, xlab='Property damage [10^9 $]')
```



```
top_crop= econ_dmg[order(-sum_crop)][1:10]
par(mar=c(4,11,1,1))
barplot2( top_crop[['sum_crop']]/1e9, names.arg = top_crop[['EVTYPE']],
          horiz=T, las=1, xlab='Crop damage [10^9 $]')
```



During floods the highest property damage is caused compared to the other weather events in the database. During droughts 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.