US Weather Event Data Analysis

Greg D

24 August 2018

Human Health and Property Damage of Weather Event Types Across the United States

Synopsis

Weather evens can cause significant human harm and economic damage. Health effects and economic cost were studied by weather event type using data from the NOAA Storm Database.

Data Processing

The database is downloaded from the source, extracted and loaded into R as a data object 'rawdata':

```
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file (url, destfile = "data.csv.bz2", method="curl")
rawdata <- read.csv("data.csv.bz2", sep=",", header=TRUE, na.strings = c("NA","",'#DIV/0!'))
```

Quick inspection of data file to ensure the above worked correctly and get an idea of what the data looks like:

```
dim(rawdata)
```

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

```
head(rawdata)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 2 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 3 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 4 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 5 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
## 6 TORNADO         0    <NA>       <NA>     <NA>     <NA>          0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0    <NA>       <NA>   14.0   100 3   0          0
## 2         NA         0    <NA>       <NA>    2.0   150 2   0          0
## 3         NA         0    <NA>       <NA>    0.1   123 2   0          0
## 4         NA         0    <NA>       <NA>    0.0   100 2   0          0
## 5         NA         0    <NA>       <NA>    0.0   150 2   0          0
## 6         NA         0    <NA>       <NA>    1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP  WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0       <NA> <NA>       <NA>      <NA>
## 2        0     2.5          K       0       <NA> <NA>       <NA>      <NA>
## 3        2    25.0          K       0       <NA> <NA>       <NA>      <NA>
## 4        2     2.5          K       0       <NA> <NA>       <NA>      <NA>
## 5        2     2.5          K       0       <NA> <NA>       <NA>      <NA>
## 6        6     2.5          K       0       <NA> <NA>       <NA>      <NA>
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806    <NA>      1
## 2     3042      8755          0          0    <NA>      2
## 3     3340      8742          0          0    <NA>      3
## 4     3458      8626          0          0    <NA>      4
## 5     3412      8642          0          0    <NA>      5
## 6     3450      8748          0          0    <NA>      6
```

This study looks only at population health and economic consequences of weather events by event type. Therefore only a subset of the columns are required.

```
procdata <- rawdata[, c("EVTYPE", "FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CROPDMG", "CRO
PDMGEXP")]
head(procdata)
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0       15    25.0          K       0       <NA>
## 2 TORNADO          0        0     2.5          K       0       <NA>
## 3 TORNADO          0        2    25.0          K       0       <NA>
## 4 TORNADO          0        2     2.5          K       0       <NA>
## 5 TORNADO          0        2     2.5          K       0       <NA>
## 6 TORNADO          0        6     2.5          K       0       <NA>
```

Note that both PROPDMG and CROPDMG have a multiplier column. These multipliers are codified, so they first must be converted to real numbers.

```
unique(procdata$PROPDMGEXP)
```

```
##  [1] K    M    <NA> B    m    +    0    5    6    ?    4    2    3    h
## [15] 7    H    -    1    8
## Levels: - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
```

```
unique(procdata$CROPDMGEXP)
```

```
## [1] <NA> M    K    m    B    ?    0    k    2
## Levels: ? 0 2 B k K m M
```

```
procdata$PDMULTIPLIER <- 0

procdata$PDMULTIPLIER[procdata$PROPDMGEXP %in% c('h','H')] <- 100
procdata$PDMULTIPLIER[procdata$PROPDMGEXP %in% c('k','K')] <- 1000
procdata$PDMULTIPLIER[procdata$PROPDMGEXP %in% c('m','M')] <- 1e+06
procdata$PDMULTIPLIER[procdata$PROPDMGEXP %in% c('b','B')] <- 1e+09
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 1] <- 1e+1
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 2] <- 1e+2
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 3] <- 1e+3
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 4] <- 1e+4
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 5] <- 1e+5
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 6] <- 1e+6
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 7] <- 1e+7
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 8] <- 1e+8
procdata$PDMULTIPLIER[is.na(procdata$PROPDMGEXP)] <- 1

procdata$PROPDAMAGEVALUE <- procdata$PROPDMG*procdata$PDMULTIPLIER


procdata$CDMULTIPLIER[procdata$CROPDMGEXP %in% c('h','H')] <- 100
procdata$CDMULTIPLIER[procdata$CROPDMGEXP %in% c('k','K')] <- 1000
procdata$CDMULTIPLIER[procdata$CROPDMGEXP %in% c('m','M')] <- 1e+06
procdata$CDMULTIPLIER[procdata$CROPDMGEXP %in% c('b','B')] <- 1e+09
procdata$CDMULTIPLIER[procdata$CROPDMGEXP == 0] <- 1
procdata$PDMULTIPLIER[procdata$PROPDMGEXP == 2] <- 1e+2
procdata$CDMULTIPLIER[is.na(procdata$CROPDMGEXP)] <- 1

procdata$CROPDAMAGEVALUE <- procdata$CROPDMG*procdata$CDMULTIPLIER
```

```
summary(procdata)
```

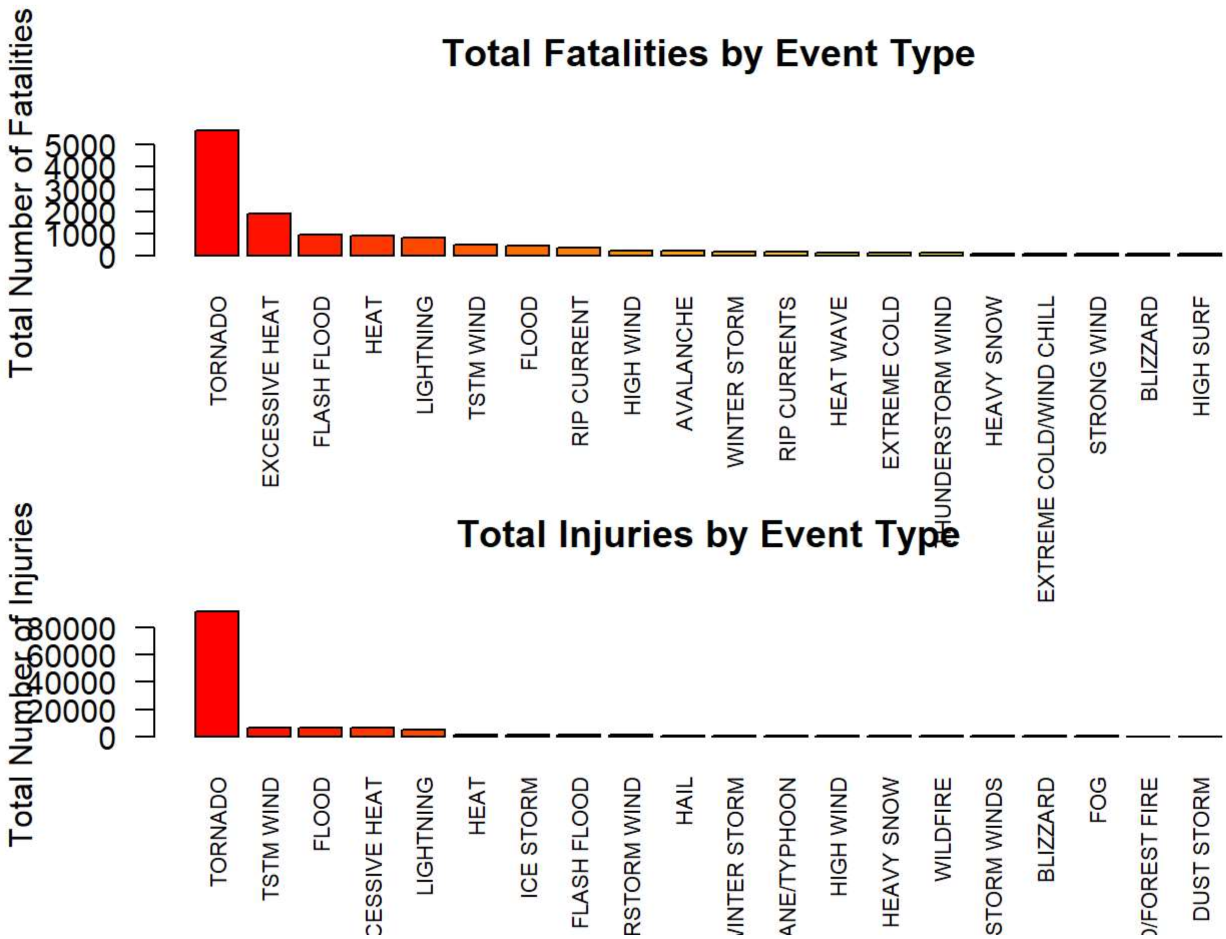
```
##                EVTYPE         FATALITIES          INJURIES        
##  HAIL             :288661   Min.   :  0.0000   Min.   :   0.0000  
##  TSTM WIND        :219940   1st Qu.:  0.0000   1st Qu.:   0.0000  
##  THUNDERSTORM WIND: 82563   Median :  0.0000   Median :   0.0000  
##  TORNADO          : 60652   Mean   :  0.0168   Mean   :   0.1557  
##  FLASH FLOOD      : 54277   3rd Qu.:  0.0000   3rd Qu.:   0.0000  
##  FLOOD            : 25326   Max.   :583.0000   Max.   :1700.0000  
##  (Other)          :170878                                         
##     PROPDMG          PROPDMGEXP        CROPDMG          CROPDMGEXP    
##  Min.   :   0.00   K      :424665   Min.   :  0.000   K      :281832  
##  1st Qu.:   0.00   M      : 11330   1st Qu.:  0.000   M      :  1994  
##  Median :   0.00   0      :   216   Median :  0.000   k      :    21  
##  Mean   :  12.06   B      :    40   Mean   :  1.527   0      :    19  
##  3rd Qu.:   0.50   5      :    28   3rd Qu.:  0.000   B      :     9  
##  Max.   :5000.00   (Other):    84   Max.   :990.000   (Other):     9  
##                    NA's   :465934                     NA's   :618413  
##   PDMULTIPLIER       PROPDAMAGEVALUE      CDMULTIPLIER    
##  Min.   :0.000e+00   Min.   :0.000e+00   Min.   :1.00e+00  
##  1st Qu.:1.000e+00   1st Qu.:0.000e+00   1st Qu.:1.00e+00  
##  Median :1.000e+00   Median :0.000e+00   Median :1.00e+00  
##  Mean   :5.754e+04   Mean   :4.746e+05   Mean   :1.25e+04  
##  3rd Qu.:1.000e+03   3rd Qu.:5.000e+02   3rd Qu.:1.00e+03  
##  Max.   :1.000e+09   Max.   :1.150e+11   Max.   :1.00e+09  
##                                          NA's   :8         
##  CROPDAMAGEVALUE    
##  Min.   :0.000e+00  
##  1st Qu.:0.000e+00  
##  Median :0.000e+00  
##  Mean   :5.442e+04  
##  3rd Qu.:0.000e+00  
##  Max.   :5.000e+09  
##  NA's   :8
```

Results

Population Harm

Population harm, in terms of both fatalities and injuries, was investigated by event type.

```
dt <- data.table(procdata)

totFatalities <- aggregate(FATALITIES ~ EVTYPE, data=procdata, FUN=sum)
totFatalities <- arrange(totFatalities, desc(FATALITIES))

totInjuries <- aggregate(INJURIES ~ EVTYPE, data=procdata, FUN=sum)
totInjuries <- arrange(totInjuries, desc(INJURIES))


par(mfrow=c(2,1))
barplot(totFatalities$FATALITIES[1:20], col =heat.colors(20), names.arg=totFatalities$EVTYPE[1:2
0], cex.names=0.7, las=2, main="Total Fatalities by Event Type", ylab="Total Number of Fatalitie
s")

barplot(totInjuries$INJURIES[1:20], col =heat.colors(20), names.arg=totInjuries$EVTYPE[1:20], ce
x.names=0.7, las=2, main="Total Injuries by Event Type", ylab="Total Number of Injuries")
```



Tornados were responsible for both the highest number of fatalities and the highest number of injuries over the studied period.

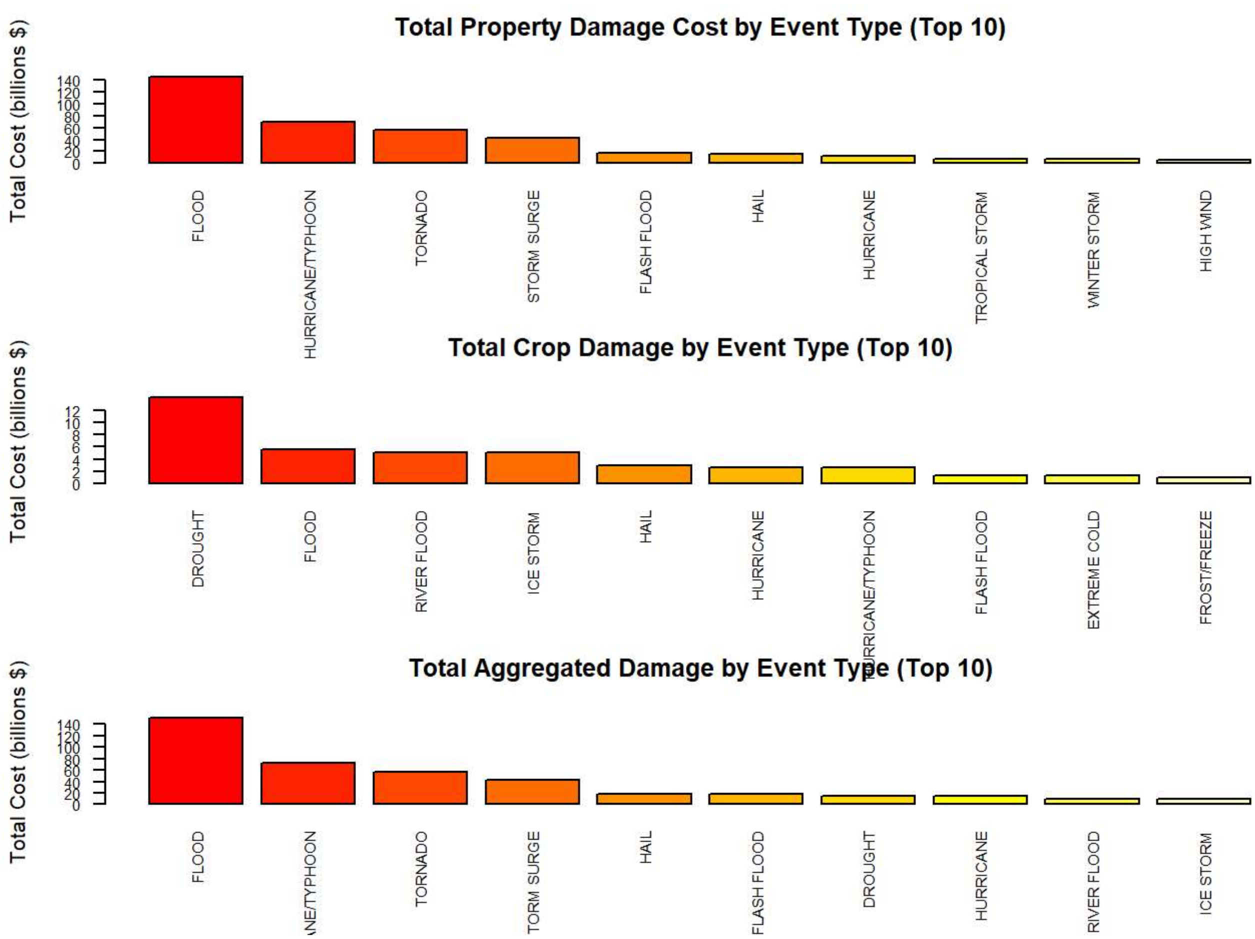
Economic Cost:

As for population harm, economic cost was investigated by event type.

The dataset separates economic damage into property damage ('PROPDMG') and crop damage ('CROPDMG').

Economic Damage

```
totPropDamage <- aggregate(PROPDAMAGEVALUE ~ EVTYPE, data=procdata, FUN=sum)
totPropDamage <- arrange(totPropDamage, desc(PROPDAMAGEVALUE))

totCropDamage <- aggregate(CROPDAMAGEVALUE ~ EVTYPE, data=procdata, FUN=sum)
totCropDamage <- arrange(totCropDamage, desc(CROPDAMAGEVALUE))

totAggDamage <- aggregate(PROPDAMAGEVALUE+CROPDAMAGEVALUE ~ EVTYPE, data=procdata, FUN=sum)
names(totAggDamage)[2]="AGGDAMAGEVALUE"
totAggDamage <- arrange(totAggDamage, desc(AGGDAMAGEVALUE))

par(mfrow=c(3,1))
barplot(totPropDamage$PROPDAMAGEVALUE[1:10]/1e+9, col=heat.colors(10), names.arg=totPropDamage$E
VTYPE[1:10], cex.names=0.7, cex.axis=0.7, las=2, main="Total Property Damage Cost by Event Type
 (Top 10)", ylab="Total Cost (billions $)")

barplot(totCropDamage$CROPDAMAGEVALUE[1:10]/1e+9, col=heat.colors(10), names.arg=totCropDamage$E
VTYPE[1:10], cex.names=0.7, cex.axis=0.7, las=2, main="Total Crop Damage by Event Type (Top 10)"
, ylab="Total Cost (billions $)")

barplot(totAggDamage$AGGDAMAGEVALUE[1:10]/1e+9, col=heat.colors(10), names.arg=totAggDamage$EVTY
PE[1:10], cex.names=0.7, cex.axis=0.7, las=2, main="Total Aggregated Damage by Event Type (Top 1
0)", ylab="Total Cost (billions $)")
```



It can be seen that floods produce the most property damage and cause the most economic damage overall, that drought causes the most damage to crops, and that property damage is of significantly greater cost than crop damage.