Public health and economic impact analysis of weather events in the United States

By Robert Merriman

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many result in fatalities, injuries, crop damage, and/or property damage.

This project involved exploring 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 negative impact.

From 1990 - 2011 the event types that caused the most damage per negative impact type across the United States were:

1. Economic consequences

a. Flooding caused the most property damage
b. Drought caused the most crop damage

2. Public harm (population health)

a. Excessive heat caused the most fatalties
b. Tornados caused the most number of injuries

Recommendations

To improve reporting and analysis of weather events, I recommend that the weather events be classified using a hierarchical classification such as group/subgroup/subsubgroup for example:

```
Tornado/
Tornado/F1
Tornado/F2
Tornado/F4
... and so on
```

The data contained 952 types of weather events, many of which were related or were a specific sub-classification of a more general event type (e.g., Tornado F1 is a specific sub-classification of the general type Tornado). While some sub-classifications are easy to group (e.g., group Tornado F1-F4 with the event type Tornado) others are not so apparent (e.g., should freezing fog be grouped with frost/freeze or should it be grouped with dense fog?). Having a hierarchical classification maintained by NOAA would enable consistent classification reporting of weather events and enable replicable reporting by event classification or group event classification. Proposing a hierarchical classification to be used is beyond the scope of this project.

References

The following references were used:

1. Assignment description
2. Source data file
3. National Climatic Data Center Storm Data FAQ
4. National Weather Service Storm Data Preparation

Data Processing

Summary of the data

The data starts in the year 1950 and ends in November 2011. Prior to 1990 only the event types Tornado and TSTM Wind had any reported data. The event types are not hierarchically classified resulting in related events being listed in the top 10 analysis as seperate events rather than looking at larger aggregate groups of events.

Reading and processing the data

The data file is a comma delimited file that was compresssed using the bzip2 algorithm. I downloaded the file and unzipped it.

```
if(!file.exists("StormData.bz2")){
    download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
                  , destfile = "StormData.bz2", method = "curl")
}
stormData <- read.csv(bzfile("StormData.bz2"), stringsAsFactors = FALSE)
```

I updated the begin date (BGN_DATE) from character type to date type.

```
library(lubridate)
stormData$BGN_DATE <- parse_date_time(stormData$BGN_DATE,orders="%m/%d/%Y %H%M%S")
detach("package:lubridate", unload=TRUE)
```

I converted the data frame into a data table, added a year field which is the year the event began, and restricted the data to states using the R built in state abbrevation data.

```
library(data.table)
stormTable <- data.table(stormData)
stormTable$year <- as.POSIXlt(stormTable$BGN_DATE)$year + 1900
stormTable <- stormTable[STATE %in% state.abb]
```

I multiplied the values in PROPDMG and CROPDMG by the corrsponding magnitude in the PROPDMGEXP and CROPDMGEXP fields respectively where “H” represented hundreds, “K” thousands, “M” millions, and “B” billions, and a numeric value such as “3” represented 1000. I did not use the other magnitude values such as “+”, “?”, etc. as it was not clear what those values represented.

```
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "H" | stormTable$PROPDMGEXP == "h" | stormTable$P
                              , stormTable$PROPDMG * 100, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "K" | stormTable$PROPDMGEXP == "k" | stormTable$P
                              , stormTable$PROPDMG * 1000, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "4")
                              , stormTable$PROPDMG * 10000, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "5")
                              , stormTable$PROPDMG * 100000, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "M" | stormTable$PROPDMGEXP == "m" | stormTable$P
                              , stormTable$PROPDMG * 1000000, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "7")
                              , stormTable$PROPDMG * 10000000, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "8")
                              , stormTable$PROPDMG * 100000000, stormTable$PROPDMG)
stormTable$PROPDMG <- ifelse((stormTable$PROPDMGEXP == "B" | stormTable$PROPDMGEXP == "b" | stormTable$P
                              , stormTable$PROPDMG * 1000000000, stormTable$PROPDMG)

stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "H" | stormTable$CROPDMGEXP == "h" | stormTable$C
                              , stormTable$CROPDMG * 100, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "K" | stormTable$CROPDMGEXP == "k" | stormTable$C
                              , stormTable$CROPDMG * 1000, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "4")
                              , stormTable$CROPDMG * 10000, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "5")
                              , stormTable$CROPDMG <- stormTable$CROPDMG * 100000, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "M" | stormTable$CROPDMGEXP == "m" | stormTable$C
                              , stormTable$CROPDMG * 1000000, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "7")
                              , stormTable$CROPDMG * 10000000, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "8")
                              , stormTable$CROPDMG * 100000000, stormTable$CROPDMG)
stormTable$CROPDMG <- ifelse((stormTable$CROPDMGEXP == "B" | stormTable$CROPDMGEXP == "b" | stormTable$C
                              , stormTable$CROPDMG * 1000000000, stormTable$CROPDMG)
```

I added a decade field:

```
stormTable$decade <- NA
stormTable$decade <- ifelse((stormTable$year >1949 & stormTable$year < 1960), 1950, stormTable$decade)
stormTable$decade <- ifelse((stormTable$year >1959 & stormTable$year < 1970), 1960, stormTable$decade)
stormTable$decade <- ifelse((stormTable$year >1969 & stormTable$year < 1980), 1970, stormTable$decade)
stormTable$decade <- ifelse((stormTable$year >1979 & stormTable$year < 1990), 1980, stormTable$decade)
stormTable$decade <- ifelse((stormTable$year >1989 & stormTable$year < 2000), 1990, stormTable$decade)
stormTable$decade <- ifelse((stormTable$year >1999 & stormTable$year < 2010), 2000, stormTable$decade)
stormTable$decade <- ifelse((stormTable$year >2009 & stormTable$year < 2020), 2010, stormTable$decade)
```

I then subset, sorted, and filtered the data to determine that only Tornado and TSTM Wind had any data prior to 1980. And prior to 1990 only Tornado, TSTM Wind, and Hail had any data. As such for determining which eventy type had the most negative impact, I filtered for year > 1989.

```
options(scipen=999, digits=0)
setkey(stormTable, EVTYPE)
propertydamage <- stormTable[,list(propertydamage = sum(PROPDMG)), by = list(EVTYPE, decade)]
cropdamage <- stormTable[,list(cropdamage = sum(CROPDMG)), by = list(EVTYPE, decade)]
```

```
fatalities <- stormTable[,list(fatalities = sum(FATALITIES)), by = list(EVTYPE, decade)]
injuries <- stormTable[,list(injuries = sum(INJURIES)), by = list(EVTYPE, decade)]

propertydamage[(propertydamage>0 & decade < 1990),]
```

```
##     EVTYPE decade propertydamage
## 1: TORNADO   1950     1516409420
## 2: TORNADO   1960     3757536860
## 3: TORNADO   1970     8024388050
## 4: TORNADO   1980    13180060840
```

```
cropdamage[(cropdamage>0 & decade < 1990),]
```

```
## Empty data.table (0 rows) of 3 cols: EVTYPE,decade,cropdamage
```

```
fatalities[(fatalities>0 & decade < 1990),]
```

```
##       EVTYPE decade fatalities
## 1:   TORNADO   1950       1419
## 2:   TORNADO   1960        942
## 3:   TORNADO   1970        998
## 4:   TORNADO   1980        522
## 5: TSTM WIND   1980        177
```

```
injuries[(injuries>0 & decade < 1990),]
```

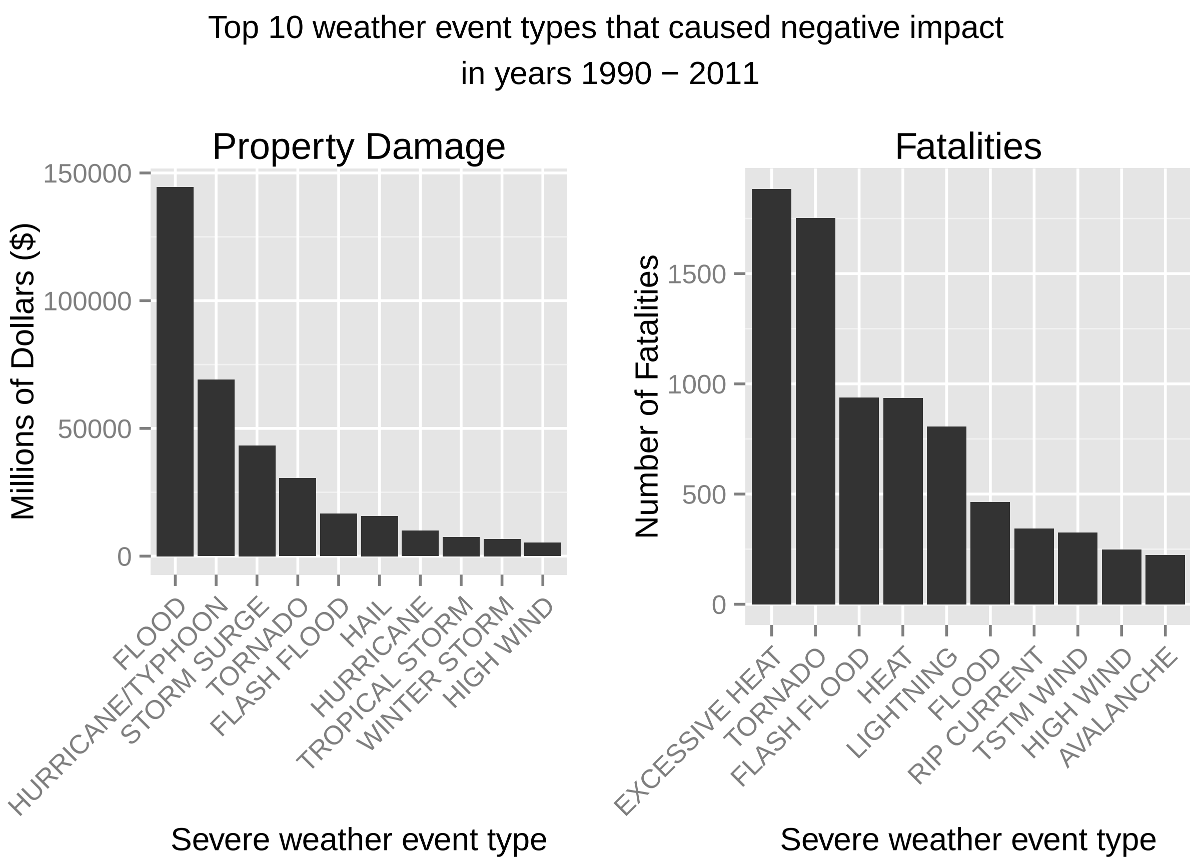
```
##       EVTYPE decade injuries
## 1:      HAIL   1980      222
## 2:   TORNADO   1950    14470
## 3:   TORNADO   1960    17265
## 4:   TORNADO   1970    21640
## 5:   TORNADO   1980    11297
## 6: TSTM WIND   1980     1935
```

Results

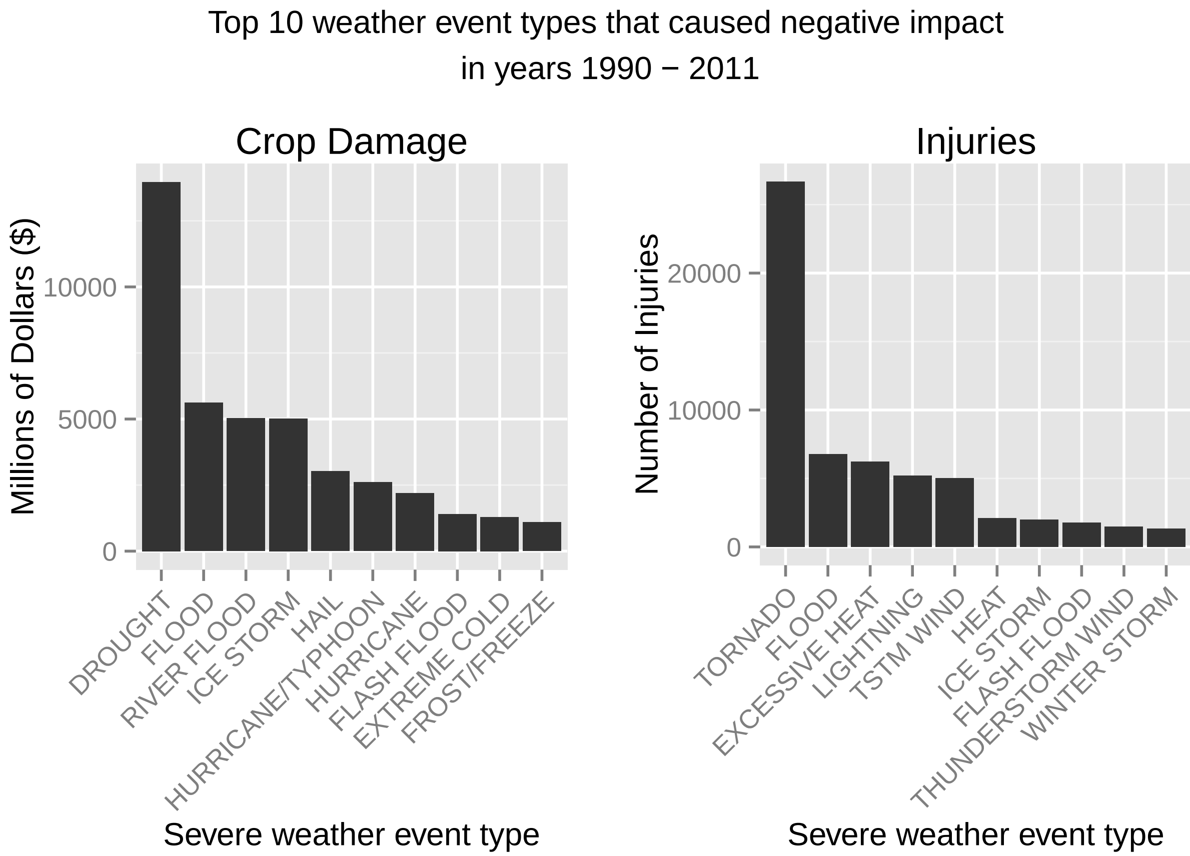
I then filtered the data to years 1990-2011, calculated the aggregate negative impact, and plotted the top 10 event types for each negative impact. To make the dollar amounts readible I used a unit of million.

```
stormTable <- stormTable[year > 1989,]
propertydamage <- stormTable[,list(propertydamage = sum(PROPDMG)), by = list(EVTYPE)]
cropdamage <- stormTable[,list(cropdamage = sum(CROPDMG)), by = list(EVTYPE)]
fatalities <- stormTable[,list(fatalities = sum(FATALITIES)), by = list(EVTYPE)]
injuries <- stormTable[,list( injuries = sum(INJURIES)), by = list(EVTYPE)]

propertydamage <- head(propertydamage[order(propertydamage, decreasing=TRUE),],10)
propertydamage$propertydamage <- propertydamage$propertydamage/1000000
cropdamage <- head(cropdamage[order(cropdamage, decreasing=TRUE),],10)
cropdamage$cropdamage <- cropdamage$cropdamage/1000000
```

```
fatalities <- head(fatalities[order(fatalities, decreasing=TRUE),],10)
injuries <- head(injuries[order(injuries, decreasing=TRUE),],10)

library(ggplot2)
library(grid)
library(gridExtra)

names <- propertydamage$EVTYPE
propertydamage$EVTYPE <- factor(propertydamage$EVTYPE, levels = names)
propertydamagePlot <- qplot(EVTYPE,propertydamage,data=propertydamage,geom="bar",stat="identity"
                            , xlab = "Severe weather event type", ylab = "Millions of Dollars ($)"
                            , main = "Property Damage") + theme(axis.text.x = element_text(angle=45,hju

names <- cropdamage$EVTYPE
cropdamage$EVTYPE <- factor(cropdamage$EVTYPE, levels = names)
cropdamagePlot <- qplot(EVTYPE,cropdamage,data=cropdamage,geom="bar",stat="identity"
                        , xlab = "Severe weather event type", ylab = "Millions of Dollars ($)"
                        , main = "Crop Damage") + theme(axis.text.x = element_text(angle=45,hjust=1))

names <- fatalities$EVTYPE
fatalities$EVTYPE <- factor(fatalities$EVTYPE, levels = names)
fatalitiesPlot <- qplot(EVTYPE,fatalities,data=fatalities,geom="bar",stat="identity"
                        ,xlab = "Severe weather event type", ylab = "Number of Fatalities"
                        , main = "Fatalities") + theme(axis.text.x = element_text(angle=45,hjust=1))

names <- injuries$EVTYPE
injuries$EVTYPE <- factor(injuries$EVTYPE, levels = names)
injuriesPlot <- qplot(EVTYPE,injuries,data=injuries,geom="bar",stat="identity"
                      , xlab = "Severe weather event type", ylab = "Number of Injuries"
                      , main = "Injuries") + theme(axis.text.x = element_text(angle=45,hjust=1))

grid.arrange(propertydamagePlot,fatalitiesPlot, nrow = 1, ncol=2
            , main="Top 10 weather event types that caused negative impact\n in years 1990 - 2011")
```




```r
grid.arrange(cropdamagePlot, injuriesPlot, nrow = 1, ncol=2
             , main="Top 10 weather event types that caused negative impact\n in years 1990 - 2011")
```



From the above plots I can see that from 1990 - 2011 the event types that caused the most damage were:

1. Economic consequence

a. Flooding caused the most property damage
b. Drought caused the most crop damage

2. Public harm (population health)

a. Excessive heat caused the most fatalties
b. Tornados caused the most number of injuries