The Economic and Public Health Impact of Various Storm Events inside the United States

1) Synopsis

In this report I explore a public domain dataset from NOAA on the impact of various storm events in the United States. I wish to answer a couple of basic questions including:

1. Which type of events are the most harmful from a population health perspective?

2. Which types of events have the greatest economic consequences?

To answer the former, I present the top causes of fatalities, non-fatal injuries in a figure below.
For the latter, the dataset contains two tracers of the total economic impact: property damage, and damage to crops. As in my answer to the first question, I present a figure detailing the economic damage caused by the events.

2) Loading and Processing the Data

Begin by loading required packages into R, and obtaining the data-set if it is not in the present working directory.

```r
library(ggplot2)
library(plyr)
library(gridExtra)
# Read in the data frame.  If the file does not exist in the current
# working directory, go find it.
if (!exists('stormdata'))
{
  filename = 'repdata%2Fdata%2FStormData.csv.bz2'
  if (!file.exists(filename))
  {
    url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    download.file(url = url, destfile = filename)
}
  stormdata <- read.csv(filename, header=TRUE, sep=',', as.is=TRUE, strip.white=TRUE)
}
```

One thing that is noticable about the data frame is that the property damage and crop damage are stored in scientific notation with the decimal stored in one column, while the exponent in a differnet column. To deal with that a function to convert the stormdata$*DMG column to a number can be applied with maaply.

```r
unique(stormdata$PROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-"
## [18] "1" "8"
```

```r
unique(stormdata$CROPDMGEXP)
```

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

```r
newexp <- function(incolumn)
{ if      (toupper(incolumn) == "H")  outexp <- 2
  else if (toupper(incolumn) == "K")  outexp <- 3
```

1

```
    else if (toupper(incolumn) == "M")  outexp <- 6
    else if (toupper(incolumn) == "B")  outexp <- 9
    else if (grepl("[0-9]", incolumn) == TRUE) outexp <- incolumn
    else  (outexp <- 0)
    return(outexp)
}
#Now get the damages in Billions of dollars
stormdata$PROPDMGVAL <- stormdata$PROPDMG*10 **as.numeric(mapply(stormdata$PROPDMGEXP, FUN='newexp'))/(
stormdata$CROPDMGVAL  <- stormdata$CROPDMG*10 **as.numeric(mapply(stormdata$CROPDMGEXP, FUN='newexp'))/

#aggregate based on injuries and fatalities in order to answer
# the questions
FatDat <- aggregate(FATALITIES~EVTYPE, stormdata, FUN=sum, na.rm = TRUE)
InjDat <- aggregate(INJURIES~EVTYPE, stormdata, FUN=sum, na.rm = TRUE)
PropDat<- aggregate(PROPDMGVAL~EVTYPE, stormdata, FUN=sum, na.rm = TRUE)
CropDat <- aggregate(CROPDMGVAL~EVTYPE, stormdata, FUN=sum, na.rm = TRUE)
```

3) Analysis

Subset the data

In order to show a legible plot, it's necessary to only show the most relevant categories. Let's take the top five for each category.

```
fatal <- FatDat[order(-FatDat$FATALITIES), ][1:5, ]
injure <- InjDat[order(-InjDat$INJURIES), ][1:5, ]
cropdmg <- CropDat[order(-CropDat$CROPDMGVAL), ][1:5, ]
propdmg <- PropDat[order(-PropDat$PROPDMGVAL), ][1:5, ]
```

Make a bar plot showing the top causes of Injuries, Fatalities.

Tornadoes by far cause the highest number of injuries. Tornadoes are also the most fatal, with heat related deaths coming in second.

```
# create a color palette for the fill colors
mycols <- c("#FF6633", "#669933", "#660000", "#6666FF")
p1 <- ggplot(fatal, aes(x=reorder(EVTYPE, -FATALITIES), weight=FATALITIES))
p1 <- p1 + geom_bar(color='black', fill=mycols[1], width=0.7)
p1 <- p1 + theme_light() +coord_flip()
p1 <- p1 + labs( y='Recorded Fatalities', x='Event Type')
p2 <- ggplot(injure, aes(x=reorder(EVTYPE, -INJURIES), weight=INJURIES))
p2 <- p2 + geom_bar(color='black', fill=mycols[2], width=0.7)
p2 <- p2  + theme_light() + coord_flip()
p2 <- p2 + labs( y='Recorded Injuries', x='Event Type' )
grid.arrange(p1,p2, ncol=1, nrow =2)
```

Repeat for Economic Damage

Aside from the economic consequences of weather events due to fatalities and injuries, weather events cause economic damage in the form of destroyed property and damaged crops. The figure shows that probably unsurspisingly, flooding was the biggest cause of property damage, while crops were most impacted by drought.
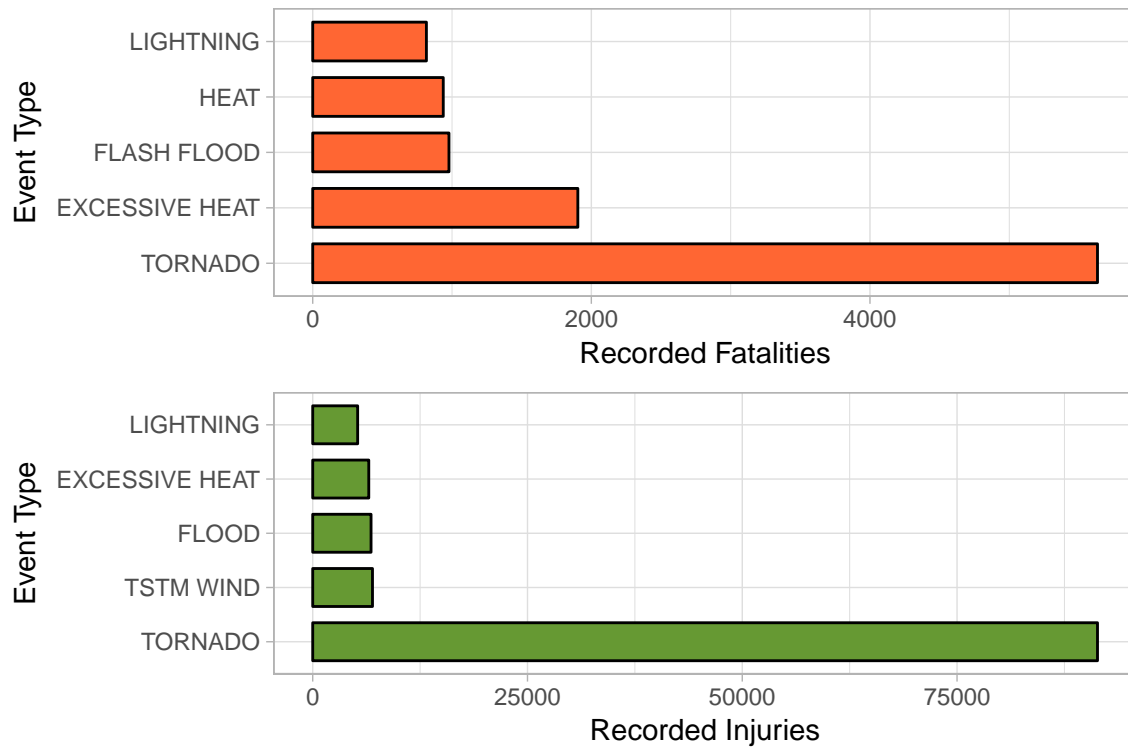
Figure 1: Top Panel: Sum of Fatalities according to event type. Bottom Panel: As in the top panel but sorted for the sum of reported inuries.

```
p3 <- ggplot(cropdmg, aes(x=reorder(EVTYPE, -CROPDMGVAL), weight=CROPDMGVAL))
p3 <- p3 + geom_bar(color='black', fill=mycols[3], width=0.7)
p3 <- p3 + theme_light()+ coord_flip()
p3 <- p3 + labs( y='Crop Damage [$Billions]', x='Event Type')
p4 <- ggplot(propdmg, aes(x=reorder(EVTYPE, -PROPDMGVAL), weight=PROPDMGVAL))
p4 <- p4 + geom_bar(color='black', fill=mycols[4], width= 0.7)
p4 <- p4 + theme_light()+ coord_flip()
p4 <- p4 + labs( y='Property Damage [$Billions]', x='Event Type' )
grid.arrange(p3,p4, ncol=1, nrow =2)
```

4) Conclusion

Tornadoes have been responsible for the greatest number of recorded deaths in the NOAA database, however flash floods and excessive heat have proven to be quite deadly as well. When it comes to non-fatal injuries, Tornadoes still lead, but other factors like thunderstorms come into play. When it comes to economic damage to crops and property, flooding is one of the most significant causes. This isn't too much of a surprise considering how much property damage can be caused by a flooded basement. When it comes to damaging crops, however, lack of water in the form of drought causes the most damage. Combining the River Flood and General Flood catergories into a single one, however, would result in a similar amount of damage.
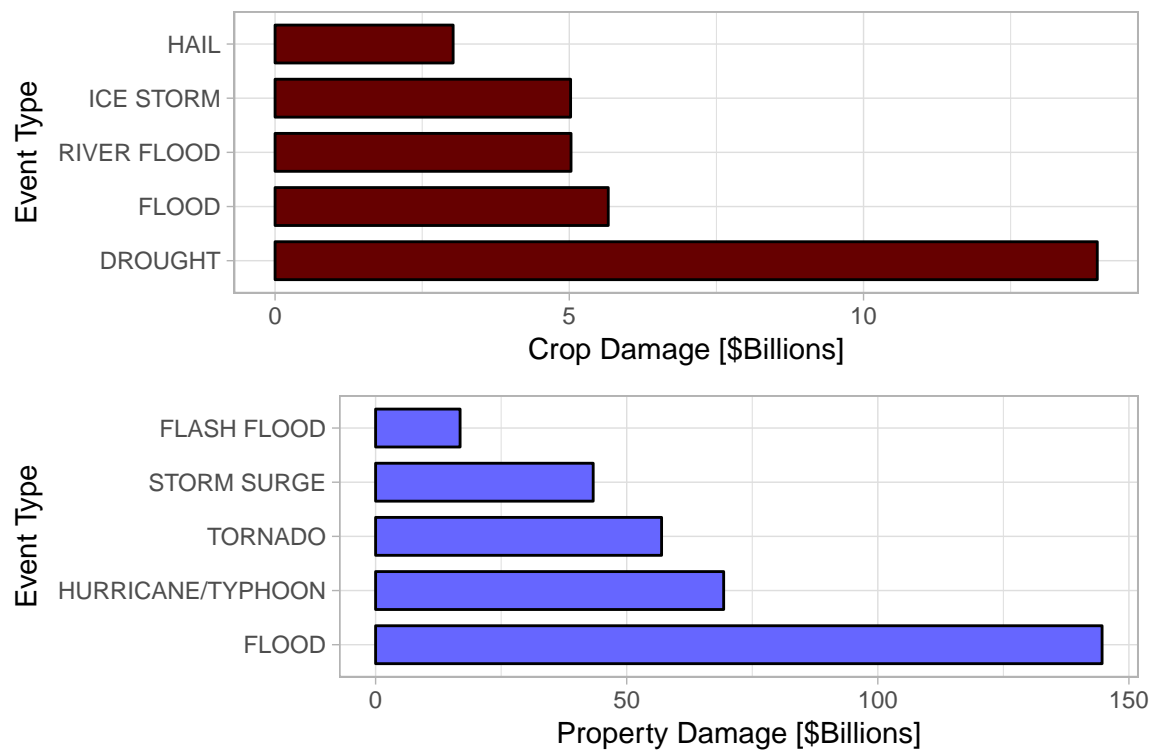
Figure 2: Top Panel: As in Figure 1. Top: Reported crop damage in billions of dollars. Bottom: Reported property damage.