Weather Events causing the Highest Health and Economic Damages according to the Stormdata Dataset

Namakuto

July 14, 2017

Synopsis

In this report, the "Stormdata.csv" file was analyzed to interpret various health and economic damages caused by various weather events. Events with the highest numbers (summed) in human injuries + fatalities as well as crop + property damages were looked at, in particular.

The Stormdata.csv is directly available from https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2. The data was collected by the National Storm Service from 1950 to 2011 and further information is available here: https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf.

Data Processing and Transformations

The Stormdata.csv was first (manually) downloaded from the aforementioned link in the synopsis. A .zip extraction tool was then used to open the the compressed .csv.bz2 file.

After following the above steps, the dataset was then loaded into R:

```
url<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"; download.file(url,dest
library(ggplot2); library(grid); library(gridExtra)
storm<-read.csv("stormdata.csv")
```

Processing for Q1 A new data.frame was made to add "Storm"'s injuries + fatalities together, into a new column.

```
storm.2<-data.frame(storm$INJURIES+storm$FATALITIES, storm$EVTYPE)
names(storm.2)<-c("fatal_injury", "event")
```

Each injury + fatality row was then aggregated (summed) according to the weather event. The summed data.frame was rearranged into descending order.

```
total.sum<-aggregate(storm.2$fatal_injury, by=list(storm.2$event), FUN=sum)
total.sum<-total.sum[order(total.sum$x, decreasing = TRUE),]
names(total.sum)<-c("event", "count")
```

- The levels of the injury+fatality counts were finally reordered, to fix the order that would be shown in the discrete barplot.
- The data.frame was changed to show only the top 20 events highest in injury and fatality, due to a high number of 0's/low counts/improved visualization.

```
total.sum$event<-factor(total.sum$event,
                        levels=total.sum$event[order(total.sum$count)])
total.sum2<-total.sum[1:20,]
```

Processing for Q2 Crop and property damage each included an extra column (2 in total) to indicate the magnitude of their "DMG" values– i.e, "H|h" units for hundreds, "m|M" units for millions, etc. Each "DMG"

value was multiplied accordingly by its (respective) magnitude unit. "DMG" values were then filtered back into the dataset:

```
K.prop<-grep(storm$PROPDMGEXP, pattern="K", value=FALSE)
storm$PROPDMG[K.prop]<-storm$PROPDMG[K.prop]*(1000)
B.prop<-grep(storm$PROPDMGEXP, pattern="B", value=FALSE)
storm$PROPDMG[B.prop]<-storm$PROPDMG[B.prop]*(10^9)
M.prop<-grep(storm$PROPDMGEXP, pattern="m|M", value=FALSE)
storm$PROPDMG[M.prop]<-storm$PROPDMG[M.prop]*(10^6)
H.prop<-grep(storm$PROPDMGEXP, pattern="h|H", value=FALSE)
storm$PROPDMG[H.prop]<-storm$PROPDMG[H.prop]*(100)

K.crop<-grep(storm$CROPDMGEXP, pattern="k|K", value=FALSE)
storm$CROPDMG[K.crop]<-storm$CROPDMG[K.crop]*(1000)
B.crop<-grep(storm$CROPDMGEXP, pattern="B", value=FALSE)
storm$CROPDMG[B.crop]<-storm$CROPDMG[B.crop]*(10^9)
M.crop<-grep(storm$CROPDMGEXP, pattern="m|M", value=FALSE)
storm$CROPDMG[M.crop]<-storm$CROPDMG[M.crop]*(10^6)
```

Units that weren't reported in hundreds, kilos, millions or billions weren't converted, and instead, left to their original (exp 1) value in the "DMG" column. They are assumed to be insignificant) given their low counts in the dataset (e.g):

```
cat("Indices with (exp 7): ", grep(storm$PROPDMGEXP, pattern="7", value = FALSE))
```

```
## Indices with (exp 7):  213319 220986 234146 234615 234940
```

```
cat("Indices with (exp 8): ", grep(storm$PROPDMGEXP, pattern="8", value = FALSE))
```

```
## Indices with (exp 8):  234228
```

CROPDMGEXP, for instance, only consisted of (exp 2) magnitudes after the alphabetical magnitudes:

```
levels(factor(storm$CROPDMGEXP))
```

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

After unit conversions, an aggregation process followed. Two data.frames were made, seperating (and arranging) weather events highest in crop damage from those highest in property damage.

```
property.sum<-aggregate(storm$PROPDMG, by=list(storm$EVTYPE), FUN=sum)
property.sum<-property.sum[order(property.sum$x, decreasing = TRUE),]
names(property.sum)<-c("event", "count")
property.sum$event<-factor(property.sum$event,
                      levels=property.sum$event[order(property.sum$count)])

crop.sum<-aggregate(storm$CROPDMG, by=list(storm$EVTYPE), FUN=sum)
crop.sum<-crop.sum[order(crop.sum$x, decreasing = TRUE),]
names(crop.sum)<-c("event", "count")
crop.sum$event<-factor(crop.sum$event,
                        levels=crop.sum$event[order(crop.sum$count)])
```

- The data.frames were then shrunken into "Top 6" lists, for ease of comparison/ visualization.

```
crop.sum2<-crop.sum[1:6,]; property.sum2<-property.sum[1:6,]
```

Results

Q1 Tornadoes had a significant impact on population health. Tornadoes caused greatly significant injury + fatality counts (summed) compared to any other weather phenomenon. The remaining top 19 weather events in summed injury + fatality counts are as follows:

```
q<-ggplot(total.sum2, aes(event,count))+geom_col()+
  ggtitle("Top 20 Weather Events Highest in Injuries + Fatalities, Summed")+
  labs(x="Weather Event", y="Number of Injuries + Fatalities")+
  theme(
    axis.text.x = element_text(margin=margin(t=20,b=-20),
                               angle=40,size=6, face="bold")
  ); q
```

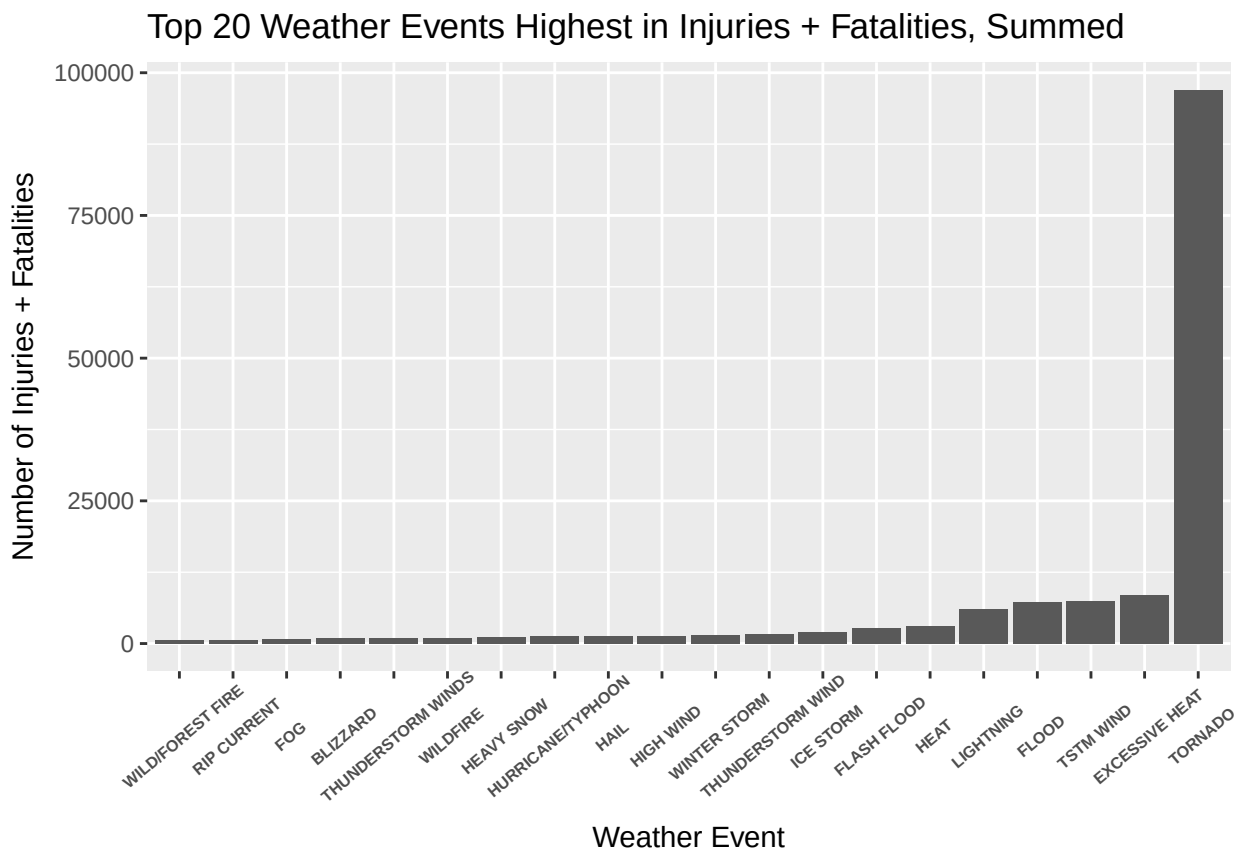

Figure 1. Barchart of the top 20 weather events highest in injuries and fatalities combined (summed).

A preview of the data.frame, in combined injury + fatality counts:

```
##              event count
## 834        TORNADO 96979
## 130 EXCESSIVE HEAT  8428
## 856      TSTM WIND  7461
## 170          FLOOD  7259
## 464      LIGHTNING  6046
## 275           HEAT  3037
```

Heat seemed to notably reduce population health right after tornadoes. But tornadoes were particularly harmful.

Q2 A plot on economic damages showed drought to have the highest impact on crop damage. Flooding, meanwhile, had the highest impact on property damage.

```
q1<-ggplot(crop.sum2,aes(event,count))+
  geom_bar(stat="identity",position="dodge", fill="dark cyan")+
  labs(title="Top 6 Weather Events Highest in Crop
Damage, Summed",
    x="Event", y="Dollars (U.S.D)")+
  theme(axis.text.x = element_text(size="7", face="bold",angle=40, margin=margin(t=10,b=-10)),
        plot.title = element_text(size="9", face="bold"))

q2<-ggplot(property.sum2,aes(event,count))+
  geom_bar(stat="identity",position="dodge", fill="brown")+
  labs(title="Top 6 Weather Events Highest in Property
Damage, Summed",
       x="Event", y="Dollars (U.S.D)")+
  theme(axis.text.x = element_text(size="6",face="bold", angle=40, margin=margin(t=20,b=-30)),
        plot.title = element_text(size="9", face="bold"))

grid.arrange(q1,q2, ncol=2)
```

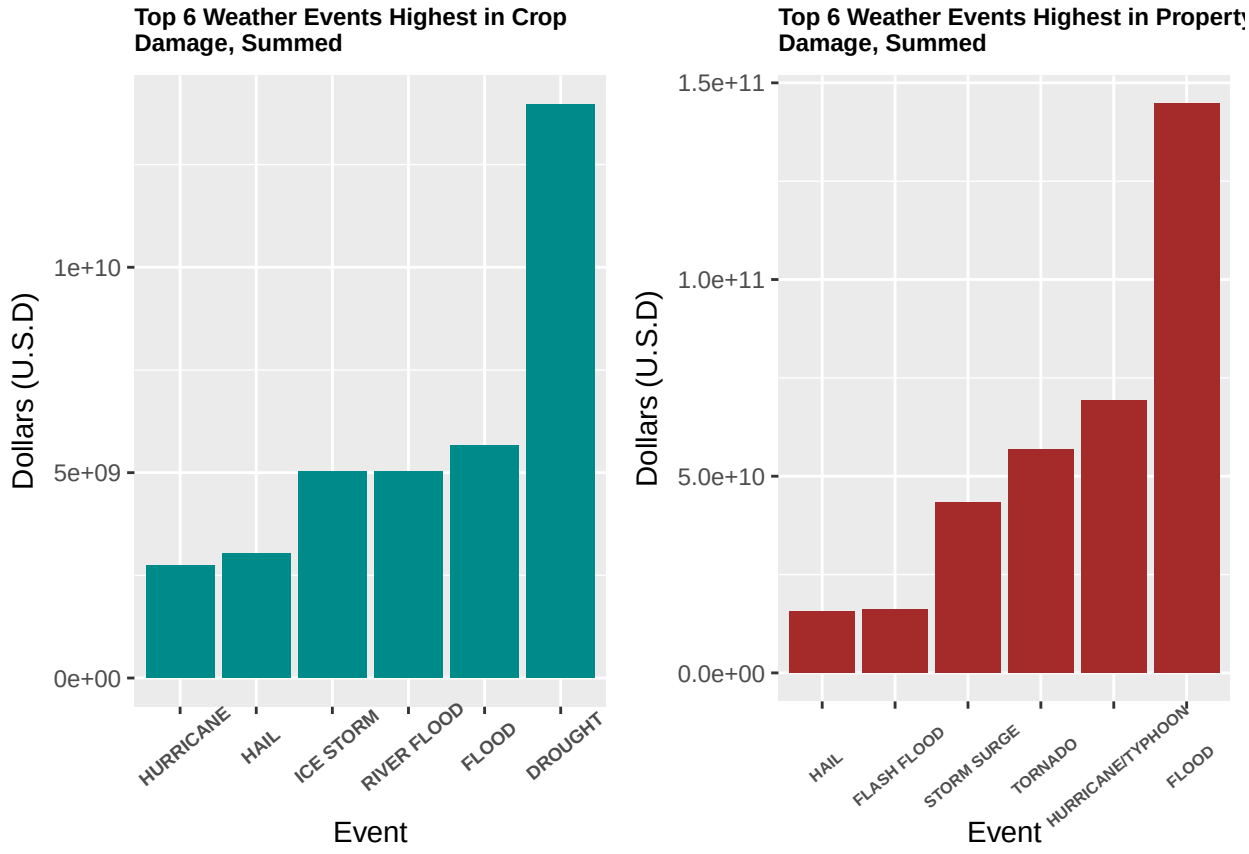

Figure 2. Barchart of the top 6 weather events highest in crop and property damage, in (assumed) U.S.D dollars.

Conclusion

Based on the first chart, tornadoes had the most significant impact on population health, both in regards to human fatalities and injuries. Drought, meanwhile, had the largest economic consequence to crops. Flooding had the largest economic consequence to property damage.