US Weather Events Historically Most Harmful to Population Health and Economy

K. Divis

Synopsis

Severe weather events can have a negative impact on both population health and economy. Below, I break down which types of weather events have historically impacted population health and the econony the most. Using data drawn from weather events in the US from 1996-2008, I look at the impact of weather on the population health indices of number of injuries and fatalities and the economic consequence indices of dollars of damage to property and crops. I found that the types of weather events that cause the most economic damage (e.g., floods) are not indentical to those that cause the most population health damage (e.g., heat or tornados).

Please note that this project was completed as part of and under the guidelines of the Johns Hopkins Data Science specialization Reproducible Research class on Coursera.

Data Processing

```
if(!file.exists("./stormData.csv.bz2")){
    fileURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    download.file(fileURL,destfile="./stormData.csv.bz2")
}

storm <- read.csv("./stormData.csv.bz2")
```

Step 1: Load the data directly from the link provided on the course website.

Step 2: Subset the raw data, including only data from 1996-2008. According to the NOAA (the source of the weather data), only a subset of the events were reported prior to 1996. Since all events are of interest, only data from 1996+ can be used without introducing bias. (See http://www.ncdc.noaa.gov/stormevents/details.jsp).

```
start96=grep("1996",storm$BGN_DATE)
storm96=data.frame()
storm96=storm[start96[1]:dim(storm)[1],]
```

Step 3: Filter the data into 2 separate datasets. One only includes instances where injuries and/or fatalities occurred (health dataset). The other only includes instances where crop damage and/or property damage occurred (economic dataset).

```
library(dplyr)
storm96health=filter(storm96, INJURIES>0 | FATALITIES>0)      #Health dataset
storm96ec=filter(storm96, PROPDMG>0 | CROPDMG>0)              #Economic dataset
```

Step 4: Calculate totals for each event type in health and economic datasets "H/h/K/k/M/m/B/b" labels are converted to their appropriate numeric value (e.g., 100) and included in the total sum for that event.

Health dataset

```
#Injuries
injSums=tapply(storm96health$INJURIES,list(storm96health$EVTYPE),sum,na.rm=TRUE)
injSumsTotal=0
for (i in 1:dim(injSums)[1]){injSumsTotal[i]=injSums[[i]]}
injData=data.frame(row.names(injSums),injSumsTotal)
names(injData)[1]="Event"
injData=filter(injData,injData$injSums>0)

#Fatalities
fatSums=tapply(storm96health$FATALITIES,list(storm96health$EVTYPE),sum,na.rm=TRUE)
fatSumsTotal=0
for (i in 1:dim(fatSums)[1]){fatSumsTotal[i]=fatSums[[i]]}
fatData=data.frame(row.names(fatSums),fatSumsTotal)
names(fatData)[1]="Event"
fatData=filter(fatData,fatSums>0)

#Combined
library(plyr)
healthData=join(injData,fatData)
```

Economic dataset

```
propSums=tapply(storm96ec$PROPDMG,list(storm96ec$EVTYPE,storm96ec$PROPDMGEXP),sum,na.rm=TRUE)
propSums=propSums[,14:19]   #Only used EXP of h/H/k/K/m/M/B
propSumsTotal=NULL
for (i in 1:dim(propSums)[1]){
    propSumsTotal[i]=sum(propSums[i,1]*1000000000,propSums[i,2]*100,propSums[i,3]*100,propSums[i,4]*1000
}
propData=data.frame(row.names(propSums),propSumsTotal)
names(propData)[1]="Event"
propData=filter(propData,propSumsTotal>0)   #So now have only events with property damage > 0 (but even

#Crop Damage
cropSums=tapply(storm96ec$CROPDMG,list(storm96ec$EVTYPE,storm96ec$CROPDMGEXP),sum,na.rm=TRUE)
cropSums=cropSums[,5:9] #Only used EXP of h/H/k/K/m/M/B
cropSumsTotal=NULL
for (i in 1:dim(cropSums)[1]){
    cropSumsTotal[i]=sum(cropSums[i,1]*1000000000,cropSums[i,2]*1000,cropSums[i,3]*1000,cropSums[i,4]*10
}
cropData=data.frame(row.names(cropSums),cropSumsTotal)
names(cropData)[1]="Event"
cropData=filter(cropData,cropSumsTotal>0)


#Combined
ecData=join(propData,cropData)
```

Step 5: Group Events into Larger Categories The individual events reported in the original EVTYPE variable were grouped into 27 possible larger categories (e.g., flood, thunderstorm, lightning). Do to the messiness and inconsistancies in the original data, it was necessary to narrow the original EVTYPE options into more uniform, consistent categories. The mapping of those groupings can be found in the Sub-Event and Event columns of the ecData and healthData dataframes below.

```
labelsHealth=c("Avalanche","Ice","Snow","Snow","Fire","Coastal Flooding/Erosion/Storm","Coastal Flooding
healthData=mutate(healthData,LargeEvent=labelsHealth)
names(healthData)=c("Sub-Event","Injuries","Fatalities","Event")

labelsEc=c("Surf/Seas","Flood","Thunderstorm","Thunderstorm","Coastal Flooding/Erosion/Storm","Coastal F
ecData=mutate(ecData,LargeEvent=labelsEc)
names(ecData)=c("Sub-Event","Property Damage","Crop Damage","Event")
```

Results

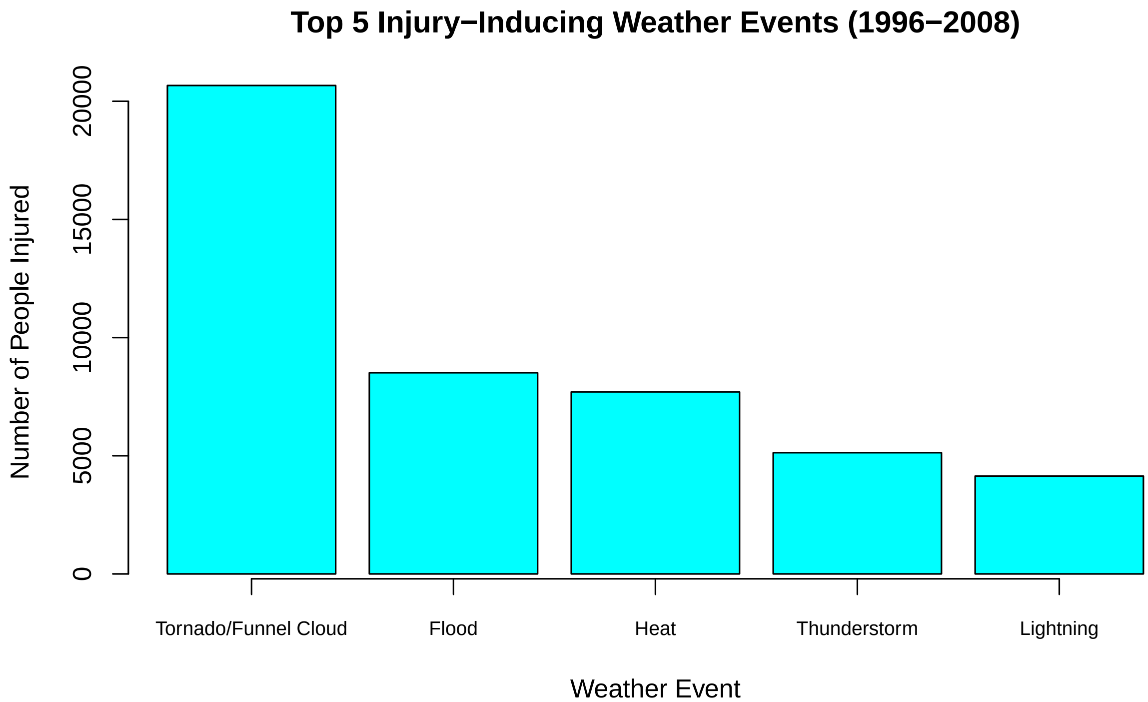
Population Health The original dataset gives two indices of the impact of severe weather on population health: injuries and fatalities. Figure 1 below shows the top 5 weather events leading to the most injuries or fatalities from 1996 to 2008. Heat events were the most fatal (followed by tornados/funnel clouds, floods, lightning, and rip currents), while tornados/funnel clouds were most likely to lead to injury (followed by floods, heat, thunderstorms, and lightning). How individual events are categorized into larger events likely influences the result of these analyses, but it appears that severe summer weather (e.g., storms leading to tornados, lightning, flooding; or extreme heat) are most detrimental to population health.

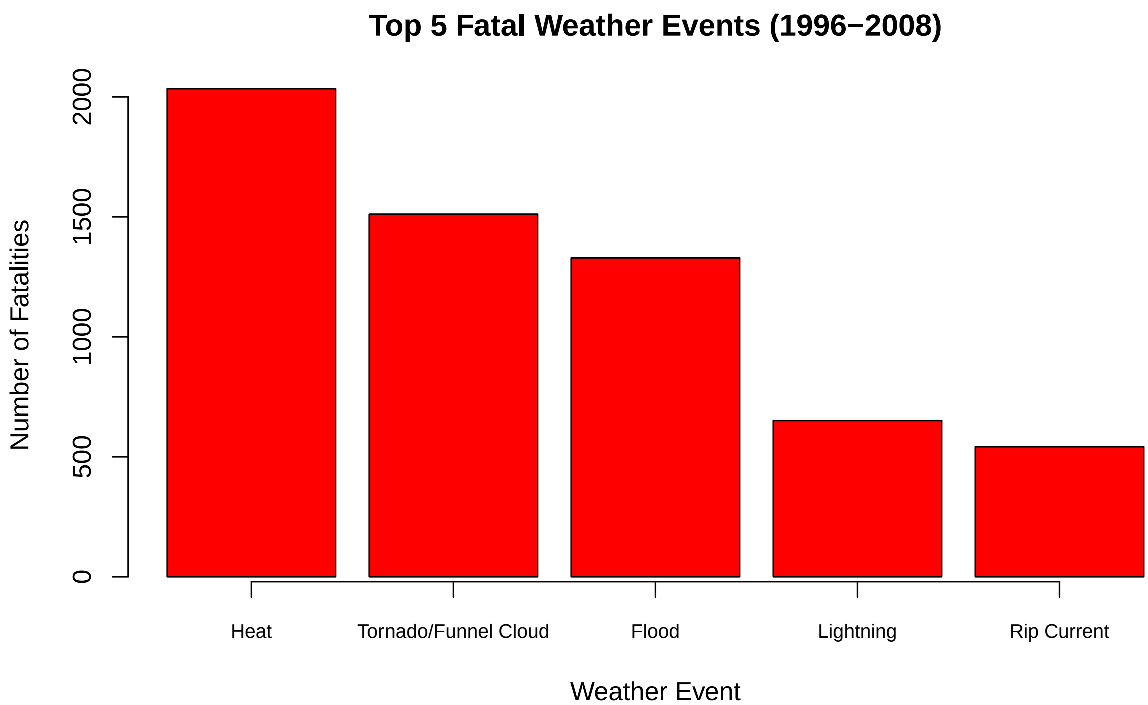
```
library(reshape2)
library(plyr)
hMelt <- melt(healthData,id="Event",measure.vars=c("Injuries","Fatalities"))
names(hMelt)=c("Event","Source","Count")

hMelt2 <- ddply(hMelt,.(Event,Source),summarize,TotalCount=sum(Count,na.rm=TRUE))
hMelt3 <- arrange(hMelt2,Source,desc(TotalCount))

#Get top 5 events in each category
hTopInj <- hMelt3[1:5,]; hTopFat <- hMelt3[26:30,]

par(mfcol = c(2,1))
barplot(hTopInj$TotalCount,names.arg=hTopInj$Event,cex.names=.75,axis.lty=1,col="cyan",
        ylab="Number of People Injured",xlab="Weather Event", main="Top 5 Injury-Inducing Weather Events
barplot(hTopFat$TotalCount,names.arg=hTopFat$Event,cex.names=.75,axis.lty=1,col="red",
        ylab="Number of Fatalities",xlab="Weather Event", main="Top 5 Fatal Weather Events (1996-2008)")
```





(Above) Figure 1: Weather events most harmful to population health

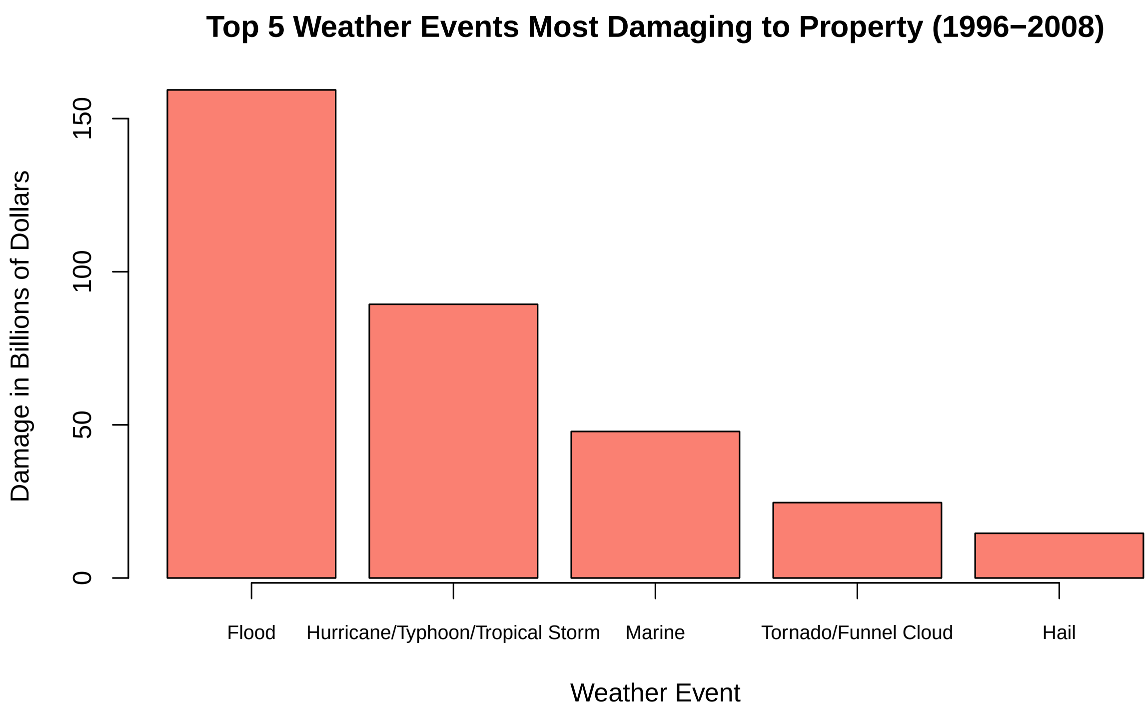
Economic Consequences The original dataset gives two indices of the economic consequences of severe weather: property damage and crop damage. Figure 2 below shows the top 5 weather events leading to the most property or crop damage from 1996-2008. Floods and hurricanes/typhoons/tropical storms caused the most damage to both property and crops. Marine, tornado/funnel cloud and hail events also led to a lot of property damge; hail, cold, and ice also led to substantial crop damage. Weather events that led to moisture damage (e.g., flooding or hurricanes) appeared to cause the greatest negative economic consequences.

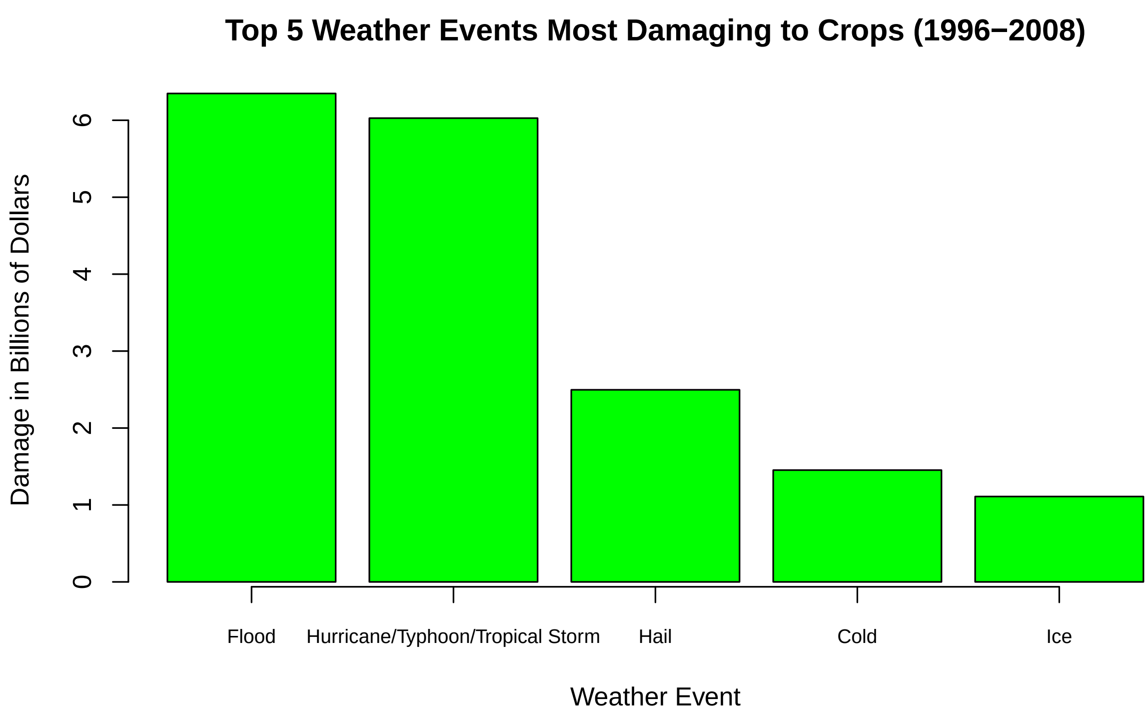
```
ecMelt <- melt(ecData,id="Event", measure.vars=c("Property Damage","Crop Damage"))
names(ecMelt)=c("Event","Source","Cost")

ecMelt2 <- ddply(ecMelt,.(Event,Source),summarize,TotalCost=sum(Cost,na.rm=TRUE))
ecMelt3 <- arrange(ecMelt2,Source,desc(TotalCost))

#Get top 5 events in each category
ecTopProp <- ecMelt3[1:5,]; ecTopCrop <- ecMelt3[28:32,]

par(mfcol = c(2,1))
barplot(ecTopProp$TotalCost/1000000000,names.arg=ecTopProp$Event,cex.names=.75,axis.lty=1,col="salmon",
        ylab="Damage in Billions of Dollars",xlab="Weather Event", main="Top 5 Weather Events Most Damag
barplot(ecTopCrop$TotalCost/1000000000,names.arg=ecTopCrop$Event,cex.names=.75,axis.lty=1,col="green",
        ylab="Damage in Billions of Dollars",xlab="Weather Event", main="Top 5 Weather Events Most Damag
```





(Above) Figure 2: Weather events with worst economic consequences

Overall, the weather events that cause population health problems and those that cause negative economic consequences do not perfectly overlap. While flooding seemed to be a problem for both, heat had a larger relative impact on population health than economic consequences. Events in the family of hurricanes were more likely to lead relatively larger economic consequences than population health detriments (perhaps due to advanced warning and evacuation).