

Explorative Analysis of Severe Weather Event Data from the NOAA Storm Database

Jens Hooge

21.6.2014

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern. In this analysis we asked the following questions:

1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

We could show that, the majority of deaths as well as injuries by severe weather events, between 1950 and 2011 in the US, were caused by hurricanes, strong winds and floods. A similar result was seen in the comparison of the average financial property and crop damages for the same time interval, with an emphasis on heatwaves in case of crop damage.

Data Processing

Loading packages

```
library(memoise)
library(ggplot2)
library(gridExtra)
```

```
## Loading required package: grid
```

Loading the Dataset

The data will be loaded using the read.csv function. Unfortunately, this function is slow for larger datasets. Therefore we will use memoisation, which provides a way of caching the results of a function so that when we call it again with the same arguments it returns the pre-computed value. (This does not seem to work, while knitting however.)

```
mem.read.csv <- memoise(read.csv)
data <- mem.read.csv(bzfile("data//FStormData.csv.bz2"))
```

Data Extraction and Data Cleaning

For the following analysis we will only extract relevant columns from the original data, remove rows with missing values and reformat numeric values describing property damage.

```
extdata <- subset(data,select=c("EVTYPE", "MAG",  
                              "FATALITIES", "INJURIES",  
                              "PROPDMG", "PROPDMGEXP",  
                              "CROPDMG", "CROPDMGEXP"))  
  
## Replace chracters describing the exponent with proper numeric values  
extdata$PROPDMGEXP <- as.numeric(chartr("kKmMbB", "336699", extdata$PROPDMGEXP))
```

```
## Warning: NAs introduced by coercion
```

```
extdata$CROPDMGEXP <- as.numeric(chartr("kKmMbB", "336699", extdata$CROPDMGEXP))
```

```
## Warning: NAs introduced by coercion
```

```
extdata <- extdata[complete.cases(extdata), ]  
  
## Exponentiate damage by its exponent  
extdata$PROPDMG <- extdata$PROPDMG ^ extdata$PROPDMGEXP  
extdata$CROPDMG <- extdata$CROPDMG ^ extdata$CROPDMGEXP  
  
## Remove exponent columns  
extdata <- subset(extdata, select = -c(PROPDMGEXP, CROPDMGEXP))
```

To get an idea about the severity of certain weather events with respect to their impact on population health, we will sum over the the number of fatalities and injuries respectively for each weather event. The resulting data frames will then be ordered by the number of fatalities and injuries and the 10 most severe weather events will be extracted.

```
fatalities <- aggregate(FATALITIES ~ EVTYPE, data=extdata, sum)  
injuries <- aggregate(INJURIES ~ EVTYPE, data=extdata, sum)  
  
fatalities <- fatalities[with(fatalities, order(-FATALITIES)), ]  
fatalities <- fatalities[1:10, ]  
injuries <- injuries[with(injuries, order(-INJURIES)), ]  
injuries <- injuries[1:10, ]
```

The property and crop damage will be averaged over the different weather events and the 10 financially most severe events will be extracted. Because the difference between the largest and smallest financial damage, theses mean values will late be plotted on a log scale.

```
avgPropertyDmg <- aggregate(PROPDMG ~ EVTYPE, data=extdata, mean)  
avgCropDmg <- aggregate(CROPDMG ~ EVTYPE, data=extdata, mean)  
  
avgPropertyDmg <- avgPropertyDmg[with(avgPropertyDmg, order(-PROPDMG)), ]  
avgPropertyDmg <- avgPropertyDmg[1:10, ]  
avgCropDmg <- avgCropDmg[with(avgCropDmg, order(-CROPDMG)), ]  
avgCropDmg <- avgCropDmg[1:10, ]
```

Results

```
fatalityPlot <- ggplot() +  
  geom_histogram(data=fatalities,  
    aes(x=reorder(EVTYPE, -FATALITIES), y=FATALITIES,  
      labels=FATALITIES),  
    stat="identity",  
    position="dodge") +  
  labs(x="", y="No. Fatalities") +  
  theme_bw() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  
  
injuryPlot <- ggplot() +  
  geom_histogram(data=injuries, aes(x=reorder(EVTYPE, -INJURIES), y=INJURIES),  
    stat="identity",  
    position="dodge") +  
  labs(x="", y = "No. Injuries") +  
  theme_bw() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  
  
grid.arrange(fatalityPlot, injuryPlot,  
  main = "Aftermath of Severe Weather Events in the US\nbetween 1950 and 2011",  
  nrow=1)
```

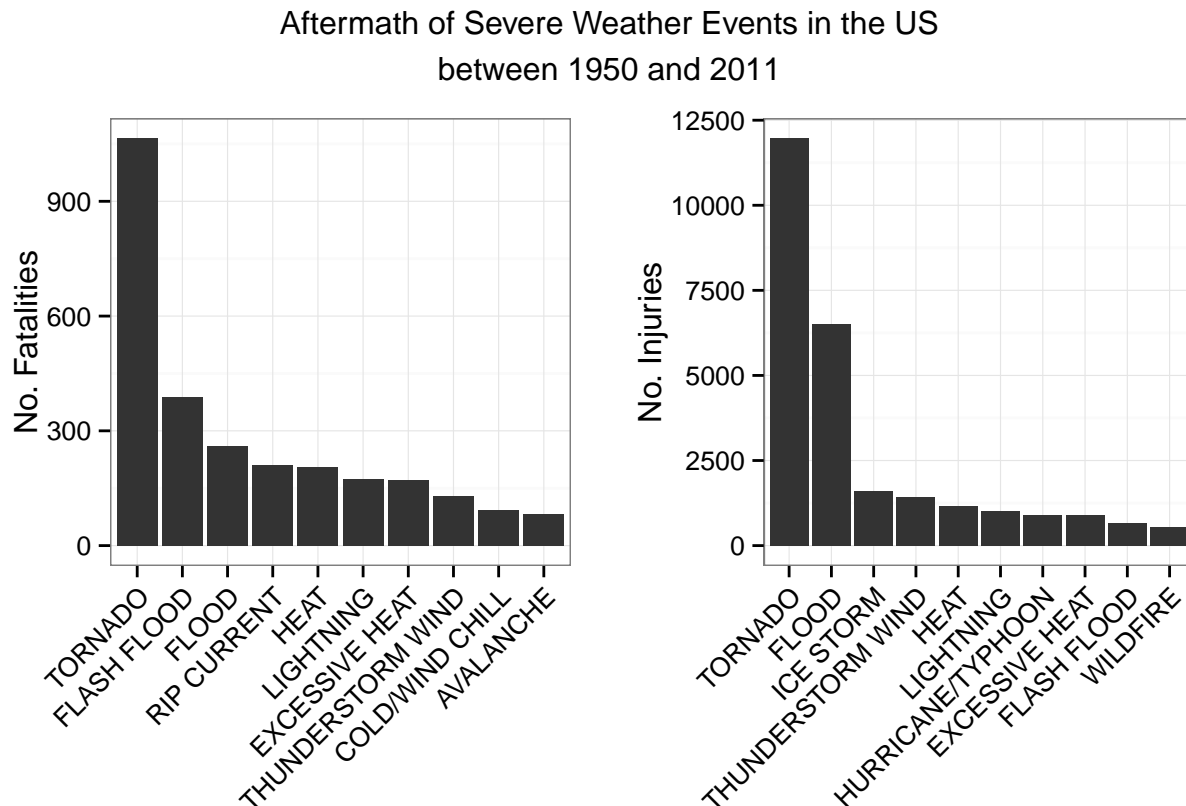


Fig. 1: This figure shows the 10 most severe weather events in the US during the documented time interval. It can be seen, that most fatalities [left] and injuries [right], by far, had been caused by tornadoes and to some degree by floods. To a minor degree, heatwaves and thunderstorms can also be made responsible for many

deaths between the years 1950 and 2011

```
propertyDmgPlot <- ggplot() +
  geom_histogram(data=avgPropertyDmg,
    aes(x=reorder(EVTYPE, -PROPDMG), y=PROPDMG,
      labels=PROPDMG),
    stat="identity",
    position="dodge") +
  scale_y_log10() +
  labs(x="", y="Avg. Property Damage [USD]") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

cropDmgPlot <- ggplot() +
  geom_histogram(data=avgCropDmg,
    aes(x=reorder(EVTYPE, -CROPDMG), y=CROPDMG,
      labels=CROPDMG),
    stat="identity",
    position="dodge") +
  scale_y_log10() +
  labs(x="", y="Avg. Crop Damage [USD]") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

grid.arrange(propertyDmgPlot, cropDmgPlot,
  main = "Average Property and Crop Damage by Severe Weather Events\nin the US between 1950 and 2011")
```

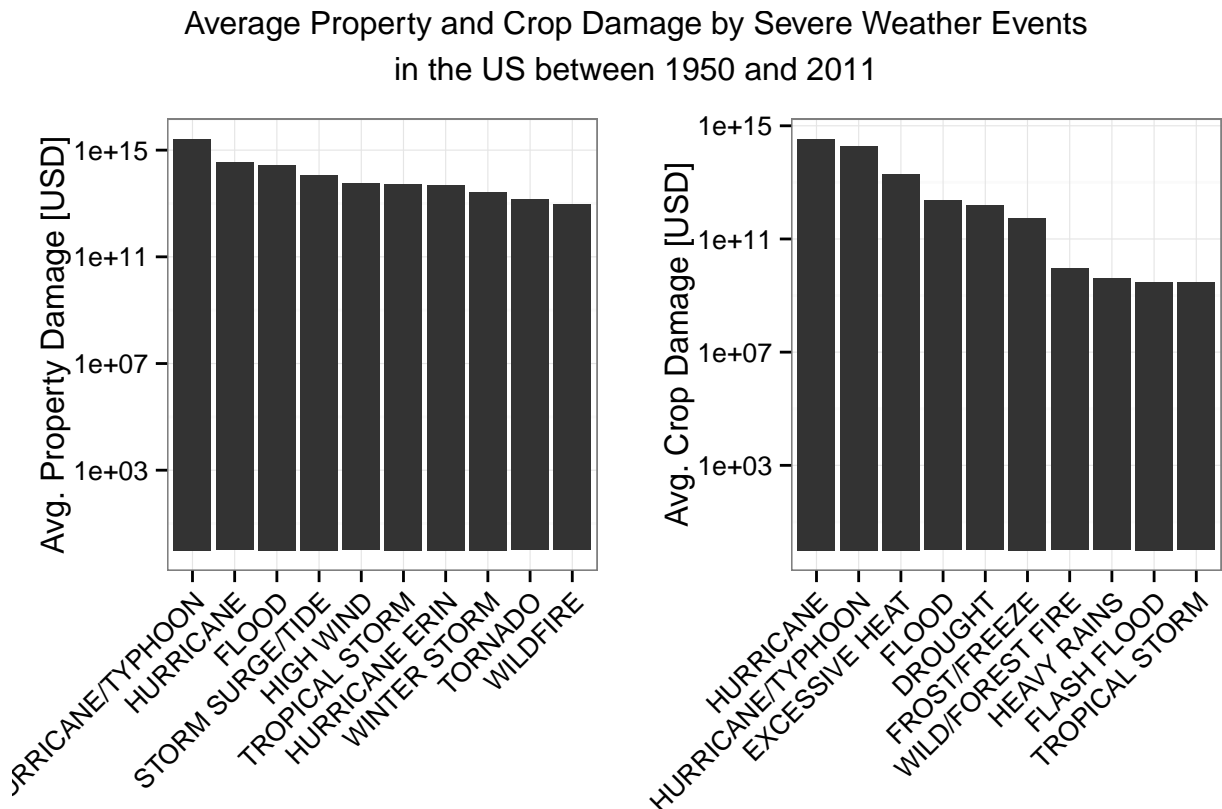


Fig. 2: This figure shows the average property damage [left] and the average crop damage [right] of the

10 financially most severe weather events in the US. As in figure 1, a strong impact of heavy winds like hurricanes, tropical storms and interconnected floods can be observed. In case of crop damage, excessive heat is responsible for harvesting deficits.