Health and economic damages caused by severe weather events

Synopsis

NOAA [http://www.noaa.gov/], the U.S. National Oceanic and Atmospheric Administration, keeps trace of storms and weather events occurred in the United States, along with a set of characteristics like location, date and time of the events, fatalities, injuries, and property damage measures.

The purpose of this study is to analyse the data from NOAA database, focusing on the negative effects caused by severe weather like storms, tornado, flood, … In particular, the research questions addressed by this analysis are:

1. Across the United States, which types of events are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

Data Processing

The data used is the file repdata_data_StormData.csv.bz2. It is a comma-separated-value file and compressed via the bzip2 algorithm to reduce its size.

Steps to process the data:

1. Download the data set from the link [https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2] e unzip it.
2. Load the data. Execute the instruction:

```
storm <- read.csv("repdata_data_StormData.csv")
```

3. Filter the data. The data set contains data from 1950 to 2011. In the earlier years there are generally fewer events recorded and fewer kind of events (only Tornado, Thunderstorm Wind and Hail from 1950 to 1996), while the recent years are more complete so the analysis is limited to data from 1996 (see details at link [http://www.ncdc.noaa.gov/stormevents/details.jsp]).

```
require(lubridate)
```

```
storm$BGN_YEAR <- year(as.Date(storm$BGN_DATE, "%m/%d/%Y %H:%M:%S"))
storm <- subset(storm, BGN_YEAR >= 1996)
```

4. Identify the variables of interest. For the analysis, the most important characteristics are:
   - EVTYPE: event type (e.g. tornado, flood, etc.)
   - BGN_YEAR: year of the event
   - FATALITIES: number directly killed
   - INJURIES: number directly injured
   - PROPDMG: property damage in whole numbers and hundredths
   - PROPDMGEXP: a multiplier where Hundred (H), Thousand (K), Million (M), Billion (B)
   - CROPDMG: crop damage in whole numbers and hundredths
   - CROPDMGEXP: a multiplier where Hundred (H), Thousand (K), Million (M), Billion (B)

```

storm <- subset(storm, select = c(EVTYPE, BGN_YEAR, FATALITIES, INJURIES, PROPDMG,
    PROPDMGEXP, CROPDMG, CROPDMGEXP))
```

5. Combines the PROPDMG and PROPDMGEXP fields to create a numeric value PROPCASH.

   Combines the CROPDMG and CROPDMGEXP fields to create a numeric value CROPCASH.

```
install.packages("car", repos = "http://cran.rstudio.com/")
library(car)
```

```
storm$PROPDMGEXP <- recode(storm$PROPDMGEXP, " ''=0; 'B'=9; 'h'=2; 'H'=2; 'K'=3; 'm'=6;
'M'=6; c('   \n-','?','+') = NA ")
storm$CROPDMGEXP <- recode(storm$CROPDMGEXP, " ''=0; 'B'=9; 'K'=3; 'k'=3; 'm'=6; 'M'=6;
'?'=NA")

storm$PROPCASH <- storm$PROPDMG * (10^as.numeric(storm$PROPDMGEXP))
storm$CROPCASH <- storm$CROPDMG * (10^as.numeric(storm$CROPDMGEXP))
```

Results

Population Health

In order to analyse the impact on population health of the different weather events, we focus on the two variables FATALITIES and INJURIES and we sum up the two characteristics in an aggregate.

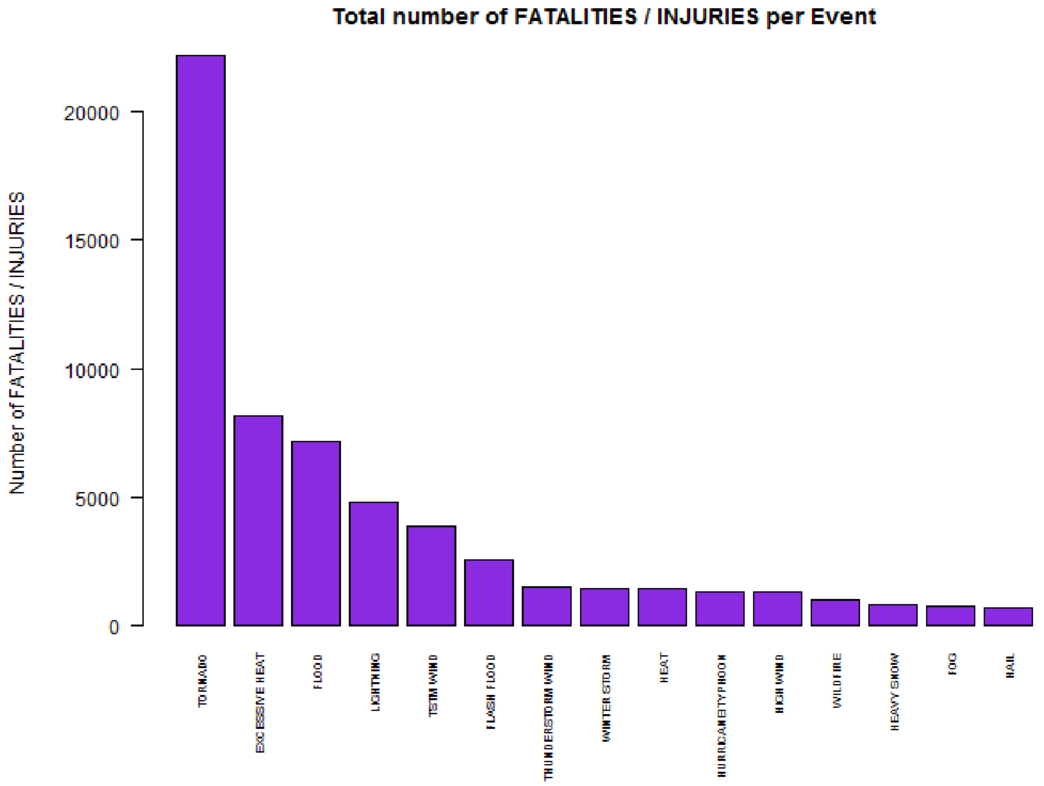
```
storm$FATAL.INJUR <- storm$FATALITIES + storm$INJURIES
fatal <- aggregate(FATAL.INJUR ~ EVTYPE, storm, sum)
fatal <- fatal[order(fatal$FATAL.INJUR, decreasing = TRUE), ]
```

```
total <- sum(fatal$FATAL.INJUR)
perc <- sum(fatal$FATAL.INJUR[1:15])/total
```

The first 15 most harmful events cause the 0.8871% of the damage in term of human health.

The following histogram describes the distribution.

```
fatal.subset <- fatal[1:15, ]
par(mar = c(6.5, 6.5, 3.5, 3.5), mgp = c(5, 1, 0))
barplot(fatal.subset$FATAL.INJUR, names.arg = fatal.subset$EVTYPE, cex.names = 0.6,
    col = "blueviolet", xlab = "", ylab = "Number of FATALITIES / INJURIES",
    main = "Total number of FATALITIES / INJURIES per Event", las = 2)
```




```
perc1 <- sum(fatal$FATAL.INJUR[1:1])/total
```

In particular, tornados are the most dangerous events and causes 0.3325% of the total damage.

Economic Effects

To measure the economic effects, we use the variables PROPCASH and CROPCASH.

We proceed in the same way of population health except this time we cannot sum up the variables because they depend differently on the kind of events.

Property and Crop Damage.

```
property <- aggregate(PROPCASH ~ EVTYPE, storm, sum)
property <- property[order(property$PROPCASH, decreasing = TRUE), ]
crop <- aggregate(CROPCASH ~ EVTYPE, storm, sum)
crop <- crop[order(crop$CROPCASH, decreasing = TRUE), ]
```

```
total1 <- sum(property$PROPCASH)
```

```
total1 <- sum(property$PROPCASH)
perc <- sum(property$PROPCASH[1:15])/total1
```

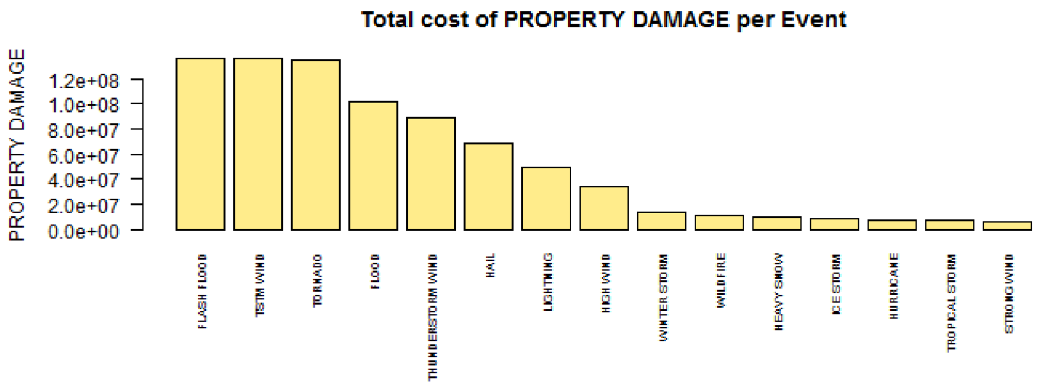
The first 15 events cause the 0.9497% of property Damage.

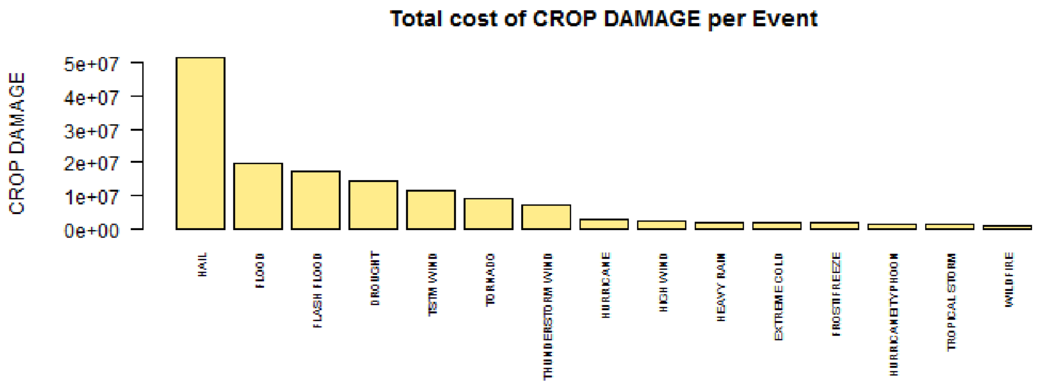
```
total2 <- sum(crop$CROPCASH)
perc <- sum(crop$CROPCASH[1:15])/total2
```

The first 15 events cause the 0.9712% of crop damage.

The following histogram describes the distribution for property and crop damage across different type of events.

```
par(mfrow = c(2, 1))
par(mar = c(6.5, 6.5, 3.5, 3.5), mgp = c(5, 1, 0))
property.subset <- property[1:15, ]
barplot(property.subset$PROPCASH, names.arg = property.subset$EVTYPE, cex.names = 0.6,
    col = "lightgoldenrod1", xlab = "", ylab = "PROPERTY DAMAGE", main = "Total cost of
PROPERTY DAMAGE per Event",
    las = 2)
crop.subset <- crop[1:15, ]
barplot(crop.subset$CROPCASH, names.arg = crop.subset$EVTYPE, cex.names = 0.6,
    col = "lightgoldenrod1", xlab = "", ylab = "CROP DAMAGE", main = "Total cost of CROP
DAMAGE per Event",
    las = 2)
```






```
perc1 <- sum(property$PROPCASH[1:1])/total1
perc2 <- sum(crop$CROPCASH[1:1])/total2
```

In particular for property damage, flash flood are the most dangerous events and causes 0.159% of damage. For crop damage, hails are the most dangerous events and causes 0.3483% of damage.

Worst events: tornado, flash flood and hail

To summarize:

- tornadoes are the most dangerous events for public health
- flash floods have the greatest economic consequences for property
- hails cause the greatest crop damage.

Moreover, these events are increasing in number during the years as we can see from the plot below.

```
library(plyr)
```

```
storm$tornado <- ifelse(storm$EVTYPE == "TORNADO", 1, 0)
tornado.sum <- storm[storm$tornado == 1, c("EVTYPE", "BGN_YEAR", "tornado")]
tornado.sum <- ddply(storm, .(BGN_YEAR), summarise, val = sum(tornado))
```

```
storm$hail <- ifelse(storm$EVTYPE == "HAIL", 1, 0)
hail.sum <- storm[storm$hail == 1, c("EVTYPE", "BGN_YEAR", "hail")]
hail.sum <- ddply(storm, .(BGN_YEAR), summarise, val = sum(hail))


storm$flood <- ifelse(storm$EVTYPE == "FLASH FLOOD", 1, 0)
flood.sum <- storm[storm$hail == 1, c("EVTYPE", "BGN_YEAR", "flood")]
flood.sum <- ddply(storm, .(BGN_YEAR), summarise, val = sum(flood))

par(mfrow = c(3, 1))
plot(tornado.sum, xlab = "Year", ylab = "Tornados", main = "Total number of Tornados",
    col = "red", type = "l")

plot(hail.sum, xlab = "Year", ylab = "Hails", main = "Total number of Hails",
    col = "green", type = "l")

plot(flood.sum, xlab = "Year", ylab = "Flash flood", main = "Total number of Flash flood",
    col = "blue", type = "l")
```