Storms and other severe weather events - Public Health and Economic Impact over cities

Edilmo Palencia

September 26, 2015

Synopsis

The NOAA (National Oceanic & Atmospheric Administration) Storm Database contains characteristics of Major weather events in the United States of America. This report use that data to discover two things: 1. Most harmful events to human health. 2. Events that have strongest damage to property and crop.

Data Processing

Before the analysis let's prepare the data. The steps to do it are as follow: * Unzip the file with the data. * Read the CSV we the data. * Discard the columns not used for the analysis. * Normalize the erros in the Event Types names using the official ones (you can found them in the documentation of the data base). This transformation is necessary because the data presents a lot of errors in the name of the events. * Apply the exponents to the economical variables and after that discard the columns of the exp variables. This transformation makes easy our future analysis.

```
# Loading zip file.
unzip.file<-bzfile("repdata-data-StormData.csv.bz2")
# Reading the CSV file.
data <- read.csv(unzip.file)
# Keep the columns necessary for the analysis.
data <- data[,c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")]
# Lets load a vector with the official Event Types
oficialEventTypes <- c("Astronomical Low Tide", "Avalanche", "Blizzard", "Coastal Flood", "Cold/Wind Ch
# Lets load a vector with the Regular Expression pattern for official Event Types
oficialEventTypesPattern <- c("Astronomical Low Tide", "Avalanche", "Blizzard", "Coastal Flood", "Cold/
# Extract the event types
levels <- as.character(data$EVTYPE)
inicialUniqueEventTypes <- length(unique(levels))
# Put to lower to easy comparison
levels <- lapply(levels, tolower)
# Put to lower the regular expression pattern
oficialEventTypesPattern <- lapply(oficialEventTypesPattern, tolower)
# Lets look for event types names that start with the official names
levelsPerOficialEventTypes <- lapply(oficialEventTypesPattern, function(x){which(grepl(paste("^", x, se
# Lets remove any suffix over the oficial event type names
for(i in 1:length(oficialEventTypesPattern)){
        for(j in levelsPerOficialEventTypes[[i]]){
                levels[j] <- oficialEventTypes[i]
        }
}
finalUniqueEventTypes <- length(unique(levels))
# Lets impute our new levels
data$EVTYPE <- factor(unlist(levels))
# Lets apply the exponent to the Property Damage variable and the Crop Damage Variable
```

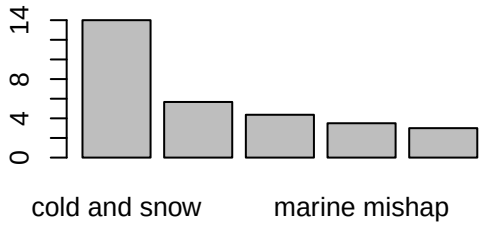
```
PROPDMG <- data$PROPDMG
CROPDMG <- data$CROPDMG
PROPDMGEXP <- as.character(data$PROPDMGEXP)
CROPDMGEXP <- as.character(data$CROPDMGEXP)
for(i in 1:length(data$PROPDMG)){
        PROPDMG[i] <- switch(PROPDMGEXP[i],
                             H = 100*PROPDMG[i],
                             h = 100*PROPDMG[i],
                             K = 1000*PROPDMG[i],
                             k = 1000*PROPDMG[i],
                             M = 1000000*PROPDMG[i],
                             m = 1000000*PROPDMG[i],
                             B = 1000000000*PROPDMG[i],
                             b = 1000000000*PROPDMG[i],
                             "+" = PROPDMG[i],
                             "-" = 0,
                             "?" = 0,
                             "0" = 10*PROPDMG[i],
                             "1" = 10*PROPDMG[i],
                             "2" = 10*PROPDMG[i],
                             "3" = 10*PROPDMG[i],
                             "4" = 10*PROPDMG[i],
                             "5" = 10*PROPDMG[i],
                             "6" = 10*PROPDMG[i],
                             "7" = 10*PROPDMG[i],
                             "8" = 10*PROPDMG[i],
                             0)
        CROPDMG[i] <- switch(CROPDMGEXP[i],
                             H = 100*CROPDMG[i],
                             h = 100*CROPDMG[i],
                             K = 1000*CROPDMG[i],
                             k = 1000*CROPDMG[i],
                             M = 1000000*CROPDMG[i],
                             m = 1000000*CROPDMG[i],
                             B = 1000000000*CROPDMG[i],
                             b = 1000000000*CROPDMG[i],
                             "+" = CROPDMG[i],
                             "-" = 0,
                             "?" = 0,
                             "0" = 10*CROPDMG[i],
                             "1" = 10*CROPDMG[i],
                             "2" = 10*CROPDMG[i],
                             "3" = 10*CROPDMG[i],
                             "4" = 10*CROPDMG[i],
                             "5" = 10*CROPDMG[i],
                             "6" = 10*CROPDMG[i],
                             "7" = 10*CROPDMG[i],
                             "8" = 10*CROPDMG[i],
                             0)
}
data$PROPDMG <- PROPDMG
data$CROPDMG <- CROPDMG
data <- data[,c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "CROPDMG")]
```

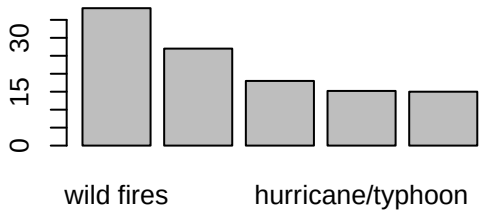
Results

Let's address the main question of this analysis:

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health? In order to answer this question lets make some plots:

```
# Computing mean fatalities per event
data.FATALITIES.avarage<-tapply(data$FATALITIES,data$EVTYPE,mean,na.rm=TRUE)
data.INJURIES.avarage<-tapply(data$INJURIES,data$EVTYPE,mean,na.rm=TRUE)
data.FATALITIESandINJURIES.avarage <- tapply(data$FATALITIES+data$INJURIES,data$EVTYPE,mean,na.rm=TRUE)
# Sorting the means
data.FATALITIES.avarage<-sort(data.FATALITIES.avarage, decreasing = TRUE)
data.INJURIES.avarage<-sort(data.INJURIES.avarage, decreasing = TRUE)
data.FATALITIESandINJURIES.avarage<-sort(data.FATALITIESandINJURIES.avarage, decreasing = TRUE)
# Taking the top five
data.FATALITIES.avarage<-data.FATALITIES.avarage[1:5]
data.INJURIES.avarage<-data.INJURIES.avarage[1:5]
data.FATALITIESandINJURIES.avarage <- data.FATALITIESandINJURIES.avarage[1:5]
# Generating plot
par(mfrow=c(2,2))
barplot(data.FATALITIES.avarage)
barplot(data.INJURIES.avarage)
barplot(data.FATALITIESandINJURIES.avarage)
```



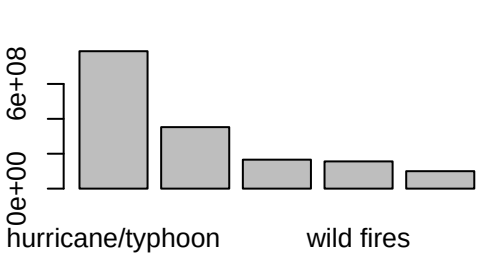


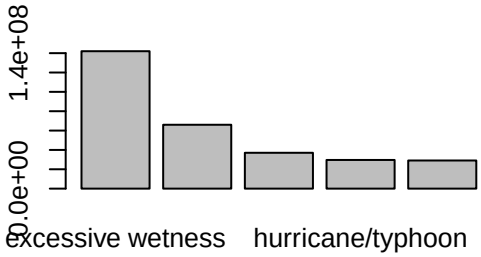
From the above reulst we can say that the most harmful events from a population health point of view: - wild fires - thunderstormw - cold and snow

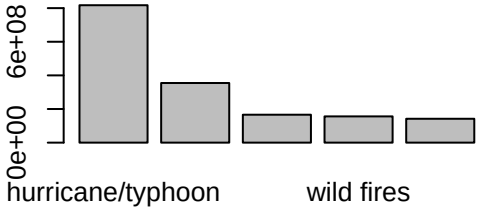
Across the United States, which types of events have the greatest economic consequences? In order to answer this question lets make some plots:

```
# Computing mean fatalities per event
data.PROPDMG.avarage<-tapply(data$PROPDMG,data$EVTYPE,mean,na.rm=TRUE)
data.CROPDMG.avarage<-tapply(data$CROPDMG,data$EVTYPE,mean,na.rm=TRUE)
```

```r
data.PROPDMGandCROPDMG.avarage <- tapply(data$PROPDMG+data$CROPDMG,data$EVTYPE,mean,na.rm=TRUE)
# Sorting the means
data.PROPDMG.avarage<-sort(data.PROPDMG.avarage, decreasing = TRUE)
data.CROPDMG.avarage<-sort(data.CROPDMG.avarage, decreasing = TRUE)
data.PROPDMGandCROPDMG.avarage<-sort(data.PROPDMGandCROPDMG.avarage, decreasing = TRUE)
# Taking the top five
data.PROPDMG.avarage<-data.PROPDMG.avarage[1:5]
data.CROPDMG.avarage<-data.CROPDMG.avarage[1:5]
data.PROPDMGandCROPDMG.avarage <- data.PROPDMGandCROPDMG.avarage[1:5]
# Generating plot
par(mfrow=c(2,2))
barplot(data.PROPDMG.avarage)
barplot(data.CROPDMG.avarage)
barplot(data.PROPDMGandCROPDMG.avarage)
```







From the above reulst we can say that the most harmful events from a population health point of view: - hurricane/typhoon - hurricane opal - excessive wetness