Severe Weather - Effects and Impact

jp4711

July 26, 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.

This data analysis involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database from 1950 to 2011. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

The data analysis in this report address the following questions:

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

# Data Processing

This analysis makes use of dplr, knitr, reshape, xtable and ggplot2 library. Documentation of dplr can be found at <http://cran.r-project.org/web/packages/dplR/dplR.pdf>

# use dplr lib   
library(dplyr)

##   
## Attaching package: 'dplyr'  
##   
## The following objects are masked from 'package:stats':  
##   
## filter, lag  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(xtable)  
library(knitr)  
library(reshape)  
library(ggplot2)

This analysis will use the following original variables:

* EVTYPE: weather event type (i.e. flood, tornado, ...)
* BGN\_DATE: beginning date of the event
* STATE: state in which the event occurred
* COUNTY: county in which the event occurred
* FATALITIES: number of human fatalities
* INJURIES: number of human injuries
* PROPDMG: a measure of the property damage
* CROPDMG: a measure of the crop damage

and to compute dollar values for damage PROPDMGEXP and CROPDMGEXP (e.g B for billions, M for millions, etc.)

### Load/Retrieve Data

# download data  
setwd("~/Courses/Data Science/repos/Reproducible Research/RepData\_PeerAssessment2")  
dataUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
dataFile <- "repdata-data-StormData.csv.bz2"  
  
if (!file.exists(dataFile)) {  
 download.file(dataUrl, dataFile, method="curl")  
}  
orgData <- read.csv(bzfile(dataFile))

The original data include **902297** records and **37** variables.

# select columns needed for this report  
data <- orgData[,c("BGN\_DATE","STATE","COUNTY","EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CROPDMG","CROPDMGEXP")]

### Compute dollar amount for property and crop damage

To use data for computation the values DMG columns have to be converted int dollar amounts.

# convertToDollar function will convert PROPDMGEXP or CROPDMGEXP   
# to the correct dollar amount (i.e. M for millions, B for billions, etc.)  
convertToDollar <- function (x) {  
 if (x == "B") {  
 1e9  
 } else if (x %in% c("m","M")) {  
 1e6  
 } else if (x %in% c("k", "K")) {  
 1e3  
 } else if (x %in% c("h", "H")) {  
 1e2  
 } else if (x %in% c("+", "-", "?")) {  
 1  
 } else {  
 0  
 }  
}

# Calculate Property and Crop Damage in dollars by converting xxxxDMGEXP   
# to the dollar amount and multiplying its dollar representative  
propDamage <- data$PROPDMG \* unlist(lapply(data$PROPDMGEXP, function(x) convertToDollar(x)))  
cropDamage <- data$CROPDMG \* unlist(lapply(data$CROPDMGEXP, function(x) convertToDollar(x)))

### Create Data Frame with columns needed for this analysis

# create data frame with dollar values as number  
data <- cbind(orgData[,c("BGN\_DATE","STATE","COUNTY","EVTYPE","FATALITIES","INJURIES")], propDamage, cropDamage)

## Compute per Event Type - Fatalities, Injuries and Damage

### 1. Total Fatalities, Injuries and Damage

totalFatalities <- sum(data$FATALITIES)  
totalInjuries <- sum(data$INJURIES)  
totalDamage <- sum(data$cropDamage + data$propDamage)  
topN\_perEvent <- 7  
topN\_State <- 10  
topN\_County <- 10  
topN\_Damage <- 10

* total # of fatalities : **1.5145 × 104**
* total # of injuries : **1.4053 × 105**
* total damage amount : **4.7642 × 1011**

### 2. Top **7** Fatalies per Event Type

dataByEventType <- group\_by(data, EVTYPE)

eventDamage <- summarise(dataByEventType,   
 fatalities = sum(FATALITIES, na.rm = TRUE),  
 injuries = sum(INJURIES, na.rm = TRUE),  
 propDamage = sum(propDamage, na.rm=TRUE),  
 cropDamage = sum(cropDamage, na.rm=TRUE),  
 totalDmg = sum(propDamage + cropDamage, na.rm=TRUE)  
)  
  
fatalitiesIdx <- order(eventDamage$fatalities, decreasing=TRUE)  
topFatalities <- eventDamage[fatalitiesIdx[1:topN\_perEvent],]

### 3. Top **7** Injuries per Event Type

injuryIdx <- order(eventDamage$injuries, decreasing=TRUE)  
topInjury <- eventDamage[injuryIdx[1:topN\_perEvent],]

## Compute per State - Fatalities, Injuries and Damage

An analysis per state was to see the impact on per state level.

### 1. Total Fatalities, Injuries and Damage by State

by\_state <- group\_by(data, STATE)  
state\_damage <- summarise(by\_state,  
 fatalities = sum(FATALITIES, na.rm=TRUE),  
 injuries = sum(INJURIES, na.rm=TRUE),  
 propDamage = sum(propDamage, na.rm=TRUE),  
 cropDamage = sum(cropDamage, na.rm=TRUE),  
 totalDmg = sum(propDamage + cropDamage, na.rm=TRUE)   
)

### 2. Top **10** Fatalies per State

fatalStateIdx <- order(state\_damage$fatalities, decreasing=TRUE)  
topFatalState <- state\_damage[fatalStateIdx[1:topN\_State],]

### 3. Top **10** Damage in dollar per State

dmgStateIdx <- order(state\_damage$totalDmg, decreasing=TRUE)  
topDmgState <- state\_damage[dmgStateIdx[1:topN\_State],]

## Compute Events with Top Damage

damageIdx <- order((eventDamage$cropDamage + eventDamage$propDamage), decreasing=TRUE)  
topDollarDmg <- eventDamage[damageIdx[1:topN\_Damage],]

# Results

## Analysis per Event Type

### Top Fatalities by Event Type

print(topFatalities[,1:2], floating=FALSE)

## Source: local data frame [7 x 2]  
##   
## EVTYPE fatalities  
## 834 TORNADO 5633  
## 130 EXCESSIVE HEAT 1903  
## 153 FLASH FLOOD 978  
## 275 HEAT 937  
## 464 LIGHTNING 816  
## 856 TSTM WIND 504  
## 170 FLOOD 470

### Top Injuries by Event Type

print(topInjury[,c(1,3)])

## Source: local data frame [7 x 2]  
##   
## EVTYPE injuries  
## 834 TORNADO 91346  
## 856 TSTM WIND 6957  
## 170 FLOOD 6789  
## 130 EXCESSIVE HEAT 6525  
## 464 LIGHTNING 5230  
## 275 HEAT 2100  
## 427 ICE STORM 1975

#kable(head(topInjury[,1:3]), format = "markdown")

### Top Economic Damage by Event Type

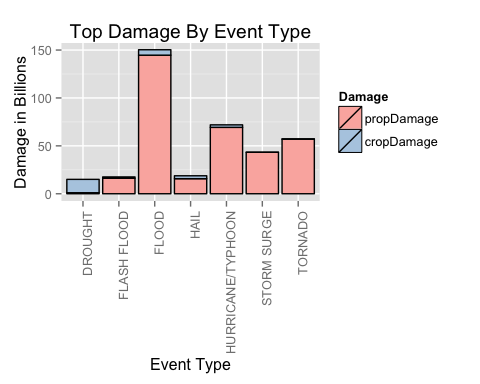
X <- topDollarDmg[,c(1,4:6)]  
X[,c(2:4)] <- X[,c(2:4)] / 1000000000  
print(X)

## Source: local data frame [10 x 4]  
##   
## EVTYPE propDamage cropDamage totalDmg  
## 170 FLOOD 144.658 5.661968 150.320  
## 411 HURRICANE/TYPHOON 69.306 2.607873 71.914  
## 834 TORNADO 56.937 0.414953 57.352  
## 670 STORM SURGE 43.324 0.000005 43.324  
## 244 HAIL 15.732 3.025954 18.758  
## 153 FLASH FLOOD 16.141 1.421317 17.562  
## 95 DROUGHT 1.046 13.972566 15.019  
## 402 HURRICANE 11.868 2.741910 14.610  
## 590 RIVER FLOOD 5.119 5.029459 10.148  
## 427 ICE STORM 3.945 5.022113 8.967

#kable(head(X), format = "markdown")

* Note: The numbers of damage are in billions.

X <- topDollarDmg[1:7,c(1,4:5)]  
X1 <- melt(X, id=(c("EVTYPE")))  
colnames(X1) <- c("EventType","Damage","Value")  
X1$Value = X1$Value / 1000000000  
   
ggplot(X1, aes(x=EventType,y=Value, fill=Damage)) +   
 geom\_bar(stat="identity", colour="black") +  
 ggtitle("Top Damage By Event Type") +  
 ylab("Damage in Billions") + xlab("Event Type") +   
 scale\_fill\_brewer(palette="Pastel1") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))

 - Note: The numbers of propDamage and cropDamage are in billions.

Obeservations:

* Tornates caused the most fatalities (over 5,000) and injuries (over 91,000).
* Floods caused the most monitary damage, over $150 billion total with over $144 billion by property damage.
* Drought has the most negative impact on crop damage.

## Analysis per State

### Top Fatalities by State

topFatalState[,1:3]

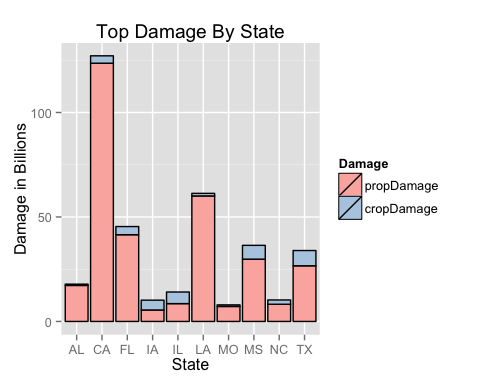
## Source: local data frame [10 x 3]  
##   
## STATE fatalities injuries  
## 20 IL 1421 5563  
## 63 TX 1366 17667  
## 51 PA 846 3223  
## 2 AL 784 8742  
## 37 MO 754 8998  
## 13 FL 746 5918  
## 38 MS 555 6675  
## 8 CA 550 3278  
## 5 AR 530 5550  
## 62 TN 521 5202

### Top Economic Damage by State

kable(head(topDmgState[,c(1,4:6)]), format = "markdown")

##   
##   
## | |STATE | propDamage| cropDamage| totalDmg|  
## |:--|:-----|----------:|----------:|---------:|  
## |8 |CA | 1.236e+11| 3.528e+09| 1.271e+11|  
## |24 |LA | 6.007e+10| 1.229e+09| 6.130e+10|  
## |13 |FL | 4.151e+10| 3.903e+09| 4.541e+10|  
## |38 |MS | 2.981e+10| 6.610e+09| 3.642e+10|  
## |63 |TX | 2.664e+10| 7.301e+09| 3.394e+10|  
## |2 |AL | 1.724e+10| 6.068e+08| 1.785e+10|

X <- topDmgState[,c(1,4:5)]  
X <- melt(X, id=(c("STATE")))  
colnames(X) <- c("State","Damage","Value")  
X$Value = X$Value / 1000000000  
   
ggplot(X, aes(x=State,y=Value, fill=Damage)) +   
 geom\_bar(stat="identity", colour="black") +  
 ggtitle("Top Damage By State") +  
 ylab("Damage in Billions") + xlab("State") +   
 scale\_fill\_brewer(palette="Pastel1")



kable(head(topDmgState[,c(1,6)]), format = "markdown")

##   
##   
## | |STATE | totalDmg|  
## |:--|:-----|---------:|  
## |8 |CA | 1.271e+11|  
## |24 |LA | 6.130e+10|  
## |13 |FL | 4.541e+10|  
## |38 |MS | 3.642e+10|  
## |63 |TX | 3.394e+10|  
## |2 |AL | 1.785e+10|

Observations:

* California has the highest damage from all states (over $127 billion), followed by Lousiana and Florida. In each the biggest bulk came from property damage.
* The top states for human fatalities and injuries are Illinois, Texas, Pennsylvania, Alabama and Missouri.

## Conclusion

The data analysis address the following questions:

* which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health across the United States.
* which types of events have the greatest economic consequences across the United States.