Reproducible Research Project 2

Antonio Quintieri

January 19, 2018

## Synopsis

National Weather Service Storm data set is read and analyzed to determine the types of storm events that are the most harmful to population health as well as the highest economic consequences. The analysis results show that the biggest source of property damage in the United States comes from Tornados, followed by Floods and Flash Flood. The highest contributor to crop damage comes from hail and is followed by flood and flash floods. The highest source of injuries and fatalities from weather events both come from tornados. The second highest fatality source comes from flood sources and is followed by lightning then wind. The second largest source of injuries comes from wind followed by lightning sources.

## Data Processing and open libraries

The following is the reading and initialization of the dataset:

#Read and initialize library and data  
 library(ggplot2)  
 library(grid)  
 library(gridExtra)  
 library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':  
##   
## combine

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

#Cache and reads data if available  
   
 storm <- read.csv("repdata\_data\_StormData.csv", stringsAsFactors=FALSE)  
head(storm)

## STATE\_\_ BGN\_DATE BGN\_TIME TIME\_ZONE COUNTY COUNTYNAME STATE  
## 1 1 4/18/1950 0:00:00 0130 CST 97 MOBILE AL  
## 2 1 4/18/1950 0:00:00 0145 CST 3 BALDWIN AL  
## 3 1 2/20/1951 0:00:00 1600 CST 57 FAYETTE AL  
## 4 1 6/8/1951 0:00:00 0900 CST 89 MADISON AL  
## 5 1 11/15/1951 0:00:00 1500 CST 43 CULLMAN AL  
## 6 1 11/15/1951 0:00:00 2000 CST 77 LAUDERDALE AL  
## EVTYPE BGN\_RANGE BGN\_AZI BGN\_LOCATI END\_DATE END\_TIME COUNTY\_END  
## 1 TORNADO 0 0  
## 2 TORNADO 0 0  
## 3 TORNADO 0 0  
## 4 TORNADO 0 0  
## 5 TORNADO 0 0  
## 6 TORNADO 0 0  
## COUNTYENDN END\_RANGE END\_AZI END\_LOCATI LENGTH WIDTH F MAG FATALITIES  
## 1 NA 0 14.0 100 3 0 0  
## 2 NA 0 2.0 150 2 0 0  
## 3 NA 0 0.1 123 2 0 0  
## 4 NA 0 0.0 100 2 0 0  
## 5 NA 0 0.0 150 2 0 0  
## 6 NA 0 1.5 177 2 0 0  
## INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES  
## 1 15 25.0 K 0   
## 2 0 2.5 K 0   
## 3 2 25.0 K 0   
## 4 2 2.5 K 0   
## 5 2 2.5 K 0   
## 6 6 2.5 K 0   
## LATITUDE LONGITUDE LATITUDE\_E LONGITUDE\_ REMARKS REFNUM  
## 1 3040 8812 3051 8806 1  
## 2 3042 8755 0 0 2  
## 3 3340 8742 0 0 3  
## 4 3458 8626 0 0 4  
## 5 3412 8642 0 0 5  
## 6 3450 8748 0 0 6

# Extracting the required data for health and economic impact analysis against weather  
  
event <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG",   
 "CROPDMGEXP")  
data <- storm[event]

## Finding crop damage

Crop damage exponents for each level was listed out and assigned those values for the crop exponent data. Invalid data was excluded by assigning the value as ‘0’. Then crop damage value was calculated by multiplying the crop damage and crop exponent value.The code for this process was listed below

# Finding the property damage exponent and levels  
unique(data$PROPDMGEXP)

## [1] "K" "M" "" "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-"  
## [18] "1" "8"

# Assigning values for the property exponent data   
data$PROPEXP[data$PROPDMGEXP == "K"] <- 1000  
data$PROPEXP[data$PROPDMGEXP == "M"] <- 1e+06  
data$PROPEXP[data$PROPDMGEXP == ""] <- 1  
data$PROPEXP[data$PROPDMGEXP == "B"] <- 1e+09  
data$PROPEXP[data$PROPDMGEXP == "m"] <- 1e+06  
data$PROPEXP[data$PROPDMGEXP == "0"] <- 1  
data$PROPEXP[data$PROPDMGEXP == "5"] <- 1e+05  
data$PROPEXP[data$PROPDMGEXP == "6"] <- 1e+06  
data$PROPEXP[data$PROPDMGEXP == "4"] <- 10000  
data$PROPEXP[data$PROPDMGEXP == "2"] <- 100  
data$PROPEXP[data$PROPDMGEXP == "3"] <- 1000  
data$PROPEXP[data$PROPDMGEXP == "h"] <- 100  
data$PROPEXP[data$PROPDMGEXP == "7"] <- 1e+07  
data$PROPEXP[data$PROPDMGEXP == "H"] <- 100  
data$PROPEXP[data$PROPDMGEXP == "1"] <- 10  
data$PROPEXP[data$PROPDMGEXP == "8"] <- 1e+08  
# Assigning '0' to invalid exponent data  
data$PROPEXP[data$PROPDMGEXP == "+"] <- 0  
data$PROPEXP[data$PROPDMGEXP == "-"] <- 0  
data$PROPEXP[data$PROPDMGEXP == "?"] <- 0  
# Calculating the property damage value  
data$PROPDMGVAL <- data$PROPDMG \* data$PROPEXP

# Exploring the crop exponent data  
unique(data$CROPDMGEXP)

## [1] "" "M" "K" "m" "B" "?" "0" "k" "2"

# Assigning values for the crop exponent data   
data$CROPEXP[data$CROPDMGEXP == "M"] <- 1e+06  
data$CROPEXP[data$CROPDMGEXP == "K"] <- 1000  
data$CROPEXP[data$CROPDMGEXP == "m"] <- 1e+06  
data$CROPEXP[data$CROPDMGEXP == "B"] <- 1e+09  
data$CROPEXP[data$CROPDMGEXP == "0"] <- 1  
data$CROPEXP[data$CROPDMGEXP == "k"] <- 1000  
data$CROPEXP[data$CROPDMGEXP == "2"] <- 100  
data$CROPEXP[data$CROPDMGEXP == ""] <- 1  
# Assigning '0' to invalid exponent data  
data$CROPEXP[data$CROPDMGEXP == "?"] <- 0  
# calculating the crop damage value  
data$CROPDMGVAL <- data$CROPDMG \* data$CROPEXP

## Finding totals of each incident by event type.

It was observed that " most harmful to population health" events are fatalities and injuries.So,only those events with fatalities and injuries were selecetd.

It was observed that " most harmful to econamic problem“” events are Property and crop damages.So,only those events with property and crop damage were selecetd.

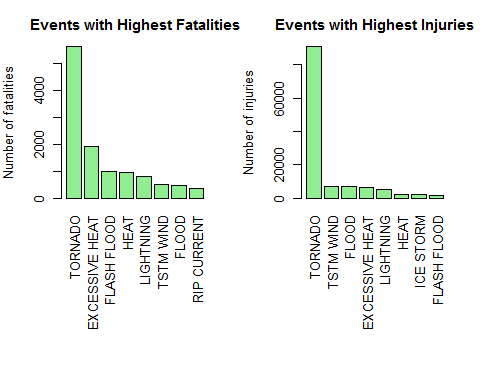
Then for each incident (Fatalities,Injuries, Property damage and Crop damage), the total values were estimated. Code for which is as follows.

# Totalling the data by event  
fatal <- aggregate(FATALITIES ~ EVTYPE, data, FUN = sum)  
injury <- aggregate(INJURIES ~ EVTYPE, data, FUN = sum)  
propdmg <- aggregate(PROPDMGVAL ~ EVTYPE, data, FUN = sum)  
cropdmg <- aggregate(CROPDMGVAL ~ EVTYPE, data, FUN = sum)

## Plotting events with highest fatalities and highest injuries.

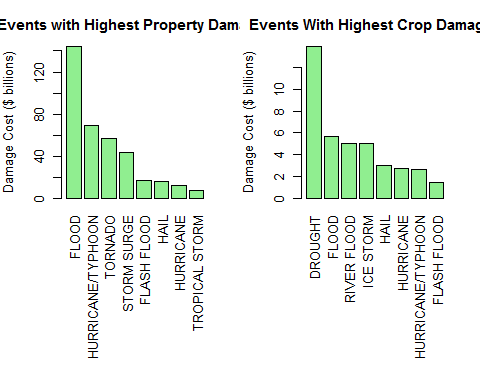
Highest fatalities and highest injuries for Top 8 events were calculated. For better understanding and comparision these values were plotted as follows.

# Listing events with highest fatalities  
fatal8 <- fatal[order(-fatal$FATALITIES), ][1:8, ]  
# Listing events with highest injuries  
injury8 <- injury[order(-injury$INJURIES), ][1:8, ]  
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)  
barplot(fatal8$FATALITIES, las = 3, names.arg = fatal8$EVTYPE, main = "Events with Highest Fatalities",   
 ylab = "Number of fatalities", col = "light green")  
barplot(injury8$INJURIES, las = 3, names.arg = injury8$EVTYPE, main = "Events with Highest Injuries",   
 ylab = "Number of injuries", col = "light green")

 ##Plotting events with highest Property damage and highest crop damage.

Highest Property damage and highest crop damage for Top 8 events were calculated. For better understanding and comparision these values were plotted as follows.

# Finding events with highest property damage  
propdmg8 <- propdmg[order(-propdmg$PROPDMGVAL), ][1:8, ]  
# Finding events with highest crop damage  
cropdmg8 <- cropdmg[order(-cropdmg$CROPDMGVAL), ][1:8, ]  
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)  
barplot(propdmg8$PROPDMGVAL/(10^9), las = 3, names.arg = propdmg8$EVTYPE,   
 main = "Events with Highest Property Damages", ylab = "Damage Cost ($ billions)",   
 col = "lightgreen")  
barplot(cropdmg8$CROPDMGVAL/(10^9), las = 3, names.arg = cropdmg8$EVTYPE,   
 main = "Events With Highest Crop Damages", ylab = "Damage Cost ($ billions)",   
 col = "lightgreen")

 ##Results and Conclusions The maximum number of fatalities and injuries are caused by Tornados. Excessive Heat follow for fatalities and Thunderstorm wind for injuries.

Floods caused the maximum property damage where as Drought caused the maximum crop damage. Second major events that caused the maximum damage was Hurricanes/Typhoos for property damage and Floods for crop damage. ## Antonio Quintieri, 18/01/18