NOAA Storm Data Analysis

Yun Yao

February 20, 2019

## 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 project explored the U.S. National Oceanic and Atmospheric Administration’s (NOAA) storm database, which 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 results of the analysis indicate that tornadoes are the most harmful with respect to population health (i.e., fatalities and injuries), whereas floods have the greatest economic consequences in the United States.

## Data Processing

This data set includes data from 1950 to 2011. The data set consists of 902,297 observations with 37 variables. The variables used for this analysis include:

1. EVTYPE - a factor variable indicating the event type (e.g. tornado, flood, etc.)
2. FATALITIES - a numerical variable indicating the number of fatalities
3. INJURIES - a numerical variable indicating the number of injuries
4. PROPDMG - a numerical variable indicating the mantissa for the value of property damage in USD
5. PROPDMGEXP - a factor variable indicating the exponent for the value of property damage in USD
6. CROPDMG - a numerical variable indicating the mantissa for the value of crop damage in USD
7. CROPDMGEXP - a factor variable indicating the exponent for the value of crop damage in USD

library(knitr)  
library(dplyr)  
library(stringr)  
library(reshape2)  
library(ggplot2)  
library(gridExtra)

Download data

if (!file.exists("./data")) {dir.create("./data")}  
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
download.file(fileUrl, destfile = "./data/storms.csv")  
  
storms <- read.csv("./data/storms.csv")  
dat <- storms[c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")]  
head(dat)

## EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP  
## 1 TORNADO 0 15 25.0 K 0   
## 2 TORNADO 0 0 2.5 K 0   
## 3 TORNADO 0 2 25.0 K 0   
## 4 TORNADO 0 2 2.5 K 0   
## 5 TORNADO 0 2 2.5 K 0   
## 6 TORNADO 0 6 2.5 K 0

Clean and manipulate data

The property damage and crop damage data are available in four columns: PROPDMG, PROPDMGEX, CROPDMG, and CROPDMGEX. The *EX columns provide the multipliers for the base dollar amount, and there four possible values in these columns: H (hundred), K (thousand), M (million), and B (billion). In order to obtain the exact dollar amount, the values in the PROPDMG and CROPDMG columns need to be multiplier by the corresponding values in the* EX columns.

dat <- dat %>% mutate(EVTYPE = str\_trim(toupper(EVTYPE)),  
 PROPDMGEXP = (toupper(PROPDMGEXP)),  
 CROPDMGEXP = (toupper(CROPDMGEXP))) %>%  
 mutate(PROPDMGEXP = ifelse(PROPDMGEXP == "H", 1E2,  
 ifelse(PROPDMGEXP == "K", 1E3,  
 ifelse(PROPDMGEXP == "M", 1E6,  
 ifelse(PROPDMGEXP == "B", 1E9, 0))))) %>%  
 mutate(CROPDMGEXP = ifelse(CROPDMGEXP == "H", 1E2,  
 ifelse(CROPDMGEXP == "K", 1E3,  
 ifelse(CROPDMGEXP == "M", 1E6,  
 ifelse(CROPDMGEXP == "B", 1E9, 0)))))  
table(dat$PROPDMGEXP)  
dat$PROPDMGEXP[dat$PROPDMGEXP == 0] <- 1  
table(dat$CROPDMGEXP)  
dat$CROPDMGEXP[dat$CROPDMGEXP == 0] <- 1  
  
dat <- dat %>% mutate(PROPDMG\_DOLLAR = PROPDMG \* PROPDMGEXP,   
 CROPDMG\_DOLLAR = CROPDMG \* CROPDMGEXP)

Analysis and calculations

totals <- dat %>% group\_by(EVTYPE) %>%  
 summarise(Fatalities = sum(FATALITIES),  
 Injuries = sum(INJURIES),  
 PropertyDamage = sum(PROPDMG\_DOLLAR),  
 CropDamage = sum(CROPDMG\_DOLLAR))  
head(totals)

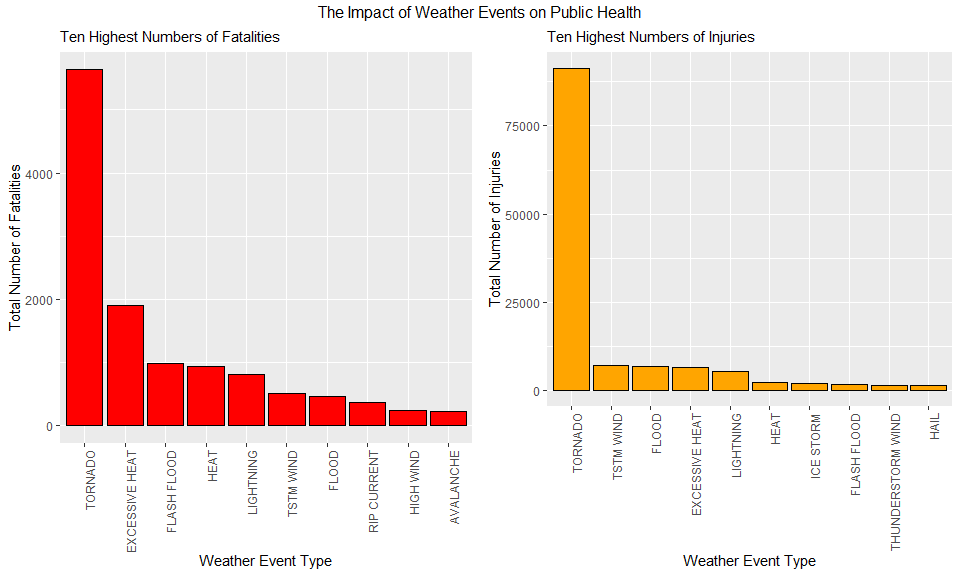
## # A tibble: 6 x 5  
## EVTYPE Fatalities Injuries PropertyDamage CropDamage  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 ? 0 0 5000 0  
## 2 ABNORMAL WARMTH 0 0 0 0  
## 3 ABNORMALLY DRY 0 0 0 0  
## 4 ABNORMALLY WET 0 0 0 0  
## 5 ACCUMULATED SNOWFALL 0 0 0 0  
## 6 AGRICULTURAL FREEZE 0 0 0 28820000

## Results

# Public health

As illustrated in the figure below, the data analysis shows that tornado is the most harmful weather event to public health, both in terms of number of fatalities and number of injuries.

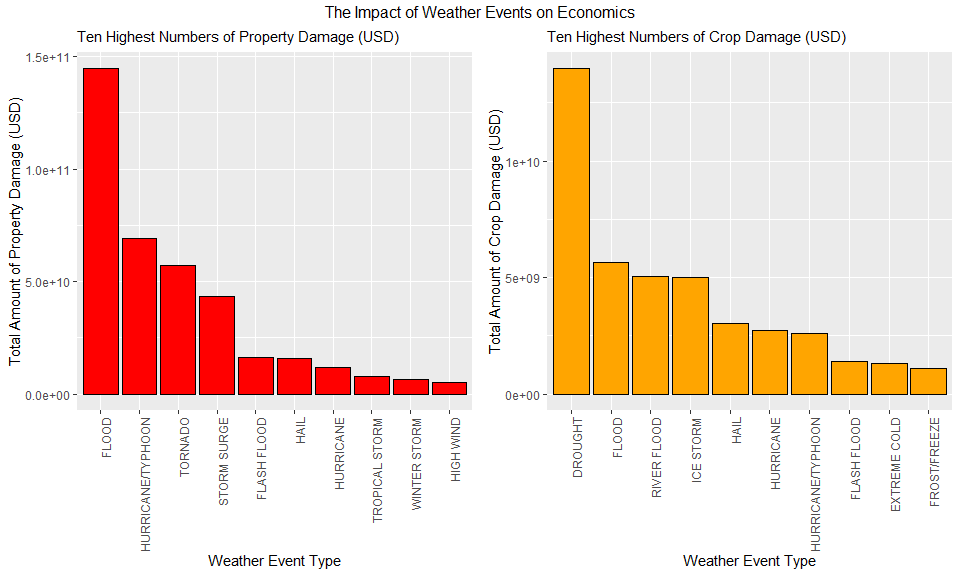
fatalities <- totals[order(totals$Fatalities, decreasing = TRUE), ][1:10, ]  
injuries <- totals[order(totals$Injuries, decreasing = TRUE), ][1:10, ]  
  
# plot - public health  
plot1 <- ggplot(data = fatalities, aes(x = reorder(EVTYPE, -Fatalities), y = Fatalities)) +   
 geom\_bar(stat = "identity", col = "black", fill = "red") +  
 xlab("Weather Event Type") +   
 ylab("Total Number of Fatalities") +  
 ggtitle("Ten Highest Numbers of Fatalities") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 theme(plot.title = element\_text(size = 11)) +  
 theme(text = element\_text(size = 11))  
  
plot2 <- ggplot(data = injuries, aes(x = reorder(EVTYPE, -Injuries), y = Injuries)) +   
 geom\_bar(stat = "identity", col = "black", fill = "orange") +  
 xlab("Weather Event Type") +  
 ylab("Total Number of Injuries") +  
 ggtitle("Ten Highest Numbers of Injuries") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 theme(plot.title = element\_text(size = 11)) +  
 theme(text = element\_text(size = 11))  
  
grid.arrange(plot1, plot2, top = "The Impact of Weather Events on Public Health",   
 layout\_matrix = matrix(c(1, 2, 1, 2), ncol = 2, byrow = TRUE))



# Economic Consequences

As illustrated in the figure below, the data analysis shows that flooding is the most harmful weather event to economics when property damage cost and crop damage cost are combined. Specifically, flooding has the greatest economic impact on property damage, whereas drought has the greatest impact on crop damage.

property <- totals[order(totals$PropertyDamage, decreasing = TRUE), ][1:10, ]  
crop <- totals[order(totals$CropDamage, decreasing = TRUE), ][1:10, ]  
  
# plot - economic consequences  
plot3 <- ggplot(data = property, aes(x = reorder(EVTYPE, -PropertyDamage), y = PropertyDamage)) +   
 geom\_bar(stat = "identity", col = "black", fill = "red") +  
 xlab("Weather Event Type") +  
 ylab("Total Amount of Property Damage (USD)") +  
 ggtitle("Ten Highest Numbers of Property Damage (USD)") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 theme(plot.title = element\_text(size = 11)) +  
 theme(text = element\_text(size = 11))  
  
plot4 <- ggplot(data = crop, aes(x = reorder(EVTYPE, -CropDamage), y = CropDamage)) +   
 geom\_bar(stat = "identity", col = "black", fill = "orange") +  
 xlab("Weather Event Type") +  
 ylab("Total Amount of Crop Damage (USD)") +  
 ggtitle("Ten Highest Numbers of Crop Damage (USD)") +  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1)) +  
 theme(plot.title = element\_text(size = 11)) +  
 theme(text = element\_text(size = 11))  
  
grid.arrange(plot3, plot4, top = "The Impact of Weather Events on Economics",   
 layout\_matrix = matrix(c(1, 2, 1, 2), ncol = 2, byrow = TRUE))



## Conculsion

The results of the analysis indicate that tornadoes are the most harmful with respect to population health (i.e., fatalities and injuries), whereas floods have the greatest economic consequences in the United States.