Storm Data Analysis with Respect to Population Health across the US

Yanal Kashou

Contents

Synopsis

This data analysis uses the NOAA Storm Database where the data analysed correlates the following:
1. Weather related events and associated repercussions on human health
2. Weather related events and their effect on property
3. Weather related events and their effect on crops

In this analysis we find the top 8 fatality and injury causing events as well as the top ten events in terms of property and crop damage caused.

Data Processing

```
# Download dataset
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
f <- file.path(getwd(), "storm_dataset.csv.bz2")
download.file(url, f)

# Read without directly decompressing .bz2 file (read.csv decompresses the .bz2
# file automatically)
storm <- read.csv("storm_dataset.csv.bz2")

# A bit of exploration
head(storm)
str(storm)
dim(storm)

# Load necessary libraries
library(plyr)
library(ggplot2)
library(reshape2)
library(qdap)
library(gridExtra)
```

Processing Health Data

```
injuries_evtype <- aggregate(INJURIES ~ EVTYPE, storm, sum)
fatalities_evtype <- aggregate(FATALITIES ~ EVTYPE, storm, sum)
```

```r
injuries_evtype.sub <- subset(injuries_evtype, injuries_evtype$INJURIES != 0)
fatalities_evtype.sub <- subset(fatalities_evtype, fatalities_evtype$FATALITIES !=
    0)

inj_evtype.desc <- arrange(injuries_evtype.sub, desc(INJURIES))
inj_evtype.desc7 <- inj_evtype.desc[1:7, ]

fat_evtype.desc <- arrange(fatalities_evtype.sub, desc(FATALITIES))
fat_evtype.desc7 <- fat_evtype.desc[1:7, ]

inj.fat.merge <- merge(inj_evtype.desc7, fat_evtype.desc7, all = TRUE)
inj.fat.desc <- arrange(inj.fat.merge, desc(FATALITIES))
inj.fat.desc
```

```
##            EVTYPE INJURIES FATALITIES
## 1         TORNADO    91346       5633
## 2 EXCESSIVE HEAT     6525       1903
## 3    FLASH FLOOD       NA        978
## 4           HEAT     2100        937
## 5       LIGHTNING    5230        816
## 6       TSTM WIND    6957        504
## 7          FLOOD     6789        470
## 8       ICE STORM    1975         NA
```

It would be inaccurate to assume that an event that caused over 100 fatalities, caused in return zero injuries. Hence we cannot replace the NA values with Zeroes. However, we can calculate the mean ratio of Injuries to Fatalities, and replace the NAs with the number from the calculated ratio.

```r
a1 <- 91345/5633
a2 <- 6525/1903
# a3 is missing Injuries count and hence reserved for ratio-based calculation a3
# = x0/978
a4 <- 2100/937
a5 <- 5230/816
a6 <- 6957/504
a7 <- 6789/470
# a8 is missing Fatalities count and hence reserved for ratio-based calculation
# a8 = 1975/y0
ratio.v <- c(a1, a2, a4, a5, a6, a7)
ratio <- mean(ratio.v)

a3 <- ratio
a8 <- ratio
x0 <- a3 * 978
x0
```

```
## [1] 9216.608
```

```r
y0 <- 1975/a8
y0
```

```
## [1] 209.5728
```

```r
# x0 is our missing Injuries count from row 3 y0 is our missing Fatalities count
# from row 8 Replace NA values with appropriately with x0 and y0
```

```r
inj.fat.desc[3, 2] <- x0
inj.fat.desc[8, 3] <- y0
```

```r
# Melt the data frames to be plotted in a single graph
melted.merge <- melt(inj.fat.desc, id.vars = "EVTYPE")
```

Processing Economic Data

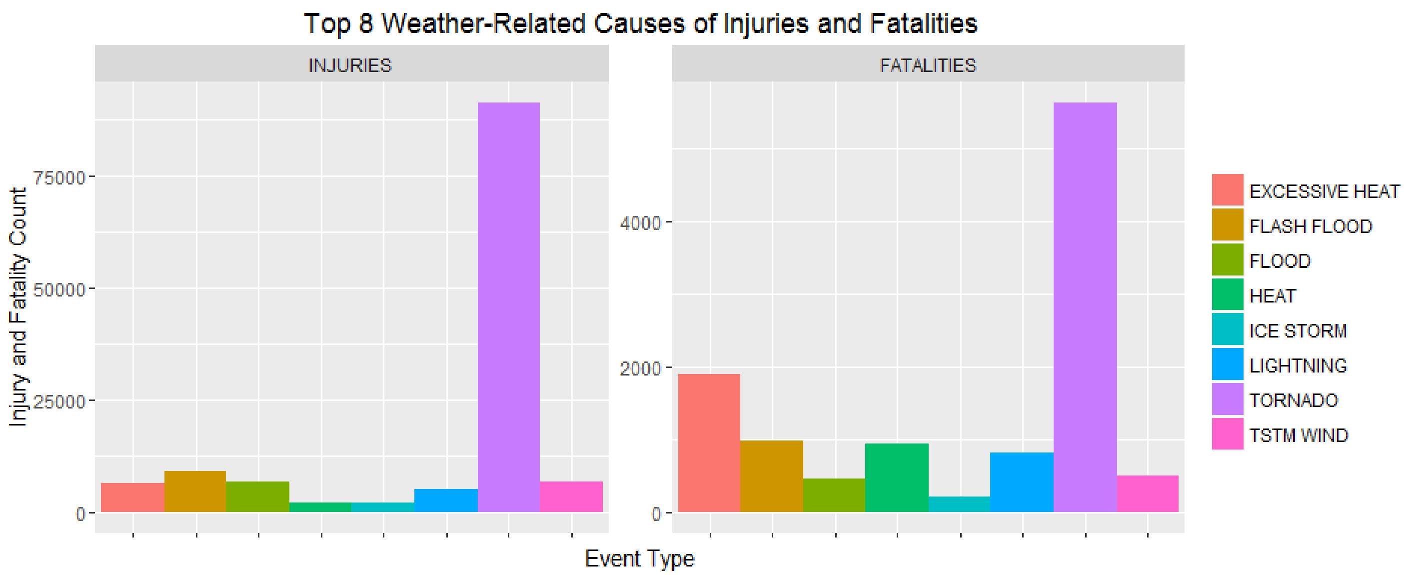
```r
# Process factors
u.prop <- unique(storm$PROPDMGEXP)
u.crop <- unique(storm$CROPDMGEXP)
u.prop

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

u.crop

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

exp.prop <- storm$PROPDMGEXP
exp.crop <- storm$CROPDMGEXP

# Map Property Damage Exponentials
prop0 <- c("B", "b", "M", "m", "K", "k", "H", "h", "-", "+", "?")
prop1 <- c(10^9, 10^9, 10^6, 10^6, 10^3, 10^3, 10^2, 10^2, 0, 0, 0)
exp.prop <- mgsub(prop0, prop1, exp.prop)

# Map Crop Damage Exponentials
crop0 <- c("B", "M", "m", "K", "k", "2", "?")
crop1 <- c(10^9, 10^6, 10^6, 10^3, 10^3, 10^2, 0)
exp.crop <- mgsub(crop0, crop1, exp.crop)

# Calculate Actual Property and Crop Damage in US Dollars
storm$prop.dmg.actual <- storm$PROPDMG * as.numeric(exp.prop)
storm$crop.dmg.actual <- storm$CROPDMG * as.numeric(exp.crop)

# Aggregate, Arrange and Select Top 10
prop.dmg <- aggregate(prop.dmg.actual ~ EVTYPE, data = storm, sum)
crop.dmg <- aggregate(crop.dmg.actual ~ EVTYPE, data = storm, sum)
prop.dmg.desc <- arrange(prop.dmg, desc(prop.dmg.actual))
crop.dmg.desc <- arrange(crop.dmg, desc(crop.dmg.actual))
prop.dmg.desc10 <- prop.dmg.desc[1:10, ]
crop.dmg.desc10 <- crop.dmg.desc[1:10, ]
```

Results

Effect of Weather-Related Events on Human Health

```r
# Plot1
p1 <- ggplot(melted.merge, aes(x = EVTYPE, y = value, fill = EVTYPE)) + geom_bar(stat =
"identity",
    width = 1) + facet_wrap(~variable, ncol = 2, scales = "free") +
scale_x_discrete(labels = abbreviate) +
    ggtitle("Top 8 Weather-Related Causes of Injuries and Fatalities") +
scale_fill_discrete(name = NULL) +
    xlab("Event Type") + ylab("Injury and Fatality Count") + theme(axis.text.x =
```

```
element_blank())
p1
```

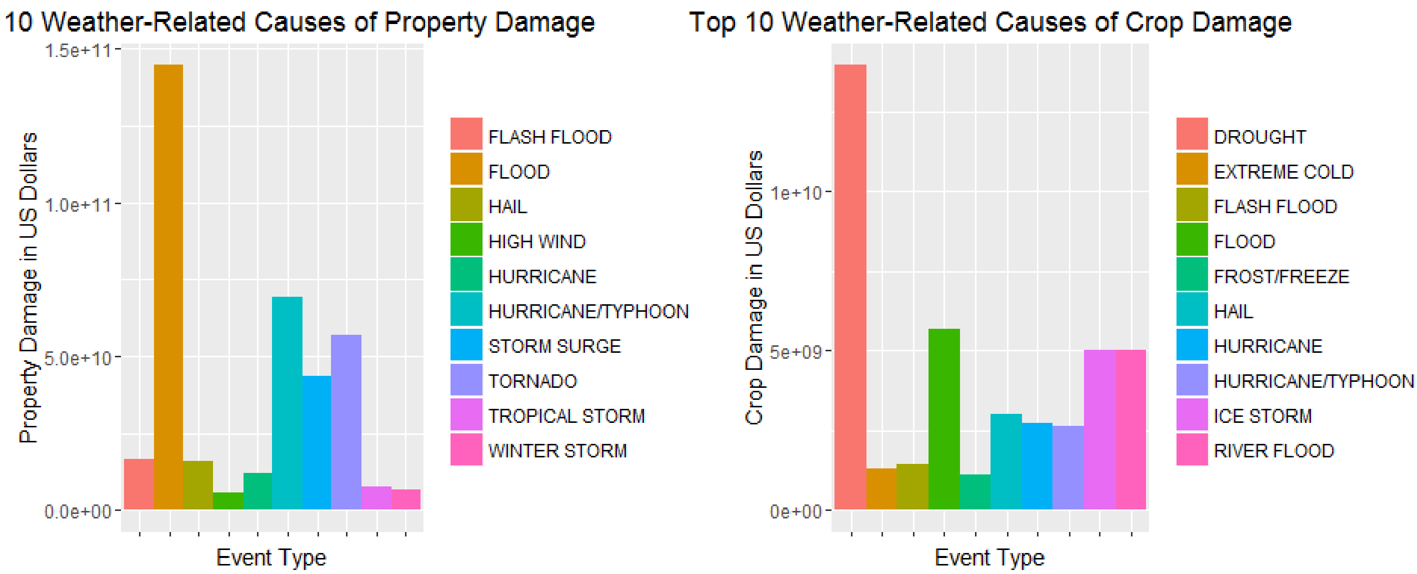


We can clearly see that Tornadoes have caused the greatest number of fatalities as well as injuries.

Effect of Weather-Related Events on Economy (Properties and Crops)

```
prop.dmg.desc10

##                EVTYPE prop.dmg.actual
## 1               FLOOD    144657709800
## 2   HURRICANE/TYPHOON     69305840000
## 3             TORNADO     56937160991
## 4         STORM SURGE     43323536000
## 5         FLASH FLOOD     16140812087
## 6                HAIL     15732267370
## 7           HURRICANE     11868319010
## 8      TROPICAL STORM      7703890550
## 9        WINTER STORM      6688497250
## 10          HIGH WIND      5270046260

crop.dmg.desc10

##                EVTYPE crop.dmg.actual
## 1             DROUGHT     13972566000
## 2               FLOOD      5661968450
## 3         RIVER FLOOD      5029459000
## 4           ICE STORM      5022113500
## 5                HAIL      3025954450
## 6           HURRICANE      2741910000
## 7   HURRICANE/TYPHOON      2607872800
## 8         FLASH FLOOD      1421317100
## 9        EXTREME COLD      1292973000
## 10       FROST/FREEZE      1094086000
```

```
p2 <- ggplot(prop.dmg.desc10, aes(x = EVTYPE, y = prop.dmg.actual, fill = EVTYPE)) +
    geom_bar(stat = "identity", width = 1) + scale_x_discrete(labels = abbreviate) +
    ggtitle("Top 10 Weather-Related Causes of Property Damage") +
scale_fill_discrete(name = NULL) +
    xlab("Event Type") + ylab("Property Damage in US Dollars") + theme(axis.text.x =
element_blank())

p3 <- ggplot(crop.dmg.desc10, aes(x = EVTYPE, y = crop.dmg.actual, fill = EVTYPE)) +
    geom_bar(stat = "identity", width = 1) + scale_x_discrete(labels = abbreviate) +
    ggtitle("Top 10 Weather-Related Causes of Crop Damage") + scale_fill_discrete(name =
NULL) +
    xlab("Event Type") + ylab("Crop Damage in US Dollars") + theme(axis.text.x =
element_blank())

grid.arrange(p2, p3, nrow = 1, ncol = 2)
```



We can clearly see that Floods, in specific Flash Floods have caused the greatest property damage in US dollars.
While Droughts caused the greatest crop damage in US dollars.