# Analyzing Storm Data to find events causing maximum consequences

*Rahulg13*

*29/12/2019*

This report explores the NOAA dataset to answer two questions - 1. Which are the events most harmful to population health? 2. Which events are associated with greatest economic consequences?

# 1. DATA PROCESSING

## 1.1 Data download and reading

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "stormdata.csv
x <- read.csv("stormdata.csv.bz2")
```

## 1.2 Data Exploration

1. On exploration of dataset, it is found there 985 event types.
2. We try to find out the events with - 1. Maximum fatalities (FATALITIES) 2. Maximum injuries (INJURIES) 3. Maximum property damage (PROPDMG) 4. Maximum crop damage (CROPDMG) For this, we find totals of fatalities, injuries, property damage and crop damage using the function "tapply". Next, we order them in a decreasing order to help find the events causing maximum damage.

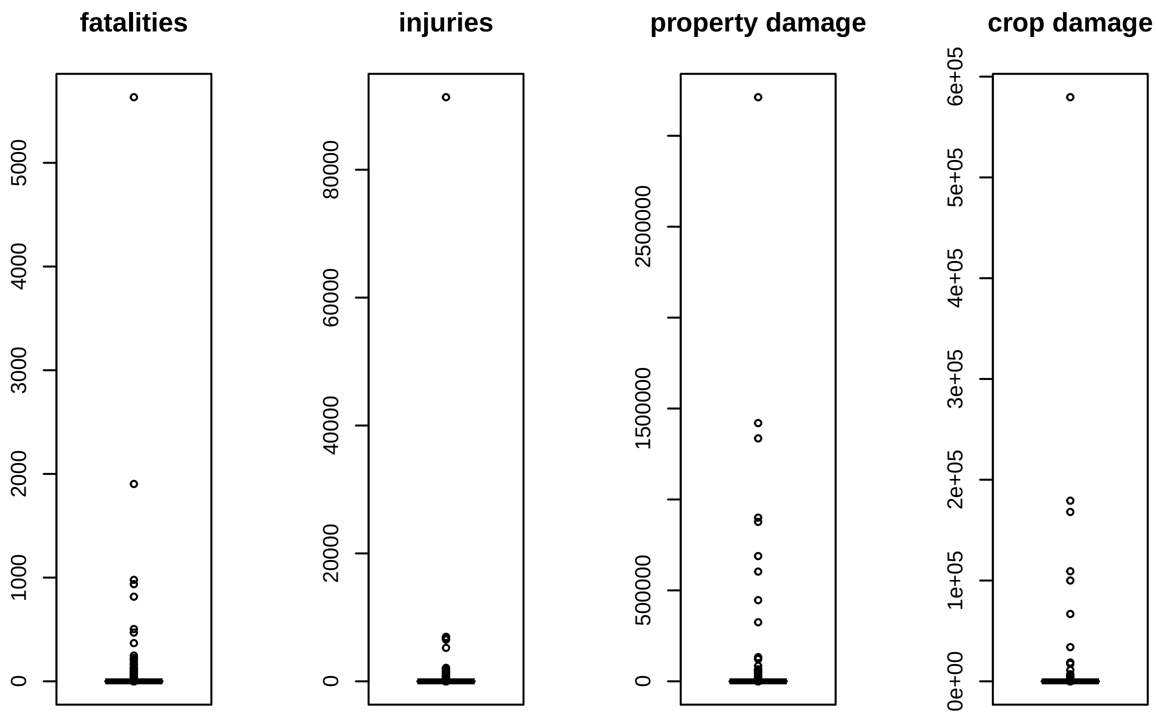
```
library(dplyr)
```

```
##
## 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(ggplot2)
eventwise_fat <- with(x, tapply(FATALITIES, EVTYPE, sum))
eventwise_inj <- with(x, tapply(INJURIES, EVTYPE, sum))
eventwise_prop <- with(x, tapply(PROPDMG, EVTYPE, sum))
eventwise_crop <- with(x, tapply(CROPDMG, EVTYPE, sum))
eventwise_fat <- eventwise_fat[order(eventwise_fat, decreasing = TRUE)]
eventwise_inj <- eventwise_inj[order(eventwise_inj, decreasing = TRUE)]
eventwise_prop <- eventwise_prop[order(eventwise_prop, decreasing = TRUE)]
eventwise_crop <- eventwise_crop[order(eventwise_crop, decreasing = TRUE)]
```

To look at the distribution of damage among events, we draw boxplots of the four relevant variables.

```
par(mfrow = c(1, 4))
boxplot(eventwise_fat, main = "fatalities")
boxplot(eventwise_inj, main = "injuries")
boxplot(eventwise_prop, main = "property damage")
boxplot(eventwise_crop, main = "crop damage")
```




Clearly, from above, the damage caused in every type is limited to a few event types where most of the damage is concentrated. To confirm the findings above, we draw the element wise quantiles. (Note 985 is the number of type of events, so 985 quantiles represent one event for each quantile.)

```
tail(quantile(eventwise_fat, prob = seq(0, 1, length.out = 985)), 5)
```

```
##  99.5934959%  99.6951220%  99.7967480%  99.8983740% 100.0000000%
##          816          937          978         1903         5633
```

```
tail(quantile(eventwise_inj, prob = seq(0, 1, length.out = 985)), 5)
```

```
##  99.5934959%  99.6951220%  99.7967480%  99.8983740% 100.0000000%
##         5230         6525         6789         6957        91346
```

```
tail(quantile(eventwise_crop, prob = seq(0, 1, length.out = 985)), 5)
```

```
## 99.5934959%  99.6951220%  99.7967480%  99.8983740% 100.0000000%
##    100018.5     109202.6     168037.9     179200.5     579596.3
```

```
tail(quantile(eventwise_prop, prob = seq(0, 1, length.out = 985)), 5)
```

```
## 99.5934959%  99.6951220%  99.7967480%  99.8983740% 100.0000000%
##    876844.2     899938.5    1335965.6    1420124.6    3212258.2
```

The result confirms that *top 5 events contribute more than 90% of damage in each category.*

## 1.3 Data in a format suitable for plotting

To detail the events, we subset the data for top 10 categories of event types in each variable category i.e. fatalities, injuries, property damage and crop damage.

```
maxfat_events <- names(eventwise_fat[1:10])
maxinj_events <- names(eventwise_inj[1:10])
x_final <- subset(x, EVTYPE %in% c(maxfat_events, maxinj_events))

maxprop_events <- names(eventwise_prop[1:10])
maxcrop_events <- names(eventwise_crop[1:10])
x_final2 <- subset(x, EVTYPE %in% c(maxprop_events, maxcrop_events))
```

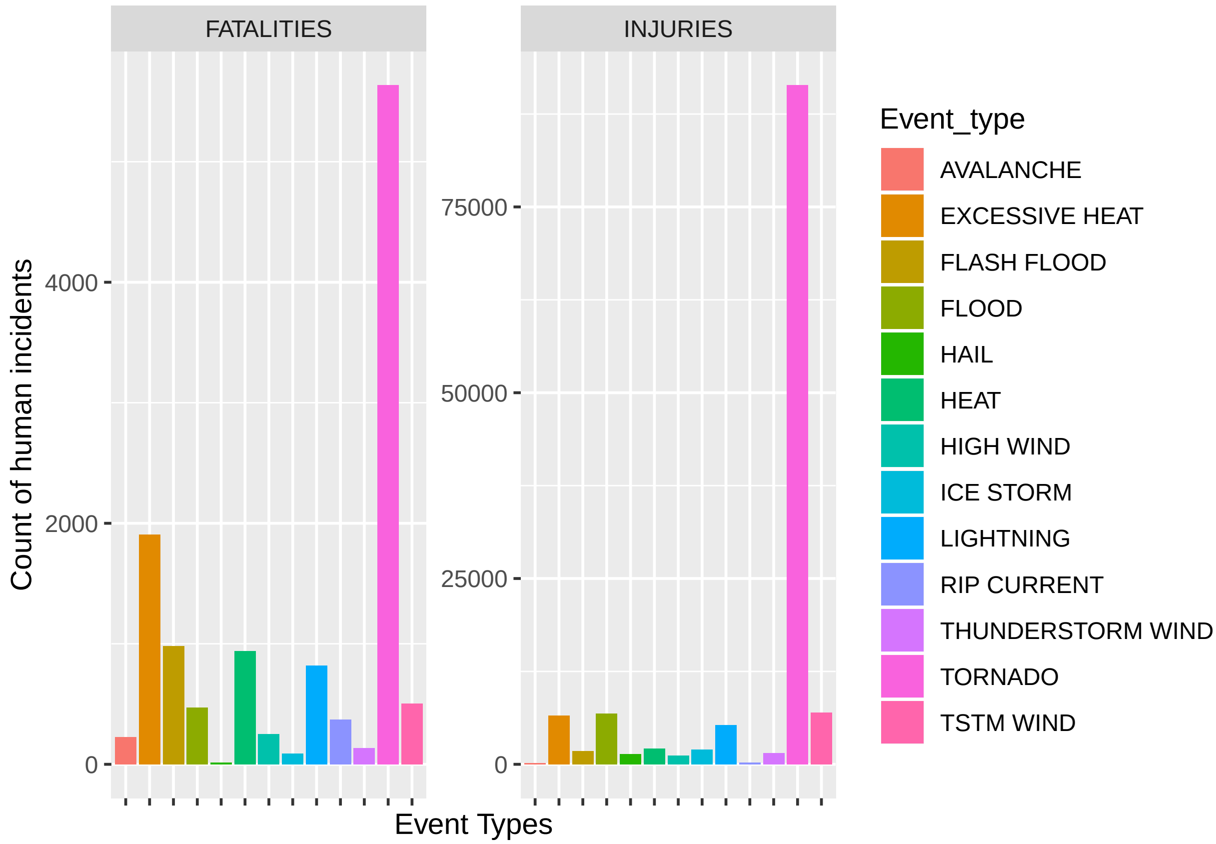
### 1.4 New dataset for plotting effect on health of human population

The dataset combines the fatalities and injuries with a separate variable/column to differentiate between the two. Event type is taken from original dataset.

```
l <- length(x_final$FATALITIES)
effect_count <- c(x_final$FATALITIES, x_final$INJURIES)
effect_type <- c(rep("FATALITIES", l), rep("INJURIES", l))
event_type <- c(x_final$EVTYPE, x_final$EVTYPE)
x3 <- data.frame(Effect_count = effect_count, Effect_type = effect_type, Event_type =  event_type)

x3$Event_type <- as.factor(x3$Event_type)
levelnames <- as.numeric(levels(x3$Event_type))
levelnames <- levels(x$EVTYPE)[levelnames]
levels(x3$Event_type) <- levelnames
```

### Plot of major events harming health of human population

```
g <- ggplot(x3) + geom_bar(aes(x = Event_type, y = Effect_count, fill = Event_type), stat = "identity")
g <- g + facet_wrap( ~ Effect_type, scales = "free")
g <- g + theme(axis.text.x = element_blank())
g <- g + xlab("Event Types") + ylab("Count of human incidents")
g
```



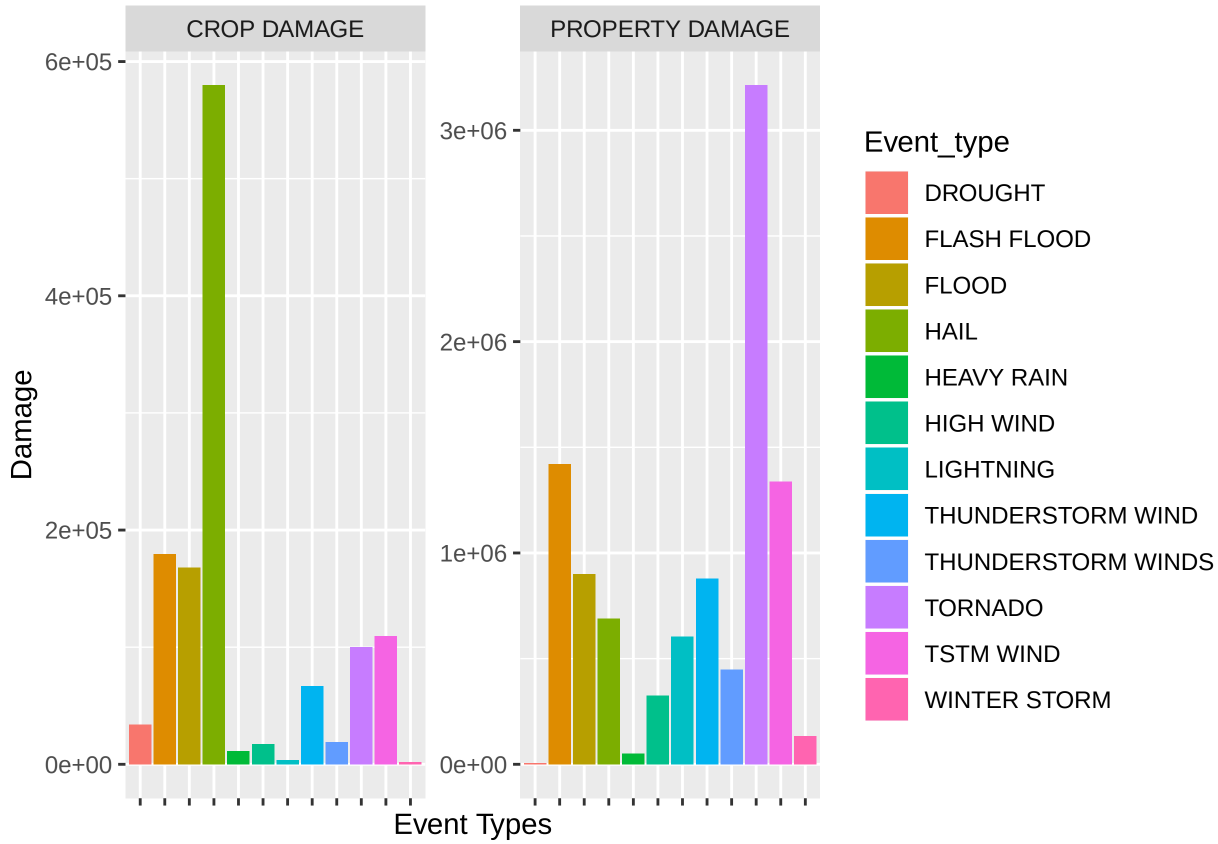

Similarly, we carry out the analysis for property damage and crop damage

### New dataset for plotting effect on damage to property and crops

```
l <- length(x_final2$PROPDMG)
effect_count <- c(x_final2$PROPDMG, x_final2$CROPDMG)
effect_type <- c(rep("PROPERTY DAMAGE", l), rep("CROP DAMAGE", l))
event_type <- c(x_final2$EVTYPE, x_final2$EVTYPE)
x3 <- data.frame(Effect_count = effect_count, Effect_type = effect_type, Event_type =  event_type)

x3$Event_type <- as.factor(x3$Event_type)
levelnames <- as.numeric(levels(x3$Event_type))
levelnames <- levels(x$EVTYPE)[levelnames]
levels(x3$Event_type) <- levelnames
```

### Plot for Property damage and Crop damage

```
#plotting event wise total damage
g <- ggplot(x3) + geom_bar(aes(x = Event_type, y = Effect_count, fill = Event_type), stat = "identity")
g <- g + facet_wrap( ~ Effect_type, scales = "free")
g <- g + theme(axis.text.x = element_blank())
g <- g + xlab("Event Types") + ylab("Damage")
g
```




# RESULTS

Based on the analysis and plots above, following conclusions can be drawn - ## 1. Health of human population -

1. Tornadoes are, by far, the biggest cause of fatalities and injuries for population. It constitute more than 50% of both fatalities and injuries.
2. Excessive Heat is the second major cause of fatalities with flash floods at the third position
3. For injuries, flash floods and hail come next to tornadoes

## 2. Economics consequences

1. For Crop damage, hail is the biggest cause. Flash floods and floods are distant second and third caused of crop damage.
2. For Property damage, tornadoes are the biggest reason. Flash floods comes distant second and TSTM winds are a close third.