Population Health and Economic Impact of Different Types of Storms in the United States of America

Balogun Stephen Taiye

SYNOPSIS

Storms and other severe weather conditions have both public health and economic impacts to the nation. Loss of lives and injuries, damages to crops and properties are huge losses the direct and indirect impact of which can have dire consequences in the longer term. This study was conducted using the storm database of the National Oceanic and Atmospheric Adminstrations (NOAA). The data was collected between 1950 and November 2011. Population health impact measures *Fatalities* and *Injuries* caused by storms while the Economic impact measures (in dollars) values of *properties* and *crops* damaged during storms.
The study shows that **Tornado** storms has the highest population health impact.
Flood storms however has the highest economic impact.

DATA PROCESSING

The downloaded file used for this studies is a `csv.bz2` zipped file named `rawData`

```
library(knitr)
opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE)

## reading in the data and unzipping the file
rawData <-
"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
if(!file.exists("rawData.csv.bz2")) {
download.file(url = rawData, destfile = "./rawData.csv.bz2")
}

library(R.utils)  ## to unzip the bz2 formatted file
if(!file.exists("rawData.csv")) {
bunzip2(filename = "./rawData.csv.bz2", destname = "./rawData.csv")
}

rawData <- read.csv("./rawData.csv")  ## reads in the csv file
```

The rawData was downloaded on 2015-08-21.
The `rawData` contains 902297 rows and 37 columns.
Next a bit of data processing is done to make the data more tidy and subset the required data for this analysis.

```
names(rawData) <- tolower(names(rawData))  ## changes column names from
capital letters to small letters
```

```
names(rawData) <- gsub(pattern = "*_", replacement = ".", x = names(rawData))
## replaces '_' in column names with '.'
library(dplyr)
requiredData <- select(rawData, c(evtype, fatalities, injuries, propdmg,
propdmgexp,
    cropdmg, cropdmgexp))
library(printr)  ## to display nicely formatted tables
head(requiredData)
```

evtype	fatalities	injuries	propdmg	propdmgexp	cropdmg	cropdmgexp
TORNADO	0	15	25.0	K	0	
TORNADO	0	0	2.5	K	0	
TORNADO	0	2	25.0	K	0	
TORNADO	0	2	2.5	K	0	
TORNADO	0	2	2.5	K	0	
TORNADO	0	6	2.5	K	0	

The `propdmgexp` and the `cropdmgexp` columns of the data represents the exponential values of properties damaged and crops damaged respectively.
The letters **[hH], K, [mM] and B** represents **hundreds, thousands millions and billions** respectively.

These letters are replaced with their corresponding exponents in numbers using the code below.

```
requiredData$propdmgexp <- gsub(pattern = "[hH]", replacement = 2,
                                x = requiredData$propdmgexp)
requiredData$propdmgexp <- gsub(pattern = "K", replacement = 3,
                                x = requiredData$propdmgexp)
requiredData$propdmgexp <- gsub(pattern = "[mM]", replacement = 6,
                                x = requiredData$propdmgexp)
requiredData$propdmgexp <- gsub(pattern = "B", replacement = 9,
                                x = requiredData$propdmgexp)
requiredData$cropdmgexp <- gsub(pattern = "[kK]", replacement = 3,
                                x = requiredData$cropdmgexp)
requiredData$cropdmgexp <- gsub(pattern = "[mM]", replacement = 6,
                                x = requiredData$cropdmgexp)
requiredData$cropdmgexp <- gsub(pattern = "B", replacement = 6,
                                x = requiredData$cropdmgexp)
```

Now that the *letter exponents* have been replaced by their *numeric* equivalents, i proceed to multiply the `propdmg` with the `propdmgexp` and the `cropdmg` with the `cropdmgexp` to get the absolute values of `propertiesDamaged` and `cropsDamaged` respectively.

```
Data <- mutate(requiredData,
               propertiesDamaged = propdmg * 10^as.numeric(propdmgexp))
Data <- mutate(Data,
               cropDamaged = cropdmg * 10^as.numeric(cropdmgexp))
```

The processed `Data` now looks like this:

```
head(Data)
```

evtype	fatalities	injuries	propdmg	propdmgexp	cropdmg	cropdmgexp	propertiesDamaged	cropDamaged
TORNADO	0	15	25.0	3	0		25000	NA
TORNADO	0	0	2.5	3	0		2500	NA
TORNADO	0	2	25.0	3	0		25000	NA
TORNADO	0	2	2.5	3	0		2500	NA
TORNADO	0	2	2.5	3	0		2500	NA
TORNADO	0	6	2.5	3	0		2500	NA

Finally for data processing, i remove the columns used for the merging since they are no longer needed. I call this new data `finalData`

```
finalData <- select(Data, -c(propdmg, propdmgexp, cropdmg, cropdmgexp))
head(finalData)
```

evtype	fatalities	injuries	propertiesDamaged	cropDamaged
TORNADO	0	15	25000	NA
TORNADO	0	0	2500	NA
TORNADO	0	2	25000	NA
TORNADO	0	2	2500	NA
TORNADO	0	2	2500	NA
TORNADO	0	6	2500	NA

RESULTS

across the US, which types of events are most harmful with respect to population health?

To answer this question, I made a subset of the final data containing the event type `evtype`, `fatalities` and `injuries`; add-up the total of `fatalities` and `injuries` for each of the `evtype` and arrange them from the most harmful to the least harmful.

```
impactOnPopulation <- finalData %>%
                        group_by(evtype) %>%
                        summarise(fatalities = sum(fatalities, na.rm = TRUE),
                        injuries = sum(injuries, na.rm = TRUE))

impactFatalities <- arrange(impactOnPopulation[, 1:2], desc(fatalities))
head(impactFatalities)
```

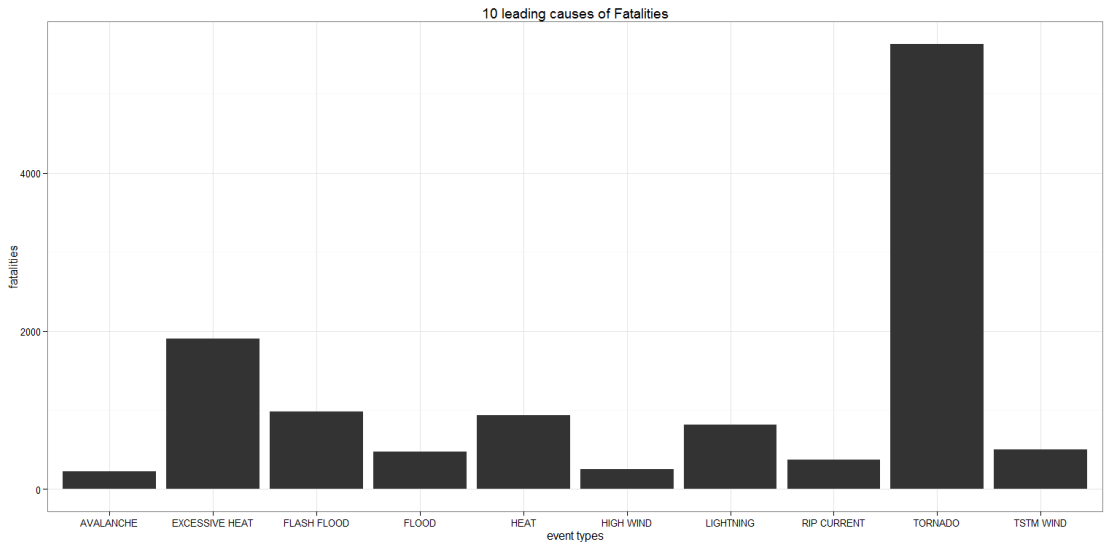
evtype	fatalities
TORNADO	5633
EXCESSIVE HEAT	1903
FLASH FLOOD	978
HEAT	937
LIGHTNING	816
TSTM WIND	504

```
impactInjuries <- arrange(impactOnPopulation[, c(1, 3)], desc(injuries))
head(impactInjuries)
```

evtype	injuries
TORNADO	91346
TSTM WIND	6957
FLOOD	6789
EXCESSIVE HEAT	6525
LIGHTNING	5230
HEAT	2100

Next I plot a bar graph of `fatalities` and `injuries` for the 10 most harmful event types using the `ggplot2` plotting system.

```
library(ggplot2)
plot1 <- ggplot(data = impactFatalities[1:10, ], aes(x = evtype, y =
fatalities))
plot1 <- plot1 + geom_bar(stat = "identity") +
        labs(title = "10 leading causes of Fatalities", x = "event types") +
        theme_bw()
print(plot1)
```

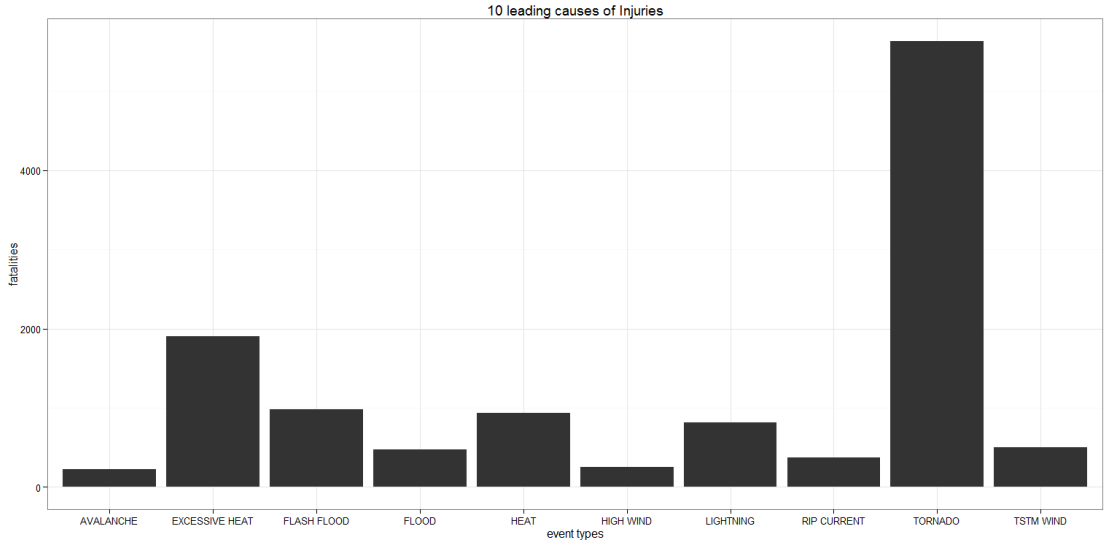



```
percent <- with(impactFatalities, round((fatalities[1] / sum(fatalities) *
100), digits = 2))
print(percent)

[1] 37.19
```

This shows that `evtype tornado` has the highest fatalities accounting for 37.19% of all fatalities.

```
plot2 <- ggplot(data = impactInjuries[1:10, ], aes(x = evtype))
plot2 <- plot1 + geom_bar(stat = "identity") +
        labs(title = "10 leading causes of Injuries", x = "event types") +
        theme_bw()
print(plot2)
```



```
percent2 <- with(impactInjuries, round((injuries[1] / sum(injuries) * 100),
digits = 2))
print(percent2)

[1] 65
```

This also shows that `evtype tornado` has the highest injuries accounting for 65% of all fatalities.
Therefore, `evtype tornado` has the most harmful population health effect of all the event types.
across the US, which types of event have the greatest economic consequence
First I made a subset of the `finalData` containing `evtype`, `propertiesDamaged`, and `cropsDamaged`; add-up the total for each of the `evtype`, then and the columns `propertiesDamaged¬ andcropdDamaged`` together and arrange the output in descending order.

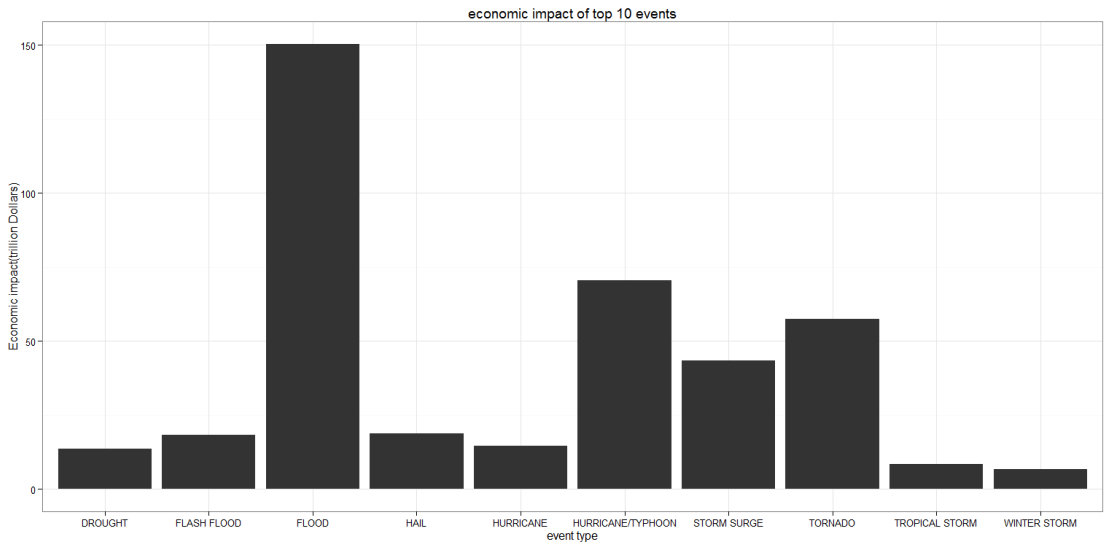
```
impactOnEconomy <- finalData %>%
                        group_by(evtype) %>%
                        summarise(propertiesDamaged = sum(propertiesDamaged,
                        na.rm = TRUE),
                        cropsDamaged = sum(cropDamaged, na.rm =TRUE))
impactOnEconomy <- mutate(impactOnEconomy,
                          impact = propertiesDamaged + cropsDamaged)
impactOnEconomy <- arrange(impactOnEconomy, desc(impact))

head(impactOnEconomy)
```

evtype	propertiesDamaged	cropsDamaged	impact
FLOOD	144657709800	5661968450	150319678250
HURRICANE/TYPHOON	69305840000	1099382800	70405222800
TORNADO	56947380614	414953270	57362333884
STORM SURGE	43323536000	5000	43323541000
HAIL	15735267456	3025954470	18761221926
FLASH FLOOD	16822673772	1421317100	18243990872

Lastly, I plot a bar graph of the economic consequences(impact) of the top ten event types.

```
plot3 <- ggplot(data = impactOnEconomy[1:10, ], aes(x = evtype, y = impact /
10^9))
plot3 <- plot3 + geom_bar(stat = "identity") +
        labs(title = "economic impact of top 10 events", x = "event type",
             y = "Economic impact(trillion Dollars)") +
        theme_bw()
print(plot3)
```




```
percent3 <- with(impactOnEconomy, round((impact[1] / sum(impact) * 100),
digits = 2))
print(percent3)

[1] 32.42
```

This also shows that `evtype flood` has the highest impact on the economy accounting for 32.42% of all economic losses.

CONCLUSION

The study shows that from the data available, `tornado` storms have the highest population health impact while `flood` has the highest economic health impact.

```
sessionInfo()

## R version 3.2.2 (2015-08-14)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 7 x64 (build 7601) Service Pack 1
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_1.0.1     printr_0.0.4      dplyr_0.4.2       R.utils_2.1.0
## [5] R.oo_1.19.0       R.methodsS3_1.7.0 knitr_1.11
```

```
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.0      magrittr_1.5     MASS_7.3-43      munsell_0.4.2
##  [5] colorspace_1.2-6 R6_2.1.0         stringr_1.0.0    highr_0.5
##  [9] plyr_1.8.3       tools_3.2.2      parallel_3.2.2   grid_3.2.2
## [13] gtable_0.1.2     DBI_0.3.1        htmltools_0.2.6  yaml_2.1.13
## [17] lazyeval_0.1.10  assertthat_0.1   digest_0.6.8     reshape2_1.4.1
## [21] formatR_1.2      codetools_0.2-14 evaluate_0.7.2   rmarkdown_0.7
## [25] labeling_0.3     stringi_0.5-5    scales_0.2.5     proto_0.3-10
```