The Impact of Weather Events in the U.S.A

Thuan G Pham

August 21, 2016

Introduction

Severe Weather events have tremendous impact to the population across the United States such as Katrina and many others.

The NOAA Storm Database tracks the severe weather events.

Using the storm data provided by NOAA which can be accessed here [https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2], this project will explore the following questions of weather events across the United States:

1. Which types of events are most harmful with respect to population health?

2. Which types of events have the greatest economic consequences?

Synopsis

The analysis of the data shows that Tornadoes are the most dangerous weather events in the U.S. with respect to fatalities follow by Excessive Heat, Flash Flood, Heat and Lightning. We also find that Tornadoes are the most dangerous weather events in the U.S. with respect to injuries follow by Thunderstorm wind, Flood, Excessive Heat, And Lightning. These events are quite harmful to the population.

The analysis of the data also shows that Floods are weather events which caused the most property damage in the U.S follow by Huricane, tornado, storm surge and Flash flood. We also find that Drought caused the most crop damage in the U.S follow by Flood, River Flood, ice storm and hail.

Load the data

```
library(plyr)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.1

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.3.1

setwd("C:/DataScience/Reproducible Research/W4 Project")
stormData <- read.csv(bzfile("repdata-data-StormData.csv.bz2"))
```

Which events are most harmful with respect to the population health?

The NOAA storm data has 2 columns which indicate fatalities and injuries. We will extract and tally these information for each event and sort based on the highest value to find which are the top 5 event types are the most harmful to the population.

```
harmfulEvents <- stormData[,c("EVTYPE","FATALITIES","INJURIES")]

harmfulEventsSummarize <- ddply(harmfulEvents, .(EVTYPE), summarize,
                    fatalities = sum(FATALITIES),
                    injuries = sum(INJURIES))

fatalities_events <-
head(harmfulEventsSummarize[order(harmfulEventsSummarize$fatalities,
decreasing = T), ], 5)

injuries_events <-
head(harmfulEventsSummarize[order(harmfulEventsSummarize$injuries, decreasing
= T), ],5)
```

The Top 5 Severe Weather Events with Highest Fatalities

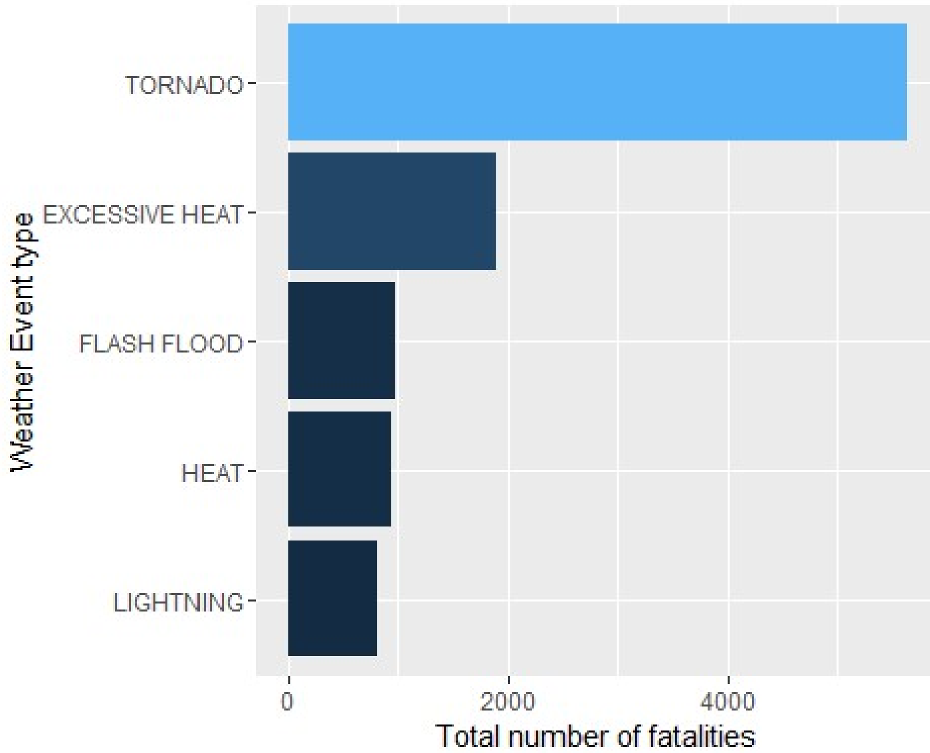
The data shows that the top 5 weather event with the highest fatalities are:

```
print(fatalities_events[,c("EVTYPE","fatalities")])

##             EVTYPE fatalities
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
```

The plot of the top 5 Severe Weather events with highest Fatalities

```
ggplot(data=fatalities_events,
             aes(x=reorder(EVTYPE, fatalities), y=fatalities,
fill=fatalities)) +
    geom_bar(stat="identity") +
    coord_flip() +
    ylab("Total number of fatalities") +
    xlab("Weather Event type") +
    theme(legend.position="none")
```



Top 5 Severe Weather Events with Highest Injuries

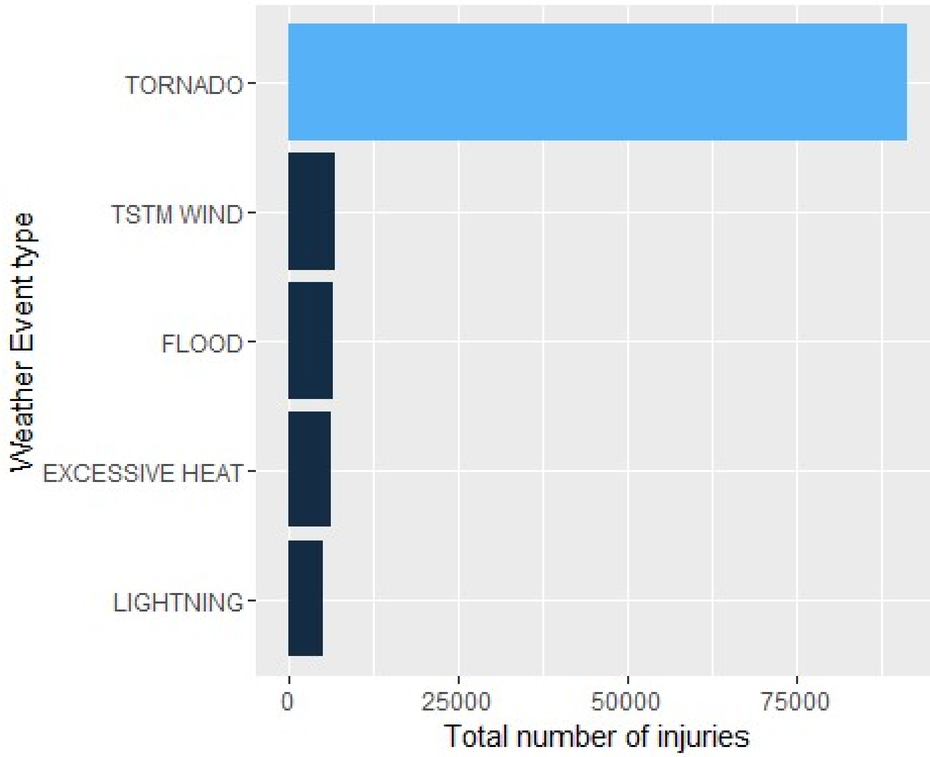
The data shows that the top 5 weather event with the highest Injuries are:

```
print(injuries_events[,c("EVTYPE","injuries")])
```

```
##              EVTYPE injuries
## 834         TORNADO    91346
## 856       TSTM WIND     6957
## 170           FLOOD     6789
## 130  EXCESSIVE HEAT     6525
## 464       LIGHTNING     5230
```

The plot of the top 5 Severe Weather events with highest Injuries

```
ggplot(data=injuries_events,
             aes(x=reorder(EVTYPE, injuries), y=injuries, fill=injuries)) +
    geom_bar(stat="identity") +
    coord_flip() +
    ylab("Total number of injuries") +
    xlab("Weather Event type") +
    theme(legend.position="none")
```



Which types of events have the greatest economic consequences?

The PROPDMG and CROPDMG columns in the storm data contain the Property Damage cost and Crop Damage cost respectively. Calculating that will allow us see which weather event types have the most economic consequences.

Calculating the Property Damage

Since the columns PROPDMGEXP and CROPDMGEXP contain the exponent indicator of the values in hundreds (h), thousands (k), millions (m), and billions (B), I am creating a function named GetDMGValue to calculate the real cost for each row based on the value of the PROPDMGEX and CROPDMGEXP.

```
GetDMGValue<- function(e,val) {
    # h -> hundred, k -> thousand, m -> million, b -> billion
    if (e %in% c('h', 'H'))
        return((10**2) * val)
    else if (e %in% c('k', 'K'))
        return ((10**3) * val)
    else if (e %in% c('m', 'M'))
        return ((10**6) * val)
    else if (e %in% c('b', 'B'))
        return ((10**9) * val)
    else if (!is.na(as.numeric(e))) # if a digit
        return(as.numeric(e))
    else if (e %in% c('', '-', '?', '+'))
```

```r
        return(0)
    else {
        stop("Invalid exponent value.")
    }
}

stormData$PROGDMGVal<-
mapply(GetDMGValue,stormData$PROPDMGEXP,stormData$PROPDMG)
stormData$CROPDMGVal<-
mapply(GetDMGValue,stormData$CROPDMGEXP,stormData$CROPDMG)
```

With the property damage cost and crop cost values added to storm data data frame, I am goin to summarize the cost for each types and sort them to select the top 5 event per categories.

```r
costlyEvents <- stormData[,c("EVTYPE","PROGDMGVal","CROPDMGVal")]

costlyEventsSummarize <- ddply(costlyEvents, .(EVTYPE), summarize,
                    propertyCost = sum(PROGDMGVal),
                    cropCost = sum(CROPDMGVal))

propCostEvents <-
head(costlyEventsSummarize[order(costlyEventsSummarize$propertyCost,
decreasing = T), ], 5)

cropCostEvents <-
head(costlyEventsSummarize[order(costlyEventsSummarize$cropCost, decreasing =
T), ],5)
```

Top 5 Severe Weather Events with Highest Property Damage

The data shows that the top 5 weather event with the highest Property Damage are:

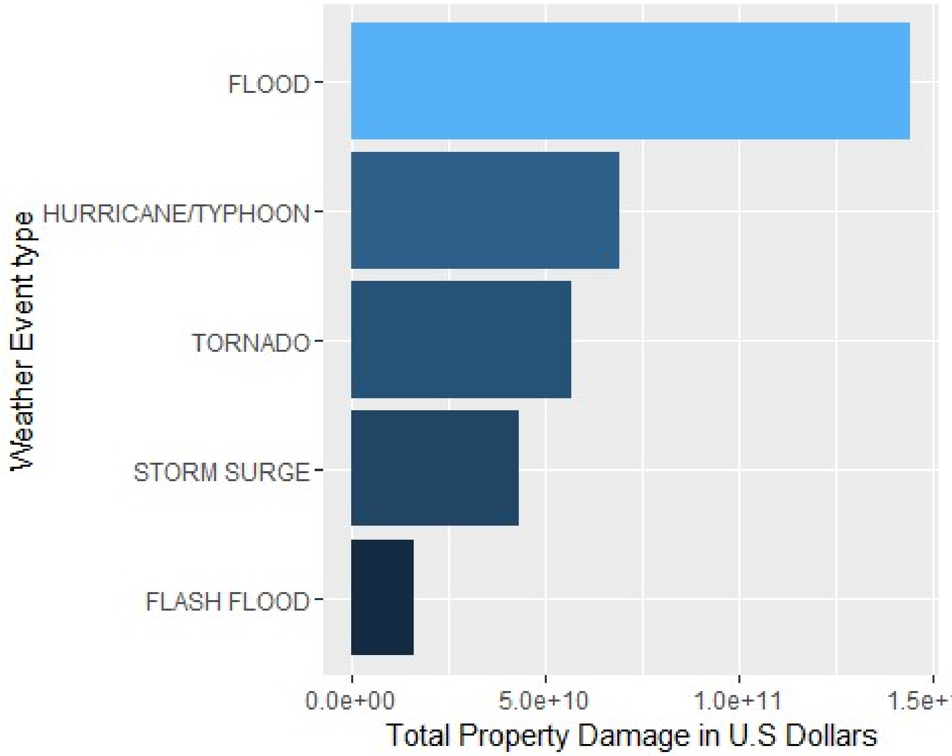
```r
print(propCostEvents[,c("EVTYPE","propertyCost")])

##                EVTYPE propertyCost
## 170             FLOOD 144657717736
## 411 HURRICANE/TYPHOON  69305840018
## 834           TORNADO  56937169417
## 670       STORM SURGE  43323536086
## 153       FLASH FLOOD  16140832926
```

The plot of the top 5 Severe Weather events with Highest Property Damage

```r
ggplot(data=propCostEvents,
             aes(x=reorder(EVTYPE, propertyCost), y=propertyCost,
fill=propertyCost)) +
    geom_bar(stat="identity") +
    coord_flip() +
    ylab("Total Property Damage in U.S Dollars") +
```

```r
    xlab("Weather Event type") +
    theme(legend.position="none")
```



The Top 5 Severe Weather Events with Highest Crop Damage

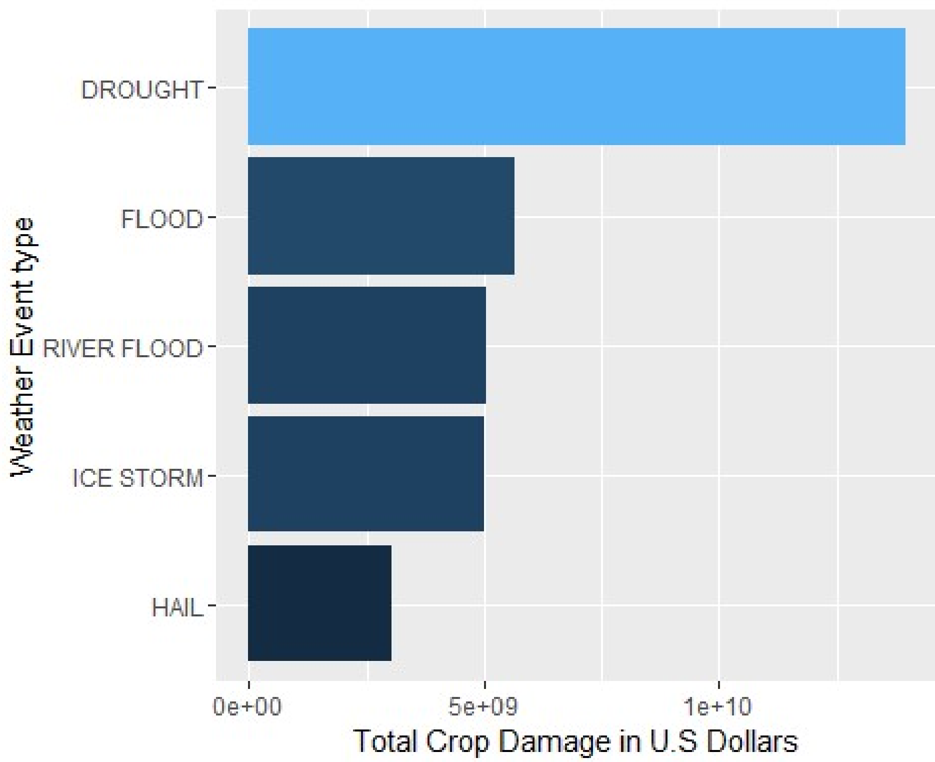
The data shows that the top 5 weather event with the highest Crop Damage are:

```r
print(cropCostEvents[,c("EVTYPE","cropCost")])
```

```
##           EVTYPE    cropCost
## 95       DROUGHT 13972566971
## 170        FLOOD  5661980154
## 590  RIVER FLOOD  5029459153
## 427    ICE STORM  5022114545
## 244         HAIL  3026160812
```

The plot of the top 5 Severe Weather events with Highest Crop Damage

```r
ggplot(data=cropCostEvents,
             aes(x=reorder(EVTYPE, cropCost), y=cropCost, fill=cropCost)) +
    geom_bar(stat="identity") +
    coord_flip() +
    ylab("Total Crop Damage in U.S Dollars") +
    xlab("Weather Event type") +
    theme(legend.position="none")
```


DROUGHT
FLOOD
RIVER FLOOD
ICE STORM
HAIL
Weather Event type
0e+00
5e+09
1e+10
Total Crop Damage in U.S Dollars