# Severe Weather Events Study using NOAA Storm Data from year 1950 till November 2011

*Dipali Bagad*

*August 9, 2017*

# Synopsis

This R markdown file analyzes Storm Data from U.S.National Oceanic and Atmospheric Administration's (NOAA) database to answer below questions about severe weather events:

1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

# Data Processing:

Step1: Download and unzip strom data stored in bz2 file to csv file. Renamed the file name to StormDataset.csv

Step2: We will read content of this file as a raw data for our analysis. We will load this data into variable StormDataset to make our data accessibility easier

Step3: We will process and analze this data to answer assignment questions.

Step4: 'EVTYPE' variable and time variable used to see which is the most harmful event respect to population health. Also, varaibles related to economic consequences will be used to see which event is the influential to economic of USA.

Step5: Results from the analysis will be generated and file will be uploaded to Rpubs.

# Reading Data

We first read data from the comma-separated-value file for data analysis i.e. StormDataset.csv file.

```
library(ggplot2)
library(lattice)

StormDataset <-read.csv("StormData.csv")

#After reading data from year 1950 to Nov, 2011 we check the first few rows rows in th
is dataset
dim(StormDataset)
```

```
## [1] 902297     37
```

```
head(StormDataset[, c(8, 23:29)])
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO
## 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
```

# 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```
#aggregate data of injuries and fatalities for each event type
Event <- with(StormDataset, aggregate(INJURIES + FATALITIES ~ EVTYPE, data=StormDatase
t, FUN = "sum"))

# Change the name for understanding
names(Event)[2] <- "Totalcasualities"

# Order the number of accidents or casualities by decending order to get top prioritiz
e events causing more casualities
Prioritize_events <- Event[order(-Event$Totalcasualities),]
Prioritize_top_events <- head(Prioritize_events)
```

# Data processing for Question 2:

2.Across the United States, which types of events have the greatest economic consequences?

For data processing we will use StormDataset variable from above which has all required dataset.

There are measurements judging economic losses, property damages and crop damages.hence we are going to use propdmg, propdmgexp, cropdmg, cropdmgexp columns to understand economic sequences. Also, we need to make sure all damages are with the same unit, i.e. million dollars, combining the information afforded by four variables propdmg, propdmgexp, cropdmg, cropdmgexp in the dataset. To perform this conversion, we first have a look at propdmgexp and cropdmgexp. We learn from the Storm Data Documentation that it does not account for the characters like “-”, “+” or “?” and the numbers like “1”, “2”, etc. so we will ignore them we are considering “h” or “H” means 102, “k” or “K” means 103, “m” or “M” means 106 and “b” or “B” means 109.

```
#understand PROPDMGEXP and CROPDMGEXP data
table(StormDataset$PROPDMGEXP)
```

```
##
##             -      ?      +      0      1      2      3      4      5
## 465934      1      8      5    216     25     13      4      4     28
##      6      7      8      B      h      H      K      m      M
##      4      5      1     40      1      6 424665      7  11330
```

```
table(StormDataset$CROPDMGEXP)
```

```
##
##             ?      0      2      B      k      K      m      M
## 618413      7     19      1      9     21 281832      1   1994
```

```
#get data only with M,K,B from CROPDMGEXP and PROPDMGEXP
# CROPDMG and PROPDMG value will be multipied accordingly with "M", "K", "B" which mea
n 1000000, 1000, 1000000000 respectively.
# Character with 'B', 'K', 'M' are assigned values accordingly and Other than these mu
ltiplier will be 1


SubsetData <- subset(StormDataset, (StormDataset$CROPDMGEXP == "M" | StormDataset$CROP
DMGEXP == "K" | StormDataset$CROPDMGEXP == "B"  | StormDataset$CROPDMGEXP == "m" | Sto
rmDataset$CROPDMGEXP == "k" | StormDataset$CROPDMGEXP == "b") | (StormDataset$PROPDMGE
XP =="M" | StormDataset$PROPDMGEXP =="K" | StormDataset$PROPDMGEXP =="B" | StormDatase
t$PROPDMGEXP =="m" | StormDataset$PROPDMGEXP =="k" | StormDataset$PROPDMGEXP =="b") )

for(i in 1:length(SubsetData$CROPDMGEXP)) {
                ifelse(SubsetData$CROPDMGEXP[i] == "M" | SubsetData$CROPDMGEXP[i] ==
"m", SubsetData$CROPDMG[i] <- SubsetData$CROPDMG[i] * 1000000,
                ifelse(SubsetData$CROPDMGEXP[i] == "K" | SubsetData$CROPDMGEXP[i] ==
"k", SubsetData$CROPDMG[i] <- SubsetData$CROPDMG[i] * 1000,
                ifelse(SubsetData$CROPDMGEXP[i] == "B" | SubsetData$CROPDMGEXP[i] ==
"b", SubsetData$CROPDMG[i] <- SubsetData$CROPDMG[i] * 1000000000, SubsetData$CROPDMG
[i] <- SubsetData$CROPDMG[i] * 1)))
}


for(i in 1:length(SubsetData$PROPDMGEXP)) {
                ifelse(SubsetData$PROPDMGEXP[i] == "M" | SubsetData$PROPDMGEXP[i] ==
"m", SubsetData$PROPDMG[i] <- SubsetData$PROPDMG[i] * 1000000,
                ifelse(SubsetData$PROPDMGEXP[i] == "K" | SubsetData$PROPDMGEXP[i] ==
"k", SubsetData$PROPDMG[i] <- SubsetData$PROPDMG[i] * 1000,
                ifelse(SubsetData$PROPDMGEXP[i] == "B" | SubsetData$PROPDMGEXP[i] ==
"b", SubsetData$PROPDMG[i] <- SubsetData$PROPDMG[i] * 1000000000, SubsetData$PROPDMG
[i] <- SubsetData$PROPDMG[i] * 1)))
}


summary(SubsetData$PROPDMGEXP)
```

```
##             -      ?      +      0      1      2      3      4      5
##   4312      0      0      0      5      0      0      1      0      2
##      6      7      8      B      h      H      K      m      M
##      0      0      0     40      0      0 424665      7  11330
```

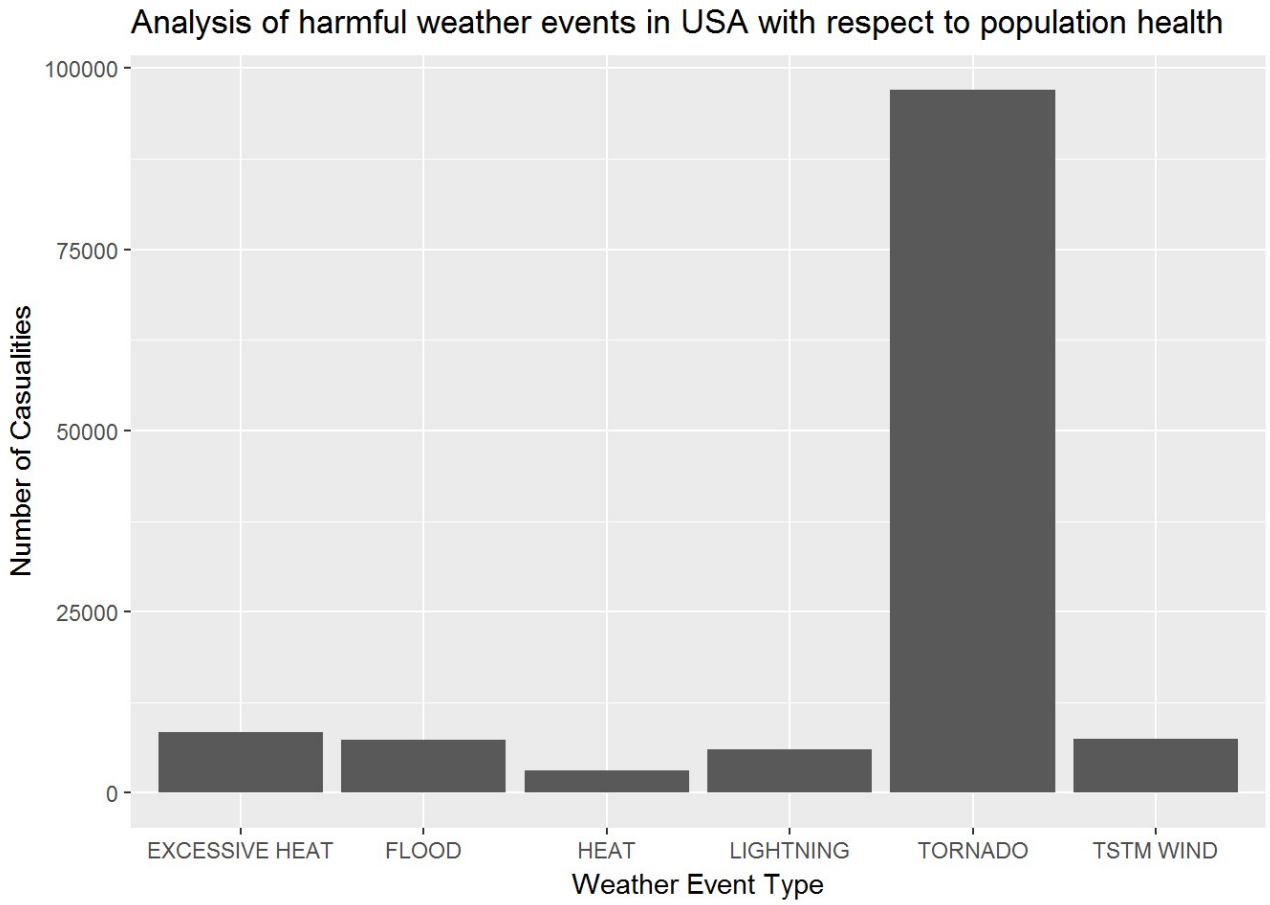
```
summary(SubsetData$CROPDMGEXP)
```

```
##            ?      0      2      B      k      K      m      M 
## 156484      5     16      0      9     21 281832      1   1994
```

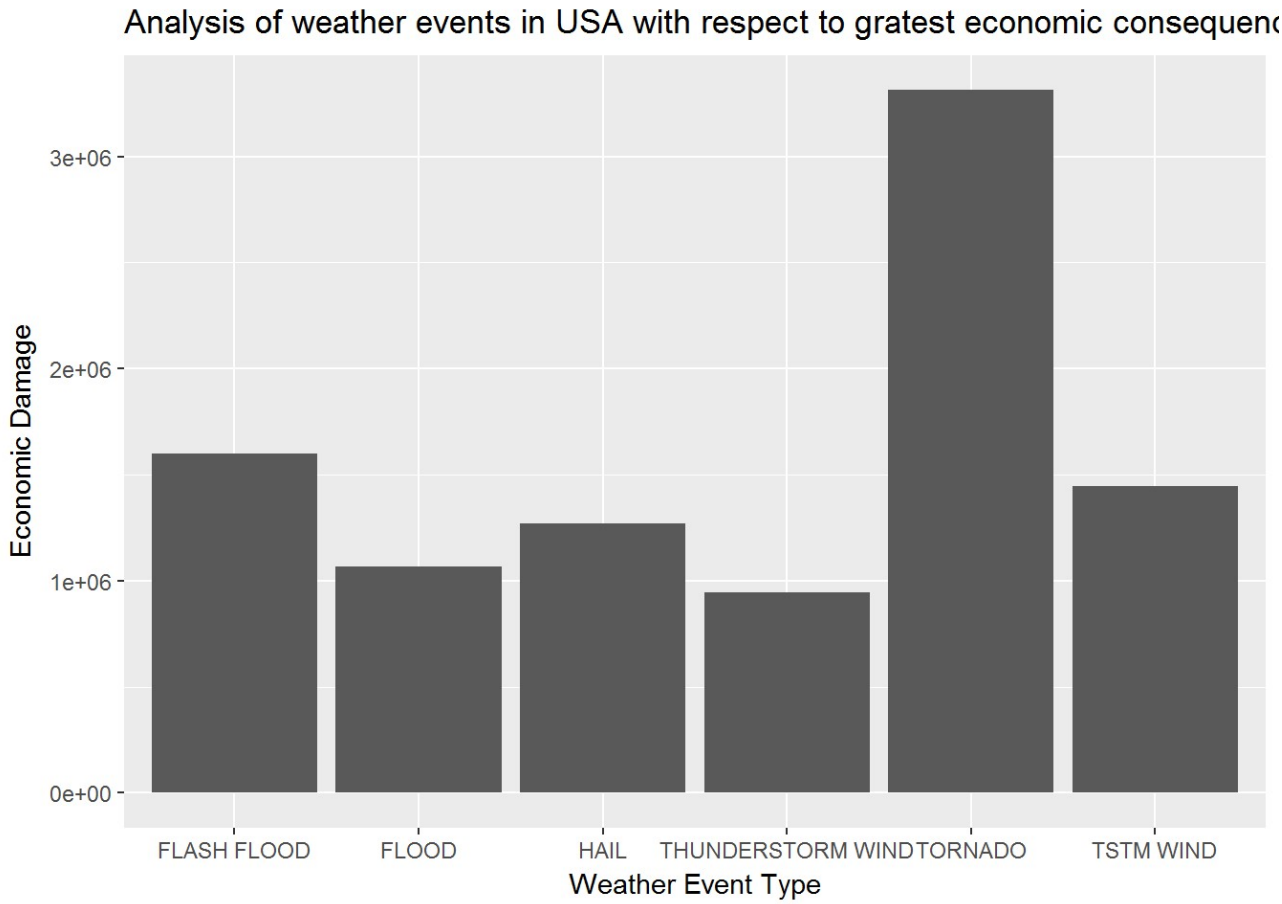
```
#aggregate final subset data for CROPDMG and PROPDMG for each event
EconomicDamage <- with(SubsetData, aggregate(CROPDMG + PROPDMG ~ EVTYPE, data=StormDat
aset, FUN = "sum"))

# Change the name of second column for understanding and set in descending order
names(EconomicDamage )[2] <- "TotalEconomicDamage"

Prioritized_EconomicDamage  <- EconomicDamage [order(-EconomicDamage$TotalEconomicDama
ge),]

# get top economic damage using head function as it is in descending order
Top_Economic_Damage <- head(Prioritized_EconomicDamage)
```

# Results

```
# Draw a graph to represent analysis
ggplot(Prioritize_top_events, aes( EVTYPE, Totalcasualities))  + geom_bar(stat = "iden
tity") +
    ylab("Number of Casualities") + xlab("Weather Event Type") + ggtitle("Analysis of
harmful weather events in USA with respect to population health")
```




```
# Draw a graph to represent analysis of economic damage
ggplot(Top_Economic_Damage, aes( EVTYPE, TotalEconomicDamage))  + geom_bar(stat = "ide
ntity") +
  ylab("Economic Damage") + xlab("Weather Event Type") + ggtitle("Analysis of weather
events in USA with respect to gratest economic consequences")
```




“Most harmful weather event in USA based on NOAA database is:”

```
Prioritize_top_events$EVTYPE[1]
```

```
## [1] TORNADO
## 985 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```

“Weather Event with gratest economic consequences in USA based on NOAA database is:”

```
Top_Economic_Damage$EVTYPE[1]
```

```
## [1] TORNADO
## 985 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```