Exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database

Synopsis

This report provides an analysis on the influence of severe weather events that can cause both public health and economic problems for communities and municipalities in the United States. # Data Processing

```r
library(dplyr)
library(R.utils)
url <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
setwd("G:\\datascience\\reproducible research\\project2")
#Check if zip has already been downloaded in projectData directory?
if(!file.exists("StormData.csv.bz2")){
download.file(url, "StormData.csv.bz2")
# unzip the file first
bunzip2("StormData.csv.bz2", "StormData.csv")
}
StormData <- read.csv("StormData.csv")
```

Across the United States, which types of events are most harmful with respect to population health?

The most harmful EVTYPE with respect to population health will be calculated with respect to the numbers of FATALITIES and INJURIES

```r
StormData$total_harm<-StormData$FATALITIES+StormData$INJURIES
head(StormData$total_harm)
```

```
## [1] 15  0  2  2  2  6
```

```r
#get the total harm per each EVTYPE
harmful_EVTYPE <- StormData %>% select(EVTYPE, total_harm) %>% group_by(EVTYPE) %>% summarise(sum=sum(t
```

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

Economic impact can be calculated with respect to the numbers of PROPDMG and CROPDMG. The actual damage indicated by PROPDMGEXP and CROPDMGEXP.

```r
key <- sort(unique(StormData$PROPDMGEXP))
key
```

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

```r
# as shown in the key there are 8 PROPDMGEXP values that can be mapped as follows
# blank,-,?,+ -> *0
# B,b  = *1^9
# H,h = *1^2
# K,K = *1^3
# M,m  = *1^6
#Levels:  - ? + 0 1 2 3 4 5 6 7 8 B h H K m M with be mapped respectively
key_value <- c(0,0,0,0,10,10,10,10,10,10,10,10,10,10^9,10^2,10^2,10^3,10^6,10^6)


StormData$property_factor <- key_value[match(StormData$PROPDMGEXP, key)]
StormData$crop_factor     <- key_value[match(StormData$CROPDMGEXP,key)]

StormData$total_damage = (StormData$PROPDMG * StormData$property_factor) +
        (StormData$CROPDMG * StormData$crop_factor)
head(StormData$total_damage)
```

```
## [1] 25000  2500 25000  2500  2500  2500
```

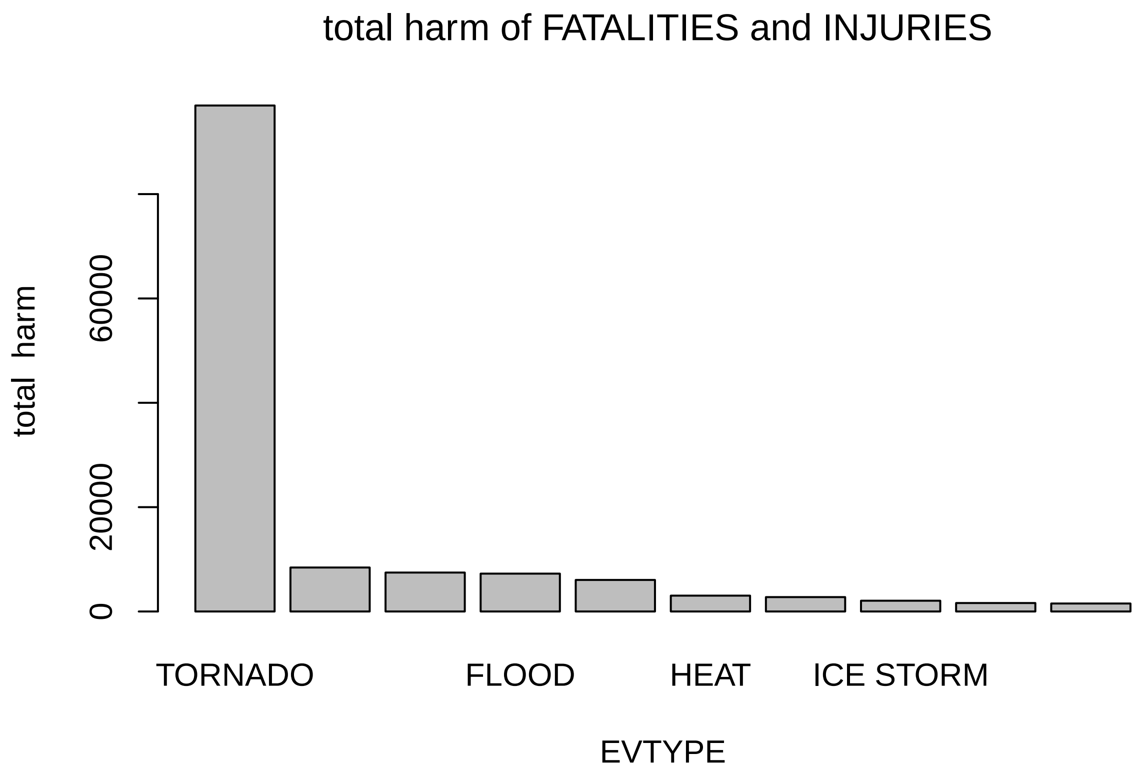
```r
#get the total damage per each EVTYPE
greatest_economic_consequences<- StormData %>% select(EVTYPE, total_damage) %>% group_by(EVTYPE) %>% sum
```

Results

```r
#the most harmfull EVTYPE "top ten"
head(harmful_EVTYPE)
```

```
## # A tibble: 6 x 2
##   EVTYPE           sum
##   <fct>          <dbl>
## 1 TORNADO        96979
## 2 EXCESSIVE HEAT  8428
## 3 TSTM WIND       7461
## 4 FLOOD           7259
## 5 LIGHTNING       6046
## 6 HEAT            3037
```

```r
barplot(height=harmful_EVTYPE$sum[1:10], names.arg=harmful_EVTYPE$EVTYPE[1:10], xlab="EVTYPE", ylab=exp
```

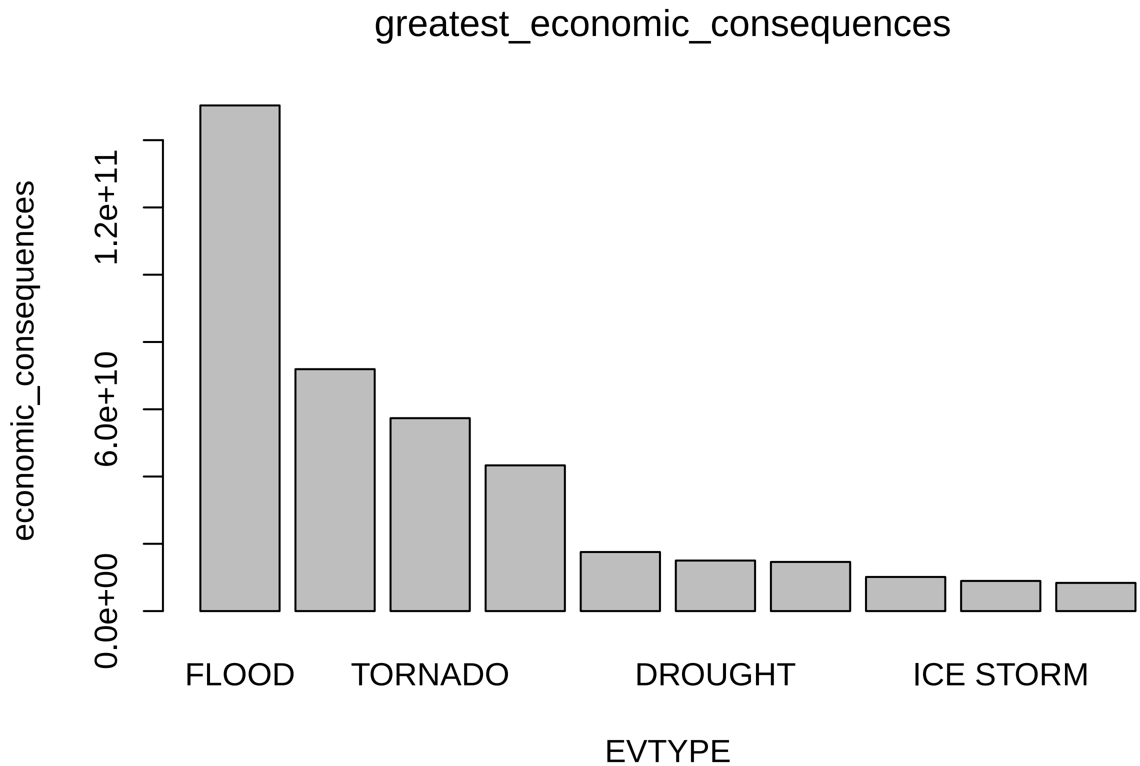


Tornadoes present the most harm with respect to population health

```
#the greatest economic consequences "top ten"
head(greatest_economic_consequences)
```

```
## # A tibble: 6 x 2
##   EVTYPE                      sum
##   <fct>                     <dbl>
## 1 FLOOD              150319678250
## 2 HURRICANE/TYPHOON   71913712800
## 3 TORNADO             57352117547
## 4 STORM SURGE         43323541000
## 5 FLASH FLOOD         17562132111
## 6 DROUGHT             15018672000
```

```
barplot(height=greatest_economic_consequences$sum[1:10], names.arg=greatest_economic_consequences$EVTYPE
```



Flood has the greatest economic consequences