Top 15 US Storm Events With Greater Consequences

Snehadrita Das

Synopsis

The basic goal of this report is to explore the NOAA Storm Database and answer some basic questions about severe weather events with certain consequences regarding the Population Health and Economy. We have taken into consideration the data from 1950 to 2011 (omitting the unrecorded data) and variables that are in connection with Health and Economy. The analysis investigation focuses on *fatalities*,*injuries*,*property and crop damage* which are key concerns for Government.
In our analysis we see that storm event type *TORNADO* causes greatest consequences in Population Health and event types like *FLOOD*, *HURRICANE/TYPHOON*, *TORNADO* and *STORM SURGE* causes a lot of property damage.

Data Processing

Data Loading and EDA

- We load the storm data for our analysis and do some exploratory data analysis.

```
url<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
destfile<-"C:/Users/Hp/Desktop/stormdata.csv.bz2"
download.file(url,destfile)  ## Downloading the data

stormdata<-read.csv("stormdata.csv.bz2")  ## reading the data

head(stormdata)  ## EDA
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL TORNADO
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5         0                                               0         NA
## 6         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0                      14.0   100 3   0          0       15    25.0
## 2         0                       2.0   150 2   0          0        0     2.5
## 3         0                       0.1   123 2   0          0        2    25.0
## 4         0                       0.0   100 2   0          0        2     2.5
## 5         0                       0.0   150 2   0          0        2     2.5
## 6         0                       1.5   177 2   0          0        6     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
## 6          K       0                                         3450      8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
## 6          0          0              6
```

```
str(stormdata)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951
0:00:00" ...
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
dim(stormdata)
```

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

Justification and data transformation

- We see that this dataset consists of a large number of observations and columns as well and this might take up some time and complications during furthur analysis procedure.
- So we filter the data set, remove the observations with unrecorded data or missing values and only keep those variables which are related to Health and Economy.

Transforming/Filtering the dataset

```
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
```

```
storm.data<- stormdata %>%
select(c("BGN_DATE","STATE","EVTYPE","FATALITIES","INJURIES", "PROPDMG","PROPDMGEXP","CROPDM
G", "CROPDMGEXP")) %>%filter(INJURIES>0 | FATALITIES>0 | PROPDMG>0 | CROPDMG>0)

dim(storm.data)  ## dimension of the new subsetted dataset
```

```
## [1] 254633      9
```

```
sum(is.na(storm.data))  ## looking for any NA values
```

```
## [1] 0
```

Dealing with Exponenets

These variables are associated with PROPDMGEXP and CROPDMGEXP which are used as exponents to interpret the numeric values for the damage. The informations regarding the interpretations are in the documentation of the database.

```
unique(storm.data$PROPDMGEXP)  ## looking for unique strings in property damage exponent
```

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

```
unique(storm.data$CROPDMGEXP)  ## looking for unique strings in crop damage exponent
```

```
## [1] ""  "M" "K" "m" "B" "?" "0" "k"
```

```
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP==""]<-10^0  ## assigning numeric values
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="+"]<-10^0
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="-"]<-10^0
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="0"]<-10^0
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="?"]<-10^0
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="1"]<-10^1
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="2"]<-10^2
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="3"]<-10^3
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="4"]<-10^4
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="5"]<-10^5
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="6"]<-10^6
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="7"]<-10^7
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="8"]<-10^8
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="h"]<-10^2
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="H"]<-10^2
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="K"]<-10^3
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="m"]<-10^6
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="M"]<-10^6
storm.data$PROPDMGEXP[storm.data$PROPDMGEXP=="B"]<-10^9

storm.data$CROPDMGEXP[storm.data$CROPDMGEXP==""]<-10^0
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="?"]<-10^0
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="0"]<-10^0
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="2"]<-10^2
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="k"]<-10^3
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="K"]<-10^3
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="m"]<-10^6
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="M"]<-10^6
storm.data$CROPDMGEXP[storm.data$CROPDMGEXP=="B"]<-10^9

storm.data$PROPDMGEXP<-as.numeric(storm.data$PROPDMGEXP)
storm.data$CROPDMGEXP<-as.numeric(storm.data$CROPDMGEXP)
```

Results

Storm Event consequences in Population Health

- We shall calculate the event type wise total number of injuries and fatalities and look for the top 15 events with greater consequences.

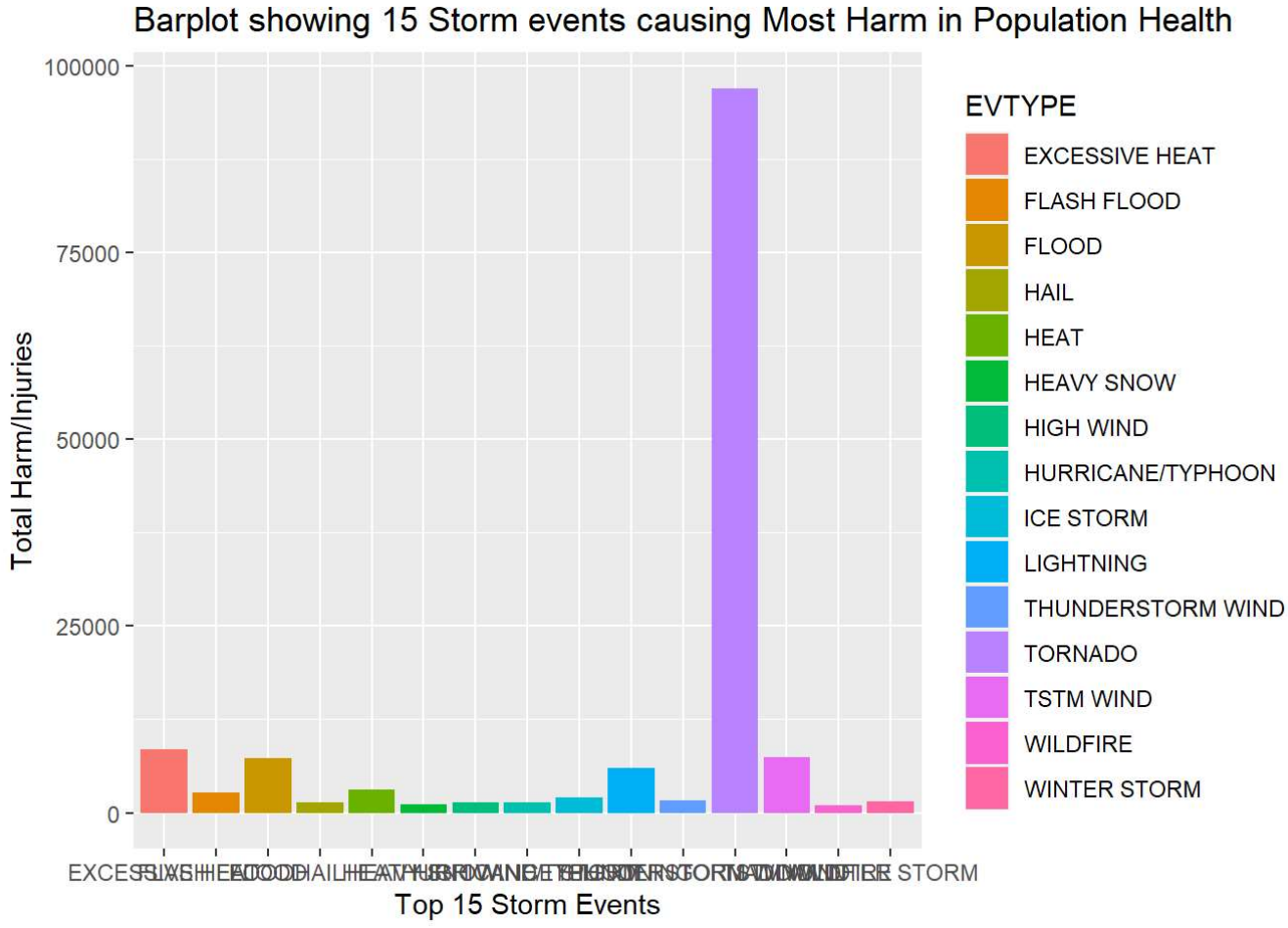
```
## Grouped total injuries and fatalities
harm.in.health<- storm.data %>% group_by(EVTYPE) %>%
     summarise(total.fatality=sum(FATALITIES),total.injury=sum(INJURIES),total.harm=sum(FATAL
ITIES)+sum(INJURIES))

## arranging the dataset from lager to smaller value of total harm/injuries
harm.in.health<-arrange(harm.in.health,desc(total.harm))

## Considering top 15 events
harm.in.health<-harm.in.health[1:15,]

library(ggplot2)

g<-ggplot(harm.in.health,aes(x=EVTYPE,y=total.harm))
g<- g + geom_bar(stat="identity",aes(fill=EVTYPE))
g<- g + ggtitle("Barplot showing 15 Storm events causing Most Harm in Population Health")
g<-g + ylab("Total Harm/Injuries")
g<-g + xlab("Top 15 Storm Events")
g
```



Storm Event consequences in Economy

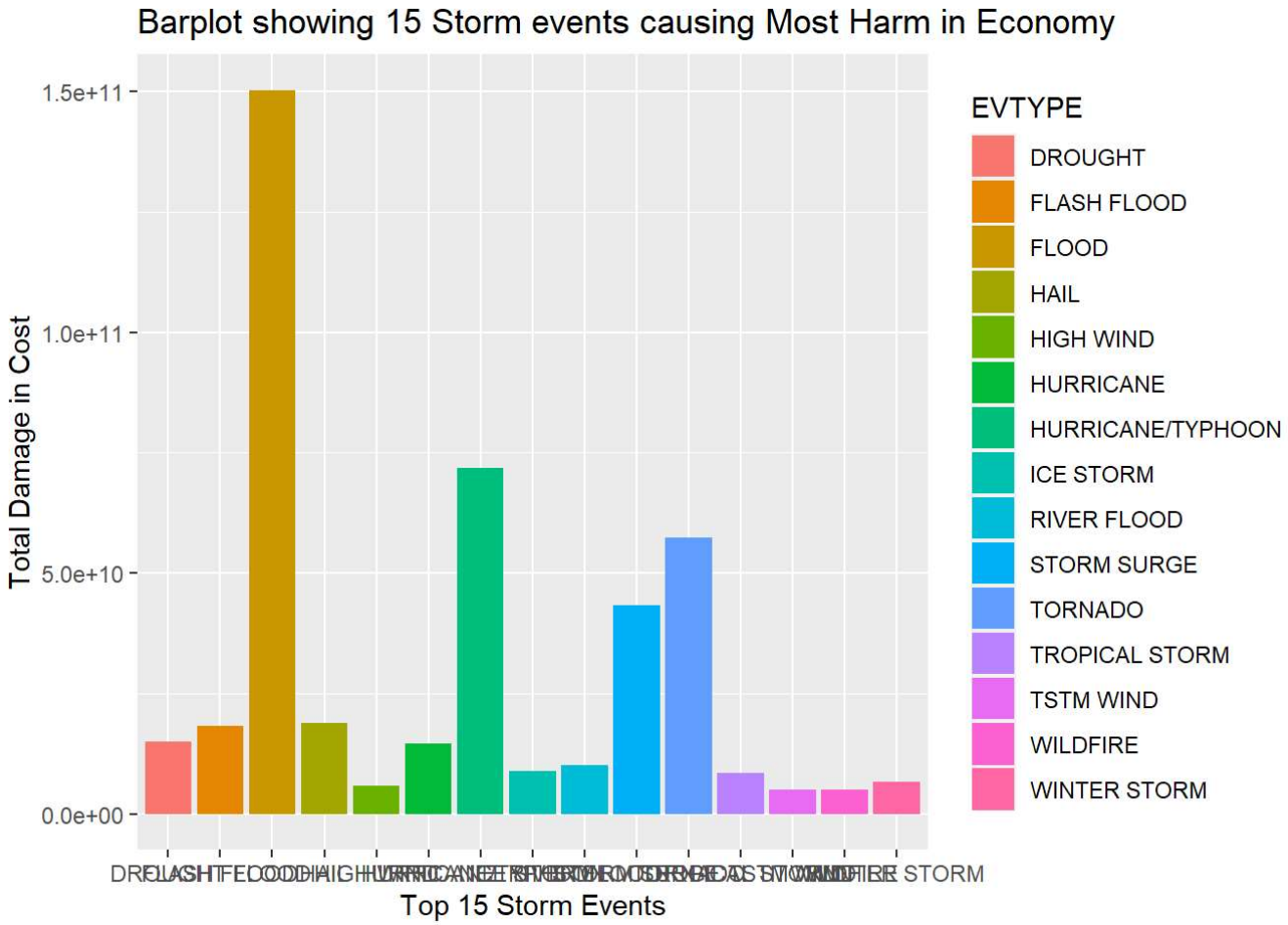
- We shall calculate the event type wise total number of crop and property damage and look for the top 15 events with greater consequences.

```
## grouped total damage
harm.in.econ<- storm.data %>%
     mutate(crop.damage=CROPDMG*CROPDMGEXP,prop.damage=PROPDMG*PROPDMGEXP) %>% group_by(EVTYP
E) %>%
     summarise(total.crop.damage=sum(crop.damage),
               total.prop.damage=sum(prop.damage),total.damage=total.crop.damage+total.prop.d
amage)

## arranging the dataset from lager to smaller value of total damage
harm.in.econ<-arrange(harm.in.econ,desc(total.damage))

## Considering top 15 events
harm.in.econ<-harm.in.econ[1:15,]


g<-ggplot(harm.in.econ,aes(x=EVTYPE,y=total.damage))
g<- g + geom_bar(stat="identity",aes(fill=EVTYPE),position = "dodge")
g<- g + ggtitle("Barplot showing 15 Storm events causing Most Harm in Economy")
g<-g+ ylab("Total Damage in Cost") + xlab("Top 15 Storm Events")
g
```



Conclusion

In our analysis we see that storm event type *TORNADO* causes greatest consequences in Population Health and event types like *FLOOD*, *HURRICANE/TYPHOON*, *TORNADO* and *STORM SURGE* causes a lot of property damage.