Exploring the NOAA Storm Database : Health and Economic impacts of Severe Weather in the US.

SUCHITRA PUDIPEDDI

Date: 31/10/2020

SYNOPSIS

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

In this report, the US National Oceanic and Atmospheric Administration (NOAA) storm database is analyzed to find out what kinds of natural phenomena have been the most damaging events in the US between 1950 and end of November 2011, in terms of damages related to people, and public and private property.

That database tracks the characteristics of major storms and weather events in the US, including when and where they occur, as well as estimates of deaths, injuries, and property damage.

The steps taken in order to generate the results are as follows:

- Loading the data using download.file
- Processing the data - seperating columns that are in interest of the analysis namely - Type of the event, fatalities, injuries, crop damage information and property damage information
- Factoring Event variable
- Finding Aggregate of the total fatalities and injuries for each type of event repestively
- Finding Aggregate of the total property damage plus crop damage in dollars for each type of event
- Arranging all the aggregates in decending order and plotting a bar graph to demonstrate the extent of damage done to health and ecomony respectively.

DATA PROCESSING:

LOADING AND READING THE DATA USING download.file and read.csv

```
#loading the data directly from the link
if(!file.exists("./data")) {dir.create("./data")}
fileURL<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(fileURL,destfile="D:\\EDUCATION\\DATA SCIENCE JH\\Course5.csv.bz2",method="curl")
#reading the data
restData<-read.csv("D:\\EDUCATION\\DATA SCIENCE JH\\Course5.csv.bz2")
```

The data is now loaded. The next step is to take a rough look at the summary of the data using str

```
#converting the type of event to a Factor variable
restData<- transform(restData,EVTYPE=as.factor(EVTYPE))
str(restData)
```

```
## '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    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 834 ...
##  $ 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 ...
```

We can now see that there are 37 variables in the raw data. For the present analysis we only need a subset of the data, essentially containing the columns that have details of:

- PROPDMG: assessment of property damage.
- PROPDMGEXP: exponent value for PROPDMG.
- CROPDMG: assessment of damage to crops.
- CROPDMGEXP: exponent value for CROPDMG.
- FATALITIES: number of deaths.
- INJURIES: number of injured people.
- EVTYPE: type of weather hazard

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.2
```

```
##
## 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
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
tidydata <- select(restData,EVTYPE,FATALITIES,INJURIES,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)
```

The next step is to convert the property damage and crop damage to amount in dollars I have chosen to store the data related to ecomonical damage in a seperate data frame called ecnomicdmg

On the following web page "How To Handle Exponent Value of PROPDMGEXP and CROPDMGEXP" there is an explanation on how to understand and use PROPDMGEXP and CROPDMGEXP to calculate the property and crop damage assessment for each row of dfm.

There it was said that these are the possible values of CROPDMGEXP and PROPDMGEXP: H, h, K, k, M, m, B, b, +, -, ?, 0, 1, 2, 3, 4, 5, 6, 7, 8, and blank-character. And these are the equivalences:

- H,h = hundreds = 100
- K,k = kilos = thousands = 1,000
- M,m = millions = 1,000,000
- B,b = billions = 1,000,000,000
- (+) = 1
- (-) = 0
- (?) = 0
- black/empty character = 0
- numeric 0..8 = 10

```
ecnomicdmg<- select(tidydata,EVTYPE,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)
ecnomicdmg$PROPDMGDOL = 0
ecnomicdmg[ecnomicdmg$PROPDMGEXP == "H", ]$PROPDMGDOL = ecnomicdmg[ecnomicdmg$PROPDMGEXP == "H", ]$PROPDMG * 10^2
ecnomicdmg[ecnomicdmg$PROPDMGEXP == "K", ]$PROPDMGDOL = ecnomicdmg[ecnomicdmg$PROPDMGEXP == "K", ]$PROPDMG * 10^3
ecnomicdmg[ecnomicdmg$PROPDMGEXP == "M", ]$PROPDMGDOL = ecnomicdmg[ecnomicdmg$PROPDMGEXP == "M", ]$PROPDMG * 10^6
ecnomicdmg[ecnomicdmg$PROPDMGEXP == "B", ]$PROPDMGDOL = ecnomicdmg[ecnomicdmg$PROPDMGEXP == "B", ]$PROPDMG * 10^9

ecnomicdmg$CROPDMGDOL = 0
ecnomicdmg[ecnomicdmg$CROPDMGEXP == "H", ]$CROPDMGDOL = ecnomicdmg[ecnomicdmg$CROPDMGEXP == "H", ]$CROPDMG * 10^2
ecnomicdmg[ecnomicdmg$CROPDMGEXP == "K", ]$CROPDMGDOL = ecnomicdmg[ecnomicdmg$CROPDMGEXP == "K", ]$CROPDMG * 10^3
ecnomicdmg[ecnomicdmg$CROPDMGEXP == "M", ]$CROPDMGDOL = ecnomicdmg[ecnomicdmg$CROPDMGEXP == "M", ]$CROPDMG * 10^6
ecnomicdmg[ecnomicdmg$CROPDMGEXP == "B", ]$CROPDMGDOL = ecnomicdmg[ecnomicdmg$CROPDMGEXP == "B", ]$CROPDMG * 10^9
```

Let's now take a look at the data

```
head(ecnomicdmg)
```

```
##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP PROPDMGDOL CROPDMGDOL
## 1 TORNADO    25.0          K       0                 25000          0
## 2 TORNADO     2.5          K       0                  2500          0
## 3 TORNADO    25.0          K       0                 25000          0
## 4 TORNADO     2.5          K       0                  2500          0
## 5 TORNADO     2.5          K       0                  2500          0
## 6 TORNADO     2.5          K       0                  2500          0
```

```
head(tidydata)
```

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

Our data is now filtered and processed The next step done is to find out solutions to the concerns of this analysis

RESULTS

TYPE OF WEATHER THAT IS MOST HARMFUL TO HUMAN HEALTH IN THE US

The steps taken to answer this question: - Finding total number of fatalities and injuries respectively for every type of event that has occured - Plotting the barplots for fatalities and injuries vs type of event respectivelu for the Top 10 values

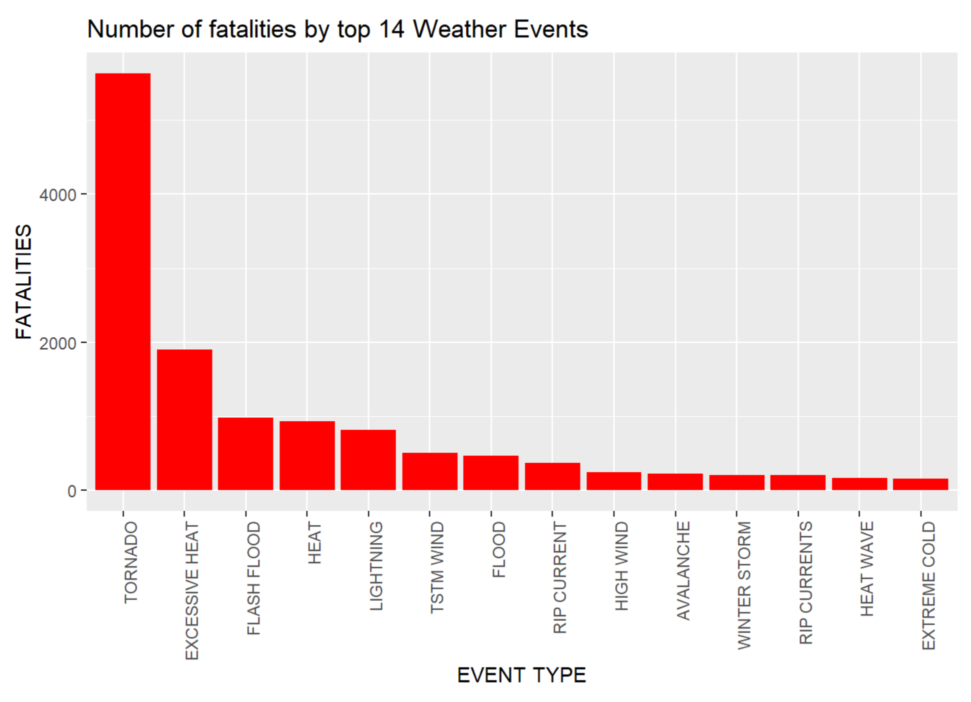
```
sumFatalities <- aggregate(FATALITIES~EVTYPE,tidydata,sum)
sumFatalities<- sumFatalities[sumFatalities[,"FATALITIES"]!=0,]
sumFatalities<- arrange(sumFatalities,desc(FATALITIES))
sumFatalities<- sumFatalities[1:14,]
sumFatalities$EVTYPE<- factor(sumFatalities$EVTYPE, levels= sumFatalities$EVTYPE)

sumInj <- aggregate(INJURIES~EVTYPE,tidydata,sum)
sumInj<- sumInj[sumInj[,"INJURIES"]!=0,]
sumInj<- arrange(sumInj,desc(INJURIES))
sumInj<- arrange(sumInj,desc(INJURIES))
sumInj<- sumInj[1:14,]
sumInj$EVTYPE<- factor(sumInj$EVTYPE, levels= sumInj$EVTYPE)

ggp<-ggplot(sumFatalities,aes(x=EVTYPE,y= FATALITIES))+
geom_bar(stat = "identity", fill = "red", las = 2) +
theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
xlab("EVENT TYPE") + ylab("FATALITIES") + ggtitle("Number of fatalities by top 14 Weather Events")
```

```
## Warning: Ignoring unknown parameters: las
```

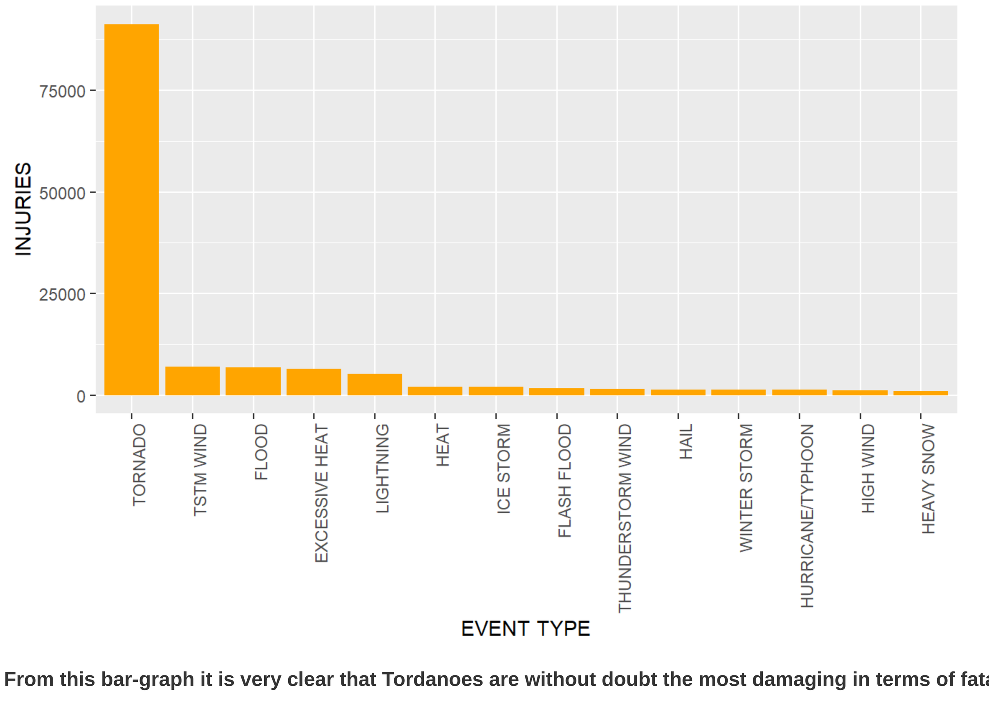
```
ggp
```




```
ggp<-ggplot(sumInj,aes(x=EVTYPE,y= INJURIES))+
  geom_bar(stat = "identity", fill = "orange", las = 2) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  xlab("EVENT TYPE") + ylab("INJURIES") + ggtitle("Number of INJURIES by top 14 Weather Events")
```

```
## Warning: Ignoring unknown parameters: las
```

```
ggp
```



From this bar-graph it is very clear that Tordanoes are without doubt the most damaging in terms of fatality as well as in term of injuries.

TYPE OF WEATHER THAT IS MOST HARMFUL FOR ECOONOMY OF THE COUNTRY U.S

The steps taken to answer this question: - Finding total amount of damage done to prpoperty and crop in terms of money in dollars for every type of event that has occured - Plotting the barplot for damage in dollars vs type of event for the Top 15 values of damage

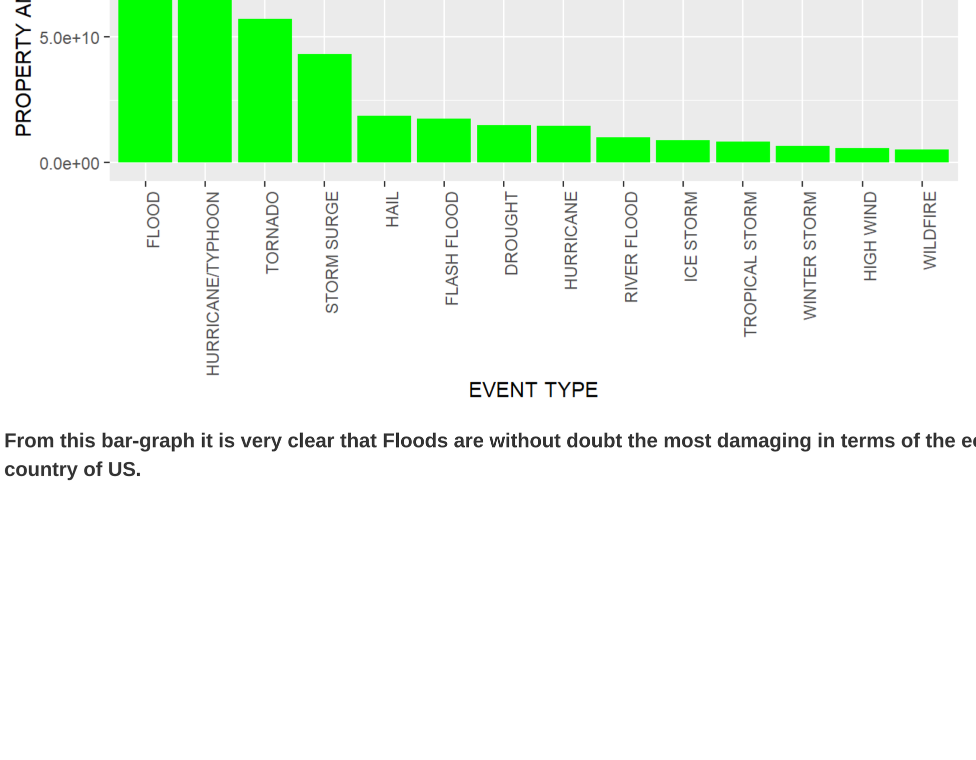
```
#Adding the amount in dollars for propery damage and crop damage
ecnomicdmg$SUMOFPC<-ecnomicdmg$PROPDMGDOL+ecnomicdmg$CROPDMGDOL

#Total damage for each type of weather event
sumpc<- aggregate(SUMOFPC~EVTYPE,ecnomicdmg,sum)
#arranding the damage in descending order
sumpc<- arrange(sumpc,desc(SUMOFPC))
sumpc<- sumpc[1:14,]
sumpc$EVTYPE<- factor(sumpc$EVTYPE, levels= sumpc$EVTYPE)

ggp<-ggplot(sumpc,aes(x=EVTYPE,y= SUMOFPC))+
geom_bar(stat = "identity", fill = "green", las = 2) +
theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
xlab("EVENT TYPE") + ylab("PROPERTY AND CROP DAMAGE") + ggtitle("Property and crop damage for top 14 weather even
ts")
```

```
## Warning: Ignoring unknown parameters: las
```

```
ggp
```



From this bar-graph it is very clear that Floods are without doubt the most damaging in terms of the economic consequences for the country of US.