Data Science Spesialization[JHU]: Reproduciple Research[Project 2]

==

The Impact of Severe Weather Events to Health and Economy in US based on NOAA database.

==

by: Ahmed Talaat, Cousera JHU

1.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.

> This project involves exploring the *U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database**. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

> In this report, we 'll study the effect of weather events on the personal and property damages. The top weather events that causes highest injuries and fatalities 'll be illustrated.

2. Data:

- The data disussed in this report is available at this Storm Data (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2).
- There is also some documentation of the database available at Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)
- National Weather Service FAQ (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf)

> Hint: The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

3. Data Processing:

1. Reading the Whole Dataset "storm table"

```
#you must have been dowloaded and unzipped the data file
storm <- read.csv("./repdata_data_StormData.csv", sep = ",", header = TRUE)
#showing data
str(storm)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383 3980 3980 ...
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 3186 ...
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 1873 4598 10592 4372
10094 1973 23873 24418 4598 ...
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ 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   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","- 1 N Albion",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels ""," 0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","- .5 NNW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels ""," CI","$AC",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
                                  "| __truncated__,..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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   : Factor w/ 436774 levels "","-2 at Deer Park\n",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

2. Cleaning the EVTYPE variable

```
#renaming and cleaning the data
storm$EVTYPE_NEW <- as.character(storm$EVTYPE)
storm[grep("THUNDERSTORM|THUDERSTORM|THUNDEERSTORM|THUNDERESTORM|THUNERSTORM|THUNDERTORM|THUNDERSTROM|TSTM",storm$EVTYPE_NE
W, ignore.case = TRUE),38] <- "THUNDERSTORM"
storm[grep("HURRICANE|TYPHOON",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "HURRICANE/TYPHOON"
storm[grep("TORNADO",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "TORNADO"
storm[grep("FLASH FLOOD|FLASHFLOOD|FLOOD FLASH",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "FLASH FLOOD"
storm[grep("HIGH WIND|HIGH WINDS",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "HIGH WIND"
storm[grep("COASTAL FLOOD",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "COASTAL FLOOD"
storm[grep("COLD",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "COLD"
storm[grep("SNOW",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "SNOW"
storm[grep("ICE|ICY",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "ICE"
storm[grep("NON-TSTM WIND",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "WIND"
storm[grep("TSTM WIND",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "TSTM WIND"
storm[grep("HAIL",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "HAIL"
storm[grep("HEAT",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "EXCESSIVE HEAT"
storm[grep("^FLOOD$|^FLOODS$|^FLOODING$",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "FLOOD"
storm[grep("FROST|FREEZE|FREEZING",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "FROST/FREEZE"
storm[grep("HEAVY RAIN|HEAVY RAINS|HVY RAIN",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "HEAVY RAIN"
storm[grep("MUD SLIDE|MUD SLIDES|MUDSLIDE|MUDSLIDES",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "MUD SLIDE"
storm[grep("WINTER WEATHER|WINTRY",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "WINTER WEATHER"
storm[grep("URBAN AND SMALL|URBAN FLOOD|URBAN FLOODING|URBAN FLOODS|URBAN SMALL|URBAN/SMALL STREAM|URBAN/SML STREAM",storm$E
VTYPE_NEW, ignore.case = TRUE),38] <- "URBAN FLOOD"
storm[grep("TROPICAL STORM",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "TROPICAL STORM"
storm[grep("^WIND$|^WINDS$",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "WIND"
storm[grep("WILDFIRE|WILD FIRE|WILD/FOREST FIRE",storm$EVTYPE_NEW, ignore.case = TRUE),38] <- "WILD FIRE"
```

Q.1 Across the United States, which types of events are most harmful with respect to population health?

To answer this question, we 'll calculate the total injuries and fatalities for each event. We 'll subset the required data and then aggregate the impacts of each event in a new data table.

```
#you should make sure that package "plyr" has installed at first.
library(plyr)
#substetting events for injuries and fatalities != 0
personal <- storm[!(storm$INJURIES  == 0 & storm$FATALITIES == 0), c(38, 23, 24)]
#subsetting the personal data into a new data frame using ddply() function
personal_impact <- ddply(personal, .(EVTYPE_NEW), summarize, impact = sum(INJURIES+FATALITIES), injury = sum(INJURIES), fata
lity = sum(FATALITIES))
personal_impact_top10 <- arrange(personal_impact,desc(impact))[1:10,]
```

Q2 Across the United States, which types of events have the greatest economic consequences?

To answer this question, we 'll calculate the values of total economy damages. We 'll aggregate the values of crops and property damages.

```
#subsetting the whole data storm
event <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG",
           "CROPDMGEXP")
data <- storm[event]

# Assigning values for the property exponent data
# B or b = Billion, M or m = Million, K or k = Thousand, H or h = Hundred). The number from one to ten represent the power o
f ten (10^The number). The symbols "-", "+" and "?" refers to less than, greater than and low certainty. Here, we ignored th
ese three symbols.
economy <- storm[!(storm$PROPDMG==0 & storm$CROPDMG==0),c(38,25,26,27,28)]

m <- cbind(names(table(economy$PROPDMGEXP)),c(0,0,0,0,0,1,2,3,4,5,6,7,8,9,2,2,3,6,6))
m <- rbind(m,c("k",3))

economy$PROPDMGEXP <- as.integer(m[,2][match(economy$PROPDMGEXP,m[,1])])
economy$CROPDMGEXP <- as.integer(m[,2][match(economy$CROPDMGEXP,m[,1])])
#making economy impact from the economy data, a new data frame using ddply() function
economy_impact <- ddply(economy,.(EVTYPE_NEW),summarize,damage=sum(PROPDMG*10^(PROPDMGEXP)+CROPDMG*10^(CROPDMGEXP)),propdmg=
sum(PROPDMG*10^(PROPDMGEXP)),cropdmg=sum(CROPDMG*10^(CROPDMGEXP)))
economy_impact_top10 <- arrange(economy_impact,desc(damage))[1:10,]
```

3. Results:

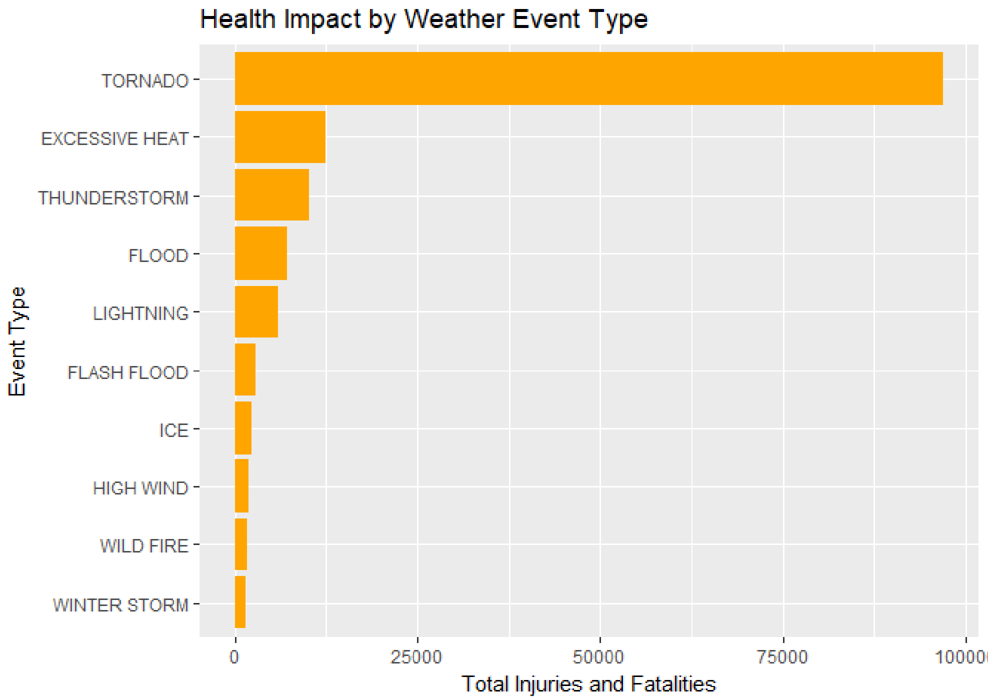
3.1: Showing the Health impacts from diffrent events

```
personal_impact_top10
```

```
##        EVTYPE_NEW impact injury fatality
## 1         TORNADO  97043  91407     5636
## 2  EXCESSIVE HEAT  12362   9224     3138
## 3    THUNDERSTORM  10299   9544      755
## 4           FLOOD   7267   6791      476
## 5       LIGHTNING   6046   5230      816
## 6     FLASH FLOOD   2837   1802     1035
## 7             ICE   2285   2183      102
## 8       HIGH WIND   1820   1523      297
## 9       WILD FIRE   1696   1606       90
## 10   WINTER STORM   1527   1321      206
```

```
library(ggplot2)

ggplot(personal_impact_top10, aes(x = reorder(EVTYPE_NEW,impact), y = impact)) + geom_bar(stat="identity", fill = "orange")
 + labs(title = "Health Impact by Weather Event Type", x = "Event Type", y = "Total Injuries and Fatalities") + coord_flip()
```

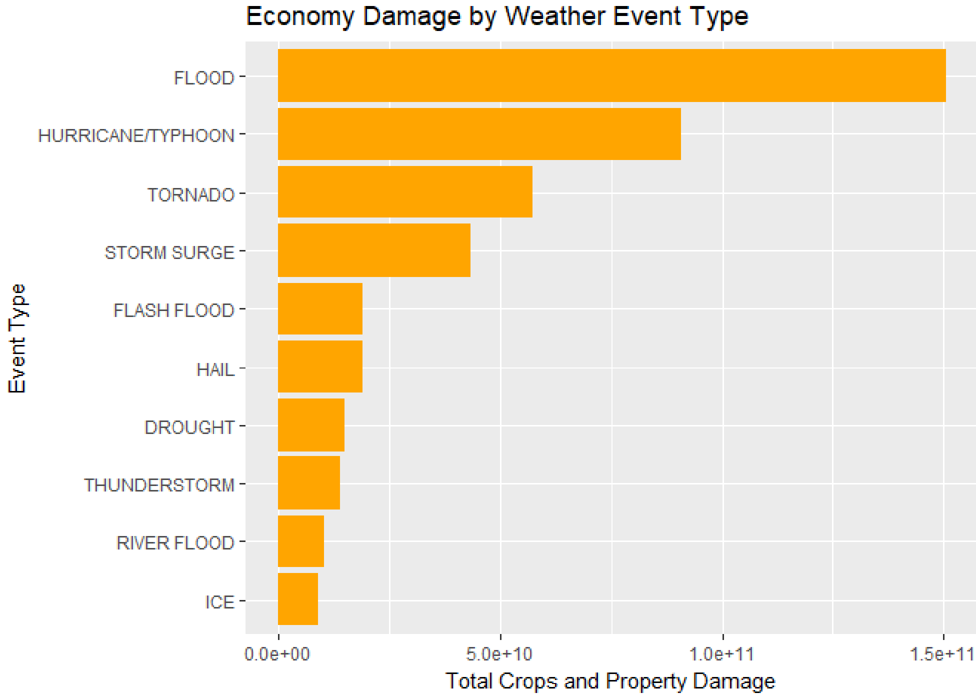


3.2: Showing the Economy impacts from diffrent events

```
economy_impact_top10
```

```
##          EVTYPE_NEW       damage      propdmg     cropdmg
## 1             FLOOD 150443429757 144772555807  5670873950
## 2  HURRICANE/TYPHOON  90872527810  85356410010  5516117800
## 3           TORNADO  57418279447  57003317927   414961520
## 4       STORM SURGE  43323541000  43323536000        5000
## 5       FLASH FLOOD  19120499246  17588302096  1532197150
## 6              HAIL  19024452136  15977564513  3046887623
## 7           DROUGHT  15018672000   1046106000 13972566000
## 8      THUNDERSTORM  14059635688  12785421700  1274213988
## 9       RIVER FLOOD  10148404500   5118945500  5029459000
## 10              ICE   8994976860   3967862560  5027114300
```

```
ggplot(economy_impact_top10, aes(x = reorder(EVTYPE_NEW,damage), y = damage)) + geom_bar(stat = "identity", fill = "orange")
 + labs(title = "Economy Damage by Weather Event Type", x = "Event Type", y = "Total Crops and Property Damage") + coord_fli
p()
```



4. Conclusion:

As we see, Tornado events have the highest harmful impacts on population health. Also, Flood events have the biggest impacts on the economical side. So, we wish that the American Govenments would take the right preventive actions to save both lifes and money.