Reproducible Research: Exploring Storm Data impacts on people and economics

Jeanna

2021-07-12

Synopsis

Storm and other severe weather events can cause both public health and economic problems for communities and municipalities.
In this project, we explore Storm Data from the U.S. National Oceanic and Atmospheric Administration(NOAA). This contains storms and weather events in the United States with time, location of the occurrences, estimations of any fatalities, injuries, and property damage. We will analyze and find out what type of events have big impacts for population health and economic consequences.

Preparation

We created two mapping csv files manually and store them in folder /data under project folder. They are used to clean event type data. For mapping data detail, they are listed in the Appendix.
- **ref_Evtype_Corrections.csv** is for correcting the typos of original event type
- **ref_Evtype_Mapping.csv** is for mapping the official events type

Before data processing, load R packages.

```
## Loading packages
library(dplyr)
library(ggplot2)
library(lubridate)
library(stringr)
library(tidyr)
library(cowplot)
```

Data Processing

In this section, Data will be processed by five steps.

- Loading data
- Converting date character to date type and reducing the data size by aggregating
- Cleaning Event Type
- Converting property damage and crop damage to million
- Calculating data

Loading data

```r
## To denote the starting time
startTime <- Sys.time()

filename <- "data/repdata_data_StormData.csv.bz2"
if (!file.exists(filename)) {
    file_url <-
        "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
    download.file(file_url, filename, method = "curl")
}
## Reading storm data set and processing the data
df.storm <- read.csv(filename)
cat("In the data file, there are",
    nrow(df.storm),
    "records on",
    ncol(df.storm),
    "variables",
    "\n"
)
```

```
## In the data file, there are 902297 records on 37 variables
```

Converting date character to date type and reducing the data size by aggregating

Convert date character to date type

```r
## Convert date character to date type
df.storm$BGN_DATE <- as.Date(df.storm$BGN_DATE, format = "%m/%d/%Y")
df.storm$END_DATE <- as.Date(df.storm$END_DATE, format = "%m/%d/%Y")
## Add Year column
df.storm$BGN_YEAR <- as.character(df.storm$BGN_DATE, format = "%Y")
```

To reduce data, aggregate fatalities, injuries, crop damage and property damage by PROPDMGEXP, CROPDMGEXP, event type and year.

```r
## Aggregate fatalities, injuries, and property damage
## by event type and beginning year
df.byevtypeyear <-
    df.storm %>% group_by(EVTYPE, BGN_YEAR, PROPDMGEXP, CROPDMGEXP) %>%
    summarize(
        sumFatalities = sum(FATALITIES),
        sumInjuries = sum(INJURIES),
        sumPropdmg = sum(PROPDMG),
        sumCropdmg = sum(CROPDMG)
    )

cat("after reduced, the data rows is",
    nrow(df.byevtypeyear),
    "\n"
)
```

```
## after reduced, the data rows is 4516
```

```
checkDataType <- unique(df.byevtypeyear[, "EVTYPE"])
cat ("The official events type are 48.",
    "However, 'EVTYPE' column has", nrow(checkDataType), "unique events",
    "\n"
)
```

```
## The official events type are 48. However, ’EVTYPE’ column has 985 unique events
```

Cleaning Event Type

TO simplify data cleaning process, two configuration files are created manually (mentioned in Section Preparation). Firstly, read file ref_Evtype_Corrections.csv and correct the typos of event type. Then, after loading data from file ref_Evtype_Mapping.csv, map event type to the official events type and remove the data without official events type.

```
## Convert EVTYPE to only the first letter to Uppercase and remove whitespace
dfNew <- df.byevtypeyear
dfNew$EVTYPE <- str_trim(str_to_title(dfNew$EVTYPE))
## convert backslash \ to forward slash /
dfNew$EVTYPE <- gsub("\\\\", "/", dfNew$EVTYPE )

## Loading Typo Mapping data manually created for correcting typos
refTypo <- read.csv("data/ref_Evtype_Corrections.csv", header = TRUE)
for (i in 1:nrow(refTypo)) {
    dfNew$EVTYPE <-
        sub(refTypo$EVTXT[i],
            refTypo$EVSUBSTR[i],
            dfNew$EVTYPE)
}

## Loading Event Type Mapping data manually created,
## which are used for simplifying remove operation and grepl operation
refEvType <- read.csv("data/ref_Evtype_Mapping.csv", header = TRUE)
## Retrieving subset event type data for removing operation and processing data
refEvTypeRemove <- refEvType [refEvType$OPER=="remove", ]
for (i in 1:nrow(refEvTypeRemove)) {
    patx <- refEvTypeRemove$SEARCHEVTYPE [i]
    searchLoc <- grepl(patx, dfNew$EVTYPE)
    dfNew <- dfNew[!(searchLoc), ]
}
## Fetching and sorting Event Type for grepl operation by priority
refEvTypeGrepl <- refEvType [refEvType$OPER=="grepl", ]
refEvTypeGrepl <- arrange(refEvTypeGrepl, PRIO, STORMEVENT)
## updating storm event with official one.
dfNew["StormEvent"] <- NA
for (i in 1:nrow(refEvTypeGrepl)) {
    patx <- refEvTypeGrepl$SEARCHEVTYPE [i]
    searchLoc <- grepl(patx, dfNew$EVTYPE)
    dfNew$StormEvent[is.na(dfNew$StormEvent) &
                       searchLoc == TRUE] <- refEvTypeGrepl$STORMEVENT[i]
}
## Remove rows with NA StormEvent
dfNew <- dfNew[!(is.na(dfNew$StormEvent)), ]

cat(
    "Removed",
    nrow(df.byevtypeyear) - nrow(dfNew),
    "uncount records, and left data has",
    nrow(dfNew),
    "records.",
    "\n"
)
```

```
## Removed 196 uncount records, and left data has 4320 records.
```

```
checkDataType <- unique(dfNew[ , c("StormEvent")])
cat ("Currently, the storm events type are",
     nrow(checkDataType),
     "\n"
)
```

```
## Currently, the storm events type are 48
```

Converting property damage and crop damage to million

For the value of column 'CROPDMGEXP' and 'PROPDMGEXP', we refer to a nice work How To Handle Exponent Value of PROPDMGEXP and CROPDMGEXP which mentioned in discussion forum by mentor Usama.

- 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

```
dfNew$PROPDMGEXP <- toupper(dfNew$PROPDMGEXP)
dfNew$CROPDMGEXP <- toupper(dfNew$CROPDMGEXP)

iNum <- 0:8
dfNew <- mutate(
    dfNew,
    PROPDMG_M = case_when(
        PROPDMGEXP == "B" ~  sumPropdmg * 1000,
        PROPDMGEXP == "M" ~  sumPropdmg,
        PROPDMGEXP == "K" ~  sumPropdmg / 1000,
        PROPDMGEXP == "H" ~  sumPropdmg / 10000,
        PROPDMGEXP %in% iNum ~ sumPropdmg / 100000,
        PROPDMGEXP == "+" ~ sumPropdmg / 1000000,
        TRUE ~ 0
    )
)
dfNew <- mutate(
    dfNew,
    CROPDMG_M = case_when(
        CROPDMGEXP == "B" ~  sumCropdmg * 1000,
        CROPDMGEXP == "M" ~  sumCropdmg,
        CROPDMGEXP == "K" ~  sumCropdmg / 1000,
        CROPDMGEXP == "H" ~  sumCropdmg / 10000,
        CROPDMGEXP %in% iNum ~ sumCropdmg / 100000,
        CROPDMGEXP == "+" ~ sumCropdmg / 1000000,
        TRUE ~ 0
    )
)
```

Calculating data

Summarize the data by official event type.

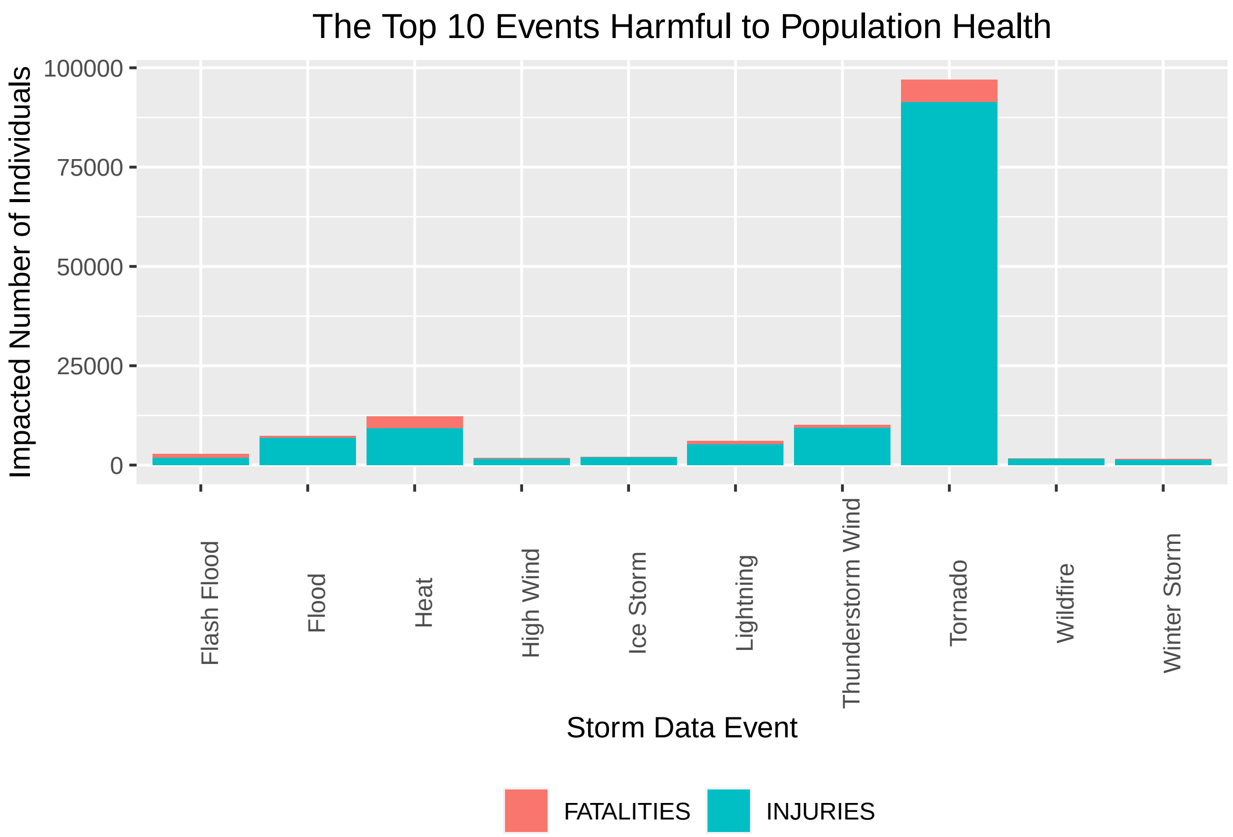
```
byEvtype <- dfNew %>%
    group_by(StormEvent) %>%
    summarise(
        DMG = round(sum(PROPDMG_M) + sum(CROPDMG_M)),
        FATALITIES = sum(sumFatalities),
        INJURIES = sum(sumInjuries)
    )
```

```
byEvtypeYear <- dfNew %>%
    group_by(StormEvent, BGN_YEAR) %>%
    summarise(
        DMG = round(sum(PROPDMG_M) + sum(CROPDMG_M)),
        FATALITIES = sum(sumFatalities),
        INJURIES = sum(sumInjuries)
    )
## Pivot data from wide to long
byEvTypeEcon <- top_n(byEvtype, 10, DMG)
byEvTypeHlth <- top_n(byEvtype, 10, (FATALITIES + INJURIES))
byEvtypeHlthPiv <-
    subset(byEvTypeHlth, select = -DMG) %>%
      pivot_longer(-StormEvent,
                   names_to = "TYPE",
                   values_to = "CONSEQVALUE")
```

Results

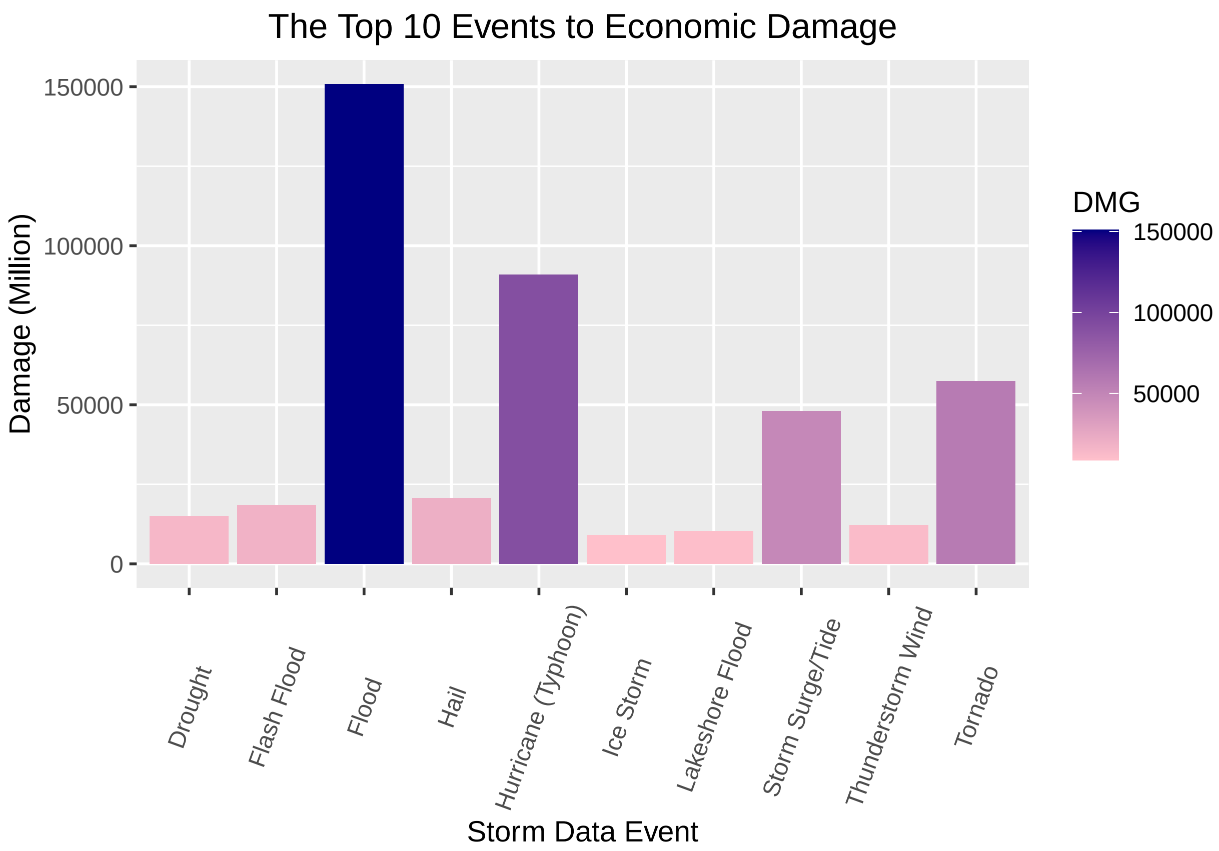
Plot the top 10 injuries and fatalities, and top 10 sum of crop damage property damage to check which type of event have the most impact.

```
g <-
    ggplot(byEvtypeHlthPiv, aes(StormEvent, CONSEQVALUE, fill = TYPE)) +
    geom_bar(stat = "identity") +
    theme(axis.text.x = element_text(angle = 90),
          plot.title = element_text(hjust = 0.5),
          legend.title = element_blank(),
          legend.position = "bottom" ) +
    ggtitle("The Top 10 Events Harmful to Population Health") +
    labs(x = "Storm Data Event", y = "Impacted Number of Individuals")
print(g)
```



From above diagram, we can conclude that Tornado is the most harmful storm event to population health.

```
## Histogram the top 10 harmful events to economic consequences
g <- ggplot(data = byEvTypeEcon, aes(x = StormEvent, y = DMG, fill = DMG)) +
    geom_bar(stat = "identity") +
    theme(axis.text.x = element_text(angle =  70, vjust = 0.5),
          plot.title = element_text(hjust = 0.5)) +
    scale_fill_gradient(low="pink",high="navy") +
    ggtitle("The Top 10 Events to Economic Damage") +
    labs(x = "Storm Data Event", y = "Damage (Million)")
print(g)
```



From above diagram, we can conclude that Flood have the greatest economic consequences.

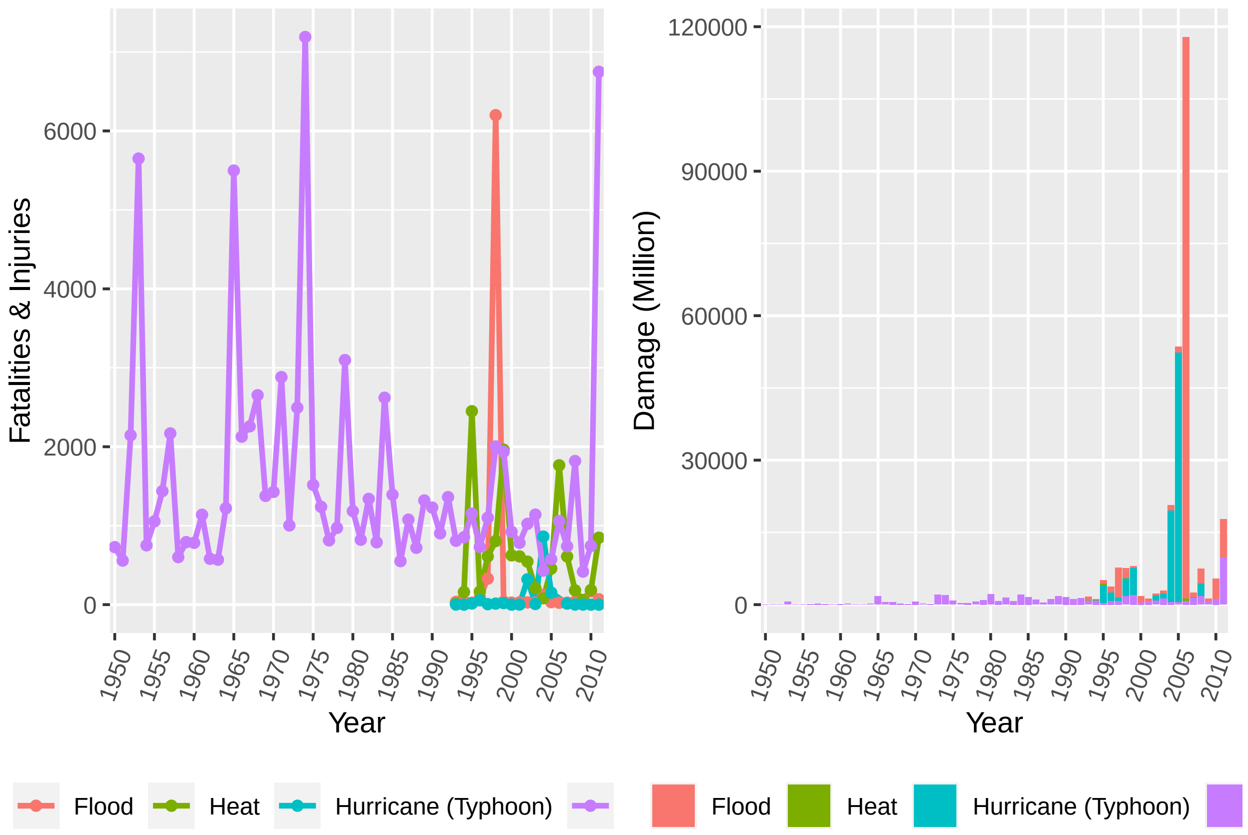
Additionally, we would like to check what is the impacts over years for the first two most influential events, for people, the first two events are Tornado/Heat, and for economics, they are Flood/Hurricane (Typhoon).

```
## Check Storm Event Tornado/Flood/Heat/Hurricane (Typhoon) impacts
byEvtypeYearTop <-
    subset(
        byEvtypeYear,
        StormEvent == "Tornado" |
            StormEvent == "Flood" |
            StormEvent == "Hurricane (Typhoon)" |
            StormEvent == "Heat"
    )

plot.yHlth <-
    ggplot(byEvtypeYearTop,
           aes(x = BGN_YEAR,
               y = FATALITIES + INJURIES,
               group = StormEvent
           )) +
    geom_line(aes(color = StormEvent), size = 1)+
    geom_point(aes(color = StormEvent), size = 1.5)+
    scale_x_discrete(breaks = seq(1950, 2015, by = 5)) +
    theme(
        axis.text.x = element_text(angle = 70, vjust = 0.5),
        legend.title = element_blank(),
        legend.position = "bottom"
    ) +
    labs(x = "Year", y = "Fatalities & Injuries")

plot.yEcon <-
    ggplot(data = byEvtypeYearTop,
           aes(x = BGN_YEAR, y = DMG, fill = StormEvent)) +
    # geom_point(size = 3, alpha = 0.5) +
    geom_bar(stat = "identity") +
    theme(
        axis.text.x = element_text(angle = 70, vjust = 0.5),
        legend.title = element_blank(),
        legend.position = "bottom"
    ) +
    scale_x_discrete(breaks = seq(1950, 2015, by = 5)) +
    labs(x = "Year", y = "Damage (Million)")

gg <-
    ggdraw() +
    draw_plot(plot.yHlth, 0, 0, 0.5, 1) +
    draw_plot(plot.yEcon, 0.5, 0, 0.5, 1)
print(gg)
```



From the plots, we can see Tornado always has impacts on people and economics since 1950. Since 1993, Flood and began to affect on economics and population's health.

```
## To denote the end time
endTime <- Sys.time()
cat("Processing time:", endTime-startTime, "\n")
```

```
## Processing time: 33.60532
```

Appendix

File **ref__Evtype__Corrections.csv** is for correcting the typos of original event type.

Table 1: Mapping the Typos Correction

EVTXT	EVSUBSTR
-	/
/	/
Abnormally	Extreme
Abnormal	Extreme
Avalance	Avalanche
Beach	Coastal
Bitter	Extreme
Coastalflood	Coastal Flood
Coastalstorm	Coastal Flood
Cool	Cold
Cstl	Coastal
Dust Devel	Dust Devil
Erosin	Erosion
Excessive	Extreme
Extremely	Extreme
Extended	Extreme
Fldg	Flooding
Fld	Flood
Floooding	Flooding
Hvy	Heavy
Hyperthermia	Hypothermia
Icestorm	Ice Storm
Lighting	Lightning
Ligntning	Lightning
Micoburst	Microburst
Mudslide	Mud Slide
Precipatation	Precipitation
Record	Extreme
Severe	Extreme
Sml	Small
Strm	Stream
Thundeerstorm	Thunderstorm
Thunderstrom	Thunderstorm
Thundertorm	Thunderstorm
Thundertsorm	Thunderstorm
Thundestorm	Thunderstorm
Thunerstorm	Thunderstorm
Thuderstorm	Thunderstorm
Thunderestorm	Thunderstorm

EVTXT	EVSUBSTR
Torndao	Tornado
Torrential	Heavy
Tstmw	Thunderstorm Wind
Tstm	Thunderstorm
Unseasonable	Extreme
Unseasonal	Extreme
Unusual	Extreme
Unusually	Extreme
Very	Extreme
Water Spout	Waterspout
Wayterspout	Waterspout
Wet	Rain
Windchill	Wind Chill
Wintery	Winter
Wintry	Winter
Wnd	Wind

File **ref__Evtype__Mapping.csv** is for mapping official events type.

Table 2: Mapping the official events type

OPER	STORMEVENT	PRIO	SEARCHEVTYPE
remove		1	Apache County
remove		1	No Severe Weather|No Extreme Weather
remove		1	Non
remove		1	Summary
remove		1	Month
remove		1	Year
remove		1	Northern Lights
remove		1	Other
grepl	Astronomical Low Tide	1	Astronomical Low Tide
grepl	Avalanche	1	Avalanche
grepl	Blizzard	1	Blizzard
grepl	Coastal Flood	1	Coastal Flood|Coastal Erosion
grepl	Cold/Wind Chill	9	Cold/Wind Chill|Cold
grepl	Debris Flow	1	Debris|Slide|Landslump|Landslide
grepl	Dense Fog	9	Fog
grepl	Dense Smoke	1	Dense Smoke|Smoke
grepl	Drought	1	Dry|Drought
grepl	Dust Devil	1	Dust Devil
grepl	Dust Storm	1	Dust Storm|Saharan Dust|Duststorm
grepl	Excessive Heat	5	Excessive Heat|Extreme Warm|Extreme High|Hot
grepl	Extreme Cold/Wind Chill	1	Extreme Wind|Extreme Cold|Extreme Low
grepl	Flash Flood	2	Flash|Drowning
grepl	Flood	8	Flood

OPER	STORMEVENT	PRIO	SEARCHEVTYPE
grepl	Frost/Freeze	5	Frost|Freeze|Ice
grepl	Funnel Cloud	1	Funnel|Wall Cloud
grepl	Freezing Fog	1	Freezing Fog|Ice Fog
grepl	Hail	1	Hail
grepl	Heat	9	Heat
grepl	Heavy Rain	9	Heavy Rain|Heavy Shower|Precipitation|Rain
grepl	Heavy Snow	1	Heavy Snow|Accumulated Snow|Extreme Snow|Heavy Wet Snow
grepl	High Surf	1	Surf
grepl	High Wind	9	High Wind|Gustnado|Wind
grepl	Hurricane (Typhoon)	1	Hurricane|Typhoon
grepl	Ice Storm	1	Ice Storm
grepl	Lake-Effect Snow	1	Lake-Effect Snow|Lake Snow
grepl	Lakeshore Flood	1	Lakeshore Flood|Lake Flood|River Flood
grepl	Lightning	1	Lightning
grepl	Marine Hail	1	Marine Hail|Marine Accident|Marine Mishap
grepl	Marine High Wind	1	Marine High Wind
grepl	Marine Strong Wind	1	Marine Strong Wind
grepl	Marine Thunderstorm Wind	1	Marine Thunderstorm Wind
grepl	Rip Current	1	Rip Current
grepl	Seiche	1	Seiche
grepl	Sleet	1	Sleet
grepl	Storm Surge/Tide	1	Surge|Tide|Tidal|Coastal Storm
grepl	Strong Wind	1	Strong Wind|Gusty Wind|Downburst
grepl	Thunderstorm Wind	1	Thunderstorm|Microburst
grepl	Tornado	1	Tornado
grepl	Tropical Depression	1	Tropical Depression
grepl	Tropical Storm	1	Tropical Storm
grepl	Tsunami	1	Tsunami
grepl	Volcanic Ash	1	Volcanic
grepl	Waterspout	1	Waterspout
grepl	Wildfire	9	Wildfire|Fire
grepl	Winter Storm	5	Winter Storm
grepl	Winter Weather	8	Snow|Winter|Freezing|Glaze|Hypothermia