Effect of Weather Events On Population Health And National Economy

Synopsis :

Objective of this report is to evaluate the impact of weather events on population health and national economy. Understand which events have the maximum effect on mentioned parameters and evaluate potential action to mitigate the effects. Data on major storms and weather events compiled by U.S. National Oceanic and Atomospheric Administration (NOAA) for the period 1950 to 2011 is used for this analysis.

Loading and Processing the Raw Data :

The data file is a comma-separated-value (csv) file compressed via the bzip2 algorithm to reduce its size. Still its a 47Mb file, which takes some time to download as well as load to R Program. The file has been downloaded from the Coursera project website link, Storm Data into the working directory to facilitate loading.

Reading in the data :

```
if (file.exists("repdata-data-StormData.Csv.bz2")) {
    wthrevents=read.csv("repdata-data-StormData.Csv.bz2")
}else {
    download.file(url="https://d396qusza40orc.cloudfront.net/
                  repdata%2Fdata%2FStormData.csv.bz2",
                  destfile="repdata-data-StormData.Csv.bz2")
    wthrevents=read.csv("repdata-data-StormData.Csv.bz2")
}
```

After reading in the data, run a check on the data to get a flavour of it.

```
dim(wthrevents) # rows & columns of the data
```

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

```
# flavour the data
head(wthrevents[,1:12]);head(wthrevents[,13:24]);head(wthrevents[,25:37])
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE
## 1 TORNADO         0
## 2 TORNADO         0
## 3 TORNADO         0
## 4 TORNADO         0
## 5 TORNADO         0
## 6 TORNADO         0
```

```
##   END_TIME COUNTY_END COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH
## 1                   0         NA         0                      14.0   100
## 2                   0         NA         0                       2.0   150
## 3                   0         NA         0                       0.1   123
## 4                   0         NA         0                       0.0   100
## 5                   0         NA         0                       0.0   150
## 6                   0         NA         0                       1.5   177
##   F MAG FATALITIES INJURIES
## 1 3   0          0       15
## 2 2   0          0        0
## 3 2   0          0        2
## 4 2   0          0        2
## 5 2   0          0        2
## 6 2   0          0        6
```

```
##   PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE
## 1    25.0          K       0                                         3040
## 2     2.5          K       0                                         3042
## 3    25.0          K       0                                         3340
## 4     2.5          K       0                                         3458
## 5     2.5          K       0                                         3412
## 6     2.5          K       0                                         3450
##   LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1      8812       3051       8806              1
## 2      8755          0          0              2
## 3      8742          0          0              3
## 4      8626          0          0              4
## 5      8642          0          0              5
## 6      8748          0          0              6
```

Change the names of the columns to lower case to facilitate handling.

```r
names(wthrevents)=tolower(names(wthrevents))
```

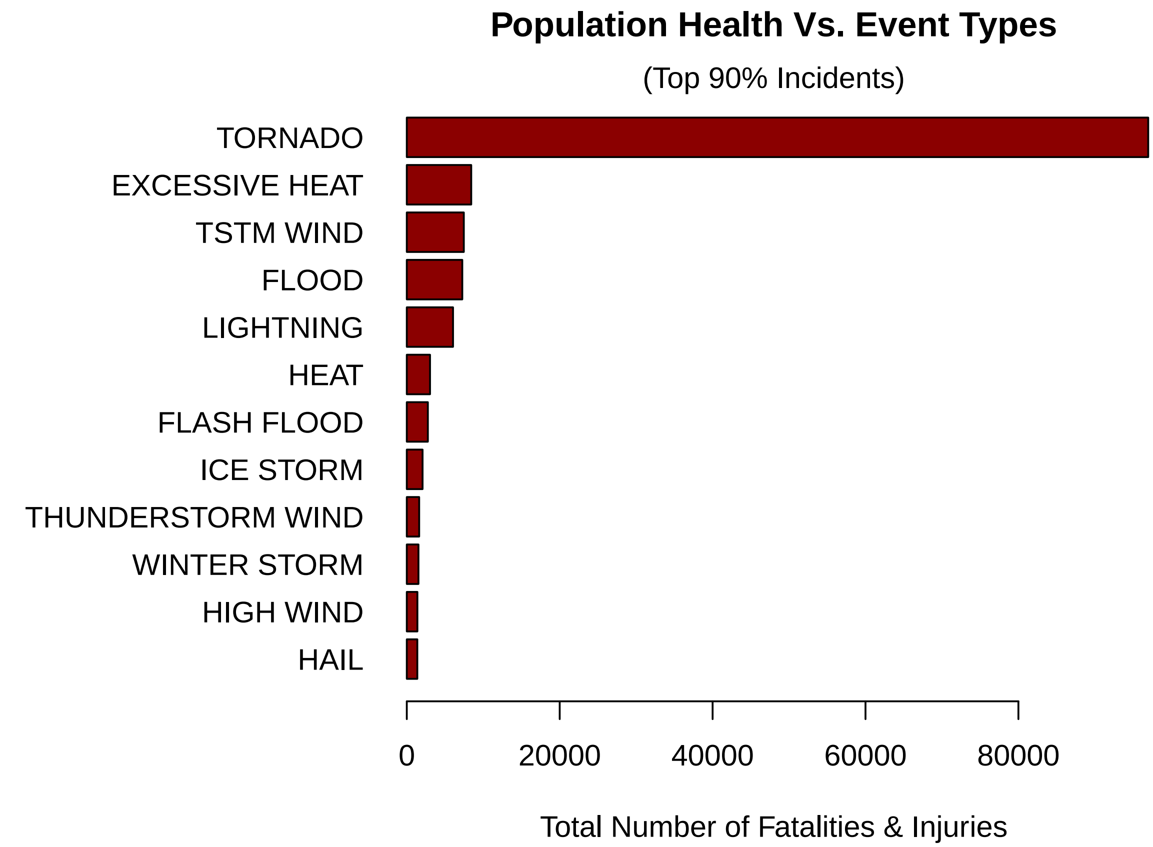
Results :

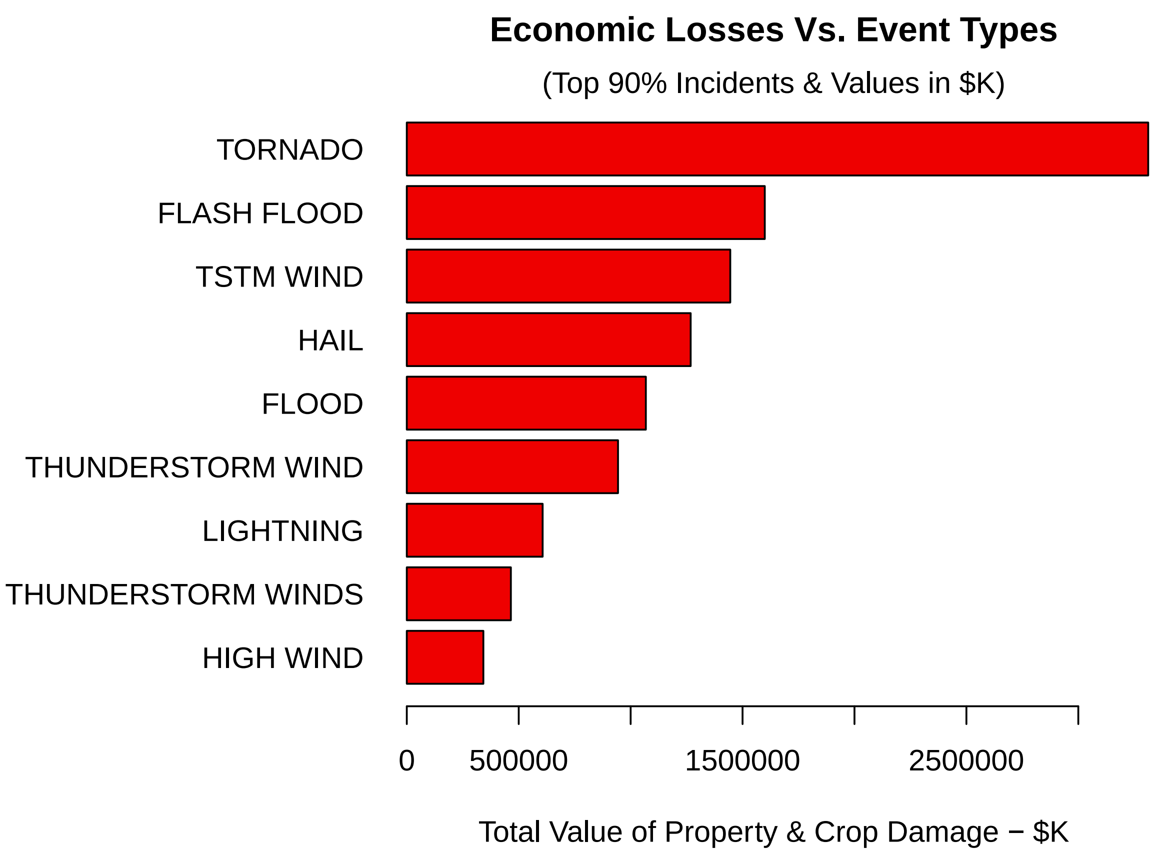
*****Note column names have been converted to lower case***

Based on the data evaluation using the head command, it is obvious that the effect on population health due to weather events can be evaluated from informaton in the columns **"fatalities"** and **"injuries"**. So, studying the sum of these columns relative to the events as contained in column **"evtype"** may yield some hypothesis. A plot of the event types which account for the top **90%** of health related incidents would give deeper insight.

Economic consequences can be assessed from information in the columns, **"propdmg"- Property Damage** and **"cropdmg" - Crop Damage**. These values are in $K. Hence, a plot of the sum of these values relative to the events that account for **90%** of total impact would also yield better insights.

*****Refer to Appendix-I for code for the plots and subsequent values***.


Population Health Vs. Event Types
(Top 90% Incidents)
TORNADO
EXCESSIVE HEAT
TSTM WIND
FLOOD
LIGHTNING
HEAT
FLASH FLOOD
ICE STORM
THUNDERSTORM WIND
WINTER STORM
HIGH WIND
HAIL
0
20000
40000
60000
80000
Total Number of Fatalities & Injuries


Economic Losses Vs. Event Types
(Top 90% Incidents & Values in $K)
TORNADO
FLASH FLOOD
TSTM WIND
HAIL
FLOOD
THUNDERSTORM WIND
LIGHTNING
THUNDERSTORM WINDS
HIGH WIND
0
500000
1500000
2500000
Total Value of Property & Crop Damage − $K

Conclusions :

The plots indicate that there are **12** events which have major impact on health and **9** events which have major impact on the economy.

Further evaluation of the events show that **8** events are common for both categories of damage. These events account for **86%** of total damage for both health and economic damages and are as follows :

```
  Common Events
1 HAIL
2 HIGH WIND
3 THUNDERSTORM WIND
4 FLASH FLOOD
5 LIGHTNING
6 FLOOD
7 TSTM WIND
8 TORNADO
```

Recommendations :

1. Further detailed study to identify patterns in event times across seasons and states can help understand and plan actions that could mitigate the harmful effects of weather events.Beginning date, end date and state columns may be used to analyse.
2. Based on outputs from any detected patterns may help identify further fields of study to identify root causes, eliminating which can potentially mitigate risks.

Appendix-I

```
## Code for assessing impact to health. fig.width=7,fig.height=10

# get the sum of health damages
healthdamage=aggregate(fatalities+injuries~evtype,wthrevents,sum)
names(healthdamage)[2]="healthdamage" #Rename the second column

# Sort on 2nd column
healthdamage=healthdamage[order(healthdamage$healthdamage,decreasing=T,na.last=T),]

# Add a column for cumulative percentage of the damage.
healthdamage$cumpercent=round(cumsum(healthdamage$healthdamage)/
                                 sum(healthdamage$healthdamage)*100,0)

# re-sort to increasing to get the barplot as required
healthdamage=healthdamage[order(healthdamage$healthdamage,na.last=T),]

## Code for assessing impact to economy.

# get the sum of economic damages
ecodamage=aggregate(propdmg+cropdmg~evtype,wthrevents,sum)
names(ecodamage)[2]="ecodamage" #Rename the second column

# Sort on 2nd column
ecodamage=ecodamage[order(ecodamage$ecodamage,decreasing=T,na.last=T),]

# Add a column for cumulative percentage of the damage.
ecodamage$cumpercent=round(cumsum(ecodamage$ecodamage)/
                              sum(ecodamage$ecodamage)*100,0)

# re-sort to increasing to get the barplot as required
ecodamage=ecodamage[order(ecodamage$ecodamage,na.last=T),]

## Now to Plot the barplots

# Set the environment

# Set left margin to increase space for axis text & fig rows to 2
par(mar=c(4,12,4,2),mfrow=c(2,1))

# bar plot for top 90% health damages
barplot(healthdamage$healthdamage[healthdamage$cumpercent<=90],
        names.arg=healthdamage$evtype[healthdamage$cumpercent<=90],
        horiz=T,las=1,col="red4",
        main="Population Health Vs. Event Types",
        xlab="Total Number of Fatalities & Injuries")
mtext("(Top 90% Incidents)")

# bar plot for top 90% economic damages
barplot(ecodamage$ecodamage[ecodamage$cumpercent<=90],
        names.arg=ecodamage$evtype[ecodamage$cumpercent<=90],
        horiz=T,las=1,col="red2",
```

```
        main="Economic Losses Vs. Event Types",
        xlab="Total Value of Property & Crop Damage - $K")
mtext("(Top 90% Incidents & Values in $K)")

# Events common to both categories of damage
comevents=intersect(healthdamage$evtype[healthdamage$cumpercent<=90],
                    ecodamage$evtype[ecodamage$cumpercent<=90])

# Percentage total damage of common events
totperdmg=round(sum(healthdamage$healthdamage[healthdamage$evtype%in%comevents]
            ,ecodamage$ecodamage[ecodamage$evtype%in%comevents])/
            (sum(healthdamage$healthdamage,ecodamage$ecodamage))*100,0)
```