Reproducible Research: Peer Assessment 2

10/8/15

An Analysis Report of Health and Economic Impact by Severe Weather Events - Based on NOAA Storm Database

Synopsis

Storm and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severs events can results in fatalities, injuries and property damage. Preventing such outcomes to the extent possible is a key concern. The U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database tracks characteristics of major storms and weather events in the United States, include when and where they occur, aswell as estimates of any fatalities, injuries and property damage. This report contains the exploratory analysis results on the health and economic impact by the severe weather events based on the data from NOAA database.

Data Processing

Loading the data

download file from URL

```
if(!file.exists('StormData.csv.bz2')){
    download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                  destfile='StormData.csv.bz2')

}

if(file.exists('StormData.csv.bz2')){
    # uncompress the file and read the CSV file into a data file
    storm <- read.csv(bzfile('StormData.csv.bz2'), header = TRUE)
}
```

```
## Warning in scan(file, what, nmax, sep, dec, quote, skip, nlines,
## na.strings, : EOF within quoted string
```

load data into R

```
head(storm)
```

```
##   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 END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

```
str(storm)
```

```
## 'data.frame':    386258 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 12348 levels "1/1/1966 0:00:00",..: 4453 4453 2791 8165 1479 1479 1504 229 2
##  $ BGN_TIME  : Factor w/ 2888 levels "000","0000","0001",..: 212 227 2405 1623 2344 2646 182 1623 264
##  $ TIME_ZONE : Factor w/ 20 levels "ADT","AST","CDT",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 10968 levels "","ABBEVILLE",..: 5199 649 2174 4159 1979 4001 739 8945 9091 2
##  $ STATE     : Factor w/ 58 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 919 levels " COASTAL FLOOD",..: 783 783 783 783 783 783 783 783 783 783 ...
##  $ 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/ 34327 levels "","- 1 N Albion",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 2675 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 2925 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/ 20887 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/ 539 levels ""," CI","$AC",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 224 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 8011 levels "","ABBEVILLE - ABBEVILLE - LAURENS - UNION - CHESTER - GREENWOO
##  $ 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/ 103810 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 ...
```

There are 902297 observations and 37 variables.

Preprocess the data

There are 7 variables we are interested regarding the two questions. They are:-

EVTYPE as a measure of event type (e.g. tornado, flood, etc.)

FATALITIES as a measure of harm to human health

INJURIES as a measure of harm to human health

PROPDMG as a measure of property damage and hence economic damage in USD

PROPDMGEXP as a measure of magnitude of property damage (e.g. thousands, millions USD, etc.)

CROPDMG as a measure of crop damage and hence economic damage in USD

CROPDMGEXP as a measure of magnitude of crop damage (e.g. thousands, millions USD, etc.)

To increase the computation speed, we can select these columns to make subsequent computation and analysis faster.

```
mydata <- storm[c(-1,-2,-3,-4,-5,-6,-7,-9,-10,-11,-12,-13,-14,-15,-16,-17,-18,-19,-20,-21,-22,-29,-30,-3
str(mydata)
```

```
## 'data.frame':    386258 obs. of  7 variables:
##  $ EVTYPE    : Factor w/ 919 levels " COASTAL FLOOD",..: 783 783 783 783 783 783 783 783 783 783 ...
##  $ 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 ...
```

Preparing the property damage data

exploring the property exponent

```r
unique(mydata$PROPDMGEXP)
```

```
##  [1] K M   B m + 0 5 6 ? 4 2 3 h 7 H - 1 8
## Levels:  - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
```

Sorting the property exponent data

```r
mydata$PROPEXP[mydata$PROPDMGEXP == "K"] <- 1000
mydata$PROPEXP[mydata$PROPDMGEXP == "M"] <- 1e+06
mydata$PROPEXP[mydata$PROPDMGEXP == ""] <- 1
mydata$PROPEXP[mydata$PROPDMGEXP == "B"] <- 1e+09
mydata$PROPEXP[mydata$PROPDMGEXP == "m"] <- 1e+06
mydata$PROPEXP[mydata$PROPDMGEXP == "0"] <- 1
mydata$PROPEXP[mydata$PROPDMGEXP == "5"] <- 1e+05
mydata$PROPEXP[mydata$PROPDMGEXP == "6"] <- 1e+06
mydata$PROPEXP[mydata$PROPDMGEXP == "4"] <- 10000
mydata$PROPEXP[mydata$PROPDMGEXP == "2"] <- 100
mydata$PROPEXP[mydata$PROPDMGEXP == "3"] <- 1000
mydata$PROPEXP[mydata$PROPDMGEXP == "h"] <- 100
mydata$PROPEXP[mydata$PROPDMGEXP == "7"] <- 1e+07
mydata$PROPEXP[mydata$PROPDMGEXP == "H"] <- 100
mydata$PROPEXP[mydata$PROPDMGEXP == "1"] <- 10
mydata$PROPEXP[mydata$PROPDMGEXP == "8"] <- 1e+08
```

give 0 to invalid exponent data, so they not count in

```r
mydata$PROPEXP[mydata$PROPDMGEXP == "+"] <- 0
mydata$PROPEXP[mydata$PROPDMGEXP == "-"] <- 0
mydata$PROPEXP[mydata$PROPDMGEXP == "?"] <- 0
```

compute the property damage value

```r
mydata$PROPDMGVAL <- mydata$PROPDMG * mydata$PROPEXP
```

Preparing the crop damage data

exploring the crop exponent data

```r
unique(mydata$CROPDMGEXP)
```

```
## [1]   M K m B ? 0 k 2
## Levels:  ? 0 2 B k K m M
```

Sorting the property exponent data

```r
mydata$CROPEXP[mydata$CROPDMGEXP == "M"] <- 1e+06
mydata$CROPEXP[mydata$CROPDMGEXP == "K"] <- 1000
mydata$CROPEXP[mydata$CROPDMGEXP == "m"] <- 1e+06
mydata$CROPEXP[mydata$CROPDMGEXP == "B"] <- 1e+09
mydata$CROPEXP[mydata$CROPDMGEXP == "0"] <- 1
mydata$CROPEXP[mydata$CROPDMGEXP == "k"] <- 1000
mydata$CROPEXP[mydata$CROPDMGEXP == "2"] <- 100
mydata$CROPEXP[mydata$CROPDMGEXP == ""] <- 1
```

give 0 to invalid exponent data, so they not count in

```r
mydata$CROPEXP[mydata$CROPDMGEXP == "?"] <- 0
```

compute the crop damage value

```r
mydata$CROPDMGVAL <- mydata$CROPDMG * mydata$CROPEXP
```

Aggregate the data by event

aggregate the data by event

```r
fatal <- aggregate(FATALITIES ~ EVTYPE, data = mydata, FUN = sum)
injury <- aggregate(INJURIES ~ EVTYPE, data = mydata, FUN = sum)
propdmg <- aggregate(PROPDMGVAL ~ EVTYPE, data = mydata, FUN = sum)
cropdmg <- aggregate(CROPDMGVAL ~ EVTYPE, data = mydata, FUN = sum)
```

Results

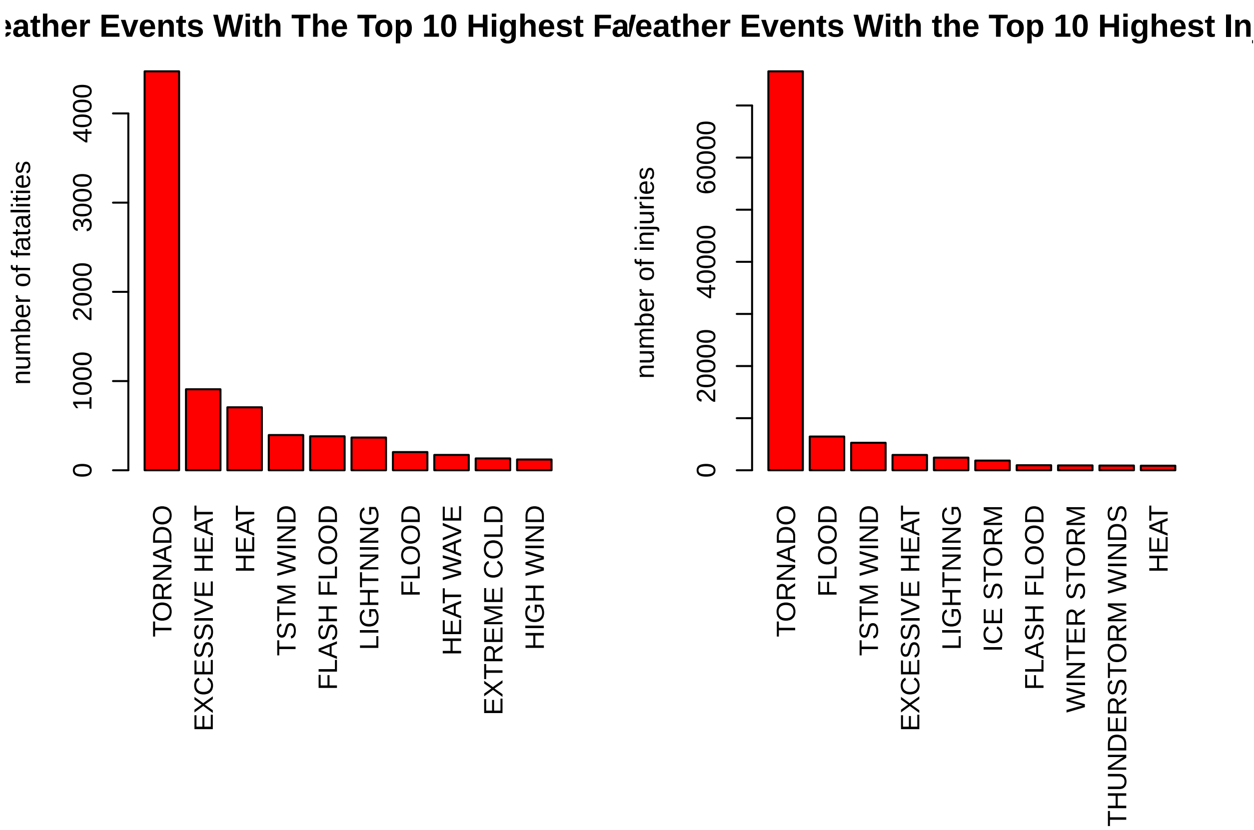
Aross the United States, Which types of events are nost harmful with respect to population health?

get top10 event with highest fatalities

```r
fatal10 <- fatal[order(-fatal$FATALITIES), ][1:10, ]
```

get top10 event with highest injuries

```r
injury10 <- injury[order(-injury$INJURIES), ][1:10, ]
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(fatal10$FATALITIES, las = 3, names.arg = fatal10$EVTYPE, main = "Weather Events With The Top 10
    ylab = "number of fatalities", col = "red")
barplot(injury10$INJURIES, las = 3, names.arg = injury10$EVTYPE, main = "Weather Events With the Top 10
    ylab = "number of injuries", col = "red")
```



The most harmful weather event to population health is Tornado.It is cause for both the highest fatalities and the highest injuries across United States.

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

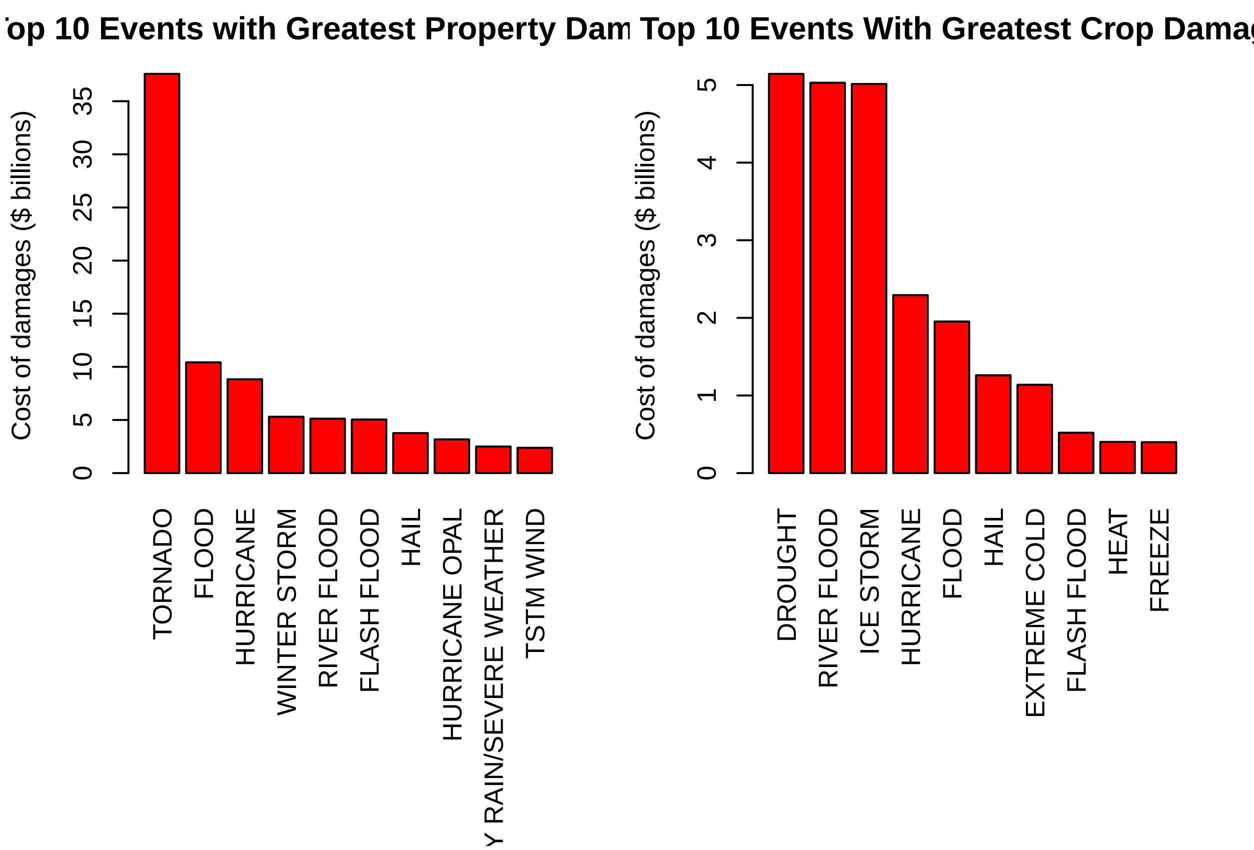
get top 10 events with highest property damage

```r
propdmg10 <- propdmg[order(-propdmg$PROPDMGVAL), ][1:10, ]
```

Get top 10 events with highest crop damage

```r
cropdmg10 <- cropdmg[order(-cropdmg$CROPDMGVAL), ][1:10, ]
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(propdmg10$PROPDMGVAL/(10^9), las = 3, names.arg = propdmg10$EVTYPE,
    main = "Top 10 Events with Greatest Property Damages", ylab = "Cost of damages ($ billions)",
```

```
    col = "red")
barplot(cropdmg10$CROPDMGVAL/(10^9), las = 3, names.arg = cropdmg10$EVTYPE,
    main = "Top 10 Events With Greatest Crop Damages", ylab = "Cost of damages ($ billions)",
    col = "red")
```



##The weather events have the greatest economic damage are: flood, drought, Tornado and Typhoon.

Across the United States, flood, tornado and typhoon have caused the greatest damage to properties.

Drought and flood come as the causes for the greatest damage to crops.