Analysis of Storms, and other severe weather events that cause both public health and economic problems in United States.

by clpong

Thursday, October 22, 2015

1. Synopsis

The National Oceanic and Atmospheric Administration (NOAA) maintains a public database for storm event. The data contains the type of storm event, details like location, date, estimates for damage to property as well as the number of human victims of the storm. In this report we investigate which type of events are the most harmful to the population and financially.

The conclusion is that the impact on humans, be it injuries or fatalities, isn't directly correlated to the ecomomic damage weather events cause. Tornado is the highest cause for injuries and fatalities, whilst flood and drought cause the greatest cost of damages on the cops and properties.

2. Data Processing

2.1 Load all the libraried used for data loading, computation and plotting

```
library(R.utils)
library(data.table)
```

2.2 Unzip the file and process it:-

The data file is downloaded and copied to the working directory.

```
bunzip2("repdata-data-StormData.csv.bz2", overwrite=T, remove=F)
data <- read.csv("repdata-data-StormData.csv")
```

We will perform the basic data analysis:

```
summary(data)
```

```
##     STATE__                  BGN_DATE             BGN_TIME
##  Min.   : 1.0   5/25/2011 0:00:00:  1202   12:00:00 AM: 10163
##  1st Qu.:19.0   4/27/2011 0:00:00:  1193   06:00:00 PM:  7350
##  Median :30.0   6/9/2011 0:00:00 :  1030   04:00:00 PM:  7261
##  Mean   :31.2   5/30/2004 0:00:00:  1016   05:00:00 PM:  6891
##  3rd Qu.:45.0   4/4/2011 0:00:00 :  1009   12:00:00 PM:  6703
##  Max.   :95.0   4/2/2006 0:00:00 :   981   03:00:00 PM:  6700
##                 (Other)          :895866   (Other)    :857229
##    TIME_ZONE          COUNTY           COUNTYNAME         STATE
##  CST    :547493   Min.   :  0.0   JEFFERSON :  7840   TX     : 83728
##  EST    :245558   1st Qu.: 31.0   WASHINGTON:  7603   KS     : 53440
##  MST    : 68390   Median : 75.0   JACKSON   :  6660   OK     : 46802
##  PST    : 28302   Mean   :100.6   FRANKLIN  :  6256   MO     : 35648
##  AST    :  6360   3rd Qu.:131.0   LINCOLN   :  5937   IA     : 31069
##  HST    :  2563   Max.   :873.0   MADISON   :  5632   NE     : 30271
##  (Other):  3631                   (Other)   :862369   (Other):621339
##                EVTYPE         BGN_RANGE           BGN_AZI
##  HAIL              :288661   Min.   :   0.000          :547332
##  TSTM WIND         :219940   1st Qu.:   0.000   N      : 86752
##  THUNDERSTORM WIND: 82563   Median :   0.000   W      : 38446
##  TORNADO           : 60652   Mean   :   1.484   S      : 37558
##  FLASH FLOOD       : 54277   3rd Qu.:   1.000   E      : 33178
##  FLOOD             : 25326   Max.   :3749.000   NW     : 24041
##  (Other)           :170878                      (Other):134990
##          BGN_LOCATI                  END_DATE             END_TIME
##                :287743                     :243411             :238978
##  COUNTYWIDE    : 19680   4/27/2011 0:00:00:  1214   06:00:00 PM:  9802
##  Countywide    :   993   5/25/2011 0:00:00:  1196   05:00:00 PM:  8314
##  SPRINGFIELD   :   843   6/9/2011 0:00:00 :  1021   04:00:00 PM:  8104
##  SOUTH PORTION:   810   4/4/2011 0:00:00 :  1007   12:00:00 PM:  7483
##  NORTH PORTION:   784   5/30/2004 0:00:00:   998   11:59:00 PM:  7184
##  (Other)       :591444   (Other)          :653450   (Other)    :622432
##    COUNTY_END COUNTYENDN       END_RANGE           END_AZI
##  Min.   :0    Mode:logical   Min.   :  0.0000          :724837
##  1st Qu.:0    NA's:902297    1st Qu.:  0.0000   N      : 28082
##  Median :0                   Median :  0.0000   S      : 22510
##  Mean   :0                   Mean   :  0.9862   W      : 20119
##  3rd Qu.:0                   3rd Qu.:  0.0000   E      : 20047
##  Max.   :0                   Max.   :925.0000   NE     : 14606
##                                                 (Other): 72096
##             END_LOCATI         LENGTH              WIDTH
##                   :499225   Min.   :   0.0000   Min.   :   0.000
##  COUNTYWIDE       : 19731   1st Qu.:   0.0000   1st Qu.:   0.000
##  SOUTH PORTION    :   833   Median :   0.0000   Median :   0.000
##  NORTH PORTION    :   780   Mean   :   0.2301   Mean   :   7.503
##  CENTRAL PORTION:   617   3rd Qu.:   0.0000   3rd Qu.:   0.000
##  SPRINGFIELD      :   575   Max.   :2315.0000   Max.   :4400.000
##  (Other)          :380536
##        F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
```

```
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG          PROPDMGEXP        CROPDMG          CROPDMGEXP
##  Min.   :   0.00          :465934   Min.   :  0.000          :618413
##  1st Qu.:   0.00   K      :424665   1st Qu.:  0.000   K      :281832
##  Median :   0.00   M      : 11330   Median :  0.000   M      :  1994
##  Mean   :  12.06   0      :   216   Mean   :  1.527   k      :    21
##  3rd Qu.:   0.50   B      :    40   3rd Qu.:  0.000   0      :    19
##  Max.   :5000.00   5      :    28   Max.   :990.000   B      :     9
##                    (Other):    84                     (Other):     9
##       WFO                                       STATEOFFIC
##         :142069                                       :248769
##  OUN    : 17393   TEXAS, North                       : 12193
##  JAN    : 13889   ARKANSAS, Central and North Central: 11738
##  LWX    : 13174   IOWA, Central                      : 11345
##  PHI    : 12551   KANSAS, Southwest                  : 11212
##  TSA    : 12483   GEORGIA, North and Central         : 11120
##  (Other):690738   (Other)                            :595920
##
ZONENAMES
##
:594029
##
:205988
##  GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
:   639
##  GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA
:   592
##  JEFFERSON - JEFFERSON
:   303
##  MADISON - MADISON
:   302
##  (Other)
:100444
##     LATITUDE      LONGITUDE        LATITUDE_E     LONGITUDE_
##  Min.   :   0   Min.   :-14451   Min.   :   0   Min.   :-14455
##  1st Qu.:2802   1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0
##  Median :3540   Median :  8707   Median :   0   Median :     0
##  Mean   :2875   Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:4019   3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   :9706   Max.   : 17124   Max.   :9706   Max.   :106220
##  NA's   :47                      NA's   :40
##                                            REMARKS           REFNUM
##                                                :287433   Min.   :     1
##                                                : 24013   1st Qu.:225575
##  Trees down.\n                                 :  1110   Median :451149
##  Several trees were blown down.\n              :   569   Mean   :451149
##  Trees were downed.\n                          :   446   3rd Qu.:676723
##  Large trees and power lines were blown down.\n:   432   Max.   :902297
##  (Other)                                       :588294
```

```
names(data)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

From all the columns, only several of them will be revelant to our analysis. These are: * EVTYPE: the type of weather event * FATALITIES: the number of fatalities * INJURIES: the number of injuries * PROPDMG: the amount of property damage (in US dollars) * PROPDMGEXP: a multiplier for PROPDMG * CROPDMG: the amount of crop damage (in US dollars) * CROPDMGEXP: a multiplier for CROPDMG

2.3 Remove the unuse columns (not needed for analysis)

```
use_columns <- c("BGN_DATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDM
G", "CROPDMGEXP")
storm_data <- data[,use_columns]
storm_data$EVTYPE <- toupper(data$EVTYPE)
storm_data <- storm_data[!grepl("Summary", data$EVTYPE), ]
```

3.Results

3.1 Most harmful events with respect to population health

Find the top 10 weather event that causes fatalities and injuries

```
dt_storm_data <- as.data.table(storm_data)
dt_fatalities <- dt_storm_data[,sum(FATALITIES),by=EVTYPE]
dt_fatalities <- dt_fatalities[order(-rank(V1))]
dt_fatalities <- dt_fatalities[1:10,]
setnames(dt_fatalities, c('Event','Total'))
dt_fatalities
```

```
##              Event Total
##  1:         TORNADO  5633
##  2: EXCESSIVE HEAT  1903
##  3:     FLASH FLOOD   978
##  4:           HEAT   937
##  5:      LIGHTNING   816
##  6:      TSTM WIND   504
##  7:          FLOOD   470
##  8:    RIP CURRENT   368
##  9:      HIGH WIND   248
## 10:      AVALANCHE   224
```

```
dt_injuries <- dt_storm_data[,sum(INJURIES), by=EVTYPE]
dt_injuries <- dt_injuries[order(-rank(V1))]
dt_injuries <- dt_injuries[1:10,]
setnames(dt_injuries, c('Event','Total'))
dt_injuries
```

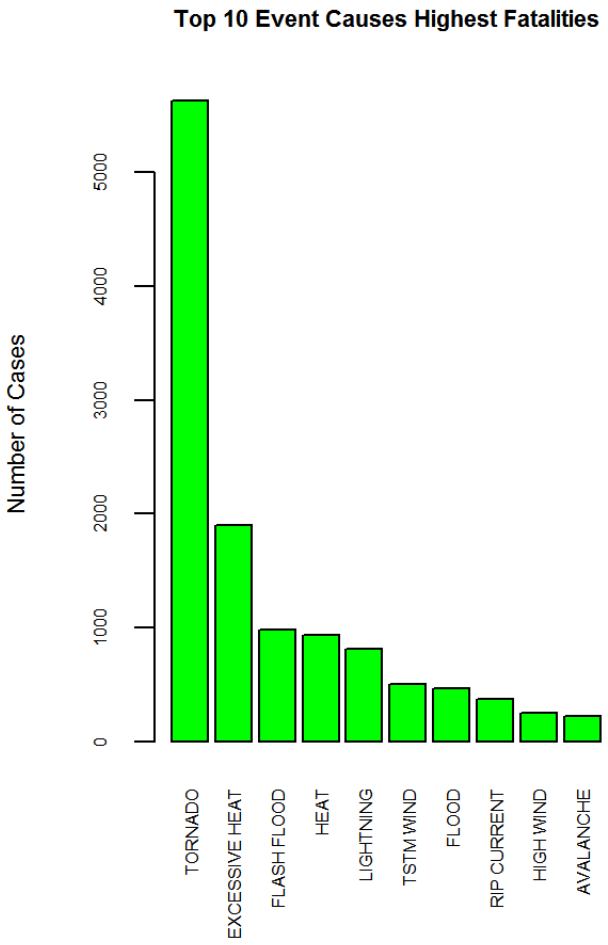
```
##                 Event Total
##  1:           TORNADO 91346
##  2:         TSTM WIND  6957
##  3:             FLOOD  6789
##  4:    EXCESSIVE HEAT  6525
##  5:         LIGHTNING  5230
##  6:              HEAT  2100
##  7:         ICE STORM  1975
##  8:       FLASH FLOOD  1777
##  9: THUNDERSTORM WIND  1488
## 10:              HAIL  1361
```

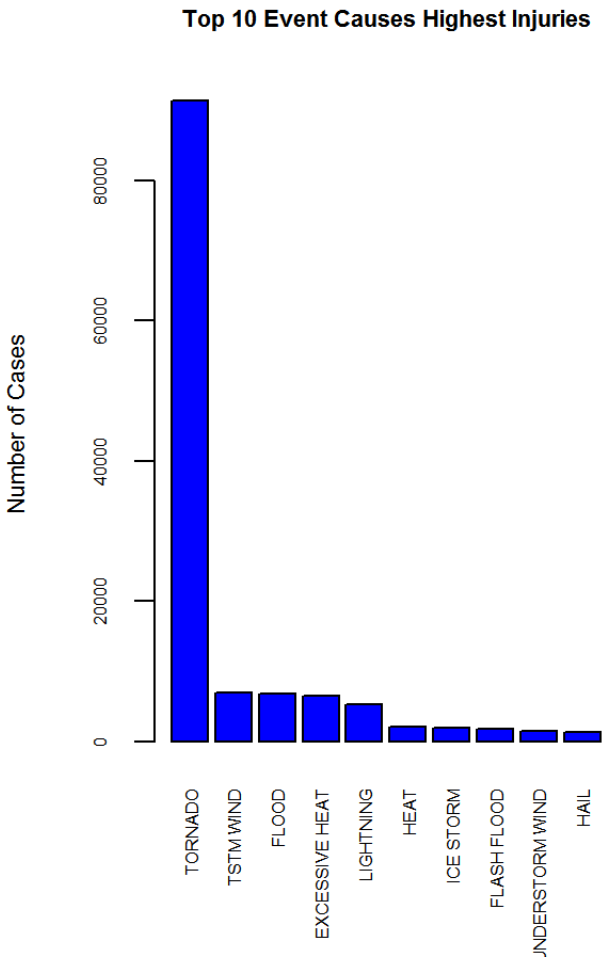
Plot graph for top 10 weather events classified by number of fatalities and injuries

```
par(mfrow=c(1,2))

barplot(dt_fatalities$Total, las=3, names.arg=dt_fatalities$Event, col="green", main='Top 10 Event Causes Highest Fatalities',ylab='Number of Cases',cex.lab=0.7,cex.main=0.7,cex.axis=0.5,cex.name=0.5)

barplot(dt_injuries$Total, las=3, names.arg=dt_injuries$Event, col="blue", main='Top 10 Event Causes Highest Injuries',ylab='Number of Cases',cex.lab=0.7,cex.main=0.7,cex.axis=0.5,cex.name=0.5)
```





From these plots, we see that tornadoes are the weather events that caused the most injuries and fatalities between 1950 and 2011 across the U.S. If we consider injuries, the 2nd and 3rd most harmful types of events are respectively thunderstorm winds and floods. In the case of fatalities, the equivalent are excessive heats and flash floods.

3.2 The events that had the greatest ecomomic consequences

Calculate the costs of the damages on crops and properties

```
# Set to upper case
dt_storm_data$CROPDMGEXP <- toupper(dt_storm_data$CROPDMGEXP)
dt_storm_data$PROPDMGEXP <- toupper(dt_storm_data$PROPDMGEXP)

# Take the unique values
cropdmg_exp <- as.character(unique(dt_storm_data$CROPDMGEXP))
propdmg_exp <- as.character(unique(dt_storm_data$PROPDMGEXP))

# Combine the unique upper case values using union
damage_exp <- union(cropdmg_exp, propdmg_exp)

damage_values <- function(x) {
    e <- list(K=3, M=6, B=9,
              "+"=0, "-"=0, "?"=0,
              "0"=0, "1"=1, "2"=2, "3"=3, "4"=4, "5"=5, "6"=6, "7"=7, "8"=8, "9"=9)

    if (x %in% names(e)) {
            int_value <- e[[x]]
    }
    else {
        int_value <- 0
    }
    return(int_value)
}

calculate_cost <- function(damage, damage_exp) {
    damage*10^damage_values(damage_exp)
}

dt_storm_data$CROPDMG_COST = mapply(calculate_cost, dt_storm_data$CROPDMG, dt_storm_data$CROPDMGEX
P)
dt_storm_data$PROPDMG_COST = mapply(calculate_cost, dt_storm_data$PROPDMG, dt_storm_data$PROPDMGEX
P)
```

Find the top 10 events having the greatest cost of damages on the crops and properties

```
dt_crops <- dt_storm_data[,sum(CROPDMG_COST), by=EVTYPE]
dt_crops <- dt_crops[order(-rank(V1))]
dt_crops<- dt_crops[1:10,]
setnames(dt_crops, c('Event','Total'))
dt_crops
```

```
##                Event       Total
##  1:           DROUGHT 13972566000
##  2:             FLOOD  5661968450
##  3:       RIVER FLOOD  5029459000
##  4:         ICE STORM  5022113500
##  5:              HAIL  3025954473
##  6:         HURRICANE  2741910000
##  7: HURRICANE/TYPHOON  2607872800
##  8:       FLASH FLOOD  1421317100
##  9:      EXTREME COLD  1312973000
## 10:      FROST/FREEZE  1094186000
```

```
dt_properties <- dt_storm_data[,sum(PROPDMG_COST), by=EVTYPE]
dt_properties <- dt_properties[order(-rank(V1))]
dt_properties <- dt_properties[1:10,]
setnames(dt_properties, c('Event','Total'))
dt_properties
```

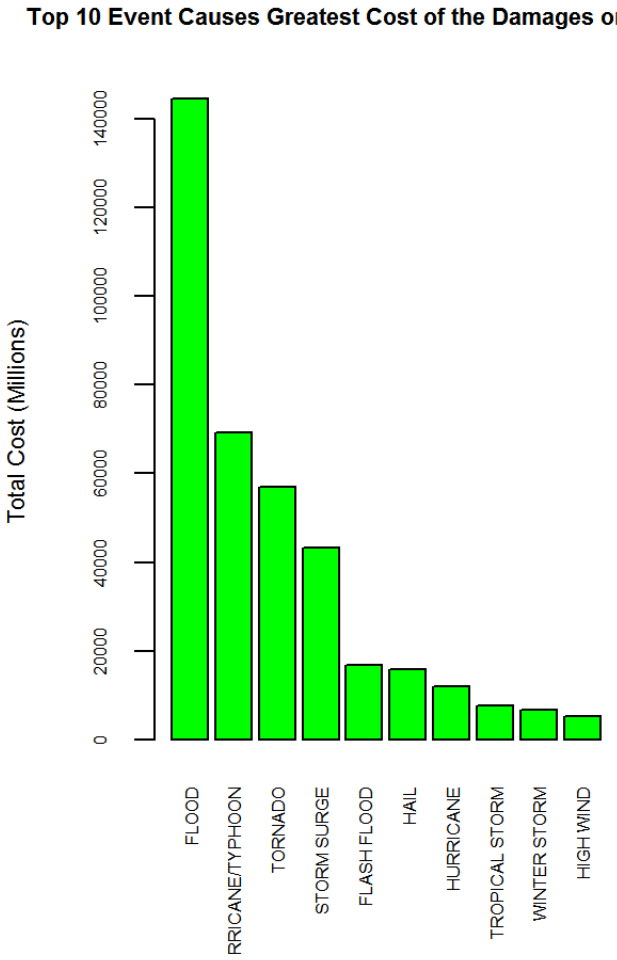
```
##                Event        Total
##  1:             FLOOD 144657709807
##  2: HURRICANE/TYPHOON  69305840000
##  3:           TORNADO  56947380677
##  4:       STORM SURGE  43323536000
##  5:       FLASH FLOOD  16822673979
##  6:              HAIL  15735267018
##  7:         HURRICANE  11868319010
##  8:    TROPICAL STORM   7703890550
##  9:      WINTER STORM   6688497251
## 10:         HIGH WIND   5270046295
```

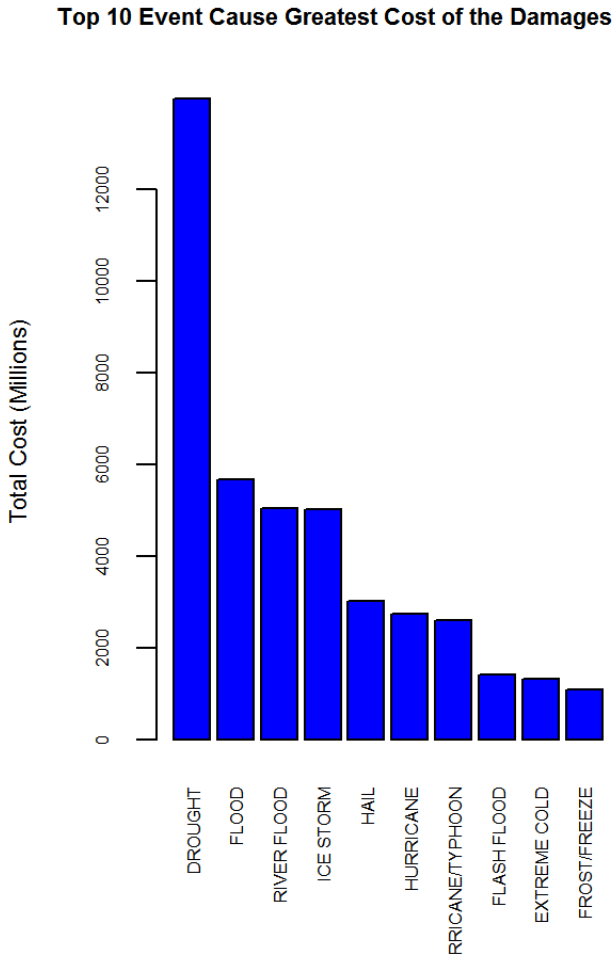
Plot graph for top 10 weather events having the greatest economic consequences

```
par(mfrow=c(1,2))

barplot(dt_properties$Total/1000000, las=3, names.arg=dt_properties$Event, col="green", main='Top 1
0 Event Causes Greatest Cost of the Damages on Properties',ylab='Total Cost (Millions)',cex.la
b=0.7,cex.main=0.7,cex.axis=0.5,cex.name=0.5)

barplot(dt_crops$Total/1000000, las=3, names.arg=dt_crops$Event, col="blue", main='Top 10 Event Cau
se Greatest Cost of the Damages on Crops',ylab='Total Cost (Millions)',cex.lab=0.7,cex.main=0.7,ce
x.axis=0.5,cex.name=0.5)
```





Based on the histograms above, we find that flood and hurricane/typhoon cause most property damage; drought and flood causes most crop damage in the United States from 1995 to 2011.

4. Conclusion

From these data, we found that excessive heat and tornado are most harmful with respect to population health, while flood, drought, and hurricane/typhoon have the greatest economic impact.