Severe weather events impacts on the US communities and municipalities

Reproducible Research: Peer Assessment 2

Eddy Delta

30/04/2017

Synopsis

This report aim is to provide an analysis of severe weather events impacts on the US communities and municipalities, from both public health and economic problems perspective. We will study fatalities, injuries, property, and crop damage weather event can generate. A better understanding of such impacts can help to prevent or handle more efficiently such outcomes. To do so, we will explore the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database.

From those information we'll try to identify :

1. Which types of events (as indicated in the **EVTYPE** variable) are most harmful with respect to population health across the United States.
2. Which types of events have the greatest economic consequences across the United States.

Data Processing

```
setwd('~/projects/training/datascience/reproducible_research/storm')

get_file <- function(file_url = 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz
    if(!dir.exists(data_dir)) dir.create(data_dir)
    file_name <- basename(file_url)
    file_path <- file.path(data_dir, file_name)
    if(!file.exists(file_path))
        download.file(file_url, file_path, quiet = TRUE)
    if(!file.exists(file_path))
        stop("Unable to download file ", file_url)
    # return the archive file name
    file_name
}
# if (!file.exists('load_data.R')) setwd('~/projects/training/datascience/reproducible_research/storm')
# source('load_data.R')

# Load data from files
file_archive <- get_file()
if(!exists('storm')){
    storm <- read.csv(file.path('data', file_archive), sep = ',', stringsAsFactors = FALSE, na.strings =
    # change the fips, SCC, Pollutant, and type column type to factor
    for(col in c('COUNTYNAME', 'STATE', 'STATEOFFIC', 'EVTYPE', 'ZONENAMES', 'WFO', 'BGN_LOCATI', 'END_
    # create begin and end date type and years
    storm$BGN_DATETIME <- strptime(paste(gsub('([0-9/]+).+', '\\1', storm$BGN_DATE), storm$BGN_TIME, st
    storm$BGN_YEAR <- as.numeric(format(storm$BGN_DATETIME, '%Y'))
    storm$END_DATETIME <- strptime(paste(gsub('([0-9/]+).+', '\\1', storm$END_DATE), storm$END_TIME, st
```

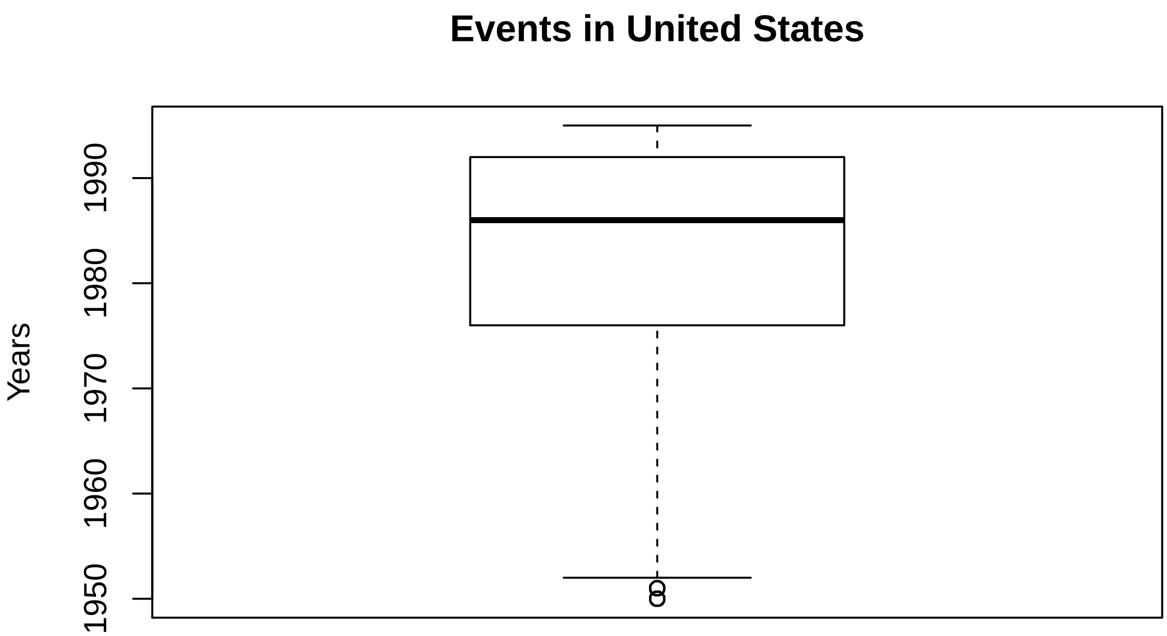
```
    storm$END_YEAR <- as.numeric(format(storm$END_DATETIME, '%Y'))
}
```

```
year_range <- range(storm$BGN_YEAR, na.rm = T)
```

The dataset contains 902297 rows and 41 columns. And log events from 1950 to 1995.

Let's first have a look at the number of events distribution across the years in United States.

```
# Build a boxplot to have an overview of the data distribution
# hist(storm$BGN_YEAR, main = 'Events in United States', xlab = 'Years', ylab = 'Events (nb)')
boxplot(storm$BGN_YEAR, main = 'Events in United States', ylab = 'Years')
```




```
evt_count <- aggregate(BGN_DATE ~ BGN_YEAR, data = storm, FUN = length)
evt_count_range <- range(evt_count$BGN_DATE, na.rm = T)
```

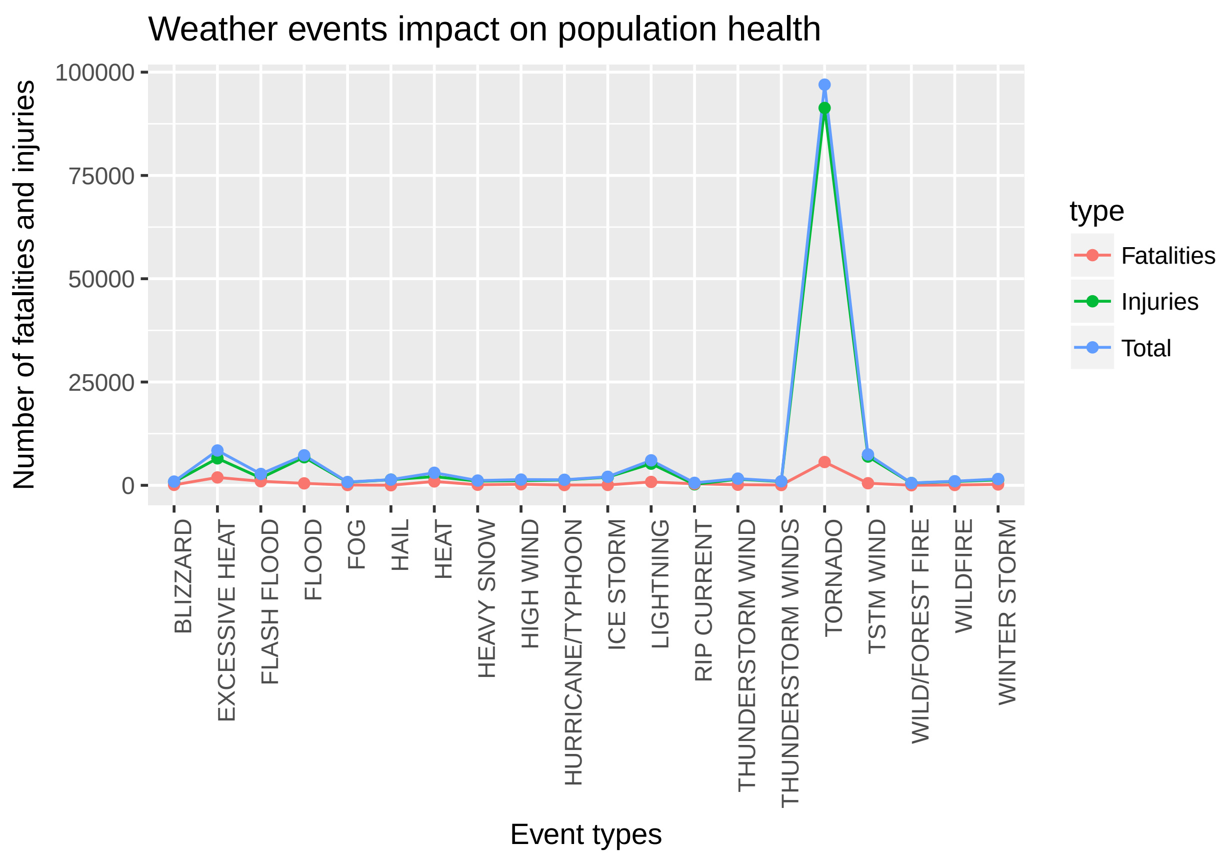
We can notice a constant increase of the count of events across the year. We are able to confirm the lack of recorded event in the early years in the dataset (from 223 to 27967).

Analysis of Severe weather events impact on population health across the United States

Let's now at the impact of the events on the population in terms of **INJURIES** and **FATALITIES**. First we need to have an idea of the most harmful events. To do so we'll have a look at the total numbers of **Injuries** and **Fatalities** per event types.

```
injuries <- aggregate(INJURIES ~ EVTYPE, storm, sum, na.rm = T)
fatalities <- aggregate(FATALITIES ~ EVTYPE, storm, sum, na.rm = T)
# put side by side the costs
harm <- merge(injuries, fatalities, by = 'EVTYPE')
harm$CASES <- harm$INJURIES + harm$FATALITIES
# sort the tables
injuries <- injuries[order(injuries$INJURIES, decreasing = T), ]
fatalities <- fatalities[order(fatalities$FATALITIES, decreasing = T), ]
harm <- harm[order(harm$CASES, decreasing = T), ]
```

```r
# reshape the table columns in a single column type
top_harm <- reshape(head(harm, 20), direction = 'long', varying = c('INJURIES', 'FATALITIES', 'CASES'),
library(ggplot2)
ggplot(data = top_harm, aes(x = EVTYPE, y = cases, group = type, colour = type)) +
    geom_line() + geom_point() +
    labs(title = 'Weather events impact on population health', x = 'Event types', y = 'Number of fatali
    theme(axis.text.x=element_text(angle = 90, hjust = 1))
```




```r
head(harm, 20)
```

```
##                  EVTYPE INJURIES FATALITIES CASES
## 834             TORNADO    91346       5633 96979
## 130      EXCESSIVE HEAT     6525       1903  8428
## 856           TSTM WIND     6957        504  7461
## 170               FLOOD     6789        470  7259
## 464           LIGHTNING     5230        816  6046
## 275                HEAT     2100        937  3037
## 153         FLASH FLOOD     1777        978  2755
## 427           ICE STORM     1975         89  2064
## 760   THUNDERSTORM WIND     1488        133  1621
## 972        WINTER STORM     1321        206  1527
## 359           HIGH WIND     1137        248  1385
## 244                HAIL     1361         15  1376
## 411   HURRICANE/TYPHOON     1275         64  1339
## 310          HEAVY SNOW     1021        127  1148
## 957            WILDFIRE      911         75   986
## 786  THUNDERSTORM WINDS      908         64   972
```

```
## 30            BLIZZARD      805         101    906
## 188                FOG      734          62    796
## 585        RIP CURRENT      232         368    600
## 955   WILD/FOREST FIRE      545          12    557
```

```
head(fatalities, 10)
```

```
##             EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

```
head(injuries, 10)
```

```
##                EVTYPE INJURIES
## 834           TORNADO    91346
## 856         TSTM WIND     6957
## 170             FLOOD     6789
## 130    EXCESSIVE HEAT     6525
## 464         LIGHTNING     5230
## 275              HEAT     2100
## 427         ICE STORM     1975
## 153       FLASH FLOOD     1777
## 760 THUNDERSTORM WIND     1488
## 244              HAIL     1361
```

TORNADO are by far the most harmful event, followed by **EXCESSIVE HEAT** and **TSTM WIND**.

TORNADO cost the most injuries, followed by **TSTM WIND** and **FLOOD**.

TORNADO the most lethals event, followed by **EXCESSIVE HEAT** and **FLASH FLOOD**.

Analysis of Severe weather events economical consequences across the United States

For this analysis we're going to pay particularly attention to **property** (**PROPDMG**) and **crop** (**CROPDMG**) damages.

We need to consider the factor which need to be applied on the values (National Weather Service Storm Data Documentation)

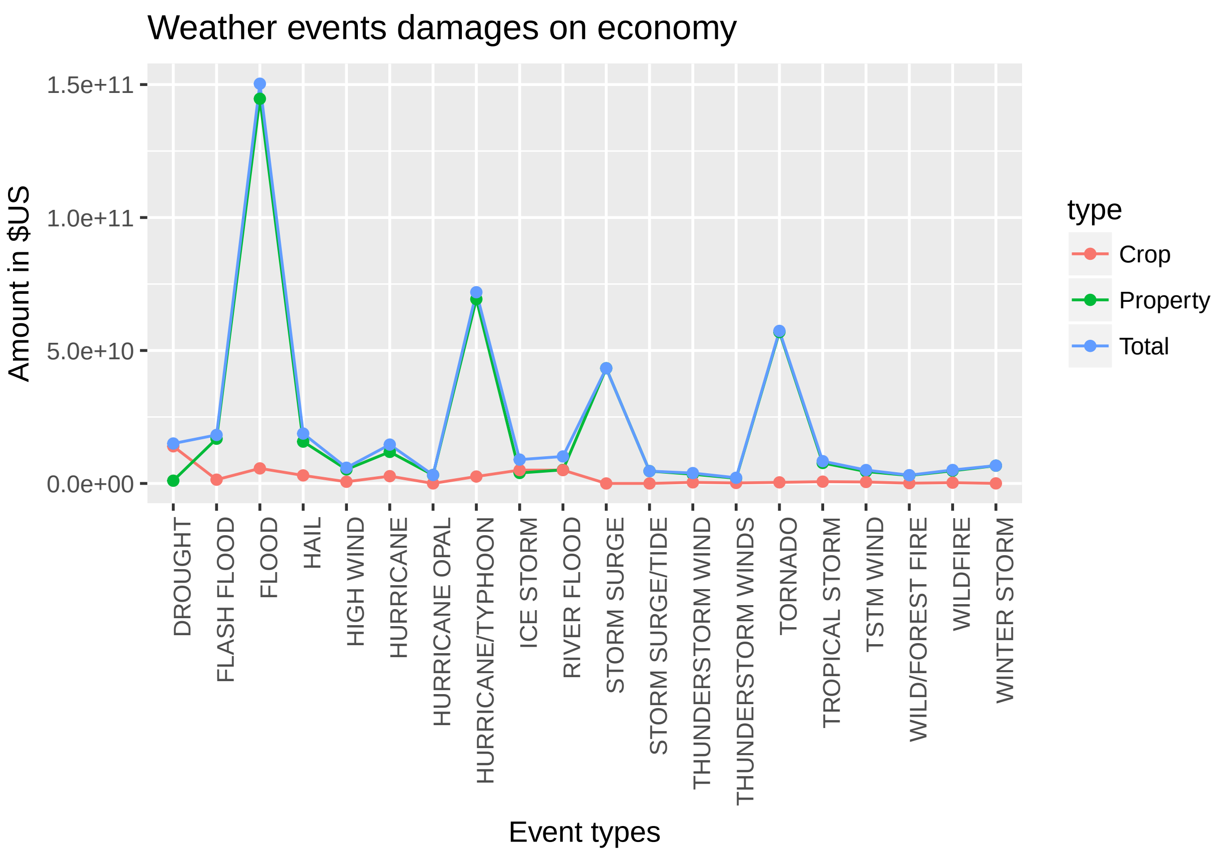
```
# get the cost on the same scale
tr <- c(H = '2', h = '2', K = '3', M = '6', m = '6', B = '9', '+' = '1', '-' = '-1', '?' = NA, '0' = '0
if(!'PROP' %in% names(storm)) storm$PROP <- with(storm, PROPDMG * 10^as.numeric(tr[PROPDMGEXP]))
if(!'CROP' %in% names(storm)) storm$CROP <- with(storm, CROPDMG * 10^as.numeric(tr[CROPDMGEXP]))
```

To do so we'll have a look at the total numbers of **property** and **crop** damages per event types.

```
prop <- aggregate(PROP ~ EVTYPE, storm, sum, na.rm = T)
crop <- aggregate(CROP ~ EVTYPE, storm, sum, na.rm = T)
# put side by side the costs
dmg <- merge(prop, crop, by = 'EVTYPE')
```

```r
dmg$AMOUNT <- dmg$PROP + dmg$CROP
# sort the tables
prop <- prop[order(prop$PROP, decreasing = T), ]
crop <- crop[order(crop$CROP, decreasing = T), ]
dmg <- dmg[order(dmg$AMOUNT, decreasing = T), ]
```

```r
# reshape the table columns in a single column type
top_dmg <- reshape(head(dmg, 20), direction = 'long', varying = c('PROP', 'CROP', 'AMOUNT'), v.names =
library(ggplot2)
ggplot(data = top_dmg, aes(x = EVTYPE, y = amount, group = type, colour = type)) +
    geom_line() + geom_point() +
    labs(title = 'Weather events damages on economy', x = 'Event types', y = 'Amount in $US') +
    theme(axis.text.x=element_text(angle = 90, hjust = 1))
```




```r
head(dmg, 20)
```

```
##                  EVTYPE         PROP        CROP       AMOUNT
## 26                FLOOD 144657709800  5661968450 150319678250
## 67    HURRICANE/TYPHOON  69305840000  2607872800  71913712800
## 109             TORNADO  56947381214   414953270  57362334484
## 92          STORM SURGE  43323536000        5000  43323541000
## 42                 HAIL  15735267456  3025537470  18760804926
## 22          FLASH FLOOD  16822673772  1421317100  18243990872
## 12              DROUGHT   1046106000 13972566000  15018672000
## 62            HURRICANE  11868319010  2741910000  14610229010
## 82          RIVER FLOOD   5118945500  5029459000  10148404500
## 69            ICE STORM   3944927860  5022113500   8967041360
```

```
## 113     TROPICAL STORM   7703890550   678346000   8382236550
## 133       WINTER STORM   6688497251    26944000   6715441251
## 57           HIGH WIND   5270046462   638571300   5908617762
## 128           WILDFIRE   4765114000   295472800   5060586800
## 117          TSTM WIND   4484928495   554007350   5038935845
## 93    STORM SURGE/TIDE   4641188000      850000   4642038000
## 99   THUNDERSTORM WIND   3483122470   414843050   3897965520
## 65      HURRICANE OPAL   3172846000    19000000   3191846000
## 126   WILD/FOREST FIRE   3001829500   106796830   3108626330
## 100 THUNDERSTORM WINDS   1944590658   190650780   2135241438
```

```r
head(prop, 10)
```

```
##                EVTYPE         PROP
## 63              FLOOD 144657709800
## 181 HURRICANE/TYPHOON  69305840000
## 335           TORNADO  56947381214
## 283       STORM SURGE  43323536000
## 51        FLASH FLOOD  16822673772
## 105              HAIL  15735267456
## 173         HURRICANE  11868319010
## 343    TROPICAL STORM   7703890550
## 402      WINTER STORM   6688497251
## 158         HIGH WIND   5270046462
```

```r
head(crop, 10)
```

```
##               EVTYPE        CROP
## 16           DROUGHT 13972566000
## 34             FLOOD  5661968450
## 98       RIVER FLOOD  5029459000
## 85         ICE STORM  5022113500
## 52              HAIL  3025537470
## 77         HURRICANE  2741910000
## 82 HURRICANE/TYPHOON  2607872800
## 30       FLASH FLOOD  1421317100
## 26      EXTREME COLD  1292973000
## 46      FROST/FREEZE  1094086000
```

FLOOD are the most expensive event, followed by **HURRICANE/TYPHOON** and **TORNADO**.

FLOOD are the most expensive event on properties, followed by **HURRICANE/TYPHOON** and **TORNADO**.

DROUGHT are the most expensive event on crop, followed by **FLOOD** and **RIVER FLOOD**.

Results

As a conclusion we can safely consider **TORNADO** as the most dangerous weather event for the population.

The **FLOOD** causes the greatest economic consequences on the US communities and municipalities in general and on the properties in particular. Meanwhile the **DROUGHT** have a huge cost on the crops.