

The Analysis of the health and economic impact of storms and other severe weather events in the USA from 1950 to 2011

Franklin Dono

10/24/2020

Executive summary

This document studies the health and economic impact of storms and other severe weather events that occurred in the United States of America from 1950 to 2011. It explores data from <https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>, which is the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. It ranks and isolates the top 10% of weather events that have the worst economic and population health impact. For the purposes of this exercise population-health-impact of an event is measured by the sum of the number of fatalities and injuries resulting directly or indirectly from the event. Economic impact on the other hand is measured by the sum of the cost of damages to property and crops.

The study identified 22 storm weather events with the worst population-health impact and 13 storm weather events with the worst economic impact over the period of study. Overall, Tornado have had the worst effect on population health over the period with 96,979 fatalities and injuries and cost of flood related damages to property and crops is at least 138 billion us dollars.

Data processing

Using R version 4.0.2, load the following libraries; dplyr, ggplot2 and lubridate for the analysis:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

Download and read the data from source <https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>. It is important to check first to ensure that a copy of the data is not already downloaded into the working directory. The data consist of 903871 observations of 37 variables

```
## The source url to the data
data.source <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

## Downloading data
if(!file.exists("StormData.csv.bz2")){
  download.file(data.source, "./StormData.csv.bz2")
}

## Reading data into R
stormdata <- read.csv("StormData.csv.bz2")
```

For the purpose of assessing the top ranking harsh weather events in respect of their adverse impact on economic activities and population health a subset of the data relevant to the objectives was selected.

```
## Sub setting
stormdata = select(stormdata, REFNUM, BGN_DATE, STATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP)

head(tibble(stormdata))
```

```
## # A tibble: 6 x 10
##   REFNUM BGN_DATE STATE EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG
##   <dbl> <chr>    <chr> <chr>      <dbl>    <dbl>    <dbl> <chr>      <dbl>
## 1     1  4/18/19~ AL   TORNA~      0      15     25    K          0
## 2     2  4/18/19~ AL   TORNA~      0       0     2.5    K          0
## 3     3  2/20/19~ AL   TORNA~      0       2     25     K          0
## 4     4  6/8/195~ AL   TORNA~      0       2     2.5    K          0
## 5     5  11/15/1~ AL   TORNA~      0       2     2.5    K          0
## 6     6  11/15/1~ AL   TORNA~      0       6     2.5    K          0
## # ... with 1 more variable: CROPDMGEXP <chr>
```

The fields are: REFNUM - Reference Number Injuries - Injuries BGN_DATE - Beginning date PROPDMG - Property damage STATE - State PROPDMGEXP - Property damage exponent (e.g H - Hundred, K - Thousand) EVTYPE - Event type CROPDMG - Crop damage FATALITIES - Fatalities CROPDMG - Crop damage exponent (e.g M - Million, B - Billion)

Transformations were applied to the “BGN_DATE” field format and “PROPDMGEXP” and “CROPDMGEXP” fields. The abbreviations H, K, M and B (not case sensitive) in the fields “PROPDMGEXP” and “CROPDMGEXP” were duly replaced by 100, 1000, 1000000 and 1000000000 respectively.

```
## Transformations
## Data and time fields
stormdata$BGN_DATE <- as.Date(mdy_hms(stormdata$BGN_DATE))

## PROPDMGEXP and CROPDMGEXP Fields
```

```

for (i in c("H", "K", "M", "B", "h", "k", "m", "b")) {
  if( i == "H" | i == "h"){
    j = 100} else if ( i == "K" | i == "k"){
    j = 1000
  } else if ( i == "M" | i == "m"){
    j = 1000000
  } else if ( i == "B" | i == "b"){
    j = 1000000000
  } else {

  }

  stormdata$PROPDMGEXP <- gsub(i, j, stormdata$PROPDMGEXP)
  stormdata$CROPDMGEXP <- gsub(i, j, stormdata$CROPDMGEXP)
}

head(tibble(stormdata))

```

```

## # A tibble: 6 x 10
##   REFNUM BGN_DATE   STATE EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG
##   <dbl> <date>     <chr> <chr>      <dbl>    <dbl>    <dbl> <chr>      <dbl>
## 1     1 1950-04-18 AL    TORNA~         0        15     25    1000         0
## 2     2 1950-04-18 AL    TORNA~         0         0     2.5    1000         0
## 3     3 1951-02-20 AL    TORNA~         0         2     25    1000         0
## 4     4 1951-06-08 AL    TORNA~         0         2     2.5    1000         0
## 5     5 1951-11-15 AL    TORNA~         0         2     2.5    1000         0
## 6     6 1951-11-15 AL    TORNA~         0         6     2.5    1000         0
## # ... with 1 more variable: CROPDMGEXP <chr>

```

Other transformations were carried out on the data to make it easy to further transform and/or draw subsets

```

## Health and economic value considerations
stormdata$FATALITIES <- as.numeric(stormdata$FATALITIES)
stormdata$INJURIES <- as.numeric(stormdata$INJURIES)
stormdata$PROPDMG <- as.numeric(stormdata$PROPDMG)
stormdata$CROPDMG <- as.numeric(stormdata$CROPDMG)
stormdata$PROPDMGEXP <- as.numeric(stormdata$PROPDMGEXP)

```

```
## Warning: NAs introduced by coercion
```

```
stormdata$CROPDMGEXP <- as.numeric(stormdata$CROPDMGEXP)
```

```
## Warning: NAs introduced by coercion
```

A subset was drawn for the purpose of assessing the impact of the weather events on the population health across the United States. The data excluded fields from the previous subset “stormdata”. The excluded fields were “PROPDMG”, “PROPDMGEXP”, “CROPDMG”, “CROPDMGEXP”. An additional field labeled “HEALTHVALUE” (health value is equals to the sum of number of fatalities and injured) was added. All rows of data where the “HEALTHVALUE” field was set to NA is removed and same is applicable to fields where the “HEALTHVALUE” is either equal to zero (0) or less.

```
## Sub setting health data
healthdata <- select(stormdata, -PROPDGMG, -PROPDGMGEXP, -CROPDMG, -CROPDMGEXP)
healthdata <- mutate(healthdata, HEALTHVALUE = FATALITIES + INJURIES)
healthdata$HEALTHVALUE <- as.numeric(healthdata$HEALTHVALUE)
healthdata <- filter(healthdata, is.na(HEALTHVALUE) == FALSE, HEALTHVALUE > 0)
head(tibble(healthdata))
```

```
## # A tibble: 6 x 7
##   REFNUM BGN_DATE   STATE EVTYPE  FATALITIES INJURIES HEALTHVALUE
##   <dbl> <date>     <chr> <chr>      <dbl>      <dbl>      <dbl>
## 1     1 1950-04-18 AL    TORNADO         0         15         15
## 2     3 1951-02-20 AL    TORNADO         0          2          2
## 3     4 1951-06-08 AL    TORNADO         0          2          2
## 4     5 1951-11-15 AL    TORNADO         0          2          2
## 5     6 1951-11-15 AL    TORNADO         0          6          6
## 6     7 1951-11-16 AL    TORNADO         0          1          1
```

A subset was also drawn for the purpose of assessing the economic impact of the weather events across the United States. The data excluded fields from the previous subset “stormdata”. The excluded fields were “FATALITIES” and “INJURIES”. An additional field labeled “ECONOMICVALUE” (which is the sum of the damage to property and crops multiplied by their exponent) was added. All rows of data where the “ECONOMICVALUE” field was set to NA is removed and same is applicable to fields where the “ECONOMICVALUE” is either equal to zero (0) or less.

```
## Sub setting economic data
economicdata <- select(stormdata, -FATALITIES, -INJURIES)
economicdata <- mutate(economicdata, ECONOMICVALUE = PROPDMG*PROPDGMGEXP + CROPDMG*CROPDMGEXP)
economicdata$ECONOMICVALUE <- as.numeric(economicdata$ECONOMICVALUE)
economicdata <- filter(economicdata, is.na(ECONOMICVALUE) == FALSE, ECONOMICVALUE > 0)
head(tibble(economicdata))
```

```
## # A tibble: 6 x 9
##   REFNUM BGN_DATE   STATE EVTYPE PROPDMG PROPDGMGEXP CROPDMG CROPDMGEXP
##   <dbl> <date>     <chr> <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 187566 1995-10-04 AL    HURRI~    0.1 1000000000      10 1000000
## 2 188271 1994-06-26 AL    THUND~     5 1000000      500 1000
## 3 187568 1995-08-03 AL    HURRI~    25 1000000      1 1000000
## 4 187570 1995-10-03 AL    HURRI~    48 1000000      4 1000000
## 5 187571 1995-10-04 AL    HURRI~    20 1000000     10 1000000
## 6 187640 1994-03-24 AL    THUND~    50 1000      50 1000
## # ... with 1 more variable: ECONOMICVALUE <dbl>
```

Data Analysis and Results

Population health Impact In order to the determine the weather event with the most adverse impact on the population health, the health data set is arranged in descending order of HEALTHVALUE. Where there is a tie it is resolved by arranging the event type is ascending order (e.g A, B, C, ...Z). The table is subsequently grouped according to the event types and and summed across the impact value field. The weather events whose health value is above the 90% quantile are selected as the events with the most population health impacts. See table below:

```
## Impact Analysis
healthimpact <- healthdata %>% group_by(EVTYPE) %>% summarise(IMPACT = sum(HEALTHVALUE)) %>% arrange(desc(IMPACT))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

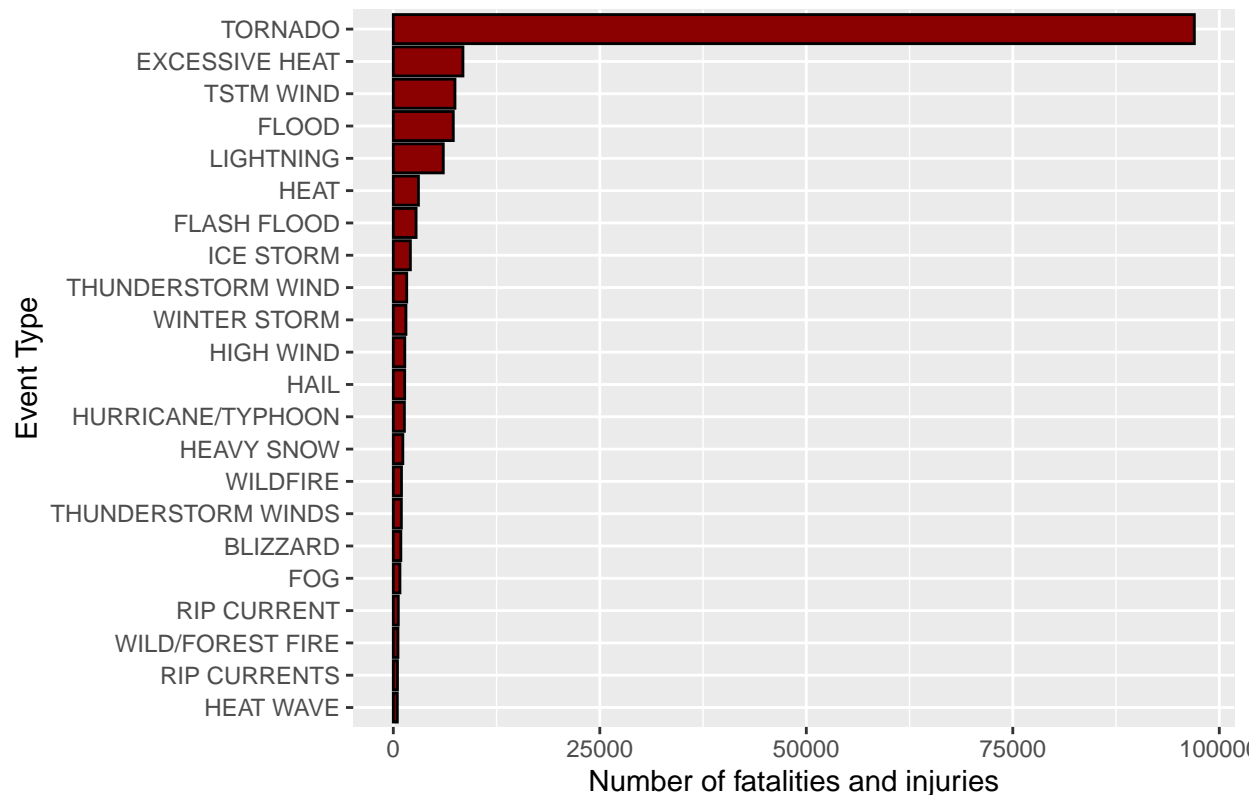
```
p <- quantile(healthimpact$IMPACT, probs = 0.9)[[1]]
healthimpact <- healthimpact[healthimpact$IMPACT > p, ]
tibble(healthimpact)
```

```
## # A tibble: 22 x 2
##   EVTYPE          IMPACT
##   <chr>          <dbl>
## 1 TORNADO        96979
## 2 EXCESSIVE HEAT  8428
## 3 TSTM WIND      7461
## 4 FLOOD         7259
## 5 LIGHTNING      6046
## 6 HEAT          3037
## 7 FLASH FLOOD    2755
## 8 ICE STORM      2064
## 9 THUNDERSTORM WIND 1621
## 10 WINTER STORM   1527
## # ... with 12 more rows
```

See plot below:

```
## plotting impact on population field
g <- ggplot(data = healthimpact, aes(x = reorder(EVTYPE,IMPACT), y = IMPACT))
g <- g + geom_bar(stat = "identity", colour = "black", fill = "darkred")
g <- g + labs(title = "Top 10% of events with the worst population health impact:1950-2011")
g <- g + theme(plot.title = element_text(hjust = 0.5))
g <- g + labs(y = "Number of fatalities and injuries", x = "Event Type")
g <- g + coord_flip()
print(g)
```

Top 10% of events with the worst population health impact:1950–



Economic impact In order to determine the weather event with the most adverse economic impact, the economic data set is arranged in descending order of ECONOMICVALUE. Where there is a tie it is resolved by arranging the event types in ascending order (e.g. A, B, C, ...Z). The table is subsequently grouped according to the event types and summed across the impact value field. The weather events whose health value is above the 90% quantile are selected as the events with the most severe economic impact. See table below:

```
economicimpact <- economicdata %>% group_by(EVTYPE) %>% summarise(IMPACT = sum(ECONOMICVALUE)) %>% arrange(desc(IMPACT))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

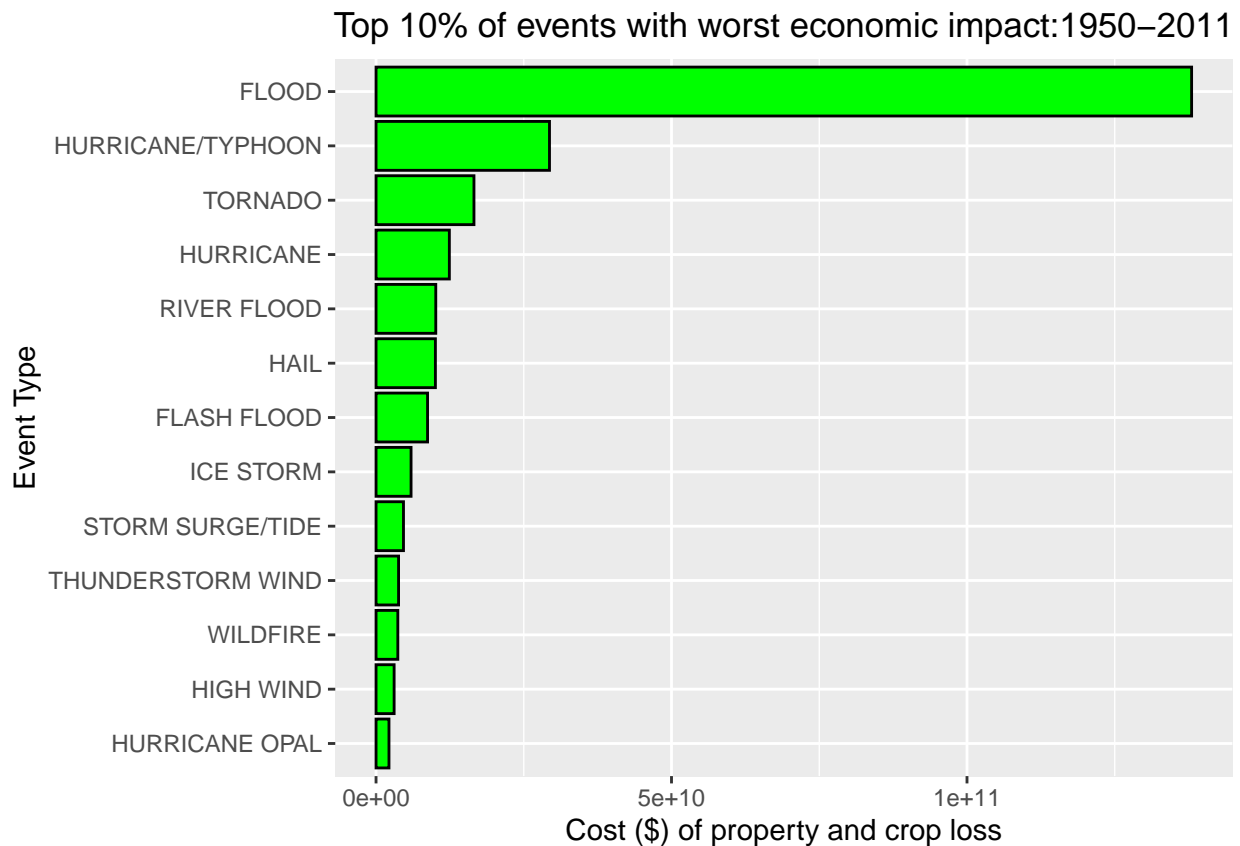
```
p <- quantile(economicimpact$IMPACT, probs = 0.9)[[1]]
economicimpact <- economicimpact[economicimpact$IMPACT > p, ]
tibble(economicimpact)
```

```
## # A tibble: 13 x 2
##   EVTYPE          IMPACT
##   <chr>          <dbl>
## 1 FLOOD          138007444500
## 2 HURRICANE/TYPHOON 29348167800
## 3 TORNADO         16570326150
## 4 HURRICANE        12405268000
## 5 RIVER FLOOD      10108369000
## 6 HAIL            10045596740
```

```
## 7 FLASH FLOOD      8715885162
## 8 ICE STORM        5925150800
## 9 STORM SURGE/TIDE 4641493000
## 10 THUNDERSTORM WIND 3813647990
## 11 WILDFIRE        3684468370
## 12 HIGH WIND       3057666640
## 13 HURRICANE OPAL  2187000000
```

See plot below:

```
#plotting economic loss
g <- ggplot(data = economicimpact, aes(x = reorder(EVTYPE, IMPACT), y = IMPACT))
g <- g + geom_bar(stat = "identity", colour = "black", fill = "green")
g <- g + labs(title = "Top 10% of events with worst economic impact:1950-2011")
g <- g + theme(plot.title = element_text(hjust = 0.5))
g <- g + labs(y = "Cost ($)" of property and crop loss", x = "Event Type")
g <- g + coord_flip()
print(g)
```



Conclusion

There is, thus, 22 storm weather events with the worst population health impact and 13 storm events with the worst economic impact over the period 1950 to 2011.