RR-AnalysisOnWeatherBasedOnNOAA

Ts. Nurul Haszeli Ahmad

9/12/2020

```
knitr::opts_chunk$set(echo = TRUE)
```

Impact of Severe Weather Events on Health and Economy in United States

Introduction

This report presented the analysis on data collected by U.S. National Oceanic and Atmospheric Administration's (NOAA). The main focus of this report is on storm data set collected between the year 1950 until the end of November 2011.

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This analysis involves exploring the **U.S. National Oceanic and Atmospheric Administration's (NOAA)** storm database starting from the year **1950 until the end of November 2011**. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. The result of the analysis presented in this document can be used as a supporting evidence for various ministries and agencies within the government and municipal who might be responsible for preparing for severe weather events and will need to prioritize resources for different types of events.

Based on the analysis; it is concluded that the severe weather events in united states highly impacting the health and economy. In terms of health, based on the number of injuries and fatalities, it is shown that tornadoes has the highest impact of all weather events. Whereas, flood causes major impact to economy due to the damages affecting properties and plantation.

Note:
1. This report was made as an assignment for the Reproducible Research Coursera course.
2. The report is made with RMarkdown and Knitr through tool Rstudio.
3. The complete code used to produce this analysis is available at github

Data

The data used in this analysis is obtained from **U.S. National Oceanic and Atmospheric Administration's (NOAA)** storm database. * Storm Data - Data set from the year 1950 until end of November 2011. * Data Description - Storm data documentation by National weather services

Library Used

The following library is used in the analysis.
1. dplyr
2. ggplot2 3. gridExtra

Data Processing

The section below describe the data processing processes. It starts with laading the necessary library and access the link where the data exist. It is then followed with decompressing the file and load data into memory or processing

Data preprocessing

1. Loading Library

```
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(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(grid)
```

2. Retrieve the file The initial steps to analyse the data is to download the required data set from the given location as listed above; section **Data**.

```
setwd(getwd())
getwd
link <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url = link, destfile = "dataset")
```

Note: *The link may change depending on the course provider and update from the data provider*

3. Data loading Upon the data was loaded into the local drive, it will loaded into the memory using *read.csv* function The *bzfile* function is used if the file downloaded is still in compressed format (bz2 extension). If the data set is already loaded, it will use cache object instead of loading it each time the Rmd file is knitted.

```
if (!exists("dataset")) {
    # Extract file if it is not already extracted
    if (!file.exists("repdata_data_StormData.csv.bz2")) {
        link <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
        download.file(url = link, destfile = "dataset")
    }

    if (!file.exists("repdata_data_StormData.csv")) {
        bunzip2("repdata_data_StormData.csv.bz2", overwrite = F)
    }

    # Read data into the variable called dataset
    dataset <- read.csv("repdata_data_StormData.csv", sep = ",", header=TRUE)
}
```

4. Take a look on the dataset

```
# view the first 3 rows of the data
head(dataset, n=3)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1       1 4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2       1 4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3       1 2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0                      14.0   100 3   0          0       15    25.0
## 2         0                       2.0   150 2   0          0        0     2.5
## 3         0                       0.1   123 2   0          0        2    25.0
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
```

```
# view the summary of the data
summary(dataset)
```

```
##     STATE__       BGN_DATE           BGN_TIME          TIME_ZONE
##  Min.   : 1.0   Length:902297      Length:902297      Length:902297
##  1st Qu.:19.0   Class :character   Class :character   Class :character
```

```
##  Median :30.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :31.2
##  3rd Qu.:45.0
##  Max.   :95.0
##
##      COUNTY       COUNTYNAME           STATE              EVTYPE
##  Min.   :  0.0   Length:902297      Length:902297      Length:902297
##  1st Qu.: 31.0   Class :character   Class :character   Class :character
##  Median : 75.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :100.6
##  3rd Qu.:131.0
##  Max.   :873.0
##
##    BGN_RANGE          BGN_AZI           BGN_LOCATI          END_DATE
##  Min.   :   0.000   Length:902297      Length:902297      Length:902297
##  1st Qu.:   0.000   Class :character   Class :character   Class :character
##  Median :   0.000   Mode  :character   Mode  :character   Mode  :character
##  Mean   :   1.484
##  3rd Qu.:   1.000
##  Max.   :3749.000
##
##    END_TIME           COUNTY_END COUNTYENDN       END_RANGE
##  Length:902297      Min.   :0    Mode:logical   Min.   :  0.0000
##  Class :character   1st Qu.:0    NA's:902297    1st Qu.:  0.0000
##  Mode  :character   Median :0                   Median :  0.0000
##                     Mean   :0                   Mean   :  0.9862
##                     3rd Qu.:0                   3rd Qu.:  0.0000
##                     Max.   :0                   Max.   :925.0000
##
##    END_AZI           END_LOCATI            LENGTH              WIDTH
##  Length:902297      Length:902297      Min.   :   0.0000   Min.   :   0.000
##  Class :character   Class :character   1st Qu.:   0.0000   1st Qu.:   0.000
##  Mode  :character   Mode  :character   Median :   0.0000   Median :   0.000
##                                        Mean   :   0.2301   Mean   :   7.503
##                                        3rd Qu.:   0.0000   3rd Qu.:   0.000
##                                        Max.   :2315.0000   Max.   :4400.000
##
##        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   Length:902297      Min.   :  0.000   Length:902297
##  1st Qu.:   0.00   Class :character   1st Qu.:  0.000   Class :character
##  Median :   0.00   Mode  :character   Median :  0.000   Mode  :character
##  Mean   :  12.06                      Mean   :  1.527
##  3rd Qu.:   0.50                      3rd Qu.:  0.000
##  Max.   :5000.00                      Max.   :990.000
##
##      WFO             STATEOFFIC         ZONENAMES            LATITUDE
```

```
##  Length:902297      Length:902297      Length:902297      Min.   :   0
##  Class :character   Class :character   Class :character   1st Qu.:2802
##  Mode  :character   Mode  :character   Mode  :character   Median :3540
##                                                           Mean   :2875
##                                                           3rd Qu.:4019
##                                                           Max.   :9706
##                                                           NA's   :47
##    LONGITUDE        LATITUDE_E     LONGITUDE_       REMARKS
##  Min.   :-14451   Min.   :   0   Min.   :-14455   Length:902297
##  1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0   Class :character
##  Median :  8707   Median :   0   Median :     0   Mode  :character
##  Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   : 17124   Max.   :9706   Max.   :106220
##                   NA's   :40
##      REFNUM
##  Min.   :     1
##  1st Qu.:225575
##  Median :451149
##  Mean   :451149
##  3rd Qu.:676723
##  Max.   :902297
##
```

```
# view the data structure. It shall has 37 columns (variables) and 902,297 rows (records).
dim(dataset)
```

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

```
# view the structure/characteristics of each column
str(dataset)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ 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: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Data analysis

Pre-analysis on the data based on the requirements The data set contains 902297 events and 37 variables. To find the answers we seek; that are the health and economy impact of the severe weather event, the parameters (or variables) that we are interested are:

1. EVTYPE

- Event type described the event happen. Below are some total events in the data set and some example of the event

```
# total unique event
length(unique(dataset$EVTYPE))
```

```
## [1] 985
```

```
# some of the events
head(unique(dataset$EVTYPE),n=5)
```

```
## [1] "TORNADO"       "TSTM WIND"     "HAIL"          "FREEZING RAIN"
## [5] "SNOW"
```

2. Health data

- Counter for the health impact is in the column **FATALITIES** and **INJURIES**

3. Monetary impact

- The monetary impact on crop and property is measured from column **PROPDMG** and **CROPDMG**

4. Other data

- Each orresponding exponents; **PROPDMGEXP** and **CROPDMGEXP**

Processing the data After we have list all the above parameters/variables that are required to perform analysis in order to find the impact on health and economy of the United States, we separate the data into two sub sets; health and economy data set.

1. Health data

```r
# getting fatalities data
healthDataset_fatal <- dataset %>% select(EVTYPE, FATALITIES) %>% group_by(EVTYPE) %>% summarise(total.
head(healthDataset_fatal, n=5)
```

```
## # A tibble: 5 x 2
##   EVTYPE         total.fatalities
##   <chr>                     <dbl>
## 1 TORNADO                    5633
## 2 EXCESSIVE HEAT             1903
## 3 FLASH FLOOD                 978
## 4 HEAT                        937
## 5 LIGHTNING                   816
```

```r
summary(healthDataset_fatal)
```

```
##     EVTYPE          total.fatalities
##  Length:985         Min.   :   0.00
##  Class :character   1st Qu.:   0.00
##  Mode  :character   Median :   0.00
##                     Mean   :  15.38
##                     3rd Qu.:   0.00
##                     Max.   :5633.00
```

```r
# getting injuries data
healthDataset_injury <- dataset %>% select(EVTYPE, INJURIES) %>% group_by(EVTYPE) %>% summarise(total.i
head(healthDataset_injury, n=5)
```

```
## # A tibble: 5 x 2
##   EVTYPE         total.injuries
##   <chr>                   <dbl>
## 1 TORNADO                 91346
## 2 TSTM WIND                6957
## 3 FLOOD                    6789
## 4 EXCESSIVE HEAT           6525
## 5 LIGHTNING                5230
```

```r
summary(healthDataset_injury)
```

```
##     EVTYPE          total.injuries
##  Length:985         Min.   :    0.0
##  Class :character   1st Qu.:    0.0
##  Mode  :character   Median :    0.0
##                     Mean   :  142.7
##                     3rd Qu.:    0.0
##                     Max.   :91346.0
```

2. Economy data

Economic data is provided in the column **PROPDMG** and **CROPDMG** and it is translated in value of USD ($) by **PROPDMGEXP** and **CROPDMGEXP** parameters. The index in the PROPDMGEXP and CROPDMGEXP can be interpreted as the following H, h -> hundreds = x100
K, K -> kilos = x1,000

M, m -> millions = x1,000,000
B,b -> billions = x1,000,000,000
(+) -> x1
(-) -> x0
(?) -> x0
blank -> x0
Note: *as described via the following link*

```
economicDataset <- dataset %>% select(EVTYPE, PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)

Symbol <- sort(unique(as.character(economicDataset$PROPDMGEXP)))
Multiplier <- c(0,0,0,1,10,10,10,10,10,10,10,10,10,10^9,10^2,10^2,10^3,10^6,10^6)
convert.Multiplier <- data.frame(Symbol, Multiplier)

economicDataset$Prop.Multiplier <- convert.Multiplier$Multiplier[match(economicDataset$PROPDMGEXP, conv
economicDataset$Crop.Multiplier <- convert.Multiplier$Multiplier[match(economicDataset$CROPDMGEXP, conv

economicDataset <- economicDataset %>% mutate(PROPDMG = PROPDMG*Prop.Multiplier) %>% mutate(CROPDMG = C

economicDataset.total <- economicDataset %>% group_by(EVTYPE) %>% summarize(TOTAL.DMG.EVTYPE = sum(TOTA

head(economicDataset.total,n=5)
```

```
## # A tibble: 5 x 2
##   EVTYPE            TOTAL.DMG.EVTYPE
##   <chr>                        <dbl>
## 1 FLOOD                150319678250
## 2 HURRICANE/TYPHOON     71913712800
## 3 TORNADO               57352117607
## 4 STORM SURGE           43323541000
## 5 FLASH FLOOD           17562132111
```

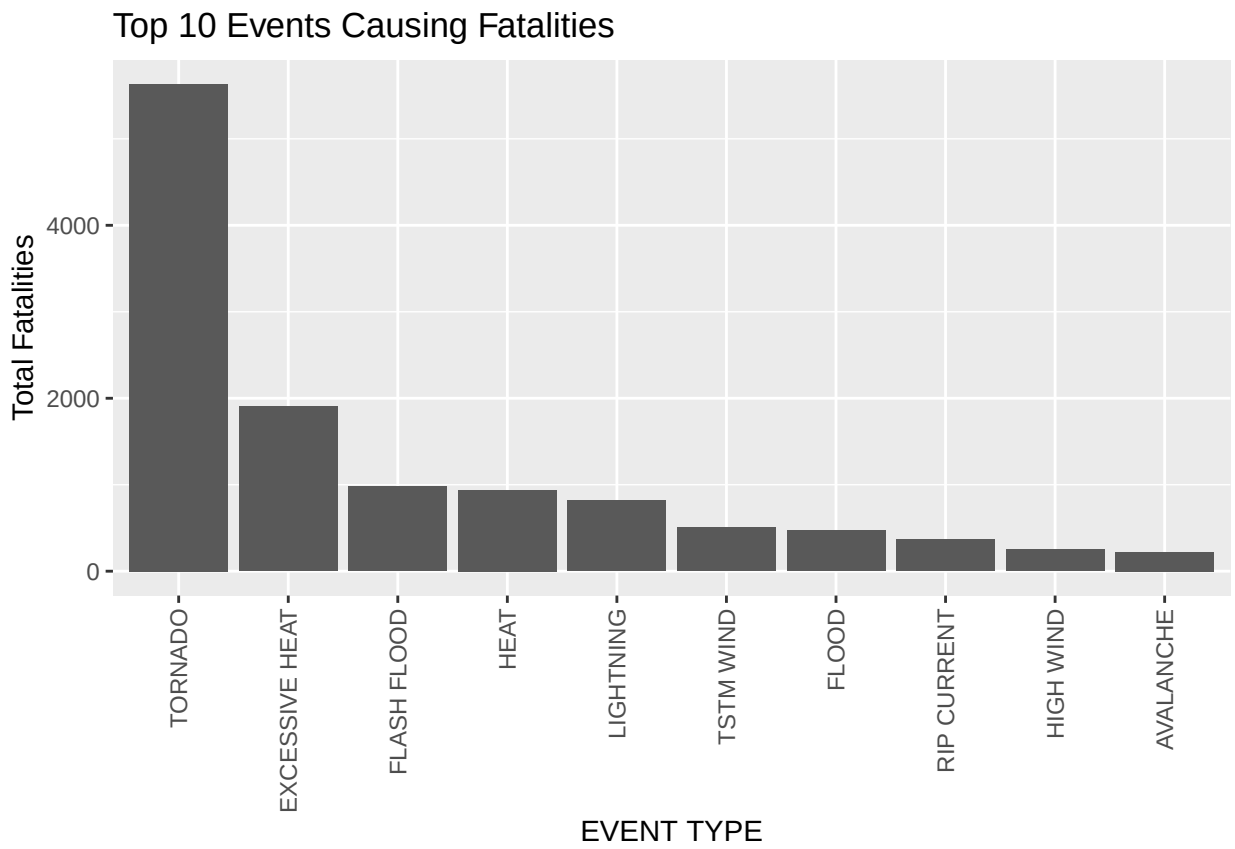
```
summary(economicDataset.total)
```

```
##     EVTYPE          TOTAL.DMG.EVTYPE
##  Length:985         Min.   :0.000e+00
##  Class :character   1st Qu.:0.000e+00
##  Mode  :character   Median :0.000e+00
##                     Mean   :4.642e+08
##                     3rd Qu.:7.500e+04
##                     Max.   :1.503e+11
##                     NA's   :4
```

Result

The following is the plot of the data processed above and result of the analysis
Health Impact Analysis #### Fatalities Impact
The following is the plot for fatalities data by Event type.

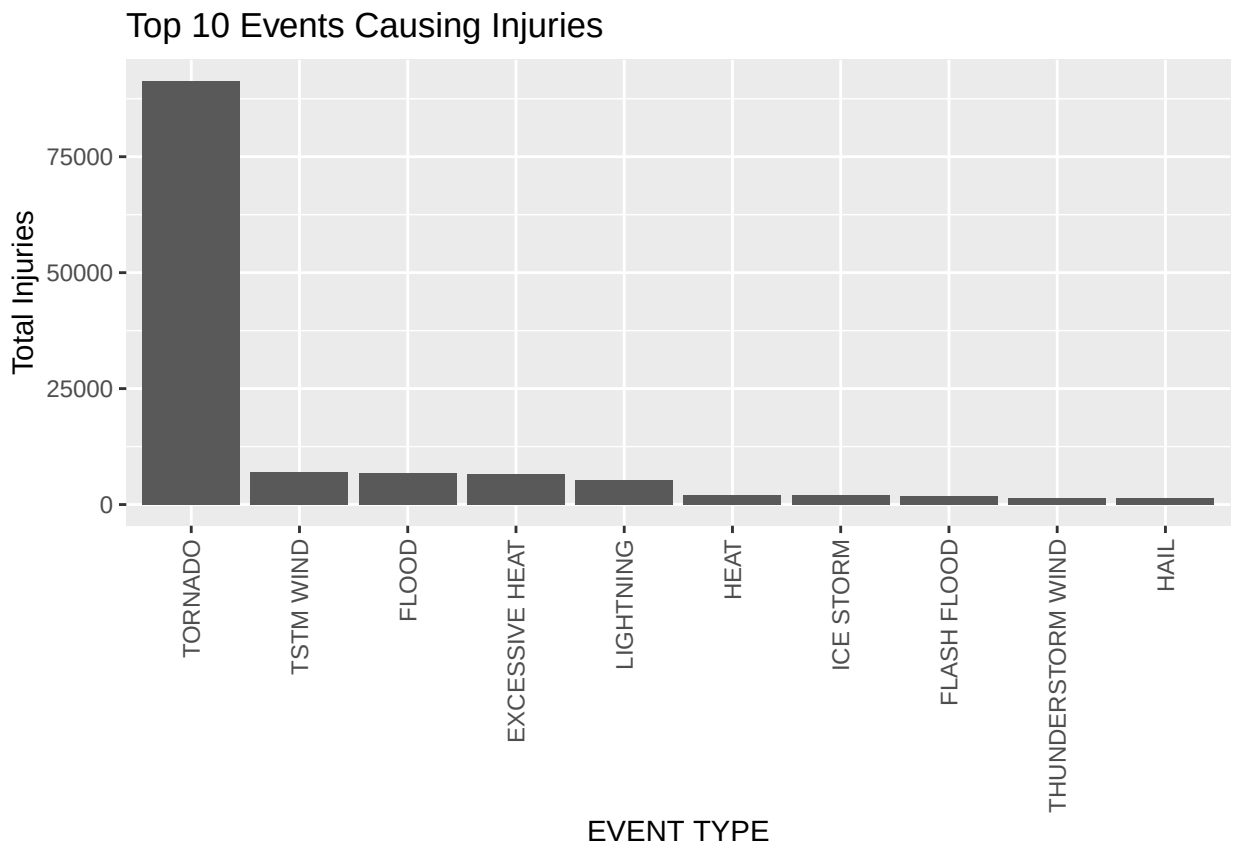
```
fatal <- ggplot(healthDataset_fatal[1:10,], aes(x=reorder(EVTYPE, -total.fatalities), y=total.fatalitie
fatal
```



Result: From the bar chart plotted above, it is shown that **tornado** contributes to the highest fatality followed by **excessive heat** and **flash flood**. The gap between tornado and the other weather events is huge, of which a clear indication that tornado is extremely dangerous compared to other.

Injuries Impact The following is the plot for injuries data by Event type.

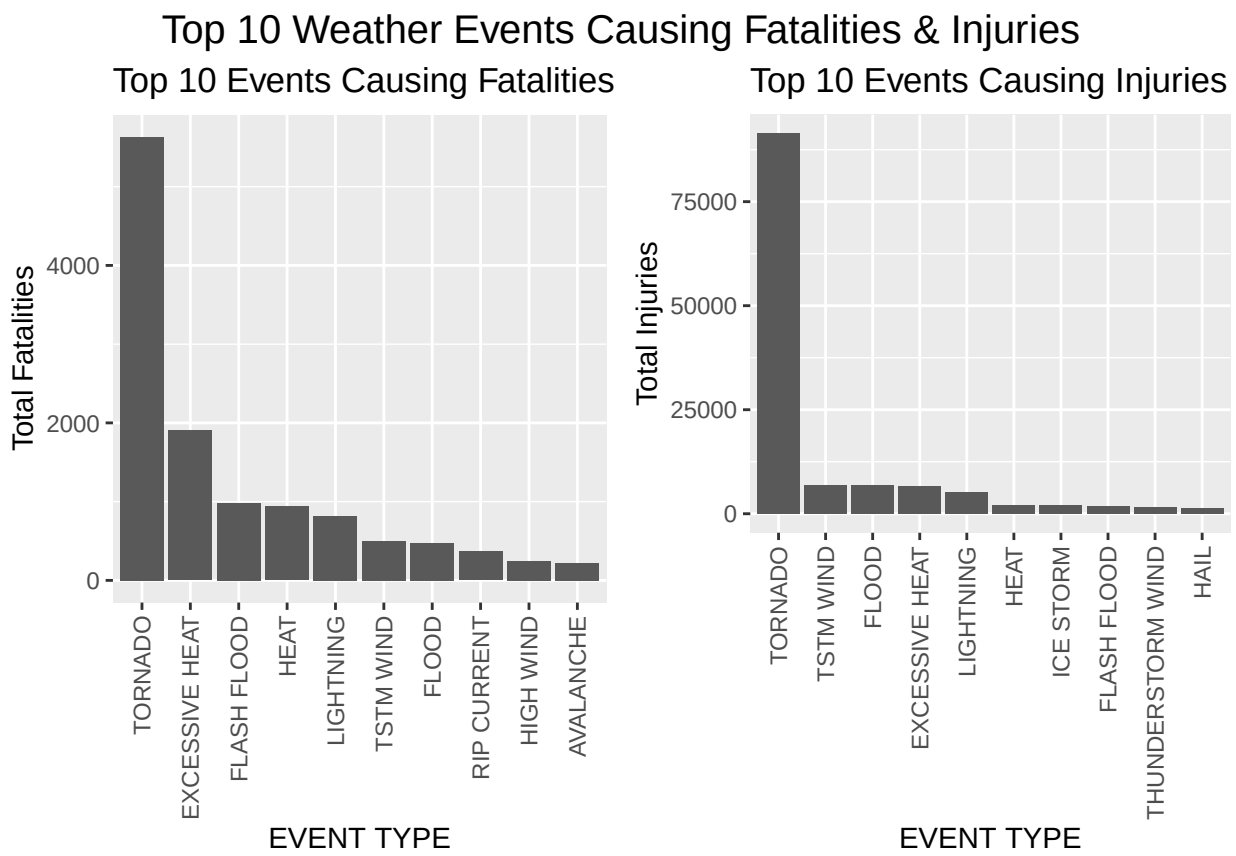
```
injury <- ggplot(healthDataset_injury[1:10,], aes(x=reorder(EVTYPE, -total.injuries), y=total.injuries)
injury
```



Result: In terms of injuries, from the bar chart plotted above, it is shown that **tornado** is still the highest contributor for the injuries due to weather events followed by **tstm wind** and **flood**. Similar to fatalities, tornado is also causing the most injuries to US citizen.

Conclusion for Health Impact

```
title <- textGrob('Top 10 Weather Events Causing Fatalities & Injuries', gp = gpar(fontsize = 15, font =
grid.arrange(fatal, injury, nrow = 1, newpage = F,top = title)
```

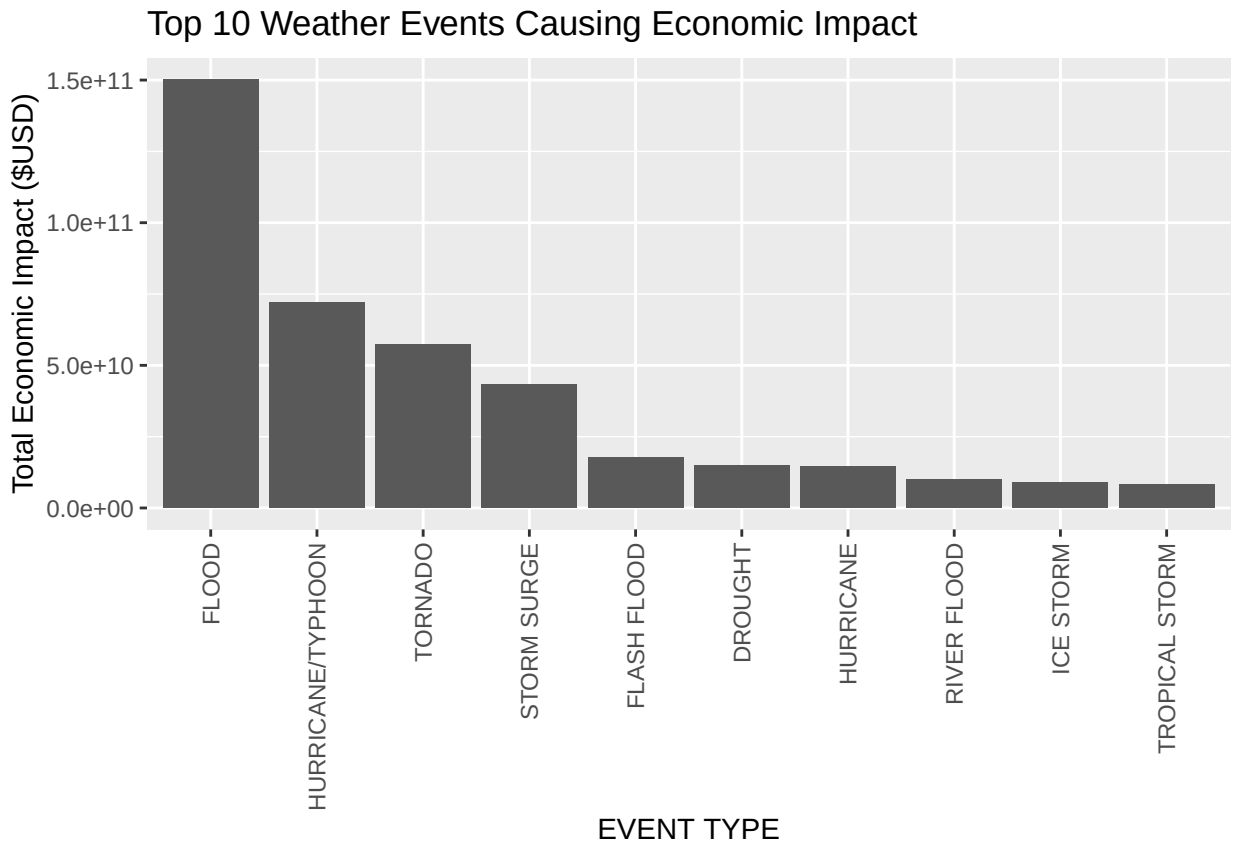


As shown in the plot above, there is no doubt that *tornado* has the highest impact to health based on the data analyse.

Economic Impact Analysis

The following bar plot shows the economic impact due to the weather events.

```
ggplot(economicDataset.total[1:10,], aes(x=reorder(EVTYPE, -TOTAL.DMG.EVTYPE), y=TOTAL.DMG.EVTYPE),fill
```



Conclusion for Economic Impact Economic impact on the other end, is mostly affected due to flood as compared to health impact.

Conclusion

From the analysis shown, it is concluded that event related to wind, mainly tornado is causing fatal and injuries to US citizen, whereas flood has the biggest impact to economy.