NOAA STORM DATA ANALYSIS

Peer Graded Assignment: Course Project 2

Sharanya Gowda

July 25, 2022

---

1. SYNOPSIS

In response to the requirements of Peer Graded Assignment: Course Project 2, I present the following **analysis on severe weather events** to the government. The detailed description and summary of **NOAA Storm Data Analysis** is provided from *Sections 1.1* to *Section 1.4*.

1.1. Background

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 project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. 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.

1.2. Objective

The *goal* of this analysis is to explore the NOAA Storm Database and answer the following questions:

- Across the United States, which types of events (as indicated in the `EVTYPE` variable) are most harmful with respect to population health?
- Across the United States, which types of events have the greatest economic consequences?

1.3. Government Furnished Information (GFI)

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site Storm Data

There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

- National Weather Service Storm Data Documentation
- National Climatic Data Center Storm Events FAQ

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

1.4. Tech Stack

The following *tools* were utilized for this assignment:

- **RStudio** - Used to edit/write the analysis and publish the completed analysis document to RPubs.
- **knitr** - knitr package used to compile the R Markdown document and convert it to HTML.

---

2. DATA PROCESSING

There should be a section titled Data Processing which describes (in words and code) how the data were loaded into R and processed for analysis. In particular, your analysis must start from the raw CSV file containing the data. You cannot do any pre-processing outside the document. If pre-processing is time-consuming you may consider using the `cache = TRUE` option for certain code chunks.

2.1. Load Packages

Load relevant packages required for the analysis.

```
library(data.table)
library(dplyr)
```

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

## The following objects are masked from 'package:data.table':
##
##     between, first, last

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

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

```
library(lubridate)
```

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

## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year

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

```r
library(tidyr)
```

2.2. Data File Download

Download data file from the website using the link Storm Data

```r
if (!file.exists("StormData.csv.bz2")) {
    fileURL <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
    download.file(fileURL, destfile='StormData.csv.bz2', method = 'curl')
}
```

2.3. Preliminary Data Analysis

Read downloaded data table.

```r
Read_Storm_Data <- read.csv(bzfile('StormData.csv.bz2'),header=TRUE, stringsAsFactors = FALSE)
```

Analyze the preliminary data.

```r
#summary(Read_Storm_Data)
head(Read_Storm_Data, 5)
```

```
##   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
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    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
## 4         0                                               0         NA
## 5         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
## 4         0                       0.0   100 2   0          0        2     2.5
## 5         0                       0.0   150 2   0          0        2     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
```

---

3. HARMFUL EVENTS

To understand which types of events are most harmful to population health, both, **injuries** and **fatalities** caused by various weather events must be analyzed.

3.1. Injuries cause by Weather Events

Calculate total number of Injuries caused by individual Weather Events.

```r
Storm_Injuries <- aggregate(Read_Storm_Data$INJURIES, by = list(Read_Storm_Data$EVTYPE), "sum")
names(Storm_Injuries) <- c("WeatherEvent", "Injuries")
Storm_Injuries_Display<- Storm_Injuries[order(-Storm_Injuries$Injuries), ][1:15, ]
Storm_Injuries_Display
```

```
##          WeatherEvent 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
## 972      WINTER STORM     1321
## 411 HURRICANE/TYPHOON     1275
## 359         HIGH WIND     1137
## 310        HEAVY SNOW     1021
## 957          WILDFIRE      911
```

3.2. Fatalities cause by Weather Events

Calculate total number of Fatalities caused by individual Weather Events.

```r
Storm_Fatalities <- aggregate(Read_Storm_Data$FATALITIES, by = list(Read_Storm_Data$EVTYPE), "sum")
names(Storm_Fatalities) <- c("WeatherEvent", "Fatalities")
Storm_Fatalities_Display <- Storm_Fatalities[order(-Storm_Fatalities$Fatalities), ][1:15, ]
Storm_Fatalities_Display
```

```
##          WeatherEvent 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
## 972      WINTER STORM        206
## 586      RIP CURRENTS        204
## 278         HEAT WAVE        172
## 140      EXTREME COLD        160
## 760 THUNDERSTORM WIND        133
```

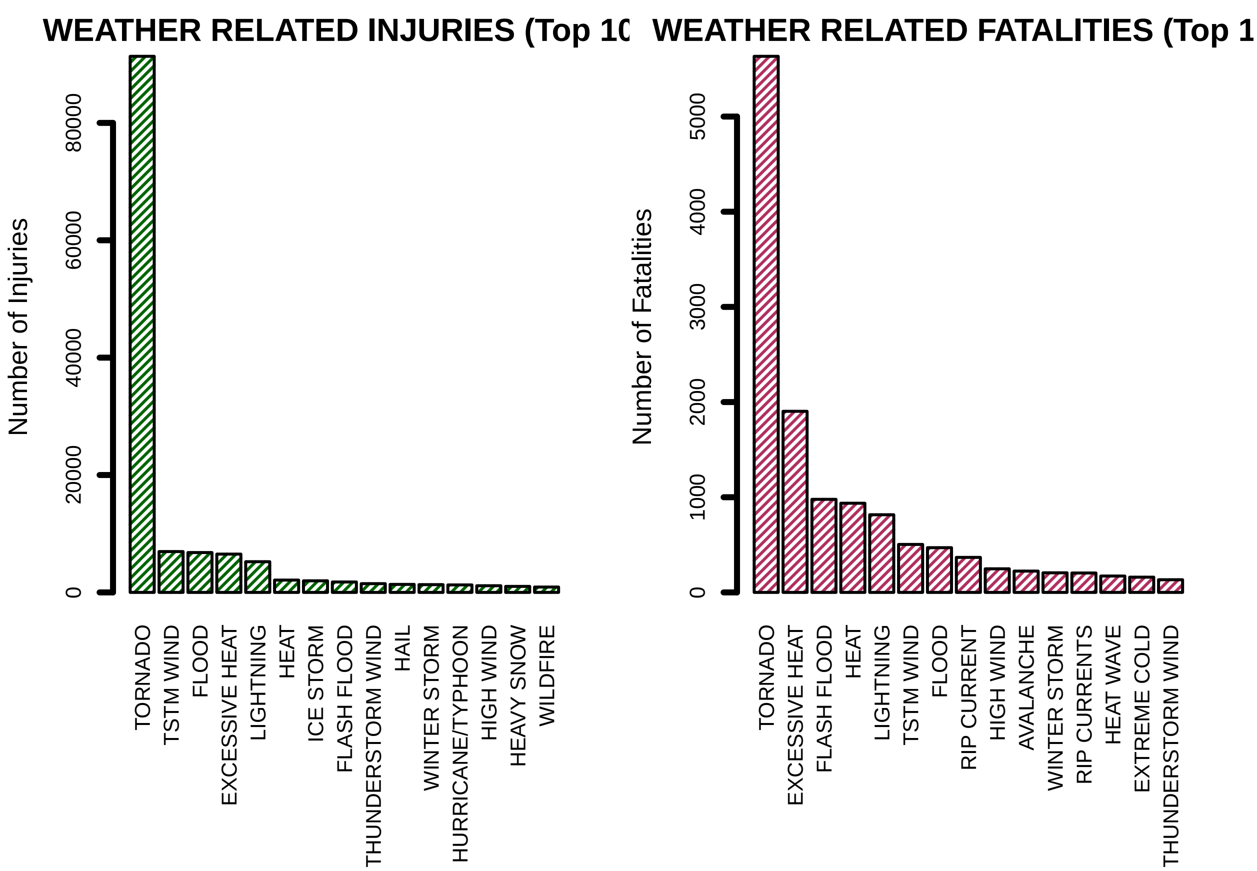
3.3. Injury and Fatality Plot

Plot Injuries and Fatalities caused by Weather Events to analyze which types of events are most harmful with respect to population health, across the United States.

```
par(mfrow = c(1, 2), mar = c(10, 4, 2, 2), las = 3, cex = 0.7, cex.main = 1.4, cex.lab = 1.2, lwd=1.5)

barplot(Storm_Injuries_Display$Injuries, names.arg = Storm_Injuries_Display$WeatherEvent, col="darkgree

barplot(Storm_Fatalities_Display$Fatalities, names.arg = Storm_Fatalities_Display$WeatherEvent, col="ma
```



****

4. WEATHER EVENTS CAUSING GREATEST ECONOMIC CRISIS

To understand which types of weather events have the greatest economic consequences, both, **crop damages** and **cost of property** must be analyzed.

4.1. Weather Events causing Crop Damages

Calculate weather events causing crop damages.

```
ec_Crop <- aggregate(Read_Storm_Data$CROPDMG, by = list(Read_Storm_Data$EVTYPE), "sum")
names(ec_Crop) <- c("WeatherEvent", "Crop")
ec_Crop_Display <- ec_Crop[order(-ec_Crop$Crop), ][1:15, ]
ec_Crop_Display
```

```
##           WeatherEvent      Crop
## 244               HAIL 579596.28
## 153        FLASH FLOOD 179200.46
## 170              FLOOD 168037.88
## 856          TSTM WIND 109202.60
## 834            TORNADO 100018.52
## 760  THUNDERSTORM WIND  66791.45
## 95             DROUGHT  33898.62
## 786 THUNDERSTORM WINDS  18684.93
## 359          HIGH WIND  17283.21
## 290         HEAVY RAIN  11122.80
## 212       FROST/FREEZE   7034.14
## 140       EXTREME COLD   6121.14
## 848     TROPICAL STORM   5899.12
## 402          HURRICANE   5339.31
## 164     FLASH FLOODING   5126.05
```

4.2. Weather Events causing Property Damage

Calculate weather events causing property damage.

```
ec_Property <- aggregate(Read_Storm_Data$PROPDMG, by = list(Read_Storm_Data$EVTYPE), "sum")
names(ec_Property) <- c("WeatherEvent", "Property")
ec_Property_Display <- ec_Property[order(-ec_Property$Property), ][1:15, ]
ec_Property_Display
```

```
##           WeatherEvent   Property
## 834            TORNADO 3212258.16
## 153        FLASH FLOOD 1420124.59
## 856          TSTM WIND 1335965.61
## 170              FLOOD  899938.48
## 760  THUNDERSTORM WIND  876844.17
## 244               HAIL  688693.38
## 464          LIGHTNING  603351.78
## 786 THUNDERSTORM WINDS  446293.18
## 359          HIGH WIND  324731.56
## 972       WINTER STORM  132720.59
## 310         HEAVY SNOW  122251.99
## 957           WILDFIRE   84459.34
## 427          ICE STORM   66000.67
## 676        STRONG WIND   62993.81
## 376         HIGH WINDS   55625.00
```

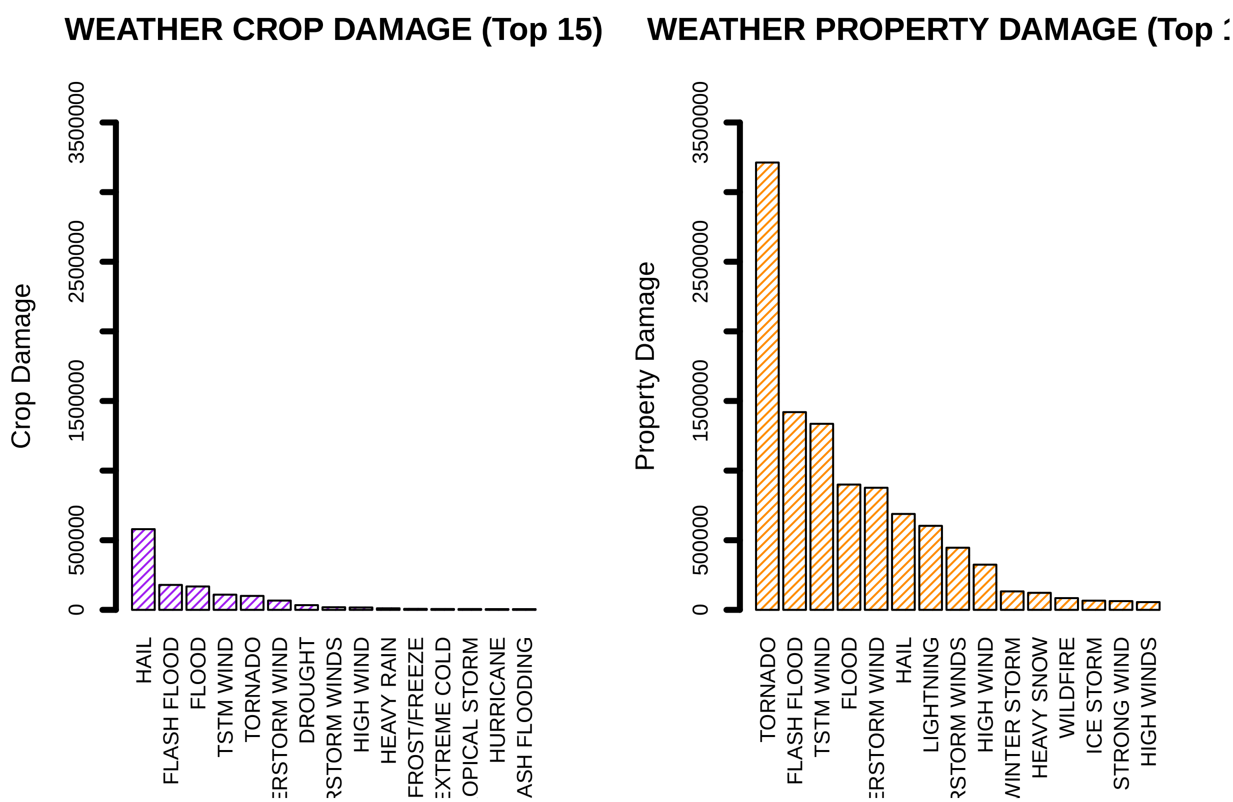
4.3. Crop Damages and Cost of Property Plot

Calculate total economic consequences caused by cost of property.

```
par(mfrow = c(1, 2), mar = c(7, 5, 7, 2), las = 3, cex = 0.7, cex.main = 1.4, cex.lab = 1.2)

barplot(ec_Crop_Display$Crop, names.arg = ec_Crop_Display$WeatherEvent, col="purple",density=30,lwd=3,ma

barplot(ec_Property_Display$Property, names.arg = ec_Property_Display$WeatherEvent, col="darkorange",de
```



4.4. Total Damage

Total Damage = Crop Damage + Property Damage

```
Total_Damage <- aggregate(Read_Storm_Data$CROPDMG+Read_Storm_Data$PROPDMG, by = list(Read_Storm_Data$EV
names(Total_Damage) <- c("WeatherEvent", "TotalDamage")
Total_Damage_Display <- Total_Damage[order(-Total_Damage$TotalDamage), ][1:15, ]
Total_Damage_Display
```

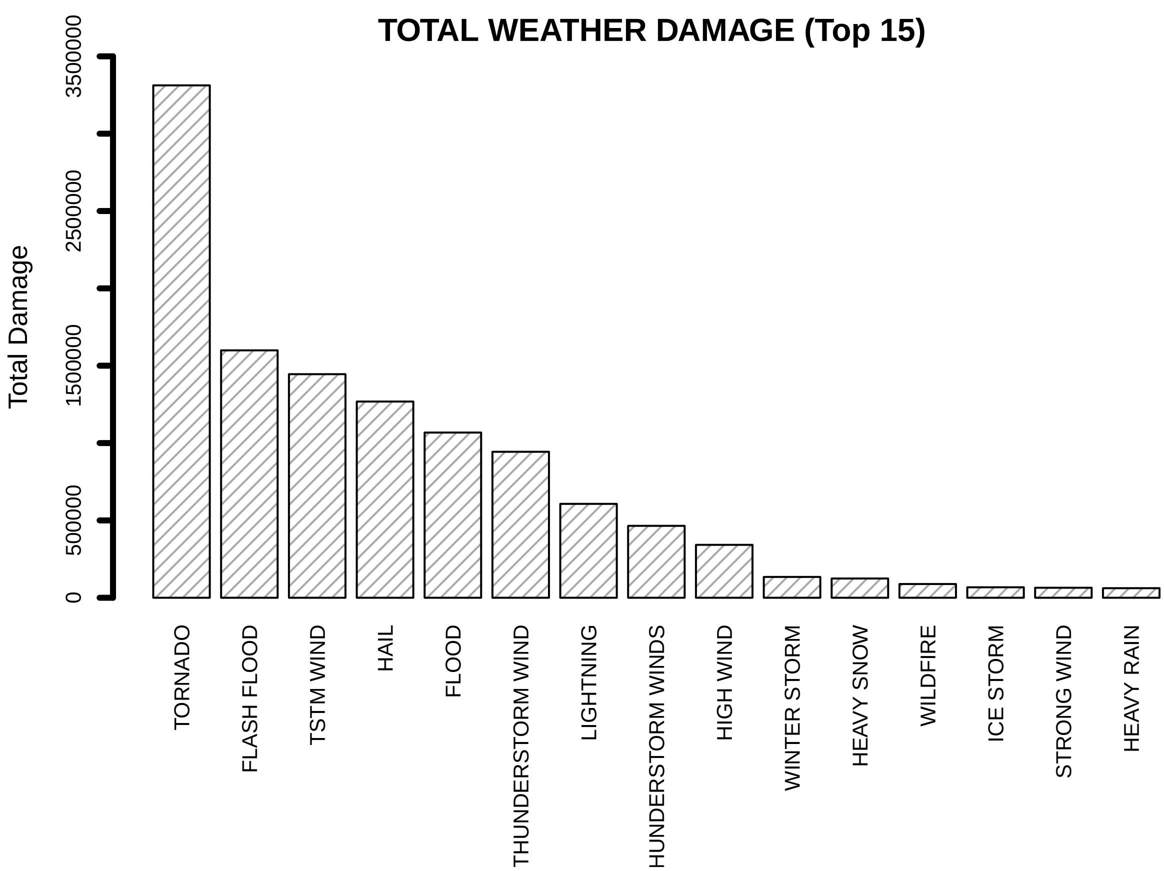
```
##         WeatherEvent TotalDamage
## 834          TORNADO  3312276.68
## 153      FLASH FLOOD  1599325.05
## 856        TSTM WIND  1445168.21
## 244             HAIL  1268289.66
```

```
## 170              FLOOD  1067976.36
## 760  THUNDERSTORM WIND   943635.62
## 464          LIGHTNING   606932.39
## 786 THUNDERSTORM WINDS   464978.11
## 359          HIGH WIND   342014.77
## 972       WINTER STORM   134699.58
## 310         HEAVY SNOW   124417.71
## 957           WILDFIRE    88823.54
## 427          ICE STORM    67689.62
## 676        STRONG WIND    64610.71
## 290         HEAVY RAIN    61964.94
```

4.5. Total Damage Plot

Total Damage = Crop Damage + Property Damage

```
par(mfrow = c(1,1), mar = c(10, 4, 2, 2), las = 3, cex = 0.7, cex.main = 1.4, cex.lab = 1.2)
barplot(Total_Damage_Display$TotalDamage, names.arg = Total_Damage_Display$WeatherEvent, col="darkgray"
```



5. RESULTS

The analysis on the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database shows us that:

- The weather event that is the *most harmful* with respect to population health across the United States are **Tornados**.
  - The *second most harmful* weather event is **Excessive Heat**.
- The weather event that causes the *greatest economic consequences* are **Tornados**.
  - The *second greatest cause* for economic concequences are **Flash Floods**.