

# Impact of extreme weather events on human health and estimating fatalities

Manjari singh

2024-06-30

## 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 project involves exploring the NOAA's storm database, which 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 basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events, such as:

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

The result indicated that Tornado followed by Excessive heat, Tstm Wind and Flooding are the ones having greatest impact on population health. Moreover, Flood followed by Hurricane/typhoon, Tornado and Storm Surge are the ones having greatest impact on economic.

## Data Processing

set directory, load data, and see data structure

```
setwd("C:/Users/MSingh/coursera/assignment/data")
noaa_data <- read.csv("repdata_data_StormData.csv", na.strings = TRUE)
str(noaa_data)
```

```
## 'data.frame': 903870 obs. of 40 variables:
## $ STATE__ : chr "1" "1" "1" "1" ...
## $ BGN_DATE : chr "4/18/1950 0:00" "4/18/1950 0:00" "2/20/1951 0:00" "6/8/1951 0:00" ...
## $ BGN_TIME : chr "130" "145" "1600" "900" ...
## $ TIME_ZONE : chr "CST" "CST" "CST" "CST" ...
## $ COUNTY : chr "97" "3" "57" "89" ...
## $ COUNTYNAME: chr "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
## $ STATE : chr "AL" "AL" "AL" "AL" ...
## $ EVTYPE : chr "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ BGN_RANGE : chr "0" "0" "0" "0" ...
## $ BGN_AZI : chr "" "" "" "" ...
## $ BGN_LOCATI: chr "" "" "" "" ...
## $ END_DATE : chr "" "" "" "" ...
## $ END_TIME : chr "" "" "" "" ...
## $ COUNTY_END: chr "0" "0" "0" "0" ...
## $ COUNTYENDN: chr "" "" "" "" ...
## $ END_RANGE : chr "0" "0" "0" "0" ...
## $ END_AZI : chr "" "" "" "" ...
## $ END_LOCATI: chr "" "" "" "" ...
## $ LENGTH : chr "14" "2" "0.1" "0" ...
## $ WIDTH : chr "100" "150" "123" "100" ...
## $ F : chr "3" "2" "2" "2" ...
## $ MAG : chr "0" "0" "0" "0" ...
## $ FATALITIES: chr "0" "0" "0" "0" ...
## $ INJURIES : chr "15" "0" "2" "2" ...
## $ PROPDMG : chr "25" "2.5" "25" "2.5" ...
## $ PROPDMGEXP: chr "K" "K" "K" "K" ...
## $ CROPDGMG : chr "0" "0" "0" "0" ...
## $ CROPDGMGEXP: chr "" "" "" "" ...
## $ WFO : chr "" "" "" "" ...
## $ STATEOFFIC: chr "" "" "" "" ...
## $ ZONENAMES : chr "" "" "" "" ...
## $ LATITUDE : chr "3040" "3042" "3340" "3458" ...
## $ LONGITUDE : chr "8812" "8755" "8742" "8626" ...
## $ LATITUDE_E: chr "3051" "0" "0" "0" ...
## $ LONGITUDE_: chr "8806" "0" "0" "0" ...
## $ REMARKS : chr "" "" "" "" ...
## $ REFNUM : chr "1" "2" "3" "4" ...
## $ X : chr "" "" "" "" ...
## $ X.1 : chr "" "" "" "" ...
## $ X.2 : chr "" "" "" "" ...
```

since all columns are of character datatype, we need to tranform few columns which has o be used during data analysis ### transform datatype of required columns to numeric type

```
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
```

```
noaa_data <- noaa_data %>% mutate_at(c('INJURIES', 'FATALITIES', 'PROPDGMG', 'CROPDMG'), as.numer  
ic)
```

```
## Warning: There were 4 warnings in `mutate()`.  
## The first warning was:  
## i In argument: `INJURIES = .Primitive("as.double")(INJURIES)`.  
## Caused by warning:  
## ! NAs introduced by coercion  
## i Run `dplyr::last_dplyr_warnings()` to see the 3 remaining warnings.
```

```
str(noaa_data)
```

```
## 'data.frame':  903870 obs. of  40 variables:
## $ STATE__ : chr  "1" "1" "1" "1" ...
## $ BGN_DATE : chr  "4/18/1950 0:00" "4/18/1950 0:00" "2/20/1951 0:00" "6/8/1951 0:00" ...
## $ BGN_TIME : chr  "130" "145" "1600" "900" ...
## $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
## $ COUNTY : chr  "97" "3" "57" "89" ...
## $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
## $ STATE : chr  "AL" "AL" "AL" "AL" ...
## $ EVTYPE : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ BGN_RANGE : chr  "0" "0" "0" "0" ...
## $ BGN_AZI : chr  "" "" "" "" ...
## $ BGN_LOCATI: chr  "" "" "" "" ...
## $ END_DATE : chr  "" "" "" "" ...
## $ END_TIME : chr  "" "" "" "" ...
## $ COUNTY_END: chr  "0" "0" "0" "0" ...
## $ COUNTYENDN: chr  "" "" "" "" ...
## $ END_RANGE : chr  "0" "0" "0" "0" ...
## $ END_AZI : chr  "" "" "" "" ...
## $ END_LOCATI: chr  "" "" "" "" ...
## $ LENGTH : chr  "14" "2" "0.1" "0" ...
## $ WIDTH : chr  "100" "150" "123" "100" ...
## $ F : chr  "3" "2" "2" "2" ...
## $ MAG : chr  "0" "0" "0" "0" ...
## $ FATALITIES: num  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 25 25 ...
## $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
## $ CROPDGMG : num  0 0 0 0 0 0 0 0 0 ...
## $ CROPDGMGEXP: chr  "" "" "" "" ...
## $ WFO : chr  "" "" "" "" ...
## $ STATEOFFIC: chr  "" "" "" "" ...
## $ ZONENAMES : chr  "" "" "" "" ...
## $ LATITUDE : chr  "3040" "3042" "3340" "3458" ...
## $ LONGITUDE : chr  "8812" "8755" "8742" "8626" ...
## $ LATITUDE_E: chr  "3051" "0" "0" "0" ...
## $ LONGITUDE_: chr  "8806" "0" "0" "0" ...
## $ REMARKS : chr  "" "" "" "" ...
## $ REFNUM : chr  "1" "2" "3" "4" ...
## $ X : chr  "" "" "" "" ...
## $ X.1 : chr  "" "" "" "" ...
## $ X.2 : chr  "" "" "" "" ...
```

## Events which are most harmful with respect to population health

In order to determine which events are most harmful to population health, examine injuries and fatalities data for each event type

```
casualties <- aggregate(cbind(noaa_data$INJURIES, noaa_data$FATALITIES) ~ noaa_data$EVTYPE, data
= noaa_data, sum)
## Change column names for better presentation
colnames(casualties) <- c("EventType", "Injuries", "Fatalities")
```

Events that causes the most harmful impact to population health will be ones that have high Injuries and Fatalities values:

```
## Add new column Injuries_Fatalities which is the sum of the other two
casualties$Injuries_Fatalities <- casualties$Injuries + casualties$Fatalities
## Sort data by Injuries_Fatalities
casualties <- casualties[with(casualties, order(-Injuries_Fatalities)),]
```

Top 10 events area:

```
head(casualties, 10)
```

##	EventType	Injuries	Fatalities	Injuries_Fatalities
## 834	TORNADO	90671	5593	96264
## 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

## Events have the greatest economic consequences

Economic consequences are determined by total property damage values (PROPDMG, PROPDMGEXP) and total crop damage value (CROPDMG, CROPDMGEXP). PROPDMGEXP and CROPDMGEXP have values of "H", "K", "M", "B" (hundred, thousand, million, billion respectively multiplier). create a function to calculate this multiplier:

```
get_exp_multiplier <- function(m) {  
  if (m %in% c('H', 'h'))  
    return (100)  
  else if (m %in% c('K', 'k'))  
    return (1000)  
  else if (m %in% c('M', 'm'))  
    return (1000000)  
  else if (m %in% c('B', 'b'))  
    return (1000000000)  
  else if (!is.na(as.numeric(m)))  
    return (as.numeric(m) * 10)  
  else  
    return(1)  
}
```

Then calculate total property and crop damage values:

```
## Calculate total property damage  
prop_dmg_mul <- sapply(noaa_data$PROPDMGEXP, FUN=get_exp_multiplier)
```

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion  
  
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
noaa_data$total_prop_dmg <- noaa_data$PROPDMG * prop_dmg_mul  
## Get crop damage multiplier  
crop_dmg_mul <- sapply(noaa_data$CROPDMGEXP, FUN=get_exp_multiplier)
```

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
noaa_data$total_crop_dmg <- noaa_data$CROPDMG * crop_dmg_mul
```

Group them by event type

```
storm_econ_dmg <- aggregate(cbind(noaa_data$total_prop_dmg, noaa_data$total_crop_dmg) ~ noaa_data$EVTYPE, data = noaa_data, sum)
## Change column names for better presentation
colnames(storm_econ_dmg) <- c("EventType", "Property_Dmg", "Crop_Dmg")
```

Events that have the greatest economic consequences will be ones that have high property damage and crop damage values

```
## Add new column Prop_Crop_Dmg which is the sum of the other two
storm_econ_dmg$Prop_Crop_Dmg <- storm_econ_dmg$Property_Dmg + storm_econ_dmg$Crop_Dmg
## Sort data by Prop_Crop_Dmg
storm_econ_dmg <- storm_econ_dmg[with(storm_econ_dmg, order(-Prop_Crop_Dmg)),]
```

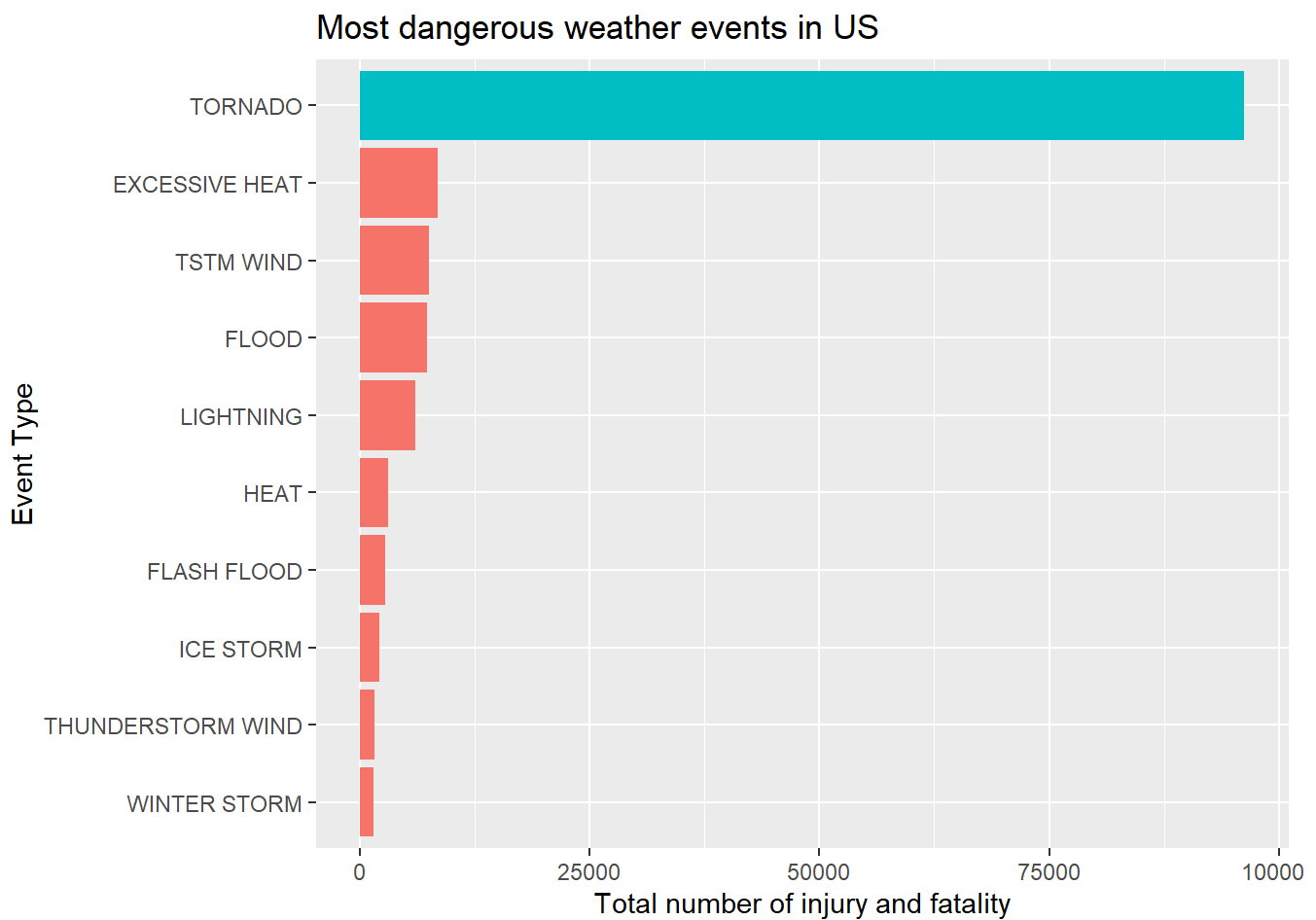
Here are top 10 events:

```
head(storm_econ_dmg, 10)
```

##	EventType	Property_Dmg	Crop_Dmg	Prop_Crop_Dmg
## 170	FLOOD	144657709807	5661968450	150319678257
## 411	HURRICANE/TYPHOON	69305840000	2607872800	71913712800
## 834	TORNADO	55845959553	414953110	56260912663
## 670	STORM SURGE	43323536000	5000	43323541000
## 244	HAIL	15731438777	3025954453	18757393230
## 153	FLASH FLOOD	16140817485	1421317100	17562134585
## 95	DROUGHT	1046106000	13972566000	15018672000
## 402	HURRICANE	11868319010	2741910000	14610229010
## 590	RIVER FLOOD	5118945500	5029459000	10148404500
## 427	ICE STORM	3944927810	5022113500	8967041310

#Results -From our analysis, Tornado is the one that has greatest impact on population health with over 96,000 injuries and fatalities, followed by Excessive heat, Tstm Wind and Flooding which caused around 8000 injuries and fatalities in each.

```
library(ggplot2)
## Add colors
casualties$FillColor = "blue"
casualties[1,]$FillColor = "red"
g1 <- ggplot(data = head(casualties, 10), aes(x = reorder(EventType, Injuries_Fatalities), y = I
njuries_Fatalities, fill = FillColor)) + geom_bar(stat = "identity") + coord_flip() + theme(lege
nd.position="none")
g1 <- g1 + labs( x = "Event Type", y = "Total number of injury and fatality", title = "Most dang
erous weather events in US")
g1
```



- From our analysis, Flood is the one that has greatest impact on economic with \$150 billions in property and crop damage, followed by Hurricane/typhoon, Tornado and Storm Surge which caused damage of \$72, \$57, \$43 billions respectively.

```

library(ggplot2)
## Add colors
storm_econ_dmg$FillColor = "blue"
storm_econ_dmg[1,]$FillColor = "red"
g2 <- ggplot(data = head(storm_econ_dmg, 10), aes(x = reorder(EventType, Prop_Crop_Dmg), y = Prop_Crop_Dmg / 10^9, fill = FillColor)) + geom_bar(stat = "identity") + coord_flip() + theme(legend.position="none")
g2 <- g2 + labs( x = "Event Type", y = "Property and Crop Damange (in billion dollars)", title = "Economy impact weather events in US")
g2

```

