US NOAA Storm Data Analysis

Tandin Dorji

2023-08-02

Data processing

As this analysis is not a complete data analysis project, it will not follow the standard delineation of data work. This section, while titled data processing will only deal with importing the data. The actual data processing work will be performed later in the exploratory data analysis section based on the need to process the data to answer specific questions.

Load libraries and data

- Load libraries
- Read data

```
# load/install required libraries
require(tidyverse)
if(!require(treemap)) {
        install.packages("treemap")
        library(treemap)
}


# read and check original dataset
sd <- read.csv("Dataset/StormData.csv", header = TRUE)
dim(sd)
```

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

```
# head(sd)
# str(sd)
names(sd)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Check data and preprocess

- Check data
- Select required variables

```
# select only required variables for analysis
storm <- sd[, c(8, 23:28, 36)]
rm(sd)
gc()
```

```
##            used  (Mb) gc trigger   (Mb)  max used   (Mb)
## Ncells  1512257  80.8    4646155  248.2   2613589  139.6
## Vcells 34965931 266.8  226346949 1726.9 242497797 1850.2
```

```
head(storm)
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP REMARKS
## 1 TORNADO          0       15    25.0          K       0
## 2 TORNADO          0        0     2.5          K       0
## 3 TORNADO          0        2    25.0          K       0
## 4 TORNADO          0        2     2.5          K       0
## 5 TORNADO          0        2     2.5          K       0
## 6 TORNADO          0        6     2.5          K       0
```

(Exploratory) Data analysis

In this section, we will do an exploratory analysis of the storm data to look at events which are most harmful to population health and economy.

Events that are most harmful to population health

- Aggregate data by event type to find total fatality and injury
- Generate plots for above

```
# Across the United States, which types of events
# (as indicated in the EVTYPE variable) are most harmful
# with respect to population health?
fatality <- storm %>%
        group_by(EVTYPE) %>%
        summarise(Total.Fatalities = sum(FATALITIES)) %>%
        arrange(desc(Total.Fatalities))


lrow <- nrow(fatality) # number of observation for data frame subsetting/indexing


fatality.summ <- data.frame("Event.Type" = c((fatality$EVTYPE[1:10]), "Others"),
                            "Fatality" = c(fatality$Total.Fatalities[1:10],
                                           sum(fatality$Total.Fatalities[11:lrow])))
```

```
injury <- storm %>%
        group_by(EVTYPE) %>%
        summarise(Total.Injuries = sum(INJURIES)) %>%
        arrange(desc(Total.Injuries))


injury.summ <- data.frame("Event.Type" = c((injury$EVTYPE[1:10]), "Others"),
                            "Injury" = c(injury$Total.Injuries[1:10],
                                           sum(injury$Total.Injuries[11:lrow])))


odr <- seq(1:11) # to reorder x-axis/bars from highest to lowest, others


fatality.plot <- ggplot(fatality.summ,
               aes(x = reorder(Event.Type, odr), y = Fatality)) +
        geom_bar(stat = "identity", aes(fill = Event.Type)) +
        labs(title = "Fatality by event type") +
        theme(
                axis.text.x = element_text(angle = 45, hjust = 1),
                axis.title.x = element_blank(),
                legend.position = "none"
        )


injury.plot <- ggplot(injury.summ,
                      aes(x = reorder(Event.Type, odr), y = Injury)) +
        geom_bar(stat = "identity", aes(fill = Event.Type)) +
        labs(title = "Injury by event type") +
        theme(
                axis.text.x = element_text(angle = 45, hjust = 1),
                axis.title.x = element_blank(),
                legend.position = "none"
        )
```

Economic impact analysis

In this subsection, we will find the events that have the greatest economic consequences. For this analysis, the total economic value of the damages due to an event will be taken as the sum of nominal dollar value of damage to property and crops.

First, let us check the above data to see if we have the required data to perform this calculation.

Check Data and Impute Values

```
table(storm$CROPDMGEXP)
```

```
##
##              ?      0      2      B      k      K      m      M
## 618413       7     19      1      9     21 281832      1   1994
```

```r
incorrect <- storm[storm$CROPDMGEXP %in% c("?", "0", "2"), 6:7]
table(incorrect)
```

```
##        CROPDMGEXP
## CROPDMG ? 0 2
##      0  7 7 1
##      5  0 6 0
##      20 0 1 0
##      25 0 2 0
##      50 0 2 0
##      60 0 1 0
```

```r
sum(table(incorrect))
```

```
## [1] 27
```

```r
# boxplot(storm$CROPDMG, horizontal = TRUE)
# hist(storm$CROPDMG)
```

All crop damage values with crop damage magnitude as blank, ?, 0 or 2 will be replaced by zero and xp recoded as 0. ‘k’ and ‘m’ will be recoded as K and M respectively for uniformity.

```r
storm[storm$CROPDMGEXP %in% c("", "?", "0", "2"), 6:7] <- 0
storm[storm$CROPDMGEXP %in% c("k"), 7] <- "K"
storm[storm$CROPDMGEXP %in% c("m"), 7] <- "M"
table(storm$CROPDMGEXP)
```

```
##
##      0      B      K      M
## 618440      9 281853   1995
```

```r
sum(table(storm$CROPDMGEXP))
```

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

It is confirmed that all 902,297 rows of data have crop damage magnitude as either 0, B, K or M.

We will do a similar transformation for property damage value and magnitudes. All property damage values with property damage magnitude other than K, M/m, or B will be replaced by zero and xp recoded as 0. ‘k’ and ‘m’ will be recoded as K and M respectively for uniformity

```r
table(storm$PROPDMGEXP)
```

```
##
##             -      ?      +      0      1      2      3      4      5      6
## 465934      1      8      5    216     25     13      4      4     28      4
##      7      8      B      h      H      K      m      M
##      5      1     40      1      6 424665      7  11330
```

```
storm[!(storm$PROPDMGEXP %in% c("K", "m", "M", "B")), 4:5] <- 0
storm[storm$PROPDMGEXP %in% c("k"), 5] <- "K"
storm[storm$PROPDMGEXP %in% c("m"), 5] <- "M"
table(storm$PROPDMGEXP)
```

```
##
##      0      B      K      M
## 466255     40 424665  11337
```

```
sum(table(storm$PROPDMGEXP))
```

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

As we now have crop and property damage figures in numbers and magnitude in characters, we will transform the magnitude to their numeric equivalent to obtain the total dollar value.

Transform data to get dollar values

For this, we will first create two new variables to store the numeric values of the damage magnitudes and then calculate the total dollar value of damages per event by multiplying the damage amounts with their magnitudes. For ease of presentation, the total damage amount will be computed in billions of nominal dollars.

```
Prop.XP <- case_match(storm$PROPDMGEXP,
                      "K" ~ 1e3,
                      "M" ~ 1e6,
                      "B" ~ 1e9,
                      .default = 1)

Crop.XP <- case_match(storm$CROPDMGEXP,
                      "K" ~ 1e3,
                      "M" ~ 1e6,
                      "B" ~ 1e9,
                      .default = 1)
storm <- cbind(storm, Prop.XP, Crop.XP)
storm$Damage.Value <-  (storm$PROPDMG * storm$Prop.XP +
                               storm$CROPDMG * storm$Crop.XP)/1e9
```

With this we now have the total value of damage for all of the recorded events. This data can now be used to find the total economic consequence of these events as follows.

Group data and generate treemap

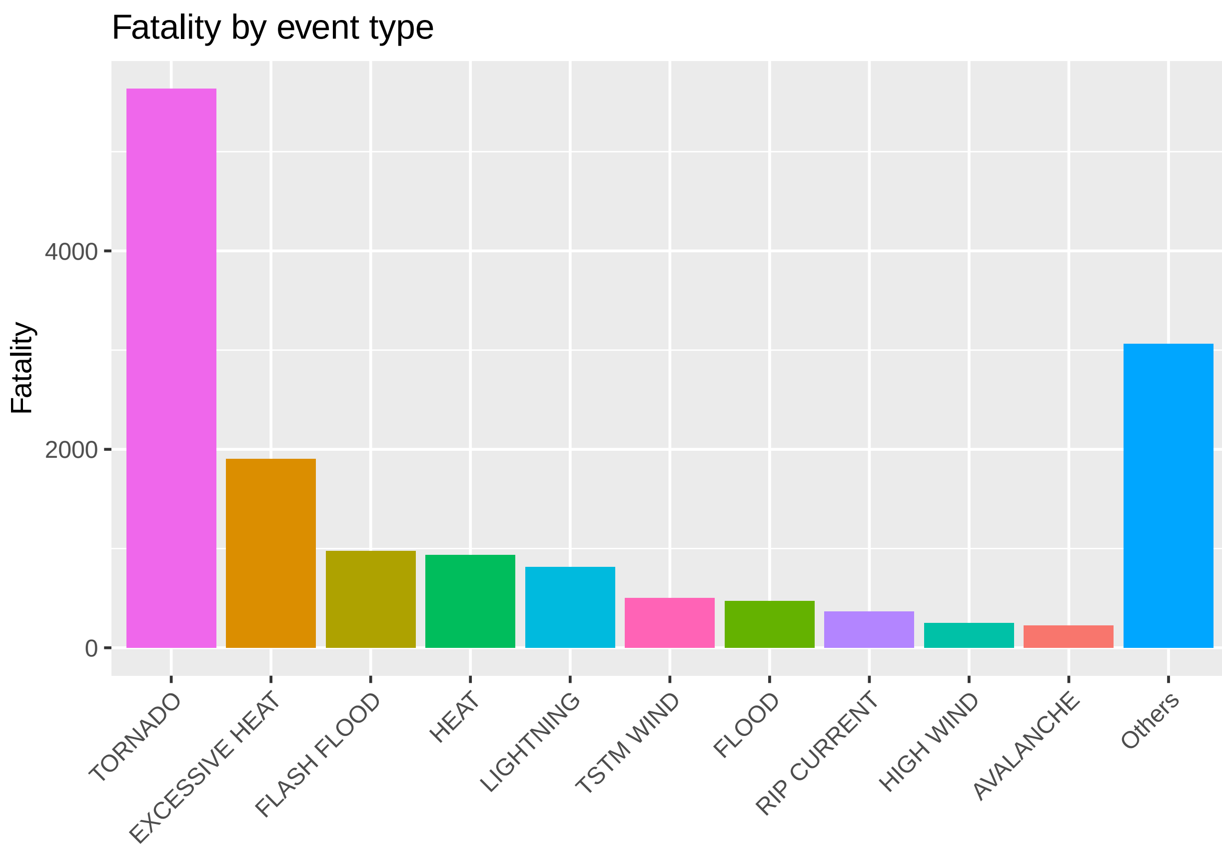
Among other options, we will use a treemap to visually see, at a glance the events that have had the most economic consequence. For this, we will aggregate the total damage values of all events by their event type and find the percentage of the total damage attributable to the event.

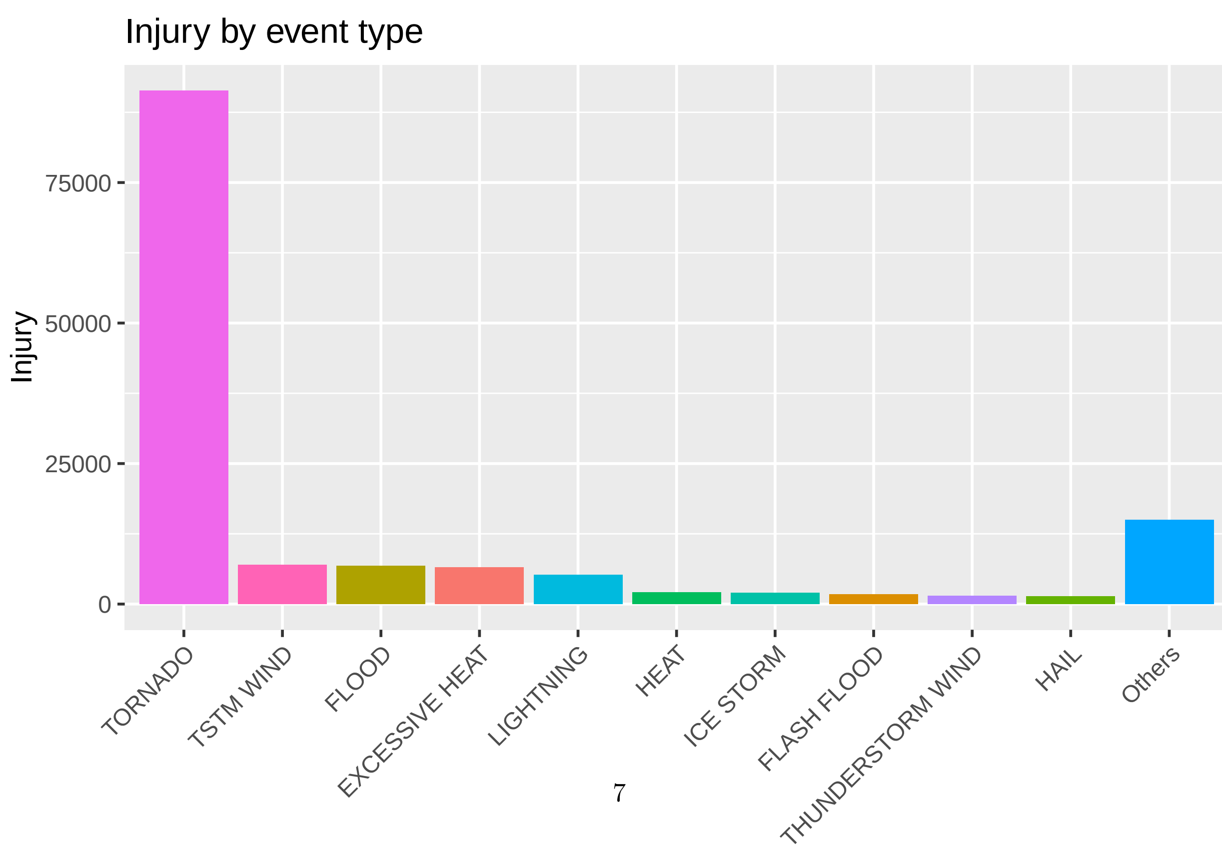
```
economic.plot <- storm %>%
        group_by(EVTYPE) %>%
        summarise(DMG.VAL = sum(Damage.Value)) %>%
```

```
        mutate(Event.Percent =
                       paste0(EVTYPE, "\n",
                              round(DMG.VAL/sum(DMG.VAL)*100, 2), "%" ))
```

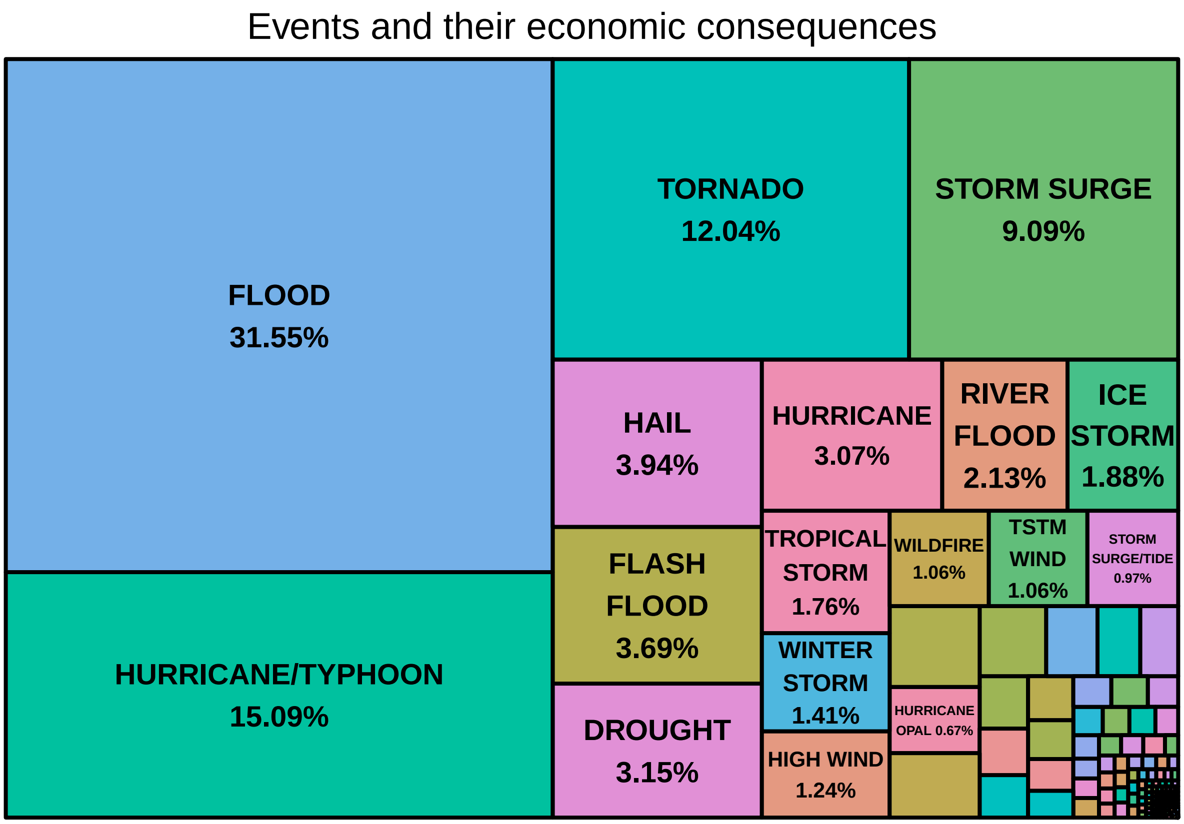
Results

Fatality and Injury





Economic consequences



Discussion of result

This brief analysis of the US NOAA storm data offers some useful insights on the impact of natural disasters on population health and economy. Tornado appears to be the top event that causes the highest death, injury and third highest damage to property/crop. Next, flashflood and flood are top causes of economic damage and is among the top then causes of fatality and injury. While heat, excessive heat and lightning are highly fatal and injurious events, they do not have significant economic impact. Overall, the top ten events causes more fatality, injury and economic damage than the rest 975 events combined.