Types of severe weather events in the United States that have the greatest impact

Steffen Zou Weilun

3 November 2020

I will explore the U.S. National Oceanic and Atmospheric Administration's (NOAA) Storm Database to address the following questions about severe weather events:

1. Across the United States, which types of events are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

From on the results of the analysis, it shows that tornado is the most harmful with respect to population health, followed by thunderstorm wind and excessive heat. It also shows that flood has the greatest economic consequences, followed by hurricane/typhoon and tornado.

```
library(dplyr)
library(ggplot2)
```

Data Processing

```
fileUrl = 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
fileName = 'repdata_data_StormData.csv.bz2'

if(!file.exists(fileName)) {
    download.file(url = fileUrl, dest = fileName)
}

stormData <- read.csv(fileName, na.strings = '')
```

As there are 2 columns (i.e. FATALITIES and INJURIES) which represents harm with respect to population health, I will create a new column with the sum of the numbers from both of these columns.

The columns for property damage and crop damage (i.e. PROPDMG and CROPDMG) contain numbers with different scales, noted in other columns (i.e. PROPDMGEXP and CROPDMGEXP) as "K" for thousands, "M" for millions, and "B" for billions. I will create 2 new columns that will contain numeric values of the same scale in order to be used for analysis.

```
stormData2 <-
    rename_with(stormData, tolower) %>%
    mutate(harms = fatalities + injuries
           , propdmgfull =
               case_when(
                   is.na(propdmgexp) ~ propdmg,
                   propdmgexp %in% c('k', 'K') ~ propdmg * 1000,
                   propdmgexp %in% c('m', 'M') ~ propdmg * 1000000,
                   propdmgexp %in% c('b', 'B') ~ propdmg * 1000000000,
                   TRUE ~ propdmg
               )
```

```
            , cropdmgfull =
                case_when(
                    is.na(cropdmgexp) ~ cropdmg,
                    cropdmgexp %in% c('k', 'K') ~ cropdmg * 1000,
                    cropdmgexp %in% c('m', 'M') ~ cropdmg * 1000000,
                    cropdmgexp %in% c('b', 'B') ~ cropdmg * 1000000000,
                    TRUE ~ cropdmg
                )
            , economicdmg = propdmgfull + cropdmgfull
    ) %>%
    select(evtype, fatalities, injuries, harms, propdmgfull
           , cropdmgfull, economicdmg)
```

There are rows with the same event types which are named differently (e.g. “HURRICANE”, “TYPHOON”, “HURRICANE/TYPHOON”). I will update them to have the same value.

The event names are referenced from the storm data event table, which is found in the Storm Data Documentation, which can be downloaded from https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf.

```
stormData2$evtype <- toupper(trimws(gsub(' +', ' ', stormData2$evtype)))

events <- c('Astronomical Low Tide', 'Marine High Wind', 'Marine Strong Wind'
            , 'Marine Thunderstorm Wind'
            , 'Coastal Flood', 'Debris Flow'
            , 'Dense Fog', 'Dense Smoke', 'Dust Devil', 'Dust Storm'
            , 'Excessive Heat', 'Flash Flood', 'Funnel Cloud', 'Freezing Fog'
            , 'Heavy Rain', 'Heavy Snow', 'High Surf', 'High Wind', 'Ice Storm'
            , 'Lakeshore Flood', 'Lake-Effect Snow', 'Marine Hail'
            , 'Rip Current', 'Strong Wind', 'Thunderstorm Wind'
            , 'Tropical Depression', 'Tropical Storm', 'Volcanic Ash'
            , 'Winter Storm', 'Winter Weather'
            , 'Extreme Cold/Wind Chill', 'Cold/Wind Chill', 'Frost/Freeze'
            , 'Storm Surge/Tide', 'Hurricane/Typhoon'
            , 'Avalanche', 'Blizzard', 'Drought', 'Flood', 'Hail', 'Heat'
            , 'Lightning', 'Seiche', 'Sleet', 'Tornado', 'Tsunami'
            , 'Waterspout', 'Wildfire')

for (i in 1:length(events)) {
    rePattern = paste('^', toupper(events[i]), sep = '')
    matchRows <- grepl(rePattern, stormData2$evtype)
    selectedRows <- matchRows & is.na(stormData2$evtypenew)
    stormData2[selectedRows, 'evtypenew'] <- toupper(events[i])
}

eventStartWith <- c('Extreme Cold', 'Extreme Wind Chill'
                    , 'Cold', 'Wind Chill'
                    , 'Frost', 'Freeze', 'Storm Surge'
                    , 'hurricane', 'typhoon'
                    , 'tstm wind')

eventFinal <- c('Extreme Cold/Wind Chill', 'Extreme Cold/Wind Chill'
                , 'Cold/Wind Chill', 'Cold/Wind Chill'
                , 'Frost/Freeze', 'Frost/Freeze', 'Storm Surge/Tide'
                , 'Hurricane/Typhoon', 'Hurricane/Typhoon'
```

```
                , 'Thunderstorm Wind')

for (i in 1:length(eventStartWith)) {
    rePattern = paste('^', toupper(eventStartWith[i]), sep = '')
    matchRows <- grepl(rePattern, stormData2$evtype)
    selectedRows <- matchRows & is.na(stormData2$evtypenew)
    stormData2[selectedRows, 'evtypenew'] <- toupper(eventFinal[i])
}

stormData2[is.na(stormData2$evtypenew), 'evtypenew'] <-
    stormData2[is.na(stormData2$evtypenew), 'evtype']


mostHarmfulEvTypes <- group_by(stormData2, evtypenew) %>%
    summarize(totalharms = sum(harms)) %>%
    arrange(desc(totalharms)) %>%
    head(n = 10)

mostEconDmgEvTypes <- group_by(stormData2, evtypenew) %>%
    summarize(totalecondmg = sum(economicdmg)) %>%
    arrange(desc(totalecondmg)) %>%
    head(n = 10)
```

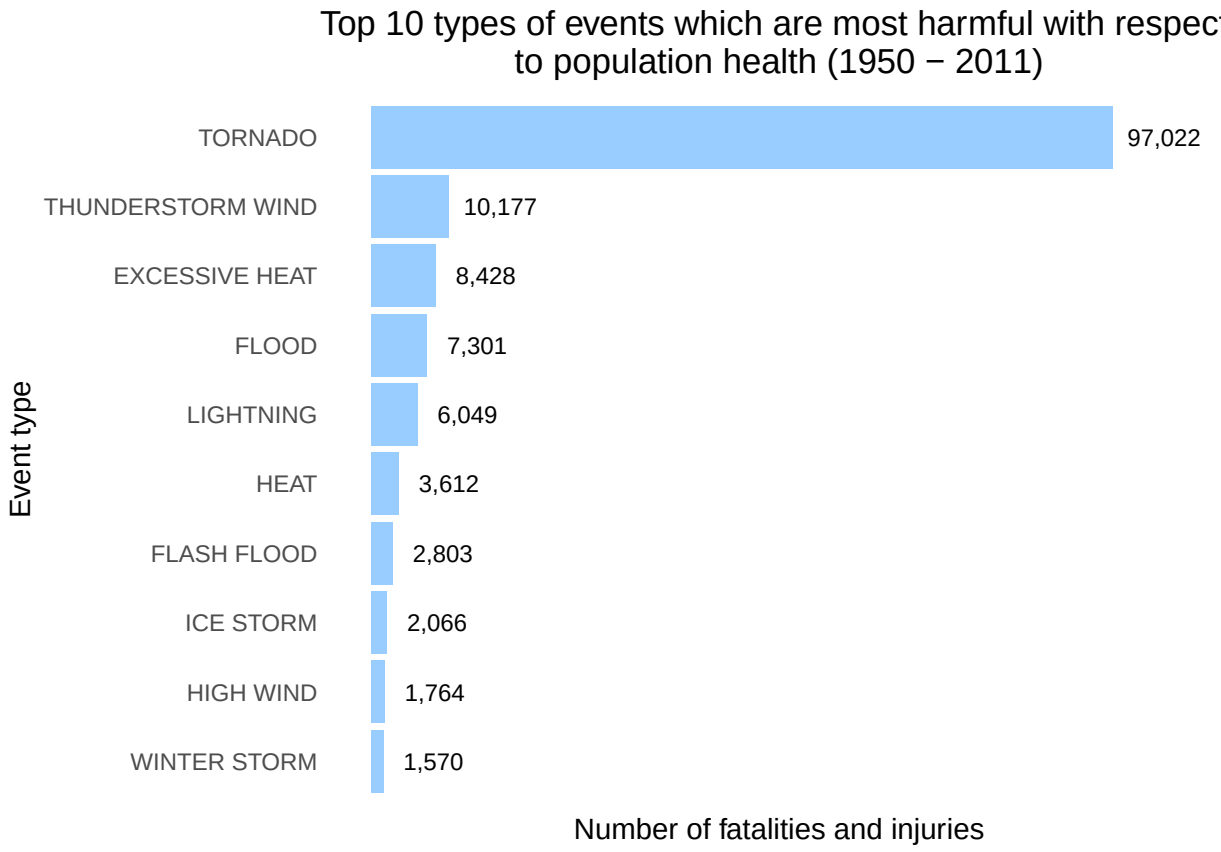
Results

```
mostHarmfulEvTypes$evtypenew <- with(mostHarmfulEvTypes
                                     , factor(evtypenew, levels = rev(evtypenew)))

title = 'Top 10 types of events which are most harmful with respect
to population health (1950 - 2011)'

xlim <- max(mostHarmfulEvTypes$totalharms) * 1.1

ggplot(data = mostHarmfulEvTypes, aes(x = totalharms, y = evtypenew)) +
    geom_col(fill = '#99ccff') +
    ggtitle(title) +
    labs(x = 'Number of fatalities and injuries', y = 'Event type') +
    geom_text(aes(label = format(totalharms, big.mark = ','))
              , hjust = -0.2, size = 3.3) +
    coord_cartesian(xlim = c(0, xlim)) +
    theme_minimal() +
    theme(panel.grid = element_blank()
          , plot.title = element_text(hjust = 0.5)
          , axis.text.x = element_blank()
          )
```



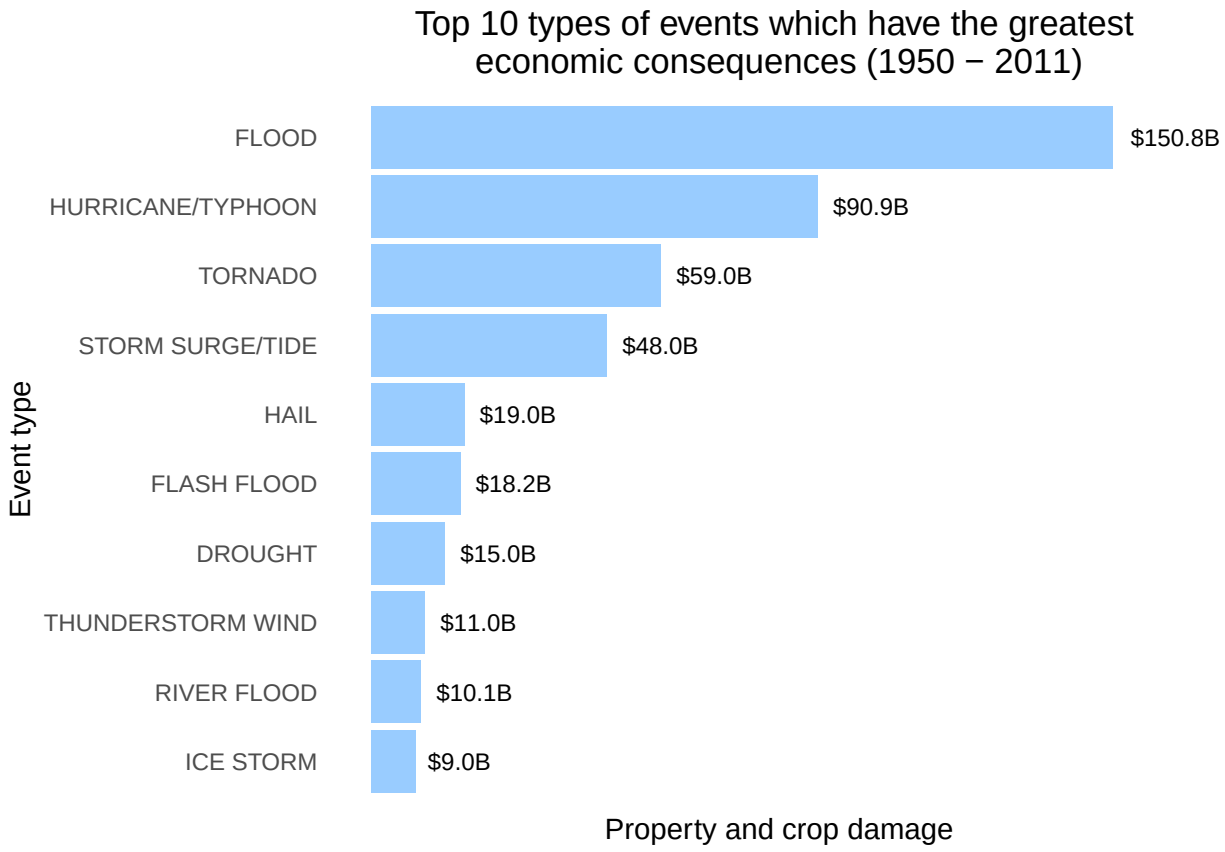
From this bar plot, it shows that tornado is the most harmful with respect to population health, followed by thunderstorm wind and excessive heat.

```
mostEconDmgEvTypes$evtypenew <- with(mostEconDmgEvTypes
                                     , factor(evtypenew, levels = rev(evtypenew)))

title = 'Top 10 types of events which have the greatest
economic consequences (1950 - 2011)'

xlim <- max(mostEconDmgEvTypes$totalecondmg) * 1.1

ggplot(data = mostEconDmgEvTypes, aes(x = totalecondmg, y = evtypenew)) +
    geom_col(fill = '#99ccff') +
    ggtitle(title) +
    labs(x = 'Property and crop damage', y = 'Event type') +
    geom_text(aes(label = paste('$'
                                , format(round(totalecondmg / 1000000000, 1)
                                         , big.mark = ',', trim = TRUE)
                                , 'B', sep = ''))
              , hjust = -0.2, size = 3.3) +
    coord_cartesian(xlim = c(0, xlim)) +
    theme_minimal() +
    theme(panel.grid = element_blank()
          , plot.title = element_text(hjust = 0.5)
          , axis.text.x = element_blank()
          )
```



From this bar plot, it shows that flood has the greatest economic consequences, followed by hurricane/typhoon and tornado.