Population Health and Economic Consequences of Weather Events in US

Alan Wong

4/2/2020

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 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.

Specifically, the following questions will be addressed:
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?

Table of Contents

Data

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.

Data Processing

Data Preview

Data is downloaded from the NOAA's storm database, and is then loaded and previewed.

```
# Download Data
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
file.name <- "storm_data.csv.bz2"

download.file(url, file.name)

# Read Data
df <- read.csv(file.name)
print(head(df))
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

Data Transformation

Not all columns in the data set are relevant to our analysis. We select only the relevant columns from our data.

```
library(dplyr)
library(magrittr)
```

```
df <- df %>%
  dplyr::select(BGN_DATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)
```

Since damage values are denoted in different scales, we have to first align them into a common scale of USD.

```
df <- df %>%
  mutate(PROPDMGVAL = case_when(
    PROPDMGEXP == '+' ~ PROPDMG * 10^0,
    PROPDMGEXP %in% paste(seq(0,8)) ~ PROPDMG * 10^1,
    PROPDMGEXP %in% c('H', 'h') ~ PROPDMG * 10^2,
    PROPDMGEXP %in% c('K', 'k') ~ PROPDMG * 10^3,
    PROPDMGEXP %in% c('M', 'm') ~ PROPDMG * 10^6,
    PROPDMGEXP %in% c('B', 'b') ~ PROPDMG * 10^9,
    TRUE ~ 0),
    CROPDMGVAL = case_when(
    CROPDMGEXP == '+' ~ CROPDMG * 10^0,
    CROPDMGEXP %in% paste(seq(0,8)) ~ CROPDMG * 10^1,
    CROPDMGEXP %in% c('H', 'h') ~ CROPDMG * 10^2,
    CROPDMGEXP %in% c('K', 'k') ~ CROPDMG * 10^3,
    CROPDMGEXP %in% c('M', 'm') ~ CROPDMG * 10^6,
    CROPDMGEXP %in% c('B', 'b') ~ CROPDMG * 10^9,
    TRUE ~ 0)
  )
```

Next, we can calculate the total amount of damages, including property and crops damage.

```
df <- df %>%
  mutate(TOTAL.ECONOMIC.IMPACT = PROPDMGVAL + CROPDMGVAL)
```

We can also calculate the total amount of casualties caused by the weather events.

```
df <- df %>%
  mutate(TOTAL.HEALTH.IMPACT = FATALITIES + INJURIES)
```

Here, we consider only dates starting from 1/1/1996 as stated in the documentation.

```
df <- df %>%
  filter(as.Date(BGN_DATE, '%m/%d/%Y') >= as.Date('1/1/1996', '%m/%d/%Y'))
```

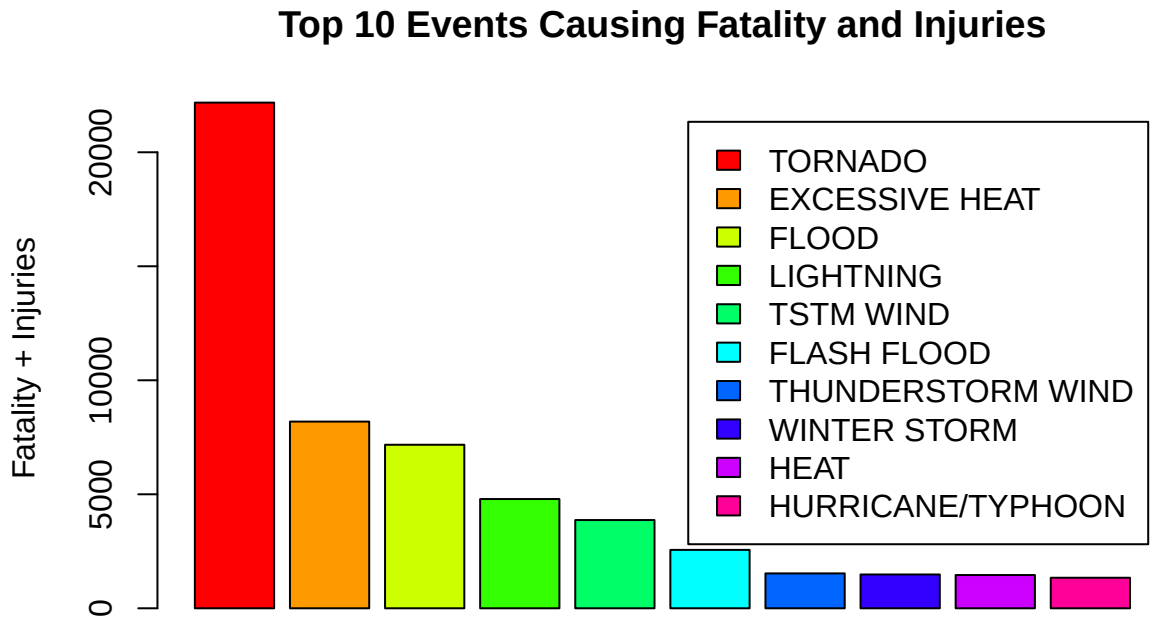
Results

Event Types Most Harmful to Population Health

```
top.pop.health.summary.df <- df %>%
  group_by(EVTYPE) %>%
  summarise(TOTAL.HEALTH.IMPACT = sum(TOTAL.HEALTH.IMPACT, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(TOTAL.HEALTH.IMPACT)) %>%
  head(10)

barplot(
  top.pop.health.summary.df$TOTAL.HEALTH.IMPACT,
  col = rainbow(10),
  legend.text = top.pop.health.summary.df$EVTYPE,
  ylab="Fatality + Injuries",
```

```
  main="Top 10 Events Causing Fatality and Injuries"
  )
```

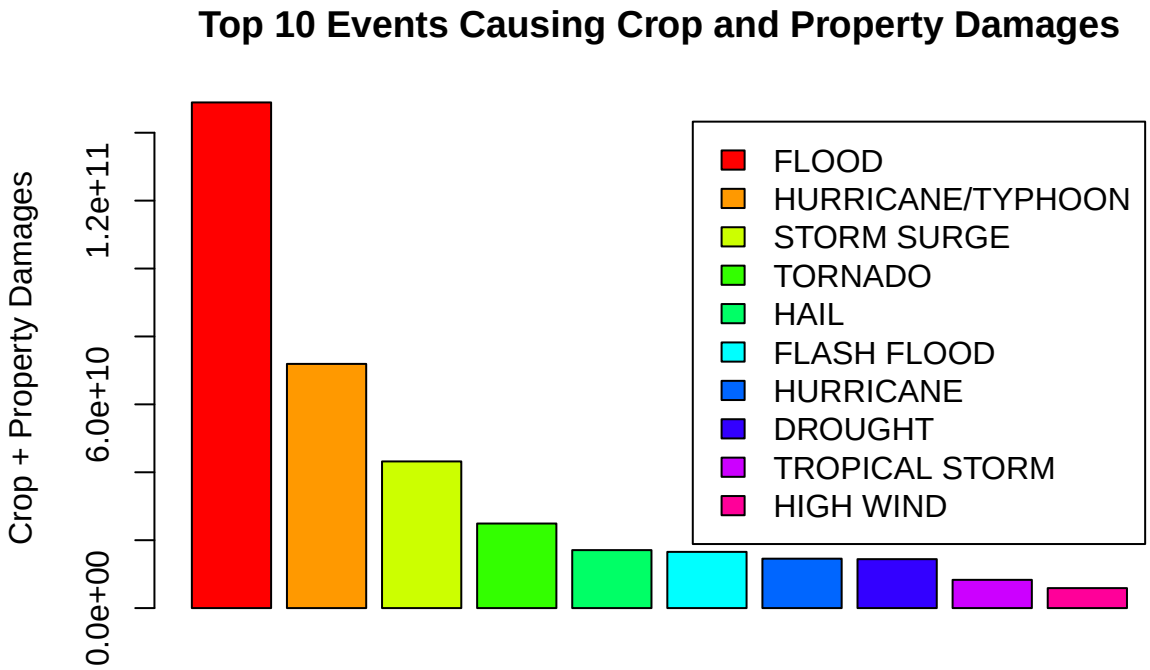


Based on the plots, we can conclude that Tornadoes, Excessive Heat, and Floods are the most harmful events to population health.

Event Types with Greatest Economic Consequences

```
top.economic.impact.summary.df <- df %>%
  group_by(EVTYPE) %>%
  summarise(TOTAL.ECONOMIC.IMPACT = sum(TOTAL.ECONOMIC.IMPACT, na.rm = TRUE)) %>%
  ungroup() %>%
  arrange(desc(TOTAL.ECONOMIC.IMPACT)) %>%
  head(10)

barplot(
  top.economic.impact.summary.df$TOTAL.ECONOMIC.IMPACT,
  col = rainbow(10),
  legend.text = top.economic.impact.summary.df$EVTYPE,
  ylab="Crop + Property Damages",
  main="Top 10 Events Causing Crop and Property Damages"
  )
```



Based on the plots, we can conclude that Floods, Hurricane/Typhoon, and Storm Surge causes the greatest economic consequences.