# Influence of Weather Events on Population Health and Economy by Event Type

Jonathan Lehmkuhl

2023-03-24

## SYNOPSIS

The U.S. National Oceanic and Atmospheric Administration's (**NOAA**) storm 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. The database that we used can be found here.

We want to investigate, which types of weather events are most harmful to population health and economy. These insights, for example, could be used to prepare for predicted weather events more effectively.

In our analysis we restricted our selfs on the event types flash flood, flood, hail, snow, wind, lightning, thunderstorm wind and tornado. Carrying out the analysis shown below, we found out that tornadoes and lightnings should be prepared for with great care. These are by far most severe with respect to population health. Economcially, a damage should be expected for every event type. However, tornadoes and lightning again create the most damage.

## DATA PREPROCESSING

We download the data file and read it into a data frame.

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "data.csv.bz2
data <- read.csv("data.csv.bz2")

names(data)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"     "COUNTYNAME" "STATE"      "EVT
## [10] "BGN_AZI"    "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN" "END_RANGE"  "END_
## [19] "LENGTH"     "WIDTH"      "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"    "PRO
## [28] "CROPDMGEXP" "WFO"        "STATEOFFIC" "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONG
## [37] "REFNUM"
```

We choose **FATALITIES** and **INJURIES** as the most relevant variables to investigate the impact on population health and **PROPDMG** (property damage) as the most relevant variable with respect to economy. Moreover, we want to look at **EVTYPE** (event type) and **BGN_DATE** (date of event). To keep our analysis more simple and focused on the essentials, we reduce our data frame to these variables.

```
suppressMessages(library(dplyr))
data_reduced <- select(data, BGN_DATE, EVTYPE, FATALITIES, INJURIES, PROPDMG)
head(data_reduced)
```

```
##             BGN_DATE  EVTYPE FATALITIES INJURIES PROPDMG
## 1  4/18/1950 0:00:00 TORNADO          0       15    25.0
## 2  4/18/1950 0:00:00 TORNADO          0        0     2.5
```

```
## 3  2/20/1951 0:00:00 TORNADO          0        2    25.0
## 4   6/8/1951 0:00:00 TORNADO          0        2     2.5
## 5 11/15/1951 0:00:00 TORNADO          0        2     2.5
## 6 11/15/1951 0:00:00 TORNADO          0        6     2.5
```

Since we want to find out, which types of event have the most impact, let us look on the available event types.

```
evtypes <- data_reduced$EVTYPE
length(unique(evtypes))
```

```
## [1] 985
```

As we see, there are a total of 985 event types.

```
head(table(evtypes))
```

```
## evtypes
##    HIGH SURF ADVISORY         COASTAL FLOOD           FLASH FLOOD             LIGHTNING
##                     1                     1                     1                     1
##       TSTM WIND (G45)
##                     1
```

However, as we can see in the table, there are event types which only occur a small number of times. To make generalizable statements about weather events, we only want to look at event types for which a large amount of data is available.

```
ev_table <- table(evtypes)
ev_table[ev_table > 15000]
```

```
## evtypes
##        FLASH FLOOD              FLOOD               HAIL         HEAVY SNOW          HIGH WIND
##              54277              25326             288661              15708              20212
##  THUNDERSTORM WIND THUNDERSTORM WINDS            TORNADO          TSTM WIND
##              82563              20843              60652             219940
```

We decide to focus on the types of events for which more than fifteen thousand instances exist. We see that these are ten types of events. However, there are three categories all representing the event type thunderstorm wind, which we want to merge together.

```
data_reduced[data_reduced$EVTYPE %in% c("TSTM WIND", "THUNDERSTORM WINDS"), "EVTYPE"] <- "THUNDERSTORM W
ev_table <- table(data_reduced$EVTYPE)
ev_table <- ev_table[ev_table > 15000]

df <- filter(data_reduced, EVTYPE %in% names(ev_table))
df$EVTYPE <- as.factor(df$EVTYPE)
levels(df$EVTYPE)
```

```
## [1] "FLASH FLOOD"       "FLOOD"             "HAIL"              "HEAVY SNOW"        "HIGH WIND"
## [6] "LIGHTNING"         "THUNDERSTORM WIND" "TORNADO"
```

We get our final data frame **df**, upon which we are going to build our analysis. **df** only consists of the event types **FLASH FLOOD**, **FLOOD**, **HAIL**, **HEAVY SNOW**, **HIGH WIND**, **LIGHTNING**, **THUNDERSTORM WIND** and **TORNADO**.

# RESULTS

## IMPACT ON POPULATION HEALTH

To evaluate the impact that weather events have on population health, let us first look at injuries. To determine how harmful an event type is with respect to injuries, we compute the mean and the maximal number of injuries for each event type. In addition, we compute the probability that there is at least one injured person for an event.

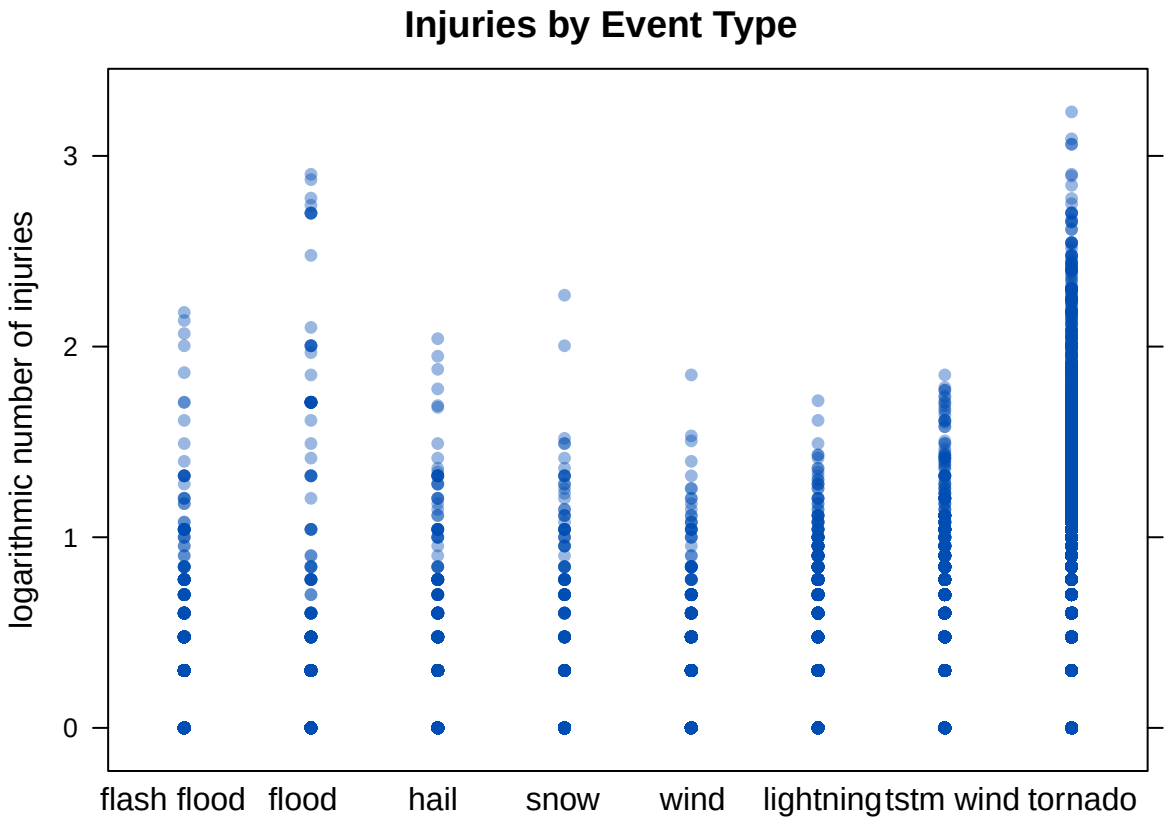
```
df %>% group_by(EVTYPE) %>% summarize(INJ_MEAN = mean(INJURIES), INJ_MAX = max(INJURIES), PROB = mean(as
```

```
## # A tibble: 8 x 4
##   EVTYPE            INJ_MEAN INJ_MAX     PROB
##   <fct>                <dbl>   <dbl>    <dbl>
## 1 TORNADO             1.51       1700 0.127
## 2 LIGHTNING           0.332        51 0.178
## 3 FLOOD               0.268       800 0.00612
## 4 HEAVY SNOW          0.0650      185 0.00828
## 5 HIGH WIND           0.0563       70 0.0203
## 6 FLASH FLOOD         0.0327      150 0.00698
## 7 THUNDERSTORM WIND   0.0289       70 0.0112
## 8 HAIL                0.00471     109 0.000970
```

We see that for tornado and lightning more than ten percent of the events lead to injured persons. All other event types have probabilities of two percent or less. Tornado and lightning also have the highest average values and tornado has the highest maximum of injured persons.

```
library(lattice)

xyplot(log10(INJURIES + 1) ~ EVTYPE, data = df, xlab = c("flash flood", "flood", "hail", "snow", "wind"
```




Plotting all injuries, we see again that tornadoes seem to be most dominant with respect to injuries.

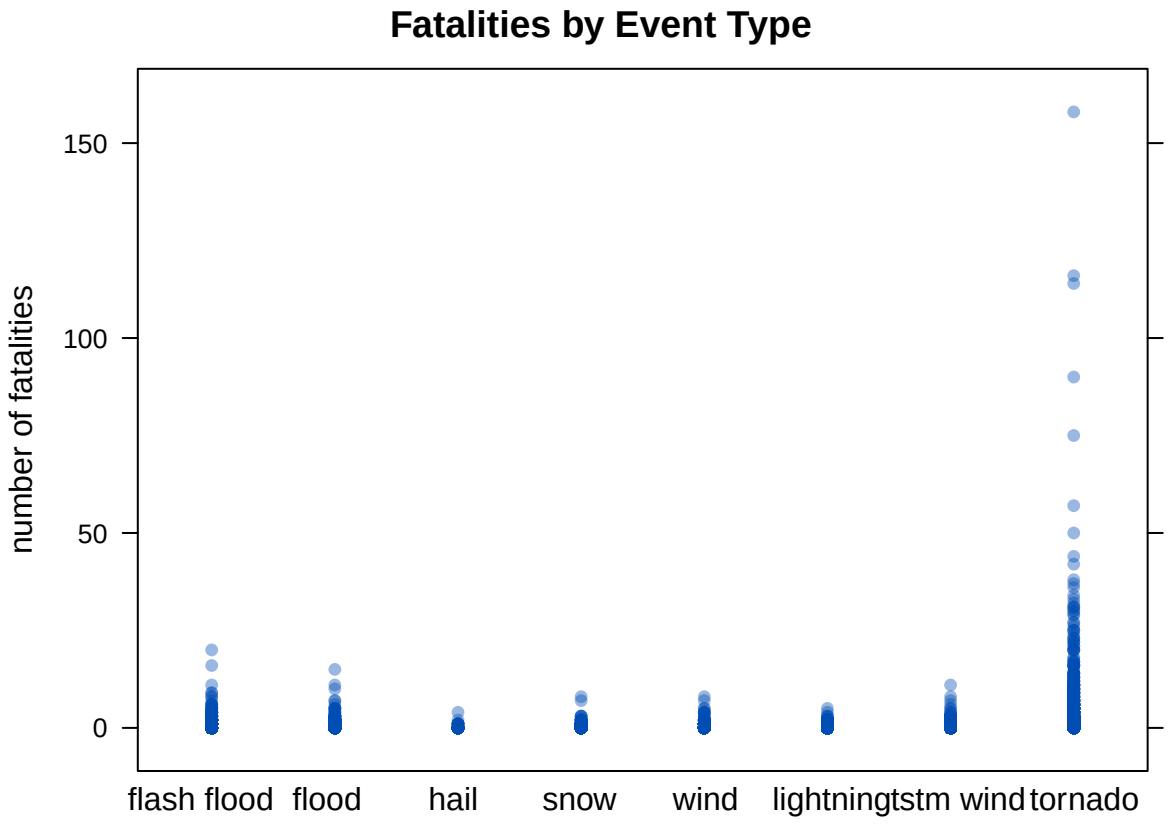
We now look at fatalities by event type. We compute the same table as above.

```
df %>% group_by(EVTYPE) %>% summarize(FAT_MEAN = mean(FATALITIES), FAT_MAX = max(FATALITIES), PROB = mea
```

```
## # A tibble: 8 x 4
##   EVTYPE             FAT_MEAN FAT_MAX      PROB
##   <fct>                 <dbl>   <dbl>     <dbl>
## 1 TORNADO           0.0929        158 0.0264
## 2 LIGHTNING         0.0518          5 0.0482
## 3 FLOOD             0.0186         15 0.0118
## 4 FLASH FLOOD       0.0180         20 0.0117
## 5 HIGH WIND         0.0123          8 0.00905
## 6 HEAVY SNOW        0.00809         8 0.00579
## 7 THUNDERSTORM WIND 0.00217        11 0.00175
## 8 HAIL              0.0000520       4 0.0000381
```

We see that considering fatalities, tornado and lightning again have the highest probabilities and the highest mean.

```
xyplot(FATALITIES ~ EVTYPE, data = df, xlab = c("flash flood", "flood", "hail", "snow", "wind", "lightni
```




Plotting all fatalities, we again see the dominance of tornadoes.

In total, we conclude that tornadoes are by far the most dangerous events with respect to population health. The second event type that should be prepared for with great care are lightnings.

## IMPACT ON ECONOMY

To evaluate the impact of weather events on the economy, we look at the mean and maximum property damage for each event type.

```
df %>% group_by(EVTYPE) %>% summarize(ECO_MEAN = mean(PROPDMG), ECO_MAX = max(PROPDMG), ECO_PROB = mean
```

```
## # A tibble: 8 x 4
##   EVTYPE            ECO_MEAN ECO_MAX ECO_PROB
##   <fct>                <dbl>   <dbl>    <dbl>
## 1 TORNADO              53.0     4410   0.644
## 2 LIGHTNING            38.3     1000   0.653
## 3 FLOOD                35.5     3000   0.383
## 4 FLASH FLOOD          26.2     5000   0.379
## 5 HIGH WIND            16.1     3000   0.264
## 6 THUNDERSTORM WIND     8.22    5000   0.358
## 7 HEAVY SNOW            7.78     950   0.0787
## 8 HAIL                  2.39     970   0.0798
```

We see that tornado and lightning lead to the highest property damage on average, namely 53.000 and 38.300 US Dollars. Flood, flash flood and high wind also have an average damage of more than 10.000 US Dollars. Tornado and lightning are the only events that in most cases have some property damage. But for flood, flash flood, high wind and thunderstorm wind there also is property damage in more than one quarter of the cases.

Thus, as with respect to population health, tornadoes and lightning do the most damage. But in contrast to population health, for all event types an economical damage should be expected.