# Analysis of the Impact of storms and other severe weather events in public health

Abel Gutierrez-Cruz

23/11/2021

## Abstract

Several weather events have a huge impact on the public health and the economy of the population. For this reason, it is important to analyze the behavior of each kind of natural disaster in each state. The outputs will be useful in the decision-making. This analysis has a focus to discover the weather events with the highest impact on public health in each state. To find them variables like injuries and fatalities were considered. Furthermore, the dates with the highest implications in public health also were analyzed.

## Data processing

Once the file was downloaded, the data had read with the help of 'read.csv" method and assigned to the variable "data". The empty string was established like NA value in the data set.

```
setwd("./")
data <- read.csv("./dataCourseProject2.csv", na.strings = "")
```

### Remove useless data

After that, columns with only NA values were determined to their elimination.

```
ColumnsNA <- apply(data, 2, function(x){all(is.na(x))})
data <- data[, !ColumnsNA]
```

To can use the dates in this analysis, date format was given to the BGN_DATE column.

```
data$BGN_DATE <- mdy_hms(data$BGN_DATE)
```

Finally, with the purpose of have a fast analysis, useless columns (to the current study) were removed of the data set.

```
data <- data[, -c(26:36)]
```

### Transform the data to be analyzed

**Data to analyze the severe weather events with the highest impact in each state.** In the first place, the data were grouped by state and weather event. After that, the data about injuries and fatalities were summed for each group. To select the highest values of injuries and fatalities for each state, the data was filtered.

```
dataAnalysis <- data %>% group_by(STATE, EVTYPE) %>%
  summarize(Injuries = sum(INJURIES, na.rm = TRUE), Fatalities = sum(FATALITIES))
```

```
## `summarise()` has grouped output by 'STATE'. You can override using the `.groups` argument.
```

```r
dataInj <- dataAnalysis %>% filter(Injuries == max(Injuries)) %>% filter(Injuries != 0)

dataFat <- dataAnalysis %>% filter(Fatalities == max(Fatalities)) %>% filter(Fatalities != 0)
```

**Data to analyze the date of the most severe weather events.** The data was filtered with a threshold of the 99% quantile in the values of injuries.

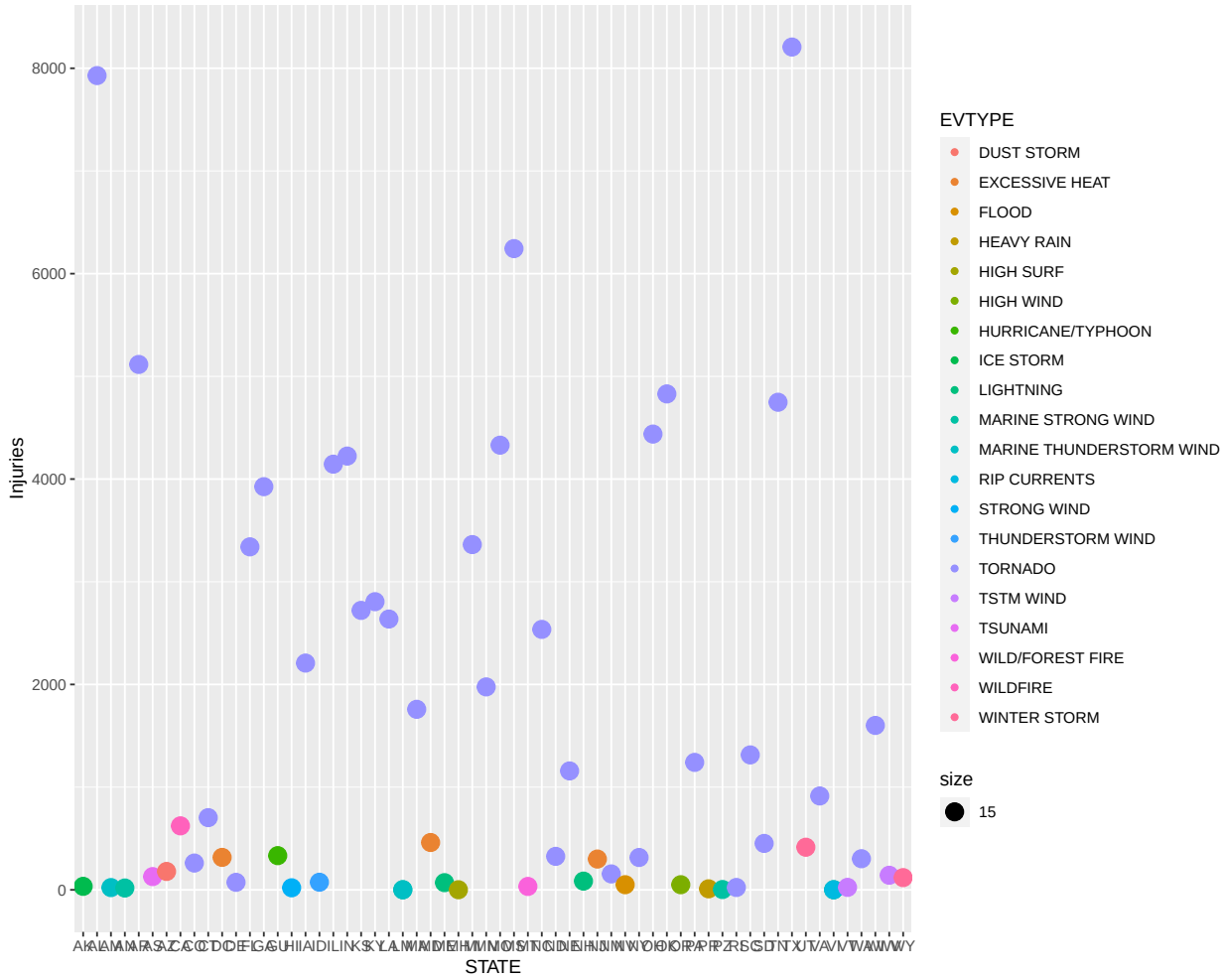
```r
dataAnalysisDates <- data %>% filter(INJURIES > 2) %>%
  filter(INJURIES > quantile(INJURIES, probs = 0.99))
```

## Results

Dismissing the values equal to 0, we can see that the weather events that caused more injuries in the population in US during all the time in monitoring were the tornadoes. They also were the most common events.

Figure 1. Natural disasters with the highest number of injuries in each state in the US.

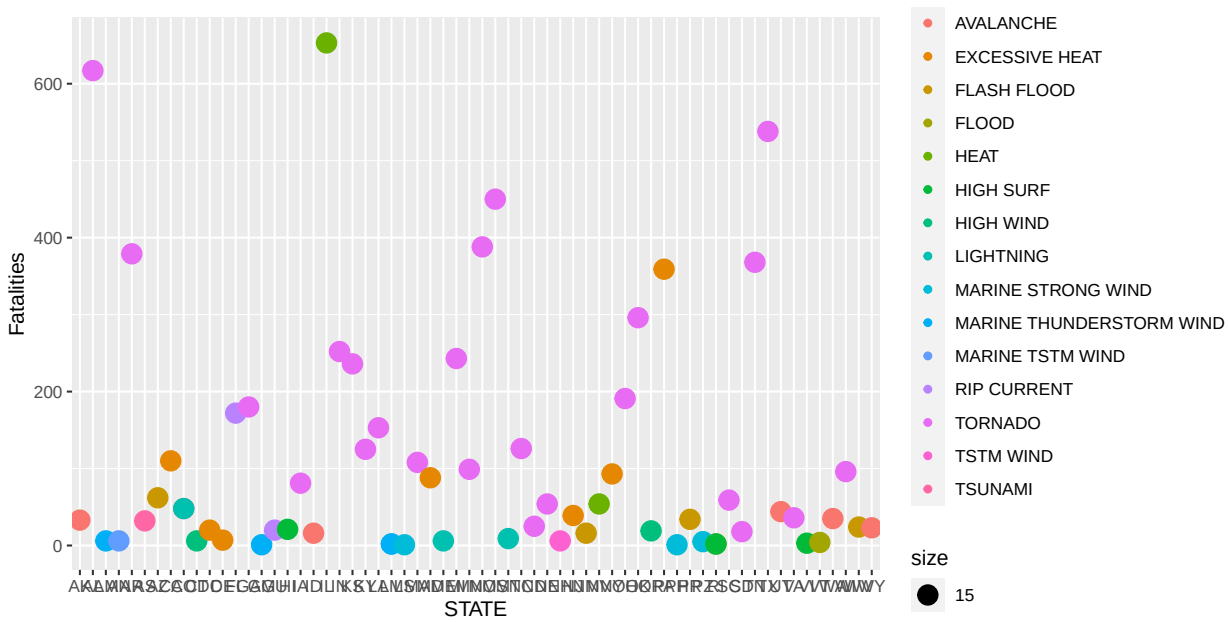
```r
ggplot(data = dataInj, mapping= aes(x=STATE, y=Injuries, color=EVTYPE, size=15)) + geom_point()
```




On the other hand, if we consider the weather events that caused more fatalities in the population we also can see marine thunderstorm wind and avalanches.

Figure 2. Natural disasters with the highest number of fatalities in each state in the US.

```
ggplot(data = dataFat, mapping= aes(x=STATE, y=Fatalities, color=EVTYPE, size=15)) + geom_point()
```




Furthermore, the dates when the highest number of natural disasters happened in the US were in the 70's, 90's and 2000.

Figure 3. Number the injuries vs date.

```
ggplot(data = dataAnalysisDates, mapping = aes(x=BGN_DATE, y=INJURIES)) + geom_line(color="red")
```