

The U.S. NOAA Storm Data Analysis by type of harmful events, and by events with greatest economic consequences

Liliana

1/27/2019

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.

The goal of this report is to answer to the following questions about severe weather events:

- (i) Across the United States, which types of events (as indicated in the *EVTYPE* variable) are most harmful with respect to population health?
- (ii) Across the United States, which types of events have the greatest economic consequences?

This analysis shows that the tornadoes are the most harmful events for population health in the United States, floods caused worst damages for properties, and hurricane caused big damages on crops.

Data Processing

The data for this project are in the form of a comma-separated-value file compressed via the bzip2 algorithm. Data can be downloaded from the Storm Data. In addition, there is some documentation available at the National Weather Service Storm Data Documentation where there are provided details how some of the variables are constructed/defined.

First, clear R environment and load necessary libraries.

```
rm(list=ls())  
library(dplyr); library(knitr); library(ggplot2); library(tidyr); library(car)
```

Download the file.

```
if(!file.exists("./StormData")){dir.create("./StormData")}  
  
fileUrl <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
download.file(fileUrl, destfile = "./StormData/repdataFdataFStormData.csv.bz2", method = "curl", mode =
```

Read the data.

```
storm <- read.csv(bzfile("./StormData/repdataFdataFStormData.csv.bz2"),  
                 header = TRUE, sep = ",", stringsAsFactors = FALSE, na.strings="NA")
```

Extract variables which are relevant to our analysis:

```
relevant.storm.data <- storm %>%
  select(STATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)
str(relevant.storm.data)
```

```
## 'data.frame':    902297 obs. of  8 variables:
## $ STATE      : chr  "AL" "AL" "AL" "AL" ...
## $ EVTYPE     : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
## $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
## $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
## $ PROPDMGEXP: chr   "K" "K" "K" "K" ...
## $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
## $ CROPDMGEXP: chr   "" "" "" "" ...
```

Rename variables as follow:

- STATE (US state abbreviation) -> State
- EVTYPE (Type of event) -> Event
- FATALITIES (Number of fatalities) -> Fatalities
- INJURIES (Number of injuries) -> Injuries
- PROPDMG (Amount of property damage in orders of magnitude) -> Property.Damage.Magnitude
- PROPDMGEXP (Order of magnitude for property damage in \$H, K, M, B) -> Property.Damage.Expenses
- CROPDMG (Amount of crop damage in orders of magnitude) -> Crop.Damage.Magnitude
- PROPDMGEXP (Order of magnitude for crop damage in \$H, K, M, B) -> Crop.Damage.Expenses

where abbreviations H, K, M, B represent H = hundreds, K = thousands, M = millions, B = billions.

```
names(relevant.storm.data) <- c("State", "Event", "Fatalities", "Injuries", "Property.Damage.Magnitude")
```

To avoid duplicates and various problems due to mixt uppercase/lowercase, spelling errors and events with many sub-categories, renames are performed for the events.

```
relevant.storm.data$Event <- gsub("^(HEAT).*", "Heat", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(RECORD HEAT).*", "Heat", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(EXTREME HEAT).*", "Heat", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(Heat).*", "Heat", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(EXCESSIVE HEAT).*", "Heat", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(TSTM).*", "Thunder.Storm", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(THUNDERSTORM).*", "Thunder.Storm", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(LIGHTNING).*", "Lightning", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(TROPICAL STORM).*", "Tropical.Storm", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(FLASH FLOOD).*", "Flood", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(FLOOD).*", "Flood", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(WIND).*", "WIND", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(STRONG WIND).*", "Wind", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HIGH WIND).*", "Wind", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HURRICANE).*", "Hurricane", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(SNOW).*", "Snow", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HEAVY SNOW).*", "Snow", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(FIRE).*", "Fire", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(WILD/FOREST FIRE).*", "Fire", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(WILDFIRE).*", "Fire", relevant.storm.data$Event)
```

```

relevant.storm.data$Event <- gsub("^(WILD FIRES).*", "Fire", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HAIL).*", "Hail", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(BLIZZARD).*", "Blizzard", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(COLD).*", "Cold", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(WINTER WEATHER).*", "Cold", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(EXTREME COLD).*", "Cold", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(RIP).*", "Rip", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(FOG).*", "Fog", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(DENSE FOG).*", "Fog", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(AVALANCHE).*", "Avalanche", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(AVALANCHE).*", "Avalanche", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(RAIN).*", "Rain", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HEAVY RAIN).*", "Rain", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HIGH SURF).*", "Surf", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(HEAVY SURF).*", "Surf", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(SURF).*", "Surf", relevant.storm.data$Event)
relevant.storm.data$Event <- gsub("^(TORNADO).*", "Tornado", relevant.storm.data$Event)

```

Results

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

Determine how many casualties (injuries + fatalities) were caused by type of event, and arrange them in descending order:

```

casualties <- relevant.storm.data %>%
  group_by(Event) %>%
  mutate(Total = Fatalities + Injuries) %>%
  summarize(Fatalities = sum(Fatalities),
            Injuries = sum(Injuries),
            Total = sum(Total)) %>%
  arrange(desc(Total))
casualties[1:6, ]

```

```

## # A tibble: 6 x 4
##   Event      Fatalities Injuries Total
##   <chr>          <dbl>    <dbl> <dbl>
## 1 Tornado         5658     91364 97022
## 2 Heat            3119      9224 12343
## 3 Thunder.Storm    710      9508 10218
## 4 Flood           1513      8591 10104
## 5 Lightning        817      5232  6049
## 6 Wind            403      1772  2175

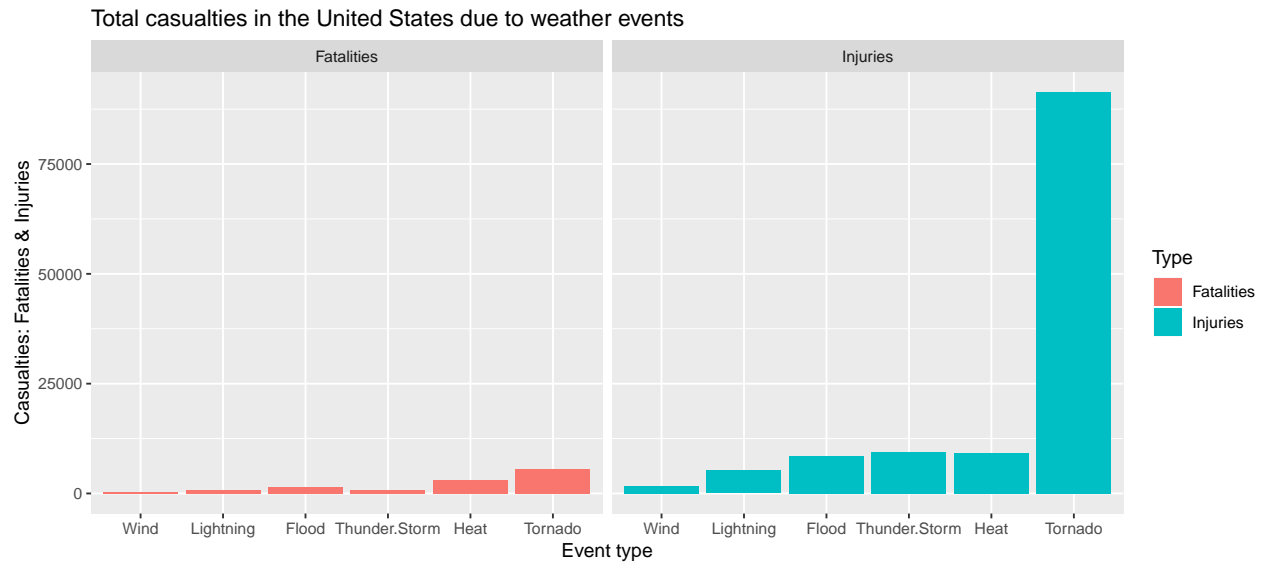
```

Plot the casualties (injuries & fatalities) by type of event:

```

casualties <- gather(casualties[1:6, 1:3], Type, Total, Fatalities:Injuries)
plot1 <- ggplot(casualties, aes(x = reorder(Event, +Total), y = Total, fill = Type)) +
  geom_bar(stat = "identity") +
  facet_grid(.~Type, scales = "free", space="free") +
  labs(title = "Total casualties in the United States due to weather events",
       x = "Event type", y = "Casualties: Fatalities & Injuries")
plot1

```



This analysis shows clearly that tornadoes are the most harmful events for population health in the United States, causing the highest number of fatalities and injuries.

2. Across the United States, which types of events have the greatest economic consequences?

To determine numerical value of damages for both properties and crops, expenses are converted from character to numeric:

```
relevant.storm.data$Property.Damage.Expenses <- as.numeric(recode(as.character(relevant.storm.data$Property.Damage.Expenses),
  "'0'=1; '1'=10; '2'=10^2; '3'=10^3; '4'=10^4; '5'=10^5; '6'=10^6; '7'=10^7; '8'=10^8; 'b'=10^9; 'h'=10^2; 'k'="))
relevant.storm.data$Crop.Damage.Expenses <- as.numeric(recode(as.character(relevant.storm.data$Crop.Damage.Expenses),
  "'0'=1; '1'=10; '2'=10^2; '3'=10^3; '4'=10^4; '5'=10^5; '6'=10^6; '7'=10^7; '8'=10^8; 'b'=10^9; 'h'=10^2; 'k'="))
```

Calculate amount of the property and crop damages by type of event, and arrange them in descending order:

```
damages <- relevant.storm.data %>%
  group_by(Event) %>%
  summarize(Property = sum(Property.Damage.Magnitude * Property.Damage.Expenses, na.rm = TRUE),
             Crop = sum(Crop.Damage.Magnitude * Crop.Damage.Expenses, na.rm = TRUE),
             Total = Property + Crop) %>%
  arrange(desc(Total))
damages[1:6, ]
```

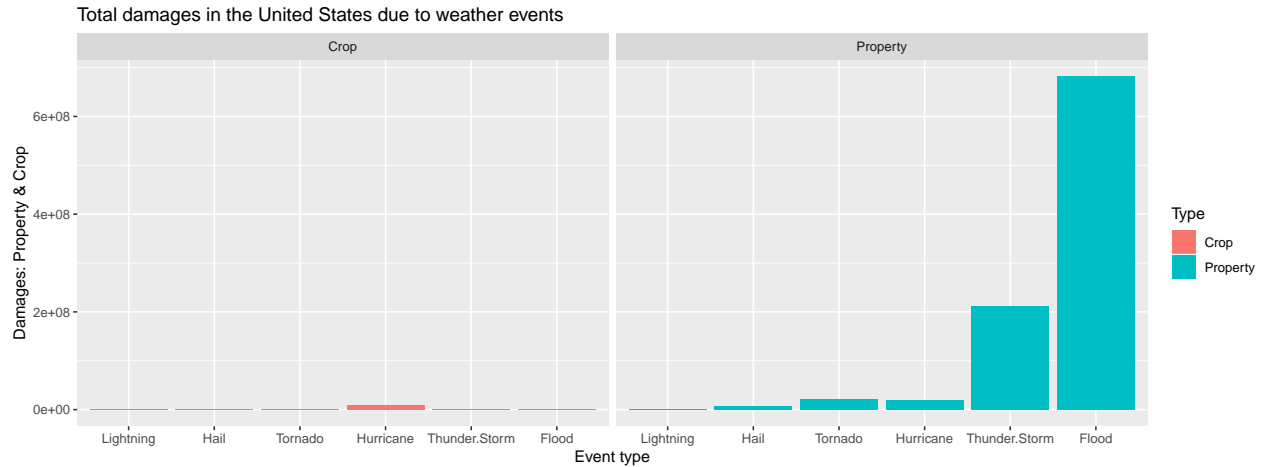
```
## # A tibble: 6 x 4
##   Event      Property      Crop      Total
##   <chr>      <dbl>      <dbl>      <dbl>
## 1 Flood      682475270.    10000 682485270.
## 2 Thunder.Storm 211137443.    4080 211141523.
## 3 Hurricane    20000000 10000000 30000000
## 4 Tornado      21720194.     160 21720354.
## 5 Hail         7900236.   417020  8317256.
## 6 Lightning    1720146.      0  1720146.
```

Plot damages for properties and crops by type of event:

```
damages<- gather(damages[1:6, 1:3], Type, Total, Property:Crop)
plot2 <- ggplot(damages, aes(x = reorder(Event, +Total), y = Total, fill = Type)) +
```

```
geom_bar(stat = "identity") +
facet_grid(~Type, scales = "free", space="free") +
labs(title = "Total damages in the United States due to weather events",
x = "Event type", y = "Damages: Property & Crop")
```

plot2



This analysis shows clearly that floods caused worst damages for properties, while hurricane caused big damages on crops. Overall floods represent the worst economic consequences damages in the United States.