C5W4 Assignment

Nic Neo

8 July 2020

NOAA Storm Database Analysis

Synopsis

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 basic goal of this assignment is to explore the NOAA Storm Database and answer the following questions:

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

Data

The data for this assignment comes in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file here.

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

Data Processing

```
#Load packages
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```r
# Get data from the web
URLFile <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
filename <- "repdata_data_StormData.csv"
if (!file.exists(filename)) {
  download.file(URLFile, filename, method = "curl")
}
# Read data
data <- read.csv(bzfile(filename), header = TRUE)
# Extract only the relevant columns
data <- data[, c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG",
           "CROPDMGEXP")]
# Convert event strings to uppercase in the event where the same event types are not spelled correctly
data <- mutate_at(data, "EVTYPE", toupper)
```

Results

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

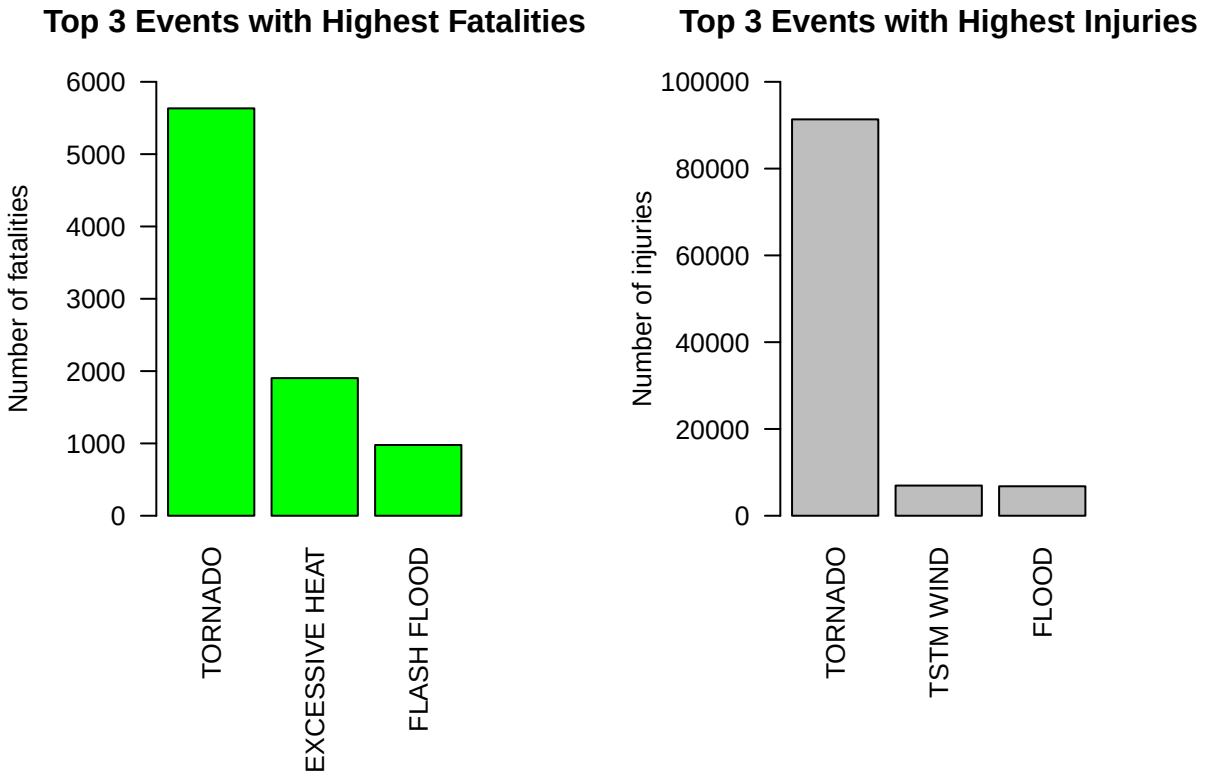
Calculating the total fatalities and injuries per event

```r
options(scipen=999)
# Calculate total fatalities per event
fatalitiesTotal <- data %>%
  group_by(EVTYPE) %>%
  summarize(sum(FATALITIES))
colnames(fatalitiesTotal) = c("event", "fatalities")
# Order and subset only to top 3 events
fatalitiesTotal <- fatalitiesTotal[order(fatalitiesTotal$fatalities, decreasing = TRUE), ]
fatalTop3 <- fatalitiesTotal[1:3, ]
# Calculate total injuries per event
injuriesTotal <- data %>%
  group_by(EVTYPE) %>%
  summarize(sum(INJURIES))
colnames(injuriesTotal) = c("event", "injuries")
# Order and subset only to top 5 events
injuriesTotal <- injuriesTotal[order(injuriesTotal$injuries, decreasing = TRUE), ]
injuriesTop3 <-injuriesTotal[1:3, ]
```

Rank top 3 most harmful events with a barplot

```r
par(mfrow = c(1,2), mar = c(10, 6, 4, 4), mgp = c(4, 1, 0), cex = 0.8)
barplot(fatalTop3$fatalities,
        names.arg = fatalTop3$event,
        main = "Top 3 Events with Highest Fatalities",
        ylab = "Number of fatalities",
        ylim = range(pretty(c(0, fatalTop3$fatalities))),
        las = 2,
        col = "green"
        )
barplot(injuriesTop3$injuries,
        names.arg = injuriesTop3$event,
```

```
        main = "Top 3 Events with Highest Injuries",
        ylab = "Number of injuries",
        ylim = range(pretty(c(0, injuriesTop3$injuries))),
        las = 2,
        col = "grey"
        )
```



As shown in the graph above, tornado, excessive heat, flash flood are the most harmful in terms of fatalities. Meanwhile, tornado, tstm wind, flood are the most harmful in terms of injuries.

In general, according to both the plots, tornadoes are the most harmful with respect to population health.

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

Examining the economic data

```
unique(data$PROPDMGEXP)
```

```
##  [1] K M   B m + 0 5 6 ? 4 2 3 h 7 H - 1 8
## Levels:  - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
```

```r
unique(data$CROPDMGEXP)
```

```
## [1]   M K m B ? 0 k 2
## Levels:  ? 0 2 B k K m M
```

Cleaning of the economic data

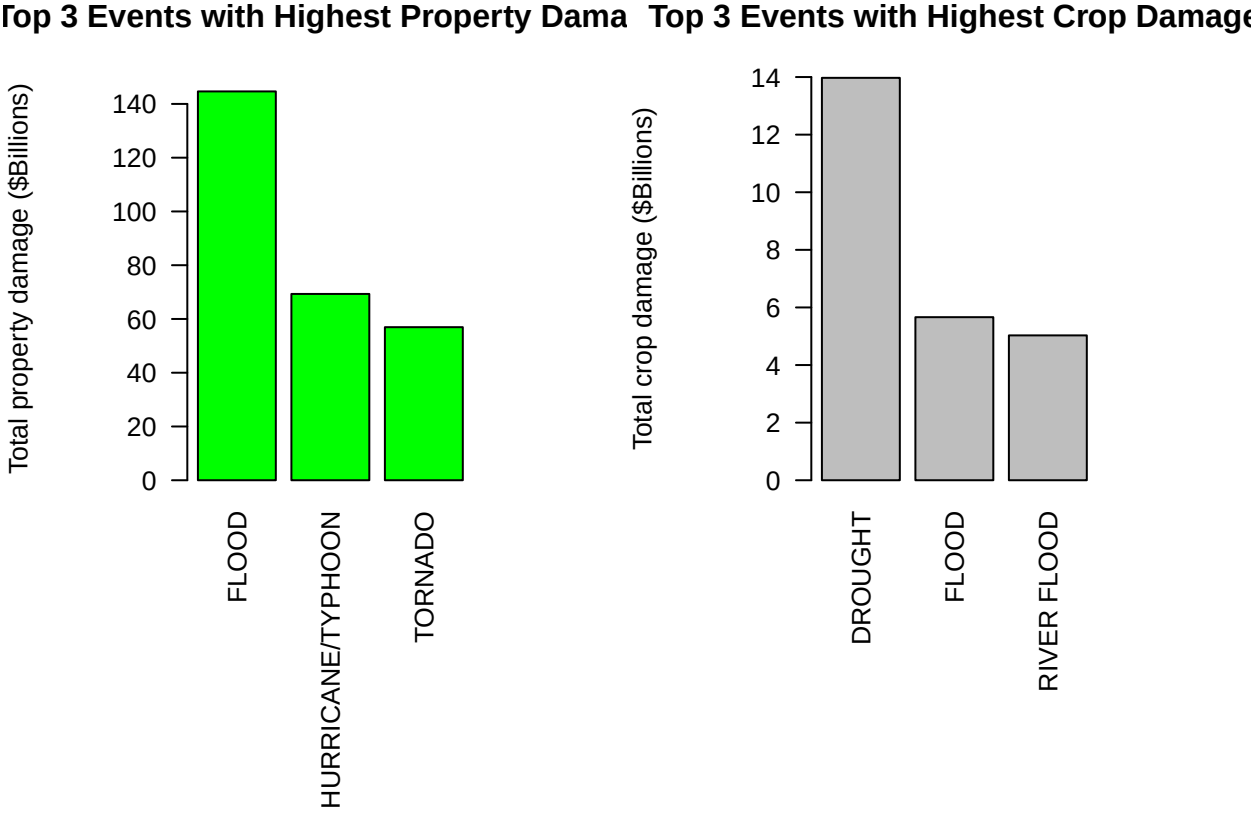
```r
# Change all to uppercase letters
data <- mutate_at(data, "PROPDMGEXP", toupper)
data <- mutate_at(data, "CROPDMGEXP", toupper)
# Replace symbols with 0
data$PROPDMGEXP[data$PROPDMGEXP %in% c("", "+", "-", "?")] = "0"
data$CROPDMGEXP[data$CROPDMGEXP %in% c("", "+", "-", "?")] = "0"
# Convert letters to numeric exponents
data$PROPDMGEXP[data$PROPDMGEXP %in% "B"] = "9"
data$PROPDMGEXP[data$PROPDMGEXP %in% "M"] = "6"
data$PROPDMGEXP[data$PROPDMGEXP %in% "K"] = "3"
data$PROPDMGEXP[data$PROPDMGEXP %in% "H"] = "2"
data$PROPDMGEXP <- 10^(as.numeric(data$PROPDMGEXP))
data$CROPDMGEXP[data$CROPDMGEXP %in% "B"] = "9"
data$CROPDMGEXP[data$CROPDMGEXP %in% "M"] = "6"
data$CROPDMGEXP[data$CROPDMGEXP %in% "K"] = "3"
data$CROPDMGEXP[data$CROPDMGEXP %in% "H"] = "2"
data$CROPDMGEXP <- 10^(as.numeric(data$CROPDMGEXP))
# Multiply damage by corresponding exponent and add to new columns
data$TOTAL.PROPDMG <- data$PROPDMG * data$PROPDMGEXP
data$TOTAL.CROPDMG <- data$CROPDMG * data$CROPDMGEXP
```

Total property damage and total crop damage per event

```r
# Calculate total property damage per event
totalpropdmg <- data %>%
  group_by(EVTYPE) %>%
  summarize(sum(TOTAL.PROPDMG))
colnames(totalpropdmg) = c("event", "propdmg")
totalpropdmg$propdmg <- totalpropdmg$propdmg / 10^9 #Scale to billions
# Top 3 events
totalpropdmg <- totalpropdmg[order(totalpropdmg$propdmg, decreasing = TRUE), ]
propTop3 <- totalpropdmg[1:3, ]
# Calculate total crop damage per event
totalcropdmg <- data %>%
  group_by(EVTYPE) %>%
  summarize(sum(TOTAL.CROPDMG))
colnames(totalcropdmg) = c("event", "cropdmg")
totalcropdmg$cropdmg <- totalcropdmg$cropdmg / 10^9 #Scale to billions
# Top 3 events
totalcropdmg <- totalcropdmg[order(totalcropdmg$cropdmg, decreasing = TRUE), ]
cropTop3 <- totalcropdmg[1:3, ]
```

Create plot of the 3 most harmful events

```r
par(mfrow = c(1,2), mar = c(11, 6, 4, 5), mgp = c(5, 1, 0), cex = 0.8)
barplot(propTop3$propdmg,
        names.arg = propTop3$event,
        main = "Top 3 Events with Highest Property Damage",
        ylab = "Total property damage ($Billions)",
        ylim = range(pretty(c(0, propTop3$propdmg))),
        las = 2,
        col = "green"
        )
barplot(cropTop3$cropdmg,
        names.arg = cropTop3$event,
        main = "Top 3 Events with Highest Crop Damage",
        ylab = "Total crop damage ($Billions)",
        ylim = range(pretty(c(0, cropTop3$cropdmg))),
        las = 2,
        col = "grey"
        )
```



As shown in the graph above, flood, hurricane/typhoon, tornado have the greatest economic consequences in terms of property damage. Meanwhile, drought, flood, river flood have the greatest economic consequences in terms of crop damage.