Analysis of the NOAA Storm Database

Gilda M Castro

7/8/2020

Synopsis

This is the Coursera assignment of *Reproducible Research* and its description is "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."

We are analysing the data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database to 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?

The result of the analysis shows that tornadoes are the most harmful event type with the greates economic consequences in the US.

Data Processing

The data has been downloaded previously into the local machine. The information is loaded in the variable *storm_data*.

```
storm_data <- read.csv("repdata_data_StormData.csv")
storm_data$UPPER_EVTYPE <- toupper(storm_data$EVTYPE)
```

We need to load *ggplot2* and *dplyr*.

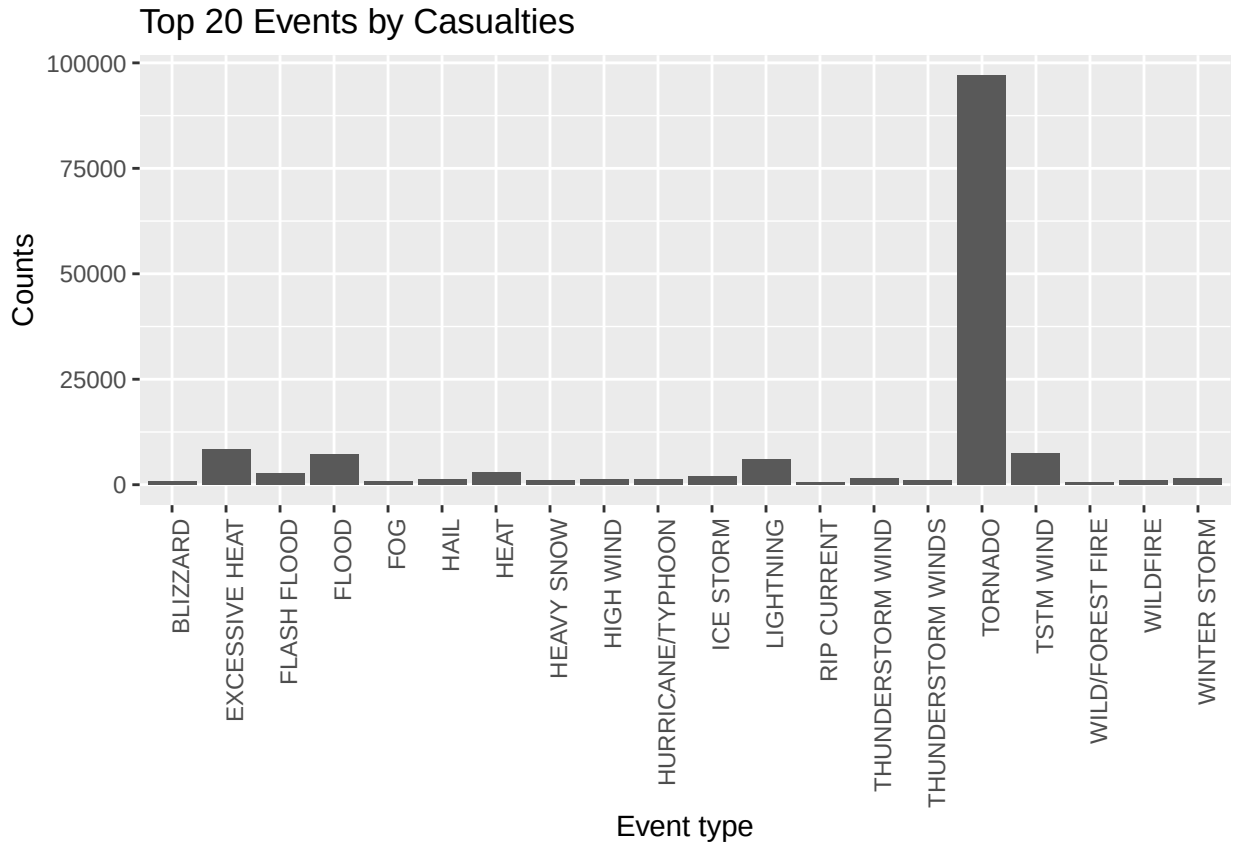
```
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

The data provides fatalities (FATALITIES in the data) and injuries (INJURIES in the data). I will consider these as most harmfull with respect to population health. The calculation of the casualties is the additoin of fatalities and injuries.

```
events <- aggregate(list(fatalities=storm_data$FATALITIES,injuries=storm_data$INJURIES) , by= list(even
events_ordered <- events[order( -(events$fatalities+events$injuries)),]
events_top20 <- head(events_ordered, 20)
events_top20$casualties <- events_top20$fatalities+events_top20$injuries

plot_casualties <- ggplot(events_top20, aes(x=event, y=casualties))

plot_casualties + geom_bar(stat="identity") + labs(title="Top 20 Events by Casualties", x= "Event type"
```



Summarizing the information, we can denote that the Tornadoes are the most dangerous events in the US as they affect the most the Health of the persons due to the fact that they produce the most casualties.

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

We need to understand the economic consequences of the damages caused by the events. In this case, the documentation: https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf, page 12, states that *damage* is measured in dollars.

The datase doesn't have a dollar amount per se, but it has four columns referring to *damage* which are: PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP

The EXP columns specify the unit, like K for Thousands, B for Billion, M for Millions.

```
storm_data_economic <- storm_data[c("UPPER_EVTYPE", "CROPDMG", "CROPDMGEXP", "PROPDMG", "PROPDMGEXP")]
#Set the values to lower case, in case there are different
storm_data_economic$PROPDMGEXP = tolower(storm_data_economic$PROPDMGEXP)
storm_data_economic$CROPDMGEXP = tolower(storm_data_economic$CROPDMGEXP)

exponent <- function(units) {
  if (units == 'k') 1000
  else if (units == 'm') 1000000
  else if (units == 'b') 1000000000
  else 1
}

#Setting default values in empty cells
storm_data_economic$CROPDMG[is.na(storm_data_economic$CROPDMG)] <- 0
storm_data_economic$PROPDMG[is.na(storm_data_economic$PROPDMG)] <- 0

#Adding a Value column
storm_data_economic$CROPDMGEXP_VAL <- exponent(storm_data_economic$CROPDMGEXP)
```

```
## Warning in if (units == "k") 1000 else if (units == "m") 1e+06 else if (units
## == : the condition has length > 1 and only the first element will be used

## Warning in if (units == "m") 1e+06 else if (units == "b") 1e+09 else 1: the
## condition has length > 1 and only the first element will be used

## Warning in if (units == "b") 1e+09 else 1: the condition has length > 1 and only
## the first element will be used
```

```
storm_data_economic$PROPDMGEXP_VAL <- exponent(storm_data_economic$PROPDMGEXP)
```

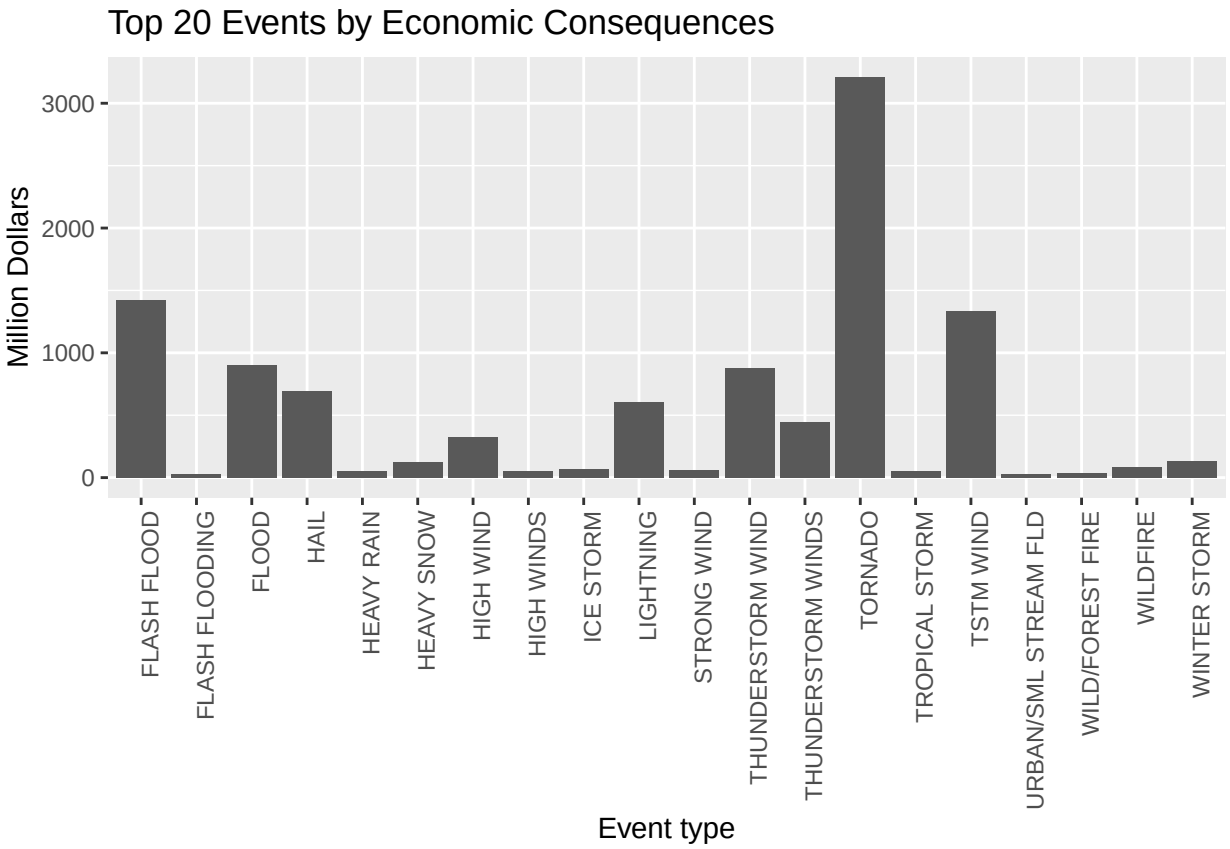
```
## Warning in if (units == "k") 1000 else if (units == "m") 1e+06 else if (units
## == : the condition has length > 1 and only the first element will be used
```

```
#Calculate the damage value in dollars
storm_data_economic$CROPDMG_VAL = storm_data_economic$CROPDMG * storm_data_economic$CROPDMGEXP_VAL
storm_data_economic$PROPDMG_VAL = storm_data_economic$PROPDMG * storm_data_economic$PROPDMGEXP_VAL
```

Now we can aggregate the data by event type. The damage is measured by adding cropdmg and propdmg.

```
events_economic <- aggregate(list(crop_dmg=storm_data_economic$CROPDMG_VAL,prop_dmg=storm_data_economic$

events_economic_ordered <- events_economic[order( -(events_economic$crop_dmg+events_economic$prop_dmg)),]
events_economic_top20 <- head(events_economic_ordered, 20)

events_economic_top20$damage <- events_economic_top20$crop_dmg + events_economic_top20$prop_dmg

events_economic_top20$damage_M <- (events_economic_top20$crop_dmg + events_economic_top20$prop_dmg) / 1
```

```
plot_damage <- ggplot(events_economic_top20, aes(x=event, y=damage_M))

plot_damage + geom_bar(stat="identity") + labs(title="Top 20 Events by Economic Consequences", x= "Even
```



Results

Tornadoes are the most harmful event type with respect to population health and also have the greatest economic consequences in the US.