Finding the events that most affect property and health

vinod

18/05/2020

#Introduction 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 basic goal of this assignment is to explore the NOAA Storm Database and answer two basic questions about severe weather events.
1.Across the United States, which types of events are most harmful with respect to population health?
2.Across the United States, which types of events have the greatest economic consequences?

##Downloading and reading the data and loading the required libraries

```
#url = 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
#download.file(url, 'data.csv', 'curl')
data = read.csv('data.csv')
library(dplyr)
library(ggplot2)
library(tidyr)
```

##Selecting the only features which are required

```
data1 = data %>%
    select(EVTYPE, MAG, FATALITIES, INJURIES,
            PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)
```

##Data Cleaning There will be two types of damages due to severe events human and material damages.
Human damages in this dataset is measured in terms of 'FATALITIES' and 'INJURIES'.
Material damages in this dataset is measured in terms of 'property damage' and 'crop damage'.

Value of the material damage (for each property and crop) is given in two separate columns, one column contains only face value of the damage and the other column contains the multiple(in terms of 'k', 'm' and 'b' etc) of the face value. So in order to get the actual value we have to multiply face value with its multiple.

```r
#calculating the multiple value of the property damage
data1[!tolower(data1$PROPDMGEXP) %in% c('k','m','b'), 'PROPDMGEXP'] = 0
data1$PROPDMGEXP = gsub('k', 1000, x = tolower(data1$PROPDMGEXP))
data1$PROPDMGEXP = gsub('m', 1000000, x = tolower(data1$PROPDMGEXP))
data1$PROPDMGEXP = gsub('b', 1000000000, x = tolower(data1$PROPDMGEXP))
data1$PROPDMGEXP = as.numeric(data1$PROPDMGEXP)
#calculating the actual value of the property damage
data1$propdmgexp = data1$PROPDMG * data1$PROPDMGEXP
#removing the features that are not required
data1$PROPDMG = NULL
data1$PROPDMGEXP = NULL

#calculating the multiple value of the crop damage
data1[!tolower(data1$CROPDMGEXP) %in% c('k','m','b'), 'CROPDMGEXP'] = 0
data1$CROPDMGEXP = gsub('k', 1000, x = tolower(data1$CROPDMGEXP))
data1$CROPDMGEXP = gsub('m', 1000000, x = tolower(data1$CROPDMGEXP))
data1$CROPDMGEXP = gsub('b', 1000000000, x = tolower(data1$CROPDMGEXP))
data1$CROPDMGEXP = as.numeric(data1$CROPDMGEXP)
#calculating the actual value of the crop damage
data1$cropdmgexp = data1$CROPDMG * data1$CROPDMGEXP
#removing the features that are not required
data1$CROPDMG = NULL
data1$CROPDMGEXP = NULL
```

Answering the question-which types of events have the greatest health consequences?

Let us find out the total fatalities, total injuries and total health damage(fatalities+injuries) due to each type of events in all years combined.

```r
r1 = data1 %>% group_by(EVTYPE) %>%
    summarise(total_fatalities = sum(FATALITIES, na.rm = T),
              total_injuries = sum(INJURIES, na.rm = T)) %>%
    mutate(total_health_damage = total_fatalities + total_injuries)
```

Now, we will find out the event that has resulted in maximum damage for each of the damage types(fatalities, injuries, fatalities+injuries)

```r
r2 = gather(r1, key = 'damagetype', value = 'damagevalue', -EVTYPE)
r3 = r2 %>% group_by(damagetype) %>% summarise(damagevalue = max(damagevalue))
r3 = as.data.frame(r3)
r4 = r2[(r2$damagetype %in% r3[,'damagetype']) & (r2$damagevalue %in% r3[,'damagevalue']),]
r4 = as.data.frame(r4)
#removing the objects that are not required
#rm(r1, r2, r3)
```

Plotting the graph to answer the question number 1.

```
ggplot(data = r4, aes(x = damagetype, y = damagevalue))+
    geom_point() +
    geom_text(aes(label = paste(EVTYPE, damagevalue, sep = '-')),
            vjust = 1.3) +
    xlab('type of health damage') +
    ylab('No of persons affected') +
    ggtitle('Event with max damage in health, fatalities, fatalities+health')
```

Event with max damage in health, fatalities, fatalities+health
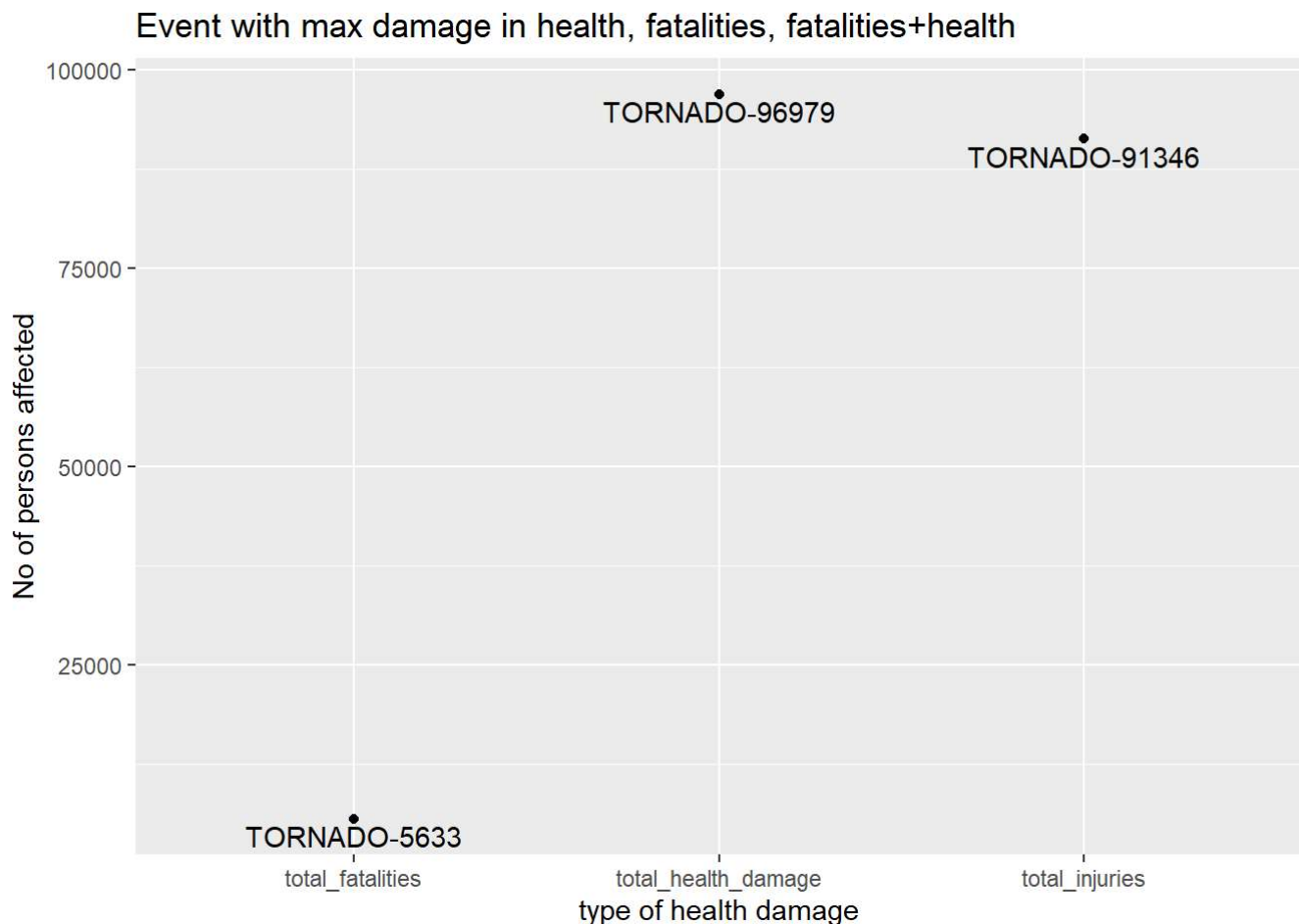
As we can see from the above plot, the event that has caused maximum fatalites with 5633 persons , maximum injuries with 91346 people getting injuried and maximum number of fatalities+injuries combined with 96979 persons affected is 'TORNADO'.

##Answering the question-which types of events have the greatest economic consequences?
Let us find out the total property damage, total crop damage and total damage(property+crop) due to each type of events in all years combined.

```
q1 = data1 %>% group_by(EVTYPE) %>%
    summarise(total_property_damage = sum(propdmgexp, na.rm = T),
            total_crop_damage = sum(cropdmgexp, na.rm = T)) %>%
    mutate(total_damage = total_property_damage + total_crop_damage)
```

Now, we will find out the event that has resulted in maximum damage for each of the damage types(property, crops, property+crops)

```
q2 = gather(q1, key = 'damagetype', value = 'damagevalue', -EVTYPE)
q3 = q2 %>% group_by(damagetype) %>% summarise(damagevalue = max(damagevalue))
q3 = as.data.frame(q3)
q4 = q2[(q2$damagetype %in% q3[,'damagetype']) & (q2$damagevalue %in% q3[,'damagevalue']),]
#converting the value in to Billion USD
q4$damagevalue = round(q4$damagevalue/10^9, 1)
q4 = as.data.frame(q4)
#removing the objects that are not required
#rm(q1, q2, q3)
```
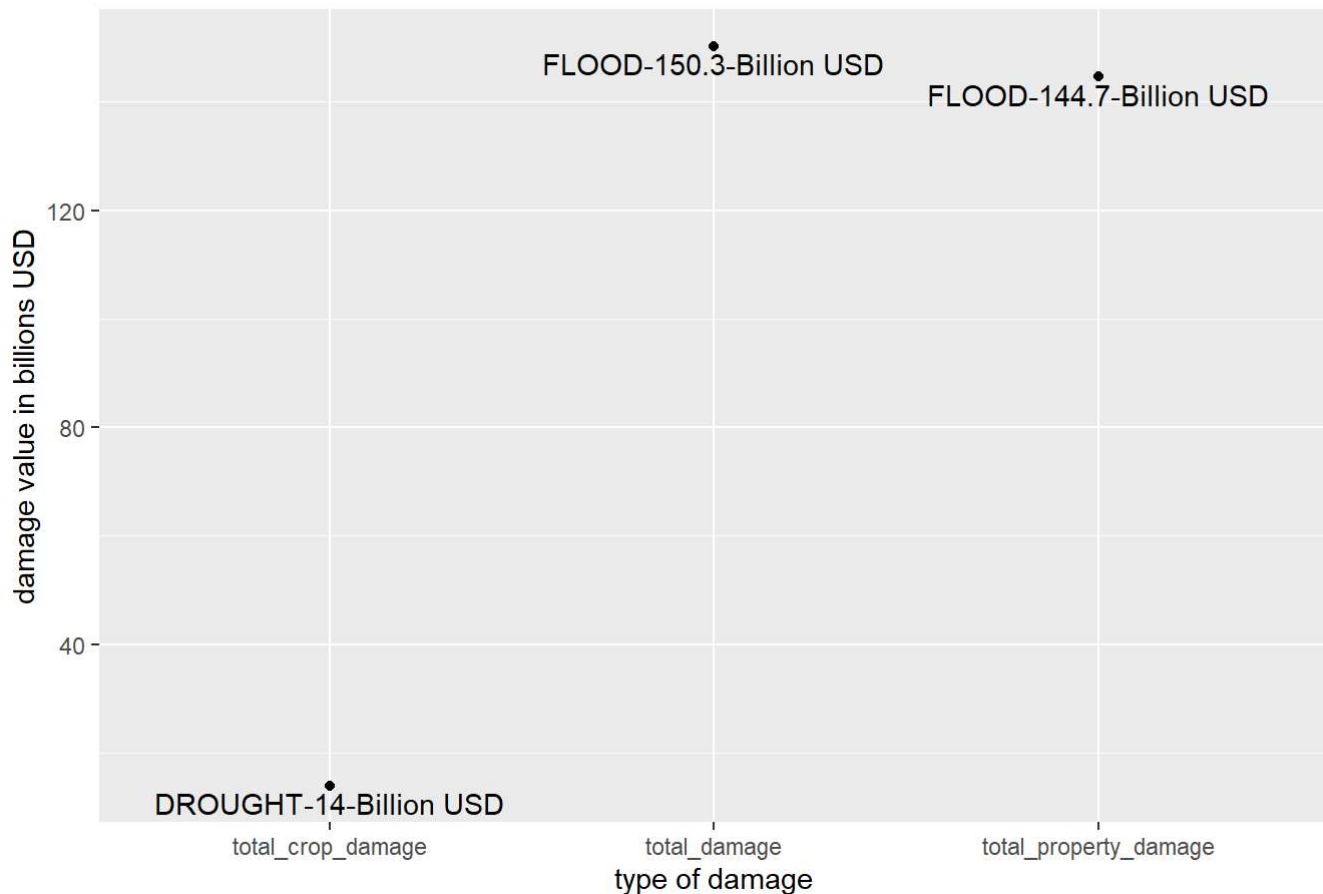
Plotting the graph to answer the question number 2.

```
ggplot(data = q4, aes(x = damagetype, y = damagevalue))+
      geom_point() +
      geom_text(aes(label = paste(EVTYPE, damagevalue, 'Billion USD', sep = '-')), vjust = 1.3)
  +

      xlab('type of damage') +
      ylab('damage value in billions USD') +
      ggtitle('Event with max damage in property, crops, property+crops')
```


As we can see from the above plot, the event which has caused maximum property damage with value of 144.7 billion USD is 'FLOOD', the even which has caused maximum crop damage with value of 14 billion USD is 'DROUGHT', and the event which has caused maximum total damage(property+crops) is also 'FLOOD'.