Exploring impact of various events w.r.t economy and human health in United States using U.S.(NOAA) storm database

Doris Muriungi

12/13/2020

```
knitr::opts_chunk$set(fig.path='Figs/')
```

__Synopsis__

The analysis focusses on the determiantion of the most harmful events to human health and those with the highest economic consequences in the United States.Data for the analysis was obtained from U.S.National Oceanic and Atmospheric Administration's (NOAA) storm database.Data in National Climatic Data Center (NCD) is received from the National Weather Service. The National Weather service receives their information from a variety of sources, which include but are not limited to: county, state and federal emergency management officials, local law enforcement officials, skywarn spotters, NWS damage surveys, newspaper clipping services, the insurance industry and the general public.

__Data Processing__

```r
#Loading libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

Loading storm data

```
#Reading the data
url<-
"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url,destfile = "C:/Users/DMuriungi/Desktop/Doris/Doris
DOCs/dm/Data Science/Reproducible Research/RepR_Project2/StormData.csv.bz2")
storm_data<-read.csv(bzfile("C:/Users/DMuriungi/Desktop/Doris/Doris
DOCs/dm/Data Science/Reproducible Research/RepR_Project2/StormData.csv.bz2"))
str(storm_data)

## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951
0:00:00" "6/8/1951 0:00:00" ...
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ 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  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
```

```
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

Most harmful event types with respect to population health in the United States

The variables considered in the analysis are as follows: -Population health related variables are **Fatalities** and **Injuries** -Adverse weather related variable **Event type**

Injuries and event type

```
INJ<-storm_data %>%
        select(INJURIES,EVTYPE) %>%
        group_by(EVTYPE) %>%
        summarise(Total_INJ=sum(INJURIES,na.rm = TRUE)) %>%
        arrange(desc(Total_INJ))

## `summarise()` ungrouping output (override with `.groups` argument)

INJ_data<-INJ[1:10,]
```

Fatalites and event type

```
FATL<-storm_data %>%
        select(FATALITIES,EVTYPE) %>%
        group_by(EVTYPE) %>%
        summarise(Total_FATL=sum(FATALITIES,na.rm = TRUE)) %>%
        arrange(desc(Total_FATL))

## `summarise()` ungrouping output (override with `.groups` argument)

FATL_data<-FATL[1:10,]
```

Events with the greatest economic consequences across the United States

The variables considered in the analysis are as follows: -Economic aspect related variables are approximate **Property damage** and **Crop damage** -Adverse weather related variable **Event type**

Property damage and event type

```
PROP<-storm_data %>%
        select(PROPDMG,EVTYPE) %>%
        group_by(EVTYPE) %>%
        summarise(Total_Property_Damage=sum(PROPDMG,na.rm = TRUE))

## `summarise()` ungrouping output (override with `.groups` argument)

PROP <- storm_data %>%
        mutate(PROP = ifelse(toupper(PROPDMGEXP) =='K', PROPDMG*1000,
```

```
ifelse(toupper(PROPDMGEXP) =='M', PROPDMG*1000000, ifelse(toupper(PROPDMGEXP)
== 'B', PROPDMG*1000000000, ifelse(toupper(PROPDMGEXP) == 'H', PROPDMG*100,
PROPDMG)))))

PROP<-storm_data %>%
        select(PROPDMG,EVTYPE) %>%
        group_by(EVTYPE) %>%
        summarise(Total_Property_Damage=sum(PROPDMG,na.rm = TRUE)) %>%
        arrange(desc(Total_Property_Damage))

## `summarise()` ungrouping output (override with `.groups` argument)

PROP_data<-PROP[1:10,]
```

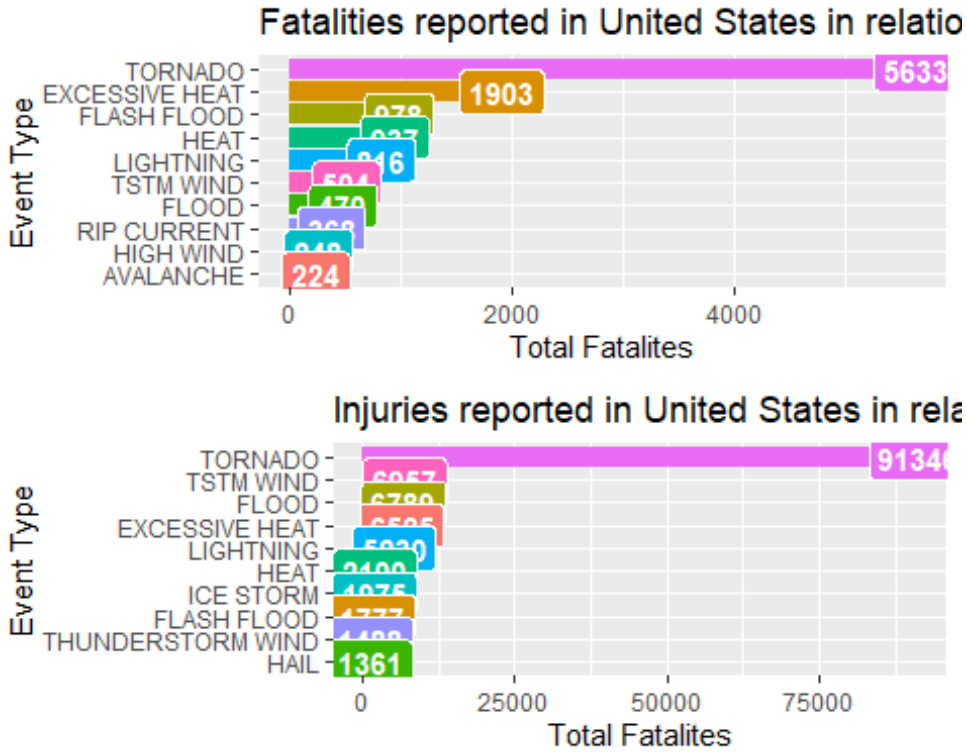
Crop Damage and event type

```
CROP<-storm_data %>%
        select(CROPDMG,EVTYPE) %>%
        group_by(EVTYPE) %>%
        summarise(Total_Crop_Damage=sum(CROPDMG,na.rm = TRUE))

## `summarise()` ungrouping output (override with `.groups` argument)

CROP <- storm_data%>%
        mutate(CROP = ifelse(toupper(CROPDMGEXP) =='K', CROPDMG*1000,
ifelse(toupper(CROPDMGEXP) =='M', CROPDMG*1000000, ifelse(toupper(CROPDMGEXP)
== 'B', CROPDMG*1000000000, ifelse(toupper(CROPDMGEXP) == 'H', CROPDMG*100,
CROPDMG)))))

CROP<-storm_data %>%
        select(CROPDMG,EVTYPE) %>%
        group_by(EVTYPE) %>%
        summarise(Total_Crop_Damage=sum(CROPDMG,na.rm = TRUE)) %>%
        arrange(desc(Total_Crop_Damage))

## `summarise()` ungrouping output (override with `.groups` argument)

CROP_data<-CROP[1:10,]
```

Results Visualizing Injuries and Fatalities with event types

```
#Plotting Fatalities with event type
plot1<-
ggplot(FATL_data,aes(reorder(EVTYPE,Total_FATL),Total_FATL,fill=EVTYPE))+
        geom_bar(stat = "identity")+
        geom_label(aes(EVTYPE,Total_FATL,label=Total_FATL),
                   color="white",fontface="bold")+
        xlab("Event Type")+ ylab("Total Fatalites")+
        ggtitle("Fatalities reported in United States in relation to Event
Type")+
        theme(legend.position = "none")+
        coord_flip()
```

```
#Plotting Injuries with event type
plot2<-ggplot(INJ_data,aes(reorder(EVTYPE,Total_INJ),Total_INJ,fill=EVTYPE))+
        geom_bar(stat = "identity")+
        geom_label(aes(EVTYPE,Total_INJ,label=Total_INJ),
                   color="white",fontface="bold")+
        xlab("Event Type")+ ylab("Total Fatalites")+
        ggtitle("Injuries reported in United States in relation to Event
Type")+
        theme(legend.position = "none")+
        coord_flip()
grid.arrange(plot1,plot2,ncol=1)
```

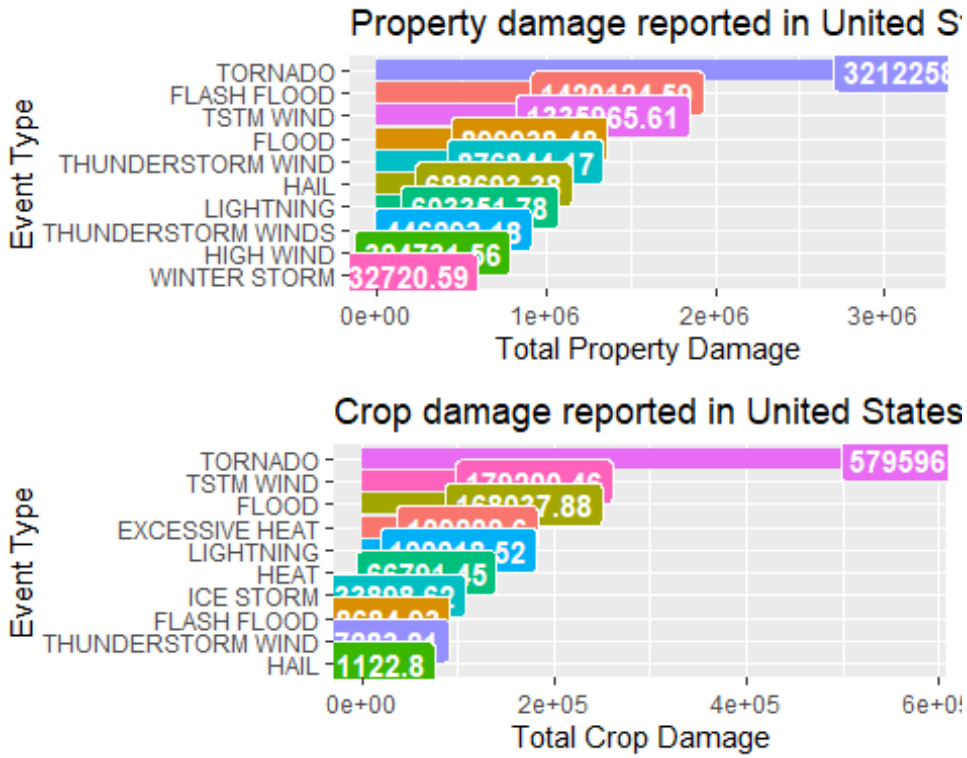


From the above plots, **Tornado** is the weather event with the most adverse effects to human population health recording the highest overall number of fatalities and injuries accross United States.

Visualizing Property damage and Crop damage with events

```
#Plotting Property damage with event type
plot3<-
ggplot(PROP_data,aes(reorder(EVTYPE,Total_Property_Damage),Total_Property_Dam
age,fill=EVTYPE))+
        geom_bar(stat = "identity")+

geom_label(aes(EVTYPE,Total_Property_Damage,label=Total_Property_Damage),
                   color="white",fontface="bold")+
```

```
        xlab("Event Type")+ ylab("Total Property Damage")+
        ggtitle("Property damage reported in United States in relation to
Event Type")+
        theme(legend.position = "none")+
        coord_flip()

#Plotting Crop damaage with event type
plot4<-
ggplot(INJ_data,aes(reorder(EVTYPE,CROP_data$Total_Crop_Damage),CROP_data$Tot
al_Crop_Damage,fill=EVTYPE))+
        geom_bar(stat = "identity")+

geom_label(aes(EVTYPE,CROP_data$Total_Crop_Damage,label=CROP_data$Total_Crop_
Damage),
                   color="white",fontface="bold")+
        xlab("Event Type")+ ylab("Total Crop Damage")+
        ggtitle("Crop damage reported in United States in relation to Event
Type")+
        theme(legend.position = "none")+
        coord_flip()
grid.arrange(plot3,plot4,ncol=1)
```



From the above plots, **Tornado** is the weather event with the most adverse effects to on the economy recording the highest overall property and crop damages accross United States.