Reproducible Research - Week 4 Project 2

Joaquin Arango

12/28/2020

#NOAA Storm Database

For this project we will look at the NOAA Storm Database and answer questions about severe weather events and its impacts on health and ecomonics.

Synopsis

Severe weather events can cause grave health and economic impacts across the US. Events that can results in injuries, damages or even death. The NOAA database keeps track of major serve events in the US.

##Questions: This data analysis must address 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 Processing and Retrieval

```
#set the environment
library(knitr)
library(plyr)
library(ggplot2)
library(lattice)
library(data.table)
library(grid)
library(gridExtra)

File_data <- "repdata_data_StormData.csv.bz2"

# reading data
df <- read.csv(file = File_data, header=TRUE, sep=",")
```

Select Useful data for Analysis once you read up on NOAA's Storm Data Documentation

https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf

```
## variables selected
##Event Type (EVTYPE), Begin Date of Event (BGN_DATE), Event Human Fatalities (FATALITIES), Event Human
df <- df[ , c("EVTYPE", "BGN_DATE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPI
#changing date to POSIXct date
df$BGN_DATE <- as.POSIXct(df$BGN_DATE, format="%m/%d/%Y %H:%M:%S")
#take a look at current Data structure
str(df)
```

```
## 'data.frame':    902297 obs. of  8 variables:
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_DATE  : POSIXct, format: "1950-04-18" "1950-04-18" ...
##  $ 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  "" "" "" "" ...
```

```
#look at list of events
head(unique(df$EVTYPE))
```

```
## [1] "TORNADO"               "TSTM WIND"             "HAIL"
## [4] "FREEZING RAIN"         "SNOW"                  "ICE STORM/FLASH FLOOD"
```

Question 1

Best way to answer this question is to look at fatalities and injuries as they are the most harmful with respect to health.

Data Processing for Question 1:

```
#add up fatalities by event type
fatalities <- aggregate(FATALITIES ~ EVTYPE, data=df, sum)
injuries <- aggregate(INJURIES ~ EVTYPE, data=df, sum)
#arrange fatalities and injuries from highest to lowest by event type
fatalities <- arrange(fatalities, desc(FATALITIES),EVTYPE)[1:10,]
injuries <- arrange(injuries, desc(INJURIES), EVTYPE)[1:10,]

#convert EVTYPE to factor so we can analyze
fatalities$EVTYPE <-factor(fatalities$EVTYPE, level = fatalities$EVTYPE)
injuries$EVTYPE <- factor(injuries$EVTYPE, levels=injuries$EVTYPE)
```

Question 1 Tables showing fatalities and injures by most common events.

```
#to show table of fatalities
fatalities
```

```
##            EVTYPE FATALITIES
## 1         TORNADO       5633
## 2  EXCESSIVE HEAT       1903
## 3     FLASH FLOOD        978
```

```
## 4            HEAT        937
## 5       LIGHTNING        816
## 6       TSTM WIND        504
## 7           FLOOD        470
## 8     RIP CURRENT        368
## 9       HIGH WIND        248
## 10      AVALANCHE        224
```

```
#to show table of injuries
injuries
```

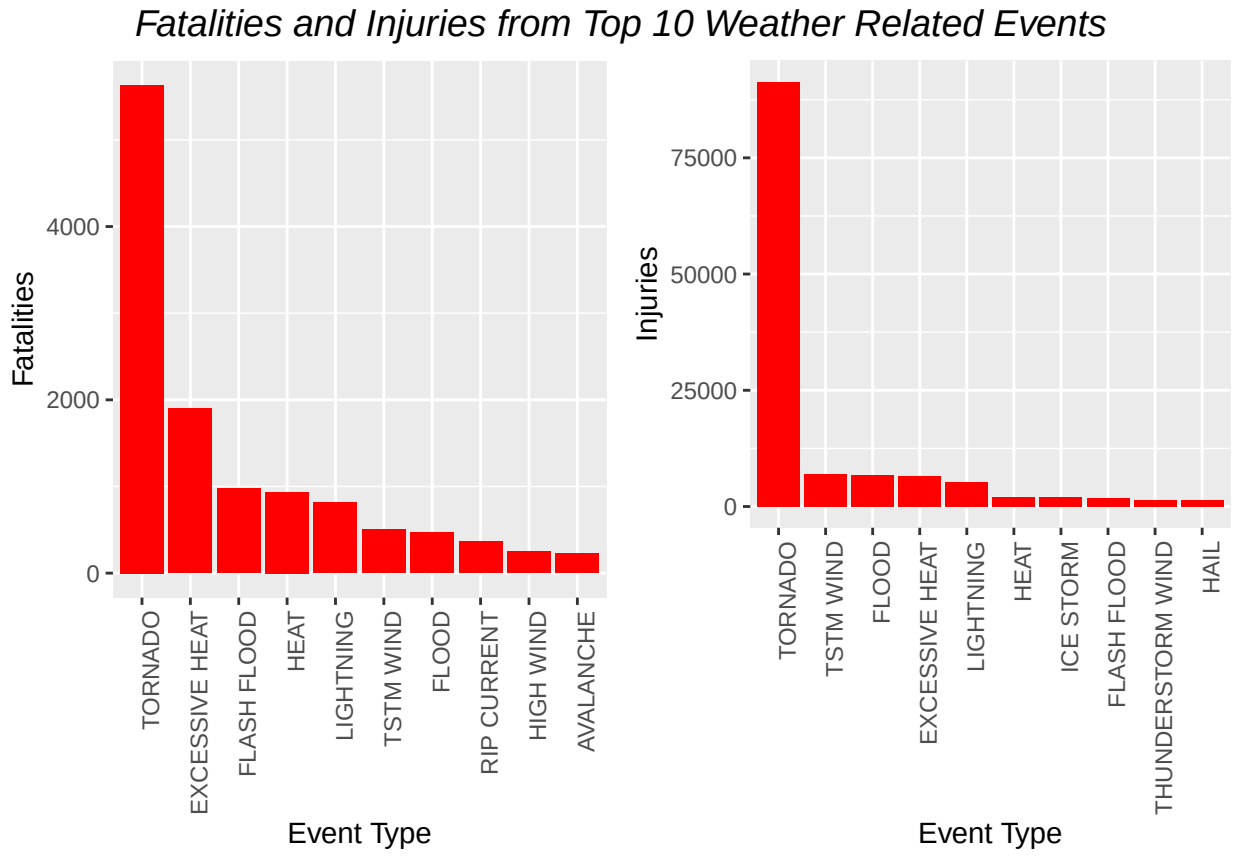
```
##               EVTYPE INJURIES
## 1            TORNADO    91346
## 2          TSTM WIND     6957
## 3              FLOOD     6789
## 4     EXCESSIVE HEAT     6525
## 5          LIGHTNING     5230
## 6               HEAT     2100
## 7          ICE STORM     1975
## 8        FLASH FLOOD     1777
## 9  THUNDERSTORM WIND     1488
## 10             HAIL     1361
```

Question 1 Plots

```
#plot fatalities
fatalitiesbyevent <- ggplot(fatalities, aes(x=EVTYPE, y=FATALITIES))+ geom_bar(stat = "identity", fill=

#plot injuries
injuriesbyevent <- ggplot(injuries, aes(x=EVTYPE, y=INJURIES))+ geom_bar(stat = "identity", fill="red",

#plot both side by side
grid.arrange(fatalitiesbyevent, injuriesbyevent, ncol=2, nrow=1,
top = textGrob("Fatalities and Injuries from Top 10 Weather Related Events", gp=gpar(fontsize=14, font=
```



Question 2

Since question 2 asked for events with the greatest economic damage, we will use PROPDMGEXP and CROPDMGEXP, since they are in character format, we will need to convert them into numeric values.

```
#turn from character into integer
numPROPDMG <- mapvalues(df$PROPDMGEXP,      c("K","M","","B","m","+","0","5","6","?","4","2","3","h","7
numCROPDMG <- mapvalues(df$CROPDMGEXP, c("","M","K","m","B","?","0","k","2"),
c( 1,1e6,1e3,1e6,1e9,1,1,1e3,1e2))
#Numeric Property and Crop Damage
df$TOTAL_PROPDMG <- as.numeric(numPROPDMG) * df$PROPDMG #property
df$TOTAL_CROPDMG <- as.numeric(numCROPDMG) * df$CROPDMG #crop

#show columns
colnames(df)
```

```
##  [1] "EVTYPE"         "BGN_DATE"       "FATALITIES"     "INJURIES"
##  [5] "PROPDMG"        "PROPDMGEXP"     "CROPDMG"        "CROPDMGEXP"
##  [9] "TOTAL_PROPDMG" "TOTAL_CROPDMG"
```

```
#lets create a total damage amount with property and crop
df$TOTALDMG <- df$TOTAL_PROPDMG + df$TOTAL_CROPDMG

#add up the total property and crop damage by event type
```

```
propdamage <- aggregate(TOTAL_PROPDMG ~ EVTYPE, data=df, sum)
cropdamage <- aggregate(TOTAL_CROPDMG ~ EVTYPE, data=df, sum)

#add total damages by event type
totaldamage <- aggregate(TOTALDMG ~ EVTYPE, data=df, sum)

#arrange the damages from highest to lowest for property and crop and total

propdamage <-arrange(propdamage, desc(propdamage$TOTAL_PROPDMG), EVTYPE)[1:10,]
cropdamage <- arrange(cropdamage, desc(cropdamage$TOTAL_CROPDMG),EVTYPE)[1:10,]
totaldamage <- arrange(totaldamage, desc(totaldamage$TOTALDMG), EVTYPE)[1:10,]

#change weather event type to factor variable

propdamage$EVTYPE <- factor(propdamage$EVTYPE, levels= propdamage$EVTYPE)
cropdamage$EVTYPE <- factor(cropdamage$EVTYPE, levels=cropdamage$EVTYPE)
totaldamage$EVTYPE <- factor(totaldamage$EVTYPE, levels=totaldamage$EVTYPE)
```

Question 2 Tables showing property and crop damage by most common events.

```
#show table of Property Damage
propdamage
```

```
##               EVTYPE TOTAL_PROPDMG
## 1              FLOOD  144657709807
## 2  HURRICANE/TYPHOON   69305840000
## 3            TORNADO   56947380677
## 4        STORM SURGE   43323536000
## 5        FLASH FLOOD   16822673979
## 6               HAIL   15735267513
## 7          HURRICANE   11868319010
## 8     TROPICAL STORM    7703890550
## 9       WINTER STORM    6688497251
## 10         HIGH WIND    5270046295
```

```
#plot of Property Damage
propplotdamage <- ggplot(propdamage, aes(x=EVTYPE, y=TOTAL_PROPDMG)) + geom_bar(stat = "identity", fill=
#plot(propplotdamage)
```

```
#show table of Crop Damage
cropdamage
```

```
##               EVTYPE TOTAL_CROPDMG
## 1            DROUGHT   13972566000
## 2              FLOOD    5661968450
## 3        RIVER FLOOD    5029459000
## 4          ICE STORM    5022113500
## 5               HAIL    3025954473
## 6          HURRICANE    2741910000
## 7  HURRICANE/TYPHOON    2607872800
## 8        FLASH FLOOD    1421317100
## 9       EXTREME COLD    1292973000
## 10      FROST/FREEZE    1094086000
```

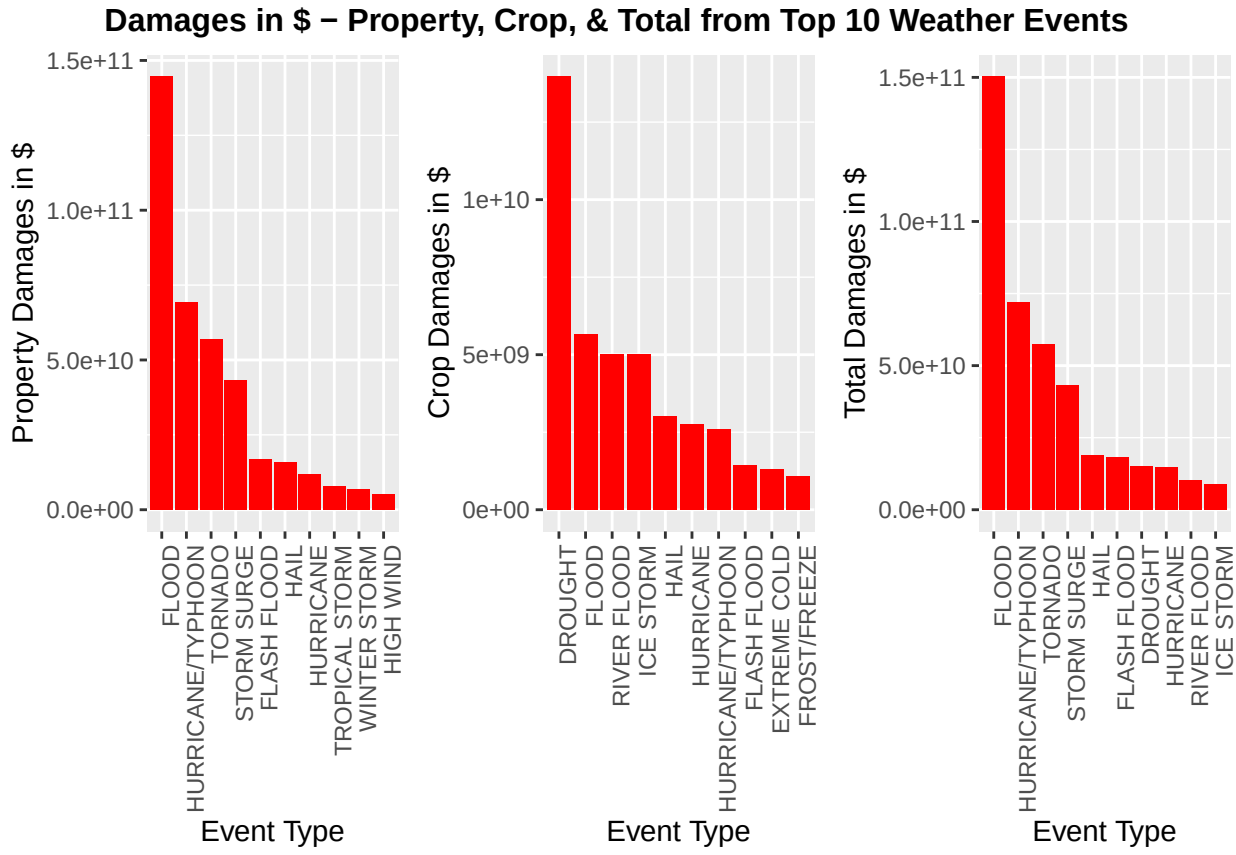
```
#plot of Crop Damage
cropplotdamage <-ggplot(cropdamage, aes(x=EVTYPE, y=TOTAL_CROPDMG)) + geom_bar(stat = "identity", fill=
#plot(cropplotdamage)
```

```
#show table of Total Damage
totaldamage
```

```
##                EVTYPE     TOTALDMG
## 1               FLOOD 150319678257
## 2  HURRICANE/TYPHOON  71913712800
## 3             TORNADO  57362333947
## 4         STORM SURGE  43323541000
## 5                HAIL  18761221986
## 6         FLASH FLOOD  18243991079
## 7             DROUGHT  15018672000
## 8           HURRICANE  14610229010
## 9         RIVER FLOOD  10148404500
## 10          ICE STORM   8967041360
```

```
#plot of Total Damage
totalplotdamage <-ggplot(totaldamage, aes(x=EVTYPE, y=TOTALDMG)) + geom_bar(stat = "identity", fill="re
#plot(totalplotdamage)

grid.arrange(propplotdamage, cropplotdamage, totalplotdamage, ncol=3, nrow=1,
     top = textGrob("Damages in $ - Property, Crop, & Total from Top 10 Weather Events ",gp=gpar(fontsi
```



Results

Question 1: Tornadoes caused the most fatalities, tornadoes caused the most significant harm out of all types of severe weather events.Heat and flash flooding are second and third leading causes of fatalities.

Tornado also cause more injuries. Thunderstorm related events (wind, lightning, flooding) combined are a significant injury contributor as well.

Question 2:

Floods caused the most significant total damage for property and crop.

While property damage is most significant with flooding, Hurricanes/Typhoons, and Tornadoes, and Storm Surge are also strong contributors to property damage as well

Crop damage occurs with flooding, but drought events are the most significant cause of economic harm. For crops flooding, ice storms, and hail are also noted as main events to economic harm.