Impact of Severe Weather Events on Health and Economy

Xiaoyi (Leo) Liu

Synopsis

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.

In this project, we will explore 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.

Data Processing

First, we download the dataset, upzip it and read the csv file.

```
setwd("~/Desktop/CourseraReproducibleResearchProject2")

if (!"stormData.csv.bz2" %in% dir())
        {
        download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfi
}
bunzip2("stormData.csv.bz2", overwrite = T, remove = F)
```

```
StormData = read.csv('stormData.csv', sep = ",")
head(StormData)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
```

```
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

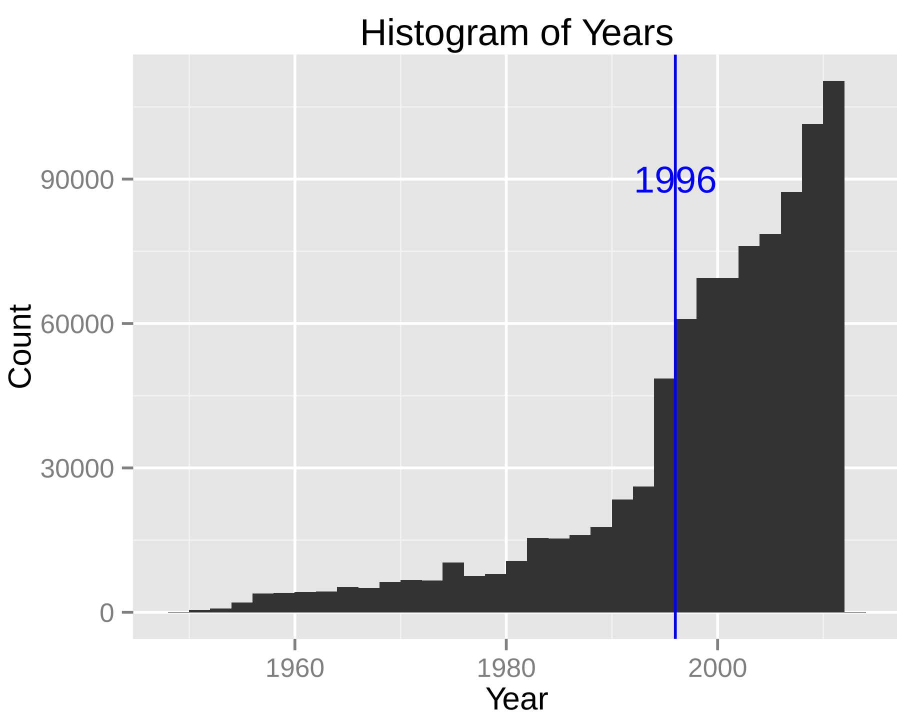
```r
dim(StormData)
```

```
## [1] 902297     37
```

There are 902297 rows and 37 columns in the dataset, and the events start from 1950 until Nov. 2011. We first add a new column **'Year'** which records the year in the end date numerically. Then, we plot a histogram of the years when those events started to happen.

```r
StormData$YEAR <- as.numeric(format(as.Date(StormData$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y"))

HistYear = ggplot(aes(x = YEAR), data = StormData) + geom_histogram(binwidth = 2) + xlab('Year') + ylab
HistYear
```



It can be seen that at early dates before 1996, the number of events is usually smaller than 60000 for each year, while the dates after 1996 have plenty of event records. For the purpose of a more consistent balance of event types across recent years, we decide to use the subset of the data from 1996 to 2011.

```r
StormData = subset(StormData, YEAR >= 1996)
dim(StormData)
```

```
## [1] 653530     38
```

Now, the refined dataset has 653530 rows now. The proceeding analysis will be based on the events after 1996.

Impact on Public Health

To find out which event causes the most harmful damages on the public health, we will compute the number of fatalities and injuries caused by the severe weather events and find out the top 5 ones.

```r
SevereEventsForFatalities = aggregate(StormData$FATALITIES, by = list(StormData$EVTYPE), FUN = 'sum')
SevereEventsForInjuries = aggregate(StormData$INJURIES, by = list(StormData$EVTYPE), FUN = 'sum')
Top5SevereEventsForFatalities = SevereEventsForFatalities[order(-SevereEventsForFatalities$x), ][1 : 5,
Top5SevereEventsForInjuries = SevereEventsForInjuries[order(-SevereEventsForInjuries$x), ][1 : 5, ]
names(Top5SevereEventsForFatalities) = c("EVTYPE_FOR_FATALITIES", "COUNT")
names(Top5SevereEventsForInjuries) = c("EVTYPE_FOR_INJURIES", "COUNT")
Top5SevereEventsForFatalities
```

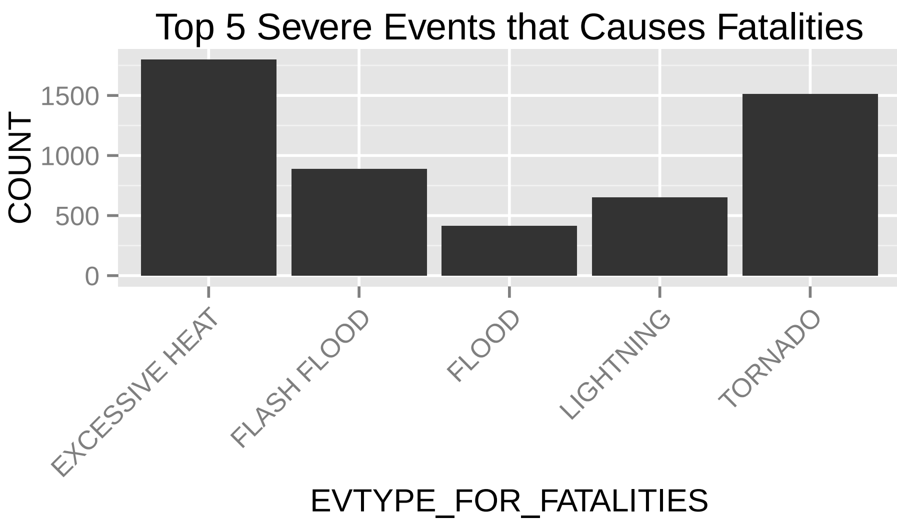
```
##     EVTYPE_FOR_FATALITIES COUNT
## 81         EXCESSIVE HEAT  1797
## 426               TORNADO  1511
## 98            FLASH FLOOD   887
## 224             LIGHTNING   651
## 102                 FLOOD   414
```

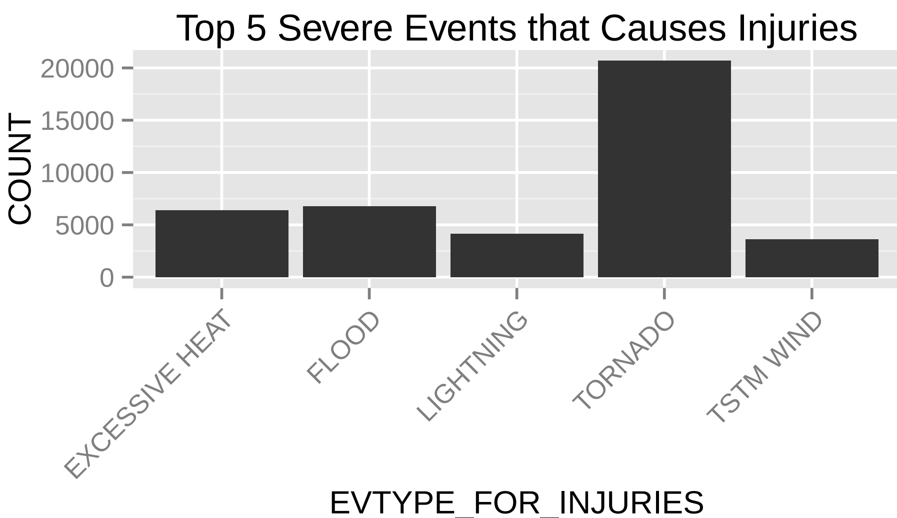
```r
Top5SevereEventsForInjuries
```

```
##     EVTYPE_FOR_INJURIES COUNT
## 426             TORNADO 20667
## 102               FLOOD  6758
## 81       EXCESSIVE HEAT  6391
## 224           LIGHTNING  4141
## 434           TSTM WIND  3629
```

Now, we plot the total fatalities and total injuries affected by top 5 severe weather events.

```r
FatalPlot = ggplot(aes(x = EVTYPE_FOR_FATALITIES, y = COUNT), data = Top5SevereEventsForFatalities) + ge
InjurPlot = ggplot(aes(x = EVTYPE_FOR_INJURIES, y = COUNT), data = Top5SevereEventsForInjuries) + geom_
grid.arrange(FatalPlot, InjurPlot, nrow = 2)
```





From above, we can see that:

- Excessive heat and tornado cause most fatalities;
- Tornado and flood cause most injuries.

Impact on Economy

Similar to the data analysis above, we will compute the number of property damage and crop damage caused by the severe weather events and find out the top 5 severe ones.

```
StormData$PROPDMGEXP = as.character(StormData$PROPDMGEXP)
StormData$PROPDMGEXP[toupper(StormData$PROPDMGEXP) == 'B'] = "9"
StormData$PROPDMGEXP[toupper(StormData$PROPDMGEXP) == 'M'] = "6"
StormData$PROPDMGEXP[toupper(StormData$PROPDMGEXP) == 'K'] = "3"
StormData$PROPDMGEXP[toupper(StormData$PROPDMGEXP) == 'H'] = "2"
StormData$PROPDMGEXP = as.numeric(StormData$PROPDMGEXP)
StormData$PROPDMGEXP[is.na(StormData$PROPDMGEXP)] = 0
StormData$PROPDMG = StormData$PROPDMG * 10^StormData$PROPDMGEXP
```

```
StormData$CROPDMGEXP = as.character(StormData$CROPDMGEXP)
StormData$CROPDMGEXP[toupper(StormData$CROPDMGEXP) == 'B'] = "9"
StormData$CROPDMGEXP[toupper(StormData$CROPDMGEXP) == 'M'] = "6"
StormData$CROPDMGEXP[toupper(StormData$CROPDMGEXP) == 'K'] = "3"
StormData$CROPDMGEXP[toupper(StormData$CROPDMGEXP) == 'H'] = "2"
StormData$CROPDMGEXP = as.numeric(StormData$CROPDMGEXP)
StormData$CROPDMGEXP[is.na(StormData$CROPDMGEXP)] = 0
StormData$CROPDMG = StormData$CROPDMG * 10^StormData$CROPDMGEXP

SevereEventsForProp = aggregate(StormData$PROPDMG, by = list(StormData$EVTYPE), FUN = 'sum')
SevereEventsForCrop = aggregate(StormData$CROPDMG, by = list(StormData$EVTYPE), FUN = 'sum')
Top5SevereEventsForProp = SevereEventsForProp[order(-SevereEventsForProp$x), ][1 : 5, ]
Top5SevereEventsForCrop = SevereEventsForCrop[order(-SevereEventsForCrop$x), ][1 : 5, ]
names(Top5SevereEventsForProp) = c("EVTYPE_FOR_PROP", "COUNT")
names(Top5SevereEventsForCrop) = c("EVTYPE_FOR_CROP", "COUNT")
Top5SevereEventsForProp
```

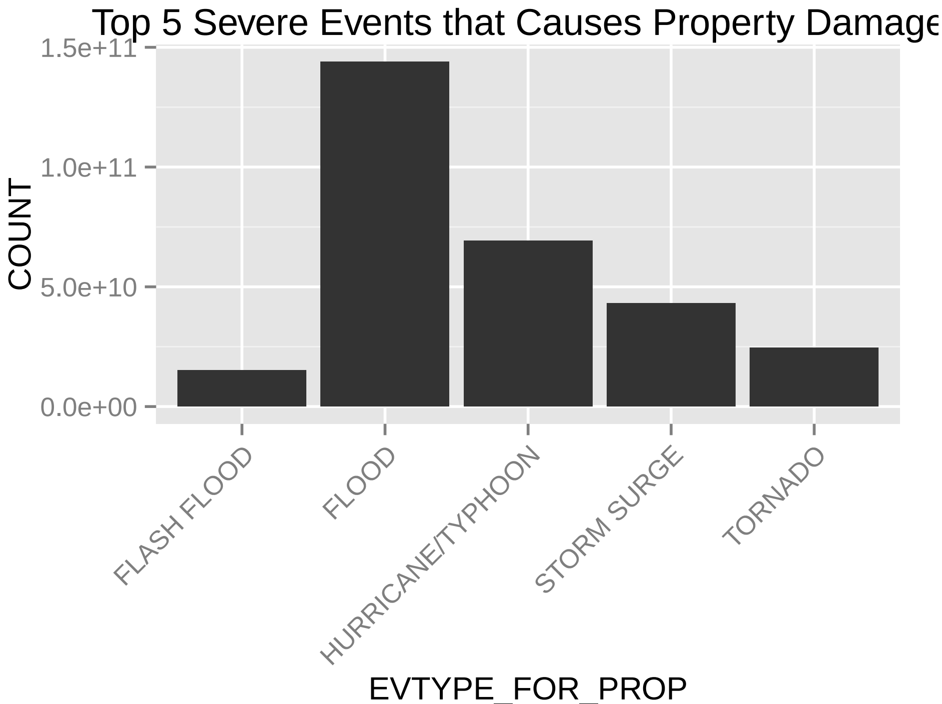
```
##       EVTYPE_FOR_PROP        COUNT
## 102             FLOOD 143944833550
## 185 HURRICANE/TYPHOON  69305840000
## 342       STORM SURGE  43193536000
## 426           TORNADO  24616945710
## 98        FLASH FLOOD  15222203910
```

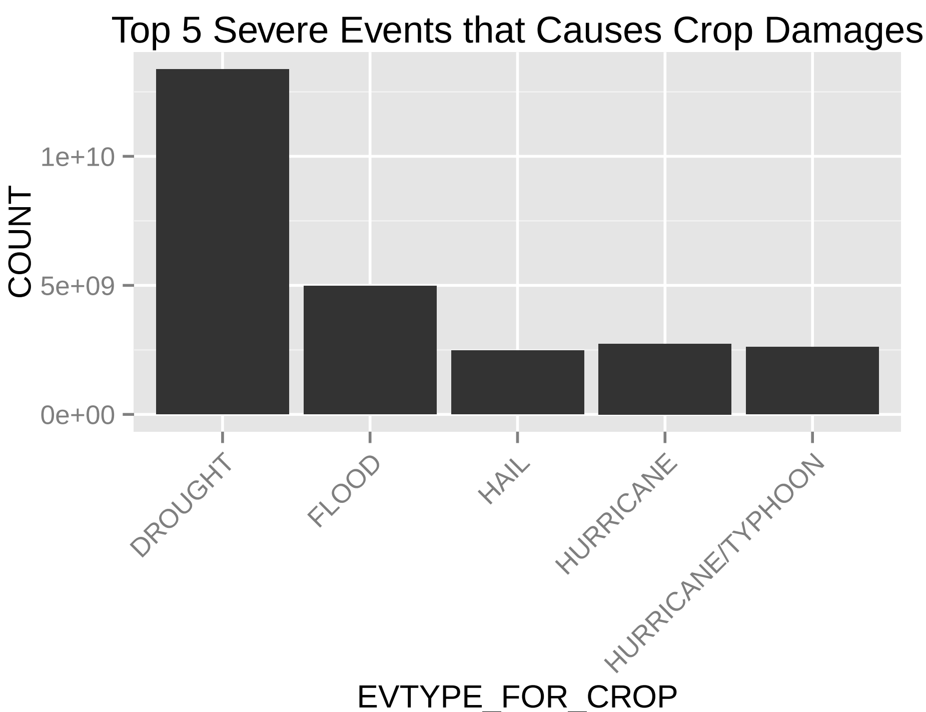
```
Top5SevereEventsForCrop
```

```
##       EVTYPE_FOR_CROP       COUNT
## 63            DROUGHT 13367566000
## 102             FLOOD  4974778400
## 183         HURRICANE  2741410000
## 185 HURRICANE/TYPHOON  2607872800
## 142              HAIL  2476029450
```

We also plot the total property and crop damages caused by top 5 severe weather events.

```
PropPlot = ggplot(aes(x = EVTYPE_FOR_PROP, y = COUNT), data = Top5SevereEventsForProp) + geom_bar(stat =
CropPlot = ggplot(aes(x = EVTYPE_FOR_CROP, y = COUNT), data = Top5SevereEventsForCrop) + geom_bar(stat =
grid.arrange(PropPlot, CropPlot, nrow = 2)
```





From above, we can see that:

- Flood and hurricane/typhoon cause most property damages;
- Drought and flood cause most crop damages.