Health and Economic Impact of Weather Events in the US

Note 1 (Important):

In my computer environment, packages related with R markdown or Knit document does not installed normally with errors, so I will submit this document instead.

Note 2:

Answers are checked to **red color and bold style**, and **codes** are checked to **blue color and bold style**.

---

1. 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.

---

2. Libraries to use for solving question

```
library(data.table); # to use fread() statement
library(ggplot2); # to use ggplot() and related statements
library(sqldf); # to use sqldf() statement
```

---

3. Basic information of given data

The analysis was performed on Storm Events Database, provided by National Climatic Data Center. The data is a comma-separated-value(csv) file.

First, I downloaded this file from this site, and read the data into my workspace in R.

```
storm <- fread("repdata%2Fdata%2FStormData.csv", head=T);
```

dim(storm);

```
[1] 902297  37
```

This file has 902,297 rows and 37 columns, and through str() statement, we can know which column has what information like below.

str(storm);

```
Classes 'data.table' and '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          : chr  "3" "2" "2" "2" ...
 $ 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 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

---

4. Question and Answer

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

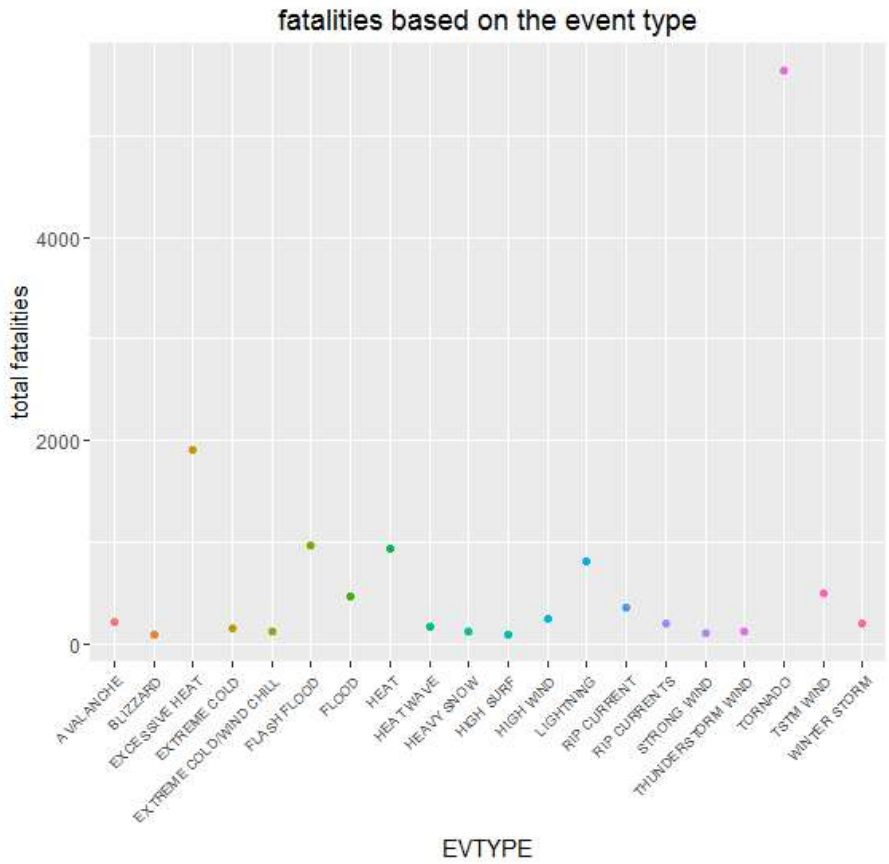
Before the analysis, the data need preprocessing. Some variables related harmfulness need to be grouped by event type(EVTYPE), so sqldf() statement can be used.

```
fatalities <- sqldf('select EVTYPE,
                            sum(FATALITIES) as "SUM_FATALITIES"
                            from storm
                            group by EVTYPE
                            order by EVTYPE');
fatalities$SUM_FATALITIES <- as.numeric(fatalities$SUM_FATALITIES);
injuries <- sqldf('select EVTYPE,
                          sum(INJURIES) as "SUM_INJURIES"
                          from storm
                          group by EVTYPE
                          order by EVTYPE');
injuries$SUM_INJURIES <- as.numeric(injuries$SUM_INJURIES);
```

Next, through this two dataset, we can figure out plots for solving the question.

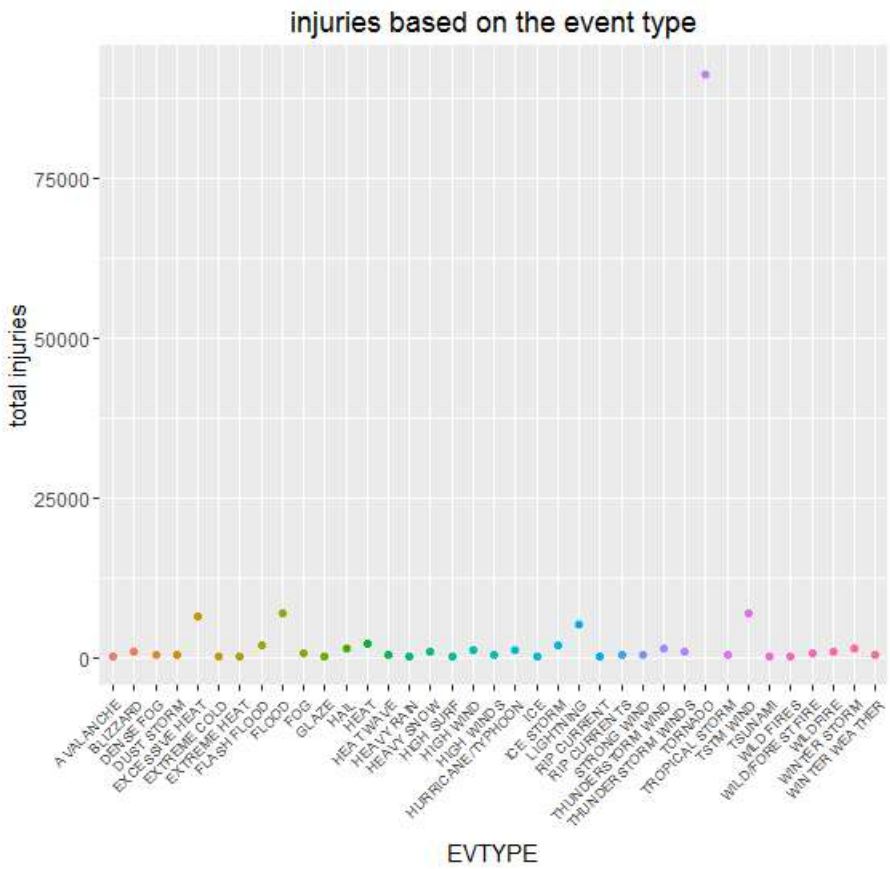
fatalities

```
plot <- ggplot(subset(fatalities,SUM_FATALITIES>100), aes(EVTYPE, SUM_FATALITIES));
plot <- plot + geom_point(aes(color=EVTYPE)) + theme(legend.position="none", axis.text.x =
element_text(angle = 45, hjust = 1, size=7)) +
  labs(x="EVTYPE") + labs(y="total fatalities") + ggtitle("fatalities based on the event type");
print(plot); # just plotting that the fatalities is over 100
```



[Figure 1. Fatalities based on the event type]

```
### injuries
plot <- ggplot(subset(injuries,SUM_INJURIES>100), aes(EVTYPE, SUM_INJURIES));
plot <- plot + geom_point(aes(color=EVTYPE)) + theme(legend.position="none", axis.text.x = element_text(angle = 45, hjust = 1, size=7)) + labs(x="EVTYPE") + labs(y="total injuries") + ggtitle("injuries based on the event type");
print(plot); # just plotting that the injuries is over 100
```



[Figure 1. Injuries based on the event type]

Through [Figure 1] and [Figure 2], **the event type that has the most fatalities and injuries is all "Tornado".**

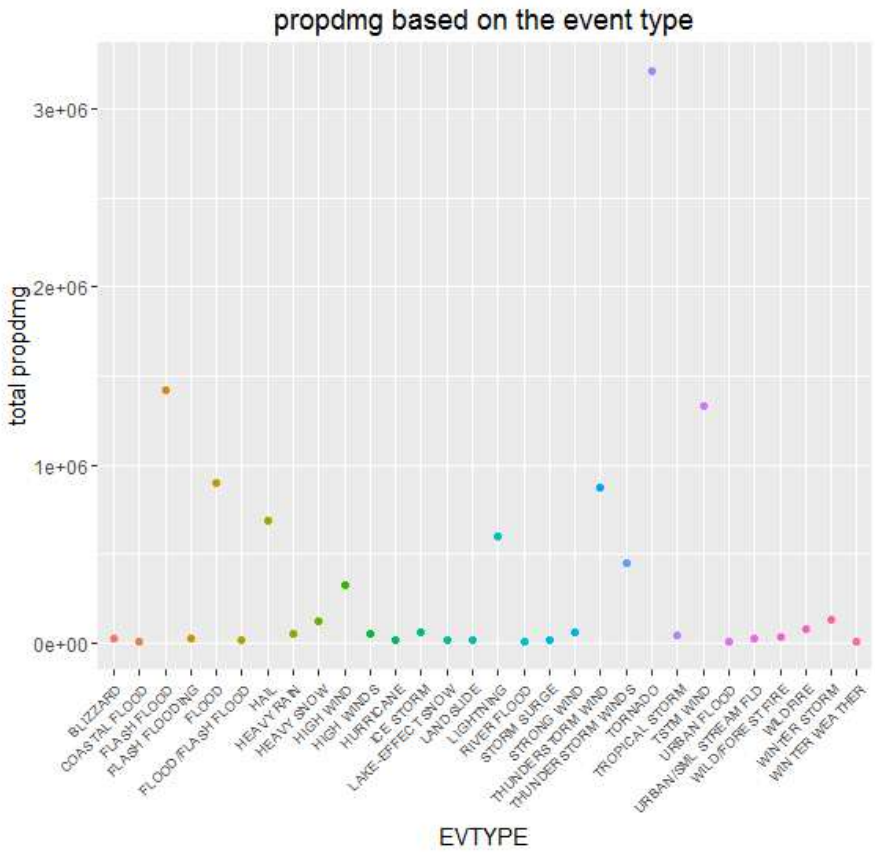
2) Across the United States, which types of events have the greatest economic consequences?

```
propdmg <- sqldf('select EVTYPE,
                         sum(PROPDMG) as "SUM_PROPDMG"
                         from storm
                         group by EVTYPE
                         order by EVTYPE');
propdmg$SUM_PROPDMG <- as.numeric(propdmg$SUM_PROPDMG);
cropdmg <- sqldf('select EVTYPE,
                        sum(CROPDMG) as "SUM_CROPDMG"
                        from storm
                        group by EVTYPE
                        order by EVTYPE');
cropdmg$SUM_CROPDMG <- as.numeric(cropdmg$SUM_CROPDMG);
```

By the same method with Question 1, I could just simply find the EVTYPE that has the most property damage or crop damage by using similar codes and figure out plots.

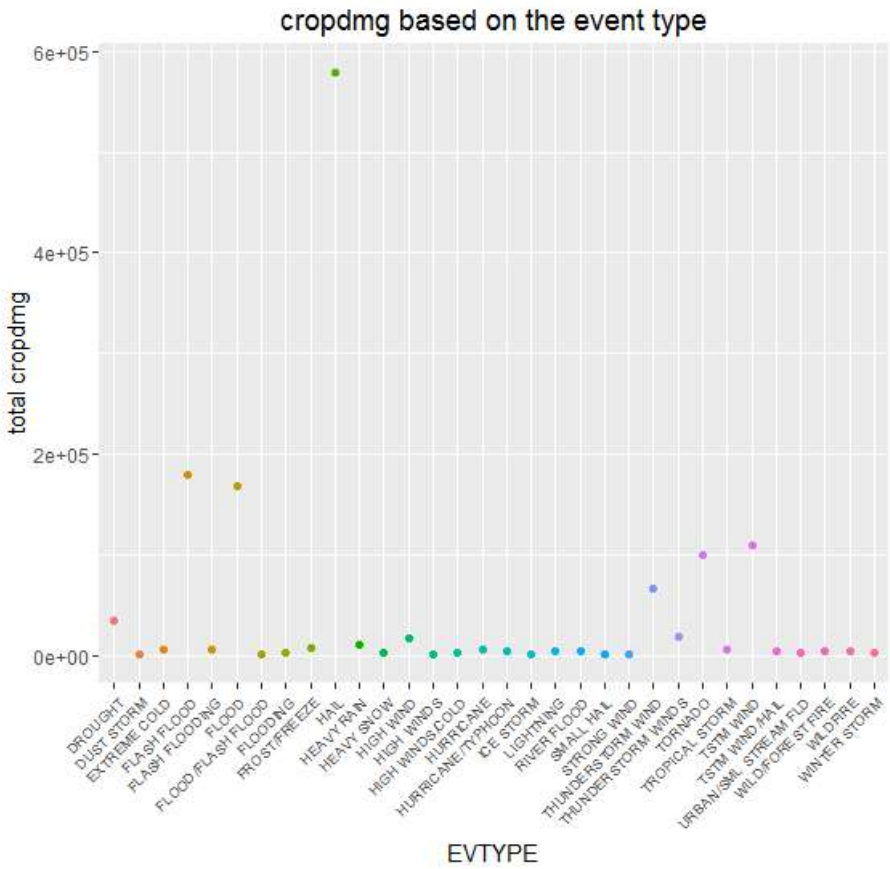
progdmg

```
plot <- ggplot(subset(propdmg,SUM_PROPDMG>mean(SUM_PROPDMG)), aes(EVTYPE, SUM_PROPDMG));
plot <- plot + geom_point(aes(color=EVTYPE)) + theme(legend.position="none", axis.text.x = element_text(angle = 45, hjust = 1, size=7)) +
  labs(x="EVTYPE") + labs(y="total propdmg") + ggtitle("propdmg based on the event type");
print(plot); # just plotting that the propdmg is over mean
```



[Figure 3. Propdmg based on the event type]

```
### cropdmg
plot <- ggplot(subset(cropdmg,SUM_CROPDMG>mean(SUM_CROPDMG)), aes(EVTYPE, SUM_CROPDMG));
plot <- plot + geom_point(aes(color=EVTYPE)) + theme(legend.position="none", axis.text.x = element_text(angle = 45, hjust = 1, size=7)) +
  labs(x="EVTYPE") + labs(y="total cropdmg") + ggtitle("cropdmg based on the event type");
print(plot); # just plotting that the cropdmg is over mean
```



[Figure 4. Cropdmg based on the event type]

Through [Figure 3] and [Figure 4], **the event type that has the most property damage is "Tornado", the event type that has the most crop damage is "Hail".**