Reproducible Research Week 4 Peer graded Assignment

This report focuses on assesing the social and economic impact of different weather events in the USA. Specifically, we are inspecting which weather events cause more injuries, fatalities and economic losses. The dataset provided starts from 1950 and ends in the year 2011. Naturally, earlier years had more missing values due to the inferior recording tools and practices at the time.

First, we import all the necessary libraries:

```
library(ggplot2)
library(grid)
library(gridExtra)
```

Now we download our dataset :

```
myfile <- read.csv("repdata_data_StormData.csv")
```

Prepping the social impact data

1. Taking only events with one or more fatlity/injury
2. Splitting data into subsets based on the consequence
3. Aggregate the data to get sum of impacts for each event
4. Reorder data in terms of sum

```
fatalitiesData <- subset(myfile, FATALITIES >0)
injuriesData <- subset(myfile, INJURIES >0)
fatalitiesSum<- aggregate(FATALITIES ~ EVTYPE, fatalitiesData, sum, na.rm= TRUE)
injuriesSum <- aggregate(INJURIES~ EVTYPE, injuriesData, sum, na.rm= TRUE)
fatalitiesSum  <- fatalitiesSum[order(fatalitiesSum$FATALITIES, decreasing = TRUE),]
injuriesSum    <- injuriesSum[order(injuriesSum$INJURIES, decreasing = TRUE),]
```

Prepping the economic impact data

1. Again, taking only events where cost was >0
2. Replacing charecter data with numerical.
3. Adding a separate column with the total damage value.
4. Aggregating and re-ordering.

```
economyData<- subset(myfile, PROPDMG > 0 | CROPDMG > 0)
economyData$PROPDMGEXP   <- sub("h|H", "2", economyData$PROPDMGEXP)
economyData$PROPDMGEXP   <- sub("k|K", "3", economyData$PROPDMGEXP)
economyData$PROPDMGEXP   <- sub("m|M", "6", economyData$PROPDMGEXP)
economyData$PROPDMGEXP   <- sub("b|B", "9", economyData$PROPDMGEXP)
economyData$CROPDMGEXP   <- sub("h|H", "2", economyData$CROPDMGEXP)
economyData$CROPDMGEXP   <- sub("k|K", "3", economyData$CROPDMGEXP)
economyData$CROPDMGEXP   <- sub("m|M", "6", economyData$CROPDMGEXP)
economyData$CROPDMGEXP   <- sub("b|B", "9", economyData$CROPDMGEXP)
economyData$propDMGQuant <- as.numeric(economyData$PROPDMG) * 10 ^ as.numeric(economyData$PROPDMGEXP)
```

```
## Warning: NAs introduced by coercion
```

```
economyData$cropDMGQuant <- as.numeric(economyData$CROPDMG) * 10 ^ as.numeric(economyData$CROPDMGEXP)
```

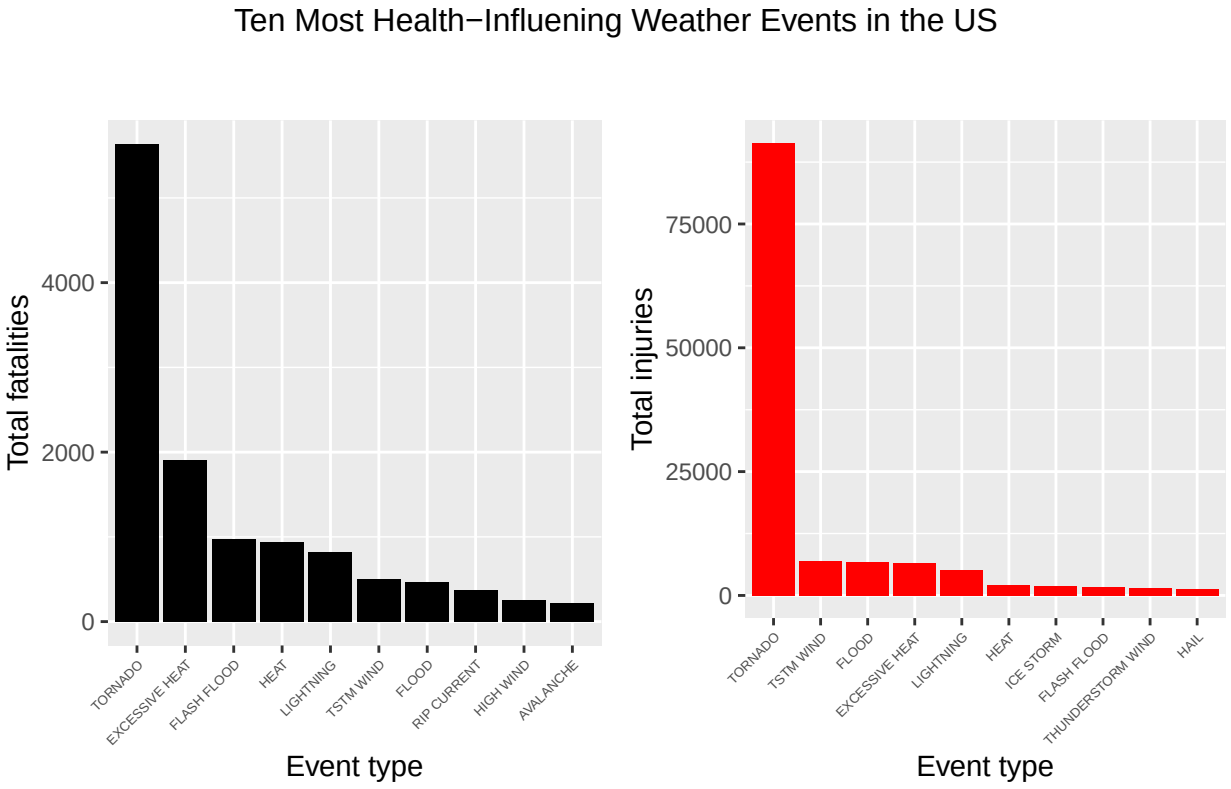
```
## Warning: NAs introduced by coercion
```

```
economyData$totalLoss    <- economyData$propDMGQuant + economyData$cropDMGQuant
economySum                <- aggregate(totalLoss ~ EVTYPE, economyData, sum, na.rm = TRUE)
economySum                <- economySum[order(economySum$totalLoss, decreasing = TRUE),]
economySum$totalLoss    <- economySum$totalLoss/10e6
head(economySum)
```

```
##               EVTYPE totalLoss
## 23             FLOOD 13800.744
## 61 HURRICANE/TYPHOON  2934.817
## 97           TORNADO  1657.033
## 56         HURRICANE  1240.527
## 74       RIVER FLOOD  1010.837
## 37              HAIL  1004.860
```

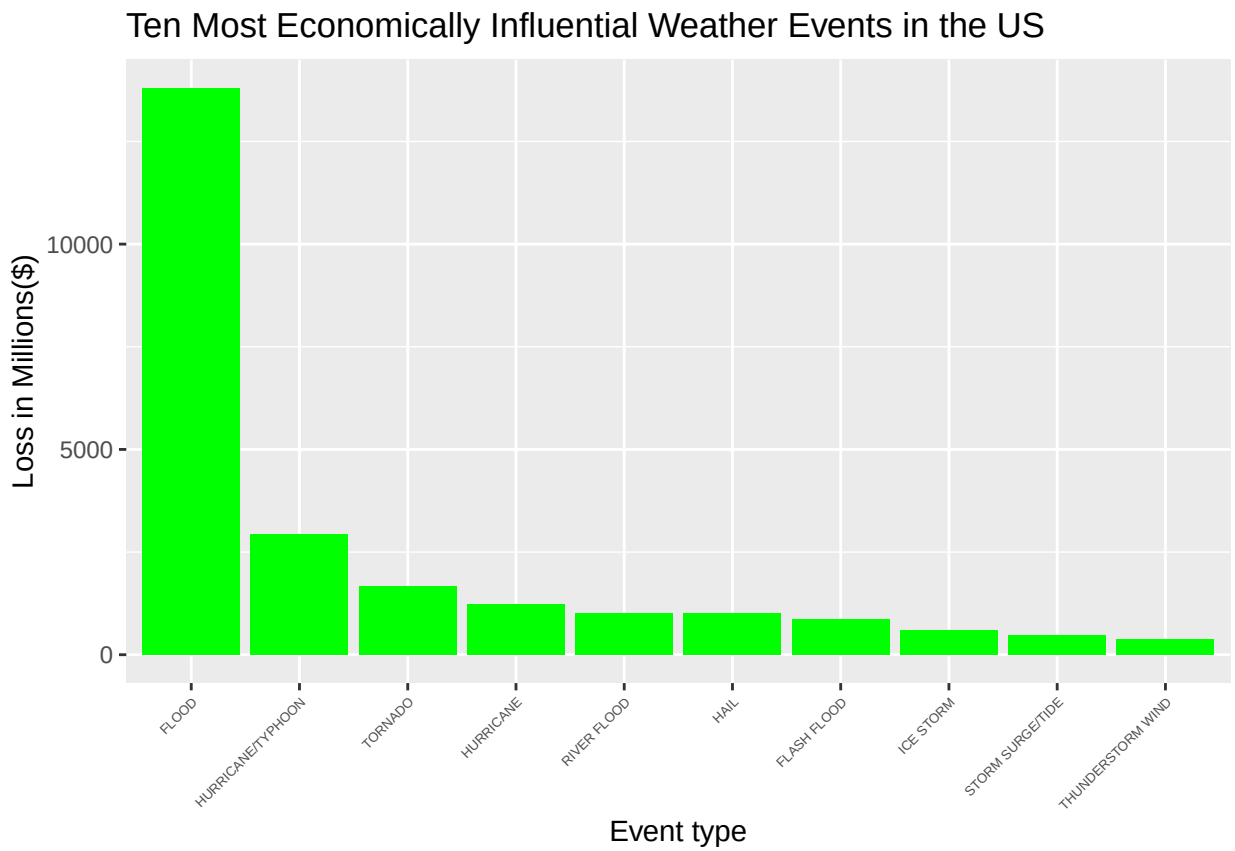
Now we have some clean data that can be used to analyze the different weather events and there impacts. We will take the top 10 events to illustrate this in a concise and readable form:

```
fatalPlot <- ggplot(data=head(fatalitiesSum,10), aes(x=reorder(EVTYPE, -FATALITIES), y=FATALITIES)) +
   geom_bar(fill="black",stat="identity")  +
    ylab("Total fatalities") + xlab("Event type") +
    theme(legend.position="none") + theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1))

injPlot <- ggplot(data=head(injuriesSum,10), aes(x=reorder(EVTYPE, -INJURIES), y=INJURIES)) +
   geom_bar(fill="red",stat="identity")  +
    ylab("Total injuries") + xlab("Event type") +
    theme(legend.position="none") + theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1))

pushViewport(viewport(layout = grid.layout(2, 2, heights = unit(c(1, 4), "null"))))
grid.text("Ten Most Health-Influening Weather Events in the US", vp = viewport(layout.pos.row = 1,
layout.pos.col = 1:2))
print(fatalPlot, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
print(injPlot, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
```



From the above plots we can see that a tornado would be most responsible for fatalities and injuries by a long way. For both categories there are different weather events contributing

```
economyPlot <- ggplot(data=head(economySum,10), aes(x=reorder(EVTYPE, -totalLoss), y=totalLoss)) +
   geom_bar(fill="green",stat="identity")  +
   ylab("Loss in Millions($)") + xlab("Event type") +
   theme(legend.position="none") + theme(axis.text.x = element_text(size = 5, angle = 45, hjust = 1)) +
   ggtitle("Ten Most Economically Influential Weather Events in the US")

grid.arrange(economyPlot, ncol = 1, nrow = 1)
```



From the plots we can see that floods are the most costly weather events by a long way.