title: "Reproducible Research Course Project 2"
output:
pdf_document: default

html_document: default

Synopsis

Storms and other severe w eather 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 w eather events in the United States, including w hen and w here they occur, as w ell as estimates of any fatalities, injuries, and property damage.

The analysis result show s that "Tornado" is the most harmful to population health, and "Flood" has the greatest economic consequences.

read the data

```
library(ggplot2)
library(dplyr)
fileURL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(fileURL,"Storm.csv.bz2")
data <- read.csv("Storm.csv.bz2")
```

Data Processing

Summary

```
dim(data)
names(data)
```

Get the variables related to our goal

1. evtype: w eather event type
2. FATALITIES: number of deaths
3. INJURIES: number of injuries
4. PROPDMG/propdmgexp/CROPDMG/cropdmgexp: property damage

```
sub_storm <- select(data, EVTYPE, FATALITIES:CROPDMGEXP)
```

First, w e have to replacement the categories for the event types

Then, for the property damage w e have to translate the values in propdmgexp and cropdmgexp:
Hh = 100
Kk = 1,000

Mm = 1,000,000
Bb = 1,000,000,000
number = number
others = 0

```
replace_types <- c("TORNADO","WIND","HAIL","HEAT","FLOOD","RAIN","STORM","HURRICANE", "SNOW", "LIGHTNING", "ICE")
sub_storm$event_type <- "OTHER"
for (i in replace_types){
  sub_storm$event_type[grep(i, sub_storm$EVTYPE, ignore.case = TRUE)] <- i
}

sub_storm$PROPDMGEXP[is.na(sub_storm$PROPDMGEXP)] <- 0
sub_storm$PROPDMGEXP[(sub_storm$PROPDMGEXP =='')|(sub_storm$PROPDMGEXP =='+')|(sub_storm$PROPDMGEXP =='?')|(sub_storm$PROP
sub_storm$PROPDMGEXP[(sub_storm$PROPDMGEXP =='h')|(sub_storm$PROPDMGEXP =='H')] <- 2
sub_storm$PROPDMGEXP[(sub_storm$PROPDMGEXP =='k')|(sub_storm$PROPDMGEXP =='K')] <- 3
sub_storm$PROPDMGEXP[(sub_storm$PROPDMGEXP =='m')|(sub_storm$PROPDMGEXP =='M')] <- 6
sub_storm$PROPDMGEXP[(sub_storm$PROPDMGEXP =='b')|(sub_storm$PROPDMGEXP =='B')] <- 9


sub_storm$CROPDMGEXP[is.na(sub_storm$CROPDMGEXP)] <- 0
sub_storm$CROPDMGEXP[(sub_storm$CROPDMGEXP =='')|(sub_storm$CROPDMGEXP =='+')|(sub_storm$CROPDMGEXP =='?')|(sub_storm$CROP
sub_storm$CROPDMGEXP[(sub_storm$CROPDMGEXP =='h')|(sub_storm$CROPDMGEXP =='H')] <- 2
sub_storm$CROPDMGEXP[(sub_storm$CROPDMGEXP =='k')|(sub_storm$CROPDMGEXP =='K')] <- 3
sub_storm$CROPDMGEXP[(sub_storm$CROPDMGEXP =='m')|(sub_storm$CROPDMGEXP =='M')] <- 6
sub_storm$CROPDMGEXP[(sub_storm$CROPDMGEXP =='b')|(sub_storm$CROPDMGEXP =='B')] <- 9

sub_storm$PROPDMG <- as.numeric(as.character(sub_storm$PROPDMG))
sub_storm$PROPDMGEXP <- as.numeric(as.character(sub_storm$PROPDMGEXP))
sub_storm$propertydmg <- sub_storm$PROPDMG*(10^sub_storm$PROPDMGEXP)

sub_storm$CROPDMG <- as.numeric(as.character(sub_storm$CROPDMG))
sub_storm$CROPDMGEXP <- as.numeric(as.character(sub_storm$CROPDMGEXP))
sub_storm$cropdmg <- sub_storm$CROPDMG*(10^sub_storm$CROPDMGEXP)
```

Sum the values into health_damage and economic_damage

```
sub_storm$FATALITIES <- as.numeric(as.character(sub_storm$FATALITIES))
sub_storm$INJURIES <- as.numeric(as.character(sub_storm$INJURIES))

health <- aggregate(x=list(health_damage = sub_storm$FATALITIES + sub_storm$INJURIES), by=list(event_type=sub_storm$event_
health <- health[order(health$health_damage, decreasing=TRUE),]


economic<- aggregate(x=list(economic_damage = sub_storm$propertydmg + sub_storm$cropdmg), by=list(event_type=sub_storm$eve
economic<- economic[order(economic$economic_damage, decreasing=TRUE),]
```

Get top 10 event type with highest health and economic damage

```
order1 <- health[1:10,]
order2 <- economic[1:10,]
```

Show the results in the plots

```
g1 <- ggplot(order1, aes(x=event_type,y=health_damage,fill=event_type))
g1 <- g1 + geom_bar(stat = "identity")
g1 <- g1 + labs(x="Event Type", y="Heal Damage", title="Top 10 events with highest Health Damage")
```

```
g1 <- g1 +  theme(axis.text.x =  element_text(size = 7,  angle = 20))
g1
```

```` ```{r } ````
```
g2 <- ggplot(order2, aes(x=event_type,y=log(economic_damage),fill=event_type))
g2 <- g2 + geom_bar(stat = "identity")
g2 <- g2 + labs(x="Event Type", y="Economic Damage", title="Top 10 events with highest Economic Damage")
g2 <- g2 + theme(axis.text.x = element_text(size = 7, angle = 20))
g2 <- g2 + scale_y_continuous(breaks = NULL)
g2
```
```` ``` ````