Strom Data Analysis
Reproducible Research - Peer Assessment 2
Analysis of the impact of Severe Weather Events on Public Health and Economy in the United States using US NOAA Strom Database.

SYNOPSIS

Storm and other severe weather events can cause both public health and economic problems for communities and municipalities. The U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database tracks characteristics of major storms and weather events in the United States, include when and where they occur, as well as estimates of any fatalities, injuries and property damage. The events in the database start in the year 1950 and end in November 2011.

The project involves exploring and analysing NOAA's storm database and finding answers to the following two questions:

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
Across the United States, which types of events have the greatest economic consequences?
This report contains results on the health and economic impact by the severe weather events based on the data from NOAA database. Tornado has caused the highest Fatalities and Injuries while Flood, Drought, Tornado and Typhoon are the top 3 events which caused greatest property damages.

DATA PROCESSING

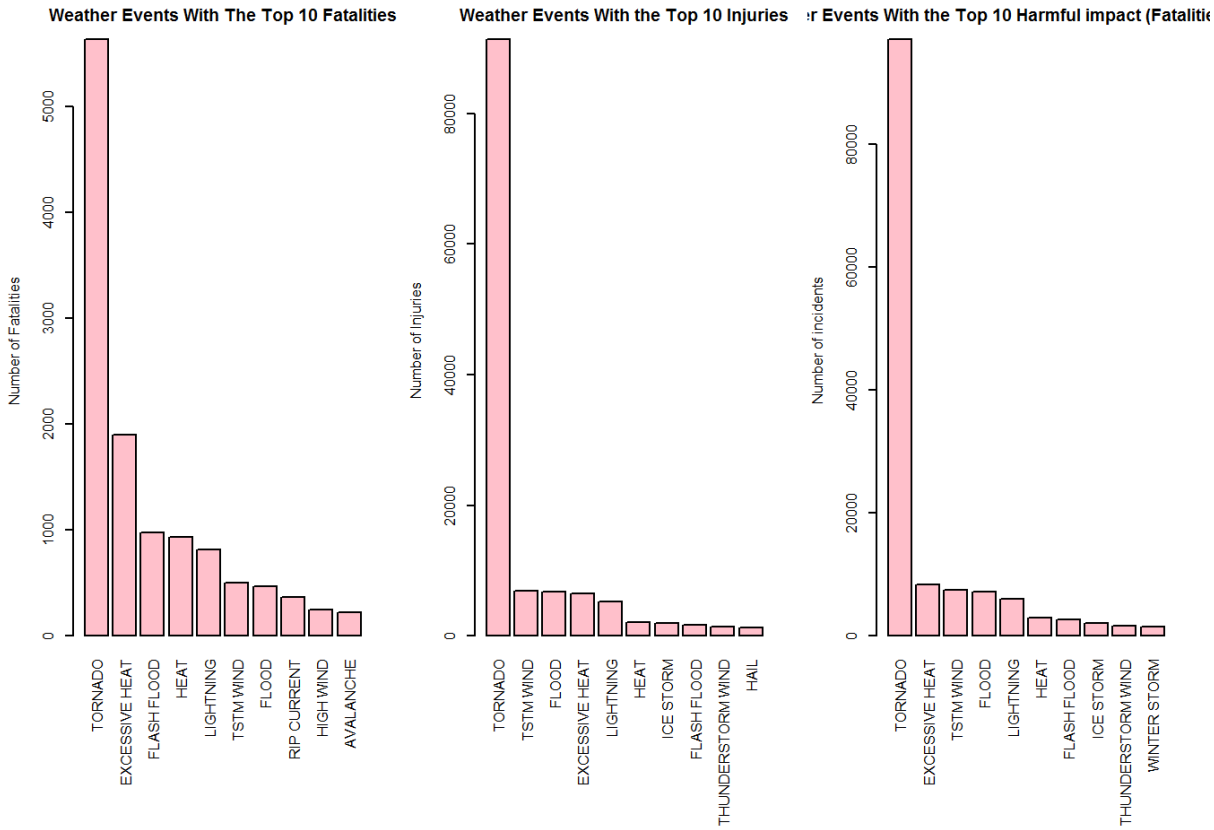
```
library(ggplot2)
library(plyr)

# Download file from URL provided in the assignment

setwd("D:/Data specialist course")

if (!file.exists("stormData.csv.bz2"))
{
  url <-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file(url,"stormData.csv.bz2")
}

# Unzip file
if (!file.exists("stormData.csv"))
{
   bunzip2("stormData.csv.bz2", "stormData.csv", remove = FALSE)
}

# load data into dataset stromData
stormData <- read.csv("stormData.csv")
dim(stormData)
```

```
names(stormData)

#Step 2: Extracting required data

stormDataNew <-
stormData[c("EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CROPDMG","CROPD
MGEXP")]
str(stormDataNew)

head(stormDataNew)

#Step 3: Identifying events that are most harmful with respect to population health

Fatal <- aggregate(FATALITIES ~ EVTYPE, data = stormDataNew, FUN = sum)
Injury <- aggregate(INJURIES ~ EVTYPE, data = stormDataNew, FUN = sum)
TotalHarm <- aggregate(FATALITIES + INJURIES ~ EVTYPE, data = stormDataNew, FUN = sum)

names(TotalHarm)[2] <- "Total"


# Extract top 10 events with highest fatalities
FatalTop10 <- Fatal[order(-Fatal$FATALITIES), ][1:10, ]

# Extract top 10 event with highest injuries
InjuryTop10 <- Injury[order(-Injury$INJURIES), ][1:10, ]

# Extract top 10 event with highest harm (injuries+fatalities)
TotalHarmTop10 <- TotalHarm[order(-TotalHarm$Total), ][1:10, ]

#Step 4: Identifying events that have the greatest economic consequences

unique(stormDataNew$PROPDMGEXP)

stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "h"] <- 100
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "H"] <- 100
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "K"] <- 1000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "M"] <- 1000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "m"] <- 1000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "B"] <- 1000000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == ""] <- 1
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "0"] <- 1
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "1"] <- 10
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "2"] <- 100
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "3"] <- 1000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "4"] <- 10000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "5"] <- 100000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "6"] <- 1000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "7"] <- 10000000
```

```
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "8"] <- 100000000

# Invalid exponent values ("=", "-", "?") are given value zero
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "+"] <- 0
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "-"] <- 0
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "?"] <- 0

stormDataNew$PROPDMGVAL <- stormDataNew$PROPDMG * stormDataNew$PROPEXP

unique(stormDataNew$CROPDMGEXP)

stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == ""] <- 1
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "0"] <- 1
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "2"] <- 100
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "K"] <- 1000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "k"] <- 1000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "M"] <- 1000000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "m"] <- 1000000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "B"] <- 1000000000

# Invalid exponent values ("=", "-", "?") are given value zero

stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "?"] <- 0

stormDataNew$CROPDMGVAL <- stormDataNew$CROPDMG * stormDataNew$CROPEXP

PropertyDamage <- aggregate(PROPDMGVAL ~ EVTYPE, data = stormDataNew, FUN = sum)
CropDamage <- aggregate(CROPDMGVAL ~ EVTYPE, data = stormDataNew, FUN = sum)
TotalDamage <- aggregate(PROPDMGVAL + CROPDMGVAL ~ EVTYPE, data = stormDataNew, FUN =
sum )
names(TotalDamage)[2] <- "Total"

# Extract top 10 events with highest Property Damage
PropertyDamageTop10 <- PropertyDamage [order(-PropertyDamage$PROPDMGVAL), ][1:10, ]

# Extract top 10 events with highest Crop Damage
CropDamageTop10 <- CropDamage [order(-CropDamage$CROPDMGVAL), ][1:10, ]

# Extract  top 10 events with highest Total Damage
TotalDamageTop10 <- TotalDamage [order(-TotalDamage$Total), ][1:10, ]
```

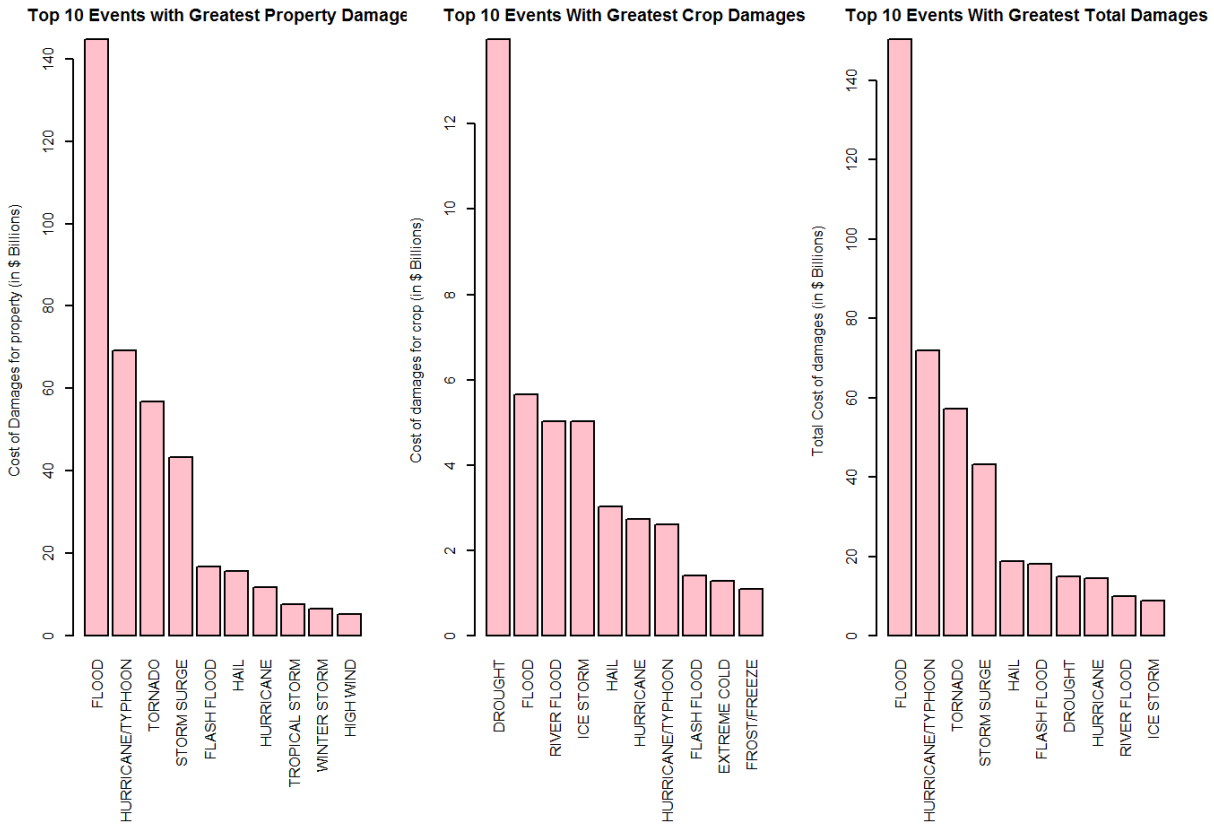
#Plot 1 : Bar graph of the top 10 events based on Fatalities and Injuries

```
par(mfrow = c(1, 3), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.5)
barplot( FatalTop10$FATALITIES,
      las = 3,
      names.arg = FatalTop10$EVTYPE,
      main = "Weather Events With The Top 10 Fatalities",
      ylab = "Number of Fatalities",
      col = "pink")

barplot(InjuryTop10$INJURIES,
      las = 3,
      names.arg = InjuryTop10$EVTYPE,
      main = "Weather Events With the Top 10 Injuries",
      ylab = "Number of Injuries",
      col = "pink")

barplot(TotalHarmTop10$Total,
      las = 3,
      names.arg = TotalHarmTop10$EVTYPE,
      main = "Weather Events With the Top 10 Harmful impact (Fatalities+Injuries)",
      ylab = "Number of incidents",
      col = "pink")
```



#Plot 2: Bar graph of the top 10 events based on damage to Property and Crop

```
par(mfrow = c(1, 3), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.5)
barplot(PropertyDamageTop10$PROPDMGVAL/(10^9),
        las = 3,
        names.arg = PropertyDamageTop10$EVTYPE,
        main = "Top 10 Events with Greatest Property Damages",
        ylab = "Cost of Damages for property (in $ Billions)",
        col = "pink")

barplot(CropDamageTop10$CROPDMGVAL/(10^9),
        las = 3,
        names.arg = CropDamageTop10$EVTYPE,
        main = "Top 10 Events With Greatest Crop Damages",
        ylab = "Cost of damages for crop (in $ Billions)",
        col = "pink")

barplot(TotalDamageTop10$Total/(10^9),
        las = 3,
        names.arg = TotalDamageTop10$EVTYPE,
        main = "Top 10 Events With Greatest Total Damages",
        ylab = "Total Cost of damages (in $ Billions)",
        col = "pink")
```



CONCLUSIONS

Across United States, based on the data provided for storms between 1950 and Nov 2011
Tornado has caused the highest Fatalities and Injuries (As seen from the Plot1)

Flood, Drought, Tornado and Typhoon are the top 3 events which caused greatest property damages (As seen from the plot 2)

Drought, Flood, River Flood and Ice Storm are the top 3 events which caused greatest crop damages (As seen from the plot 2)

Flood has caused maximum total damage (Property + Crop) (As seen from the plot 2) and has the greatest economic consequence.