Reproducible Research Peer Assignment 2

Title: Anaylsis the effect of Storms and other severe weather events on public health and economic.

Synopsis

This project is to analysis the few top types of Storms and other severa weather events that bring damages to the people (fatalities and injuries) and monetary damages (properties and crops damages). This analysis is based on the storm data from U.S. National Oceanic Atmospheric Administration (NOAA), which recorded the happening timing, fatalities, injuries and property damage of storms and weather events in United States span from 1950 to 2011. In this analysis, **Tornado** is the most harmful event to population health as shown in plots below, while **Flood** turned out to be the event that caused the greatest economy consequences.

Libraries loading and Data processing.

1. Load necessarily libraries.

```
library(knitr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lubridate)
library(plyr)
```

```
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following object is masked from 'package:lubridate':
##
##     here
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

2. Load and process data file.

```
# Assumption:
# - Set the working directory in source File location .
# - Download the "repdata-data-StormData.csv.bz2" data file from the download link indicated in Courser
subsetData <- read.table(bzfile("repdata-data-StormData.csv.bz2"), header = TRUE, sep = ",")

# Convert into factor after converting to upper case
subsetData$EVTYPE <- toupper(subsetData$EVTYPE)
subsetData$EVTYPE <- as.factor(subsetData$EVTYPE)
```

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

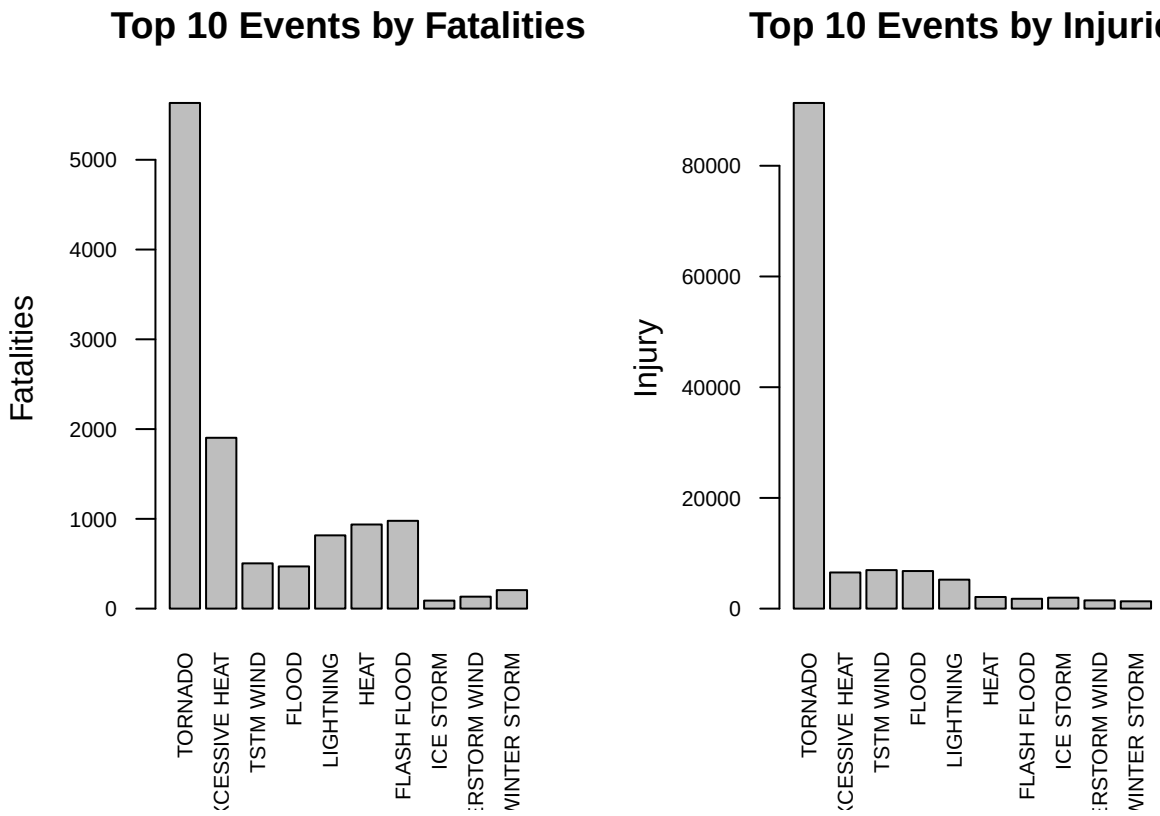
Result

```
# Subset the retrieved data to the data that just includes the events or damages and health.
subset_Item <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
subsetData <- subsetData[subset_Item]

# Use FATALITIES != 0 and INJURIES != 0 as a filter to subset the harmful factors to population.
df1 <- subset(subsetData, !(FATALITIES == 0 & INJURIES == 0), select = c(EVTYPE, FATALITIES, INJURIES))

# Create new data frame by split the analysed data and apply "summarize" and "sum" function on it.
harmfulFactor_1 <- ddply( df1,.(EVTYPE),summarize, sum_FATALITIES = c(sFATALITIES = sum(FATALITIES)),
                          sum_INJURIES = sum(INJURIES), total_sum = sum(INJURIES) + sum(FATALITIES) )


# Display the top 10 harmful event types.
top10_HarmfulEvent_1 <-head( arrange(harmfulFactor_1, desc(total_sum)), n = 10)
top10_HarmfulEvent_1
```

```
##               EVTYPE sum_FATALITIES sum_INJURIES total_sum
## 1            TORNADO           5633        91346     96979
## 2     EXCESSIVE HEAT           1903         6525      8428
## 3          TSTM WIND            504         6957      7461
## 4              FLOOD            470         6789      7259
## 5          LIGHTNING            816         5230      6046
## 6               HEAT            937         2100      3037
## 7        FLASH FLOOD            978         1777      2755
## 8          ICE STORM             89         1975      2064
## 9  THUNDERSTORM WIND            133         1488      1621
## 10      WINTER STORM            206         1321      1527
```

```
# Plot a bar chart to show top 10 harmful events that affects population health.
# Set Graphical parameters of a vector form c( nr, nc), Subsequent figures will be drawn in an nr-by-nc
par(mfrow = c(1, 2))

# Plot the Top 10 Events by FATALITIES on the left side.
barplot(top10_HarmfulEvent_1$sum_FATALITIES, names.arg = top10_HarmfulEvent_1$EVTYPE, xlab = "",
        ylab = "Fatalities", main = "Top 10 Events by Fatalities", las = 2,cex.names = 0.7,cex.axis = 0
```

```r
# Plot the Top 10 Events by INJURIES on the right side.
barplot(top10_HarmfulEvent_1$sum_INJURIES, names.arg = top10_HarmfulEvent_1$EVTYPE, xlab = "",
        ylab = "Injury", main = "Top 10 Events by Injuries", las = 2, cex.names = 0.7, cex.axis = 0.7)
```



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

Result

```r
# Data is process to present the actual exponential damage to properties.
subsetData$PROPDMGEXP <- as.character(subsetData$PROPDMGEXP)
subsetData$PROPDMGEXP[grep("K", subsetData$PROPDMGEXP)] <- "1000"
subsetData$PROPDMGEXP[grep("M", subsetData$PROPDMGEXP)] <- "1000000"
subsetData$PROPDMGEXP[grep("m", subsetData$PROPDMGEXP)] <- "1000000"
subsetData$PROPDMGEXP[grep("B", subsetData$PROPDMGEXP)] <- "1000000000"
others <- subsetData$PROPDMGEXP %in% c("1000","1000000","1000000000") == F
subsetData$PROPDMGEXP[others == T] <- "1"
subsetData$PROPDMGEXP <- as.numeric(subsetData$PROPDMGEXP)

# Data is process to present the actual exponential damage to crops.
subsetData$CROPDMGEXP <- as.character(subsetData$CROPDMGEXP)
subsetData$CROPDMGEXP[grep("K", subsetData$CROPDMGEXP)] <- "1000"
subsetData$CROPDMGEXP[grep("M", subsetData$CROPDMGEXP)] <- "1000000"
```

```
subsetData$CROPDMGEXP[grep("m", subsetData$CROPDMGEXP)] <- "1000000"
subsetData$CROPDMGEXP[grep("B", subsetData$CROPDMGEXP)] <- "1000000000"
others <- subsetData$CROPDMGEXP %in% c("1000","1000000","1000000000") == F
subsetData$CROPDMGEXP[others == T] <- "1"
subsetData$CROPDMGEXP <- as.numeric(subsetData$CROPDMGEXP)

# Create new data frame to store actual damage.
subsetData$ACTPORPDMG <- subsetData$PROPDMG * subsetData$PROPDMGEXP
subsetData$ACTCROPDMG <- subsetData$CROPDMG * subsetData$CROPDMGEXP

# Create new data frame to record the total of injuries and fatalities.
subsetData$TOTAL_HARM <- subsetData$INJURIES + subsetData$FATALITIES


# Create new data subsets that contains event brings most damages to properties.
sum_propDmg <- aggregate(subsetData[,"ACTPORPDMG"], by = list(subsetData$EVTYPE), FUN = sum, na.rm = TRU
names(sum_propDmg) <- c("EVTYPE","ACTPORPDMG")

# Arrange the components of the subset data "sum_propDmg" according the ACTPORPDMG column in decreasing
sum_propDmg <- sum_propDmg[ order(sum_propDmg$ACTPORPDMG, decreasing = TRUE), ]

# To find out which Event Type that causes most damages to properties.
sum_propDmg$EVTYPE[which.max(sum_propDmg$ACTPORPDMG)]
```

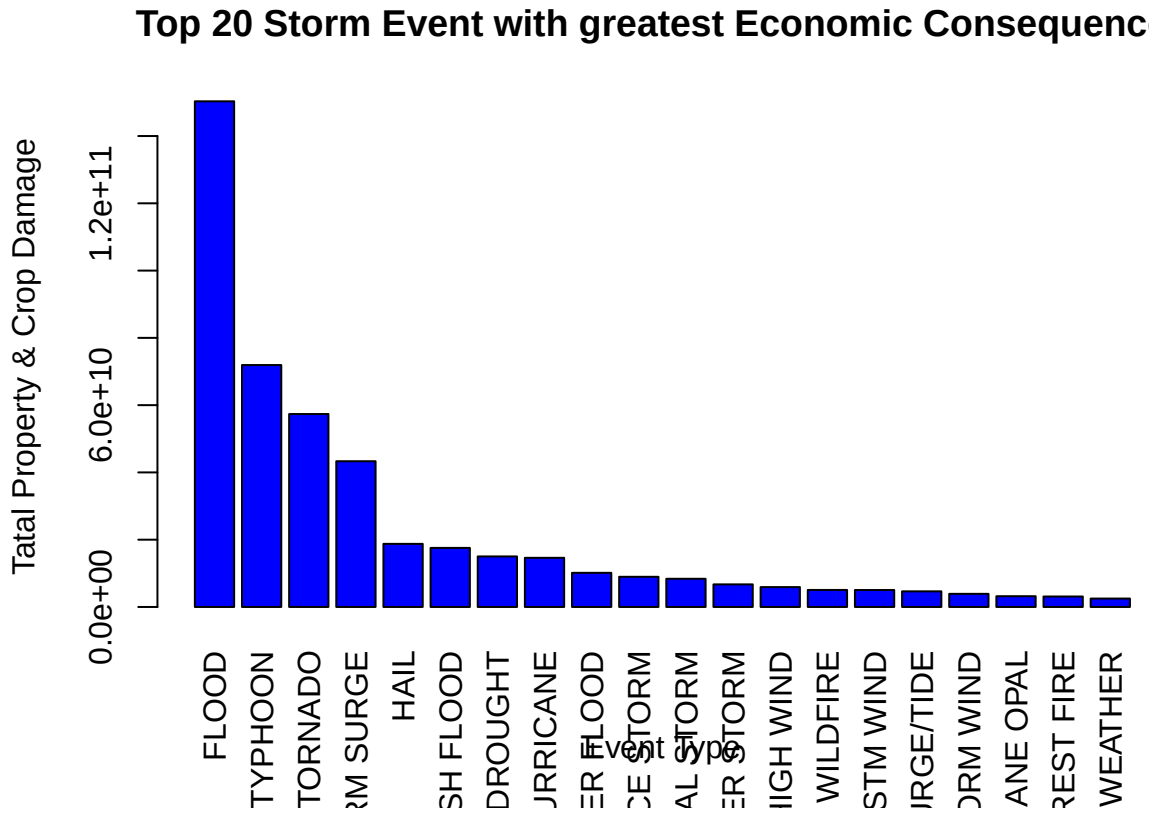
```
## [1] FLOOD
## 898 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```

```
# Create new data subsets that contains event brings most damages to crops.
sum_cropDmg <- aggregate( subsetData[,"ACTCROPDMG"], by=list(subsetData$EVTYPE), FUN = sum, na.rm = TRU
names(sum_cropDmg) <- c("EVTYPE","ACTCROPDMG")

# Arrange the components of the subset data "sum_propDmg" according the ACTPORPDMG column in decreasing
sum_cropDmg <- sum_cropDmg[ order(sum_cropDmg$ACTCROPDMG, decreasing = TRUE), ]

# To find out the Event Type that causes most damages to crops.
sum_cropDmg$EVTYPE[which.max(sum_cropDmg$ACTCROPDMG)]
```

```
## [1] DROUGHT
## 898 Levels:    HIGH SURF ADVISORY  COASTAL FLOOD ... WND
```

```
# Create new data frame for storing total damage.
subsetData$TOTAL_DMG <- subsetData$ACTPORPDMG + subsetData$ACTCROPDMG

# Create new data subsets that contains total damage according the list of Event type.
sum_Events <- aggregate( subsetData[,"TOTAL_DMG"], by = list(subsetData$EVTYPE), FUN = sum, na.rm = TRU

# Set new vector name for this newly created subset data.
names(sum_Events) <- c("EVTYPE","TOTAL_DMG")

# Arrange the components of the subset data "sum_Events" according the TOTA_DMG column in decreasing or
sum_Events <- sum_Events[ order(sum_Events$TOTAL_DMG, decreasing = TRUE), ]
```

```r
# Plot bar graph for top 20 total economic consequences.
barplot( height = sum_Events$TOTAL_DMG[1:20], names.arg = sum_Events$EVTYPE[1:20],
         las = 3, xlab="Event Type", ylab = "Tatal Property & Crop Damage",
         main = "Top 20 Storm Event with greatest Economic Consequences", col = "blue")
```



Summary

Based on the analysis and shown in plots, **Tornado** is the event type that causes most fatalities and injuries. However **Flood** is the greatest economic consequences and it is observable in the plot of Top 20 event type that causes Economic Consequences. **Flood** is also the event type that bring more damages to properties and **Drought** is the event type that causes more damages to crop.