Reproducible Research - Course Project 2

1 - Project Title:

Impact Analysis US Natural Disaster to Human Health and Economy.

---

2 - Synopsis:

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.

---

3 - Data Processing:

3.1 - Data Loading:

Extract the data into a dataframe.

```
    download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData
.csv.bz2",
                  dest="tmp.bz2",
                  method="curl")
    data <- read.csv(bzfile("tmp.bz2"),
                     header=TRUE,
                     sep=",",
                     stringsAsFactors=FALSE)
```

3.2 - Data Subsetting:

Subset the dataset on the parameters of interest.

```
    # Change parameter names to lowercase.
```

```
    colnames(data) <- tolower(colnames(data))


    # Subset on the parameters of interest.
    data <- subset(x=data,
                   subset=(evtype != "?" &
                               (injuries > 0 | fatalities > 0 | propdmg > 0 | cropd
mg > 0)),
                   select=c("evtype",
                            "fatalities",
                            "injuries",
                            "propdmg",
                            "propdmgexp",
                            "cropdmg",
                            "cropdmgexp"))
```

3.3 - Data Cleansing:

Map the property and crop damage exponent alphabetic multipliers to numeric values.

```
    # Change all damage exponents to uppercase.
    data$propdmgexp <- toupper(data$propdmgexp)
    data$cropdmgexp <- toupper(data$cropdmgexp)


    # Map property damage alphanumeric exponents to numeric values.
    propDmgKey <-  c("\"\"" = 10^0,
                     "-" = 10^0,
                     "+" = 10^0,
                     "0" = 10^0,
                     "1" = 10^1,
                     "2" = 10^2,
                     "3" = 10^3,
                     "4" = 10^4,
                     "5" = 10^5,
                     "6" = 10^6,
```

```
                    "7" = 10^7,
                    "8" = 10^8,
                    "9" = 10^9,
                    "H" = 10^2,
                    "K" = 10^3,
                    "M" = 10^6,
                    "B" = 10^9)
    data$propdmgexp <- propDmgKey[as.character(data$propdmgexp)]
    data$propdmgexp[is.na(data$propdmgexp)] <- 10^0

    # Map crop damage alphanumeric exponents to numeric values
    cropDmgKey <-  c("\"\"" = 10^0,
                    "?" = 10^0,
                    "0" = 10^0,
                    "K" = 10^3,
                    "M" = 10^6,
                    "B" = 10^9)
    data$cropdmgexp <- cropDmgKey[as.character(data$cropdmgexp)]
    data$cropdmgexp[is.na(data$cropdmgexp)] <- 10^0
```

3.4 - Human Heath Data Processing:

Select the applicable health columns from the dataset, then calculate the total number of fatalities and injuries per event type.

```
    # Aggregate number of fatalities and injuries per evtype into healthData datafr
ame
    healthData <- aggregate(cbind(fatalities, injuries) ~ evtype, data=data, FUN=su
m)
    # Add total column to healthData
    healthData$total <- healthData$fatalities + healthData$injuries
```

Find the event types corresponding with the the highest health impacts.

```
    # Remove rows with zero health impact
    healthData <- healthData[healthData$total > 0, ]
    # Sort health data in descending order
```

```
    healthData <- healthData[order(healthData$total, decreasing=TRUE), ]

    # Re-label the rows

    rownames(healthData) <- 1:nrow(healthData)

    # Create dataframe of highest health impacting event types and append an "other
" event type as a catchall

    # for everything else

    healthDataTop <- healthData[1:10, ]
```

3.5 - Economic Data Processing:

Combine the damage and damage exponent multiplier parameters into the single parameters propertyloss and croploss.

```
    # Combine propdmg and propdmgexp parameters into a single parameter called prop
ertyloss.

    data$propertyloss <- data$propdmg * data$propdmgexp

    # Combine cropdmg and cropdmgexp parameters into a single parameter called crop
loss.

    data$croploss <- data$cropdmg * data$cropdmgexp
```

Select the applicable economic columns from the dataset, then calculate the total amount of property loss and crop loss per event type.

```
    # Aggregate amount of proploss and croploss per evtype into economicData datafr
ame

    economicData <- aggregate(cbind(propertyloss, croploss) ~ evtype, data=data, FU
N=sum)

    # Add total loss column to economicData

    economicData$total <- economicData$propertyloss + economicData$croploss
```

Find the event types corresponding with the highest economic impacts.

```
    # Remove rows with zero economic impact

    economicData <- economicData[economicData$total > 0, ]

    # Sort the economy data in descending order

    economicData <- economicData[order(economicData$total, decreasing=TRUE), ]

    # Re-label the rows

    rownames(economicData) <- tolower(rownames(economicData))

    # Create dataframe of highest economy impacting event types

    economicDataTop <- economicData[1:10, ]
```

4 - Analysis Results:

Figure 4.1: Top 10 Health Impact Event Types

Plot of the ten event types with the highest fatality counts plus an eleventh catchall event type that combines the total fatality counts of all other event types.

```
    # Load necessary libraries

    library(reshape2)

    library(ggplot2)



    # Melt the data

    healthDataTopMelt <- melt(healthDataTop, id.vars="evtype")



    # Create chart

    healthChart <- ggplot(healthDataTopMelt, aes(x=reorder(evtype, -value), y=value
))

    # Plot data as bar chart

    healthChart = healthChart + geom_bar(stat="identity", aes(fill=variable), posit
ion="dodge")

    # Format y-axis scale and set y-axis label

    healthChart = healthChart + scale_y_sqrt("Frequency Count")

    # Set x-axis label

    healthChart = healthChart + xlab("Event Type")

    # Rotate x-axis tick labels

    healthChart = healthChart + theme(axis.text.x = element_text(angle=45, hjust=1)
)

    # Set chart title

    healthChart = healthChart + ggtitle("Pareto Chart of Top 10 US Storm Health Imp
acts")

    # Display the chart

    print(healthChart)
```

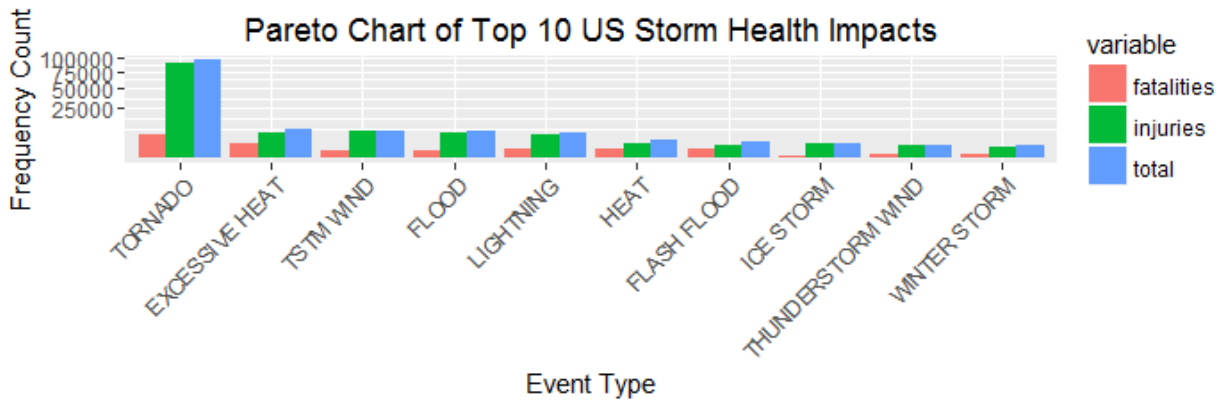
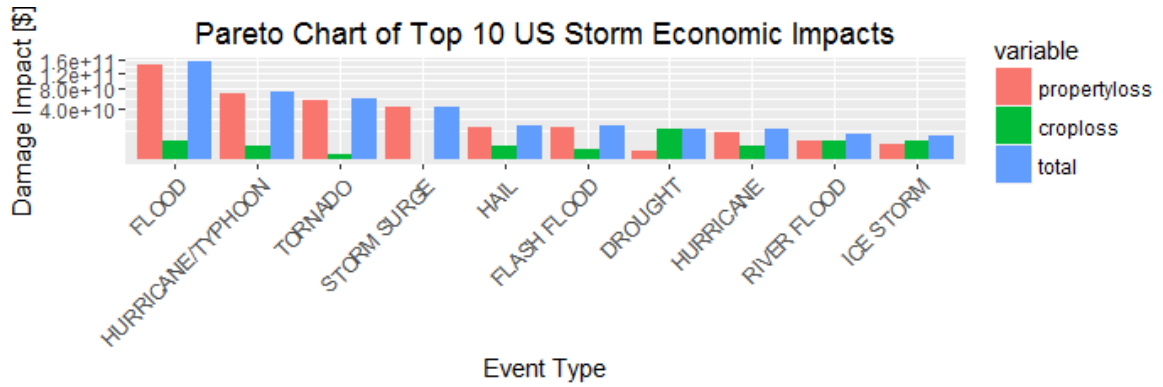


Figure 4.2: Top 10 Economic Impact Event Types

Plot of the ten event types with the highest economic impacts.

```
    # Load necessary libraries

    library(reshape2)

    library(ggplot2)



    # Melt the data

    economicDataTopMelt <- melt(economicDataTop, id.vars="evtype")



    # Create chart

    economicChart <- ggplot(economicDataTopMelt, aes(x=reorder(evtype, -value), y=v
alue))

    # Add bars

    economicChart <- economicChart + geom_bar(stat="identity", aes(fill=variable),
position="dodge")

    # Format y-axis scale and set y-axis label

    economicChart <- economicChart + scale_y_sqrt("Damage Impact [$]")

    # Set x-axis label

    economicChart <- economicChart + xlab("Event Type")

    # Rotate x-axis tick labels

    economicChart <- economicChart + theme(axis.text.x = element_text(angle=45, hju
st=1))

    # Set chart title
```

```
    economicChart <- economicChart + ggtitle("Pareto Chart of Top 10 US Storm Econo
mic Impacts")

    # Display the chart

    print(economicChart)
```



---

5 - Conclusions

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

From the graph, Tornadoes are lead most harmful to deaths and injuries out of all event types.

5.2 - Across the United States, which types of events have the greatest economic consequences?

Flooding is responsible for the largest proportion of total economic impact out of all event types.