Reproducible Research Project 2

Induction

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.

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events.

The followong analysis investigates Fatalities to: - Health (injuries and fatalities) - Properties and Crops (economic consequences)

1. Across the United States, which types of events are most harmful with respect to population health?

Data Processing

```
#importing library for plot
library(ggplot2)
Url<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
#Downloading the Data
download.file(Url,destfile = "storm_data.csv")
storm_data<-read.csv("storm_data.csv")
head(storm_data)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL TORNADO
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5         0                                               0         NA
## 6         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0                      14.0   100 3   0          0       15    25.0
```

```
## 2         0                    2.0   150 2   0          0        0     2.5
## 3         0                    0.1   123 2   0          0        2    25.0
## 4         0                    0.0   100 2   0          0        2     2.5
## 5         0                    0.0   150 2   0          0        2     2.5
## 6         0                    1.5   177 2   0          0        6     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
## 6          K       0                                         3450      8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
## 6          0          0              6
```

Exploring Columns

```
colnames(storm_data)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Extracting Necessary columns

```
storm_event<-storm_data[, c("BGN_DATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "C
summary(storm_event)
```

```
##    BGN_DATE           EVTYPE          FATALITIES          INJURIES
##  Length:902297      Length:902297      Min.   :  0.0000   Min.   :   0.0000
##  Class :character   Class :character   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Mode  :character   Mode  :character   Median :  0.0000   Median :   0.0000
##                                        Mean   :  0.0168   Mean   :   0.1557
##                                        3rd Qu.:  0.0000   3rd Qu.:   0.0000
##                                        Max.   :583.0000   Max.   :1700.0000
##     PROPDMG         PROPDMGEXP           CROPDMG         CROPDMGEXP
##  Min.   :   0.00   Length:902297      Min.   :  0.000   Length:902297
##  1st Qu.:   0.00   Class :character   1st Qu.:  0.000   Class :character
##  Median :   0.00   Mode  :character   Median :  0.000   Mode  :character
##  Mean   :  12.06                      Mean   :  1.527
##  3rd Qu.:   0.50                      3rd Qu.:  0.000
##  Max.   :5000.00                      Max.   :990.000
```

5 Events that contributes to most injuries and Fatalities

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
Total_injuries<-storm_event %>% group_by(EVTYPE) %>% summarise(FATALITIES = sum(FATALITIES),
            INJURIES = sum(INJURIES), totals = sum(FATALITIES) + sum(INJURIES))

Total_injuries<-Total_injuries[order(-Total_injuries$FATALITIES),]
head(Total_injuries,5)
```

```
## # A tibble: 5 x 4
##   EVTYPE         FATALITIES INJURIES totals
##   <chr>               <dbl>    <dbl>  <dbl>
## 1 TORNADO              5633    91346  96979
## 2 EXCESSIVE HEAT       1903     6525   8428
## 3 FLASH FLOOD           978     1777   2755
## 4 HEAT                  937     2100   3037
## 5 LIGHTNING             816     5230   6046
```

Reshaping The data for plots

```
library(reshape)
```

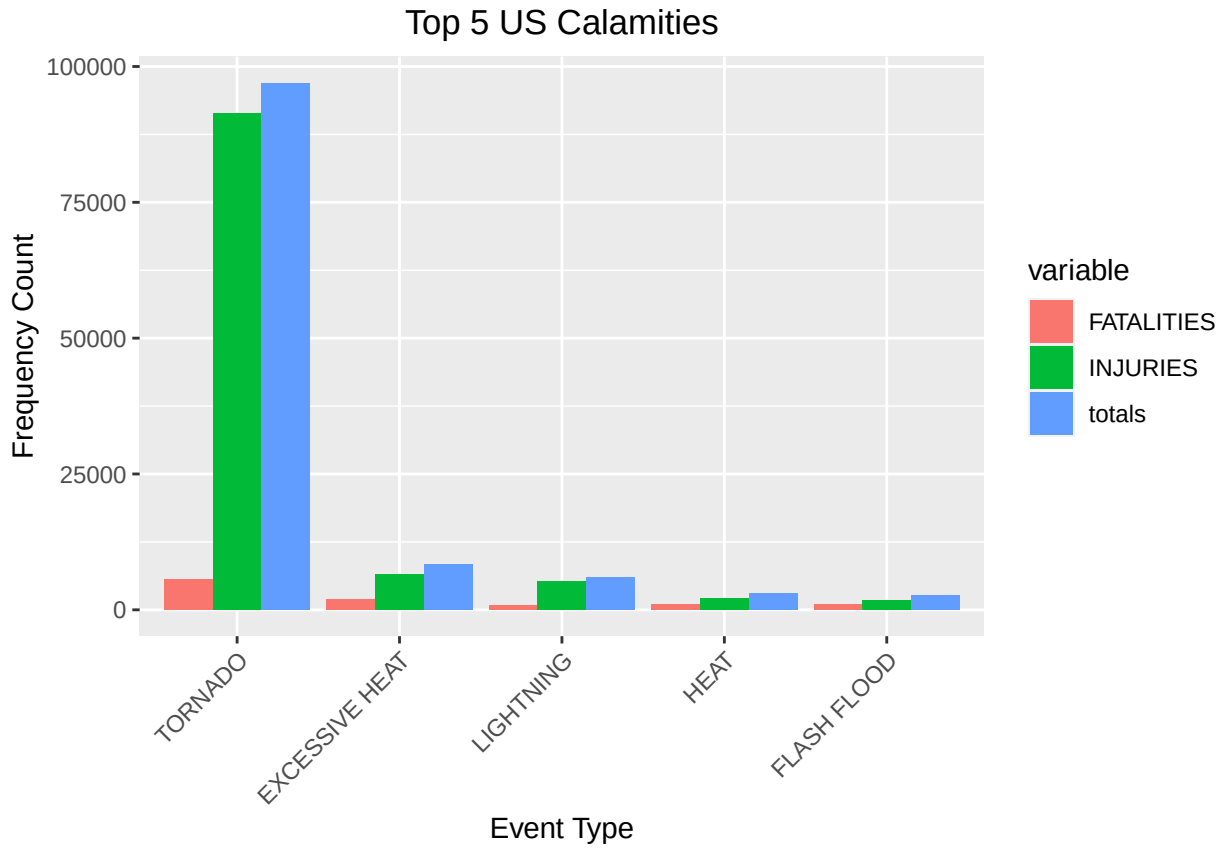
```
## Warning: package 'reshape' was built under R version 4.3.2

##
## Attaching package: 'reshape'

## The following object is masked from 'package:dplyr':
##
##     rename
```

```
New_data<-as.data.frame(head(Total_injuries,5))
#Reshaping the dataset for plot
df<-melt(New_data, id.vars="EVTYPE")
colnames(df)
```

```
## [1] "EVTYPE"   "variable" "value"
```

Results

```
# Create chart
ggplot(df,aes(x=reorder(EVTYPE,-value),y=value, fill=variable))+
    geom_bar( stat = "identity",position="dodge")+ylab("Frequency Count")+ theme(plot.title = element_t
  ggtitle("Top 5 US Calamities") + theme(plot.title = element_text(hjust = 0.5))
```



- The bar chart clearly shows that tornadoes are the primary cause of injuries and fatalities resulting from natural disasters in the US.

2. Which types of events have the greatest economic consequences in US?

Filtering DATA

```
unique(storm_event$PROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
```

```
unique(storm_event$CROPDMGEXP)
```

```
## [1] ""  "M" "K" "m" "B" "?" "0" "k" "2"
```

```
# Map property damage alphanumeric exponents to numeric values.
storm_event$PROPDMGEXP_number <-  recode(storm_event$PROPDMGEXP,
                  " " = 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,
                  .default = 10^0)

# Map crop damage alphanumeric exponents to numeric values
storm_event$CROPDMGEXP_number <-  recode(storm_event$CROPDMGEXP,
                                  " " = 10^0,
                                  "?" = 10^0,
                                  "0" = 10^0,
                                  "K" = 10^3,
                                  "M" = 10^6,
                                  "B" = 10^9,
                                  .default = 10^0)

#Cost of Damaged Property and Crop
storm_event$PropCost<-storm_event$PROPDMG * storm_event$PROPDMGEXP_number
storm_event$CropCost<-storm_event$CROPDMG * storm_event$CROPDMGEXP_number
```

Preparing Dataset for plotting

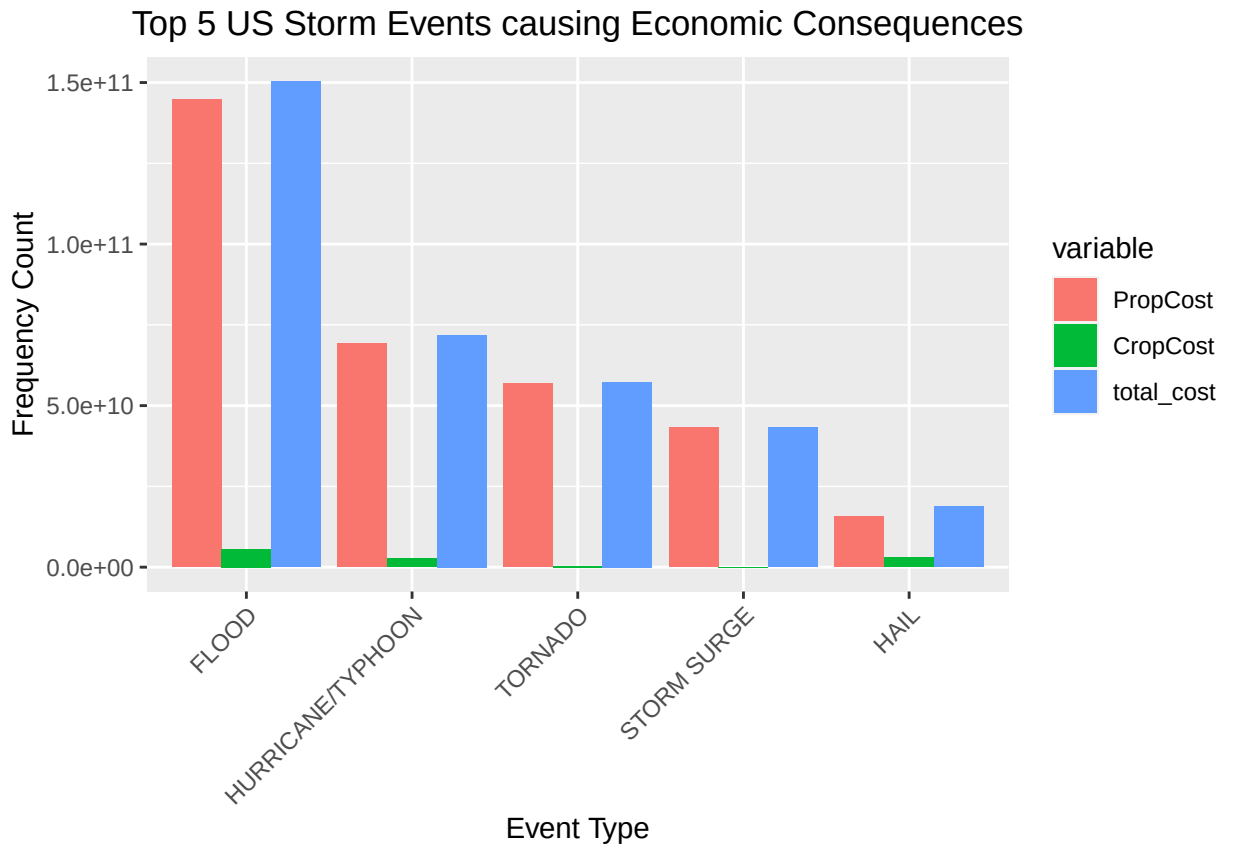
```
TotalCost<-storm_event %>% group_by(EVTYPE) %>% summarise(PropCost = sum(PropCost),
            CropCost = sum(CropCost), total_cost = sum(PropCost) + sum(CropCost))

TotalCost<-TotalCost[order(-TotalCost$total_cost),]
```

```
New_cost<-as.data.frame(head(TotalCost,5))
df_cost<-melt(New_cost, id.vars="EVTYPE")
colnames(df_cost)
```

```
## [1] "EVTYPE"   "variable" "value"
```

Results

```
ggplot(df_cost,aes(x=reorder(EVTYPE,-value),y=value, fill=variable))+
    geom_bar( stat = "identity",position="dodge")+ylab("Frequency Count")+ theme(plot.title = element_t
  theme(plot.title = element_text(hjust = 0.5))
```



-In the United States, floods cause the most financial damage from natural disasters, with hurricanes or typhoons coming in a close second.