Weather Data Analysis

Lucas Falcão Monteiro

10/08/2021

Analyzing harmful weather events

Synopsis

The National Oceanic & Atmospheric Administration publishes data regarding weather-related events and their impacts on the health of the population and the economy. This project aims to analyze the data and identify which types of events pose the greatest threat to the USA. These events were then inspected more closely to provide a better description of how they may impact the lives and the economy of Americans.

Loading and processing Raw Data

The data is available from the National Oceanic & Atmospheric Administration website. The data was compiled by the instructors from the Reproducible Research course from Coursera and made available in the following link. We unzip and read the data into the variable **noaa**.

```
noaa <- read.csv('repdata_data_StormData.csv.bz2')
```

Our analysis will focus on the health and economic effects of different types of weather events. Therefore, we keep only the columns that reference that information: *EVTYPE* (type of event), *Fatalities* and *Injuries* (effects over population health), *PROPDMG* and *PROPDMGEXP* (property damage estimates, including the order of magnitude) and *CROPDMG* and *CROPDMGEXP* (crop damage estimates, including magnitude).

```
noaa <- noaa[,c('EVTYPE','FATALITIES','INJURIES','PROPDMG','PROPDMGEXP','CROPDMG','CROPDMGEXP')]
#Adding an index row
index <- c(1:nrow(noaa))
noaa$INDEX <- index
```

We can rectify the damage columns including the thousands, millions or billions orders of magnitude and delete those columns. The characters 'K', 'M' and 'B' represent thousands, millions or billions of dollars worth of property or crops. The entries with incorrect characters had the damage set to NA.

```
noaakp <- noaa[noaa$PROPDMGEXP=='K',]
noaakp <- rbind(noaakp,noaa[noaa$PROPDMGEXP=='k',])
noaakp$PROPDMG <- as.numeric(noaakp$PROPDMG) * 1000
noaamp <- noaa[noaa$PROPDMGEXP=='M',]
noaamp <- rbind(noaamp,noaa[noaa$PROPDMGEXP=='m',])
noaamp$PROPDMG <- as.numeric(noaamp$PROPDMG) * 1000000
noaabp <- noaa[noaa$PROPDMGEXP=='B',]
noaabp <- rbind(noaabp,noaa[noaa$PROPDMGEXP=='b',])
```

```
noaabp$PROPDMG <- as.numeric(noaabp$PROPDMG) * 1000000000
noaaop <- noaa[!(noaa$PROPDMGEXP %in% c('K','k','M','m','B','b')),]
noaaop$PROPDMG <- NA
noaa1 <- rbind(noaakp,noaamp,noaabp,noaaop)

noaakp <- noaa[noaa1$CROPDMGEXP=='K',]
noaakp <- rbind(noaakp,noaa[noaa1$CROPDMGEXP=='k',])
noaakp$CROPDMG <- as.numeric(noaakp$CROPDMG) * 1000
noaamp <- noaa[noaa$CROPDMGEXP=='M',]
noaamp <- rbind(noaamp,noaa[noaa$CROPDMGEXP=='m',])
noaamp$CROPDMG <- as.numeric(noaamp$CROPDMG) * 1000000
noaabp <- noaa1[noaa$CROPDMGEXP=='B',]
noaabp <- rbind(noaabp,noaa[noaa$CROPDMGEXP=='b',])
noaabp$CROPDMG <- as.numeric(noaabp$CROPDMG) * 1000000000
noaaop <- noaa[!(noaa$CROPDMGEXP %in% c('K','k','M','m','B','b')),]
noaaop$PROPDMG <- NA
noaa2 <- rbind(noaakp,noaamp,noaabp,noaaop)

noaa <- merge(noaa1,noaa2,by='INDEX')
noaa <- cbind(noaa$EVTYPE.x,noaa$FATALITIES.x,noaa$INJURIES.x,noaa$PROPDMG.x,noaa$CROPDMG.y)
colnames(noaa) <- c('EVTYPE','FATALITIES','INJURIES','PROPDMG','CROPDMG')
noaa <- data.frame(noaa)
noaa$EVTYPE <- toupper(noaa$EVTYPE)
noaa$FATALITIES <- as.numeric(noaa$FATALITIES)
noaa$INJURIES <- as.numeric(noaa$INJURIES)
noaa$PROPDMG <- as.numeric(noaa$PROPDMG)
noaa$CROPDMG <- as.numeric(noaa$CROPDMG)
```

Finally, we want to filter occurrences where we have at least 10 events under the same category. This will get rid of entries that are not very representative.

```
entries <- table(noaa$EVTYPE)
entries <- entries[entries>=10]
noaa <- noaa[noaa$EVTYPE %in% names(entries),]
```

Results

Identifying the most harmful events

The mean number of fatalities may be calculated from the table.

```
rank <- aggregate(noaa$FATALITIES,by=list(noaa$EVTYPE),mean,na.rm=TRUE)
colnames(rank) <- c('EVTYPE','Avg.Fatality')
rank <- rank[order(rank$Avg.Fatality,decreasing=TRUE),]
head(rank,10)
```

```
##                         EVTYPE Avg.Fatality
## 67                        HEAT    7.1782178
## 30                EXTREME HEAT    4.3636364
## 68                   HEAT WAVE    2.3129252
## 193 UNSEASONABLY WARM AND DRY    2.2307692
## 25              EXCESSIVE HEAT    1.7322503
```

```
## 91       HYPOTHERMIA/EXPOSURE    1.1666667
## 103           LOW TEMPERATURE    1.0000000
## 122                 RAIN/SNOW    0.8000000
## 133              RIP CURRENT    0.7982833
## 88             HURRICANE ERIN    0.7692308
```

We can mean repeat this treatment with respect to injuries.

```
rank <- aggregate(noaa$INJURIES,by=list(noaa$EVTYPE),mean,na.rm=TRUE)
colnames(rank) <- c('EVTYPE','Avg.Injury')
rank <- rank[order(rank$Avg.Injury,decreasing=TRUE),]
head(rank,10)
```

```
##                 EVTYPE Avg.Injury
## 90   HURRICANE/TYPHOON  15.407895
## 220 WINTER WEATHER MIX  13.300000
## 67                HEAT   9.123762
## 30        EXTREME HEAT   7.045455
## 25      EXCESSIVE HEAT   5.462094
## 207 WATERSPOUT/TORNADO   5.250000
## 68           HEAT WAVE   5.149660
## 56               GLAZE   5.023256
## 110       MIXED PRECIP   2.600000
## 92                 ICE   2.264463
```

With regard to economic impacts, we can analyze the effect over property and crops.

```
rank <- aggregate(noaa$PROPDMG,by=list(noaa$EVTYPE),mean,na.rm=TRUE)
colnames(rank) <- c('EVTYPE','Avg.Property.Damage')
rank <- rank[order(rank$Avg.Property.Damage,decreasing=TRUE),]
head(rank,10)
```

```
##                 EVTYPE Avg.Property.Damage
## 90    HURRICANE/TYPHOON          1065123793
## 89       HURRICANE OPAL           396605750
## 139 SEVERE THUNDERSTORM           200893333
## 161         STORM SURGE           164679117
## 87            HURRICANE           105556770
## 187             TYPHOON            91133077
## 88       HURRICANE ERIN            57244444
## 212           WILDFIRES            40100000
## 40                FLOOD            36223991
## 180      TROPICAL STORM            34276012
```

```
rank <- aggregate(noaa$CROPDMG,by=list(noaa$EVTYPE),mean,na.rm=TRUE)
colnames(rank) <- c('EVTYPE','Avg.Crop.Damage')
rank <- rank[order(rank$Avg.Crop.Damage,decreasing=TRUE),]
head(rank,10)
```

```
##                       EVTYPE Avg.Crop.Damage
## 14            DAMAGING FREEZE        22809700
## 90          HURRICANE/TYPHOON        14396974
```

```
## 88              HURRICANE ERIN        10472769
## 87                   HURRICANE        10048503
## 42           FLOOD/RAIN/WINDS         9409400
## 16                     DROUGHT         7195198
## 140 SEVERE THUNDERSTORM WINDS         2902900
## 52               FROST/FREEZE         2769891
## 1         AGRICULTURAL FREEZE         2404068
## 45                     FREEZE         2039534
```

We can conclude that the three most dangerous events for public health and the economy are: * Hurricanes and typhoons, ranking consistently high on all types of damage; * Extreme heat, droughts or wildfires, ranking high among casualties and injuries, and also causing losses to agriculture; * Extreme cold, including freezes and snow, which are also top ranking among casualties and injuries, and crop damages.

Summarizing the data for these events

We can gather the events that appeared on the tables below under the same category to simplify the analysis.

```
noaa[(noaa$EVTYPE=='HURRICANE ERIN'|noaa$EVTYPE=='HURRICANE OPAL'|noaa$EVTYPE=='HURRICANE'|noaa$EVTYPE==
noaa[(noaa$EVTYPE=='EXTREME HEAT'|noaa$EVTYPE=='HEAT WAVE'|noaa$EVTYPE=='EXCESSIVE HEAT'|noaa$EVTYPE=='
noaa[(noaa$EVTYPE=='HYPOTHERMIA/EXPOSURE'|noaa$EVTYPE=='LOW TEMPERATURE'|noaa$EVTYPE=='RAIN/SNOW'|noaa$
noaa[(noaa$EVTYPE=='DAMAGING FREEZE'|noaa$EVTYPE=='FROST/FREEZE'|noaa$EVTYPE=='AGRICULTURAL FREEZE'|noa
```

The 5-number summary (including the mean) for Hurricane/Typhoons is included below:

```
summary(noaa[noaa$EVTYPE=='HURRICANE/TYPHOON',2:5])
```

```
##    FATALITIES          INJURIES           PROPDMG              CROPDMG
##  Min.   : 0.0000   Min.   :  0.000   Min.   :5.000e+03   Min.   :        0
##  1st Qu.: 0.0000   1st Qu.:  0.000   1st Qu.:4.288e+05   1st Qu.:        0
##  Median : 0.0000   Median :  0.000   Median :6.950e+06   Median :        0
##  Mean   : 0.4433   Mean   :  3.144   Mean   :3.234e+08   Mean   : 10058945
##  3rd Qu.: 0.0000   3rd Qu.:  0.000   3rd Qu.:7.200e+07   3rd Qu.:     5000
##  Max.   :13.0000   Max.   :780.000   Max.   :1.693e+10   Max.   :500000000
##                                      NA's   :123
```

This means most events do not cause casualties. However, there are cases where hundreds of injuries may occur. Additionally, these events almost always cause extensive economic damage, on the order of hundreds of millions of dollars. This is mostly hard to prevent due to the nature of Hurricanes.

Regarding excessive heat events:

```
summary(noaa[noaa$EVTYPE=='HEAT',2:5])
```

```
##    FATALITIES          INJURIES           PROPDMG            CROPDMG
##  Min.   :  0.000   Min.   :  0.000   Min.   :      0   Min.   :        0
##  1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.:      0   1st Qu.:        0
##  Median :  0.000   Median :  0.000   Median :      0   Median :        0
##  Mean   :  2.364   Mean   :  5.761   Mean   : 135367   Mean   :   242310
##  3rd Qu.:  1.000   3rd Qu.:  0.000   3rd Qu.:      0   3rd Qu.:        0
##  Max.   :583.000   Max.   :437.000   Max.   :3800000   Max.   :492400000
##                                      NA's   :1849
```

```
summary(noaa[noaa$EVTYPE=='WILDFIRES',2:5])
```

```
##    FATALITIES    INJURIES    PROPDMG            CROPDMG
##  Min.   :0    Min.   :0   Min.   :  500000   Min.   :     0
##  1st Qu.:0    1st Qu.:0   1st Qu.:50000000   1st Qu.:     0
##  Median :0    Median :0   Median :50000000   Median :     0
##  Mean   :0    Mean   :0   Mean   :40100000   Mean   : 33333
##  3rd Qu.:0    3rd Qu.:0   3rd Qu.:50000000   3rd Qu.:     0
##  Max.   :0    Max.   :0   Max.   :50000000   Max.   :500000
##                           NA's   :10
```

```
summary(noaa[noaa$EVTYPE=='DROUGHT',2:5])
```

```
##    FATALITIES    INJURIES           PROPDMG             CROPDMG
##  Min.   :0    Min.   :0.000000   Min.   :        0   Min.   :        0
##  1st Qu.:0    1st Qu.:0.000000   1st Qu.:        0   1st Qu.:        0
##  Median :0    Median :0.000000   Median :    39500   Median :        0
##  Mean   :0    Mean   :0.002309   Mean   : 10965045   Mean   :  7195198
##  3rd Qu.:0    3rd Qu.:0.000000   3rd Qu.:  8000000   3rd Qu.:        0
##  Max.   :0    Max.   :4.000000   Max.   :645150000   Max.   :578850000
##                                  NA's   :1620
```

These are apparently events that occur more frequently, and mostly causes no noticeable damage, with the exception of wildfires. Extreme versions of heat waves may be devastating, though.Hundreds of people may die and hundred million dollars may be lost from a single event.

Regarding excessive cold events:

```
summary(noaa[noaa$EVTYPE=='COLD',2:5])
```

```
##    FATALITIES        INJURIES         PROPDMG           CROPDMG
##  Min.   :0.0000   Min.   : 0.000   Min.   :   5000   Min.   :0
##  1st Qu.:0.0000   1st Qu.: 0.000   1st Qu.:  50000   1st Qu.:0
##  Median :0.0000   Median : 0.000   Median : 500000   Median :0
##  Mean   :0.3077   Mean   : 1.411   Mean   : 528360   Mean   :0
##  3rd Qu.:1.0000   3rd Qu.: 0.000   3rd Qu.: 500000   3rd Qu.:0
##  Max.   :2.0000   Max.   :62.000   Max.   :5000000   Max.   :0
##                                    NA's   :249
```

```
summary(noaa[noaa$EVTYPE=='FROST/FREEZE',2:5])
```

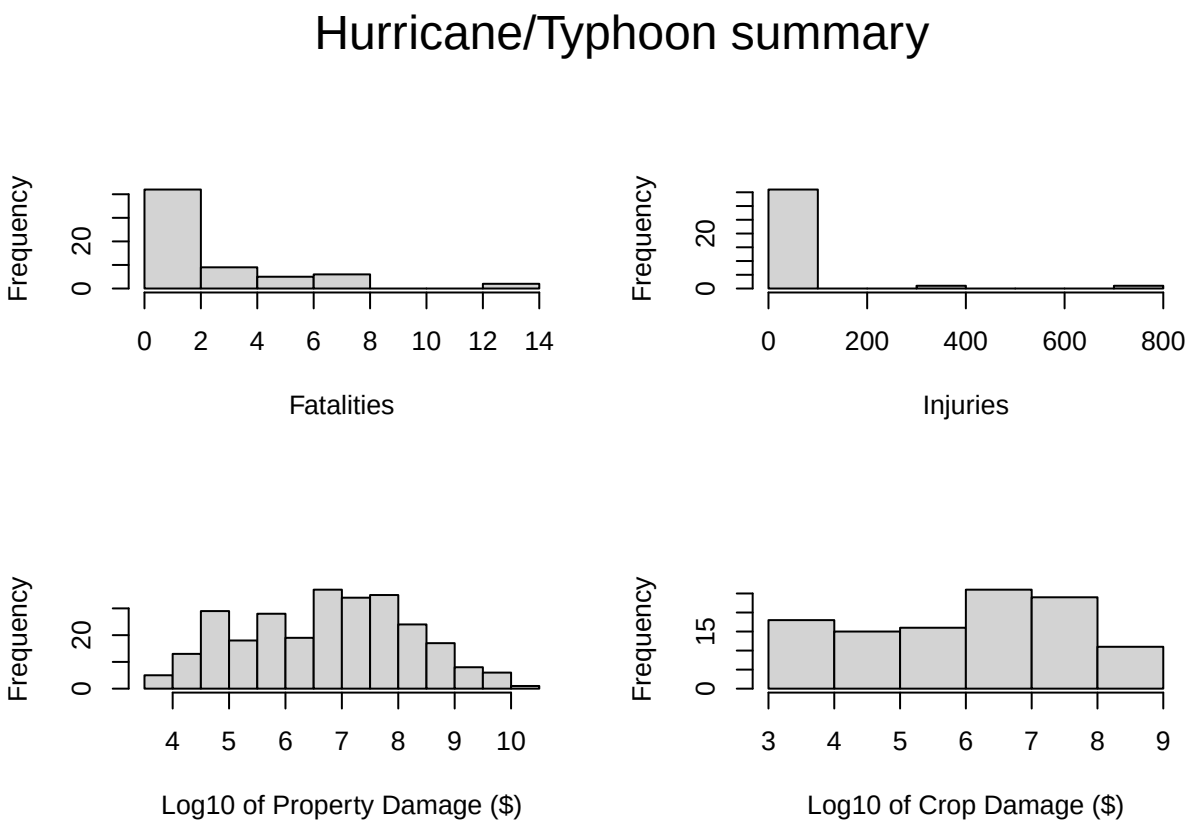
```
##    FATALITIES          INJURIES    PROPDMG           CROPDMG
##  Min.   :0.000000   Min.   :0   Min.   :      0   Min.   :        0
##  1st Qu.:0.000000   1st Qu.:0   1st Qu.:      0   1st Qu.:        0
##  Median :0.000000   Median :0   Median :      0   Median :        0
##  Mean   :0.003677   Mean   :0   Mean   : 257215   Mean   :  3071550
##  3rd Qu.:0.000000   3rd Qu.:0   3rd Qu.:      0   3rd Qu.:        0
##  Max.   :1.000000   Max.   :0   Max.   :8000000   Max.   :286000000
##                                 NA's   :465
```

In comparison to the previous categories, these events seem less deadly, but they still pose great economic threats.

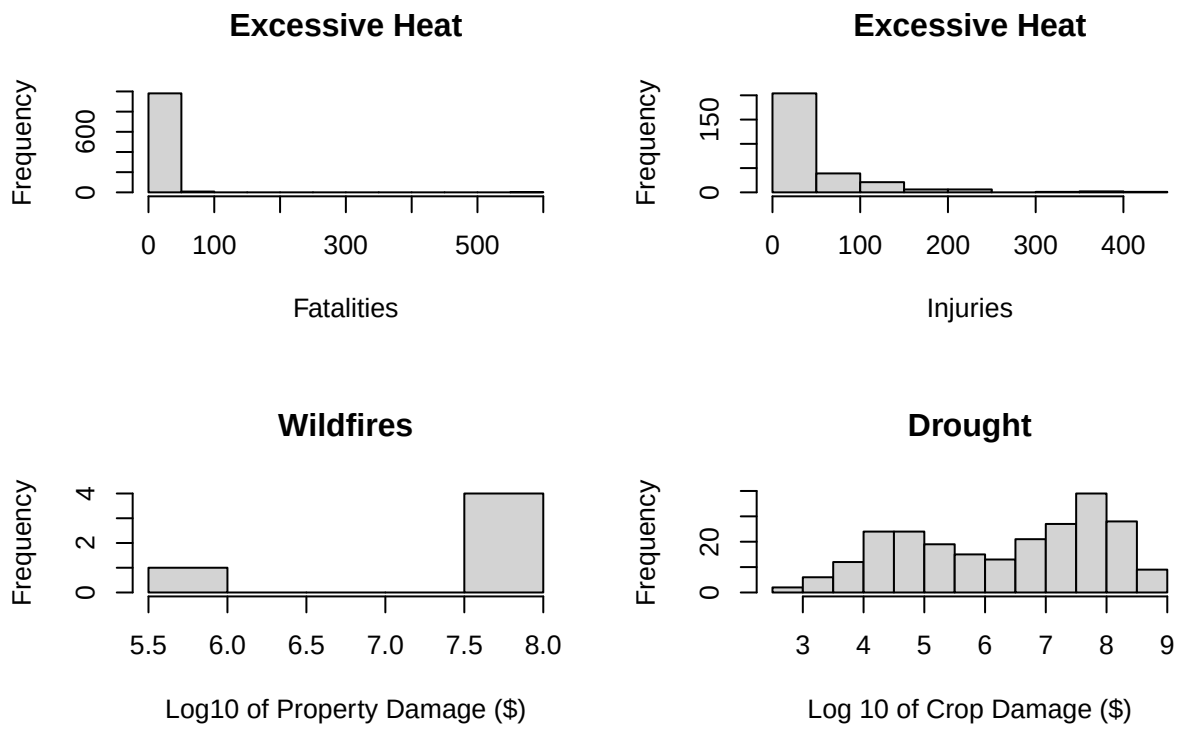
Visualizing the data

Excluding the events where there was no damage, we can plot histograms to visualize the distribution of data.

```
par(mfrow=c(2,2),oma=c(0,0,2,0))
hist(noaa[(noaa$EVTYPE=='HURRICANE/TYPHOON')& (noaa$FATALITIES!=0),'FATALITIES'],main='',xlab='Fatalitie
hist(noaa[(noaa$EVTYPE=='HURRICANE/TYPHOON')& (noaa$INJURIES!=0),'INJURIES'],main='',xlab='Injuries')
hist(log10(noaa[(noaa$EVTYPE=='HURRICANE/TYPHOON')& (noaa$PROPDMG!=0),'PROPDMG']),main='',xlab='Log10 o
hist(log10(noaa[(noaa$EVTYPE=='HURRICANE/TYPHOON')& (noaa$CROPDMG!=0),'CROPDMG']),main='',xlab='Log10 o
mtext('Hurricane/Typhoon summary',outer=TRUE,cex=1.5)
```

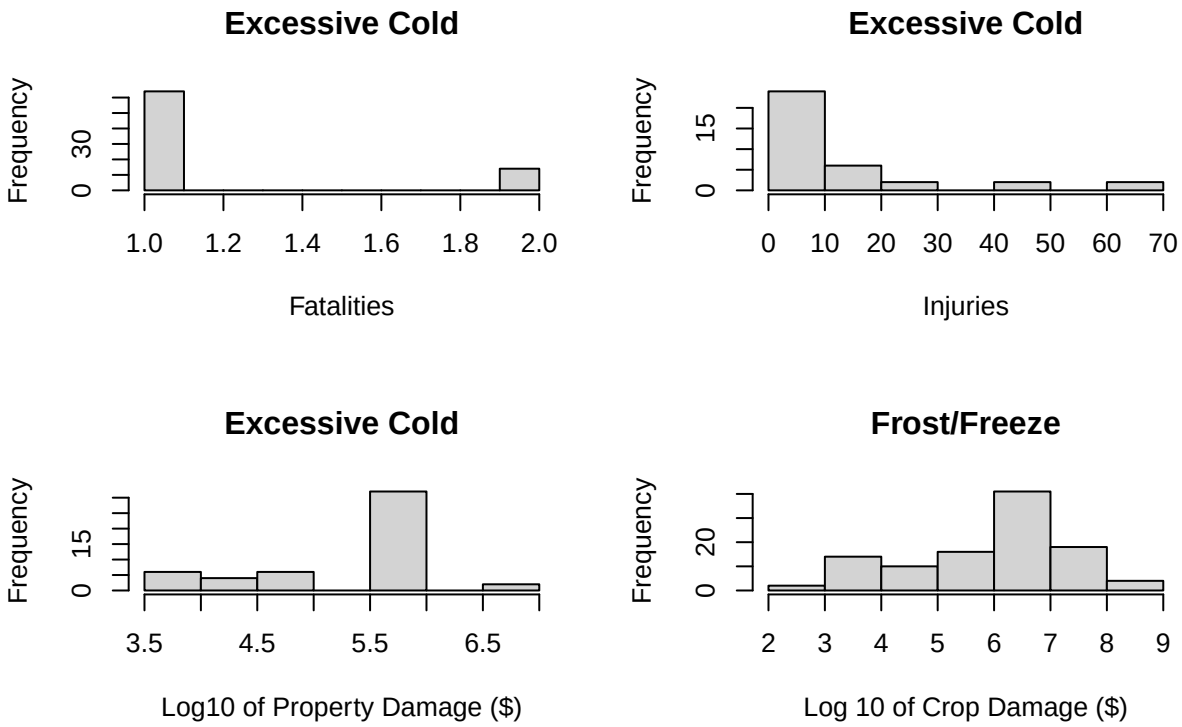



```
par(mfrow=c(2,2),oma=c(0,0,2,0))
hist(noaa[(noaa$EVTYPE=='HEAT')& (noaa$FATALITIES!=0),'FATALITIES'],main='Excessive Heat',xlab='Fatalit
hist(noaa[(noaa$EVTYPE=='HEAT')& (noaa$INJURIES!=0),'INJURIES'],main='Excessive Heat',xlab='Injuries')
hist(log10(noaa[(noaa$EVTYPE=='WILDFIRES')& (noaa$PROPDMG!=0),'PROPDMG']),main='Wildfires',xlab='Log10 
hist(log10(noaa[(noaa$EVTYPE=='DROUGHT')& (noaa$CROPDMG!=0),'CROPDMG']),main='Drought',xlab='Log 10 of 
mtext('Excessive heat events summary',outer=TRUE,cex=1.5)
```




```
par(mfrow=c(2,2),oma=c(0,0,2,0))
hist(noaa[(noaa$EVTYPE=='COLD')& (noaa$FATALITIES!=0),'FATALITIES'],main='Excessive Cold',xlab='Fatalit
hist(noaa[(noaa$EVTYPE=='COLD')& (noaa$INJURIES!=0),'INJURIES'],main='Excessive Cold',xlab='Injuries')
hist(log10(noaa[(noaa$EVTYPE=='COLD')& (noaa$PROPDMG!=0),'PROPDMG']),main='Excessive Cold',xlab='Log10 
hist(log10(noaa[(noaa$EVTYPE=='FROST/FREEZE')& (noaa$CROPDMG!=0),'CROPDMG']),main='Frost/Freeze',xlab='
mtext('Excessive cold events summary',outer=TRUE,cex=1.5)
```

Excessive cold events summary



Discussion

The data has informed us that hurricanes/typhoons cause extensive property damage, usually from hundred thousand dollars to a billion dollars, and not only must they be tracked and forecast, but also there must be governmental incentives to research focusing on damage reduction. Hurricanes are also even more malicious because they occur in areas with agricultural crops, causing even more economic danger and threatening food supply.

Extreme weather such as intensive heat or cold are also a threat. The data suggests most of these events cause no damage, but the more acute versions have devastating potential, as seen in 2021 in Canada with hundreds of deaths due to a heat wave, or the death of millions of sea animals in the Pacific coast of the USA. These events are becoming increasingly common due to global warming and is imperative that the countries unite to fight it.