Exploring the NOAA Storm Database

Synopsis

Data from the U.S. National Oceanic and Atmospheric Administration's Storm database was analyzed to determine the public health effects and economic impact of various disaster events. The data analyzed spans January 1996 to November 2011. For public health effects, fatalities and injuries are considered. For economic effects, property and crop damage are analyzed. A ranking of 13 event categories by their health and economic impacts are provided.

Data Processing

Gathering the Data

The data analyzed is from the U.S. National Oceanic and Atmospheric Administration's storm database. This database includes information on major storms and weather events from 1950 to November 2011. The following are steps to download and extract the NOAA data.

1. Load the necessary libraries and adjust settings.

```
library(R.utils)
library(plyr)
library(sqldf)

## To prevent R from showing scientific notation in output
options("scipen" = 20)
```

2. Download storm data from the NOAA website. Extract the contents of the archive to a CSV file.

```
## Download the NOAA data
file_url <- 'http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2'
dest_file <- paste(getwd(),'repdata_data_StormData.csv.bz2',sep='/')

if (!file.exists(dest_file)) {
    download.file(url = file_url, destfile = dest_file)

    ## Extract the contents
    bunzip2(filename = dest_file)
}
```

3. Load the CSV data into a data frame.

```
noaa_data <- read.csv(file = 'repdata_data_StormData.csv')
```

Filtering the data

For this analysis, we will only be considering the health and economic impact of recorded events. The following fields will be included in the analysis:

```
* Events
    - EVTYPE
* Health Effects
    - FATALITIES
    - INJURIES
* Economic Effects
    - PROPDMG
    - PROPDMGEXP
    - CROPDMG
    - CROPDMGEXP
```

Furthermore, the NOAA Storm Events Database had only included a comprehensive record of *all* events since January 1996. Before then, only a limited set of events were recorded. From January 1955 to January 1996 only tornado, thunderstorm wind, and hail events were recorded, and from January 1950 to January 1955, only tornado events were recorded. We must restrict our analysis to events recorded after January 1996.

```
get_amount <- function(x, exp) {
    ## Return amount `x` multiplied by a constant factor `exp`
    ret <- x
    if (toupper(exp) == 'M') {
        ret <- ret * 1000000
```

```
    }
    else if (toupper(exp) == 'K') {
        ret <- ret * 1000
    }
    ret
}

## only extract data since January 1996
all_events_data <- noaa_data
all_events_data$BGN_DATE <- as.Date(all_events_data$BGN_DATE,
                                    format = '%m/%d/%Y %H:%M:%S')
all_events_data <- subset(all_events_data,
                          all_events_data$BGN_DATE >= as.Date("1996-01-01"))

## Only extract relevant columns
filtered_data <- data.frame(EVTYPE=all_events_data$EVTYPE,
                            FATALITIES=all_events_data$FATALITIES,
                            INJURIES=all_events_data$INJURIES,
                            PROPDMG=mapply(get_amount,
                                           all_events_data$PROPDMG,
                                           all_events_data$PROPDMGEXP),
                            CROPDMG=mapply(get_amount,
                                           all_events_data$CROPDMG,
                                           all_events_data$CROPDMGEXP))

## No longer need `all_events_data`
rm(all_events_data)
```

We would like to gather and rank the events based on certain criteria. However, the events (EVTYPE) are not logged consistently. For example, consider the events "SNOW", "SNOW/ICE", and "ICE STORM/FLASH FLOOD". These are not 3 distinct events, but rather three groupings of events containing distinct, overlapping events. Therefore, we must extract the appropriate distinct event(s) from each EVTYPE entry.

For our analysis, we shall restrict the possible events to:

```
* Tornadoes
* Oceanic Events (includes hurricanes, tropical storms, and typhoons)
* Floods
* Rain (includes events labelled as "precipitation")
* High Winds
* Blizzards
* Snow/Ice Events
* Extreme Cold (includes only temperature measurements, not considering snow/ice.  Does include "freeze" events)
* Extreme Heat
* Extreme Dry Conditions
* Fires
* Fog Events
* Dust Events
* Volcanic Events
```

```
categorized_data <- list()
patterns <- c('TORNADO',
              'HURRICANE|TROPICAL|TYPHOON',
              'FLOOD',
              'RAIN|PRECIP',
              'WIND',
              'BLIZZARD',
              'SNOW|ICE',
              'COLD|FREEZE|LOW TEMP',
              'HEAT|HIGH TEMP',
              'DRY',
              'FIRE',
              'FOG',
              'VOLCAN'
             )
```

```
# Shorten the category names for clean plots
event_names <- patterns
event_names[[2]] <- 'OCEANIC'
event_names[[4]] <- 'RAIN'
event_names[[7]] <- 'SNOW/ICE'
event_names[[8]] <- 'COLD'
event_names[[9]] <- 'HEAT'
event_names[[13]] <- 'VOLCANIC'

## Re-categorize the data
for (i in seq_along(patterns)) {
    event_data <- subset(filtered_data, grepl(pattern = patterns[[i]],
                                               x = filtered_data$EVTYPE,
                                               ignore.case = TRUE))
    event_data$EVTYPE = event_names[[i]]
    categorized_data[[i]] <- event_data
}
categorized_data <- ldply(categorized_data)

## Filtered set is no longer needed
rm(filtered_data)
```

Analysis of Data

Once the data has been properly categorized, we can rank the events based on their overall effects.

Health effects

For health effects, we will gather:

```
* Total fatalities
* Total injuries
```

```
health_data <- sqldf('SELECT EVTYPE,
                             SUM(FATALITIES) AS TOTAL_FATAL,
                             AVG(FATALITIES) AS AVG_FATAL,
                             SUM(INJURIES) AS TOTAL_INJURE,
                             AVG(INJURIES) AS AVG_INJURE
                      FROM categorized_data
                      GROUP BY EVTYPE')
```

Economic effects

For economic effects, we will gather:

```
* Total property damage
* Average property damage per event
* Total crop damage
* Average crop damage per event
* Total absolute monetary loss (property + crop damage)
* Average absolute monetary loss (property + crop damage)
```

```
econ_data <- sqldf('SELECT EVTYPE,
                           SUM(PROPDMG) AS TOTAL_PROPDMG,
                           AVG(PROPDMG) AS AVG_PROPDMG,
                           SUM(CROPDMG) AS TOTAL_CROPDMG,
                           AVG(CROPDMG) AS AVG_CROPDMG
                    FROM categorized_data
                    GROUP BY EVTYPE')
econ_data$TOTAL_ABSDMG <- econ_data$TOTAL_PROPDMG + econ_data$TOTAL_CROPDMG
econ_data$AVG_ABSDMG <- econ_data$AVG_PROPDMG + econ_data$AVG_CROPDMG
```

Results

Below are figures for the health and economic impacts of disaster events. Raw figures are included in the appendix.

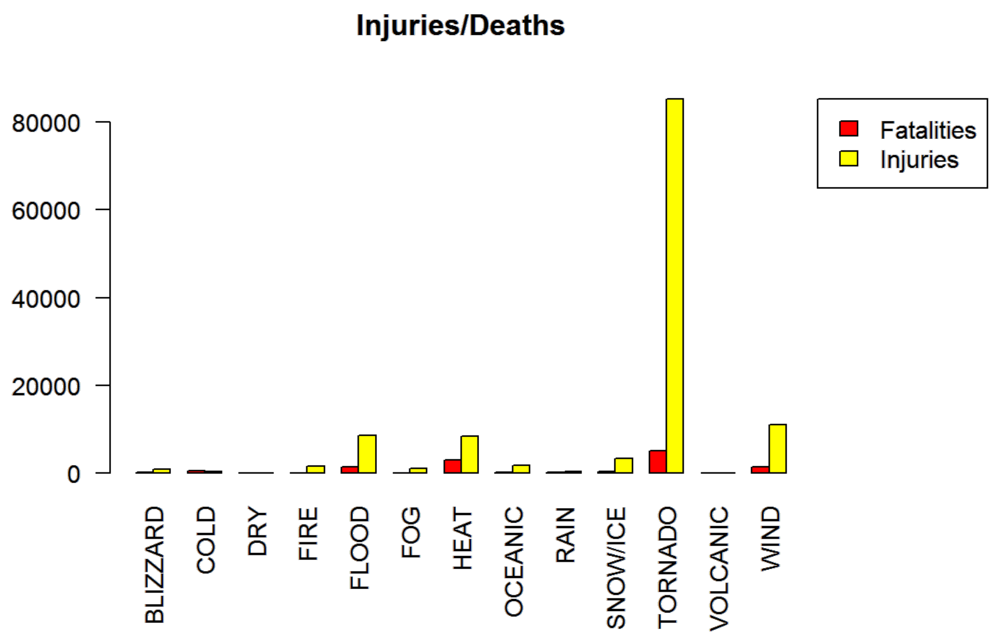
Health Impact of Events

```
health_matrix <- data.matrix(health_data[,c(2,4)])
rownames(health_matrix) <- health_data$EVTYPE

health_matrix_t <- t(health_matrix)

## For numbers
tornado_stats <- health_matrix_t[,11]
heat_stats <- health_matrix_t[,5]
top_stats <- health_matrix_t[,c(5,7,13)]
rest_stats <- health_matrix_t[,c(-5,-7,-11,-13)]

par(mar = c(7.1, 4.1, 5.1, 7.1), xpd = TRUE)
barplot(health_matrix_t,
        col = heat.colors(length(rownames(health_matrix_t))),
        main = 'Injuries/Deaths',
        width = 2,
        beside = TRUE,
        las = 2)
legend("topright",
       inset = c(-0.25, 0),
       fill = heat.colors(length(rownames(health_matrix_t))),
       legend = c('Fatalities','Injuries'))
```



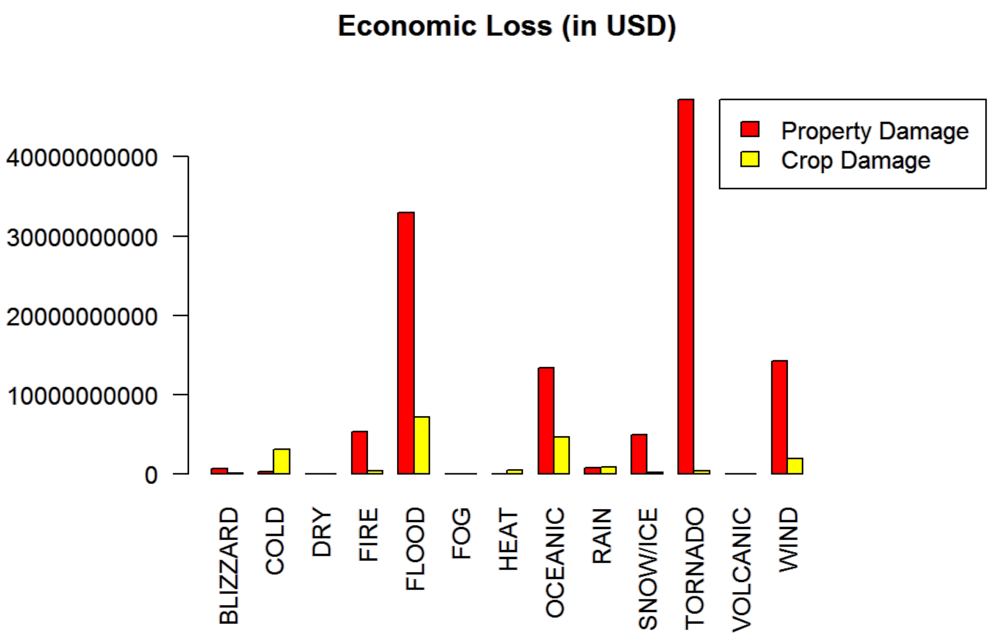
The figure above shows the total fatalities and injuries caused by disaster events since 1996. Heat events dominate total fatalities, with 1398 fatalities, while tornadoes dominate injuries, with 85247 injuries. Floods, high winds, tornadoes and extreme heat events dominate both counts overall, with 5801 fatalities and 28100 injuries (combined). All other events combined totalled 1320 fatalities and 9088 injuries.

Economic Effects of Events

```
econ_matrix <- data.matrix(econ_data[,c(2,4)])
rownames(econ_matrix) <- econ_data$EVTYPE
econ_matrix_t <- t(econ_matrix)

par(mar = c(7.1, 6.25, 5.1, 7.1), xpd = TRUE)
barplot(econ_matrix_t,
        col = heat.colors(length(rownames(econ_matrix_t))),
        main = 'Economic Loss (in USD)',
        width = 2,
        beside = TRUE,
        las = 2)
legend("topright",
```

```
        inset = c(-0.25, 0),
        fill = heat.colors(length(rownames(econ_matrix_t))),
        legend = c('Property Damage','Crop Damage'))
```

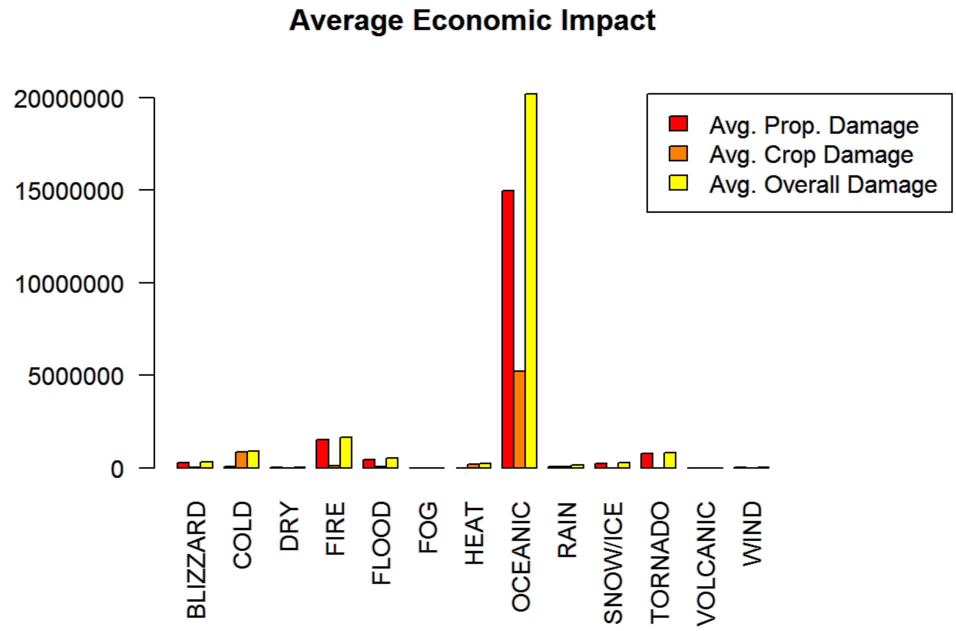


This figure shows the total economic loss for all events since 1996. Tornadoes again occupy a top spot, here for property damage. Meanwhile, floods are the leading cause of crop damage. Oceanic events, fires, and high winds are also significant contributers to property damage, while extreme cold events contribute a large portion of crop failures. All other events appear to be negligible.

```
avgs <- econ_data[,c(1,3,5,7)]
avg_matrix <- data.matrix(avgs[,2:4])
rownames(avg_matrix) <- health_data$EVTYPE

avg_matrix_t <- t(avg_matrix)

par(mar = c(7.1, 6.25, 5.1, 7.1), xpd = TRUE)
barplot(avg_matrix_t,
        col = heat.colors(length(rownames(avg_matrix_t))),
        main = 'Average Economic Impact',
        width = 2,
        beside = TRUE,
        las = 2)
legend("topright",
       inset = c(-0.25, 0),
       fill = heat.colors(length(rownames(avg_matrix_t))),
       legend = c('Avg. Prop. Damage',
                  'Avg. Crop Damage',
                  'Avg. Overall Damage'))
```



When looking at the averages for economic losses, a different picture emerges. While tornadoes occupied a top spot for total property damage, per event they cost very little. Instead, we see that Oceanic events (hurricanes, typhoons, and tropical storms) are the most costly events.

Appendix

Health data

```
health_data
```

```
##      EVTYPE TOTAL_FATAL   AVG_FATAL TOTAL_INJURE AVG_INJURE
## 1  BLIZZARD         100 0.040783034          806 0.32871126
## 2      COLD         445 0.123371223          319 0.08843915
## 3       DRY          32 0.104234528           29 0.09446254
## 4      FIRE          84 0.024172662         1492 0.42935252
## 5     FLOOD        1398 0.018506016         8563 0.11335266
## 6       FOG          81 0.046498278         1077 0.61825488
## 7      HEAT        3039 1.364616075         8475 3.80556803
## 8   OCEANIC         197 0.220357942         1715 1.91834452
## 9      RAIN         115 0.010181496          330 0.02921647
## 10 SNOW/ICE         266 0.014118147         3320 0.17621145
## 11  TORNADO        5074 0.086696511        85247 1.45656631
## 12 VOLCANIC           0 0.000000000            0 0.00000000
## 13     WIND        1364 0.004022531        11062 0.03262261
```

Economic data

```
econ_data
```

```
##      EVTYPE TOTAL_PROPDMG  AVG_PROPDMG TOTAL_CROPDMG   AVG_CROPDMG
## 1  BLIZZARD     662171950   270053.813     112060000   45701.46819
## 2      COLD     253159450    70185.597    3092416550  857337.55204
## 3       DRY       6732600    21930.293         15000      48.85993
## 4      FIRE    5314376103  1529316.864     393484630  113232.98705
## 5     FLOOD   32933515907   435957.215    7136790105   94473.21532
## 6       FOG      24849500    14264.925             0       0.00000
## 7      HEAT      19182550     8613.628     504469280  226524.14926
## 8   OCEANIC   13373295440 14958943.445    4666012802 5219253.69296
## 9      RAIN     759763195    67265.444     898602800   79557.57415
## 10 SNOW/ICE    4977010412   264158.506     161697405    8582.66481
## 11  TORNADO   47187127330   806259.224     388100520    6631.24970
## 12 VOLCANIC        500000    17241.379             0       0.00000
```

```
## 13     WIND  14297138115    42163.255    1960070538    5780.38438
##    TOTAL_ABSDMG  AVG_ABSDMG
## 1     774231950   315755.28
## 2    3345576000   927523.15
## 3       6747600    21979.15
## 4    5707860733  1642549.85
## 5   40070306012   530430.43
## 6      24849500    14264.93
## 7     523651830   235137.78
## 8   18039308241 20178197.14
## 9    1658365995   146823.02
## 10   5138707817   272741.17
## 11  47575227850   812890.47
## 12       500000    17241.38
## 13  16257208653    47943.64
```