A comparison of impact severity for different types of storms.

Synopsis

The aim of this report is to determine which types of storms have the greatest impact on human health and have the greatest economic consequences. Data was taken from the National Oceanic and Atmospheric Administration’s (NOAA) database. This report shows that, on average, wild fires cause the most damage to human health per event. Tornadoesresult in the greatest damage to human health overall. For economic impact,hurricanes have the highes cost per incident, while floods have the highest cost overall.

Data Processing

Data was taken from the NOAA at https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 . It was loaded in using:

```
file <- read.csv("~/Coursera/repdata_data_StormData.csv.bz2")
```

and the following libraries were loaded:

```
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(lattice)
```

In order to determine which events had the greatest impact to human health, I added up all fatalities and injuries for each type of event. I then extracted the top 10 events.

```
danger_file <- filter(file, FATALITIES > 0 | INJURIES > 0)

danger <- danger_file %>%

        group_by(EVTYPE) %>%

                summarise(TOTAL_FATALITIES = sum(FATALITIES),

                          TOTAL_INJURIES = sum(INJURIES),

                          AVG_FATALITIES = mean(FATALITIES),

                          AVG_INJURIES = mean(INJURIES),

                          NUMBER_OF_EVENTS = n()) %>%

        mutate(AVG_HARM = AVG_FATALITIES + AVG_INJURIES,

                          TOTAL_HARM = TOTAL_FATALITIES

                                                + TOTAL_INJURIES) %>%

        select(EVTYPE, NUMBER_OF_EVENTS, TOTAL_HARM,

                          AVG_HARM) %>%

        arrange(desc(AVG_HARM))

total_danger <- arrange(danger, desc(TOTAL_HARM))

top_10_avg_danger <- danger[1:10, ]

top_10_total_danger <- total_danger[1:10, ]
```

In order to determine which events had the greatest econonomic consequences, I replaced all of the exponents of the PROPDMGEXP AND CROPDMGEXP columns with their numeric value (B/b = 1000000000,M/m =1000000, K/k = 1000, H/h = 100, all else = 1). This number was then multiplied by the value given in the PROPDMG AND CROPDMG columns. Finally, I added up all the property damage and crop damage for each type of event. I then extracted the top 10 events.

```
subset_file <- filter(file, PROPDMG > 0 | CROPDMG > 0)

subset_file <- select(subset_file, EVTYPE, PROPDMG,PROPDMGEXP,

                CROPDMG, CROPDMGEXP)

subset_file$PROPDMGEXP <- gsub("B", 1000000000, subset_file$PROPDMGEXP,

                               ignore.case = TRUE)
```

```
subset_file$PROPDMGEXP <- gsub("M", 1000000, subset_file$PROPDMGEXP,
                               ignore.case = TRUE)
subset_file$PROPDMGEXP <- gsub("K", 1000, subset_file$PROPDMGEXP,
                               ignore.case = TRUE)
subset_file$PROPDMGEXP <- gsub("H", 100, subset_file$PROPDMGEXP,
                               ignore.case = TRUE)
subset_file$CROPDMGEXP <- gsub("B", 1000000000, subset_file$CROPDMGEXP,
                               ignore.case = TRUE)
subset_file$CROPDMGEXP <- gsub("M", 1000000, subset_file$CROPDMGEXP,
                               ignore.case = TRUE)
subset_file$CROPDMGEXP <- gsub("K", 1000, subset_file$CROPDMGEXP,
                               ignore.case = TRUE)
subset_file$CROPDMGEXP <- gsub("H", 100, subset_file$CROPDMGEXP,
                               ignore.case = TRUE)
subset_file$PROPDMGEXP <- as.numeric(subset_file$PROPDMGEXP)
```

```
## Warning: NAs introduced by coercion
```

```
subset_file$CROPDMGEXP <- as.numeric(subset_file$CROPDMGEXP)
```

```
## Warning: NAs introduced by coercion
```

```
subset_file[is.na(subset_file)] <- 1
subset_file <- mutate(subset_file, EV_PROPDMG = PROPDMG*PROPDMGEXP,
                      EV_CROPDMG = CROPDMG*CROPDMGEXP,
                      EV_EXPENSE = EV_PROPDMG + EV_CROPDMG)
expense <- subset_file %>%
        group_by(EVTYPE) %>%
                summarise(TOTAL_PROPDMG = sum(EV_PROPDMG),
                          TOTAL_CROPDMG = sum(EV_CROPDMG),
                          TOTAL_EXPENSE = sum(EV_EXPENSE),
                          AVG_PROPDMG = mean(EV_PROPDMG),
```

```
                            AVG_CROPDMG = mean(EV_CROPDMG),
                            AVG_EXPENSE = mean(EV_EXPENSE),
                            NUM_EV = n()) %>%
        select(EVTYPE, NUM_EV, TOTAL_EXPENSE, AVG_EXPENSE) %>%
        arrange(desc(AVG_EXPENSE))
total_economic <- arrange(expense, desc(TOTAL_EXPENSE))
top_10_avg_expense <- expense[1:10, ]
top_10_total_expense <- total_economic[1:10, ]
```

Results

As the following chart shows, wildfires, per event, are the most dangerous type of event.

```
head(top_10_avg_danger, 10)
```

```
## Source: local data frame [10 x 4]
##
##                       EVTYPE NUMBER_OF_EVENTS TOTAL_HARM AVG_HARM
## 1                 WILD FIRES                1        153    153.0
## 2                    TSUNAMI                2        162     81.0
## 3                  Heat Wave                1         70     70.0
## 4          HURRICANE/TYPHOON               26       1339     51.5
## 5      TROPICAL STORM GORDON                1         51     51.0
## 6          WINTER WEATHER MIX               2         68     34.0
## 7  UNSEASONABLY WARM AND DRY                1         29     29.0
## 8               THUNDERSTORMW               1         27     27.0
## 9              WINTER STORMS                1         27     27.0
## 10               RECORD HEAT                2         52     26.0
```

```
par(mai = c(1.2,2.2,1,1))
barplot(top_10_avg_danger$AVG_HARM, horiz = TRUE,
        names.arg = top_10_avg_danger$EVTYPE, col = "blue", las = 1,
```

```
        cex.names = .5,
        main = "Top 10 Dangerous Events by Average",
        xlab = "Average Fatalities and Injuries")
```

Tornadoes, on the other hand, are the most dangerous overall.

```
head(top_10_total_danger, 10)
```

```
## Source: local data frame [10 x 4]
##
##                EVTYPE NUMBER_OF_EVENTS TOTAL_HARM  AVG_HARM
## 1             TORNADO             7928      96979 12.232467
## 2      EXCESSIVE HEAT              678       8428 12.430678
## 3           TSTM WIND             2930       7461  2.546416
## 4               FLOOD              410       7259 17.704878
## 5           LIGHTNING             3305       6046  1.829349
## 6                HEAT              209       3037 14.531100
## 7         FLASH FLOOD              931       2755  2.959184
## 8           ICE STORM               95       2064 21.726316
## 9   THUNDERSTORM WIND              682       1621  2.376833
## 10       WINTER STORM              227       1527  6.726872
```

```
barplot(top_10_total_danger$TOTAL_HARM, horiz = TRUE,
        names.arg = top_10_total_danger$EVTYPE, col = "green", las = 1,
        main = "Top 10 Dangerous Events by Total",
        xlab = "Total Fatalities and Injuries",
        cex.names = .5)
```

Another concern with storms is the economic impact they have. Floods were found to have the greatest overall economic impact, while hurricanes have the greatest impact per incident.

```
head(top_10_total_expense)

## Source: local data frame [6 x 4]
##
##              EVTYPE NUM_EV TOTAL_EXPENSE  AVG_EXPENSE
## 1             FLOOD  10058  150319678257   14945285.2
## 2 HURRICANE/TYPHOON     70   71913712800 1027338754.3
## 3           TORNADO  39361   57352114164    1457079.7
## 4       STORM SURGE    173   43323541000  250425092.5
## 5              HAIL  25969   18758221880     722331.3
## 6       FLASH FLOOD  20659   17562129394     850095.8

barplot(top_10_total_expense$TOTAL_EXPENSE, horiz = TRUE,
        names.arg = top_10_total_expense$EVTYPE, col = "red", las = 1,
        main = "Top 10 Expensive Events by Total",
        xlab = "Total Expense",
        cex.names = .5)
```

```
head(top_10_avg_expense)

## Source: local data frame [6 x 4]
##
##                       EVTYPE NUM_EV TOTAL_EXPENSE AVG_EXPENSE
## 1  HEAVY RAIN/SEVERE WEATHER      1    2500000000  2500000000
## 2 TORNADOES, TSTM WIND, HAIL      1    1602500000  1602500000
## 3          HURRICANE/TYPHOON     70   71913712800  1027338754
## 4             HURRICANE OPAL      8    3191846000   398980750
## 5                STORM SURGE    173   43323541000   250425092
## 6        SEVERE THUNDERSTORM      7    1205560000   172222857
```