Impact of Storms and other Severe Weather Events on Public Health and Economy

```r
library(ggplot2)
library(gridExtra)
```

```
## Loading required package: grid
```

Synopsis

The United States National Oceanic and Atmospheric Administration (NOAA) maintains a database of significant weather events which have occured in the U.S. from 1950 until 2011.

In this data analysis we explore the NOAA database to answer the following questions:

1. Which types of events in the above database are most harmful with respect to population health (fatalities and injuries)?
2. Which types of events in the above database have the greatest economic consequences (property and crop damage)?

The data was recorded over a 61 years period and accordingly of mixed quality. Some cleanup and consolidation was necessary to derive the results to the above questions.

The results are as follows:

- Tornados are the number one harm to population health.
- Floods and droughts have the greatest economic consequences.

Reading the Data File

The data was made available as a bzip2-compressed CSV file. The contents of the file is loaded into R as follows:

```r
stormData <- read.csv("data/repdata-data-StormData.csv.bz2")
newStormData <- stormData
```

Data Processing

First, we get an overview of the data:

```r
head(stormData)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
```

```
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

Extract relevant data

From the above output and additional documents (see 1 and 2) we can determine that we only require the following columns for our final analysis:

- EVTYPE – The specific type of event,
- FATALITIES and INJURIES – The impact on population health,
- PROPDMG and CROPDMG – The economic impact of property and crop damage.

The latter two are not directly recorded as numerical values but instead are given as mulitples of factors (POPDMGEXP, CROPDMGEXP, see below).

We can now reduce the data set to the relevant information as follows:

```
relevantColumns <- c("EVTYPE", "FATALITIES", "INJURIES", "CROPDMG", "CROPDMGEXP", "PROPDMG", "PROPDMGEXP
relevantStormData <- stormData[, relevantColumns]
head(relevantStormData)
```

```
##    EVTYPE FATALITIES INJURIES CROPDMG CROPDMGEXP PROPDMG PROPDMGEXP
## 1 TORNADO          0       15       0               25.0          K
## 2 TORNADO          0        0       0                2.5          K
## 3 TORNADO          0        2       0               25.0          K
## 4 TORNADO          0        2       0                2.5          K
```

```
## 5 TORNADO          0        2       0                2.5          K
## 6 TORNADO          0        6       0                2.5          K
```

Data conversion

In order to work with numerical amounts and determine a ranking of damages, the actual numerical amounts have to be calculated. However there are some confusing levels for the multiples used in both property and crop damage:

```
levels(stormData$PROPDMGEXP)
```

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

```
levels(stormData$CROPDMGEXP)
```

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

Upon further exploring the data set it could be established that the number of unclear, confusing and "NA" data entires in the relvevant column is very small against the number of valid entries.

Thus we could further reduce the data set to only include meaningful values and also convert multiples into actual numerical values:

```
multiples <- c('k', 'K', 'm', 'M', 'b', 'B')
relevantStormData <- relevantStormData[which(relevantStormData$PROPDMGEXP
                                %in% c("","H","K","M","B")),]
relevantStormData <- relevantStormData[which(relevantStormData$CROPDMGEXP
                                %in% c("","H","K","M","B")),]
expFactors <- c(1000, 1000, 1000000, 1000000, 1000000000, 1000000000)
names(expFactors) <- multiples
multiplyFactors <- function(val, exp) val * expFactors[exp]
relevantStormData$PROPDMG <- apply(relevantStormData[,c("PROPDMG","PROPDMGEXP")], 1, function(x) multip

relevantStormData$CROPDMG <- apply(relevantStormData[,c("CROPDMG","CROPDMGEXP")], 1, function(x) multip
```

Fatalities and injuries

For the final results, we calculate the sum of fatalities and injuries by event, order them and only record the top 10 results.

```
# Aggregate fatalities damages by event.
fatalities <- aggregate(FATALITIES ~ EVTYPE, relevantStormData, sum)
fatalities <- fatalities[order(-fatalities$FATALITIES),]
fatalities <- fatalities[1:10, ]
# Aggregate injuries damages by event.
injuries <- aggregate(INJURIES ~ EVTYPE, relevantStormData, sum)
injuries <- injuries[order(-injuries$INJURIES),]
injuries <- injuries[1:10, ]

fatalities
```

```
##            EVTYPE FATALITIES
## 830        TORNADO       5630
## 129 EXCESSIVE HEAT       1903
## 152    FLASH FLOOD        978
## 272           HEAT        937
## 460      LIGHTNING        816
## 852      TSTM WIND        504
## 168          FLOOD        470
## 581    RIP CURRENT        368
## 356      HIGH WIND        246
## 19       AVALANCHE        224
```

```r
injuries
```

```
##                EVTYPE INJURIES
## 830           TORNADO    91285
## 852         TSTM WIND     6957
## 168             FLOOD     6789
## 129    EXCESSIVE HEAT     6525
## 460         LIGHTNING     5230
## 272              HEAT     2100
## 423         ICE STORM     1975
## 152       FLASH FLOOD     1777
## 756 THUNDERSTORM WIND     1488
## 241              HAIL     1360
```

From these results it can clearly be seen that Tornados are the number one cause for injuries and fatalities.

Property and crop damage

The same methodology is applied to property and crop damage:

```r
# Aggregate property damages by event.
propdmg <- aggregate(PROPDMG ~ EVTYPE, relevantStormData, sum)
propdmg <- propdmg[order(-propdmg$PROPDMG),]
propdmg <- propdmg[1:10, ]

# Aggregate crop damages by event.
cropdmg <- aggregate(CROPDMG ~ EVTYPE, relevantStormData, sum)
cropdmg <- cropdmg[order(-cropdmg$CROPDMG),]
cropdmg <- cropdmg[1:10, ]
propdmg
```

```
##                EVTYPE   PROPDMG
## 62              FLOOD 1.447e+11
## 177 HURRICANE/TYPHOON 6.931e+10
## 330           TORNADO 5.693e+10
## 279       STORM SURGE 4.332e+10
## 50        FLASH FLOOD 1.614e+10
## 103              HAIL 1.573e+10
## 169         HURRICANE 1.187e+10
## 338    TROPICAL STORM 7.704e+09
```

```
## 396      WINTER STORM 6.688e+09
## 155         HIGH WIND 5.270e+09
```

```
cropdmg
```

```
##               EVTYPE   CROPDMG
## 16           DROUGHT 1.397e+10
## 34             FLOOD 5.662e+09
## 97       RIVER FLOOD 5.029e+09
## 84         ICE STORM 5.022e+09
## 52              HAIL 3.001e+09
## 76         HURRICANE 2.742e+09
## 81 HURRICANE/TYPHOON 2.608e+09
## 30       FLASH FLOOD 1.421e+09
## 26      EXTREME COLD 1.293e+09
## 46      FROST/FREEZE 1.094e+09
```

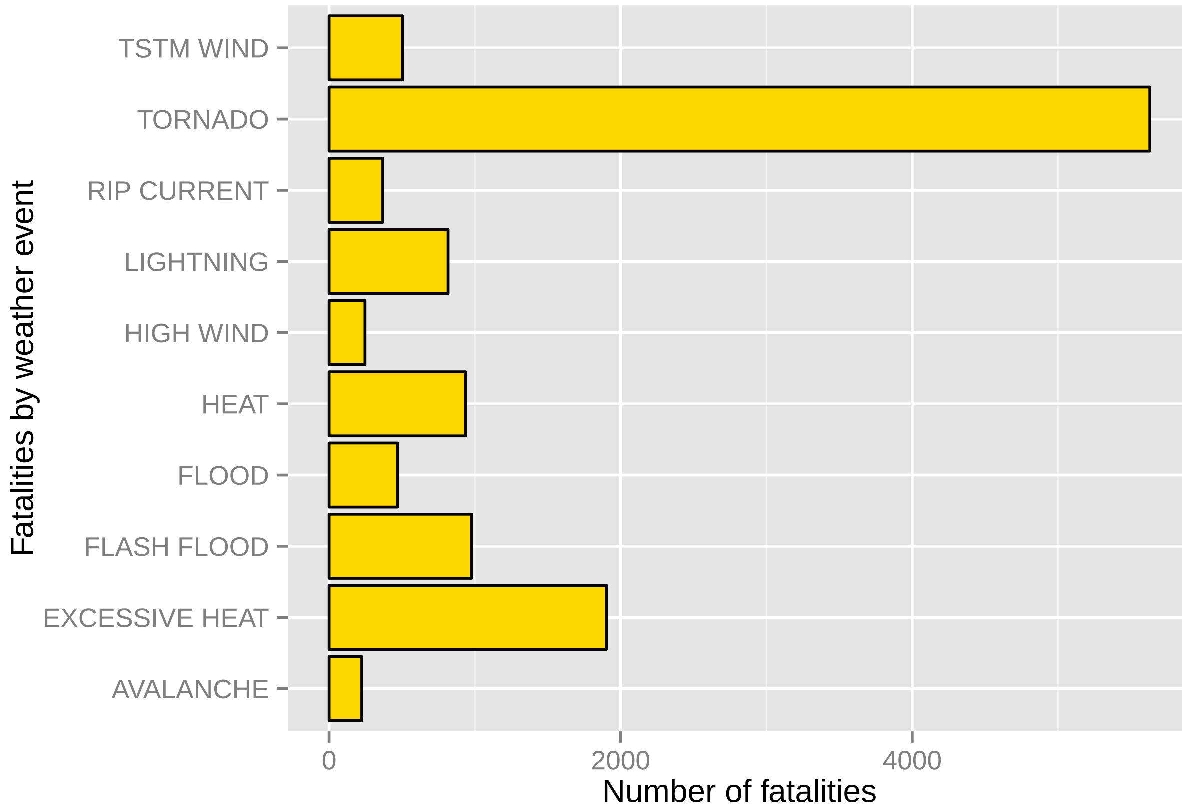
These results show that floods are the number one threat to property, and drought is the most significant threat to crop.

Results

1. Most harmful events with regards to population health across the US

Below is a plot summarising the most harmful events with regards to population health (fatalities) in the US:

```
ggplot(fatalities, aes(x=EVTYPE, y=FATALITIES)) +
    geom_bar(stat="identity", aes(fill=FATALITIES),
             fill="#FCD700", colour="black") +
    labs(x="Fatalities by weather event", y="Number of fatalities") + coord_flip()
```



Another exploratory plot shows the same result are obtained for injirues: Tornadoes.

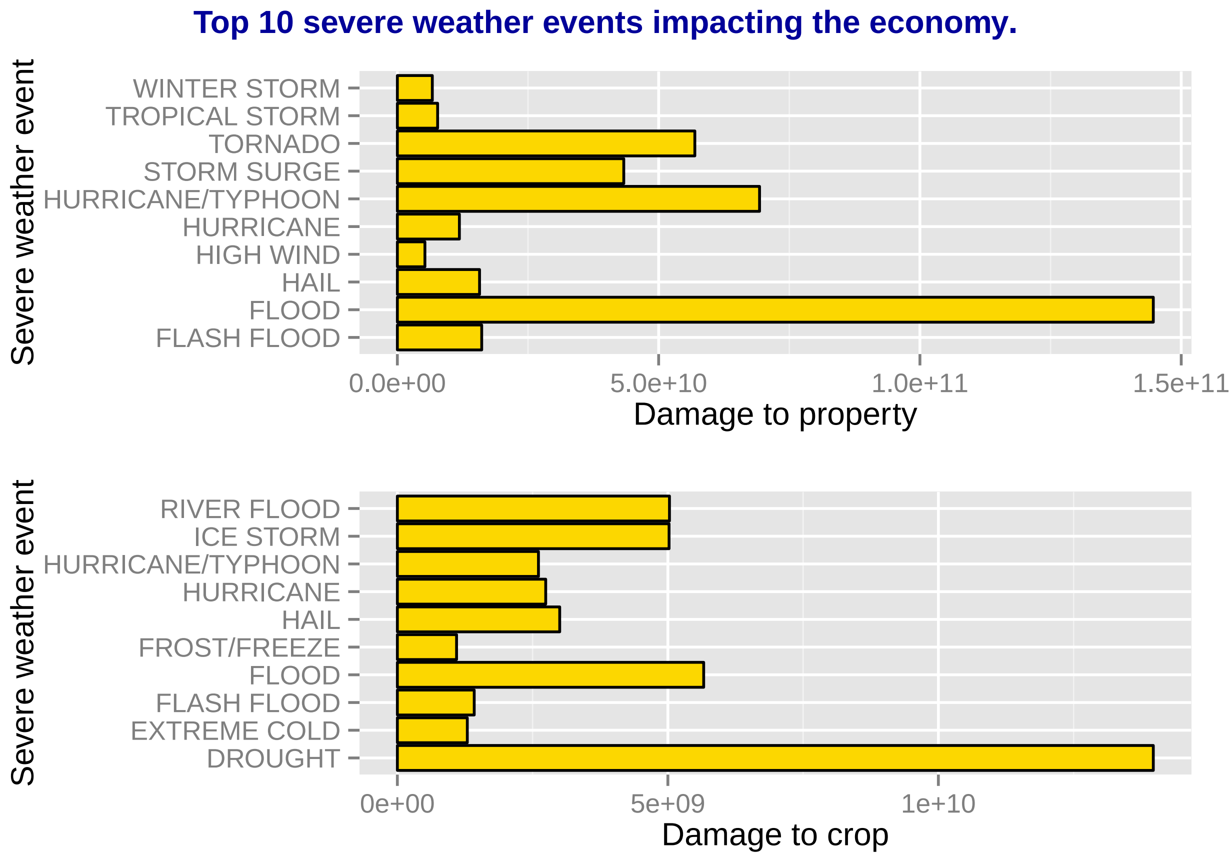
2. Events with the greatest economic consequences

Similarily, the impact of severe weather events can be summarised in a plot:

```
p1 <- ggplot(propdmg, aes(x=EVTYPE, y=PROPDMG)) +
    geom_bar(stat="identity", aes(fill=PROPDMG),
             fill="#FCD700", colour="black") +
    labs(x="Severe weather event", y="Damage to property") + coord_flip()

p2 <- ggplot(cropdmg, aes(x=EVTYPE, y=CROPDMG)) +
    geom_bar(stat="identity", aes(fill=CROPDMG),
             fill="#FCD700", colour="black") +
    labs(x="Severe weather event", y="Damage to crop") + coord_flip()

grid.arrange(p1, p2, main=textGrob("Top 10 severe weather events impacting the economy.", gp=gpar(cex=1
```



References

1. National Weather Service Storm Data Documentation
2. National Climatic Data Center Storm Events FAQ