Analysis of Weather Events And Impact On Public Health And Economics In USA (Using NOAA Database)

Reproducible Research Peer Assessment 2

Author: Mehul Patel
Date: July 25, 2015

1. Assignment

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. You must use the database to answer the questions below and show the code for your entire analysis. Your analysis can consist of tables, figures, or other summaries. You may use any R package you want to support your analysis.

2. Synopsis

This report consists in analyzing the NOAA storm database containing data on extreme climate events. This data was collected during the period from 1950 through 2011. The purpose of this analysis is to answer the following two questions:

- Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
- Across the United States, which types of events have the greatest economic consequences?

After the analysis, the conclusion is that,
1) Tornado is the most dangerous climate event with more than 5600 deaths and 91400 injuries and
2) Floods have caused the most significant economic damage - more than 157 billion USD.

Data Processing

Assuming that the data file is already in the working directory, let's do the following:

```
data <- read.csv(bzfile("repdata-data-StormData.csv.bz2"), header = TRUE,
    stringsAsFactors = FALSE)
```

Using the events specified in the NOAA's documentation, let's define all of them in one vector for ease of processing.

```
events <- c("Astronomical Low Tide", "Avalanche", "Blizzard", "Coastal Flood",
    "Cold/Wind Chill", "Debris Flow", "Dense Fog", "Dense Smoke", "Drought",
    "Dust Devil", "Dust Storm", "Excessive Heat", "Extreme cold/Wind Chill",
    "Flash Flood", "Flood", "Freezing", "Frost/Freeze", "Funnel Cloud", "Hail",
    "Heat", "Heavy Rain", "Heavy Snow", "High Surf", "High Wind", "Hurricane/Typhoon",
    "Ice Storm", "Lakeshore Flood", "Lake-Effect Snow", "Lightning", "Marine Hail",
    "Marine High Wind", "Marine Strong Wind", "Marine Thunderstorm Wind", "Rip Current",
    "Seiche", "Sleet", "Storm Tide", "Strong Wind", "Thunderstorm Wind", "Tornado",
    "Tropical Depression", "Tropical Storm", "Tsunami", "Volcanic Ash", "Waterspout",
    "Wildfire", "Winter Storm", "Winter Weather")
```

Some events are combined events separated with a slash (e.g 'Hurricane/Typhoon'). Use regular expressions to extract either a combined event (Hurricane/Typhoon) or any part of it (Hurricane or Typhoon).

```
regex_events <- c("Astronomical Low Tide|Low Tide", "Avalanche", "Blizzard",
    "Coastal Flood", "Cold/Wind Chill", "Debris Flow", "Dense Fog", "Dense Smoke",
    "Drought", "Dust Devil", "Dust Storm", "Excessive Heat", "Extreme cold/Wind Chill|Extreme Cold|Wind
    "Flash Flood", "Flood", "Freezing", "Frost/Freeze|Frost|Freeze", "Funnel Cloud",
    "Hail", "Heat", "Heavy Rain", "Heavy Snow", "High Surf", "High Wind", "Hurricane/Typhoon|Hurricane|
    "Ice Storm", "Lakeshore Flood", "Lake-Effect Snow", "Lightning", "Marine Hail",
    "Marine High Wind", "Marine Strong Wind", "Marine Thunderstorm Wind|Marine tstm Wind",
    "Rip Current", "Seiche", "Sleet", "Storm Tide", "Strong Wind", "Thunderstorm Wind|tstm wind",
    "Tornado", "Tropical Depression", "Tropical Storm", "Tsunami", "Volcanic Ash",
    "Waterspout", "Wildfire", "Winter Storm", "Winter Weather")
```

The following columns are going to be used for analysis as they correspond to exactly what we are looking for:
- EVTYPE : Type of event
- FATALITIES : Number of fatalities
- INJURIES : Number of injuries
- PROPDMG : Amount of property damaged
- PROPDMGEXP : Order of Mangitude of property damage(k for thousand)
- CROPDMG : Amoung of crop damaged
- CROPDMGEXP : Order of Magnitude of crop damage(M for Million)

```
options(scipen = 999)  # fixed notation of numbers instead of scientific numbers
cleanData <- data.frame(EVTYPE = character(0), FATALITIES = numeric(0), INJURIES = numeric(0),
    PROPDMG = numeric(0), PROPDMGEXP = character(0), CROPDMG = numeric(0), CROPDMGEXP = character(0))

    for (i in 1:length(events)) {
    rows <- data[grep(regex_events[i], ignore.case = TRUE, data$EVTYPE),
        ]
    rows <- rows[, c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP",
        "CROPDMG", "CROPDMGEXP")]
    cleanDataName <- c(rep(events[i], nrow(rows)))
    rows <- cbind(rows, cleanDataName)
    cleanData <- rbind(cleanData, rows)
}
```

Take into account the order of magnitude of property and crop damage (H = hundreds, K = thousands, M = millions, B= billions)

```
cleanData[(cleanData$PROPDMGEXP == "K" | cleanData$PROPDMGEXP == "k"), ]$PROPDMGEXP <- 3
cleanData[(cleanData$PROPDMGEXP == "M" | cleanData$PROPDMGEXP == "m"), ]$PROPDMGEXP <- 6
cleanData[(cleanData$PROPDMGEXP == "B" | cleanData$PROPDMGEXP == "b"), ]$PROPDMGEXP <- 9
cleanData[(cleanData$CROPDMGEXP == "K" | cleanData$CROPDMGEXP == "k"), ]$CROPDMGEXP <- 3
cleanData[(cleanData$CROPDMGEXP == "M" | cleanData$CROPDMGEXP == "m"), ]$CROPDMGEXP <- 6
cleanData[(cleanData$CROPDMGEXP == "B" | cleanData$CROPDMGEXP == "b"), ]$CROPDMGEXP <- 9
```

Compute combined economic damage (property damage + crops damage)

```
suppressWarnings(cleanData$PROPDMG <- cleanData$PROPDMG * 10^as.numeric(cleanData$PROPDMGEXP))
suppressWarnings(cleanData$CROPDMG <- cleanData$CROPDMG * 10^as.numeric(cleanData$CROPDMGEXP))

# compute combined economic damage (property damage + crops damage)
```

```
suppressWarnings(totalEcoDamage <- cleanData$PROPDMG + cleanData$CROPDMG)
cleanData <- cbind(cleanData, totalEcoDamage)
cleanData <- cleanData[, c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "CROPDMG",
    "cleanDataName", "totalEcoDamage")]
```

Now the data is ready for plotting.

RESULTS

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

Fatalies and Injuries:

- Data for fatalities

```
fatalities <- aggregate(FATALITIES ~ cleanDataName, data = cleanData, FUN = sum)
fatalities <- fatalities[order(fatalities$FATALITIES, decreasing = TRUE), ]
# 10 most harmful causes of fatalities
fatalitiesMax <- fatalities[1:10, ]
print(fatalitiesMax)
```

```
##              cleanDataName FATALITIES
## 38                 Tornado       5661
## 19                    Heat       3138
## 11          Excessive Heat       1922
## 14                   Flood       1525
## 13             Flash Flood       1035
## 28               Lightning        817
## 37       Thunderstorm Wind        753
## 33             Rip Current        577
## 12 Extreme cold/Wind Chill        382
## 23               High Wind        299
```

- Data for injuries

```
injuries <- aggregate(INJURIES ~ cleanDataName, data = cleanData, FUN = sum)
injuries <- injuries[order(injuries$INJURIES, decreasing = TRUE), ]
# 10 most harmful causes of injuries
injuriesMax <- injuries[1:10, ]
print(injuriesMax)
```

```
##        cleanDataName INJURIES
## 38           Tornado    91407
## 37 Thunderstorm Wind     9493
## 19              Heat     9224
## 14             Flood     8604
## 11    Excessive Heat     6525
## 28         Lightning     5232
## 25         Ice Storm     1992
```

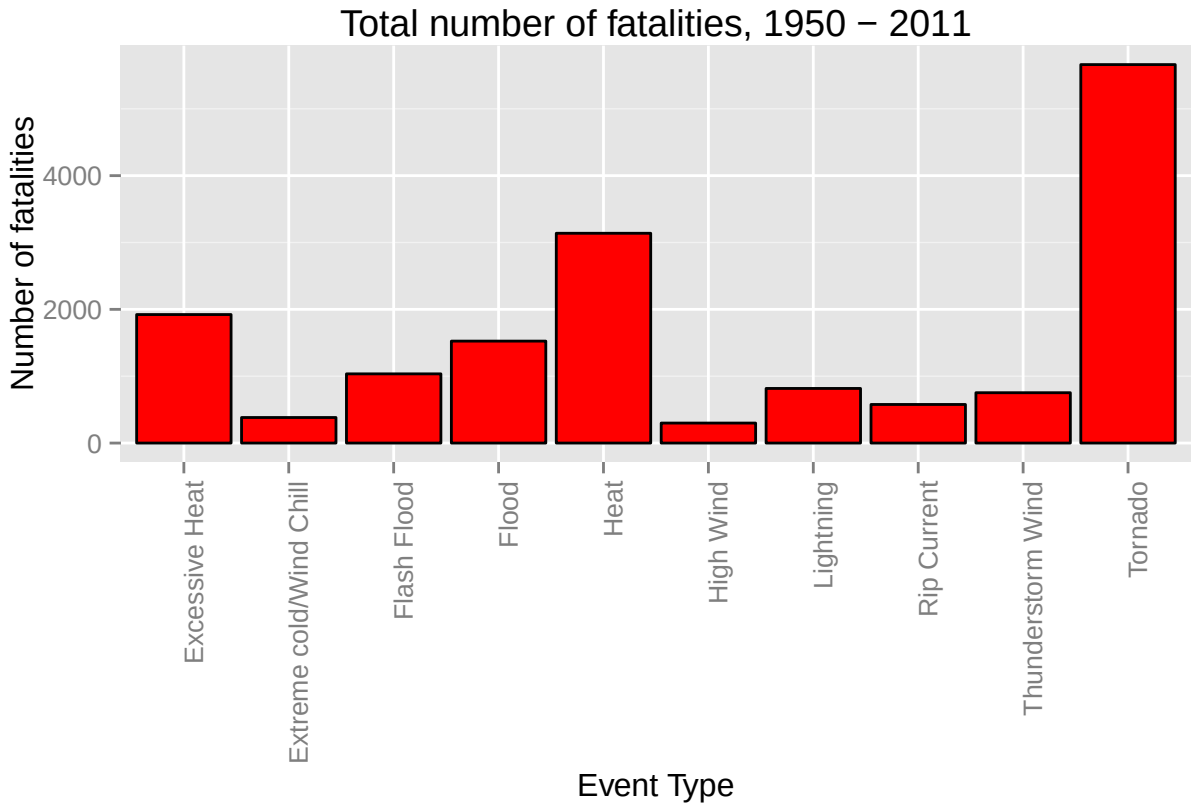
```
## 13       Flash Flood     1802
## 23         High Wind     1523
## 18              Hail     1467
```

Load the library to be used for plotting

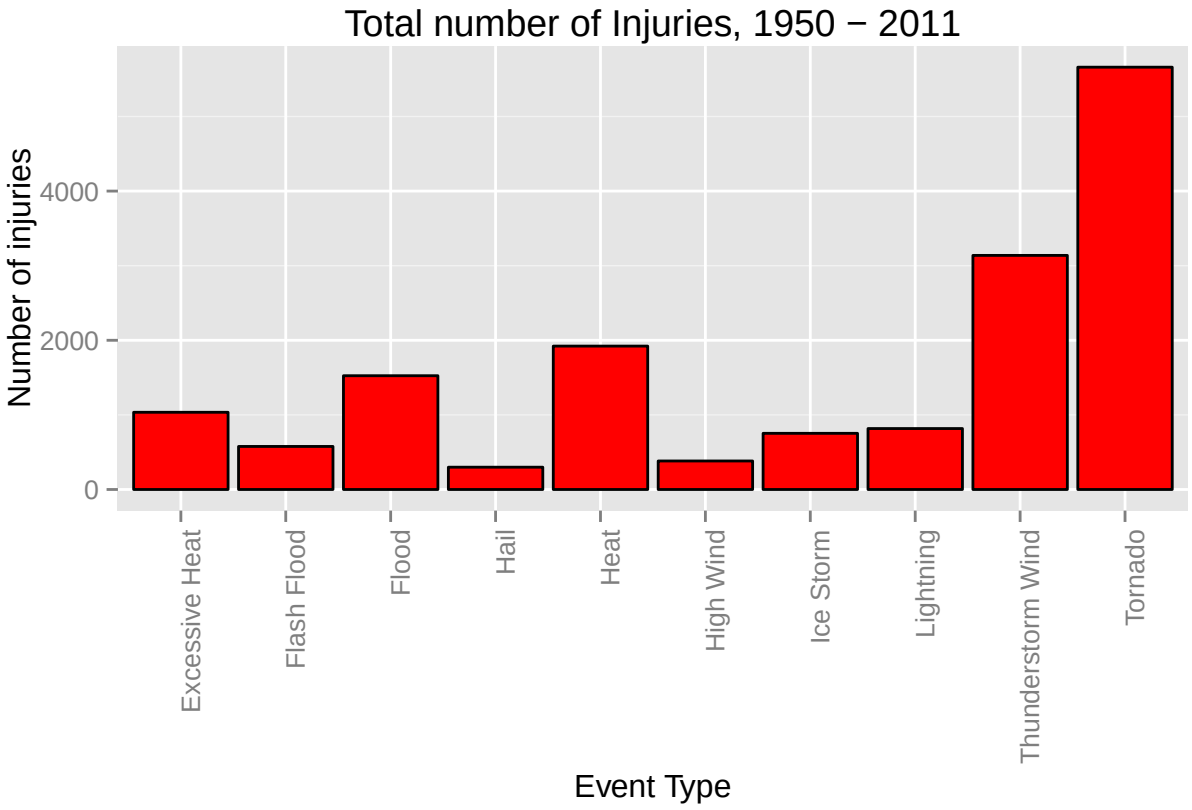
```
library(ggplot2)
```

Plotting for 10 most dangerous events for each type of damage:

```
# fatalities
ggplot(data = fatalitiesMax, aes(x = fatalitiesMax$cleanDataName, y = fatalitiesMax$FATALITIES)) +
    geom_bar(colour = "black", fill = "red", stat = "identity") + xlab("Event Type") +
    ylab("Number of fatalities") + ggtitle("Total number of fatalities, 1950 - 2011") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```
# injuries
ggplot(data = injuriesMax, aes(x = injuriesMax$cleanDataName, y = fatalitiesMax$FATALITIES)) +
    geom_bar(colour = "black", fill = "red", stat = "identity") + xlab("Event Type") +
    ylab("Number of injuries") + ggtitle("Total number of Injuries, 1950 - 2011") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Question 2: Across the United States, which types of events have the greatest economic consequences?

Property and crops combined economic damage

- Data for property damage

```
propdmg <- aggregate(PROPDMG ~ cleanDataName, data = cleanData, FUN = sum)
propdmg <- propdmg[order(propdmg$PROPDMG, decreasing = TRUE), ]
# 10 most harmful causes of injuries
propdmgMax <- propdmg[1:10, ]
print(propdmgMax)
```

```
##        cleanDataName      PROPDMG
## 14             Flood 168212215588
## 24 Hurricane/Typhoon  85356410010
## 38           Tornado  58603317864
## 18              Hail  17622990956
## 13       Flash Flood  17588791878
## 37 Thunderstorm Wind  11575228673
## 40    Tropical Storm   7714390550
## 45      Winter Storm   6749997251
## 23         High Wind   6166300000
## 44          Wildfire   4865614000
```

- Data for crop damage

```
cropdmg <- aggregate(CROPDMG ~ cleanDataName, data = cleanData, FUN = sum)
cropdmg <- cropdmg[order(cropdmg$CROPDMG, decreasing = TRUE), ]
# 10 most harmful causes of injuries
cropdmgMax <- cropdmg[1:10, ]
print(cropdmgMax)
```

```
##              cleanDataName     CROPDMG
## 8                  Drought 13972621780
## 14                   Flood 12380109100
## 24       Hurricane/Typhoon  5516117800
## 25               Ice Storm  5022113500
## 18                    Hail  3114212870
## 16            Frost/Freeze  1997061000
## 13             Flash Flood  1532197150
## 12 Extreme cold/Wind Chill  1313623000
## 37       Thunderstorm Wind  1255947980
## 19                    Heat   904469280
```

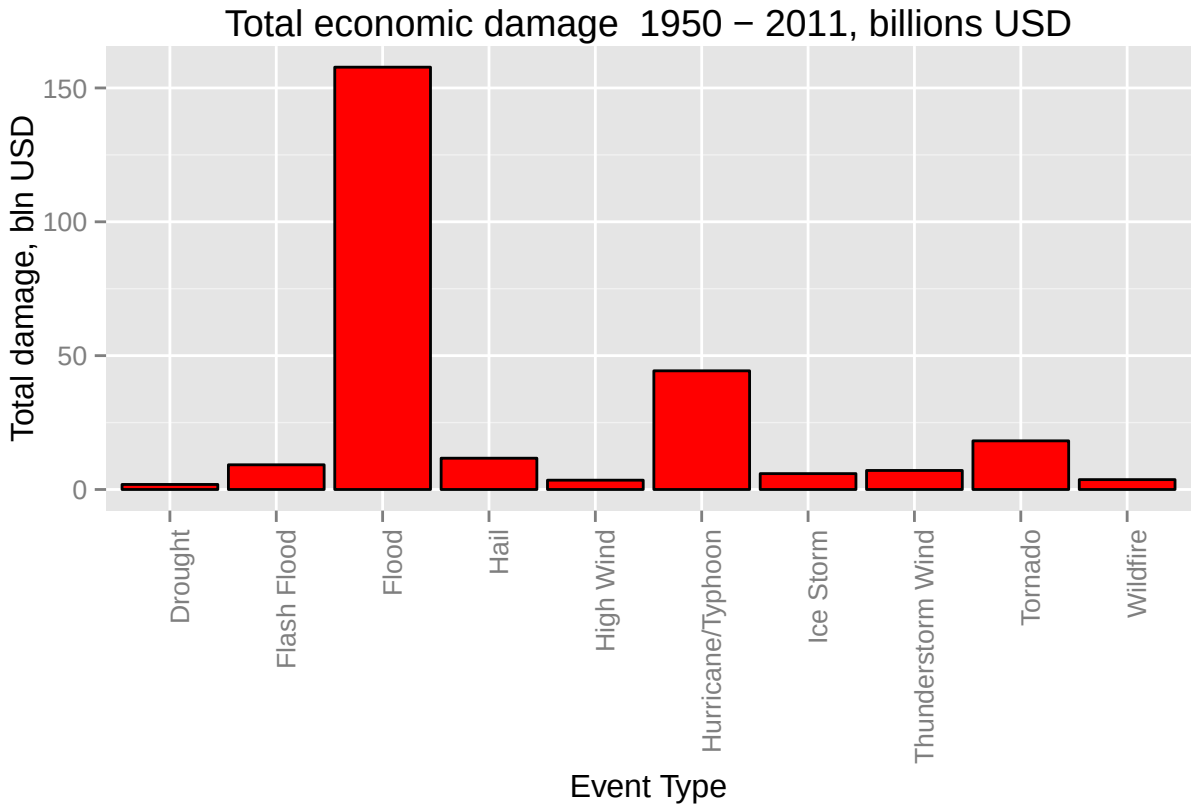
- Total economic damage

```
ecodmg <- aggregate(totalEcoDamage ~ cleanDataName, data = cleanData, FUN = sum)
ecodmg <- ecodmg[order(ecodmg$totalEcoDamage, decreasing = TRUE), ]
# 10 most harmful causes of property damage
ecodmgMax <- ecodmg[1:10, ]
print(ecodmgMax)
```

```
##        cleanDataName totalEcoDamage
## 14             Flood   157764680787
## 24 Hurricane/Typhoon    44330000800
## 38           Tornado    18172843863
## 18              Hail    11681050140
## 13       Flash Flood     9224527227
## 37 Thunderstorm Wind     7098296330
## 25         Ice Storm     5925150850
## 44          Wildfire     3685468370
## 23         High Wind     3472442200
## 8            Drought     1886667000
```

Plot the total economic damage graph. Separate information on damage to property and to crops is listed above.

```
# total economic damage (property + crops)
ggplot(data = ecodmgMax, aes(x = ecodmgMax$cleanDataName, y = ecodmgMax$totalEcoDamage/10^9)) +
    geom_bar(colour = "black", fill = "red", stat = "identity") + xlab("Event Type") +
    ylab("Total damage, bln USD") + ggtitle("Total economic damage  1950 - 2011, billions USD") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Total number of distinct climatic events as per the documentation that were extracted form the data set:

```
nrow(ecodmg)
```

```
## [1] 46
```

Data for 46 events as per documentation was extracted. Two events were not found.

CONCLUSION

Tornados have caused the greatest number of fatalities - 5,661 and injuries - 91,407 followed by Heat in terms of fatalities 3,138 (9,224 injuries slightly less than Thunderstorm Wind 9,493 injuries which is the second harrmful cause in terms of injuries).

Floods have caused the most significant economic damage 157,764,680,787 USD (combined for property loss and crops damage) followed by Hurricanes and Typhoons - 44,330,000,800 USD