PA2

Krupa

Wednesday, July 22, 2015

Reproducible Research - Peer Assessment 2

U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database analysis on the health and economic impact by the severe weather events

SYNOPSIS

Storm and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severs events can results in fatalities, injuries and property damage. Preventing such outcomes to the extent possible is a key concern. The U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database tracks characteristics of major storms and weather events in the United States, include when and where they occur, as well as estimates of any fatalities, injuries and property damage.

The project involves exploring and analysing NOAA's storm database. This report contains results on the health and economic impact by the severe weather events based on the data from NOAA database.

DATA PROCESSING

Loading Data into local dataset

```
# Download file from URL provided in the assignment

setwd("D:/Data specialist course/Reproducible research/Peer Review/PA2")

if (!file.exists("stormData.csv.bz2"))
{
  url <-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file(url,"stormData.csv.bz2")
}

# Unzip file
if (!file.exists("stormData.csv"))
{
    bunzip2("stormData.csv.bz2", "stormData.csv", remove = FALSE)
}

# load data into dataset stromData
stormData <- read.csv("stormData.csv")
dim(stormData)
```

```
## [1] 902297     37
```

Subsetting required data for analysis

```
names(stormData)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

```
stormDataNew <- stormData[c("EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CR
OPDMG","CROPDMGEXP")]
str(stormDataNew)
```

```
## 'data.frame':    902297 obs. of  7 variables:
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834
834 834 834 834 834 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 17
...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
head(stormDataNew)
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0       15    25.0          K       0
## 2 TORNADO          0        0     2.5          K       0
## 3 TORNADO          0        2    25.0          K       0
## 4 TORNADO          0        2     2.5          K       0
## 5 TORNADO          0        2     2.5          K       0
## 6 TORNADO          0        6     2.5          K       0
```

Exploring the Property Exponent Data (Factor)

```
unique(stormDataNew$PROPDMGEXP)
```

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

Rehash the Property Exponent Data

```
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "K"] <- 1000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "M"] <- 1000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == ""] <- 1
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "B"] <- 1000000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "m"] <- 1000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "0"] <- 1
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "5"] <- 100000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "6"] <- 1000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "4"] <- 10000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "2"] <- 100
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "3"] <- 1000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "h"] <- 100
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "7"] <- 10000000
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "H"] <- 100
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "1"] <- 10
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "8"] <- 100000000

# Invalid exponent values ("=", "-", "?") are given value zero
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "+"] <- 0
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "-"] <- 0
stormDataNew$PROPEXP[stormDataNew$PROPDMGEXP == "?"] <- 0
```

Calculate Property Damage Value (PROPDMGVAL)

```
stormDataNew$PROPDMGVAL <- stormDataNew$PROPDMG * stormDataNew$PROPEXP
```

Exploring the Crop Exponent Data (Factor)

```
unique(stormDataNew$CROPDMGEXP)
```

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

Rehash the Crop Exponent Data

```
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "M"] <- 1000000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "K"] <- 1000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "m"] <- 1000000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "B"] <- 1000000000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "0"] <- 1
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "k"] <- 1000
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "2"] <- 100
stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == ""] <- 1

# Invalid exponent values ("=", "-", "?") are given value zero

stormDataNew$CROPEXP[stormDataNew$CROPDMGEXP == "?"] <- 0
```

Calculate Crop Damage Value (CROPDMGVAL)

```
stormDataNew$CROPDMGVAL <- stormDataNew$CROPDMG * stormDataNew$CROPEXP
```

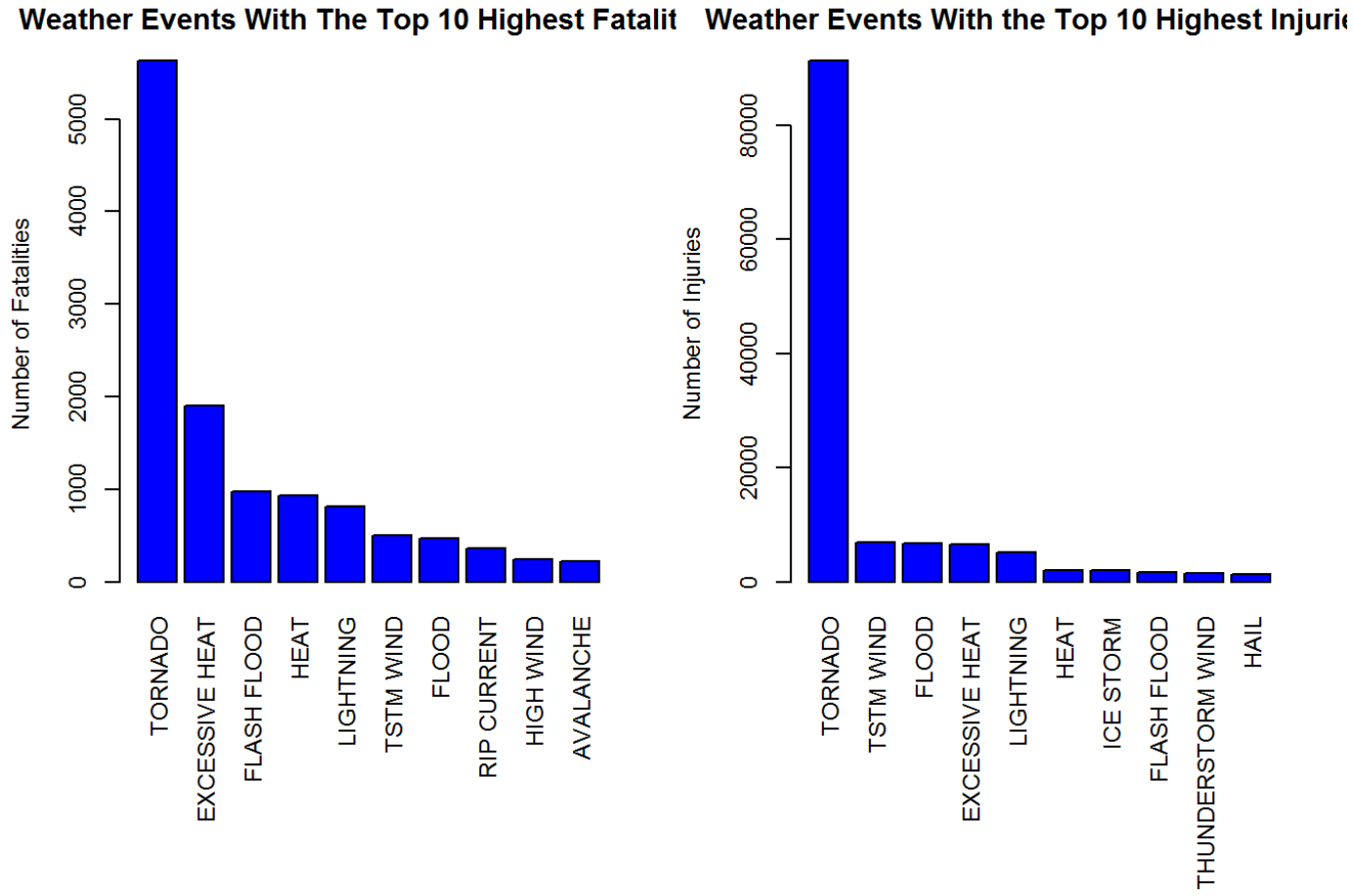
Aggregate the storm data based on event

```
Fatal <- aggregate(FATALITIES ~ EVTYPE, data = stormDataNew, FUN = sum)
Injury <- aggregate(INJURIES ~ EVTYPE, data = stormDataNew, FUN = sum)
PropertyDamage <- aggregate(PROPDMGVAL ~ EVTYPE, data = stormDataNew, FUN = sum)
CropDamage <- aggregate(CROPDMGVAL ~ EVTYPE, data = stormDataNew, FUN = sum)

# Extract top 10 events with highest fatalities
FatalTop10 <- Fatal[order(-Fatal$FATALITIES), ][1:10, ]

# Extract top 10 event with highest injuries
InjuryTop10 <- Injury[order(-Injury$INJURIES), ][1:10, ]

# Extract top 10 events with highest Property Damage
PropertyDamageTop10 <- PropertyDamage [order(-PropertyDamage$PROPDMGVAL), ][1:10, ]
# Extract  top 10 events with highest Crop Damage
CropDamageTop10 <- CropDamage [order(-CropDamage$CROPDMGVAL), ][1:10, ]
```

RESULTS

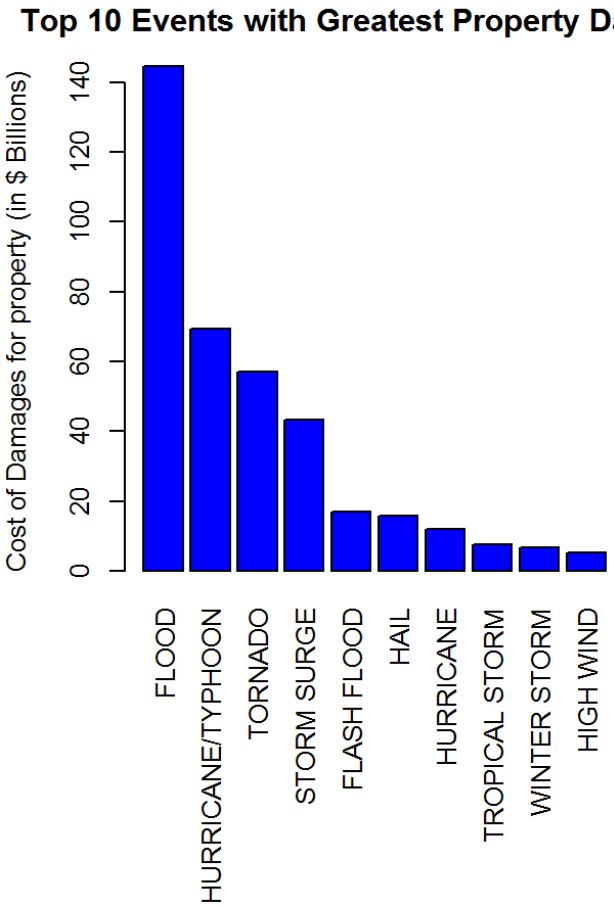
Plot 1 : Bar graph of the top 10 events based on Fatalities and Injuries

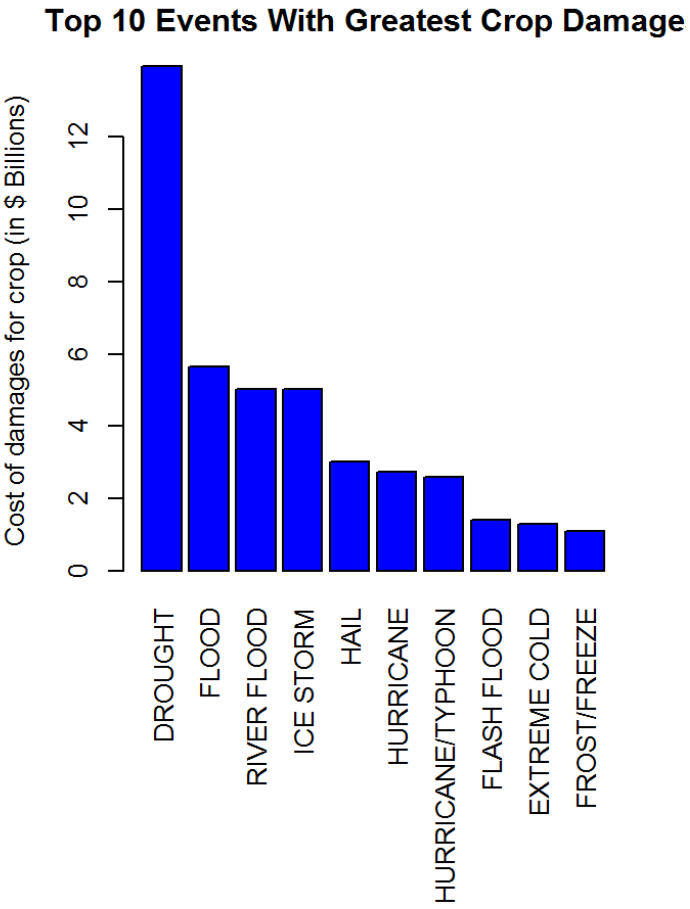
```
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.75)
barplot( FatalTop10$FATALITIES,
        las = 3,
        names.arg = FatalTop10$EVTYPE,
        main = "Weather Events With The Top 10 Highest Fatalities",
        ylab = "Number of Fatalities",
        col = "blue")

barplot(InjuryTop10$INJURIES,
        las = 3,
        names.arg = InjuryTop10$EVTYPE,
        main = "Weather Events With the Top 10 Highest Injuries",
        ylab = "Number of Injuries",
        col = "blue")
```



Plot 2: Bar graph of the top 10 events based on damage to Property and Crop

```
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(PropertyDamageTop10$PROPDMGVAL/(10^9),
        las = 3,
        names.arg = PropertyDamageTop10$EVTYPE,
        main = "Top 10 Events with Greatest Property Damages",
        ylab = "Cost of Damages for property (in $ Billions)",
        col = "blue")

barplot(CropDamageTop10$CROPDMGVAL/(10^9),
        las = 3,
        names.arg = CropDamageTop10$EVTYPE,
        main = "Top 10 Events With Greatest Crop Damages",
        ylab = "Cost of damages for crop (in $ Billions)",
        col = "blue")
```





CONCLUSIONS

Across United States, based on the data provided for storms between 1950 and Nov 2011

- Tornado has caused the highest Fatalities and Injuries (As seen from the Plot1)
- Flood, Drought, Tornado and Typhoon are the top 3 events which caused greatest property damages (As seen from the plot 2)
- Drought, Flood, River Flood and Ice Storm are the top 3 events which caused greatest crop damages (As seen from the plot 2)