StormDataAnalysis

Ashima Jain

Sunday, April 03, 2016

Reproducible Research Peer Assessment 2

Impact of weather events on human population and economy in the US

Introduction

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

Synopsis

In this project, we have tried to find how the different weather events have repercussive effects both on human population as well as at the economic front. With this report, one can conclude that tornadoes are majorly responsible for the maximum number of fatalities and injuries. Flooding and high surf cause most of the damage to the crops while tornadoes again in case of damage to property. The report further contains plots showing the top 11 events causing damage to life, property and crop.

Data Processing

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size.

```
filename <- "repdata-data-StormData.csv.bz2"
if (!file.exists("repdata-data-StormData.csv")) {
  bzip2(filename)
}

# Load dataset
data <- read.csv('repdata-data-StormData.csv')
head(data)
```

```
##   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
```

```
summary(data)
```

```
##     STATE__                  BGN_DATE             BGN_TIME
##  Min.   : 1.0   5/25/2011 0:00:00:  1202   12:00:00 AM: 10163
##  1st Qu.:19.0   4/27/2011 0:00:00:  1193   06:00:00 PM:  7350
##  Median :30.0   6/9/2011 0:00:00 :  1030   04:00:00 PM:  7261
##  Mean   :31.2   5/30/2004 0:00:00:  1016   05:00:00 PM:  6891
##  3rd Qu.:45.0   4/4/2011 0:00:00 :  1009   12:00:00 PM:  6703
##  Max.   :95.0   4/2/2006 0:00:00 :   981   03:00:00 PM:  6700
##                 (Other)          :895866   (Other)    :857229
##    TIME_ZONE          COUNTY           COUNTYNAME         STATE
##  CST    :547493   Min.   :  0.0   JEFFERSON :  7840   TX     : 83728
##  EST    :245558   1st Qu.: 31.0   WASHINGTON:  7603   KS     : 53440
##  MST    : 68390   Median : 75.0   JACKSON   :  6660   OK     : 46802
##  PST    : 28302   Mean   :100.6   FRANKLIN  :  6256   MO     : 35648
##  AST    :  6360   3rd Qu.:131.0   LINCOLN   :  5937   IA     : 31069
##  HST    :  2563   Max.   :873.0   MADISON   :  5632   NE     : 30271
##  (Other):  3631                   (Other)   :862369   (Other):621339
##                EVTYPE         BGN_RANGE           BGN_AZI
##  HAIL              :288661   Min.   :   0.000          :547332
##  TSTM WIND         :219940   1st Qu.:   0.000   N      : 86752
##  THUNDERSTORM WIND: 82563   Median :   0.000   W      : 38446
##  TORNADO           : 60652   Mean   :   1.484   S      : 37558
##  FLASH FLOOD       : 54277   3rd Qu.:   1.000   E      : 33178
##  FLOOD             : 25326   Max.   :3749.000   NW     : 24041
##  (Other)           :170878                      (Other):134990
##          BGN_LOCATI                  END_DATE             END_TIME
##                :287743                     :243411             :238978
##  COUNTYWIDE    : 19680   4/27/2011 0:00:00:  1214   06:00:00 PM:  9802
##  Countywide    :   993   5/25/2011 0:00:00:  1196   05:00:00 PM:  8314
##  SPRINGFIELD   :   843   6/9/2011 0:00:00 :  1021   04:00:00 PM:  8104
##  SOUTH PORTION:   810   4/4/2011 0:00:00 :  1007   12:00:00 PM:  7483
##  NORTH PORTION:   784   5/30/2004 0:00:00:   998   11:59:00 PM:  7184
##  (Other)       :591444   (Other)          :653450   (Other)    :622432
##    COUNTY_END COUNTYENDN       END_RANGE           END_AZI
##  Min.   :0    Mode:logical   Min.   :  0.0000          :724837
##  1st Qu.:0    NA's:902297    1st Qu.:  0.0000   N      : 28082
##  Median :0                   Median :  0.0000   S      : 22510
##  Mean   :0                   Mean   :  0.9862   W      : 20119
##  3rd Qu.:0                   3rd Qu.:  0.0000   E      : 20047
##  Max.   :0                   Max.   :925.0000   NE     : 14606
##                                                 (Other): 72096
##             END_LOCATI         LENGTH              WIDTH
##                   :499225   Min.   :   0.0000   Min.   :   0.000
##  COUNTYWIDE       : 19731   1st Qu.:   0.0000   1st Qu.:   0.000
##  SOUTH PORTION    :   833   Median :   0.0000   Median :   0.000
##  NORTH PORTION    :   780   Mean   :   0.2301   Mean   :   7.503
##  CENTRAL PORTION  :   617   3rd Qu.:   0.0000   3rd Qu.:   0.000
##  SPRINGFIELD      :   575   Max.   :2315.0000   Max.   :4400.000
##  (Other)          :380536
##        F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
```

```
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG          PROPDMGEXP        CROPDMG          CROPDMGEXP
##  Min.   :   0.00          :465934   Min.   :  0.000          :618413
##  1st Qu.:   0.00   K      :424665   1st Qu.:  0.000   K      :281832
##  Median :   0.00   M      : 11330   Median :  0.000   M      :  1994
##  Mean   :  12.06   0      :   216   Mean   :  1.527   k      :    21
##  3rd Qu.:   0.50   B      :    40   3rd Qu.:  0.000   0      :    19
##  Max.   :5000.00   5      :    28   Max.   :990.000   B      :     9
##                    (Other):    84                     (Other):     9
##       WFO                                       STATEOFFIC
##         :142069                                       :248769
##  OUN    : 17393   TEXAS, North                       : 12193
##  JAN    : 13889   ARKANSAS, Central and North Central: 11738
##  LWX    : 13174   IOWA, Central                      : 11345
##  PHI    : 12551   KANSAS, Southwest                  : 11212
##  TSA    : 12483   GEORGIA, North and Central         : 11120
##  (Other):690738   (Other)                            :595920
##
ZONENAMES
##
:594029
##
:205988
##  GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
:   639
##  GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA
:   592
##  JEFFERSON - JEFFERSON
:   303
##  MADISON - MADISON
:   302
##  (Other)
:100444
##     LATITUDE      LONGITUDE        LATITUDE_E     LONGITUDE_
##  Min.   :   0   Min.   :-14451   Min.   :   0   Min.   :-14455
##  1st Qu.:2802   1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0
##  Median :3540   Median :  8707   Median :   0   Median :     0
##  Mean   :2875   Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:4019   3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   :9706   Max.   : 17124   Max.   :9706   Max.   :106220
##  NA's   :47                      NA's   :40
##                                          REMARKS           REFNUM
##                                              :287433   Min.   :     1
##                                              : 24013   1st Qu.:225575
##  Trees down.\n                               :  1110   Median :451149
##  Several trees were blown down.\n            :   569   Mean   :451149
##  Trees were downed.\n                        :   446   3rd Qu.:676723
##  Large trees and power lines were blown down.\n:   432   Max.   :902297
```

```
##  (Other)                                    :588294
```

```
names(data) <- tolower(names(data))
length(levels(data$evtype))
```

```
## [1] 985
```

```
evtype <- tolower(levels(data$evtype))

# Processing of event types
data$damagesource <- NA
data[grepl("precipitation|rain|hail|drizzle|wet|percip|burst|depression|fog|wall cloud", data$evtype, ignore.case = TRUE), "damageSource"] <- "Precipitation & Fog"
data[grepl("wind|storm|wnd|hurricane|typhoon", data$evtype, ignore.case = TRUE), "damagesource"] <- "Wind & Storm"
data[grepl("slide|erosion|slump", data$evtype, ignore.case = TRUE), "damagesource"] <- "Landslide & Erosion"
data[grepl("warmth|warm|heat|dry|hot|drought|thermia|temperature record|record temperature|record high", data$evtype, ignore.case = TRUE), "damageSource"] <- "Heat & Drought"
data[grepl("cold|cool|ice|icy|frost|freeze|snow|winter|wintry|wintery|blizzard|chill|freezing|avalanche|glaze|sleet", data$evtype, ignore.case = TRUE), "damageSource"] <- "Snow & Ice"
data[grepl("flood|surf|blow-out|swells|fld|dam break", data$evtype, ignore.case = TRUE), "damagesource"] <- "Flooding & High Surf"
data[grepl("seas|high water|tide|tsunami|wave|current|marine|drowning", data$evtype, ignore.case = TRUE), "damagesource"] <- "High seas"
data[grepl("dust|saharan", data$evtype, ignore.case = TRUE), "damagesource"] <- "Dust & Saharan winds"
data[grepl("tstm|thunderstorm|lightning", data$evtype, ignore.case = TRUE), "damagesource"] <- "Thunderstorm & Lightning"
data[grepl("tornado|spout|funnel|whirlwind", data$evtype, ignore.case = TRUE), "damagesource"] <- "Tornado"
data[grepl("fire|smoke|volcanic", data$evtype, ignore.case = TRUE), "damagesource"] <- "Fire & Volcanic activity"
data <- data[complete.cases(data[, "damagesource"]), ]
data$damagesource <- as.factor(data$damagesource)

toPowerTen <- function(n){
  if(n %in% c('h','H'))
    return(2)
  else if(n %in% c('k','K'))
    return(3)
  else if(n %in% c('m', 'M'))
    return(6)
  else if(n %in% c('b','B'))
    return(9)
  else if(n %in% c('','-','?','+'))
    return(0)
  else if(is.numeric(n))
    return(n)
}

computeValue <- function(n, exp){
  if(is.numeric(n))
     n <- as.numeric(n * 10^(toPowerTen(exp)))
  else
     n <-0
  n
}
```

```
# Finding a numeric value of the economic damages caused
data$propdamage <- mapply(computeValue, data$propdmg, data$propdmgexp)
data$cropdamage <- mapply(computeValue, data$cropdmg, data$cropdmgexp)
storage.mode(data$propdamage) <- "integer"
```

```
## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion
```

```
## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion

## Warning in storage.mode(data$propdamage) <- "integer": NAs introduced by
## coercion
```

```
storage.mode(data$cropdamage) <- "integer"
```

```
## Warning in storage.mode(data$cropdamage) <- "integer": NAs introduced by
## coercion
```

```
## Warning in storage.mode(data$cropdamage) <- "integer": NAs introduced by
## coercion
```

```
data <- data[complete.cases(data[, "propdamage"]), ]
data <- data[complete.cases(data[, "cropdamage"]), ]
```

Impact on Public Health

The impact on human population is measured by the number of fatalities and injuries occurred.

```
library(plyr)
humandamage<- ddply(data, .(damagesource), summarize, fatalities = sum(fatalities), injuries = sum
(injuries))
fatal <- head(humandamage[order(humandamage$fatalities, decreasing = T), ], 10)
rownames(fatal)<-NULL
injure <- head(humandamage[order(humandamage$injuries, decreasing = T), ], 10)
rownames(injure)<-NULL
```

Impact on crops and property

The impact of the weather events on the economy of the country is measured by the damage caused on property and crops.

```
econdamage <- ddply(data, .(damagesource), summarize, propdamage = sum(as.numeric(propdamage)), cro
pdamage = sum(as.numeric(cropdamage)))
property <- econdamage[order(econdamage$propdamage, decreasing = T), ]
rownames(property)<-NULL
crop <- econdamage[order(econdamage$cropdamage, decreasing = T), ]
rownames(crop) <- NULL
```

Results

The weather events which lead to a large number of fatalities and injuries are stored in fatal and injure respectively.

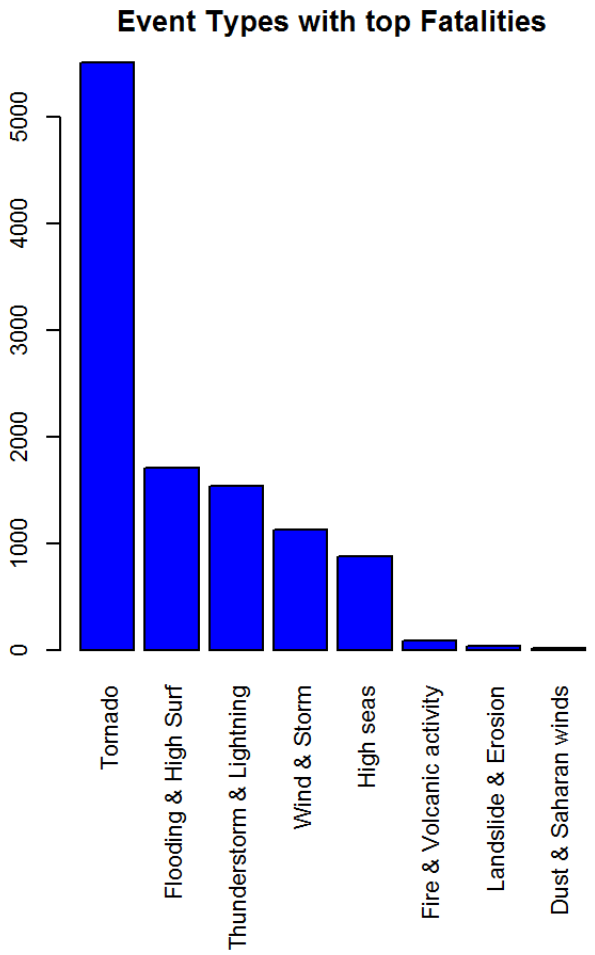
```
fatal
```

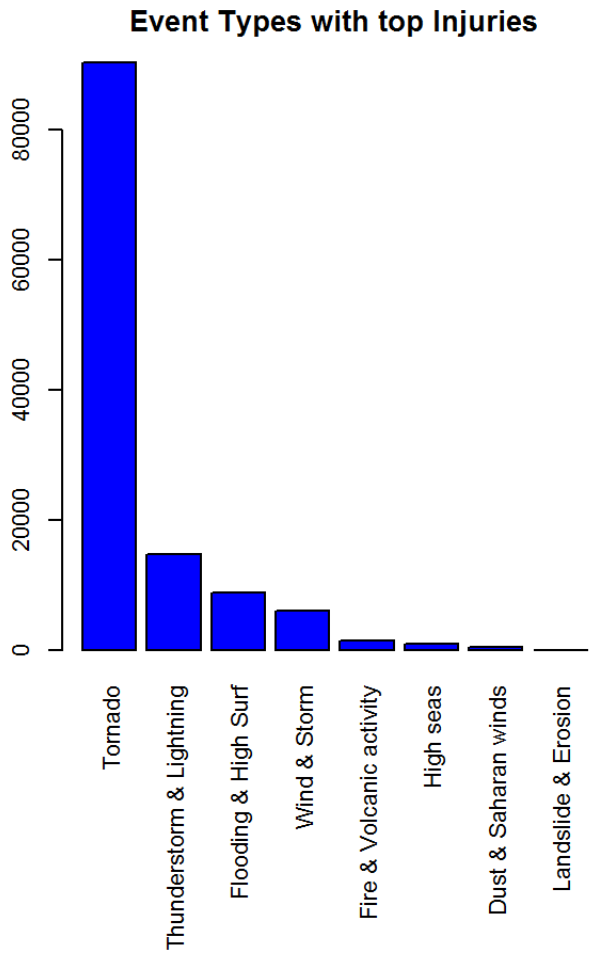
```
##               damagesource fatalities injuries
## 1                  Tornado       5504    90264
## 2     Flooding & High Surf       1715     8931
## 3 Thunderstorm & Lightning       1546    14774
## 4             Wind & Storm       1132     6147
## 5                High seas        884     1119
## 6 Fire & Volcanic activity         90     1608
## 7      Landslide & Erosion         44       55
## 8     Dust & Saharan winds         24      483
```

```
injure
```

```
##               damagesource fatalities injuries
## 1                  Tornado       5504    90264
## 2 Thunderstorm & Lightning       1546    14774
## 3     Flooding & High Surf       1715     8931
## 4             Wind & Storm       1132     6147
## 5 Fire & Volcanic activity         90     1608
## 6                High seas        884     1119
## 7     Dust & Saharan winds         24      483
## 8      Landslide & Erosion         44       55
```

```
par(mfrow=c(1,2),cex=0.7, mar = c(12, 4, 3, 2))
barplot(fatal$fatalities, names.arg=fatal$damagesource,las=3, col="blue", main="Event Types with to
p Fatalities")
barplot(injure$injuries, names.arg=injure$damagesource,las=3, col="blue", main="Event Types with to
p Injuries")
```





The above plot summarises the effect of the weather events on public health with tornado being on the top.

The weather events which lead to major damage to crop and property are stored in crop and property respectively.

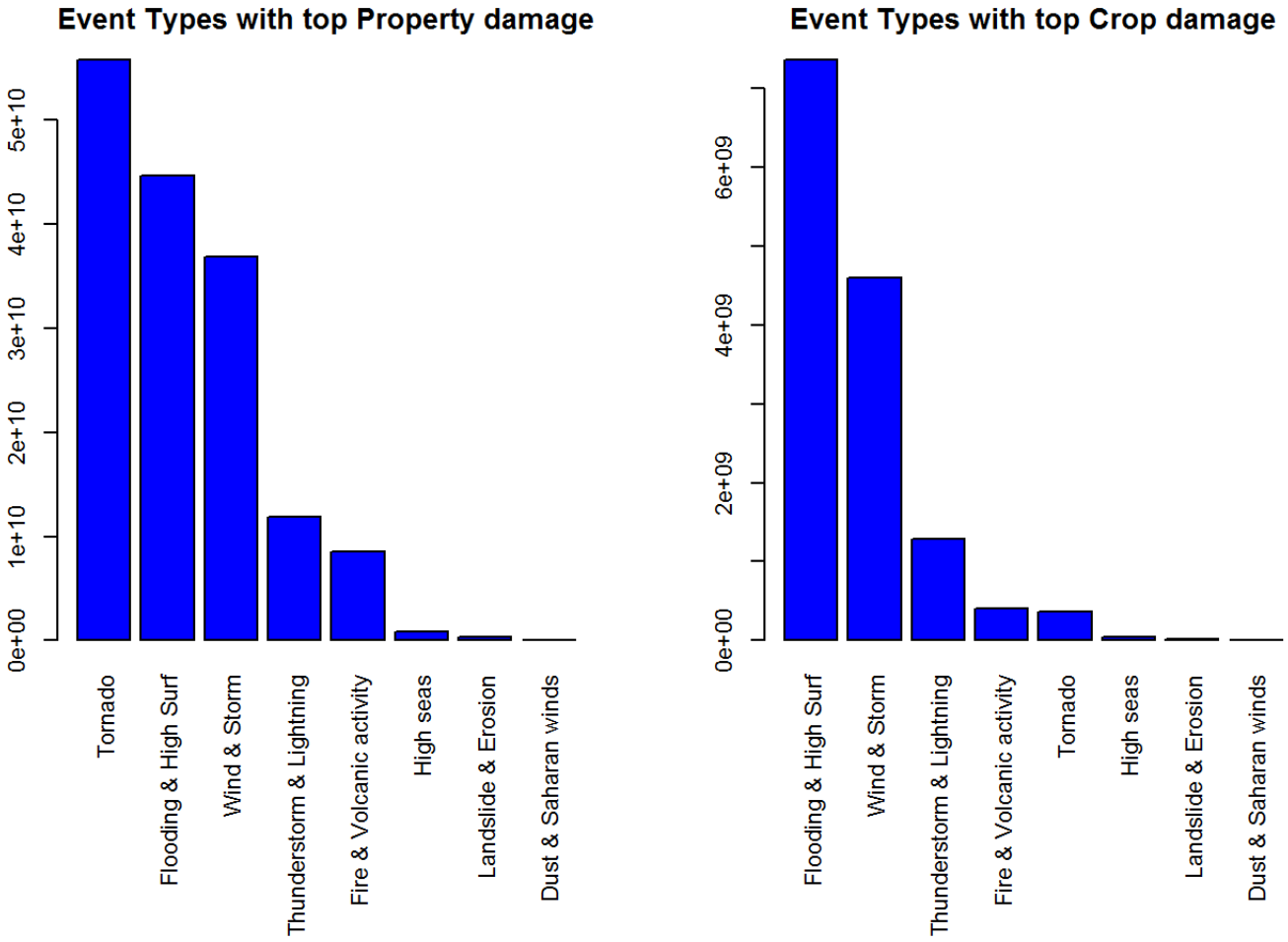
```
crop
```

```
##               damagesource  propdamage cropdamage
## 1     Flooding & High Surf 44703597200 7356977200
## 2             Wind & Storm 36916488314 4602403350
## 3 Thunderstorm & Lightning 11909781237 1283798498
## 4 Fire & Volcanic activity  8502228498  403281630
## 5                  Tornado 55802706092  367458360
## 6                High seas   809787890   46622500
## 7      Landslide & Erosion   328262100   20017000
## 8     Dust & Saharan winds     6337630    3600000
```

```
property
```

```
##               damagesource  propdamage cropdamage
## 1                  Tornado 55802706092  367458360
## 2      Flooding & High Surf 44703597200 7356977200
## 3              Wind & Storm 36916488314 4602403350
## 4 Thunderstorm & Lightning 11909781237 1283798498
## 5 Fire & Volcanic activity  8502228498  403281630
## 6                 High seas   809787890   46622500
## 7       Landslide & Erosion   328262100   20017000
## 8      Dust & Saharan winds     6337630    3600000
```

```
par(mfrow=c(1,2),cex=0.7, mar = c(12, 4, 3, 2))
barplot(property$propdamage, names.arg=property$damagesource,las=3, col="blue", main="Event Types w
ith top Property damage")
barplot(crop$cropdamage, names.arg=crop$damagesource,las=3, col="blue", main="Event Types with top
Crop damage")
```



The above plot summarises the effect of the weather events on the economy.Tornado causes most damage to property while flooding and high surf to crop.