Assignment: Reproducible Research Week 4 Course Project 2

OVERVIEW

- Weather events cause public health and economic problems for communities and municipalities. Severe events result in fatalities, injuries, and damage. Predicting and/or preventing these outcomes is a primary objective.
- This analysis examines the damaging effects of severe weather conditions (e.g. hurricanes, tornadoes, thunderstorms, floods, etc.) on human populations and the economy in the U.S. from 1950 to 2011.
- As a result, the analysis will highlight the severe weather events associated with the greatest impact on the economy and population health.

SYNOPSIS

- This is an exploration of 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, which type of event, as well as the estimates of relevant fatalities, injuries, and various forms of damage.
- The dataset used in this project is provided by the U.S. National Oceanic and Atmospheric Administration (NOAA).
- This analysis discovered that tornados are responsible for a maximum number of fatalities and injuries.
- This analysis also discovered that floods are responsible for maximum property damage, while Droughts cause maximum crop damage.

Objective: Explore the NOAA Storm Database to help answer important questions about severe weather events.

DATA PROCESSING

DATA PREP

DP1.1 Install packages & Load libraries

Install packages,

```
# load libraries ...

library(R.utils)
library(rmarkdown)
library(knitr)
Sys.setlocale("LC_TIME", "English")
```

DP1.2 Download the storm data file into the designated working directory folder

```
temp <- tempfile()

##Performing the download
if(!file.exists("/stormData.csv.bz2")){
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2F
StormData.csv.bz2", destfile="./stormData.csv.bz2")
}
##Uncompressing the file
if(!file.exists("stormdata.csv"))
{
  bunzip2("stormData.csv.bz2","stormdata.csv",remove=F)
}

# DP1.3 loading the data & reading the file

storm <- read.csv("stormdata.csv",header=TRUE,sep=",")
summary(storm)
```

```
##      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.   :   0.0
000  
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0
000  
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0
000  
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1
557  
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0
000  
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0
000  
##  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.:2255
75
##  Trees down.\n                                          :  1110   Median :4511
49
##  Several trees were blown down.\n                       :   569   Mean   :4511
49
##  Trees were downed.\n                                   :   446   3rd Qu.:6767
23
##  Large trees and power lines were blown down.\n:   432   Max.   :9022
97
##  (Other)                                                :588294
```

```
names(storm)
```

```
##  [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"
```

RESULTS

QUESTION 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

Q1.1 Variable selection (reducing the data set to only needed columns and variables)

```
variables<-c("EVTYPE","FATALITIES","INJURIES","PROPDMG", "PROPDMGEXP","C
ROPDMG","CROPDMGEXP")
strmdata<-storm[variables]


dim(strmdata)
```

```
## [1] 902297      7
```

```
## [1] "EVTYPE"     "FATALITIES" "INJURIES"   "PROPDMG"    "PROPDMGEXP"
## [6] "CROPDMG"    "CROPDMGEXP"
```

```
# Q1.2 Reviewing events that cause the most fatalities ( The Top-10 Fata
lities by Weather Event )


## Procedure = aggregate the top 10 fatalities by the event type and sor
t the output in descending order


Fatalities <- aggregate(FATALITIES ~ EVTYPE, data = strmdata, FUN = sum)
Top10_Fatalities <- Fatalities[order(-Fatalities$FATALITIES), ][1:10, ]
```

```
Top10_Fatalities
```

```
##              EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

```
# Q1.3 Reviewing events that cause the most injuries ( The Top-10 Injuri
es by Weather Event )


## Procedure = aggregate the top 10 injuries by the event type and sort
the output in descending order


Injuries <- aggregate(INJURIES ~ EVTYPE, data = strmdata, FUN = sum)
Top10_Injuries <- Injuries[order(-Injuries$INJURIES), ][1:10, ]
Top10_Injuries
```

```
##                 EVTYPE INJURIES
## 834            TORNADO    91346
## 856          TSTM WIND     6957
## 170              FLOOD     6789
## 130     EXCESSIVE HEAT     6525
## 464          LIGHTNING     5230
## 275               HEAT     2100
## 427          ICE STORM     1975
## 153        FLASH FLOOD     1777
```

```
## 760 THUNDERSTORM WIND      1488
## 244                HAIL      1361
```

```
# Q1.4 Plot of Top 10 Fatalities & Injuries for Weather Event Types ( Po
pulation Health Impact )

## Proecedure = plot graphs showing the top 10 fatalities and injuries

par(mfrow=c(1,2),mar=c(10,3,3,2))
barplot(Top10_Fatalities$FATALITIES,names.arg=Top10_Fatalities$EVTYPE,la
s=2,col="purple",ylab="fatalities",main="Top 10 fatalities")
barplot(Top10_Injuries$INJURIES,names.arg=Top10_Injuries$EVTYPE,las=2,co
l="purple",ylab="injuries",main="Top 10 Injuries")
```

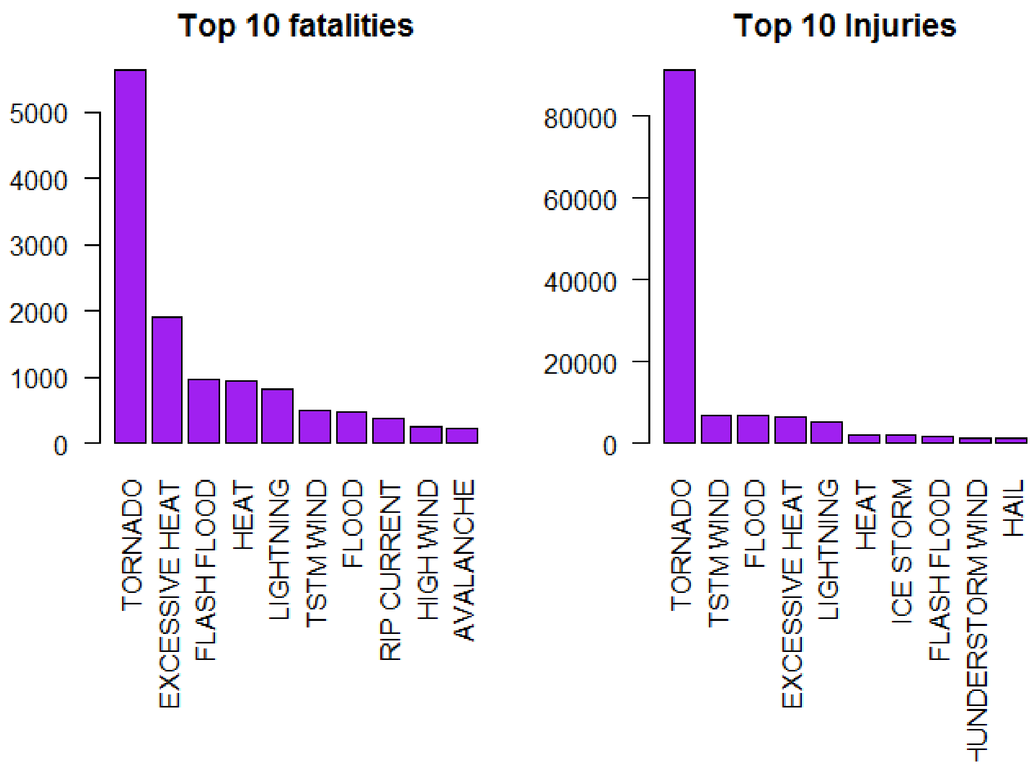

Figure 1-The weather event responsible for the highest fatalities and injuries is the 'Tornado'

```r
### QUESTION 2. Across the United States, which types of events have the greatest economic consequences?


# An analysis of the weather events responsible for the greatest economic consequences


## Hypothesis: Economic consequences means damages. The two significant types of damage typically caused by weather events include 'properties and crops'



# Q2.1 Data Exploration & Findings ...


# Upon reviewing the column names, the property damage(PROPDMG) and crop damage(CROPDMG) columns both have another related column titled 'exponents' (i.e - PROPDMGEXP and CROPDMGEXP respectively).

# As a result, let's convert the exponent columns into numeric data for the calculation of total property and crop damages encountered.


# Q2.2 Defining & Calcuating [ Property Damage ]


## Property damage exponents for each level listed out & assigned those values for the property exponent data.

## Invalid data was excluded by assigning the value as '0'.

## Then, the property damage value was calculated by multiplying the property damage and property exponent value.


unique(strmdata$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
```

```r
# Assigning values for the property exponent strmdata
strmdata$PROPEXP[strmdata$PROPDMGEXP == "K"] <- 1000
strmdata$PROPEXP[strmdata$PROPDMGEXP == "M"] <- 1e+06
strmdata$PROPEXP[strmdata$PROPDMGEXP == ""] <- 1
strmdata$PROPEXP[strmdata$PROPDMGEXP == "B"] <- 1e+09
```

```r
strmdata$PROPEXP[strmdata$PROPDMGEXP == "m"] <- 1e+06
strmdata$PROPEXP[strmdata$PROPDMGEXP == "0"] <- 1
strmdata$PROPEXP[strmdata$PROPDMGEXP == "5"] <- 1e+05
strmdata$PROPEXP[strmdata$PROPDMGEXP == "6"] <- 1e+06
strmdata$PROPEXP[strmdata$PROPDMGEXP == "4"] <- 10000
strmdata$PROPEXP[strmdata$PROPDMGEXP == "2"] <- 100
strmdata$PROPEXP[strmdata$PROPDMGEXP == "3"] <- 1000
strmdata$PROPEXP[strmdata$PROPDMGEXP == "h"] <- 100
strmdata$PROPEXP[strmdata$PROPDMGEXP == "7"] <- 1e+07
strmdata$PROPEXP[strmdata$PROPDMGEXP == "H"] <- 100
strmdata$PROPEXP[strmdata$PROPDMGEXP == "1"] <- 10
strmdata$PROPEXP[strmdata$PROPDMGEXP == "8"] <- 1e+08

# Assigning '0' to invalid exponent strmdata
strmdata$PROPEXP[strmdata$PROPDMGEXP == "+"] <- 0
strmdata$PROPEXP[strmdata$PROPDMGEXP == "-"] <- 0
strmdata$PROPEXP[strmdata$PROPDMGEXP == "?"] <- 0

# Calculating the property damage value
strmdata$PROPDMGVAL <- strmdata$PROPDMG * strmdata$PROPEXP



# Q2.3 Defining & Calcuating [ Crop Damage ]


## Crop damage exponents for each level listed out & assigned those valu
es for the crop exponent data.
## Invalid data was excluded by assigning the value as '0'.
## Then, the crop damage value was calculated by multiplying the crop da
mage and crop exponent value.


unique(strmdata$CROPDMGEXP)
```

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

```
# Assigning values for the crop exponent strmdata
strmdata$CROPEXP[strmdata$CROPDMGEXP == "M"] <- 1e+06
strmdata$CROPEXP[strmdata$CROPDMGEXP == "K"] <- 1000
strmdata$CROPEXP[strmdata$CROPDMGEXP == "m"] <- 1e+06
strmdata$CROPEXP[strmdata$CROPDMGEXP == "B"] <- 1e+09
strmdata$CROPEXP[strmdata$CROPDMGEXP == "0"] <- 1
strmdata$CROPEXP[strmdata$CROPDMGEXP == "k"] <- 1000
strmdata$CROPEXP[strmdata$CROPDMGEXP == "2"] <- 100
strmdata$CROPEXP[strmdata$CROPDMGEXP == ""] <- 1

# Assigning '0' to invalid exponent strmdata
strmdata$CROPEXP[strmdata$CROPDMGEXP == "?"] <- 0

# calculating the crop damage
strmdata$CROPDMGVAL <- strmdata$CROPDMG * strmdata$CROPEXP



# Q2.4 Property Damage Summary

## Procedure = aggregate the property damage by the event type and sort the output it in descending order

prop <- aggregate(PROPDMGVAL~EVTYPE,data=strmdata,FUN=sum,na.rm=TRUE)
prop <- prop[with(prop,order(-PROPDMGVAL)),]
prop <- head(prop,10)
print(prop)
```

```
##                EVTYPE   PROPDMGVAL
## 170             FLOOD 144657709807
## 411 HURRICANE/TYPHOON  69305840000
## 834           TORNADO  56947380617
## 670       STORM SURGE  43323536000
## 153       FLASH FLOOD  16822673979
## 244              HAIL  15735267513
## 402         HURRICANE  11868319010
## 848    TROPICAL STORM   7703890550
```

```
## 972       WINTER STORM   6688497251
## 359          HIGH WIND   5270046260
```

```
# Q2.5 Crop Damage Summary

## Procedure = aggregate the crop damage by the event type and sort the output it in descending order

crop <- aggregate(CROPDMGVAL~EVTYPE,data=strmdata,FUN=sum,na.rm=TRUE)
crop <- crop[with(crop,order(-CROPDMGVAL)),]
crop <- head(crop,10)
print(crop)
```

```
##                EVTYPE  CROPDMGVAL
## 95            DROUGHT 13972566000
## 170             FLOOD  5661968450
## 590       RIVER FLOOD  5029459000
## 427         ICE STORM  5022113500
## 244              HAIL  3025954473
## 402         HURRICANE  2741910000
## 411 HURRICANE/TYPHOON  2607872800
## 153       FLASH FLOOD  1421317100
## 140      EXTREME COLD  1292973000
## 212      FROST/FREEZE  1094086000
```

```
# Q2.6 Plot of Top 10 Property & Crop damages by Weather Event Types ( Economic Consequences )

##plot the graph showing the top 10 property and crop damages

par(mfrow=c(1,2),mar=c(11,3,3,2))
barplot(prop$PROPDMGVAL/(10^9),names.arg=prop$EVTYPE,las=2,col="gold",ylab="Prop.damage(billions)",main="Top10 Prop.Damages")
```

```
barplot(crop$CROPDMGVAL/(10^9),names.arg=crop$EVTYPE,las=2,col="gold",yl
ab="Crop damage(billions)",main="Top10 Crop.Damages")
```

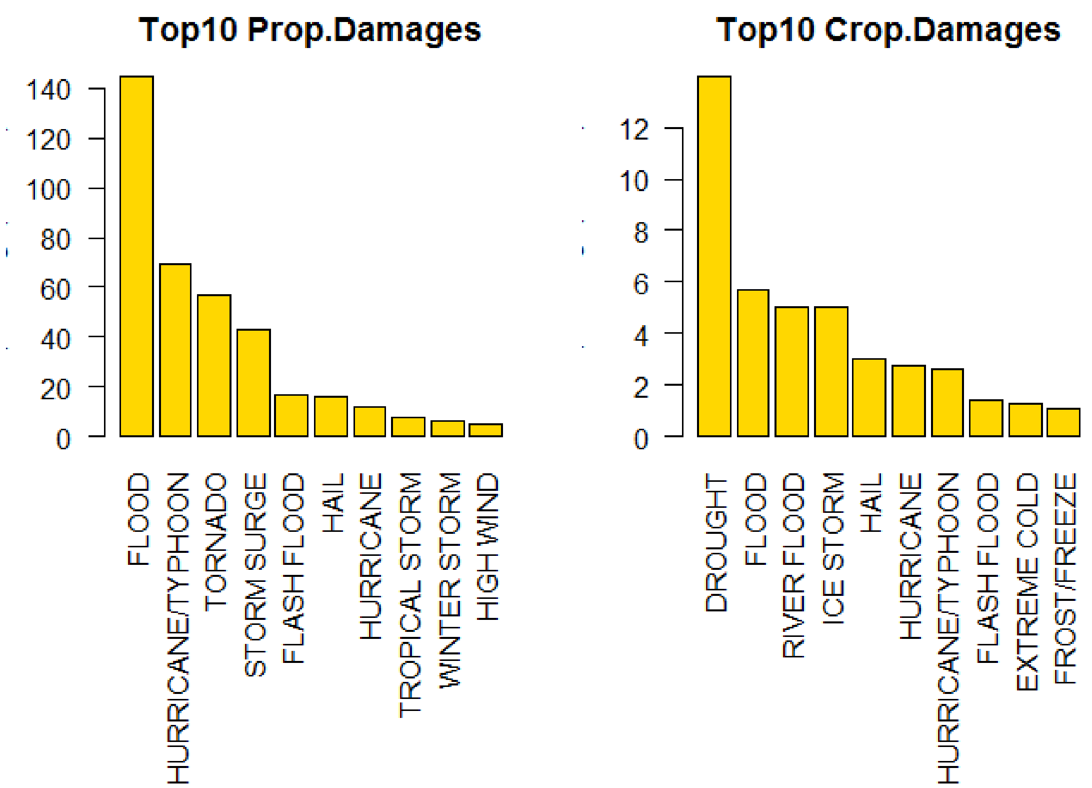

Figure 2-'Floods' are responsbile for the highest property damage while 'droughts' cause the greatest crop damage.

Summary of Conclusions

- Tornados are responsible for the maximum number of fatalities and injuries, followed by Excessive Heat for fatalities and Thunderstorm wind for injuries.
- Floods are responsbile for maximum property damage, while Droughts cause maximum crop damage. Second major events that caused the maximum damage was Hurricanes/Typhoos for property damage and Floods for crop damage.