Reproducible Research - Course Project 2

Health and Economic Consequences of Weather Events in United States

Synopsis

The report addresses questions regarding the most hurtful weather type events in terms of deaths, injuries, property and crop damage

Data Processing

Load needed packages

```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
##     date

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following object is masked from 'package:lubridate':
##
##     here

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 6523 6523 4242 11116 2224 2224 2260 383
##  $ BGN_TIME  : Factor w/ 3608 levels "00:00:00 AM",..: 272 287 2705 1683 2584 3186 242 1683 3186 318
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",..: 13513 1
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","  Christiansburg",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","  0900CST",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","  CANTON","  TULIA",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ 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 ...
##  $ WFO       : Factor w/ 542 levels "","  CI","%SD",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels "","
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : Factor w/ 436781 levels "","\t","\t\t",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...

## # A tibble: 6 x 37
##   STATE__ BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE EVTYPE
##     <dbl> <fct>    <fct>    <fct>      <dbl> <fct>      <fct> <fct>
## 1       1 4/18/19~ 0130     CST           97 MOBILE     AL    TORNA~
## 2       1 4/18/19~ 0145     CST            3 BALDWIN    AL    TORNA~
## 3       1 2/20/19~ 1600     CST           57 FAYETTE    AL    TORNA~
```

```
## 4       1 6/8/195~ 0900     CST           89 MADISON    AL    TORNA~
## 5       1 11/15/1~ 1500     CST           43 CULLMAN    AL    TORNA~
## 6       1 11/15/1~ 2000     CST           77 LAUDERDALE AL    TORNA~
## # ... with 29 more variables: BGN_RANGE <dbl>, BGN_AZI <fct>,
## #   BGN_LOCATI <fct>, END_DATE <fct>, END_TIME <fct>, COUNTY_END <dbl>,
## #   COUNTYENDN <lgl>, END_RANGE <dbl>, END_AZI <fct>, END_LOCATI <fct>,
## #   LENGTH <dbl>, WIDTH <dbl>, F <int>, MAG <dbl>, FATALITIES <dbl>,
## #   INJURIES <dbl>, PROPDMG <dbl>, PROPDMGEXP <fct>, CROPDMG <dbl>,
## #   CROPDMGEXP <fct>, WFO <fct>, STATEOFFIC <fct>, ZONENAMES <fct>,
## #   LATITUDE <dbl>, LONGITUDE <dbl>, LATITUDE_E <dbl>, LONGITUDE_ <dbl>,
## #   REMARKS <fct>, REFNUM <dbl>
```

Questions

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

```r
# Subset the data using only the values needed
var2use<-c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
storm<-storm[var2use]
str(storm)
```

```
## 'data.frame':    902297 obs. of  7 variables:
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 834 834 834 834 834 8
##  $ 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 ...
```

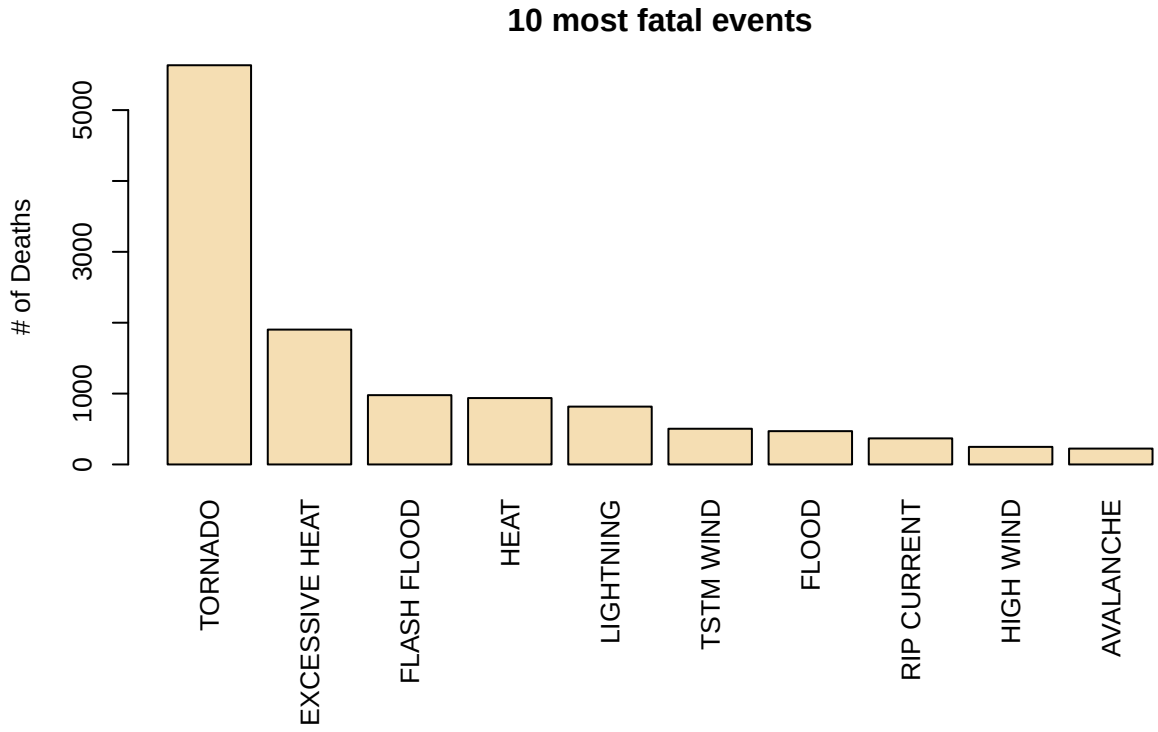
```r
head(tbl_df(storm))
```

```
## # A tibble: 6 x 7
##   EVTYPE  FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
##   <fct>        <dbl>    <dbl>   <dbl> <fct>        <dbl> <fct>
## 1 TORNADO          0       15    25   K                0 ""
## 2 TORNADO          0        0     2.5 K                0 ""
## 3 TORNADO          0        2    25   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 ""
```

```r
# For this question I am going to find which EVTYPE leads to more FATALITES and INJURIES and present th
total_deaths<-aggregate(FATALITIES ~ EVTYPE, data=storm, FUN = sum)

total_injuries<-aggregate(INJURIES ~ EVTYPE, data=storm, FUN = sum)

total_deaths_sorted<- total_deaths %>% arrange(-total_deaths$FATALITIES)

total_injuries_sorted<- total_injuries %>% arrange(-total_injuries$INJURIES)

most_fatal_events<-head(total_deaths_sorted,10)
most_fatal_events
```

```
##            EVTYPE FATALITIES
## 1         TORNADO       5633
## 2  EXCESSIVE HEAT       1903
## 3     FLASH FLOOD        978
## 4            HEAT        937
## 5       LIGHTNING        816
## 6       TSTM WIND        504
## 7           FLOOD        470
## 8     RIP CURRENT        368
## 9       HIGH WIND        248
## 10      AVALANCHE        224
```

```
par(mfrow = c(1,1), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(most_fatal_events$FATALITIES, names.arg = most_fatal_events$EVTYPE,las = 3,col="wheat",main="10
```




```
dev.copy(png, "fatal-events.png", width = 480, height = 480)
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

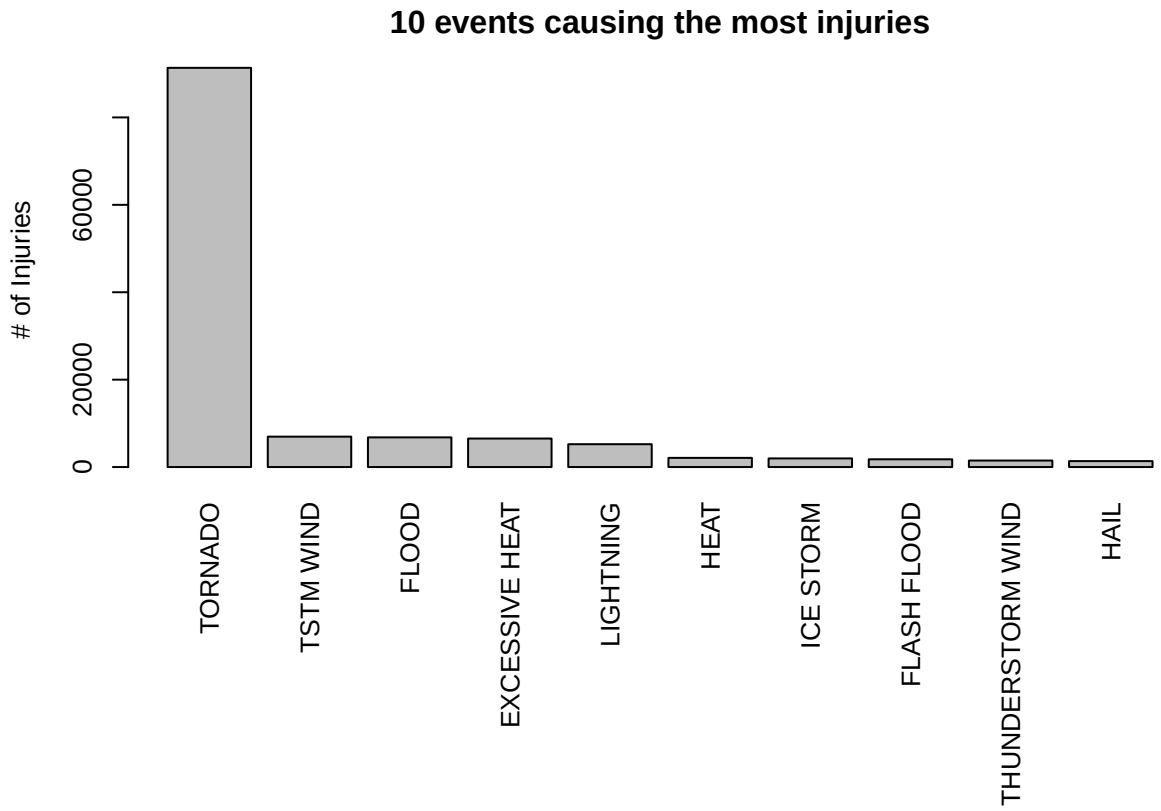
```
## pdf
##   2
```

```
most_injuries_events<-head(total_injuries_sorted,10)
most_injuries_events
```

```
##              EVTYPE INJURIES
## 1           TORNADO    91346
## 2         TSTM WIND     6957
## 3             FLOOD     6789
## 4    EXCESSIVE HEAT     6525
```

```
## 5           LIGHTNING     5230
## 6                HEAT     2100
## 7           ICE STORM     1975
## 8         FLASH FLOOD     1777
## 9  THUNDERSTORM WIND     1488
## 10              HAIL     1361
```

```
par(mfrow = c(1,1), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(most_injuries_events$INJURIES, names.arg = most_injuries_events$EVTYPE,las = 3,col="grey", main=
```




```
dev.copy(png, "injuries-events.png", width = 480, height = 480)
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

```
## pdf
##   2
```

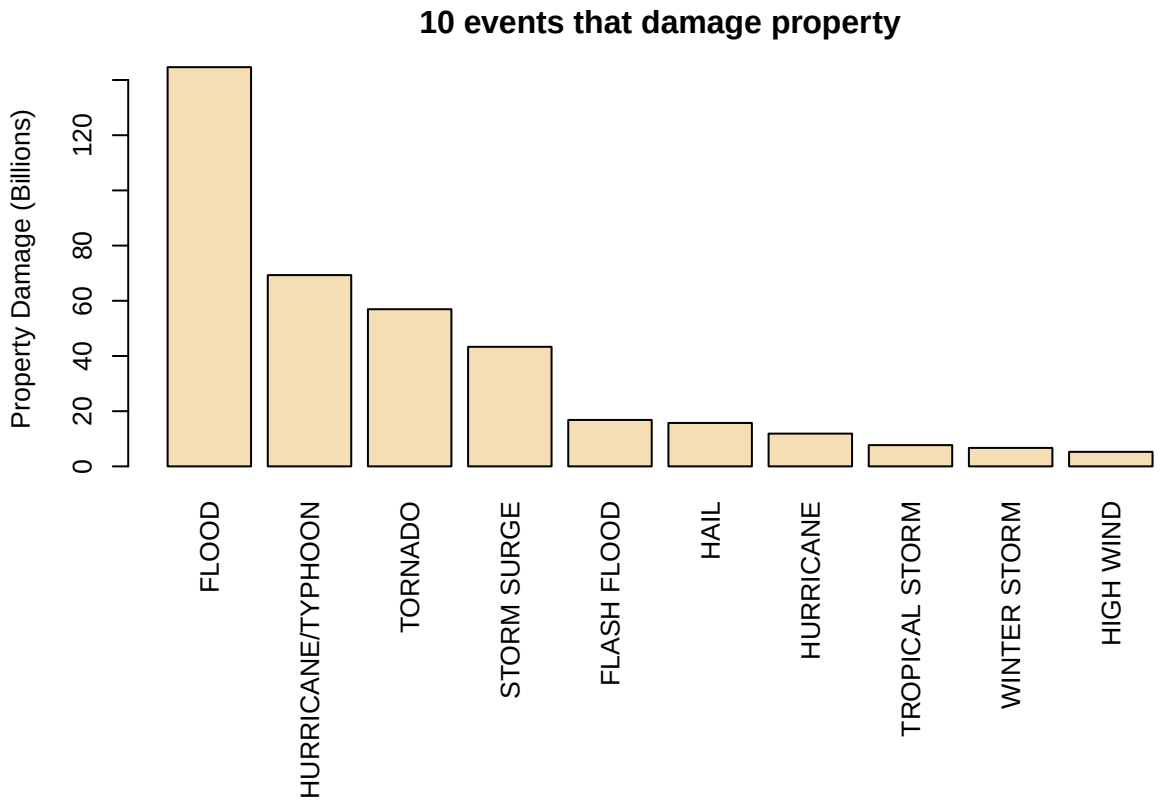
2. Across the United States, which types of events have the greatest economic consequences?

```
# For this question I am going to find which EVTYPE lead to more property and crop damage (PROPDMGTOTAL

#first I need to create a total variable for property and crop damage
# PROPDMGTOTAL=PROPDMG+PROPDMGEXP
storm$PROPDMGEXP <- mapvalues(storm$PROPDMGEXP, from = c("K", "M","", "B", "m", "+", "0", "5", "6", "?"
storm$PROPDMGEXP <- as.numeric(as.character(storm$PROPDMGEXP))
storm$PROPDMGTOTAL <- (storm$PROPDMG * storm$PROPDMGEXP)/1000000000
```

```r
prop_damage<-aggregate(PROPDMGTOTAL ~ EVTYPE, data=storm, FUN = sum)

#CROPDMGTOTAL=CROPDMG+CROPDMGEXP
storm$CROPDMGEXP <- mapvalues(storm$CROPDMGEXP, from = c("","M", "K", "m", "B", "?", "0", "k","2"), to =
storm$CROPDMGEXP <- as.numeric(as.character(storm$CROPDMGEXP))
storm$CROPDMGTOTAL <- (storm$CROPDMG * storm$CROPDMGEXP)/1000000000

crop_damage<-aggregate(CROPDMGTOTAL ~ EVTYPE, data=storm, FUN = sum)

total_prop_damage<- prop_damage %>% arrange(-prop_damage$PROPDMGTOTAL)

total_crop_damage<- crop_damage %>% arrange(-crop_damage$CROPDMGTOTAL)

most_prop_damage_events<-head(total_prop_damage,10)
most_prop_damage_events
```

```
##               EVTYPE PROPDMGTOTAL
## 1              FLOOD   144.657710
## 2  HURRICANE/TYPHOON    69.305840
## 3            TORNADO    56.947381
## 4        STORM SURGE    43.323536
## 5        FLASH FLOOD    16.822674
## 6               HAIL    15.735268
## 7          HURRICANE    11.868319
## 8     TROPICAL STORM     7.703891
## 9       WINTER STORM     6.688497
## 10         HIGH WIND     5.270046
```

```r
par(mfrow = c(1,1), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(most_prop_damage_events$PROPDMGTOTAL, names.arg = most_prop_damage_events$EVTYPE,las = 3,col="wh
```




```
dev.copy(png, "prop_damage.png", width = 480, height = 480)
```

```
## quartz_off_screen
##                 3
```

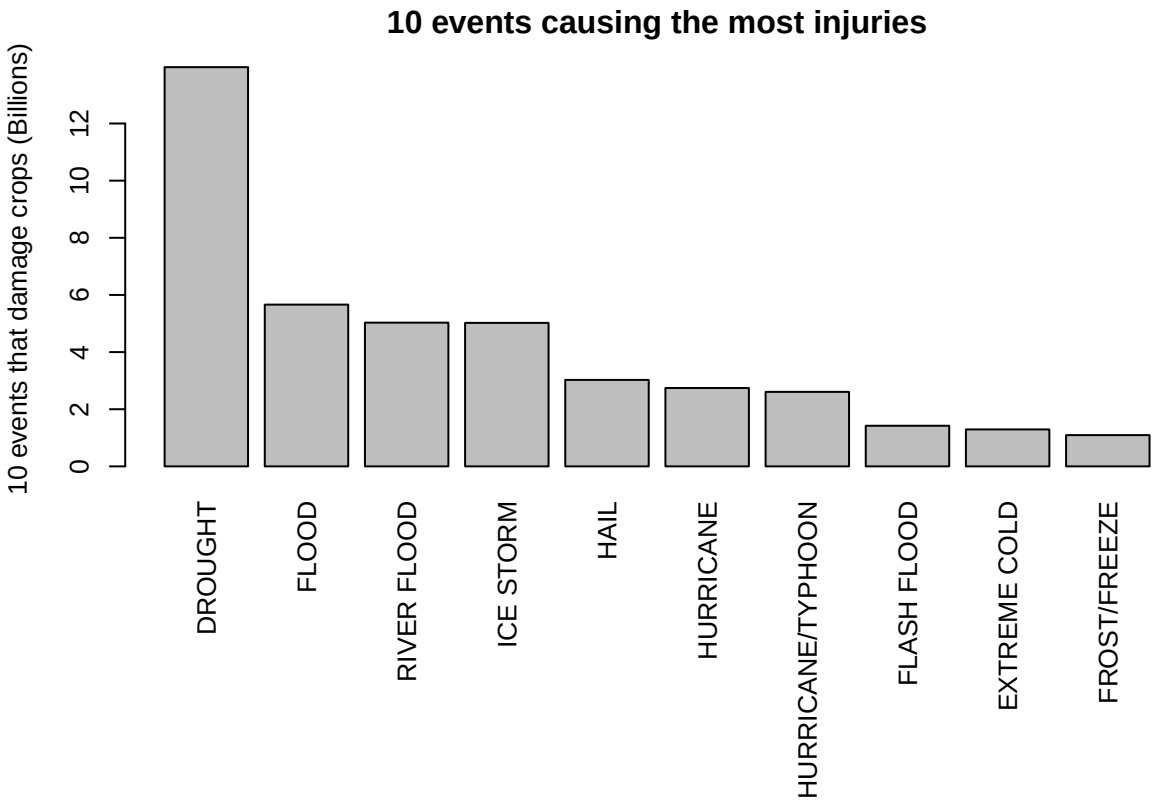
```
dev.off()
```

```
## pdf
##   2
```

```
most_crop_damage_events<-head(total_crop_damage,10)
most_crop_damage_events
```

```
##               EVTYPE CROPDMGTOTAL
## 1            DROUGHT    13.972566
## 2              FLOOD     5.661968
## 3        RIVER FLOOD     5.029459
## 4          ICE STORM     5.022113
## 5               HAIL     3.025954
## 6          HURRICANE     2.741910
## 7  HURRICANE/TYPHOON     2.607873
## 8        FLASH FLOOD     1.421317
## 9       EXTREME COLD     1.292973
## 10      FROST/FREEZE     1.094086
```

```
par(mfrow = c(1,1), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(most_crop_damage_events$CROPDMGTOTAL, names.arg = most_crop_damage_events$EVTYPE,las = 3,col="g
```




```
dev.copy(png, "crop_damage.png", width = 480, height = 480)
```

```
## quartz_off_screen
##                 3
```

```
dev.off()
```

```
## pdf
##   2
```

Results

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

The analysis and graphs indicate the most harmful type of weather phenomena with respect to death are *Tornados* Excessive Heat *Flash Flood while in terms of injuries are* Tornados *TSTM Wind* Flood

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

The most harmful events in terms of property damage are *Flood* Hurricanes/Typhoons *Tornados while in terms of crop damage are* Drought *Flood* River Flood