Coursera Reproducible Research: Course Project 2

Pranay Sarkar

7th September, 2017

Health/Economic Consequences in U.S. caused by Storms ans Weather Events

GitHub: GitHub Link

1 - Introduction

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

This Data Science 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.

Look to link: Storm Data Documentation. At pag.51 of this document, there is an example, evolving Hurricane Andrew where the powerful winds resulted in 4 fatalities, 50 injuries, $13B in property damage and $ 750M in crop damage (Notation Example pag.51: FLZ018-021 >023 24 0325EST 4 50 13B 750M Hurricane/Typhoon 0900EST)

2 - Synopsis

This report address questions related to Weather Events and Storms in U.S. that are most damaging in terms of Fatalities, Injuries and damages to properties and crop.

The two mainly questions to be answered are:

1 - Which types of events are most harmful with respect to population health?

2 - Which types of events have the greatest economic consequences?

3 - Loading the Data

3.1 - R libraries

```r
getwd()
```

```
## [1] "/home/psarkar/Documents/RepData_PeerAssessment2-master"
```

```r
library(knitr)
library(markdown)
library(rmarkdown)
```

```r
library(plyr)
library(stats)
```

3.2 - Loading NOAA data into R

```r
storm <- read.csv(file = "repdata-data-StormData.csv", header = TRUE, sep = ",")
dim(storm)
```

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

```r
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"
```

```r
str(storm)
```

```
## '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 "10/10/1954 0:00:00",..: 6523 6523 4213 11116 1426 1426 1462 28
##  $ BGN_TIME  : Factor w/ 3608 levels "000","0000","00:00:00 AM",..: 212 257 2645 1563 2524 3126 122 1
##  $ 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 "?","ABNORMALLY DRY",..: 830 830 830 830 830 830 830 830 830 830
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","E","Eas","EE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","?","(01R)AFB GNRY RNG AL",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","10/10/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","?","0000",..: 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 "","(0E4)PAYSON ARPT",..: 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 "","2","43","9V9",..: 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 "","  ","  ","   ",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

3.3 - Wrangling the Data

Defining variables that will be used:

- EVTYPE: Event Type (Tornados, Flood, ….)
- FATALITIES: Number of Fatalities
- INJURIES: Number of Injuries
- PROGDMG: Property Damage
- PROPDMGEXP: Units for Property Damage (magnitudes - K,B,M)
- CROPDMG: Crop Damage
- CROPDMGEXP: Units for Crop Damage (magnitudes - K,BM,B)

```
varsNedeed <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP")
storm <- storm[varsNedeed]
dim(storm)
```

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

```
names(storm)
```

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

```
str(storm)
```

```
## 'data.frame':    902297 obs. of  7 variables:
##  $ EVTYPE    : Factor w/ 985 levels "?","ABNORMALLY DRY",..: 830 830 830 830 830 830 830 830 830 830
##  $ 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 ...
```

3.3.1 - Calculating Total for Property Damage

- Refactor of variable PROPDNGEXP

```
unique(storm$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
```

```
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
```

- Refactor of variable CROPDMGEXP variable

```
unique(storm$CROPDMGEXP)
```

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

```
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
```

4 - Processing the Data:

Lets answer the question about Which Types of Events are most Harmful for population HEALTH? The variables involved are FATALITIES and INJURIES.

4.1 - Events are most harmful to population Health?

The item 2.6 (page 9) of National Weather Service Storm Data documentation describes about Fatalities and Injuries. So, it is necessary to assess these Variables to define which of EVENTS (EVTYPE) are most harmful. Look to link: Storm Data Documentation

4.1.1 - Total Number of Fatalities per Event

```
sumFatalities <- aggregate(FATALITIES ~ EVTYPE, data = storm,  FUN="sum")
dim(sumFatalities)  ## 985 observations
```

```
## [1] 985   2
```

- Ordering Number of Fatalities and defining the top 10 Weather events

```
fatalities10events <- sumFatalities[order(-sumFatalities$FATALITIES), ][1:10, ]
dim(fatalities10events)
```

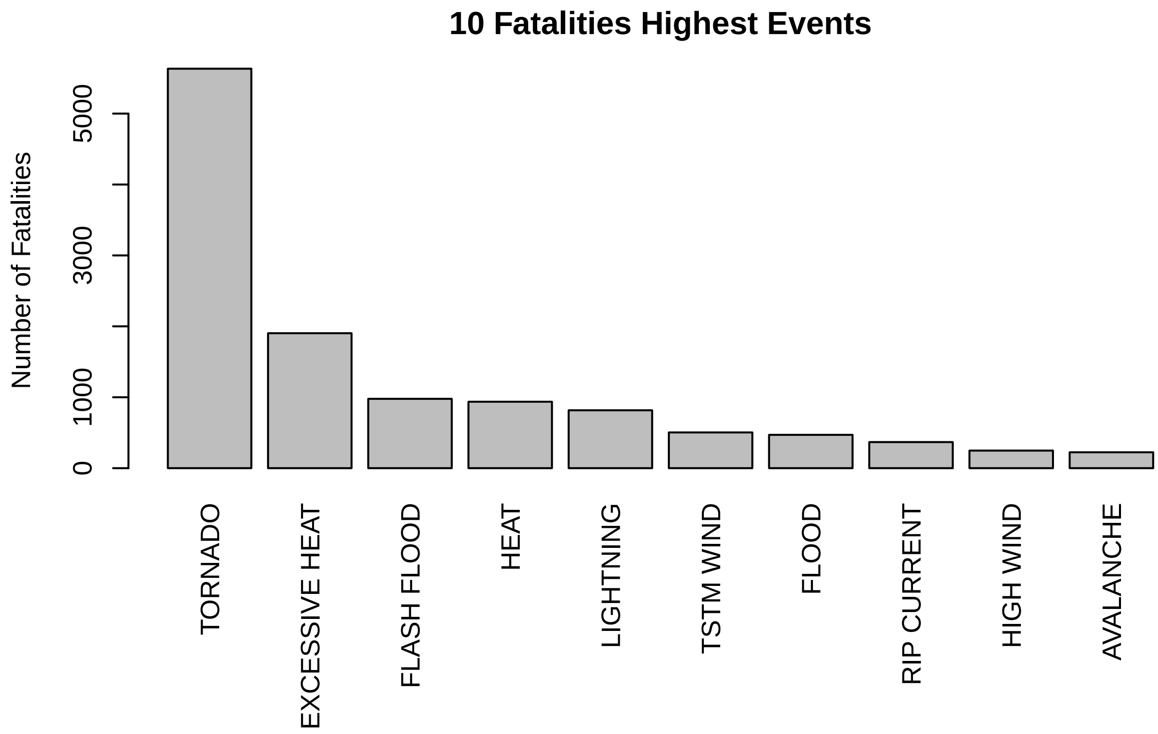
```
## [1] 10  2
```

```
fatalities10events
```

```
##             EVTYPE FATALITIES
## 830        TORNADO       5633
## 123 EXCESSIVE HEAT       1903
## 147    FLASH FLOOD        978
## 269           HEAT        937
## 452      LIGHTNING        816
## 854      TSTM WIND        504
## 164          FLOOD        470
## 581    RIP CURRENT        368
## 354      HIGH WIND        248
## 11       AVALANCHE        224
```

- BarPlot of the 10 Fatalities Events most harmful to population Health.

```
par(mfrow = c(1,1), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(fatalities10events$FATALITIES, names.arg = fatalities10events$EVTYPE, las = 3, main = "10 Fatal
```




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

```
## png
##   3
```

```r
dev.off()
```

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

4.1.2 - Total Number of Injuries per Event

Using the same reasoning of Fatalities, let's evaluate the Number of Injuries per type of Event (EVTYPE)

```r
sumInjuries <- aggregate(INJURIES ~ EVTYPE, data = storm,  FUN="sum")
dim(sumInjuries)  ## 985 observations
```

```
## [1] 985   2
```

- Ordering Number of INJURIES and defining the top 10 Weather events in this category

```r
injuries10events <- sumInjuries[order(-sumInjuries$INJURIES), ][1:10, ]
dim(injuries10events)
```

```
## [1] 10  2
```

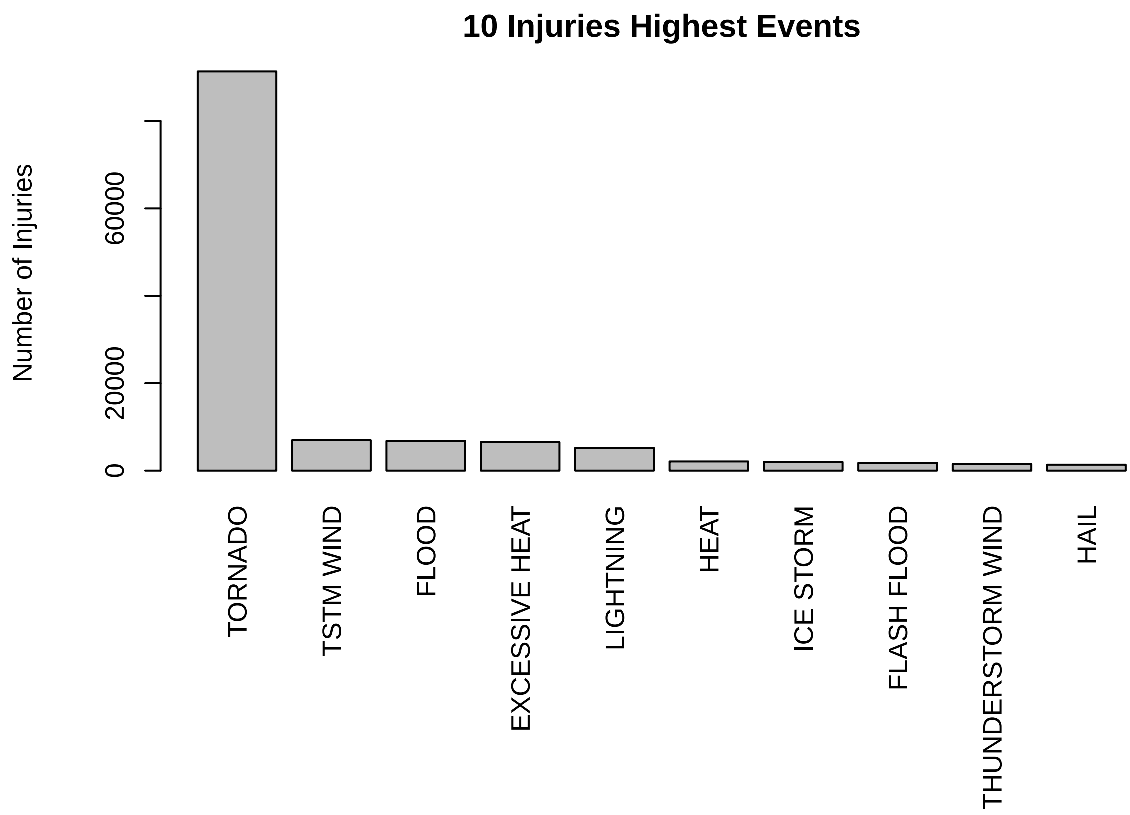
```r
injuries10events
```

```
##                  EVTYPE INJURIES
```

```
## 830          TORNADO    91346
## 854        TSTM WIND     6957
## 164            FLOOD     6789
## 123   EXCESSIVE HEAT     6525
## 452        LIGHTNING     5230
## 269             HEAT     2100
## 424        ICE STORM     1975
## 147      FLASH FLOOD     1777
## 759 THUNDERSTORM WIND     1488
## 238             HAIL     1361
```

- BarPlot of the 10 INJURIES Events most harmful to population Health.

```
par(mfrow = c(1,1), mar = c(12, 6, 3, 2), mgp = c(4, 1, 0), cex = 0.8)
barplot(injuries10events$INJURIES, names.arg = injuries10events$EVTYPE, las = 3, main = "10 Injuries Hi
```




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

```
## png
##   3
```

```
dev.off()
```

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

4.2 - Which type of Events have the greatest Economic consequences?

The item 2.7 (page 12 and the APPENDIX B) of National Weather Service Storm Data documentation describes about Damage. Two variables, PROPDMG (for Property Damage) and CROPDMG (for Crop Damage) are used to represent these losts. If you want, read more about theses damages, please connect with National Weather Service using the link Storm Data Documentation

4.2.1 - Property Damage

- Calculating Property Damage for type of Event

```
sumPropertyDamage <- aggregate(PROPDMGTOTAL ~ EVTYPE, data = storm,  FUN="sum")
dim(sumPropertyDamage)  ## 985 observations
```

```
## [1] 985   2
```

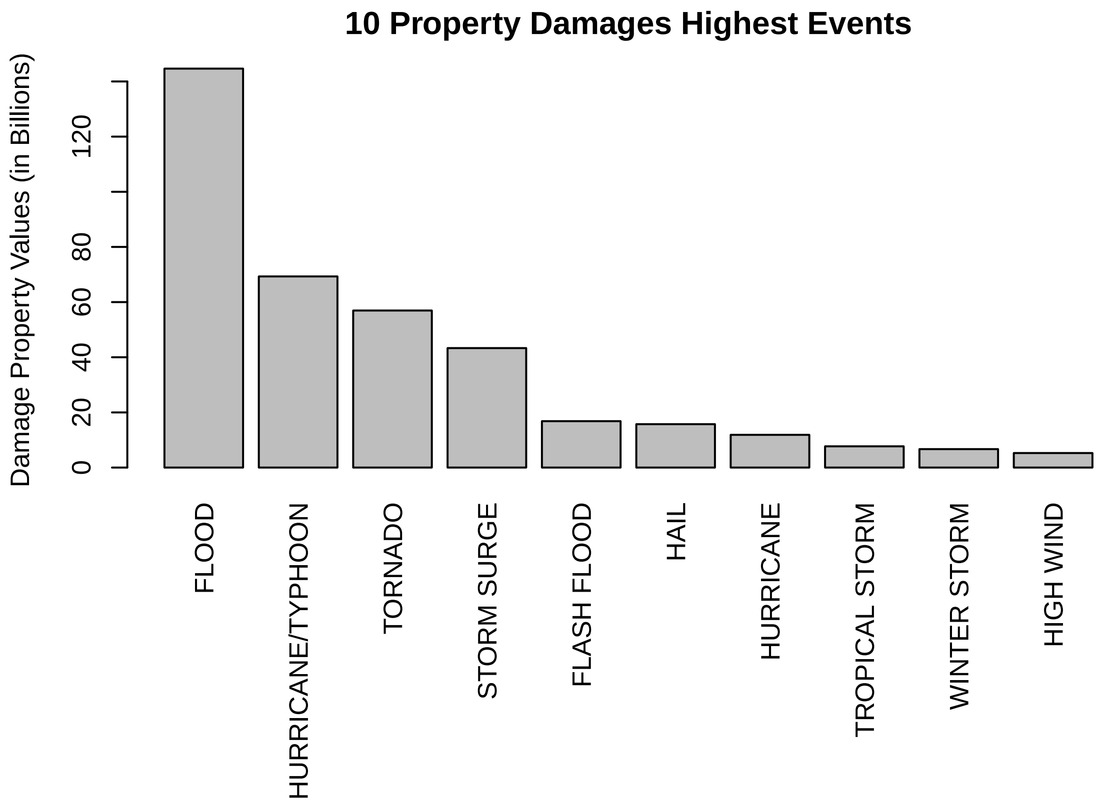
- We have 985 observations, which is a great number of Events to present in a Plot.
- Lets stay with the 10 most Property Damage Events

```
propdmg10Total <- sumPropertyDamage[order(-sumPropertyDamage$PROPDMGTOTAL), ][1:10, ]
propdmg10Total
```

```
##                EVTYPE PROPDMGTOTAL
## 164             FLOOD   144.657710
## 406 HURRICANE/TYPHOON    69.305840
## 830           TORNADO    56.947381
## 666       STORM SURGE    43.323536
## 147       FLASH FLOOD    16.822674
## 238              HAIL    15.735268
## 397         HURRICANE    11.868319
## 844    TROPICAL STORM     7.703891
## 972      WINTER STORM     6.688497
## 354         HIGH WIND     5.270046
```

- BarPlot of the 10 Proprietary Damage Events most harmful to population economic.

```
par(mfrow = c(1,1), mar = c(12, 6, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(propdmg10Total$PROPDMGTOTAL, names.arg = propdmg10Total$EVTYPE, las = 3, main = "10 Property Da
```




```
dev.copy(png, "propdmg-total.png", width = 480, height = 480)
```

```
## png
##   3
```

```
dev.off()
```

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

4.2.2 - Crop Damage

- Calculating Crop Damage for type of Event

```
sumCropDamage <- aggregate(CROPDMGTOTAL ~ EVTYPE, data = storm,  FUN="sum")
dim(sumCropDamage) ## 985 observations
```

```
## [1] 985   2
```

- We have 985 observations, which is a great number of Events to present in a Plot.
- Lets stay with the 10 most Crop Damage Events

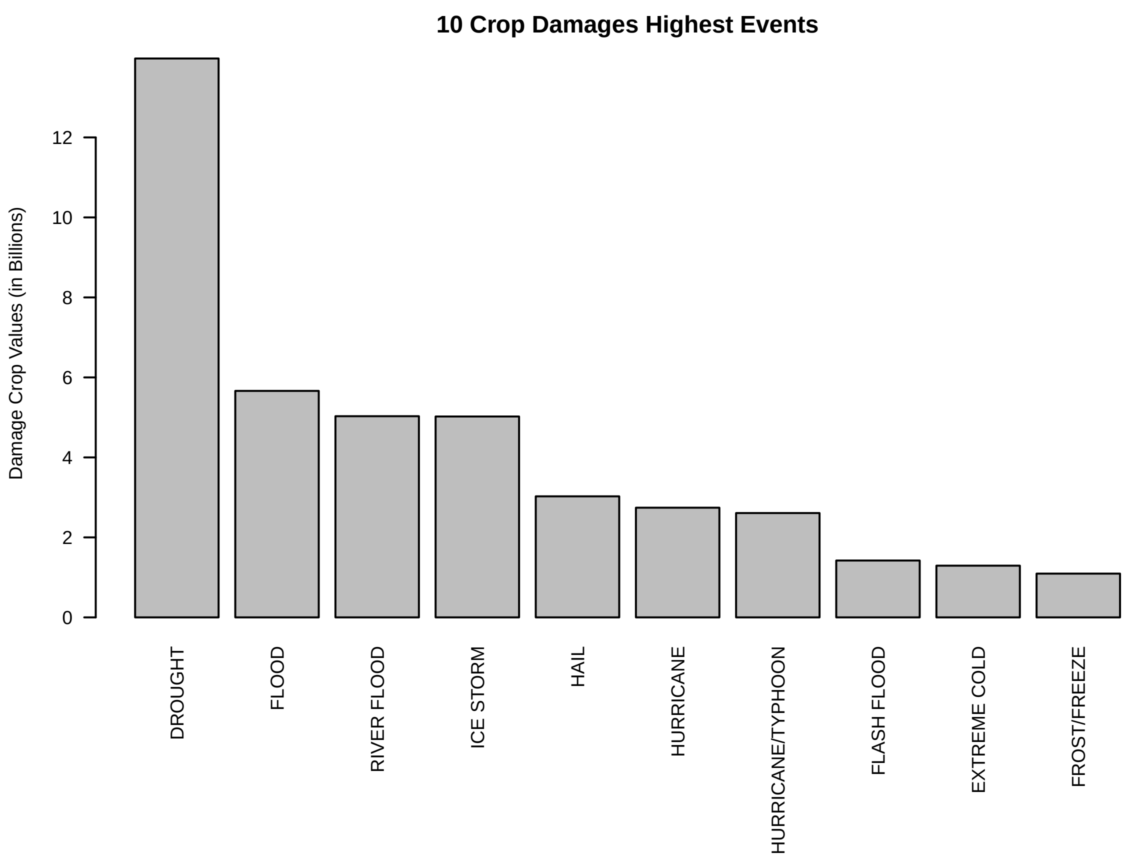
```
cropdmg10Total <- sumCropDamage[order(-sumCropDamage$CROPDMGTOTAL), ][1:10, ]
cropdmg10Total
```

```
##                EVTYPE CROPDMGTOTAL
## 88            DROUGHT    13.972566
## 164             FLOOD     5.661968
## 586       RIVER FLOOD     5.029459
```

```
## 424         ICE STORM     5.022113
## 238              HAIL     3.025954
## 397         HURRICANE     2.741910
## 406 HURRICANE/TYPHOON     2.607873
## 147       FLASH FLOOD     1.421317
## 133      EXTREME COLD     1.292973
## 206      FROST/FREEZE     1.094086
```

- BarPlot of the 10 Crop Damage Events most harmful to population economic.

```r
par(mfrow = c(1,1), mar = c(10, 6, 3, 2), mgp = c(3, 1, 0), cex = 0.6)
barplot(cropdmg10Total$CROPDMGTOTAL, names.arg = cropdmg10Total$EVTYPE, las = 2, main = "10 Crop Damage
```




```r
dev.copy(png, "cropdmg-total.png", width = 480, height = 480)
```

```
## png
##   3
```

```r
dev.off()
```

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

5 - Results and Conclusions

5.1 - Answering Question 1

As demonstrated by the Graphs, Tornados causes the greatest number of Fatalities and Injuries.

Specifically in FATALITIES, after Tornados, EXCESSIVE HEAT, FLASH FLOOD and HEAT are the next ones.

Specifically in INJURIES, after tornados we have TSTM WIND, FLOOD and EXCESSIVE HEAT.

5.2 - Answering Question 2

Floods are the Weather Event that cause most Property Damage, followed by Hurrucanes.

Drought are the Weather Event that causes most Crop damages, follwed by Flood.

5.3 - Conclusions

Based on evidences demonstrated earlier - tornados, floods and droughts have priorities to minize the impact in human and economic costs of Weather Events. Government and society always have to be alert and prepared for each type of events. For safety, it's important for the population to know what to do during these kind of events.